Pacific Journal of Mathematics

Volume 244 No. 1 January 2010

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982) F. Wolf (1904–1989)

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

Regular subscription rate for 2009: US$450 a year (10 issues). Special rate: US$225 a year to individual members of supporting institutions. 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.

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. 244, No. 1, 2010

A FREE BOUNDARY ISOPERIMETRIC PROBLEM IN HYPERBOLIC 3-SPACE BETWEEN PARALLEL HOROSPHERES

ROSA MARIA BARREIRO CHAVES,
MÁRCIO F. DA SILVA AND RENATO H. L. PEDROSA

We investigate the isoperimetric problem of finding the regions of prescribed volume with minimal boundary area between two parallel horospheres in hyperbolic 3-space (the part of the boundary contained in the horospheres is not included). We reduce the problem to the study of rotationally invariant regions and obtain the possible isoperimetric solutions by studying the behavior of the profile curves of the rotational surfaces with constant mean curvature in hyperbolic 3-space. We also classify all the connected compact rotational surfaces M of constant mean curvature that are contained in the region between two horospheres, have boundary ∂M either empty or lying on the horospheres, and meet the horospheres perpendicularly along their boundary.

1. Introduction

Geometric isoperimetric problems, (upper) estimates for the volume of regions of a given fixed boundary volume, and the dual problems play an important role in analysis and geometry. There are both isoperimetric inequalities, common in analysis, and actual classification of optimal geometric objects, like the round ball in Euclidean geometry. We will be interested in the study of a relative free-boundary isoperimetric problem in hyperbolic 3-space between two parallel horospheres. A survey of recent results in isoperimetric problems is [Ritoré and Ros 2002].

For a Riemannian manifold M^n, we state the classical isoperimetric problem as follows: Classify, up to congruency by the isometry group of M, the (compact) regions $\Omega \subseteq M$ enclosing a fixed volume that have minimal boundary volume. The relevant concepts of volume are those of geometric measure theory: Regions and their boundaries are respectively n- and $(n-1)$-rectifiable subsets of M; see [Morgan 2009].

MSC2000: 53A10, 49Q10.
Keywords: constant mean curvature surfaces, hyperbolic space, isoperimetric problem.
Da Silva thanks CAPES and CNPq for financial support during his PhD studies, which led in part to this paper.

If M has a boundary, the part of $\partial\Omega$ included in the interior of M will be called the *free boundary* of Ω, and the other part will be called the *fixed boundary*. One may specify how the fixed boundary of Ω is included in the computation of the boundary volume functional. In this paper, we will not consider the volume of the fixed boundary of Ω to be part of the boundary volume functional. We will see in Section 3 that this implies that the angle of contact between the interior boundary of Ω and ∂M is $\pi/2$ (when this contact occurs). Such problems are related to the geometry of stable drops in capillarity (the angle of contact depends, as mentioned, on how one considers the volume of the fixed boundary in the computation of the boundary volume functional). For a discussion, see [Finn 1986].

This work is motivated by the well-known results of Athanassenas [1987] and Vogel [1987], which imply that between two parallel planes in Euclidean space $\mathbb{R}^3$, a (stable) soap bubble touching both walls perpendicularly is a straight cylinder orthogonal to the planes, and may only exist down to a certain minimal enclosing volume depending on the distance between the planes. Below that value, only half-spheres touching one of the planes or whole spheres not touching either plane occur, and the cylinders become unstable. A new proof of this fact can be found in [Pedrosa and Ritoré 1999], where the authors study the analogous problem in higher-dimensional Euclidean spaces.

In this paper we study the analogous relative isoperimetric problem between two parallel horospheres in the hyperbolic space $\mathbb{H}^3(-1)$. We will use the upper half-space model $\mathbb{R}^3_+$, in which parallel horospheres are represented by horizontal Euclidean 2-planes of $\mathbb{R}^3_+$. We will classify the possible isoperimetric solutions.

The existence of isoperimetric regions in the manifold with boundary (B, g), the slab composed of the two horospheres and the region between them, may be obtained by adapting a result of Morgan [2009] (applicable since B/G is a compact space, where G is the subgroup of the isometry group of $\mathbb{H}^3(-1)$ leaving B invariant). Regarding the regularity of the free boundary, well-known results about the lower codimension bounds of the singular subset imply that it must be regular, and in fact analytic.

In Section 2, we define basic notions in the model $\mathbb{R}^3_+$, like geodesics, totally geodesic surfaces, umbilical surfaces and rotational surfaces. We also use the area and volume functionals to more precisely formulate the isoperimetric problem.

In Section 3 we get some basic geometric properties of isoperimetric regions; for instance, their (free) boundaries must have constant mean curvature and, when they touch the bounding horospheres, the contact angle must be $\pi/2$. We also discuss their rotational invariance.

In Section 4 we investigate the tangency of profile curves for the rotational surfaces with constant mean curvature, to determine the possible isoperimetric regions between the two parallel horospheres. We discuss in detail the existence of isoperimetric regions and the regularity of the free boundary part, and we prove the following result:

Theorem 1.1. *Let c_1, c_2 be positive real constants such that $c_1 < c_2$, and let $\mathcal{F}_{c_1,c_2} = \{(x, y, z) \in \mathbb{R}^3_+ : c_1 \le z \le c_2\}$. Let $V > 0$, and let $\mathcal{C}_{c_1,c_2,V}$ be the set of $\Omega \subset \mathcal{F}_{c_1,c_2}$ with volume $|\Omega| = V$ and boundary volume (area) $A(\Omega \cap \mathring{\mathcal{F}}_{c_1,c_2}) < \infty$, where we suppose that Ω is connected, compact and 3-rectifiable in $\mathcal{F}_{c_1,c_2}$, and has as boundary (between the horospheres) an embedded, orientable, 2-rectifiable surface. Let*

$$A_{c_1,c_2,V} = \inf\{A(\Omega \cap \mathring{\mathcal{F}}_{c_1,c_2}) : \Omega \in \mathcal{C}_{c_1,c_2,V}\}.$$

(1) *There exists $\Omega \in \mathcal{C}_{c_1,c_2,V}$ such that $A(\Omega \cap \mathring{\mathcal{F}}_{c_1,c_2}) = A_{c_1,c_2,V}$. The free boundaries are analytic surfaces.*

(2) *If Ω has minimal boundary volume between the horospheres, the free boundary of Ω is either*
   (a) *of catenoid cousin type or umbilical with $H = 1$,*
   (b) *of equidistant type or umbilical with $0 < H < 1$, or*
   (c) *of onduloid type or umbilical with $H > 1$.*

Remark 1.2. We give details of this description in Section 4. The hyperbolic distance $d = \ln(c_2/c_1)$ between the horospheres could determine which region among cases (a)–(c) is the isoperimetric solution. It is still not clear, however, which from among (a)–(c) would be solutions for a given d. (In [Athanassenas 1987], the classification of isoperimetric solutions depending on d is fully answered for the analogous problem in $\mathbb{R}^3$.) In some cases, as in Figure 1, we know by fixing the lower horosphere at $z = 1/2$ that umbilical surfaces with $H = 1$ cannot be solutions when the upper horosphere is at level $z < 1$. In the general case, the question is still open because it is necessary to study the stability of the surfaces (a)–(c) (see [Barbosa et al. 1988] for the notion of stability in this context).

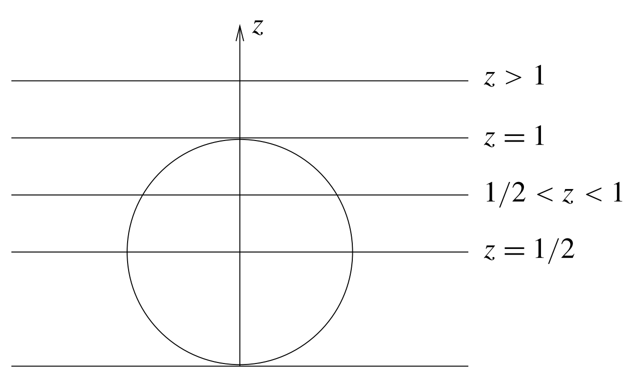

Figure 1. A case in which it is possible to decide.

Remark 1.3. Theorem 1.1 shows how the situation in hyperbolic geometry differs from that in Euclidean 3-space. In $\mathbb{R}^3$, we also have rotationally invariant surfaces of catenoid and onduloid type, but they cannot appear as boundaries of optimizing tubes, even though in higher dimensions, hypersurfaces generated by onduloids in Euclidean space are known to occur as boundaries of optimal tubes connecting two parallel hyperplanes; see [Pedrosa and Ritoré 1999].

Corollary 1.4. *Let M be a connected compact rotational surface of constant mean curvature in hyperbolic* 3*-space. Suppose M is contained in the region between two horospheres, and that the boundary ∂M is either empty or lies on the horospheres, and meets them perpendicularly along its boundary. Then M is either*

(1) *of catenoid cousin type or umbilical with $H = 1$, or*

(2) *of equidistant type or umbilical with $0 < H < 1$, or*

(3) *of onduloid type or umbilical with $H > 1$.*

2. Preliminaries

Let $\mathcal{L}^4 = (\mathbb{R}^4, g)$ be the 4-dimensional Lorentz space endowed with the metric $g(x, y) = x_1 y_1 + x_2 y_2 + x_3 y_3 - x_4 y_4$ and the 3-dimensional hyperbolic space

$$\mathbb{H}^3(-1) := \{p = (x_1, x_2, x_3, x_4) \in \mathcal{L}^4 : g(p, p) = -1,\ x_4 > 0\}.$$

We use the upper half-space model $\mathbb{R}^3_+ := \{(x, y, z) \in \mathbb{R}^3; z > 0\}$ for $\mathbb{H}^3(-1)$, endowed with the metric

$$\frac{1}{z^2}(dx^2 + dy^2 + dz^2). \tag{2-1}$$

Let $\phi : \Sigma \to \mathbb{R}^3_+$ be an isometric immersion of a compact surface Σ with nonempty boundary $\partial\Sigma$, and let Γ be a curve in $\mathbb{R}^3_+$. If ϕ is a diffeomorphism of $\partial\Sigma$ onto Γ, we say that Γ is the boundary of ϕ; if ϕ has constant mean curvature H, we say that Σ is an H-surface with boundary Γ. We identify Σ with its image by ϕ and $\partial\Sigma$ with the curve Γ.

The plane $z = 0$ is called the infinity boundary of $\mathbb{R}^3_+$ and denoted $\partial_\infty\mathbb{R}^3_+$. The geodesics of $\mathbb{R}^3_+$ are represented by vertical Euclidean lines and half-circles orthogonal to $\partial_\infty\mathbb{R}^3_+$ and contained in $\mathbb{R}^3_+$. The totally geodesic surfaces have constant mean curvature $H = 0$ and are represented by vertical Euclidean planes and hemispheres orthogonal to $\partial_\infty\mathbb{R}^3_+$ and contained in $\mathbb{R}^3_+$.

The horizontal Euclidean translations and the rotations around a vertical geodesic are isometries of $\mathbb{R}^3_+$. We have two families of isometries associated to one point $p_0 \in \partial_\infty\mathbb{R}^3_+$, the Euclidean homotheties centered at p_0 with factor $k > 0$, called hyperbolic translations through a geodesic α perpendicular to $\partial_\infty\mathbb{R}^3_+$ at p_0, and the hyperbolic reflections with respect to a totally geodesic surface P.

When P is a hemisphere orthogonal to $\partial_\infty \mathbb{R}^3_+$ centered at p_0 and of radius $r > 0$, the hyperbolic reflections are Euclidean inversions centered at p_0 that fix P. When P is a vertical Euclidean plane, they are Euclidean reflections with respect to P.

Now we describe the umbilical surfaces of $\mathbb{R}^3_+$; see for example [López 1999].

- *Totally geodesic surfaces* are represented by vertical Euclidean planes in $\mathbb{R}^3_+$ and the hemispheres in $\mathbb{R}^3_+$ perpendicular to the plane $z = 0$. They have $H = 0$.
- *Geodesic spheres* are represented by Euclidean spheres entirely contained in $\mathbb{R}^3_+$. They have $H > 1$ (the mean curvature vector points to the interior). If ρ is the hyperbolic radius of a geodesic sphere, then $H = \coth\rho$.
- *Horospheres* are represented by horizontal Euclidean planes in $\mathbb{R}^3_+$ and Euclidean spheres in $\mathbb{R}^3_+$ that are tangent to $\partial_\infty \mathbb{R}^3_+$. They have $H = 1$; the mean curvature vector points upwards in the case of horizontal planes and to the interior in the case of spheres.
- *Equidistant surfaces* are represented by the intersection of $\mathbb{R}^3_+$ with the planes in $\mathbb{R}^3$ that are neither parallel nor perpendicular to the plane $z = 0$ and by (pieces of) Euclidean spheres that are not entirely contained in $\mathbb{R}^3_+$ and are neither tangent nor perpendicular to the plane $z = 0$. They have $0 < H < 1$, and the mean curvature vector points to the totally geodesic surface they are equidistant to.

In our study, the (spherical) rotational surfaces of $\mathbb{R}^3_+$ play an important role since the solutions of the isoperimetric problem must be rotationally invariant. They are defined as surfaces invariant by a subgroup of isometries whose principal orbits are (Euclidean) circles.

Let Π_1 and Π_2 be horospheres represented by distinct parallel horizontal Euclidean planes, and let $\Pi = \Pi_1 \cup \Pi_2$. Let $\mathcal{F} = \mathcal{F}(\Pi_1, \Pi_2)$ be the closed slab between them, and let $\phi : \Sigma \to \mathcal{F}$ be an isometric immersion of a compact, connected, embedded and orientable C^2 surface with boundary $\Gamma = \partial\Sigma$ and the property that $\phi(\Gamma) \subset \Pi$. (Later we will see that the image under ϕ of the interior of the surface Σ will not touch Π if Σ is the boundary of an optimal domain in our variational problem, but this is not part of the general situation yet.)

Now we fix notation for some well-known geometric invariants related to isometric immersions. We (locally) identify Σ with $\phi(\Sigma)$ and $X(p) \in T_p\Sigma$ with $d\phi_p(X(p)) \subset \mathbb{R}^3_+$. We have the decomposition $T_p(\mathbb{R}^3_+) = T_p(\Sigma) \oplus N_p(\Sigma)$ into the tangent and normal spaces to Σ at p. Choose an orientation for Σ, and let N be the (positive) unitary normal field along the immersion ϕ. If $X(p) \in T_p(\mathbb{R}^3_+)$, we may write $X(p) = X(p)^T + X(p)^N = X(p)^T + \alpha N(p)$, where $\alpha \in \mathbb{R}$.

Let $\langle \cdot, \cdot \rangle$ be the metric induced on Σ by the immersion ϕ, let $\overline{\nabla}$ be the Riemannian connection of the ambient space $\mathbb{R}^3_+$, and let ∇ be the induced Riemannian connection on Σ. Let $X, Y \in \mathfrak{X}(\Sigma)$ be C^∞ vector fields. Then $\nabla_X Y = (\overline{\nabla}_X Y)^T$ and $\mathcal{B}(X, Y) = (\overline{\nabla}_X Y)^N$ are, as usual, the *induced connection* on Σ and the *second fundamental form* of the immersion given by $\mathcal{B}$. We also have the *Weingarten operator* $A_N(Y) = -(\overline{\nabla}_Y N)^T$, so that $\langle A_N(X), Y\rangle = \langle \mathcal{B}(X, Y), N\rangle$. Finally, the mean curvature of the immersion ϕ is $H = 1/2\,\mathrm{trace}(A_N)$.

Definition 2.1. A variation of ϕ is a smooth map $F : (-\epsilon, \epsilon) \times \Sigma \to \mathbb{R}^3_+$ such that for all $t \in (-\epsilon, \epsilon)$ the map $\phi_t : \Sigma \to \mathbb{R}^3_+$, $p \mapsto F(t, p)$ is an immersion and satisfies $\phi_0 = \phi$.

For $p \in \Sigma$, we define the *variation vector field* of F by $X(p) = \partial\phi_t(p)/\partial t|_{t=0}$ and the *normal variation function* of F by $f(p) = \langle X(p), N(p)\rangle$. We say that the variation F is *normal* if X is normal to ϕ at each point; we say F has compact support if X has compact support. For a variation with compact support and for small values of t, we have that ϕ_t is an immersion of Σ in $\mathbb{R}^3_+$. In this case the *area function* $A : (-\epsilon, \epsilon) \to \mathbb{R}$ is given by

$$A(t) = \int_\Sigma dA_t = \int_\Sigma \sqrt{\det((d\phi_t)^*(d\phi_t))}\, dA,$$

where dA is the area element of Σ. The function $A(t)$ is the area of Σ with the metric induced by ϕ_t. We also define the *volume function* $V : (-\epsilon, \epsilon) \to \mathbb{R}$ by

$$V(t) = -\int_{[0,t]\times\Sigma} F^* d(\mathbb{R}^3_+),$$

where $d(\mathbb{R}^3_+)$ is the volume element of $\mathbb{R}^3_+$ and $F^* d(\mathbb{R}^3_+)$ is the pull-back of $d(\mathbb{R}^3_+)$ by F. The function $V(t)$ does not actually represent the volume of some region with $\phi_t(\Sigma)$ as boundary, but of a "tubular neighborhood" along $\phi(\Sigma)$ between $\phi(\Sigma)$ and $\phi_t(\Sigma)$. The sign is related to the net change with respect to the normal field defining the orientation; for example, contracting a sphere in $\mathbb{R}^3$, which means moving it in the direction of the mean curvature vector, gives the expected negative sign for $V(t)$.

Definition 2.2. Let $F : (-\epsilon, \epsilon) \times \Sigma \to \mathbb{R}^3_+$ be a variation of ϕ. We say F preserves volume if $V(t) = V(0)$ (which is equal to zero) for all $t \in (-\epsilon, \epsilon)$. We say F is admissible if $F(\partial\Sigma) \subset \Pi$ for all $t \in (-\epsilon, \epsilon)$.

Definition 2.3. We say that the immersion ϕ is stationary if $A'(0) = 0$ for all admissible variations that preserve volume.

Remark 2.4. Suppose Ω is a (compact) regular region in the slab $\mathcal{F}$ between the horospheres Π. Then by taking Σ in Definition 2.2 as the (embedded regular) free boundary of Ω, we may extend the variational approach above to produce a variation $\Omega(t)$ of Ω by embedded domains (for small t), such that the condition $V(t) = 0$ in Definition 2.2 is equivalent to holding $|\Omega(t)|$ equal to $\Omega(0)$ along the variation. This justifies saying that the variation "preserves volume" in the definition above.

We now restate our problem. Let Π_1 and Π_2 be two parallel horospheres in $\mathbb{R}^3$, and let $\mathcal{F} = \mathcal{F}(\Pi_2, \Pi_2)$ be the (closed) slab between them.

Isoperimetric problem for $\mathcal{F}$. For a fixed volume, find the domains $\Omega \subset \mathcal{F}$ that have minimal free boundary area.

Definition 2.5. A (compact) minimizing region Ω for this problem will be called an *isoperimetric domain* or *region* in $\mathcal{F}$.

More precisely, one looks to classify and describe geometrically the isoperimetric regions (as a function of the volume), that is, to find the isoperimetric profile (minimal free boundary area as a function of the prescribed volume) for $\mathcal{F}$.

3. First results about the isoperimetric solutions

Here, we characterize the stationary immersions according to Definition 2.3. The formulas below for the first variations of the area and volume functions are well known. For an immersed surface with boundary, the *exterior conormal* is the vector field along the boundary given as follows: In the tangent plane of Σ at $p \in \partial\Sigma = \Gamma$, take the outward unitary vector orthogonal to the tangent vector to Γ at p.

Proposition 3.1. *Let F be a variation of ϕ with variational field X and compact support in Σ. Then*

(1) $A'(0) = -2\int_\Sigma Hf\,dA + \int_\Gamma \langle X, \nu\rangle d\Gamma$, *where ν is the unitary exterior conormal*, *dA is the element of area of Σ and $d\Gamma$ is the element of length of Γ induced by ϕ*;

(2) $V'(0) = -\int_\Sigma f\,dA$, *where* $f(p) = \langle X(p), N(p)\rangle$.

Proof. Although the formula of the variation of the area functional is well known (see [Barbosa et al. 1988]), here we show a different way to deduce it. From the definition of $A(t)$, we obtain

$$A'(t) = \int_\Sigma \Big(\frac{1}{2\sqrt{\det((d\phi_t)^* d\phi_t)}} \det((d\phi_t)^* d\phi_t) \times \operatorname{trace}\big(((d\phi_t)^* d\phi_t)^{-1} \circ \tfrac{d}{dt}((d\phi_t)^* d\phi_t)\big)\Big)\,dA.$$

Since ϕ_0 is the inclusion of Σ in $\mathbb{R}^3_+$, $d\phi_0$ is the inclusion of the respective tangent spaces and $d\phi_0^*$ is the orthogonal projection on $T\Sigma$.

By evaluating $A'(t)$ for $t = 0$, we get

$$A'(0) = \int_\Sigma \tfrac{1}{2}\operatorname{trace}\big(\tfrac{d}{dt}\big|_{t=0}((d\phi_t)^* d\phi_t)\big)\,dA.$$

We apply the symmetry lemma for ∇^{ϕ} along the immersion and get

$$\frac{d}{dt}\Big|_{t=0}(d\phi_t) = \nabla^{\phi}\frac{\partial\phi_t}{\partial t}\Big|_{t=0} = \nabla^{\phi}X.$$

Then

$$A'(0) = \int_{\Sigma} \tfrac{1}{2}\,\mathrm{trace}\big((\nabla^{\phi}X)^*\big|_{T\Sigma} + \mathrm{proj}_{T\Sigma}\,\nabla^{\phi}X\big)dA = \int_{\Sigma}\mathrm{trace}(\mathrm{proj}_{T\Sigma}\,\nabla^{\phi}X)dA,$$

where $\mathrm{proj}_{T\Sigma}$ denotes the projection on $T\Sigma$.

By decomposing the variational field as $X = X^T + X^N$, the projections of the tangent and normal components of $\nabla^{\phi}(X)$ on $T\Sigma$ are

$$\mathrm{proj}_{T\Sigma}\,\nabla^{\phi}(X^T) = \nabla(X^T) \quad \text{and} \quad \mathrm{proj}_{T\Sigma}\,\nabla^{\phi}(X^N) = -\mathscr{A}_{X^N},$$

where $\mathscr{A}_{X^N}$ is the Weingarten operator on Σ. Therefore

$$A'(0) = \int_{\Sigma}(\mathrm{div}\,X^T - 2\langle X^N, HN\rangle)dA.$$

We apply Stokes's theorem and get

$$A'(0) = \int_{\Gamma}\langle X^T, \nu\rangle d\Gamma - 2\int_{\Sigma}\langle X^N, HN\rangle dA = -2\int_{\Sigma} Hf\,dA + \int_{\Gamma}\langle X, \nu\rangle d\Gamma.$$

The first variation of volume given in (2) is standard and its proof will be omitted; see [Barbosa et al. 1988]. □

From the next result we conclude that the boundary of our isoperimetric region must be an H-surface that contacts the horospheres Π_1 and Π_2 perpendicularly.

Theorem 3.2. *Let $\phi : \Sigma \to \mathbb{R}^3_+$ be an immersion with boundary $\Gamma = \partial\Sigma$. Let $\Pi = \Pi_1 \cup \Pi_2$ be the horospheres containing Γ. Then ϕ is stationary if and only if it has constant mean curvature and intersects Π (if it does) perpendicularly along Γ.*

Proof. We may show the reverse implication by adapting the proof of [Barbosa and do Carmo 1984, Proposition 2.7]. To show that ϕ meets Π perpendicularly along Γ if ϕ is stationary, we take an admissible variation Φ that preserves volume with variational field X, and we take $p_0 \in \partial\Sigma$. Suppose by contradiction that $\langle X(p_0), \nu(p_0)\rangle \neq 0$. By continuity there is a neighborhood $U = W_1 \cap \partial\Sigma$ of p_0 such that $\langle X(p), \nu(p)\rangle > 0$ for all $p \in U$, where W_1 is a neighborhood of p_0 in Σ. Take $q \in \mathring{\Sigma} \setminus W_1$, let W_2 be a neighborhood of q disjoint from W_1, and let $\mathscr{P}$ be a partition of unity on $W_1 \bigcup W_2$. There exists a differentiable function $\xi_1 : W_1 \to \mathbb{R}$ such that $\xi_1(W_1) \subset [0, 1]$ and with support $\mathrm{supp}\,\xi_1 \subset W_1$. We may also take a differentiable map $\xi_2 : W_2 \to \mathbb{R}$ such that $\xi_2(W_2) \subset [0, 1]$, $\mathrm{supp}\,\xi_2 \subset W_2$ and

$$\int_{W_1}\xi_1 f\,dW_1 + \int_{W_2}\xi_2 f\,dW_2 = 0.$$

Define a variation $\Phi_\xi : (-\epsilon, \epsilon) \times \Sigma \to \mathbb{R}$ with compact support on $W_1 \bigcup W_2$ by

$$\Phi^t_\xi(p) = \Phi_\xi(t, p) = \begin{cases} \Phi(\xi_1 t, p) & \text{if } p \in W_1, \\ \Phi(\xi_2 t, p) & \text{if } p \in W_2. \end{cases}$$

Note that Φ_ξ is admissible because Φ is admissible.

If $f_\xi(p)$ denotes the normal component of the variation vector, we have

$$\int_\Sigma f_\xi(p)\, dA = \int_{W_1} \xi_1(p) f(p)\, dW_1 + \int_{W_2} \xi_2(p) f(p)\, dW_2 = 0,$$

and Φ_ξ preserves volume.

For this variation we have

$$0 = A'(0) = -2H \int_\Sigma f_\xi\, dA + \int_{W_1} \xi_1 \langle X, \nu\rangle\, d\Gamma = \int_{W_1} \xi_1 \langle X, \nu\rangle\, d\Gamma > 0,$$

which is a contradiction. Then for all $p \in \partial\Sigma$, it follows that $\langle X, \nu\rangle(p) = 0$. □

Next we show that the isoperimetric domains are rotationally invariant.

We need some symmetrization principle for H-surfaces. By taking the hyperbolic version of Aleksandrov's principle of reflection (for further references and details, see [Aleksandrov 1962]) and using [Barbosa and Sa Earp 1998] to specialize to the case of reflection planes, we get the next result. A detailed proof may be found in [López 2006].

Theorem 3.3. *Suppose Σ is a compact, connected, orientable and embedded H-surface of class C^2 that lies between two parallel horospheres Π_1 and Π_2 in $\mathbb{R}^3_+$ and has boundary $\partial\Sigma \subset \Pi_1 \bigcup \Pi_2$ (possibly empty). Then Σ is rotationally symmetric around an axis perpendicular to Π_1 and Π_2.*

We observe that the intersection of Σ with a horosphere $\mathcal{H}$ (represented by a horizontal Euclidean plane) is just a Euclidean circle. In fact, if there were two concentric circles and the isoperimetric region $\mathcal{R}$ was delimited by these circles, we would apply the Aleksandrov's reflection principle with respect to vertical Euclidean planes and get a totally geodesic symmetry plane P determined by the first contact point x_0; see Figure 2. However, P would obviously not contain the axis of symmetry of $\mathcal{R}$. See [López 2006] for a detailed proof of this fact.

4. Isoperimetric regions between horospheres in $\mathbb{R}^3_+$

We now classify the rotational H-surfaces of $\mathbb{R}^3_+$ that lie between two parallel horospheres, have boundary contained in the horospheres, and intersect the horospheres perpendicularly. In so doing, we get all possible solutions for the isoperimetric problem in hyperbolic space, since the solutions must be regions delimited by these H-surfaces. We start with important results from the thesis of Barrientos [1995].

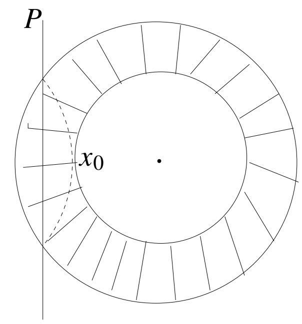

Figure 2. Excluded case: region between two concentric circles

If (ρ, θ, z) are the cylindrical coordinates of a point p in $\mathbb{R}^3_+$, then the Cartesian coordinates are given by

$$(\bar{x}, \bar{y}, \bar{z}) = e^z(\tanh\rho\cos\theta, \tanh\rho\sin\theta, \operatorname{sech}\rho). \tag{4-1}$$

For a spherical rotational surface Σ of $\mathbb{R}^3_+$ around the $\bar{z}$-axis, the orbit of a point $p = (\rho_0, \theta_0, z_0)$ is $\mathcal{R}_\varphi(p) = (\rho_0, \theta_0 + \varphi, z_0)$, where $\mathcal{R}_\varphi(p)$ denotes the rotation of p with angle φ around $\bar{z}$. If $c(s)$ is the profile curve of Σ parametrized by arclength, we can parametrize Σ as $X(s,t) = \mathcal{R}_t(c(s))$, so that its metric is $d\sigma^2 = ds^2 + U^2(s)\,dt^2$, where $U = U(s)$ is a positive function, s is the arclength parameter of $c(s)$, and $dt = d\varphi$. We call (s,t) the *natural parameters* of Σ. From (4-1), the metric (2-1) is given by

$$d\rho^2 + \sinh^2\rho\, d\theta^2 + \cosh^2\rho\, dz^2. \tag{4-2}$$

In the plane $\theta = 0$, the profile curve $c(s)$ can be locally viewed as the graph $z = \lambda(s) = \lambda(\rho(s))$. From (4-2) we have in this parameterization that $dt = d\varphi$, $ds = (1 + \lambda'^2(\rho)\cosh^2\rho)^{1/2} d\rho$,

$$U^2(s) = \sinh^2\rho(s) \quad \text{and} \quad \lambda'^2(s) = \frac{1 + U^2(s) - U'^2(s)}{(1 + U^2(s))^2}. \tag{4-3}$$

Then the natural parametrization for a rotational surface in cylindrical coordinates is

$$\sinh^2\rho(s) = U^2(s), \qquad \lambda(s) = \int_0^s \frac{\sqrt{1 + U^2(t) - U'^2(t)}}{1 + U^2(t)}\, dt, \qquad \varphi(t) = t.$$

Barrientos [1995] classified the H-surfaces of $\mathbb{R}^3_+$; another important reference is [Sterling 1987]. By setting $\zeta(s) = U^2(s)$, Barrientos found that the differential equation for the rotational H-surfaces in $\mathbb{R}^3_+$ is

$$\zeta'^2/4 = (1 - H^2)\zeta^2 + (1 + 2aH)\zeta - a^2,$$

and showed that the behavior of their profile curves is determined by the constant of integration a. After choosing the surface orientation so that $H \geq 0$, there are three cases to study: $H = 1$, $H \in [0, 1)$ and $H > 1$.

Next we give in each of these cases the natural parametrizations for a rotational H-surface in $\mathbb{R}^3_+$ generated by a curve $c(s) = (\rho(s), \lambda(s))$.

$$H = 1 \qquad \begin{cases} \sinh^2 \rho(s) = \dfrac{a^2 + (1+2a)^2 s^2}{1+2a}, \\[2ex] \lambda(s) = \displaystyle\int_0^s \frac{\sqrt{1+2a}\,(-a(1+a) + (1+2a)^2 t^2)\sqrt{a^2 + (1+2a)^2 t^2}}{(-a(1+a) + (1+2a)^2 t^2)^2 + (1+2a)^4 t^2}\, dt, \\[2ex] \varphi(t) = t. \end{cases}$$

$$H \in [0, 1) \quad \begin{cases} \sinh^2 \rho(s) = \dfrac{-A + B\cosh(2\alpha s)}{2\alpha^2}, \\[2ex] \lambda(s) = \displaystyle\int_0^s \frac{\sqrt{2}\alpha(-2a + H(-1 + B\cosh(2\alpha t)))\sqrt{-A + B\cosh(2\alpha t)}}{(-2a + H(-1 + B\cosh(2\alpha t)))^2 + \alpha^2 B^2 \sinh^2(2\alpha t)}\, dt, \\[2ex] \varphi(t) = t, \end{cases}$$

where $A = 1 + 2aH$, $B = \sqrt{1 + 4aH + 4a^2}$ and $\alpha = \sqrt{1 - H^2}$.

$$H > 1 \qquad \begin{cases} \sinh^2 \rho(s) = \dfrac{A + B\sin(2\alpha s)}{2\alpha^2}, \\[2ex] \lambda(s) = \displaystyle\int_0^s \frac{\sqrt{2}\alpha(2a + H(1 + B\sin(2\alpha t)))\sqrt{A + B\sin(2\alpha t)}}{(2a + H(1 + B\sin(2\alpha t)))^2 + \alpha^2 B^2 \cos^2(2\alpha t)}\, dt, \\[2ex] \varphi(t) = t, \end{cases}$$

where $A = 1 + 2aH$, $B = \sqrt{1 + 4aH + 4a^2}$ and $\alpha = \sqrt{H^2 - 1}$.

Now we introduce some notations and definitions used throughout this section. From (4-1) the profile curve of a rotational H-surface in $\mathbb{R}^3_+$ is given by

$$c_+(s) = e^{\lambda(s)}(\tanh \rho(s), \operatorname{sech} \rho(s)). \tag{4-4}$$

Here $\rho(s)$ and $\lambda(s)$ are determined by the suitable parametrization above. In the $H = 1$ case, $a > -1/2$. When $-1/2 < a < 0$, we say the rotational surfaces are of *catenoid cousin type*. In the $H \in [0, 1)$ case, $a \in \mathbb{R}$. When $a < 0$, we say the rotational surfaces are of *equidistant type*. In the $H > 1$ case, we have $a \geq (-H + \sqrt{H^2 - 1})/2$. When $-1/(4H) < a < 0$, we obtain the *onduloid type surfaces*. In each case, we get umbilical surfaces when $a = 0$.

By taking $\lambda = 0$ in (4-4), we get the curve $c_g(s) = (\tanh \rho(s), \operatorname{sech} \rho(s))$, which is an upper half-circle perpendicular to the $\bar{z}$-axis. Namely, it is a geodesic with Euclidean radius $r = (\tanh^2 \rho(s) + \operatorname{sech}^2 \rho(s))^{1/2} = 1$. The curve $c_g(s)$ is called the *geodesic radius*. Our analysis works up to Euclidean homotheties $\mathcal{H}_r$ for general

$r > 0$, namely

$$\mathcal{H}_r(c_+(s)) = e^{\lambda(s)}(r \tanh \rho(s), r \operatorname{sech} \rho(s));$$

these give other families of profile curves of rotational H-surfaces. For the results in this section that deal with geodesic radius, we take $r = 1$. By Theorems 3.2 and 3.3, the boundaries of the isoperimetric solutions must be rotational H-surfaces that meet the horospheres $\{z = c_1\}$ and $\{z = c_2\}$ perpendicularly. Now our goal is to determine the vertical tangency points of the profile curves of the rotational surfaces.

Definition 4.1. Suppose $c_+(s)$ is a curve parametrized by (4-4). We say that a point $c_+(s)$ is a vertical tangency point if the tangent vector at $c_+(s)$ satisfies $c'_+(s) = (0, b)$, where $b \in \mathbb{R}^*$, that is,

$$e^{\lambda(s)}(\tanh \rho(s)\lambda'(s) + \operatorname{sech}^2 \rho(s)\rho'(s)) = 0, \tag{4-5}$$

$$e^{\lambda(s)}(\operatorname{sech} \rho(s)\lambda'(s) - \operatorname{sech} \rho(s) \tanh \rho(s)\rho'(s)) = b. \tag{4-6}$$

Since $e^{\lambda(s)} > 0$, Equation (4-5) implies that

$$\tanh \rho(s)\lambda'(s) + \operatorname{sech}^2 \rho(s)\rho'(s) = 0. \tag{4-7}$$

By (4-7) we obtain the points where the vertical tangency occurs, and by (4-6) we get the direction of the tangency (upward or downward).

By applying (4-3) to (4-7) we see that, if p is a vertical tangency point with $U(s) \neq 0$, then

$$U^2(s) = U'^2(s), \tag{4-8}$$

and the roots of (4-8) give us the vertical tangency points.

Next we study the behavior of the profile curve of rotational H-surfaces and determine the possible vertical tangency points. In each case, $\rho(s)$, $\lambda(s)$ and $U(s)$ are those the of corresponding parametrization on page 11.

The case $H = 1$.

Theorem 4.2. *If $c_+(s) = e^{\lambda(s)}(\tanh \rho(s), \operatorname{sech} \rho(s))$ is the parametrization of the profile curve of a rotational H-surface in $\mathbb{R}^3_+$ with $H = 1$, then $c_+(s)$ is symmetric with respect to the geodesic radius c_g.*

Proof. By definition, $\lambda(0) = 0$, so $c_+(0) \in c_g$. If I is the Euclidean inversion through c_g, then $I(c_+(s)) = c_+(-s)$, because $\rho(s)$ is even and $\lambda(s)$ is odd. □

By the definition of $\rho(s)$,

$$\sinh \rho(s) = 0 \quad \text{if and only if} \quad a = 0 \text{ and } s = 0. \tag{4-9}$$

So $\tanh\rho(s) > 0$ if neither a nor s is zero. Furthermore $s = 0$ is the unique minimum point of $\rho(s)$.

As for $U(s)$ in the case $H = 1$, we have

$$U^2(s) = \frac{a^2 + (1+2a)^2 s^2}{1+2a} \quad \text{and} \quad U'^2(s) = \frac{(1+2a)^3 s^2}{a^2 + (1+2a)^2 s^2}. \tag{4-10}$$

By applying (4-10) to (4-8) it follows that

$$(1+2a)^4 s^4 + (2a^2(1+2a)^2 - (1+2a)^4)s^2 + a^4 = 0. \tag{4-11}$$

This is a second order equation in s^2, with discriminant

$$\Delta = (1+2a)^6(4a+1). \tag{4-12}$$

Since $1+2a > 0$, we have these facts in the $H = 1$ case:

- If $-1/2 < a < -1/4$, then (4-11) has no real roots, so there are no points of vertical tangency.
- If $a = -1/4$, there are at most two vertical tangency points

$$s = \pm 1/2. \tag{4-13}$$

- If $a > -1/4$, there are at most four vertical tangency points, given by

$$\begin{aligned} s_1 &= \frac{1+2a+\sqrt{1+4a}}{2(1+2a)}, \qquad s_2 = -s_1, \\ s_3 &= \frac{1+2a-\sqrt{1+4a}}{2(1+2a)}, \qquad s_4 = -s_3. \end{aligned} \tag{4-14}$$

Besides these pieces of information, we study the vertical tangencies as the parameter a varies.

The subcase $-1/4 \le a < 0$. In this case, we have $\lambda'(s) > 0$. If $s \ge 0$, then

$$\tanh\rho(s)\lambda'(s) + \operatorname{sech}^2\rho(s)\rho'(s) > 0,$$

and (4-7) is not possible. Since $1+2a > 0$, the roots s_1 and s_3 of (4-8) given by (4-14) are strictly positive and thus do not give vertical tangency points. The other roots $s_2, s_4 < 0$ give us the vertical tangency points with upward direction for $b > 0$ in (4-6). The left figure in Figure 3 shows the profile curve for $H = 1$ and $a = -0.2$ and the horocycles that pass through the vertical tangency. To the right, we see two parallel horospheres and the rotational surface between the horospheres that meets them perpendicularly.

In particular, if $a = -1/4$, the positive root $s = 1/2$ of (4-8) given by (4-13) does not give a vertical tangency point, and there is only one vertical tangency

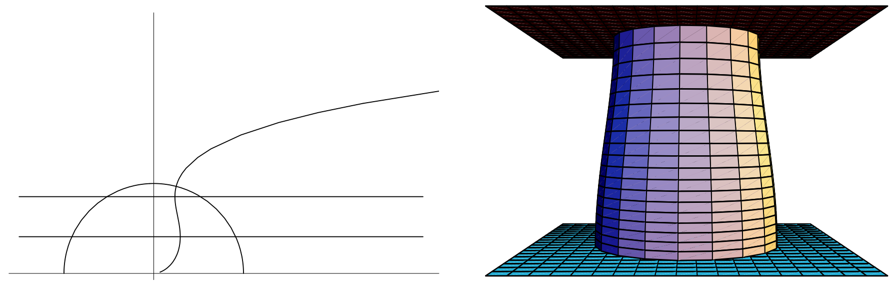

Figure 3. At left, the profile curve for $H = 1$ and $a = -0.2$, and at right the corresponding rotational surface.

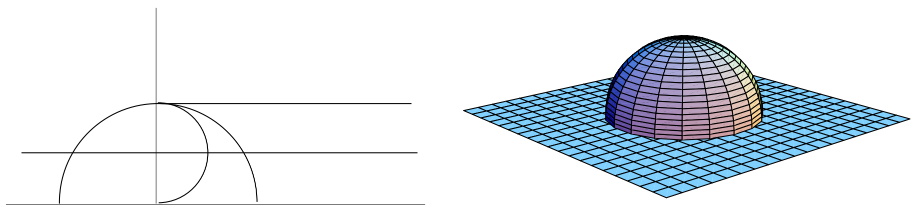

Figure 4. At left, the profile curve for $H = 1$ and $a = 0$, and at right the corresponding rotational surface.

point corresponding to $s = -1/2$. From the definitions of $\rho(s)$ and $\lambda(s)$ we get

$$\begin{aligned} &\lim_{s\to-\infty} e^{\lambda(s)} \tanh\rho(s) = 0, \qquad \lim_{s\to-\infty} e^{\lambda(s)} \operatorname{sech}\rho(s) = 0,\\ &\lim_{s\to\infty} e^{\lambda(s)} \tanh\rho(s) = \infty, \qquad \lim_{s\to-\infty} e^{\lambda(s)} \operatorname{sech}\rho(s) = \infty. \end{aligned} \tag{4-15}$$

Geometrically, one sees it's impossible to get an isoperimetric region in this case.

The subcase $a = 0$. In this case, two pieces of horocycle tangent at (0, 1) generate the umbilical surfaces with $H = 1$. They are represented by the Euclidean plane $\{z = 1\}$ or the Euclidean sphere with radius $1/2$ and tangent to $\partial\mathbb{R}^3_+$ at (0, 0, 0). In the latter case there is only one vertical tangency point and the surface boundary meets only one of the horospheres perpendicularly. In fact, by taking the upper Euclidean half-sphere that represents the horosphere, we get the possible isoperimetric solution for the umbilical case with $H = 1$. Figure 4 illustrates the situation.

The subcase $a > 0$. In this case, the profile curves have only one self-intersection. From (4-4), if $c_+(s_i) = c_+(s_j)$ is a self-intersection, then $s_i = \pm s_j$. Since the curves are symmetric with respect to c_g, the self-intersections must occur on c_g. So $\lambda(s_i) = \lambda(s_j) = 0$. By its definition in this case, we deduce that $\lambda(s)$ has a maximum at $-\sqrt{a(1+a)}/(1+2a)$ and a minimum at $\sqrt{a(1+a)}/(1+2a)$. Furthermore, $\lim_{s\to\infty} \lambda(s) = \infty$ (see [Barrientos 1995]), $\lambda(0) = 0$, and $\lambda(s)$ is an odd function.

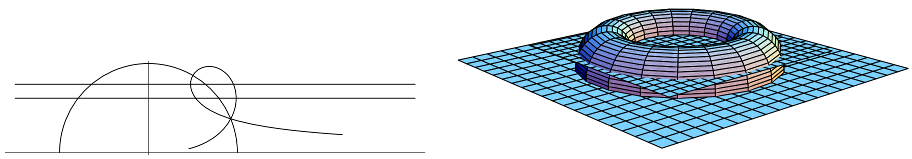

Figure 5. At left, the profile curve for $H = 1$ and $a = 1$, and at right the corresponding rotational surface. This case cannot be an isoperimetric solution.

Also, $\rho(s)$ has only one minimum at $s = 0$. So, if $s > \sqrt{a(1+a)}/(1+2a) > 0$ then $\rho'(s)$, $\lambda'(s) > 0$ and

$$\tanh \rho(s)\lambda'(s) + \operatorname{sech}^2 \rho(s)\rho'(s) > 0.$$

If however $-\sqrt{a(1+a)}/(1+2a) < s < 0$, then $\rho'(s)$, $\lambda'(s) < 0$, which implies that

$$\tanh \rho(s)\lambda'(s) + \operatorname{sech}^2 \rho(s)\rho'(s) < 0.$$

In either case, (4-7) does not hold. Since $a > 0$, the roots s_1, s_2, s_3 and s_4 of (4-8) given by (4-14) satisfy

$$s_2 < -\frac{\sqrt{a(1+a)}}{1+2a} < s_4 < 0 < s_3 < \frac{\sqrt{a(1+a)}}{1+2a} < s_1,$$

Therefore, vertical tangency is possible only for the positive roots s_2 and s_3. Since $\rho'(s_2) < 0$ and $\lambda'(s_2) > 0$, the vertical tangency for s_2 is upward. However, it is downward for s_3, since $\rho'(s_3) > 0$ and $\lambda'(s_3) < 0$. The isoperimetric solution is not possible in this case because, if the vertical tangencies did not occur at the same height, then a piece of the rotational surface would be outside of the region between the horospheres; see Figure 5.

Even if vertical tangency occurred at the same height, the intersection of the rotational H-surface with the parallel horospheres would be two concentric circles, which is not possible due to Theorem 3.3.

In summary, for $H = 1$ the boundary of the region Ω must be either a catenoid cousin-type surface as in Figure 3 or umbilical as in Figure 4

We proceed in the analogous way to study the other cases, and give here only the main equations and results.

The case $H \in [0, 1)$.

Theorem 4.3. *If $c_+(s) = e^{\lambda(s)}(\tanh \rho(s), \operatorname{sech} \rho(s))$ is the parametrization of the profile curve of a rotational H-surface in $\mathbb{R}^3_+$ with $H \in [0, 1)$, then*

(1) *$c_+(s)$ is symmetric with respect to the geodesic radius c_g, and*

(2) *the asymptotic boundary of the profile curves consists of one or two points.*

Proof. The proof of (1) is similar to that of Theorem 4.2. Barrientos [1995] showed that $\rho(s)$ is unbounded but $\lambda(s)$ is bounded and has a limit. Then we have $\lim_{|s|\to\infty} e^{\lambda(s)} \operatorname{sech} \rho(s) = 0$, and the asymptotic boundary of the profile curves consists of one or two points. □

Since $\sinh \rho(s) \geq 0$, we have by definition (page 11) that $\sinh \rho(s) = 0$ if and only if $a = 0$ and $s = 0$.

Again by definition, it follows that

$$U^2(s) = \frac{-A + B\cosh(2\alpha s)}{2\alpha^2} \quad \text{and} \quad U'^2(s) = \frac{B^2 \sinh^2(2\alpha s)}{2(-A + B\cosh(2\alpha s))}. \tag{4-16}$$

By applying (4-16) to (4-8), we get

$$B^2 H^2 \cosh^2(2\alpha s) - 2AB\cosh(2\alpha s) + A^2 + \alpha^2 B^2 = 0. \tag{4-17}$$

This is a second order equation in $\cosh(2\alpha s)$, with discriminant

$$\Delta = 4B^2(1 - H^2)^2(1 + 4aH).$$

Since $B > 0$ if $H \in (0, 1)$ and a is defined for any real, we have these facts:

- If $a < -1/(4H)$, there are no vertical tangency points.
- If $a = -1/(4H)$, there are at most two vertical tangency points

$$s = \pm\frac{1}{2\alpha} \operatorname{arccosh}(1/H). \tag{4-18}$$

- If $a > -1/(4H)$, there are at most four vertical tangency points

$$\begin{aligned} s_1 &= \frac{1}{2\alpha} \operatorname{arccosh}\left(\frac{A + (1 - H^2)\sqrt{1 + 4aH}}{BH^2}\right), \quad s_2 = -s_1, \\ s_3 &= \frac{1}{2\alpha} \operatorname{arccosh}\left(\frac{A - (1 - H^2)\sqrt{1 + 4aH}}{BH^2}\right), \quad s_4 = -s_3. \end{aligned} \tag{4-19}$$

In particular, $H = 0$ in (4-17) gives $2B\cosh(2s) - 1 - B^2 = 0$, whose solutions are $s = \pm(1/2)\operatorname{arccosh}(B^2 + 1/(2B))$.

First, let us specialize to the case that $H \in (0, 1)$; we'll treat $H = 0$ later.

The subcase $-1/(4H) \leq a < 0$. In this case, only the roots $s_2, s_4 < 0$ give us vertical tangency points with upward direction. Figure 6 shows the profile curve for $H = 0.5$ and $a = -0.25$ and the horocycles that pass through the vertical tangencies. The mean curvature vector for the part of the rotational surface in the interior of the totally geodesic (symmetry plane of the surface) points out toward the rotation axis, and so determines the isoperimetric region illustrated at right.

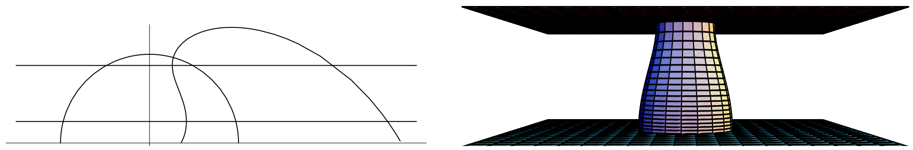

Figure 6. At left, profile curve for $H = 0.5$ and $a = -0.25$, and at right the corresponding rotational surface.

In particular, if $a = -1/(4H)$ there is only one vertical tangency point at $s = -1/(2\alpha)\operatorname{arccosh}(1/H) < 0$. Although the profile curve intersects the horocycle at another point, it is not of vertical tangency.

The subcase $a = 0$. In this case, two pieces of equidistant curve tangent at $(0, 1)$ generate the umbilical surfaces with $H \in (0, 1)$. They are represented by pieces of Euclidean spheres tangent at $(0, 0, 1)$. The vertical tangency occurs only for the equidistant profile curve that is the nearest to the rotation axis. Since the mean curvature vector of this umbilical surface points outward to the rotation axis, it determines an isoperimetric region.

The subcase $a > 0$. If $a > 0$ only the roots s_2 and s_3 correspond to vertical tangencies with directions upward in s_2 and downward in s_3. Geometrically, one sees that it is impossible to get an isoperimetric region in this case.

As for the $H = 0$ case, for $a < 0$ or $a > 0$ we get only one vertical tangency point. If $a = 0$, the rotational surface is a totally geodesic plane. Thus it is not possible to get an isoperimetric region for any $a \in \mathbb{R}$.

Finally, we conclude that for $H \in [0, 1)$ the boundary of the region Ω must be an equidistant-type surface (see Figure 6) or an umbilical surface with $H \in (0, 1)$.

The case $H > 1$.

Theorem 4.4. *If $c_+(s) = e^{\lambda(s)}(\tanh\rho(s), \operatorname{sech}\rho(s))$ is the parametrization of the profile curve of a rotational H-surface in $\mathbb{R}^3_+$ with $H > 1$, then $c_+(s)$ is a periodic curve with period π/α.*

Proof. We show that the hyperbolic length of the segment with extremes $c_+(s)$ and $c_+(s + \pi/\alpha)$ is constant for all s. Barrientos [1995] shows that

$$\rho(s + \pi/\alpha) = \rho(s) \quad \text{and} \quad \lambda(s + \pi/\alpha) = \lambda(s) + \lambda(\pi/\alpha), \tag{4-20}$$

which implies from (4-4) that $c_+(s + \pi/\alpha) = e^{\lambda(\pi/\alpha)} c_+(s)$.

We fix s_0 and parametrize the segment with extremes $c_+(s_0)$ and $c_+(s_0 + \pi/\alpha)$ by

$$\beta(t) = \left(t, \frac{t}{\sinh\rho(s_0)}\right), \quad \text{with } e^{\lambda(s_0)}\tanh\rho(s_0) \le t \le e^{\lambda(s_0)} e^{\lambda(\pi/\alpha)}\tanh\rho(s_0).$$

Therefore its hyperbolic length is $L(\beta(t)) = \lambda(\pi/\alpha)\cosh\rho(s_0)$.

The length of the segment depends only on the function $\rho(s)$ with period π/α, given in (4-20). So $L(\beta(t))$ is the same for any s_0. □

Again by definition, it follows that

$$U^2(s) = \frac{A + B\sin(2\alpha s)}{2\alpha^2} \quad \text{and} \quad U'^2(s) = \frac{B^2\cos^2(2\alpha s)}{2(A + B\sin(2\alpha s))}. \tag{4-21}$$

By applying (4-21) to (4-8), we get

$$B^2H^2\sin^2(2\alpha s) + 2AB\sin(2\alpha s) + A^2 - \alpha^2B^2 = 0.$$

This is a second order equation in $\sin(2\alpha s)$, with discriminant

$$\Delta = 4B^2(H^2 - 1)^2(1 + 4aH).$$

Since $H > 1$ and $B > 0$, this discriminant leads to these conclusions:

- If $a < -1/(4H)$, there are no vertical tangency points.
- If $a = -1/(4H)$, the possible vertical tangency points are[1]

$$\begin{aligned} s_k &= \frac{1}{2\alpha}\arcsin(1/H) + k\pi/\alpha, \quad \text{for } k \in \mathbb{Z}, \\ \tilde{s}_k &= \frac{1}{2\alpha}\widetilde{\arcsin}(1/H) + k\pi/\alpha \quad \text{for } k \in \mathbb{Z}. \end{aligned} \tag{4-22}$$

- If $a > -1/(4H)$ the possibilities are, for $k \in \mathbb{Z}$,

$$\begin{aligned} S_k &= \frac{1}{2\alpha}\arcsin(\mathcal{D}_+) + k\pi/\alpha, \quad s_k = \frac{1}{2\alpha}\arcsin(\mathcal{D}_-) + k\pi/\alpha, \\ \tilde{S}_k &= \frac{1}{2\alpha}\widetilde{\arcsin}(\mathcal{D}_+) + k\pi/\alpha, \quad \tilde{s}_k = \frac{1}{2\alpha}\widetilde{\arcsin}(\mathcal{D}_-) + k\pi/\alpha, \end{aligned}$$

  where $\mathcal{D}_\pm = (-A \pm (H^2 - 1)\sqrt{1 + 4aH})/(BH^2)$

Now we determine when the vertical tangency really occurs, depending on the geometry of the profile curve.

(1) If $-1/(4H) \le a < 0$, only the roots $\tilde{S}_k$ and $\tilde{s}_k$ give the vertical tangency points with upward direction; see Figure 7.

(2) If $a = 0$, we have tangent geodesic half-circles along the rotation axis, each of which generates a geodesic sphere in $\mathbb{R}^3_+$; these are umbilical surfaces and isoperimetric regions.

(3) If $a > 0$, we may analyze the behavior of the profile curves in the interval $]-\pi/(4\alpha), 3\pi/(4\alpha)[$, since by Theorem 4.4 they are (π/α)-periodic. It is easy to see that only the roots s_0 and $\tilde{S}_0$ give vertical tangency with directions

[1]Here $\widetilde{\arcsin}$ is the inverse sine such that $\widetilde{\arcsin}\, x \in [\pi/2, 3\pi/2)$, and arcsin is the usual inverse.

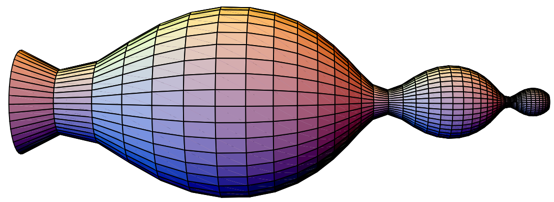

Figure 7. Hyperbolic onduloid with $H = 3$ and $a = -0.05$.

downward in s_0 and upward in $\tilde{S}_0$. This case corresponds to the so-called hyperbolic nodoids, which are not embedded surfaces.

In summary, for $H > 1$, the boundary of the region Ω can be either an onduloid-type surface as in Figure 7 or an umbilical surface.

Proof of Theorem 1.1. We start with the existence and then obtain the possible minimizing regions. By Theorem 3.3, the solutions to the isoperimetric problem have as boundaries rotationally invariant surfaces that have constant mean curvature where they are regular. But they must be regular (actually analytic), since the singularities along such boundaries must have, by well-known results, (Hausdorff) codimension at least 7, which is not possible for (2-dimensional) surfaces. Now, by results of [Morgan 1994], the existence of the isoperimetric solutions follows from the fact that $\mathcal{F}_{c_1,c_2}/G$ is compact, where G is the group of isometries of $\mathbb{R}^3_+$ whose elements leave invariant the region $\mathcal{F}_{c_1,c_2}$ between the horospheres, that is, the rotations around a vertical geodesic and the horizontal translations. The second part of Theorem 1.1 follows from the analysis of vertical tangencies done in the cases $H = 1$, $H \in [0, 1)$, and $H > 1$. □

plus .5pt

References

[Aleksandrov 1962] A. D. Aleksandrov, "Uniqueness theorems for surfaces in the large, V", *Amer. Math. Soc. Transl.* (2) **21** (1962), 412–416. MR 27 #698e Zbl 0119.16603

[Athanassenas 1987] M. Athanassenas, "A variational problem for constant mean curvature surfaces with free boundary", *J. Reine Angew. Math.* **377** (1987), 97–107. MR 88j:49027 Zbl 0604.53003

[Barbosa and do Carmo 1984] J. L. Barbosa and M. do Carmo, "Stability of hypersurfaces with constant mean curvature", *Math. Z.* **185**:3 (1984), 339–353. MR 85k:58021c Zbl 0513.53002

[Barbosa and Sa Earp 1998] J. L. Barbosa and R. Sa Earp, "Prescribed mean curvature hypersurfaces in H^{n+1} with convex planar boundary, II", pp. 43–79 in *Séminaire de Théorie spectrale et géométrie*, vol. 16, Institut Fourier, Université Grenoble I, 1998.

[Barbosa et al. 1988] J. L. Barbosa, M. do Carmo, and J. Eschenburg, "Stability of hypersurfaces of constant mean curvature in Riemannian manifolds", *Math. Z.* **197**:1 (1988), 123–138. MR 88m:53109 Zbl 0653.53045

[Barrientos 1995] J. Barrientos, *Superfícies helicoidais com curvatura média constante em espaços forma tridimensionais*, thesis, Pontifícia Universidade Católica do Rio de Janeiro, 1995.

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Mathematischen Wissenschaften **284**, Springer, New York, 1986. MR 88f:49001 Zbl 0583.35002

[López 1999] R. López, "Constant mean curvature surfaces with boundary in hyperbolic space", *Monatsh. Math.* **127**:2 (1999), 155–169. MR 2000m:53017 Zbl 0981.53053

[López 2006] R. López, "Symmetry of stationary hypersurfaces in hyperbolic space", *Geom. Dedicata* **119** (2006), 35–47. MR 2007m:58019 Zbl 1116.53008

[Morgan 1994] F. Morgan, "Clusters minimizing area plus length of singular curves", *Math. Ann.* **299**:4 (1994), 697–714. MR 95g:49083 Zbl 0805.49025

[Morgan 2009] F. Morgan, *Geometric measure theory: A beginner's guide*, 4th ed., Academic Press, San Diego, 2009. MR 2009i:49001 Zbl 05501365

[Pedrosa and Ritoré 1999] R. H. L. Pedrosa and M. Ritoré, "Isoperimetric domains in the Riemannian product of a circle with a simply connected space form and applications to free boundary problems", *Indiana Univ. Math. J.* **48**:4 (1999), 1357–1394. MR 2001k:53120

[Ritoré and Ros 2002] M. Ritoré and A. Ros, "Some updates on isoperimetric problems", *Math. Intelligencer* **24**:3 (2002), 9–14. MR 2003g:53014

[Sterling 1987] I. Sterling, "A generalization of a theorem of Delaunay to rotational W-hypersurfaces of σ_l-type in H^{n+1} and S^{n+1}", *Pacific J. Math.* **127**:1 (1987), 187–197. MR 88c:53015 Zbl 0579.53041

[Vogel 1987] T. I. Vogel, "Stability of a liquid drop trapped between two parallel planes", *SIAM J. Appl. Math.* **47**:3 (1987), 516–525. MR 88e:53010 Zbl 0627.53004

Received October 24, 2008. Revised July 23, 2009.

ROSA MARIA BARREIRO CHAVES
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO
RUA DO MATÃO, 1010
05508-090 SÃO PAULO - SP
BRAZIL
rosab@ime.usp.br

MÁRCIO F. DA SILVA
CENTRO DE MATEMÁTICA
COMPUTAÇÃO E COGNIÇÃO
UNIVERSIDADE FEDERAL DO ABC
RUA CATEQUESE, 242
09090-400 SANTO ANDRÉ - SP
BRAZIL
marcio.silva@ufabc.edu.br
http://sites.google.com/site/marcfab

RENATO H. L. PEDROSA
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
UNIVERSIDADE ESTADUAL DE CAMPINAS
PRAÇA SÉRGIO BUARQUE DE HOLANDA, 651
13083-859 CAMPINAS - SP
BRAZIL
pedrosa@ime.unicamp.br
http://www.ime.unicamp.br/~pedrosa/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

FOUR-DIMENSIONAL OSSERMAN METRICS OF NEUTRAL SIGNATURE

EDUARDO GARCÍA-RÍO, PETER B. GILKEY,
M. ELENA VÁZQUEZ-ABAL AND RAMÓN VÁZQUEZ-LORENZO

In the algebraic context, we show null Osserman, spacelike Osserman, and timelike Osserman are equivalent conditions for a model of signature (2, 2). We also classify the null Jordan Osserman models of signature (2, 2). In the geometric context, we show that a pseudo-Riemannian manifold with this signature is null Jordan Osserman if and only if either it has constant sectional curvature or it is locally a complex space form.

1. Introduction

Let $\mathcal{M} := (M, g)$ be a pseudo-Riemannian manifold. We say a tangent vector v is *spacelike*, *timelike*, or *null* if $g(v, v) > 0$, if $g(v, v) < 0$, or if $g(v, v) = 0$, respectively. Geometric properties derived from conditions on spacelike, timelike, and null vectors can have quite different meanings. For instance, the conditions of spacelike, timelike, and null geodesic completeness are nonequivalent and independent. Although spacelike and timelike conditions can sometimes become equivalent (for example, as they concern boundedness conditions on the sectional curvature), they can be quite different from similar null conditions, which are sometimes related to the conformal geometry of the manifold.

Let $R(x, y) := \nabla_x\nabla_y - \nabla_y\nabla_x - \nabla_{[x,y]}$ be the curvature operator of $\mathcal{M}$. The associated *Jacobi operator* $\mathcal{J}_R(x) : y \to R(y, x)x$ encodes much of the manifold's geometric information. The rescaling property $\mathcal{J}_R(\lambda v) = \lambda^2\mathcal{J}_R(v)$ plays a crucial role. Let $S^{\pm}(\mathcal{M})$ be the unit sphere bundles of spacelike and timelike unit tangent vectors in M, and let $N(\mathcal{M})$ be the null cone of nonzero null vectors. One says that $\mathcal{M}$ is spacelike Osserman if the eigenvalues of $\mathcal{J}_R$ are constant on $S^+(\mathcal{M})$; one says instead timelike if they are constant on $S^-(\mathcal{M})$. Normalizing the length

MSC2000: primary 53C20, 53C50; secondary 53B30.
Keywords: spacelike, timelike and null Jacobi operator, Osserman and Jordan Osserman metric, neutral signature (2, 2).
García-Río, Vázquez-Abal, and Vázquez-Lorenzo are supported by projects MTM2006-01432 and PGIDIT06PXIB207054PR (Spain). Gilkey is partially supported by the Max Planck Institute in the Mathematical Sciences (Leipzig, Germany) and by project MTM2006-01432 (Spain).

of the tangent vector to be ± 1 takes into account the above scaling of the Jacobi operator. Perhaps surprisingly, spacelike Osserman and timelike Osserman are equivalent conditions [García-Río et al. 1999; Gilkey 2001].

We say that $\mathcal{M}$ is *null Osserman* if the eigenvalues of $\mathcal{J}_R$ are constant on the null cone $N(\mathcal{M})$; with this definition, if $\mathcal{M}$ is null Osserman, then necessarily $\mathcal{J}_R(v)$ is nilpotent if $v \in N(\mathcal{M})$ and $\mathcal{J}_R(v)$ has only the eigenvalue 0. Any spacelike or timelike Osserman manifold is necessarily null Osserman; the converse can fail in general — see for example [García-Río et al. 1997] in the Lorentzian setting.

The Jordan normal form plays a crucial role in the higher signature setting — a self-adjoint linear transformation need not be determined by its eigenvalues if the metric in question is indefinite. One says that $\mathcal{M}$ is spacelike, timelike, or null Jordan Osserman if the Jordan normal form of $\mathcal{J}_R(\cdot)$ is constant on $S^+(\mathcal{M})$, on $S^-(\mathcal{M})$, or on $N(\mathcal{M})$, respectively. It is known from [Gilkey 2001; Gilkey and Ivanova 2002; 2001] that spacelike and timelike Jordan Osserman are inequivalent conditions; further neither necessarily implies the null Jordan Osserman condition.

In this paper, we concentrate on the 4-dimensional setting. Chi [1988] showed that any Riemannian Osserman 4-manifold is locally isometric to a 2-point homogeneous space; from later work [Blažić et al. 1997; García-Río et al. 1997], it follows that any Lorentzian 4-manifold has constant sectional curvature. However the situation is much more complicated in neutral signature (2, 2); there exist many examples of nonsymmetric Osserman pseudo-Riemannian manifolds of neutral signature — see [Díaz-Ramos et al. 2006b] and [García-Río et al. 1998]. Despite the results of [Alekseevsky et al. 1999; Blažić et al. 2001; Díaz-Ramos et al. 2006a; García-Río and Vázquez-Lorenzo 2001], it is still an open problem to completely describe 4-dimensional Osserman metrics of neutral signature.

It is convenient to work algebraically. Let V be a finite-dimensional real vector space that is equipped with a nondegenerate symmetric bilinear form $\langle\cdot,\cdot\rangle$ of signature (p,q). Let $A \in \otimes^4(V^*)$ be an algebraic curvature tensor on V, that is, a tensor that has the symmetries of the Riemann curvature tensor:

$$A(x,y,z,v) = -A(y,x,z,v) = A(z,v,x,y),$$
$$A(x,y,z,v) + A(y,z,x,v) + A(z,x,y,v) = 0.$$

This defines a model $\mathfrak{M} := (V, \langle\cdot,\cdot\rangle, A)$. We often prove results on the algebraic level (that is, for models), and then obtain corresponding conclusions in the geometric context. The notions spacelike unit vector, timelike unit vector, null vector, Jacobi operator, and so on extend naturally to this setting.

1.1. ***Null Osserman algebraic curvature tensors.*** Henceforth, suppose $\langle\cdot,\cdot\rangle$ is an inner product of signature (2, 2) on a 4-dimensional real vector space V. Fix an orientation of V, and let $\mathcal{B} = \{e_1, e_2, e_3, e_4\}$ be an oriented orthonormal basis for V, where e_1 and e_2 are timelike and where e_3 and e_4 are spacelike.

At the algebraic level, in signature $(2, 2)$ the conditions spacelike Osserman, timelike Osserman, spacelike Jordan Osserman and timelike Jordan Osserman are equivalent to the condition that $\mathfrak{M}$ is Einstein and self-dual with respect to a suitably chosen local orientation [Alekseevsky et al. 1999; García-Río et al. 2002]. In Section 2, we will establish Theorem 1.2, which shows that these conditions are also equivalent to null Osserman:

Theorem 1.2. *Let* $\mathfrak{M}$ *be a model of neutral signature* $(2, 2)$. *Then the following conditions are equivalent*:

(1) $\mathfrak{M}$ *is spacelike Osserman.*

(2) $\mathfrak{M}$ *is timelike Osserman.*

(3) $\mathfrak{M}$ *is spacelike Jordan Osserman.*

(4) $\mathfrak{M}$ *is timelike Jordan Osserman.*

(5) $\mathfrak{M}$ *is Einstein and self-dual for a suitably chosen local orientation.*

(6) $\mathfrak{M}$ *is null Osserman.*

Remark 1.3. The action of homothety on the null vectors is a central one in this subject. With our definition, it is immediate that $\mathfrak{M} = (V, \langle\cdot,\cdot\rangle, A)$ is null Osserman implies that 0 is the only eigenvalue of $\mathcal{J}_A$ on $N(V, \langle\cdot,\cdot\rangle)$. There is, although, an alternative, and different, formulation. One says that $\mathfrak{M}$ is *projectively null Osserman* if either $\mathfrak{M}$ is null Osserman or if given $0 \neq n_1, n_2 \in N(V, \langle\cdot,\cdot\rangle)$, there is a nonzero constant λ such that $\operatorname{Spec}(\mathcal{J}_A(n_1)) = \lambda \operatorname{Spec}(\mathcal{J}_A(n_2))$. We refer to [Brozos-Vázquez et al. 2008] for related work; we only introduce this concept for the sake of completeness as it plays no role in our development.

1.4. ***Null Jordan Osserman algebraic curvature tensors.*** Two algebraic curvature tensors will play a distinguished role. If Ψ is an skew-adjoint endomorphism of V, define the associated algebraic curvature tensor A^Ψ by setting

$$A^\Psi(x, y, z, v) := \langle\Psi y, z\rangle\langle\Psi x, v\rangle - \langle\Psi x, z\rangle\langle\Psi y, v\rangle - 2\langle\Psi x, y\rangle\langle\Psi z, v\rangle. \tag{1-1}$$

Such tensors span the linear space of all algebraic curvature tensors [Fiedler 2002].

The sectional curvature of a nondegenerate 2-plane $\pi = \operatorname{Span}\{x, y\}$ is given by

$$K_A(\pi) := \frac{A(x, y, y, x)}{\langle x, x\rangle\langle y, y\rangle - \langle x, y\rangle\langle x, y\rangle};$$

A has *constant sectional curvature* κ_0 if and only if $A = \kappa_0 A^0$, where A^0 is the algebraic curvature tensor of constant sectional curvature $+1$ defined by

$$A^0(x, y, z, v) := \langle y, z\rangle\langle x, v\rangle - \langle x, z\rangle\langle y, v\rangle. \tag{1-2}$$

We note that (1-1) and (1-2) imply that

$$\mathcal{J}_{A^\Psi}(x) : y \to 3\langle y, \Psi x\rangle \Psi x \quad \text{and} \quad \mathcal{J}_{A^0}(x) : y \to \langle x, x\rangle y - \langle x, y\rangle x. \tag{1-3}$$

Assume that Ψ is skew-adjoint. We say Ψ is an *orthogonal complex structure* if $\Psi^2 = -\operatorname{id}$ and say Ψ is an *adapted paracomplex structure* if $\Psi^2 = \operatorname{id}$. We say that a triple of skew-adjoint operators $\{\Psi_1, \Psi_2, \Psi_3\}$ is a *paraquaternionic structure* if $\Psi_1^2 = -\operatorname{id}$, $\Psi_2^2 = \operatorname{id}$, $\Psi_3^2 = \operatorname{id}$, and if $\Psi_i\Psi_j + \Psi_j\Psi_i = 0$ for $i \neq j$. We can define a paraquaternionic structure by setting

$$\begin{array}{llll} \Psi_1 e_1 = -e_2, & \Psi_1 e_2 = e_1, & \Psi_1 e_3 = e_4, & \Psi_1 e_4 = -e_3, \\ \Psi_2 e_1 = e_3, & \Psi_2 e_2 = e_4, & \Psi_2 e_3 = e_1, & \Psi_2 e_4 = e_2, \\ \Psi_3 e_1 = e_4, & \Psi_3 e_2 = -e_3, & \Psi_3 e_3 = -e_2, & \Psi_3 e_4 = e_1. \end{array} \tag{1-4}$$

Note that $\Psi_3 = \Psi_1\Psi_2$. If $\{\overline{\Psi}_1, \overline{\Psi}_2, \overline{\Psi}_3\}$ is another paraquaternionic structure on V, there is an isometry ϕ of V such that $\phi^*\overline{\Psi}_1 = \Psi_1$, $\phi^*\overline{\Psi}_2 = \Psi_2$, and $\phi^*\overline{\Psi}_3 = \pm\Psi_3$; this slight sign ambiguity plays no role in our constructions.

Let x be a spacelike or timelike vector. Then there is an orthogonal direct sum decomposition $V = \mathbb{R}x \oplus x^\perp$. Since $\mathcal{J}_A(x)x = 0$, $\mathcal{J}_A(x)$ preserves $x^\perp$. There are four different possibilities that describe the Jordan normal form of $\mathcal{J}_A(x)$ restricted to $x^\perp$ (see [Blažić et al. 2001; García-Río et al. 2002] for further details):

$$\underset{\text{Type Ia}}{\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}}, \quad \underset{\text{Type Ib}}{\begin{pmatrix} \alpha & -\beta & 0 \\ \beta & \alpha & 0 \\ 0 & 0 & \gamma \end{pmatrix}}, \quad \underset{\text{Type II}}{\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 1 & \alpha \end{pmatrix}}, \quad \underset{\text{Type III}}{\begin{pmatrix} \alpha & 0 & 0 \\ 1 & \alpha & 0 \\ 0 & 1 & \alpha \end{pmatrix}}. \tag{1-5}$$

Type Ia corresponds to a diagonalizable operator, Type Ib to an operator with a complex eigenvalue and Type II (respectively Type III) to a double (respectively triple) root of the minimal polynomial of the operator. If $\mathfrak{M}$ is spacelike, timelike, or null Osserman, then the Jordan normal form of $\mathcal{J}_A$ is constant on the spacelike and timelike unit vectors, and we classify A according to the four types above. In Section 3, we construct, up to isomorphism, all the spacelike Jordan Osserman algebraic curvature tensors and perform the analysis necessary to establish the following classification result:

Theorem 1.5. *Let $\mathfrak{M} := (V, \langle\cdot,\cdot\rangle, A)$ be a model of signature $(2, 2)$. Then $\mathfrak{M}$ is null Jordan Osserman if and only if A is of Type Ia and one of the following holds*:

(1) *There exists a constant κ_0 such that $A = \kappa_0 A^0$.*

(2) *There exists constants κ_0 and κ_J with $\kappa_J \neq 0$ such that $A = \kappa_0 A^0 + \kappa_J A^J$, where J is an orthogonal complex structure on V.*

(3) *There exists a constant $\kappa_P \neq 0$ such that $A = \kappa_P A^P$, where P is an adapted paracomplex structure on V.*

(4) *There exist constants* $\kappa_1, \kappa_2, \kappa_3$ *such that* $\kappa_2\kappa_3(\kappa_2+\kappa_1)(\kappa_3+\kappa_1) > 0$, *such that the associated eigenvalues* $\{3\kappa_1, -3\kappa_2, -3\kappa_3\}$ *are all distinct*, *and such that* $A = \kappa_1 A^{\Psi_1} + \kappa_2 A^{\Psi_2} + \kappa_3 A^{\Psi_3}$, *where* (Ψ_1, Ψ_2, Ψ_3) *is a paraquaternionic structure on* V.

Remark 1.6. The inequality $\kappa_2\kappa_3(\kappa_2+\kappa_1)(\kappa_3+\kappa_1) > 0$ is equivalent to the cross ratio satisfying

$$(0, \kappa_1, -\kappa_3, -\kappa_2) = \frac{\kappa_3(\kappa_2+\kappa_1)}{\kappa_2(\kappa_3+\kappa_1)} > 0.$$

Let $\mathbb{S}^2$ be the unit sphere in $\mathbb{R}^3$. This inequality is equivalent to the fact that the set of points $(0, -\kappa_3, -\kappa_2)$ and $(\kappa_1, -\kappa_3, -\kappa_2)$ give (via the stereographic projection) the corresponding circles in $\mathbb{S}^2$ the same orientation [Marden 2007].

1.7. ***Null Jordan Osserman manifolds.*** We characterize those neutral signature 4-manifolds that are null Jordan Osserman; null Osserman and null Jordan Osserman are not equivalent conditions, as the analysis of Section 3.7 shows. We say that $\mathcal{M}$ is locally a complex space form if it is an indefinite Kähler manifold of constant holomorphic sectional curvature. In Section 4, we will use Theorem 1.5 to establish the following geometric result:

Theorem 1.8. *If* $\mathcal{M}$ *is a connected pseudo-Riemannian manifold of neutral signature* $(2, 2)$, *then* $\mathcal{M}$ *is null Jordan Osserman if and only if either* $\mathcal{M}$ *has constant sectional curvature or* $\mathcal{M}$ *is locally a complex space form.*

Remark 1.9. There is another family of Osserman 4-manifolds with diagonalizable Jacobi operator, namely, the paracomplex space forms [Blažić et al. 2001]. Although the geometry of complex and paracomplex space forms is very similar, the Jordan–Osserman condition distinguishes them. To our knowledge, this is the first algebraic curvature condition that distinguishes these two geometries.

2. Null Osserman models of signature (2, 2)

We work in the algebraic context to prove Theorem 1.2. Here is a brief outline to this section. Previous work establishes that parts (1)–(5) are equivalent. In Section 2.1, we introduce notation and show that spacelike Osserman models are null Osserman and that null Osserman models are Einstein. Thus to complete the proof, it suffices to show null Osserman models are self-dual or anti-self-dual. In Section 2.3, we examine Einstein models. Lemma 2.4 describes the Weyl curvature operators in that setting, and Lemma 2.5 gives an alternative characterization of self-duality for an Einstein model. We use Lemma 2.5 to complete the proof of Theorem 1.2 in Section 2.6.

2.1. ***Notation.*** Let $\mathfrak{M} := (V, \langle\cdot,\cdot\rangle, A)$ be a neutral signature 4-dimensional model. We use the inner product to raise indices and to define an associated Jacobi operator $\mathcal{J}_A$, which is characterized by the identity $\langle \mathcal{J}_A(x)y, z\rangle = A(y,x,x,z)$. Let $\mathcal{B} = \{e_1, e_2, e_3, e_4\}$ be an oriented orthonormal basis for V as in Section 1.1. Let $g_{ij} := \langle e_i, e_j\rangle$, and let g^{ij} be the inverse matrix. The associated *Ricci tensor* ρ_A, the *scalar curvature* τ_A, and the *Weyl tensor* W_A are then defined by setting

$$\begin{aligned}\rho_A(x,y) &:= \textstyle\sum_{i,j=1}^4 g^{ij} A(e_i, x, y, e_j), \qquad \tau_A := \sum_{i,j=1}^4 g^{ij}\rho_A(e_i,e_j),\\ W_A(x,y,z,v) &:= A(x,y,z,v) + \tfrac{1}{6}\tau_A(\langle y,z\rangle\langle x,v\rangle - \langle x,z\rangle\langle y,v\rangle)\\ &\qquad - \tfrac{1}{2}\big(\rho_A(y,z)\langle x,v\rangle - \rho_A(x,z)\langle y,v\rangle\\ &\qquad\quad + \rho_A(x,v)\langle y,z\rangle - \rho_A(y,v)\langle x,z\rangle\big).\end{aligned}$$

Let $A_{ijkl} = A^{\mathcal{B}}_{ijkl} := A(e_i, e_j, e_k, e_l)$ denote the components of A with respect to $\mathcal{B}$, where $1 \le i, j, k, l \le 4$; we drop the dependence on $\mathcal{B}$ from the notation when there is no danger of confusion. Let $\{e^1, \dots, e^4\}$ be the dual basis for V^*. The Hodge operator $\star : \Lambda^p(V^*) \to \Lambda^{4-p}(V^*)$ is characterized by the identity

$$\phi_p \wedge \star\theta_p = \langle \phi_p, \theta_p\rangle e^1\wedge e^2\wedge e^3\wedge e^4.$$

Thus,

$$\begin{aligned}&\star(e^1\wedge e^2) = e^3\wedge e^4, \quad \star(e^1\wedge e^3) = e^2\wedge e^4, \quad \star(e^1\wedge e^4) = -e^2\wedge e^3,\\ &\star(e^2\wedge e^3) = -e^1\wedge e^4, \quad \star(e^2\wedge e^4) = e^1\wedge e^3, \quad \star(e^3\wedge e^4) = e^1\wedge e^2.\end{aligned}$$

A crucial feature of 4-dimensional geometry now enters. Since $\star^2 = \mathrm{id}$, the Hodge star induces a splitting $\Lambda^2(V^*) = \Lambda^+ \oplus \Lambda^-$ of the space of 2-forms, where

$$\Lambda^+ = \{\alpha \in \Lambda^2 : \star\alpha = \alpha\} \quad \text{and} \quad \Lambda^- = \{\alpha\in\Lambda^2 : \star\alpha = -\alpha\}$$

denote the spaces of *self-dual* and *anti-self-dual* two-forms. We have orthonormal bases $\{E_1^\mp, E_2^\mp, E_3^\mp\}$ for $\Lambda^\mp$ that are given by

$$\begin{aligned}&E_1^\mp = \tfrac{1}{\sqrt{2}}(e^1\wedge e^2 \mp e^3\wedge e^4), \quad E_2^\mp = \tfrac{1}{\sqrt{2}}(e^1\wedge e^3 \mp e^2\wedge e^4),\\ &E_3^\mp = \tfrac{1}{\sqrt{2}}(e^1\wedge e^4 \pm e^2\wedge e^3),\end{aligned}$$

where the induced inner product on $\Lambda^\mp$ has signature $(2,1)$:

$$\langle E_1^\mp, E_1^\mp\rangle = 1, \quad \langle E_2^\mp, E_2^\mp\rangle = -1, \quad \langle E_3^\mp, E_3^\mp\rangle = -1.$$

Let $W_A^\mp$ be the restriction of W_A to the spaces $\Lambda^\mp$, that is, $W_A^\mp : \Lambda^\mp \to \Lambda^\mp$, where W_A also stands for the associated Weyl curvature operator on Λ^2. We say $\mathfrak{M}$ is *self-dual* if $W_A^- = 0$ and *anti-self-dual* if $W_A^+ = 0$.

Lemma 2.2. *Let* $\mathfrak{M} = (V, \langle\cdot,\cdot\rangle, A)$ *be a model of signature* $(2,2)$.

(1) *If $\mathfrak{M}$ is spacelike Osserman, then $\mathfrak{M}$ is null Osserman.*

(2) *If $\mathfrak{M}$ is null Osserman, then $\mathfrak{M}$ is Einstein.*

Proof. Suppose first $\mathfrak{M}$ is spacelike Osserman. Set $T_j(v) := \operatorname{Tr}\{\mathcal{J}_A(v)^j\}$. Since the eigenvalues of $\mathcal{J}_A$ are constant on $S^+(V, \langle\cdot,\cdot\rangle)$, there are constants c_j such that $T_j(v) = c_j$ for $v \in S^+(V, \langle\cdot,\cdot\rangle)$. It follows since $T_j(\lambda v) = \lambda^{2j} T_j(v)$ that $T_j(v) = c_j \langle v, v\rangle^j$ for v spacelike. Since the spacelike vectors form an open subset of V, this polynomial identity holds for all $v \in V$. Thus, in particular, $T_j(v) = 0$ if $v \in N(V, \langle\cdot,\cdot\rangle)$. This implies that 0 is the only eigenvalue of $\mathcal{J}_A(v)$ and shows $\mathfrak{M}$ is null Osserman.

Now suppose $\mathfrak{M}$ is null Osserman. Let s_1 and s_2 be spacelike unit vectors. We may choose a unit timelike vector t that is perpendicular to s_1 and s_2. Let $n_i^\pm := s_i \pm t$ be null vectors. Thus $0 = \operatorname{Tr}(\mathcal{J}_A(n_i^\pm)) = \rho_A(n_i^\pm, n_i^\pm)$, and

$$0 = \rho_A(s_i \pm t, s_i \pm t) = \rho_A(s_i, s_i) + \rho_A(t,t) \pm 2\rho_A(s_i, t).$$

This implies $\rho_A(s_i, t) = 0$ and $\rho_A(s_i, s_i) + \rho_A(t,t) = 0$; in particular, one has $\rho_A(s_1, s_1) = -\rho_A(t,t) = \rho_A(s_2, s_2)$. Therefore, after rescaling, there is a constant c such that $\rho_A(s,s) = c\langle s, s\rangle$ for every spacelike vector s; this polynomial identity then continues to hold for all $s \in V$. Polarizing this identity then yields $\rho_A = c\langle\cdot,\cdot\rangle$, and hence $\mathfrak{M}$ is Einstein. $\square$

2.3. ***The Weyl tensor for an Einstein algebraic curvature tensor.*** Let

$$\begin{aligned}
\sigma_1 &= 2A_{1212} + 3\varepsilon A_{1234} + A_{1313} + A_{1414},\\
\sigma_2 &= A_{1212} + 2A_{1313} + 3\varepsilon A_{1324} - A_{1414},\\
\sigma_3 &= A_{1212} + 3\varepsilon A_{1234} - A_{1313} - 3\varepsilon A_{1324} + 2A_{1414}.
\end{aligned}$$

Then we have an immediate lemma:

Lemma 2.4. *If $\mathfrak{M}$ is Einstein, then the self-dual Weyl curvature operator W_A^+ (in which $\varepsilon = 1$) and the anti-self-dual Weyl curvature operator W_A^- (in which $\varepsilon = -1$) are given by*

$$\begin{pmatrix} \sigma_1/3 & A_{1213} + \varepsilon A_{1224} & A_{1214} - \varepsilon A_{1223}\\ -A_{1213} - \varepsilon A_{1224} & -\sigma_2/3 & -A_{1314} + \varepsilon A_{1323}\\ -A_{1214} + \varepsilon A_{1223} & -A_{1314} + \varepsilon A_{1323} & -\sigma_3/3 \end{pmatrix}.$$

The next observation is of interest in its own right:

Lemma 2.5. *If $\mathfrak{M}$ is Einstein, then the model $\mathfrak{M}$ is anti-self-dual if and only if $A^{\mathcal{B}}_{1214} - A^{\mathcal{B}}_{1223} = 0$ for every oriented orthonormal frame $\mathcal{B}$.*

Proof. If $\mathfrak{M}$ is anti-self-dual, we set $\varepsilon = 1$ in Lemma 2.4 to see $A^{\mathcal{B}}_{1214} - A^{\mathcal{B}}_{1223} = 0$. Conversely, suppose $A^{\mathcal{B}}_{1214} - A^{\mathcal{B}}_{1223} = 0$ for every $\mathcal{B}$. Define a new basis $\tilde{\mathcal{B}}$ by setting $\tilde{e}_1 = e_1,\ \tilde{e}_2 = e_2,\ \tilde{e}_3 = e_4$, and $\tilde{e}_4 = -e_3$. We then have

$$0 = -A^{\tilde{\mathscr{B}}}_{1214} + A^{\tilde{\mathscr{B}}}_{1223} = A^{\mathscr{B}}_{1213} + A^{\mathscr{B}}_{1224}.$$

Next, define $\tilde{\mathscr{B}}$ by setting $\tilde{e}_1 = e_1,\ \tilde{e}_2 = \cosh\theta e_2 + \sinh\theta e_3,\ \tilde{e}_3 = \sinh\theta e_2 + \cosh\theta e_3$, and $\tilde{e}_4 = e_4$. This yields the relation

$$0 = -A^{\tilde{\mathscr{B}}}_{1214} + A^{\tilde{\mathscr{B}}}_{1223} = \cosh\theta(-A^{\mathscr{B}}_{1214} + A^{\mathscr{B}}_{1223}) + \sinh\theta(-A^{\mathscr{B}}_{1314} + A^{\mathscr{B}}_{1323}).$$

This shows $-A^{\mathscr{B}}_{1314} + A^{\mathscr{B}}_{1323} = 0$. Thus, by Lemma 2.4,

$$W_A^+ = \frac{1}{3}\begin{pmatrix} \sigma_1^{\mathscr{B}} & 0 & 0 \\ 0 & -\sigma_2^{\mathscr{B}} & 0 \\ 0 & 0 & -\sigma_3^{\mathscr{B}} \end{pmatrix}.$$

Setting the $\tilde{e}_i$ as before yields bases for $\Lambda^\pm$ in the form

$$\tilde{E}_1^\pm = \cosh\theta E_1^\pm + \sinh\theta E_2^\pm, \quad \tilde{E}_2^\pm = \cosh\theta E_2^\pm + \sinh\theta E_1^\pm, \quad \tilde{E}_3^\pm = E_3^\pm.$$

We may compute

$$\begin{aligned} W_A^+ \tilde{E}_1^+ &= \sigma_1^{\tilde{\mathscr{B}}} \tilde{E}_1^+ = \sigma_1^{\tilde{\mathscr{B}}}(\cosh\theta E_1^+ + \sinh\theta E_2^+) \\ &= W_A^+(\cosh\theta E_1^+ + \sinh\theta E_2^+) = \sigma_1^{\mathscr{B}} \cosh\theta E_1^+ - \sigma_2^{\mathscr{B}} \sinh\theta E_2^+. \end{aligned}$$

This shows $\sigma_1^{\tilde{\mathscr{B}}} = \sigma_1^{\mathscr{B}} = -\sigma_2^{\mathscr{B}}$. A similar argument applied to the basis $\tilde{e}_1 = e_1$, $\tilde{e}_2 = \cosh\theta e_2 + \sinh\theta e_4,\ \tilde{e}_3 = e_3$, and $\tilde{e}_4 = \sinh\theta e_2 + \cosh\theta e_4$ yields $\sigma_1^{\mathscr{B}} = -\sigma_3^{\mathscr{B}}$. Since $\sigma_1^{\mathscr{B}} - \sigma_2^{\mathscr{B}} - \sigma_3^{\mathscr{B}} = 0$, it now follows that $W_A^+ = 0$. □

2.6. *Proof of Theorem 1.2.* Let $\mathfrak{M}$ be a null Osserman model. By Lemma 2.2, $\mathfrak{M}$ is Einstein. We complete the proof of Theorem 1.2 by showing $\mathfrak{M}$ is self-dual or anti-self-dual. Suppose the contrary and argue for a contradiction. As $\mathfrak{M}$ is null Osserman, $\mathscr{J}_A$ is nilpotent, so the characteristic polynomial has $p_\lambda(\mathscr{J}_A(u)) = \lambda^4$. Let

$$\begin{aligned} \mathscr{E}_1 &:= A_{1212} + 2A_{1214} - 2A_{1223} + 2A_{1234} - A_{1324} + A_{1414}, \\ Q(a,b) &:= (A_{1212} - 2A_{1214} - 2A_{1223} - 2A_{1234} + A_{1324} + A_{1414})a^4 \\ &\quad + (A_{1212} + 2A_{1214} + 2A_{1223} - 2A_{1234} + A_{1324} + A_{1414})b^4 \\ &\quad + 2(A_{1212} + 2A_{1313} - 3A_{1324} - A_{1414})a^2b^2 \\ &\quad + 4(A_{1213} - A_{1224} - A_{1314} - A_{1323})a^3b \\ &\quad + 4(A_{1213} - A_{1224} + A_{1314} + A_{1323})ab^3. \end{aligned}$$

If we take $u = ae_1 + be_2 + ae_3 + be_4$, then $\lambda^4 = p_\lambda(\mathscr{J}_A(u)) = \lambda^2(\lambda^2 - Q(a,b)\mathscr{E}_1)$. As $p_\lambda(\mathscr{J}_A(u)) = \lambda^4$, either $Q(a,b) = 0$ or $\mathscr{E}_1 = 0$. If we suppose that $\mathscr{E}_1 \neq 0$, then

$Q(a,b)$ vanishes identically for all a, b. This leads to the relations

$$A_{1213} - A_{1224} = 0, \quad A_{1214} + A_{1223} = 0, \quad A_{1314} + A_{1323} = 0,$$
$$A_{1234} + A_{1313} - 2A_{1324} - A_{1414} = 0, \qquad A_{1212} + 2A_{1313} - 3A_{1324} - A_{1414} = 0.$$

From this, we see that the matrix in Lemma 2.4 vanishes for $\varepsilon = -1$. This means that the anti-self-dual Weyl curvature operator W_A^- vanishes, so $\mathfrak{M}$ is self-dual, which is a contradiction. Thus for *any* oriented orthonormal frame, we have

$$0 = A_{1212} + 2A_{1214} - 2A_{1223} + 2A_{1234} - A_{1324} + A_{1414}. \tag{2-1}$$

Setting $\tilde{e}_1 = -e_1$, $\tilde{e}_2 = e_2$, $\tilde{e}_3 = e_3$, and $\tilde{e}_4 = -e_4$ yields

$$0 = A_{1212} - 2A_{1214} + 2A_{1223} + 2A_{1234} - A_{1324} + A_{1414}. \tag{2-2}$$

Subtracting (2-2) from (2-1) then yields the relation $0 = -A_{1214} + A_{1223}$. We may now use Lemma 2.5 to complete the proof of Theorem 1.2. □

3. Proof of Theorem 1.5

Here is a brief outline of this section. In Section 3.1, we construct, up to isomorphism, all spacelike Jordan Osserman models of signature $(2,2)$. In the remainder of Section 3, we analyze each possible Jordan normal form in some detail using the classification of (1-5). Sections 3.5–3.8 deal with Type Ia models. In Section 3.5 we study the case when all the eigenvalues are equal; this gives rise to Theorem 1.5(1). In Section 3.6, we study the case of two equal spacelike eigenvalues, and in Section 3.7, we study equal timelike and spacelike eigenvalues; these involve parts (2) and (3) of Theorem 1.5, respectively. In Section 3.8, we study Type Ia models with distinct eigenvalues; this leads to Theorem 1.5(4). We complete the proof of Theorem 1.5 by showing the remaining types do not give rise to null Jordan Osserman models. We study Type Ib models in Section 3.9, Type II models in Section 3.10, and Type III models in Section 3.11.

3.1. ***Spacelike Jordan Osserman models.*** We use the ansatz from [Gilkey and Ivanova 2001]. Let $\{\Psi_1, \Psi_2, \Psi_3\}$ be the paraquaternionic structure given in (1-4). Let $\xi_{ij} \in \mathbb{R}$ for $1 \le i \le j \le 3$, and let $\kappa_0 \in \mathbb{R}$ be given. Let

$$\begin{aligned} A_{\kappa_0,\xi} &:= \kappa_0 A^0 + \tfrac{1}{3}\xi_{11} A^{\Psi_1} + \tfrac{1}{3}\xi_{22} A^{\Psi_2} + \tfrac{1}{3}\xi_{33} A^{\Psi_3} \\ &\qquad + \tfrac{1}{3}\xi_{12} A^{\Psi_1+\Psi_2} + \tfrac{1}{3}\xi_{13} A^{\Psi_1+\Psi_3} + \tfrac{1}{3}\xi_{23} A^{\Psi_2+\Psi_3}, \\ \mathcal{J}_{\kappa_0,\xi} &:= \kappa_0 \,\mathrm{id} + \begin{pmatrix} \xi_{11}+\xi_{12}+\xi_{13} & -\xi_{12} & -\xi_{13} \\ \xi_{12} & -\xi_{22}-\xi_{12}-\xi_{23} & -\xi_{23} \\ \xi_{13} & -\xi_{23} & -\xi_{33}-\xi_{13}-\xi_{23} \end{pmatrix}. \end{aligned} \tag{3-1}$$

Lemma 3.2. *Adopt the notation established above. Let* $\mathfrak{M}_{\kappa_0,\xi} := (V, \langle\cdot,\cdot\rangle, A_{\kappa_0,\xi})$.

(1) *If $x \in S^{\pm}(V, \langle\cdot,\cdot\rangle)$, then $\mathcal{J}_{A_{\kappa_0,\xi}}(x)$ is conjugate to the matrix $\pm\mathcal{J}_{\kappa_0,\xi}$.*

(2) *The model $\mathfrak{M}_{\kappa_0,\xi}$ is spacelike and timelike Jordan Osserman.*

(3) *Let $\mathfrak{M}_i = (V, \langle\cdot,\cdot\rangle, A_i)$ be spacelike Osserman models of signature $(2, 2)$. If $\mathcal{J}_{A_1}(x)$ is conjugate to $\mathcal{J}_{A_2}(x)$ for some $x \in S^{\pm}(V, \langle\cdot,\cdot\rangle)$, then there exists an isometry ϕ of $(V, \langle\cdot,\cdot\rangle)$ such that $\phi^* A_2 = A_1$.*

Remark 3.3. Any self-adjoint map of a signature $(2, 1)$ vector space is conjugate to $\mathcal{J}_{\kappa_0,\xi}$ for some $\{\kappa_0, \xi\}$, so every spacelike Osserman model of signature $(2, 2)$ is isomorphic to one given by (3-1).

Proof. We suppose x is a spacelike unit vector; the timelike case is similar. Let $f_1 := \Psi_1 x$, $f_2 := \Psi_2 x$, and $f_3 := \Psi_3 x$. Then $\{f_1, f_2, f_3\}$ is an orthonormal basis of signature $(+,-,-)$ for $x^\perp$. Let $\mathcal{J} := \mathcal{J}_{A_{\kappa_0,\xi}}(x)$. We use (1-3) to see that

$$\begin{aligned}
\mathcal{J} f_1 &= (\kappa_0 + \xi_{11} + \xi_{12} + \xi_{13}) f_1 + \xi_{12} f_2 + \xi_{13} f_3,\\
\mathcal{J} f_2 &= -\xi_{12} f_1 + (\kappa_0 - \xi_{22} - \xi_{12} - \xi_{23}) f_2 - \xi_{23} f_3,\\
\mathcal{J} f_3 &= -\xi_{13} f_1 - \xi_{23} f_2 + (\kappa_0 - \xi_{33} - \xi_{13} - \xi_{23}) f_3.
\end{aligned}$$

Part (1) now follows; part (2) follows from part (1). Suppose that $\mathfrak{M}$ is a Type Ia spacelike Osserman model, so $\mathcal{J}_A(x) = \operatorname{diag}[\alpha, \beta, \gamma]$ for any x in $S^+(V, \langle\cdot,\cdot\rangle)$; choose the notation so $\operatorname{Ker}(\mathcal{J}_A(x) - \alpha\,\mathrm{id})$ is spacelike. It then follows from the discussion in [Blažić et al. 2001; García-Río et al. 2002] that there is an orthonormal basis $\mathcal{B}$ such that the nonzero components of the curvature tensor are

$$\begin{aligned}
&A_{1221} = A_{4334} = \alpha, && A_{1331} = A_{2442} = -\beta,\\
&A_{1441} = A_{3223} = -\gamma, && A_{1234} = (-2\alpha + \beta + \gamma)/3,\\
&A_{1423} = (\alpha + \beta - 2\gamma)/3, && A_{1342} = (\alpha - 2\beta + \gamma)/3.
\end{aligned}$$

Similar forms exist for the other types of (1-5). Thus the Jordan normal form of $\mathcal{J}_A(x)$ determines A up to the action of $\mathrm{O}(2, 2)$. Part (3) follows. □

We immediately have this:

Lemma 3.4. *A null Osserman model $\mathfrak{M}$ of signature $(2, 2)$ is null Jordan Osserman if and only if the functions $\operatorname{Rank}\{\mathcal{J}_A(\cdot)\}$ and $\operatorname{Rank}\{\mathcal{J}_A(\cdot)^2\}$ are constant on $N(V, \langle\cdot,\cdot\rangle)$.*

3.5. *Type Ia with all eigenvalues equal: $\alpha = \beta = \gamma$.* We set $A = \kappa_0 A^0$. By Lemma 3.2, the Jordan normal form is given by $\operatorname{diag}[\kappa_0, \kappa_0, \kappa_0]$. If v belongs to $N(V, \langle\cdot,\cdot\rangle)$, then $\mathcal{J}_A(v)y = -\kappa_0 \langle v, y\rangle v$, and hence $\mathfrak{M}$ is null Jordan Osserman.

3.6. ***Type Ia with two equal spacelike eigenvalues: $\beta = \gamma$ and $\alpha \neq \beta$.*** Let J be an orthogonal almost complex structure on V, and let $A = \kappa_0 A^0 + \kappa_J A^J$. The Jordan normal form is then given by $\mathrm{diag}[\kappa_0 + 3\kappa_J, \kappa_0, \kappa_0]$, which has the desired form for suitably chosen κ_0 and κ_J with $\kappa_J \neq 0$. Let $v \in N(V, \langle \cdot, \cdot \rangle)$. We have

$$\mathcal{J}_A(v)y = -\kappa_0 \langle v, y\rangle v + 3\kappa_J \langle y, Jv\rangle Jv.$$

Because $J^2 = -\,\mathrm{id}$, the vectors v and Jv are linearly independent. We note that $\langle v, v\rangle = \langle v, Jv\rangle = \langle Jv, Jv\rangle = 0$. Consequently $\mathcal{J}_A(v)v = \mathcal{J}_A(v)Jv = 0$. Since $v^\perp$ and $Jv^\perp$ are distinct 3-dimensional subspaces, we can choose y so $\langle v, y\rangle = 1$ and $\langle Jv, y\rangle = 0$. It now follows that $\mathcal{J}_A(v)y = -\kappa_0 v$, while $\mathcal{J}_A(v)Jy = 3\kappa_J Jv$. Thus $\mathcal{J}_A(v)$ has rank 2 and $\mathcal{J}_A(v)^2 = 0$. This implies A is null Jordan Osserman.

3.7. ***Type Ia with equal timelike and spacelike eigenvalues: $\alpha = \beta$ and $\beta \neq \gamma$.*** Let $A = \kappa_0 A^0 + \kappa_P A^P$, where $\kappa_P \neq 0$ and where P is an adapted paracomplex structure; the Jordan normal form is then given by $\mathrm{diag}[\kappa_0, \kappa_0 - 3\kappa_P, \kappa_0]$, which has the desired form for suitably chosen parameters. If $v \in N(V, \langle \cdot, \cdot \rangle)$, then

$$\mathcal{J}_A(v)y = -\kappa_0 \langle v, y\rangle v + 3\kappa_P \langle y, Pv\rangle Pv.$$

If $\kappa_0 = 0$, $\mathfrak{M}$ is null Jordan Osserman. Suppose $\kappa_0 \neq 0$. If $v = e_1 + Pe_1$, then $Pv = v$, so $\mathrm{Rank}\{\mathcal{J}_A(v)\} \leq 1$. On the other hand, if $v = e_1 + e_4$, then v and Pv are linearly independent, so $\mathrm{Rank}\{\mathcal{J}_A(v)\} = 2$ and $\mathfrak{M}$ is not null Jordan Osserman.

3.8. ***Type Ia with three distinct eigenvalues.*** We set $A := \sum_i \kappa_i A^{\Psi_i}$, where the triple $\{\Psi_1, \Psi_2, \Psi_3\}$ is the paraquaternionic structure of (1-4); the Jordan normal form is given by $\mathrm{diag}[3\kappa_1, -3\kappa_2, -3\kappa_3]$, which has the desired form for suitably chosen parameters with

$$\kappa_1 + \kappa_2 \neq 0, \quad \kappa_1 + \kappa_3 \neq 0, \quad \kappa_2 - \kappa_3 \neq 0.$$

Let $\tilde{e} \in S^+(V, \langle \cdot, \cdot \rangle)$, let $V_+ := \mathrm{Span}\{\tilde{e}, \Psi^1 \tilde{e}\}$, and let $V_- = V_+^\perp = \mathrm{Span}\{\Psi_2 \tilde{e}, \Psi_3 \tilde{e}\}$. We then have an orthogonal direct sum decomposition $V = V_- \oplus V_+$, where V_+ is spacelike and V_- is timelike. Decompose $v \in N(V, \langle \cdot, \cdot \rangle)$ as $v = \lambda(e_+ + e_-)$, where $e_\pm \in V_\pm$. Let $\mathfrak{M}$ be spacelike Osserman. We have $\mathcal{J}_A(v) = \lambda^2 \mathcal{J}_A(e_+ + e_-)$. Since $\mathcal{J}_A(v)$ is nilpotent, $\mathcal{J}_A(v)$ and $\mathcal{J}_A(e_+ + e_-)$ have the same Jordan normal form. Thus we may safely take $\lambda = 1$, so $v = e_+ + e_-$. Set $e = e_+$ and expand $e_- = \cos\theta \Psi_2 e + \sin\theta \Psi_3 e$. This expresses

$$v = e + \cos\theta \Psi_2 e + \sin\theta \Psi_3 e \quad \text{for } e \in S^+(V, \langle \cdot, \cdot \rangle).$$

We use the relations $\Psi_1\Psi_2 = \Psi_3$, $\Psi_1\Psi_3 = -\Psi_2$, and $\Psi_2\Psi_3 = -\Psi_1$ to see that

$$\begin{aligned}
\Psi_1 v &= 0 + \Psi_1 e - \sin\theta\Psi_2 e + \cos\theta\Psi_3 e,\\
\Psi_2 v &= \cos\theta e - \sin\theta\Psi_1 e + \Psi_2 e + 0,\\
\Psi_3 v &= \sin\theta e + \cos\theta\Psi_1 e + 0 + \Psi_3 e,
\end{aligned} \tag{3-2}$$

so that $0 = \Psi_1 v + \sin\theta\Psi_2 v - \cos\theta\Psi_3 v$. Thus the vectors $\{\Psi_1 v, \Psi_2 v, \Psi_3 v\}$ span a 2-dimensional subspace. Since $\langle\Psi_i v, \Psi_j v\rangle = 0$, $\mathrm{Span}\{\Psi_i v\} \subset \mathrm{Ker}\{\mathcal{J}_A(v)\}$. Since $\mathrm{Range}\{\mathcal{J}_A(v)\} \subset \mathrm{Span}\{\Psi_i v\}$,

$$\mathrm{Rank}\{\mathcal{J}_A(v)\} \le 2 \quad \text{and} \quad \mathcal{J}_A(v)^2 = 0.$$

Note that $\{e, \Psi_1 e, \Psi_2 v, \Psi_3 v\}$ is a basis for V. Let π_+ denote orthogonal projection on $V_+ = \mathrm{Span}\{e, \Psi_1 e\}$. Since π_+ is injective on $\mathrm{Range}\{\mathcal{J}_A(v)\} \subset \mathrm{Span}\{\Psi_2 v, \Psi_3 v\}$,

$$r(v) := \dim\mathrm{Range}\{\mathcal{J}_A(v)\} = \dim(\mathrm{Span}\{\pi_+\mathcal{J}_A(v)e, \pi_+\mathcal{J}_A(v)\Psi_1 e\}).$$

By (3-2) and the linear dependency it contains,

$$\begin{aligned}
\mathcal{J}_A(v)e &= 3\kappa_2\cos\theta\Psi_2 v + 3\kappa_3\sin\theta\Psi_3 v,\\
\mathcal{J}_A(v)\Psi_1 e &= 3\kappa_1\Psi_1 v - 3\kappa_2\sin\theta\Psi_2 v + 3\kappa_3\cos\theta\Psi_3 v,\\
\pi_+\mathcal{J}_A(v)e &= 3(\kappa_2\cos\theta(\cos\theta) + \kappa_3\sin\theta(\sin\theta))e\\
&\quad + 3(\kappa_2\cos\theta(-\sin\theta) + \kappa_3\sin\theta(\cos\theta))\Psi_1 e,\\
\pi_+\mathcal{J}_A(v)\Psi_1 e &= 3(-\kappa_2\sin\theta(\cos\theta) + \kappa_3\cos\theta(\sin\theta))e\\
&\quad + 3(\kappa_1 - \kappa_2\sin\theta(-\sin\theta) + \kappa_3\cos\theta(\cos\theta))\Psi_1 e.
\end{aligned}$$

This leads to a coefficient matrix for $\pi_+\mathcal{J}_A(v)$ on V_+ given by

$$\mathscr{C}_A(\theta) = 3\begin{pmatrix} \kappa_2\cos^2\theta + \kappa_3\sin^2\theta & (-\kappa_2 + \kappa_3)\sin\theta\cos\theta \\ (-\kappa_2 + \kappa_3)\sin\theta\cos\theta & \kappa_1 + \kappa_2\sin^2\theta + \kappa_3\cos^2\theta \end{pmatrix}.$$

We compute

$$\begin{aligned}
\tfrac{1}{9}\det(\mathscr{C}_A)(\theta) &= \kappa_1\kappa_2\cos^2\theta + \kappa_2^2\cos^2\theta\sin^2\theta + \kappa_2\kappa_3\cos^4\theta\\
&\quad + \kappa_1\kappa_3\sin^2\theta + \kappa_2\kappa_3\sin^4\theta + \kappa_3^2\sin^2\theta\cos^2\theta\\
&\quad - \kappa_2^2\sin^2\theta\cos^2\theta - \kappa_3^2\sin^2\theta\cos^2\theta + 2\kappa_2\kappa_3\sin^2\theta\cos^2\theta\\
&= \kappa_1\kappa_2\cos^2\theta + \kappa_1\kappa_3\sin^2\theta + \kappa_2\kappa_3\\
&= (\kappa_1 + \kappa_3)\kappa_2\cos^2\theta + (\kappa_1 + \kappa_2)\kappa_3\sin^2\theta.
\end{aligned}$$

Observe that $\kappa_2\kappa_3 = 0$ implies that $\det(\mathscr{C}_A)(\theta)$ vanishes for some θ, and thus $\mathfrak{M}$ is not null Jordan Osserman. Hence, since $(\kappa_1 + \kappa_3)\kappa_2$ and $(\kappa_1 + \kappa_2)\kappa_3$ are nonzero, $\det(\mathscr{C}_A)(\theta)$ never vanishes, or equivalently $\mathfrak{M}$ is null Jordan Osserman, if and only if these two real numbers have the same sign, that is, $\kappa_2\kappa_3(\kappa_1 + \kappa_3)(\kappa_1 + \kappa_2) > 0$.

3.9. ***Type Ib models.*** Let $b \neq 0$. We take a curvature tensor of the form

$$A = \tfrac{1}{3}((a-b)A^{\Psi_1} + (-b-a)A^{\Psi_2} + bA^{\Psi_1+\Psi_2} + cA^{\Psi_3}).$$

Proceeding as in the previous case, we have for any $e \in S^+(V, \langle \cdot, \cdot \rangle)$ that

$$\begin{aligned} \mathcal{J}_A(x)y &= \langle (a\Psi_1 + b\Psi_2)x, y\rangle \Psi_1 x + \langle (b\Psi_1 - a\Psi_2)x, y\rangle \Psi_2 x + c\langle \Psi_3 x, y\rangle \Psi_3 x, \\ \mathcal{J}_A(e)\Psi_1 e &= a\Psi_1 e + b\Psi_2 e,\ \mathcal{J}_A(e)\Psi_2 e = -b\Psi_1 e + a\Psi_2 e,\ \mathcal{J}_A(e)\Psi_3 e = -c\Psi_3 e. \end{aligned}$$

Thus $\mathfrak{M} := (V, \langle \cdot, \cdot \rangle, A)$ is Type Ib and any Type Ib model is isomorphic to $\mathfrak{M}$ for suitably chosen parameters. As in Section 3.8, put $v = e + \cos\theta \Psi_2 e + \sin\theta \Psi_3 e$. We compute

$$\begin{aligned} \mathcal{J}_A(v)e &= b\cos\theta \Psi_1 v - a\cos\theta \Psi_2 v + c\sin\theta \Psi_3 v, \\ \mathcal{J}_A(v)\Psi_1 e &= (a - b\sin\theta)\Psi_1 v + (b + a\sin\theta)\Psi_2 v + c\cos\theta \Psi_3 v, \\ \pi_+\mathcal{J}_A(v)e &= (-a\cos\theta(\cos\theta) + c\sin\theta(\sin\theta))e \\ &\quad + (b\cos\theta - a\cos\theta(-\sin\theta) + c\sin\theta(\cos\theta))\Psi_1 e, \\ \pi_+\mathcal{J}_A(v)\Psi_1 e &= ((b + a\sin\theta)(\cos\theta) + c\cos\theta(\sin\theta))e \\ &\quad + ((a - b\sin\theta) + (b + a\sin\theta)(-\sin\theta) + c\cos\theta(\cos\theta))\Psi_1 e. \end{aligned}$$

The coefficient matrix for $\pi_+\mathcal{J}_A(v)$ on V_+ is then given by

$$\mathscr{C}_A(\theta) = \begin{pmatrix} -a\cos^2\theta + c\sin^2\theta & b\cos\theta + (a+c)\sin\theta\cos\theta \\ b\cos\theta + (a+c)\sin\theta\cos\theta & -2b\sin\theta + (a+c)\cos^2\theta \end{pmatrix}.$$

We have $\det(\mathscr{C}_A)(\pi/2) = -2bc$ and $\det(\mathscr{C}_A)(-\pi/2) = 2bc$. If $c \neq 0$, then these signs differ and hence $\det(\mathscr{C}_A)(\theta) = 0$ for some $-\pi/2 < \theta < \pi/2$ and $\mathfrak{M}$ is not null Jordan Osserman. If $c = 0$, then $\det(\mathscr{C}_A)(\pi/2) = 0$ and $\det(\mathscr{C}_A)(0) = -a^2 - b^2 \neq 0$ and again $\mathfrak{M}$ is not null Jordan Osserman.

3.10. ***Type II models.*** We approach this case directly. Let $\mathfrak{M} = (V, \langle \cdot, \cdot \rangle, A)$ be a model of signature $(2, 2)$, where A is a Type II algebraic curvature tensor. Then the analysis of [Blažić et al. 2001; García-Río et al. 2002] shows there exists an orthonormal basis $\{e_1, e_2, e_3, e_4\}$ for V such that the nonvanishing components of A are

$$\begin{aligned} A_{1221} &= A_{4334} = \pm(\alpha - \tfrac{1}{2}), & A_{1224} &= A_{1334} = \pm\tfrac{1}{2}, \\ A_{1331} &= A_{4224} = \mp(\alpha + \tfrac{1}{2}), & A_{2113} &= A_{2443} = \mp\tfrac{1}{2}, \\ A_{1234} &= (\pm(-\alpha + \tfrac{3}{2}) + \beta)/3, & A_{1423} &= 2(\pm\alpha - \beta)/3, \\ A_{1342} &= (\pm(-\alpha - \tfrac{3}{2}) + \beta)/3, & A_{1441} &= A_{3223} = -\beta. \end{aligned}$$

Let $u = e_2 - e_3$ and let $v = e_2 + e_3$. Then

$$\mathcal{J}_A(u) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta & \beta & 0 \\ 0 & -\beta & -\beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{J}_A(v) = \begin{pmatrix} \pm 2 & 0 & 0 & \mp 2 \\ 0 & \beta & -\beta & 0 \\ 0 & \beta & -\beta & 0 \\ \pm 2 & 0 & 0 & \mp 2 \end{pmatrix}.$$

If $\beta = 0$, then $r(u) = 0$ and $r(v) = 1$; if $\beta \neq 0$, then $r(u) = 1$ and $r(v) = 2$. Thus $\mathfrak{M}$ is not null Jordan Osserman.

3.11. ***Type III models.*** For $\mathfrak{M}$ of this type, there exists by [Blažić et al. 2001; García-Río et al. 2002] an orthonormal basis $\{e_1, e_2, e_3, e_4\}$ for V such that the nonvanishing components of A are

$$\begin{aligned} &A_{1221} = A_{4334} = \alpha, && A_{1331} = A_{4224} = -\alpha, \\ &A_{1441} = A_{3223} = -\alpha, && \\ &A_{2114} = A_{2334} = -\sqrt{2}/2, && A_{3114} = -A_{3224} = \sqrt{2}/2, \\ &A_{1223} = A_{1443} = A_{1332} = -A_{1442} = \sqrt{2}/2. && \end{aligned}$$

Let $u = e_2 - e_3$ and $v = e_2 + e_3$. Then

$$\mathcal{J}_A(u) = \begin{pmatrix} 0 & -\sqrt{2} & -\sqrt{2} & 0 \\ -\sqrt{2} & \alpha & \alpha & \sqrt{2} \\ \sqrt{2} & -\alpha & -\alpha & -\sqrt{2} \\ 0 & -\sqrt{2} & -\sqrt{2} & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{J}_A(v) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \alpha & -\alpha & 0 \\ 0 & \alpha & -\alpha & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

It now follows that $r(u) = 2$ while $r(v) \leq 1$ and hence $\mathfrak{M}$ is not null Jordan Osserman. This completes the proof of Theorem 1.5. $\square$

4. The proof of Theorem 1.8

Let $\mathcal{M}$ be a null Jordan Osserman manifold of signature $(2, 2)$. First note that, by Theorem 1.5, $\mathcal{M}$ has Type Ia. Results of [Blažić et al. 2001] then show that $\mathcal{M}$ either has constant sectional curvature, is locally isometric to a complex space form, or is locally isometric to a paracomplex space form. Since the curvature tensor of a paracomplex space form of constant paraholomorphic sectional curvature κ satisfies

$$R(x, y)z = \tfrac{1}{4}\kappa(R^0(x, y)z - R^J(x, y)z),$$

this is ruled out by Theorem 1.5, thus proving Theorem 1.8. $\square$

References

[Alekseevsky et al. 1999] D. Alekseevsky, N. Blažić, N. Bokan, and Z. Rakić, "Self-duality and pointwise Osserman manifolds", *Arch. Math. (Brno)* **35**:3 (1999), 193–201. MR 2000j:53094 Zbl 1054.53082

[Blažić et al. 1997] N. Blažić, N. Bokan, and P. Gilkey, "A note on Osserman Lorentzian manifolds", *Bull. London Math. Soc.* **29**:2 (1997), 227–230. MR 97m:53111 Zbl 0865.53018

[Blažić et al. 2001] N. Blažić, N. Bokan, and Z. Rakić, "Osserman pseudo-Riemannian manifolds of signature (2, 2)", *J. Aust. Math. Soc.* **71**:3 (2001), 367–395. MR 2002g:53119 Zbl 1011.53049

[Brozos-Vázquez et al. 2008] M. Brozos-Vázquez, P. Gilkey, S. Nikčević, and U. Simon, "Projectively Osserman manifolds", *Publ. Math. Debrecen* **72**:3-4 (2008), 359–370. MR 2406926 Zbl 05367462

[Chi 1988] Q.-S. Chi, "A curvature characterization of certain locally rank-one symmetric spaces", *J. Differential Geom.* **28**:2 (1988), 187–202. MR 90a:53060 Zbl 0654.53053

[Díaz-Ramos et al. 2006a] J. C. Díaz-Ramos, E. García-Río, and R. Vázquez-Lorenzo, "Four-dimensional Osserman metrics with nondiagonalizable Jacobi operators", *J. Geom. Anal.* **16**:1 (2006), 39–52. MR 2006m:53107 Zbl 1105.53052

[Díaz-Ramos et al. 2006b] J. C. Díaz-Ramos, E. García-Río, and R. Vázquez-Lorenzo, "New examples of Osserman metrics with nondiagonalizable Jacobi operators", *Differential Geom. Appl.* **24**:4 (2006), 433–442. MR 2007h:53106 Zbl 1099.53047

[Fiedler 2002] B. Fiedler, "Determination of the structure of algebraic curvature tensors by means of Young symmetrizers", *Sém. Lothar. Combin.* **48** (2002), Art. B48d. MR 2004d:53018

[García-Río and Vázquez-Lorenzo 2001] E. García-Río and R. Vázquez-Lorenzo, "Four-dimensional Osserman symmetric spaces", *Geom. Dedicata* **88**:1-3 (2001), 147–151. MR 2002k:53137 Zbl 1008.53023

[García-Río et al. 1997] E. García-Río, D. N. Kupeli, and M. E. Vázquez-Abal, "On a problem of Osserman in Lorentzian geometry", *Differential Geom. Appl.* **7**:1 (1997), 85–100. MR 98a:53099 Zbl 0880.53017

[García-Río et al. 1998] E. García-Río, M. E. Vázquez-Abal, and R. Vázquez-Lorenzo, "Nonsymmetric Osserman pseudo-Riemannian manifolds", *Proc. Amer. Math. Soc.* **126**:9 (1998), 2771–2778. MR 98m:53089 Zbl 0939.53014

[García-Río et al. 1999] E. García-Río, D. N. Kupeli, M. E. Vázquez-Abal, and R. Vázquez-Lorenzo, "Affine Osserman connections and their Riemann extensions", *Differential Geom. Appl.* **11**:2 (1999), 145–153. MR 2000i:53030 Zbl 0940.53017

[García-Río et al. 2002] E. García-Río, D. N. Kupeli, and R. Vázquez-Lorenzo, *Osserman manifolds in semi-Riemannian geometry*, Lecture Notes in Mathematics **1777**, Springer, Berlin, 2002. MR 2003e:53052 Zbl 1005.53040

[Gilkey 2001] P. B. Gilkey, *Geometric properties of natural operators defined by the Riemann curvature tensor*, World Scientific Publishing, River Edge, NJ, 2001. MR 2002k:53052 Zbl 1007.53001

[Gilkey and Ivanova 2001] P. Gilkey and R. Ivanova, "The Jordan normal form of Osserman algebraic curvature tensors", *Results Math.* **40**:1-4 (2001), 192–204. MR 2002j:53093 Zbl 0999.53014

[Gilkey and Ivanova 2002] P. B. Gilkey and R. Ivanova, "Spacelike Jordan Osserman algebraic curvature tensors in the higher signature setting", pp. 179–186 in *Differential geometry* (Valencia, 2001), edited by O. Gil-Medrano and V. Miquel, World Scientific Publishing, River Edge, NJ, 2002. MR 2003f:53125 Zbl 1031.53037

[Marden 2007] A. Marden, *Outer circles: An introduction to hyperbolic 3-manifolds*, Cambridge University Press, 2007. MR 2008i:57001 Zbl 1149.57030

Received April 3, 2008. Revised September 15, 2009.

EDUARDO GARCÍA-RÍO
DEPARTMENT OF GEOMETRY AND TOPOLOGY
FACULTY OF MATHEMATICS
UNIVERSITY OF SANTIAGO DE COMPOSTELA
15782 SANTIAGO DE COMPOSTELA
SPAIN
eduardo.garcia.rio@usc.es

PETER B. GILKEY
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OR 97403
UNITED STATES
gilkey@uoregon.edu

M. ELENA VÁZQUEZ-ABAL
DEPARTMENT OF GEOMETRY AND TOPOLOGY
FACULTY OF MATHEMATICS
UNIVERSITY OF SANTIAGO DE COMPOSTELA
15782 SANTIAGO DE COMPOSTELA
SPAIN
elena.vazquez.abal@usc.es

RAMÓN VÁZQUEZ-LORENZO
DEPARTMENT OF GEOMETRY AND TOPOLOGY
FACULTY OF MATHEMATICS
UNIVERSITY OF SANTIAGO DE COMPOSTELA
15782 SANTIAGO DE COMPOSTELA
SPAIN
ravazlor@edu.xunta.es

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

THE GEOGRAPHY OF SYMPLECTIC 4-MANIFOLDS WITH DIVISIBLE CANONICAL CLASS

MARK J. D. HAMILTON

We consider a version of the geography question for simply connected symplectic 4-manifolds that takes as an additional parameter the divisibility of the canonical class. We also find new examples of 4-manifolds admitting several symplectic structures that are inequivalent under deformation and self-diffeomorphisms of the manifold.

We are interested in the geography of simply connected, closed, symplectic 4-manifolds whose canonical classes have a given divisibility. In general, the geography question aims at finding for any given pair of integers (x, y) a closed 4-manifold M with some *a priori* specified properties (for example, irreducible, spin, simply connected, symplectic or complex) such that the Euler characteristic $e(M)$ equals x and the signature $\sigma(M)$ equals y. This question has been considered

MSC2000: 14J29, 57N13, 57R17.
Keywords: geography, symplectic, canonical class.
This work was supported by the *Studienstiftung des deutschen Volkes* and the *Deutsche Forschungsgemeinschaft*.

for simply connected symplectic 4-manifolds both in the spin and nonspin case for example in [Gompf 1995; Park and Szabó 2000; Park 1998; 2002]; see also [Chen 1987; Fintushel and Stern 1994; Persson 1981; Persson et al. 1996]. We consider the geography question for simply connected symplectic 4-manifolds whose canonical class, considered as an element in second cohomology with integer coefficients, is divisible by a given integer $d > 1$. Since the canonical class is characteristic, the first case $d = 2$ corresponds to the general case of spin symplectic 4-manifolds.

Geography questions are often formulated in terms of the invariants c_1^2 and χ_h instead of e and σ, which for smooth closed 4-manifolds are defined by

$$c_1^2(M) = 2e(M) + 3\sigma(M) \quad \text{and} \quad \chi_h(M) = \tfrac{1}{4}(e(M) + \sigma(M)).$$

For complex 4-manifolds these numbers have the same value as the square of the first Chern class and the holomorphic Euler characteristic, making the definitions consistent.

The constructions we use here depend on generalized fibre sums of symplectic manifolds, which are also known as Gompf sums or normal connected sums [Gompf 1995; McCarthy and Wolfson 1994], in particular in the form of knot surgery [Fintushel and Stern 1998] and a generalized version of knot surgery along embedded surfaces of higher genus [Fintushel and Stern 2004]. Some details on the generalized fibre sum can be found in Section 2.

In Sections 4, 6 and 7, we consider the case $c_1^2 = 0$ and the spin and nonspin cases for $c_1^2 > 0$ and negative signature, while the case $c_1^2 < 0$ is covered at the end of Section 1. We do not consider the case of nonnegative signature, since even without a restriction on the divisibility of the canonical class, such simply connected symplectic 4-manifolds are known to be difficult to find.

As a consequence of these geography results, there often exist at the same lattice point in the (χ_h, c_1^2)-plane several simply connected symplectic 4-manifolds whose canonical classes have pairwise different divisibilities. It is natural to ask whether the same smooth 4-manifold can admit several symplectic structures with canonical classes of different divisibilities; we consider this question in Sections 8 and 9. The symplectic structures with this property are inequivalent under deformations and orientation-preserving self-diffeomorphisms of the manifold. Similar examples have been found before on homotopy elliptic surfaces by McMullen and Taubes [1999], Smith [2000] and Vidussi [2001]. Another application of the geography question to the existence of inequivalent contact structures on certain 5-manifolds can be found in [Hamilton 2008].

In the final part of this article, we give an independent construction of simply connected symplectic 4-manifolds with divisible canonical class by finding complex surfaces of general type with divisible canonical class. The construction uses branched coverings over smooth curves in pluricanonical linear systems $|nK|$.

1. General restrictions on the divisibility of the canonical class

We begin by deriving a few general restrictions for symplectic 4-manifolds admitting a symplectic structure whose canonical class is divisible by an integer $d > 1$.

Let (M, ω) be a closed, symplectic 4-manifold. The canonical class K of the symplectic form ω is defined as

$$K = -c_1(TM, J),$$

where J is an almost-complex structure compatible with ω. The self-intersection number of K is given by the formula $K^2 = c_1^2(M) = 2e(M) + 3\sigma(M)$. Since the first Chern class $c_1(TM, J)$ is characteristic, it follows by a general property of the intersection form that $c_1^2(M) \equiv \sigma(M) \bmod 8$, and hence the number

$$\chi_h(M) = \tfrac{1}{4}(e(M) + \sigma(M))$$

is an integer. If $b_1(M) = 0$, this number is equal to $\frac{1}{2}(1 + b_2^+(M))$. In particular, in this case $b_2^+(M)$ is an odd integer and $\chi_h(M) > 0$. There is a further constraint if the manifold M is spin, equivalent to the congruence $\sigma(M) \equiv 0 \bmod 16$ given by Rohlin's theorem [1952], which says that $c_1^2(M) \equiv 8\chi_h(M) \bmod 16$. In particular, $c_1^2(M)$ is divisible by 8. We say that K is divisible by an integer d if there exists a cohomology class $A \in H^2(M; \mathbb{Z})$ with $K = dA$.

Lemma 1. *Let (M, ω) be a closed symplectic 4-manifold. Suppose K is divisible by an integer d. Then $c_1^2(M)$ is divisible by d^2 if d is odd and by $2d^2$ if d is even.*

Proof. If d divides K, we can write $K = dA$, where $A \in H^2(M; \mathbb{Z})$. The equation $c_1^2(M) = K^2 = d^2A^2$ implies that $c_1^2(M)$ is divisible by d^2 in any case. If d is even, then $w_2(M) \equiv K \equiv 0 \bmod 2$; hence M is spin and the intersection form Q_M is even. This implies that A^2 is divisible by 2; hence $c_1^2(M)$ is divisible by $2d^2$. □

The case $c_1^2(M) = 0$ is special, since there are no restrictions from this lemma; see Section 4. For the general case of spin symplectic 4-manifolds ($d = 2$), we recover the constraint that c_1^2 is divisible by 8.

Further restrictions come from the adjunction formula $2g - 2 = K \cdot C + C \cdot C$, where C is an embedded symplectic surface of genus g oriented by the restriction of the symplectic form.

Lemma 2. *Let (M, ω) be a closed symplectic 4-manifold. Suppose K is divisible by an integer d.*

- *If M contains a symplectic surface of genus g and self-intersection 0, then d divides $2g - 2$.*
- *If $d \neq 1$, then M is minimal. If M is in addition simply connected, then it is irreducible.*

Proof. The first part follows immediately by the adjunction formula. If M is not minimal, it contains a symplectically embedded sphere S of self-intersection (-1). The adjunction formula can be applied and yields $K \cdot S = -1$, and hence K is indivisible. The claim of irreducibility follows from [Hamilton and Kotschick 2006, Corollary 1.4]. $\square$

The canonical class of a 4-manifold M with $b_2^+ \geq 2$ is a Seiberg–Witten basic class, that is, it has nonvanishing Seiberg–Witten invariant. This implies that only finitely many classes in $H^2(M;\mathbb{Z})$ can occur as the canonical classes of symplectic structures on M.

Theorem 3 [Li and Liu 2001]. *Let M be a* (*smoothly*) *minimal closed* 4-*manifold with $b_2^+ = 1$. Then the canonical classes of all symplectic structures on M are equal up to sign.*

If M is a Kähler surface, we can consider the canonical class of the Kähler form.

Theorem 4. *Suppose that M is a minimal Kähler surface with $b_2^+ > 1$.*

- *If M is of general type*, *then $\pm K_M$ are the only Seiberg–Witten basic classes of M.*
- *If N is another minimal Kähler surface such that $b_2^+ > 1$ and $\phi : M \to N$ is a diffeomorphism*, *then $\phi^* K_N = \pm K_M$.*

For the proofs see [Friedman and Morgan 1997; Morgan 1996; Witten 1994]. When ϕ is the identity diffeomorphism, the second part of this theorem has an immediate consequence:

Corollary 5. *Let M be a* (*smoothly*) *minimal closed* 4-*manifold with $b_2^+ > 1$. Then the canonical classes of all Kähler structures on M are equal up to sign.*

The corresponding statement is not true in general for the canonical classes of *symplectic structures* on minimal 4-manifolds with $b_2^+ > 1$. There exist such 4-manifolds M admitting several symplectic structures whose canonical classes in $H^2(M;\mathbb{Z})$ are not equal up to sign. In addition, such examples can be constructed where the canonical classes cannot be permuted by orientation-preserving self-diffeomorphisms of the manifold [McMullen and Taubes 1999; Smith 2000; Vidussi 2001], for example because they have different divisibilities as elements in integral cohomology (see the examples in Sections 8 and 9).

It is useful to define the (maximal) divisibility of the canonical class in the case that $H^2(M;\mathbb{Z})$ is torsion-free.

Definition 6. Suppose H is a finitely generated free abelian group. For $a \in H$, let

$$d(a) = \max\{k \in \mathbb{N}_0 \mid \text{there exists a nonzero element } b \in H \text{ with } a = kb\}.$$

We call $d(a)$ the *divisibility* of a (or, for emphasis, the *maximal* divisibility). The divisibility of a is 0 if and only $a = 0$. We call a *indivisible* if $d(a) = 1$.

If M is a simply connected manifold, the integral cohomology group $H^2(M; \mathbb{Z})$ is torsion-free, and $K \in H^2(M; \mathbb{Z})$ has a well-defined divisibility.

Proposition 7. *Suppose M is a simply connected closed 4-manifold that admits at least two symplectic structures whose canonical classes have different divisibilities. Then M is not diffeomorphic to a complex surface.*

Proof. The assumptions imply M has a symplectic structure whose canonical class has divisibility $\neq 1$. By Lemma 2, the manifold M is (smoothly) minimal, and by Theorem 3, it has $b_2^+ > 1$. Suppose M is diffeomorphic to a complex surface. The Kodaira–Enriques classification implies M is diffeomorphic either to an elliptic surface $E(n)_{p,q}$ with $n \geq 2$ and p, q coprime, or to a surface of general type.

Consider the elliptic surfaces $E(n)_{p,q}$ for $n \geq 2$, and denote the class of a general fibre by F. The Seiberg–Witten basic classes of these 4-manifolds are known [Fintushel and Stern 1997], and consist of the set of classes of the form kf, where f denotes the indivisible class $f = F/pq$ and k is an integer such that

$$k \equiv npq - p - q \bmod 2 \quad \text{and} \quad |k| \leq npq - p - q.$$

Suppose ω is a symplectic structure on $E(n)_{p,q}$ with canonical class K. By a theorem of Taubes [Taubes 1995a; Kotschick 1997], the inequality $K \cdot [\omega] \geq |c \cdot [\omega]|$ holds for any basic class c, with equality if and only if $K = \pm c$, and the number $K \cdot [\omega]$ is positive if K is nonzero. It follows that the canonical class of any symplectic structure on $E(n)_{p,q}$ is given by $\pm(npq - p - q) f$; hence there is only one possible divisibility. This follows for surfaces of general type by the first part of Theorem 4. $\square$

We now consider the geography question for manifolds with $c_1^2 < 0$. The next theorem is due to C. H. Taubes [1995b] in the case $b_2^+ \geq 2$ and to A.-K. Liu [1996] in the case $b_2^+ = 1$.

Theorem 8. *Let M be a closed, symplectic 4-manifold. Suppose M is minimal.*

- *If $b_2^+(M) \geq 2$, then $K^2 \geq 0$.*
- *If $b_2^+(M) = 1$ and $K^2 < 0$, then M is a ruled surface, that is, an S^2-bundle over a surface (of genus ≥ 2).*

Since ruled surfaces over irrational curves are not simply connected, any simply connected, symplectic 4-manifold M with $c_1^2(M) < 0$ is not minimal. By Lemma 2, this implies that K is indivisible, that is, $d(K) = 1$.

Let $(\chi_h, c_1^2) = (n, -r)$ be a lattice point with $n, r \geq 1$, and let M be a simply connected symplectic 4-manifold with these invariants. Since M is not minimal, we can successively blow down r (-1)-spheres in M to get a simply connected symplectic 4-manifold N with invariants $(\chi_h, c_1^2) = (n, 0)$ such that there exists a diffeomorphism $M = N \# r\,\overline{\mathbb{C}P^2}$.

Conversely, consider the manifold $M = E(n) \# r\,\overline{\mathbb{C}P^2}$. Then M is a simply connected symplectic 4-manifold with indivisible K. Since $\chi_h(E(n)) = n$ and $c_1^2(E(n)) = 0$, this implies $(\chi_h(M), c_1^2(M)) = (n, -r)$. Hence the point $(n, -r)$ can be realized by a simply connected symplectic 4-manifold.

2. The generalized fibre sum

We next recall the definition of the generalized fibre sum from [Gompf 1995; McCarthy and Wolfson 1994] and fix some notation, used in [Hamilton 2008]. Let M and N be closed oriented 4-manifolds that contain embedded oriented surfaces Σ_M and Σ_N of genus g and self-intersection 0. We choose trivializations of the form $\Sigma_g \times D^2$ for tubular neighbourhoods of the surfaces Σ_M and Σ_N. The generalized fibre sum $X = M \#_{\Sigma_M = \Sigma_N} N$ is then formed by deleting the interior of the tubular neighbourhoods and gluing the resulting manifolds M' and N' along their boundaries $\Sigma_g \times S^1$, using a diffeomorphism that preserves the meridians to the surfaces, given by the S^1 fibres, and reverses the orientation on them. The closed oriented 4-manifold can depend on the choice of trivializations and gluing diffeomorphism. The trivializations of the tubular neighbourhoods also determine push-offs of the central surfaces Σ_M and Σ_N into the boundary. Under inclusion, the push-offs determine surfaces Σ_X and Σ_X' of self-intersection 0 in the 4-manifold X. In general, these surfaces do not represent the same homology class in X but differ by a rim torus. However, if the gluing diffeomorphism is chosen so that it preserves also the Σ_g-fibres in $\Sigma_g \times S^1$, then the push-offs get identified to a well-defined surface Σ_X in X.

Suppose the surfaces Σ_M and Σ_N represent indivisible nontorsion classes in the homology of M and N. We can then choose surfaces B_M and B_N in M and N that intersect Σ_M and Σ_N at a single positive transverse point. These surfaces with a disk removed can be assumed to bound the meridians to Σ_M and Σ_N in the manifolds M' and N'; hence they sew together to give a surface B_X in X.

The second cohomology of M can be split into a direct sum

$$H^2(M;\mathbb{Z}) \cong P(M) \oplus \mathbb{Z}\Sigma_M \oplus \mathbb{Z}B_M,$$

where $P(M)$ denotes the orthogonal complement to the subgroup $\mathbb{Z}\Sigma_M \oplus \mathbb{Z}B_M$ in $H^2(M;\mathbb{Z})$ with respect to the intersection form Q_M. The restriction of the intersection form to the last two summands is given by

$$\begin{pmatrix} 0 & 1 \\ 1 & B_M^2 \end{pmatrix}.$$

This form is unimodular; hence the restriction of the intersection form to $P(M)$ (modulo torsion) is unimodular as well. There exists a similar decomposition for the second cohomology of N.

Theorem 9. *Suppose that the integral cohomology of M, N and X is torsion-free and the surfaces Σ_M and Σ_N represent indivisible classes. If rim tori do not exist in the fibre sum $X = M \#_{\Sigma_M=\Sigma_N} N$, then the second cohomology of X splits as a direct sum*

$$H^2(X;\mathbb{Z}) \cong P(X) \oplus \mathbb{Z}\Sigma_X \oplus \mathbb{Z}B_X, \quad \textit{where } P(X) \cong P(M) \oplus P(N).$$

The restriction of the intersection form Q_X to $P(X)$ is the direct sum of the restrictions of Q_M and Q_N, and the restriction to $\mathbb{Z}\Sigma_X \oplus \mathbb{Z}B_X$ is of the form

$$\begin{pmatrix} 0 & 1 \\ 1 & B_M^2 + B_N^2 \end{pmatrix}.$$

A proof for this theorem can be found in [Hamilton 2008, Section V.3.5]. It implies that there exist monomorphisms of abelian groups of both $H^2(M;\mathbb{Z})$ and $H^2(N;\mathbb{Z})$ into $H^2(X;\mathbb{Z})$ given by

$$\Sigma_M \mapsto \Sigma_X, \qquad B_M \mapsto B_X, \qquad \mathrm{Id} : P(M) \to P(M), \tag{1}$$

and similarly for N. The monomorphisms do not preserve the intersection form if B_M^2 or B_N^2 differ from B_X^2. The next lemma can be useful in checking the conditions for Theorem 9; its proof follows from [Hamilton 2008, Sections V.2 and V.3].

Lemma 10. *Let $X = M \#_{\Sigma_M=\Sigma_N} N$ be a generalized fibre sum along embedded surfaces of self-intersection* 0*. Suppose that the map on integral first homology induced by one of the embeddings, say $\Sigma_N \to N$, is an isomorphism. Then rim tori do not exist in X. If in addition one of the surfaces represents an indivisible homology class, then $H_1(X;\mathbb{Z}) \cong H_1(M;\mathbb{Z})$.*

Suppose M and N are symplectic 4-manifolds and Σ_M and Σ_N symplectically embedded. We orient both surfaces by the restriction of the symplectic forms. Then the generalized fibre sum X also admits a symplectic structure. The canonical class K_X can be calculated as follows:

Theorem 11. *Under the assumptions of Theorem 9 and the embeddings of the cohomology of M and N into the cohomology of X given by Equation* (1)*, we have*

$$K_X = K_M + K_N - (2g-2)B_X + 2\Sigma_X.$$

A proof can be found in [Hamilton 2008, Section V.5]. The formula for $g = 1$ has been proved in [Smith 2000] and a related formula for arbitrary g can be found in [Ionel and Parker 2004].

3. The knot surgery construction

We will frequently use the following construction due to Fintushel and Stern [1998]. Let K be a knot in S^3, and denote a tubular neighbourhood of K by $\nu K \cong S^1 \times D^2$. Let m be a fibre of the circle bundle $\partial\nu K \to K$, and use an oriented Seifert surface for K to define a section $l : K \to \partial\nu K$. The circles m and l are called the meridian and the longitude of K. Let M_K be the closed 3-manifold obtained by 0-Dehn surgery on K. The manifold M_K is constructed as follows: Consider $S^3 \setminus \operatorname{int} \nu K$ and let $f : \partial(S^1 \times D^2) \to \partial(S^3 \setminus \operatorname{int} \nu K)$ be a diffeomorphism that maps the circle ∂D^2 onto l. Then one defines

$$M_K = (S^3 \setminus \operatorname{int} \nu K) \cup_f (S^1 \times D^2).$$

The manifold M_K is determined by this construction uniquely up to diffeomorphism. One can show that it has the same integral homology as $S^2 \times S^1$. The meridian m, which bounds the fibre in the normal bundle to K in S^3, becomes nonzero in the homology of M_K and defines a generator for $H_1(M_K; \mathbb{Z})$. The longitude l is null-homotopic in M_K, since it bounds one of the D^2 fibres glued in. This disk fibre together with the Seifert surface of K determine a closed, oriented surface B_K in M_K that intersects m once and generates $H_2(M_K; \mathbb{Z})$.

We consider the closed, oriented 4-manifold $M_K \times S^1$. It contains an embedded torus $T_K = m \times S^1$ of self-intersection 0, which has a framing coming from a canonical framing of m. Let X be an arbitrary closed, oriented 4-manifold, which contains an embedded torus T_X of self-intersection 0, representing an indivisible homology class. Then the result of *knot surgery* on X is given by the generalized fibre sum $X_K = X \#_{T_X = T_K} (M_K \times S^1)$. Here we have implicitly chosen a trivialization of the form $T^2 \times D^2$ for the tubular neighbourhood of the torus T_X. We choose a gluing diffeomorphism that preserves both the T^2 factor and the S^1 factor on the boundaries of the tubular neighbourhoods and reverses orientation on the S^1 factor (the smooth 4-manifold X_K might depend on the choice of the framing for T_X). The embedded torus of self-intersection 0 in X_K, defined by identifying the push-offs, is denoted by T_{X_K}.

The closed surface B_K in the 3-manifold M_K determines under inclusion a closed surface in the 4-manifold $M_K \times S^1$, denoted by the same symbol. It intersects the torus T_K at a single transverse point. We also choose a surface B_X in X intersecting T_X transversely and geometrically once. Both surfaces sew together to form a surface B_{X_K} in X_K that intersects the torus T_{X_K} at a single transverse point.

We assume the cohomology of X is torsion-free. By [Fintushel and Stern 1998], it is known that there exists an isomorphism

$$H^2(X; \mathbb{Z}) \cong H^2(X_K; \mathbb{Z}) \tag{2}$$

preserving intersection forms. In the notation of Section 2 this follows because $H^2(M_K \times S^1; \mathbb{Z}) \cong \mathbb{Z}T_K \oplus \mathbb{Z}B_K$ and hence $P(M_K \times S^1) = 0$. In addition, the self-intersection number of B_{X_K} is equal to the self-intersection number of B_X, because the class B_K has zero self-intersection (it can be moved away in the S^1 direction). The claim then follows from Theorem 9 and Lemma 10.

In particular, assume that both X and $X' = X \setminus T_X$ are simply connected. Since the fundamental group of $M_K \times S^1$ is normally generated by the image of the fundamental group of T_K under inclusion, it follows that X_K is again simply connected; hence by Freedman's theorem [1982], the manifolds X and X_K are homeomorphic. However, one can show with Seiberg–Witten theory that X and X_K are in many cases not diffeomorphic [Fintushel and Stern 1998].

Suppose that K is a fibred knot, that is, there exists a fibration

$$\begin{tikzcd} S^3 \setminus \operatorname{int} \nu K \arrow[d] & \Sigma_h' \arrow[l] \\ S^1 & \end{tikzcd}$$

over the circle, where the fibres Σ_h' are punctured surfaces of genus h forming Seifert surfaces for K. Then M_K is fibred by closed surfaces B_K of genus h. This induces a fibre bundle

$$\begin{tikzcd} M_K \times S^1 \arrow[d] & \Sigma_h \arrow[l] \\ T^2 & \end{tikzcd}$$

and the torus $T_K = m \times S^1$ is a section of this bundle. By a theorem of Thurston [1976] the manifold $M_K \times S^1$ admits a symplectic form such that T_K and the fibres are symplectic. This construction can be used to do symplectic generalized fibre sums along T_K if the manifold X is symplectic and the torus T_X symplectically embedded. The canonical class of $M_K \times S^1$ can be calculated by the adjunction formula, because the fibres B_K and the torus T_K are symplectic surfaces and form a basis of $H_2(M_K \times S^1; \mathbb{Z})$. The result is $K_{M_K \times S^1} = (2h-2)T_K$. According to [Fintushel and Stern 1998], the canonical class of the symplectic 4-manifold X_K is then given by

$$K_{X_K} = K_X + 2hT_X. \tag{3}$$

See also Theorem 11.

4. Symplectic 4-manifolds with $c_1^2 = 0$

Definition 12. A closed, simply connected 4-manifold M is called a *homotopy elliptic surface* if M is homeomorphic to a relatively minimal, simply connected elliptic surface, that is, to a complex surface of the form $E(n)_{p,q}$ with p, q coprime and $n \geq 1$.

For details on the surfaces $E(n)_{p,q}$, see [Gompf and Stipsicz 1999, Section 3.3]. By definition, homotopy elliptic surfaces M are simply connected with invariants

$$c_1^2(M) = 0, \quad e(M) = 12n, \quad \sigma(M) = -8n.$$

The integer n is equal to $\chi_h(M)$. In particular, $K^2 = 0$ for symplectic homotopy elliptic surfaces. There is a converse:

Lemma 13. *Let M be a closed, simply connected, symplectic 4-manifold with $K^2 = 0$. Then M is a homotopy elliptic surface.*

Proof. Since M is almost complex, $\chi_h(M)$ is an integer. The Noether formula

$$\chi_h(M) = \tfrac{1}{12}(K^2 + e(M)) = \tfrac{1}{12}e(M)$$

implies that $e(M)$ is divisible by 12; hence $e(M) = 12k$ for some $k > 0$. Together with the equation

$$0 = K^2 = 2e(M) + 3\sigma(M),$$

it follows that $\sigma(M) = -8k$. Suppose that M is nonspin. If k is odd, then M has the same Euler characteristic, signature and type as $E(k)$. If k is even, then M has the same Euler characteristic, signature and type as the nonspin manifold $E(k)_2$. Since M is simply connected, M is homeomorphic to the corresponding elliptic surface by Freedman's theorem [1982].

Suppose that M is spin. Then the signature is divisible by 16 due to Rohlin's theorem. Hence the integer k above has to be even. Then M has the same Euler characteristic, signature and type as the spin manifold $E(k)$. Again by Freedman's theorem, the 4-manifold M is homeomorphic to $E(k)$. □

Lemma 14. *Suppose that M is a symplectic homotopy elliptic surface such that the divisibility of K is even. Then $\chi_h(M)$ is even.*

Proof. The assumption implies that M is spin. The Noether formula then shows that $\chi_h(M)$ is even, since $K^2 = 0$ and $\sigma(M)$ is divisible by 16. □

The next theorem shows that this is the only restriction on the divisibility of the canonical class K for symplectic homotopy elliptic surfaces.

Theorem 15 (homotopy elliptic surfaces). *Let n and d be positive integers. If n is odd, assume that d is odd also. Then there exists a symplectic homotopy elliptic surface (M, ω) with $\chi_h(M) = n$ whose canonical class K has divisibility d.*

Note that there is no constraint on d if n is even.

Proof. If n is equal to 1 or 2, the symplectic manifold can be realized as an elliptic surface. The canonical class of an elliptic surface $E(n)_{p,q}$ with p, q coprime is given by $K = (npq - p - q)f$, where f is indivisible and $F = pqf$ denotes the class of a general fibre. For $n = 1$ and d odd, we can take the surface $E(1)_{d+2,2}$, since

$$(d+2)2-(d+2)-2=d.$$

For $n = 2$ and d arbitrary, we can take $E(2)_{d+1} = E(2)_{d+1,1}$, since

$$2(d+1)-(d+1)-1=d.$$

We now consider the case $n \geq 1$ in general and separate the proof into several cases.

Case: $d = 2k$ and $n = 2m$ are both even with $k, m \geq 1$. Consider the elliptic surface $E(n)$. It contains a general fibre F that is an embedded symplectic torus of self-intersection 0. It also contains a rim torus R that arises from a decomposition of $E(n)$ as a fibre sum $E(n) = E(n-1) \#_F E(1)$; see [Gompf and Mrowka 1993] and Example 30. The rim torus R has self-intersection 0 and there exists a dual (Lagrangian) 2-sphere S with intersection $RS = 1$. We can assume that R and S are disjoint from the fibre F. The rim torus is in a natural way Lagrangian. By a perturbation of the symplectic form, we can assume that it becomes symplectic. We give R the orientation induced by the symplectic form. The proof consists in doing knot surgery along the fibre F and the rim torus R.

Let K_1 be a fibred knot of genus $g_1 = m(k-1)+1$. We do knot surgery along F with the knot K_1 to get a new symplectic 4-manifold M_1. The elliptic fibration $E(n) \to \mathbb{C}P^1$ has a section showing that the meridian of F, which is the S^1 fibre of $\partial \nu F \to F$, bounds a disk in $E(n) \setminus \operatorname{int} \nu F$. This implies that the complement of F in $E(n)$ is simply connected; hence the manifold M_1 is again simply connected. By the knot surgery construction the manifold M_1 is homeomorphic to $E(n)$. The canonical class is given by formula (3):

$$K_{M_1} = (n-2)F + 2g_1F = (2m-2+2mk-2m+2)F = 2mkF.$$

Here we have identified the cohomology of M_1 and $E(n)$ under the isomorphism in Equation (2). The rim torus R still exists as an embedded oriented symplectic torus in M_1 with a dual 2-sphere S because we can assume that the knot surgery takes place in a small neighbourhood of F disjoint from R and S. In particular, the complement of R in M_1 is simply connected. Let K_2 be a fibred knot of genus $g_2 = k$, and let M be the result of knot surgery on M_1 along R. Then M is a simply connected symplectic 4-manifold homeomorphic to $E(n)$. The canonical class of M is given by $K = 2mkF + 2kR$.

The cohomology class K is divisible by $2k$. The sphere S sews together with a Seifert surface for the knot K_2 to give a surface C in M with $C \cdot R = 1$ and $C \cdot F = 0$; hence $C \cdot K = 2k$. Therefore the divisibility of K is precisely $d = 2k$.

Case: $d = 2k+1$ and $n = 2m+1$ are both odd with $k \geq 0$ and $m \geq 1$. We consider the elliptic surface $E(n)$ and do a similar construction. Let K_1 be a fibred knot of genus $g_1 = 2km + k + 1$, and do knot surgery along F as above. We get a simply connected symplectic 4-manifold M_1 with canonical class

$$\begin{aligned} K_{M_1} &= (n-2)F + 2g_1 F \\ &= (2m+1-2+4km+2k+2)F = (4km+2k+2m+1)F \\ &= (2m+1)(2k+1)F. \end{aligned}$$

Next we consider a fibred knot K_2 of genus $g_2 = 2k+1$ and do knot surgery along the rim torus R. The result is a simply connected symplectic 4-manifold M homeomorphic to $E(n)$ with canonical class $K = (2m+1)(2k+1)F + 2(2k+1)R$, which is divisible by $(2k+1)$. The same argument as above shows that there is a surface C in M with $C \cdot K = 2(2k+1)$. We claim that the divisibility of K is precisely $(2k+1)$: This follows because M is still homeomorphic to $E(n)$ by the knot surgery construction. Since n is odd, the manifold M is not spin and hence 2 does not divide K. (An explicit surface with odd intersection number can be constructed from a section of $E(n)$ and a Seifert surface for the knot K_1. This surface has self-intersection number $-n$ and intersection number $(2m+1)(2k+1)$ with K.)

To cover the remaining case $m = 0$ (corresponding to $n = 1$), we can do knot surgery on the elliptic surface $E(1)$ along a general fibre F with a knot K_1 of genus $g_1 = k+1$. The resulting manifold M_1 has canonical class

$$K_{M_1} = -F + (2k+2)F = (2k+1)F.$$

Case: $d = 2k+1$ is odd and $n = 2m$ is even with $k \geq 0$ and $m \geq 1$. We consider the elliptic surface $E(n)$ and first perform a logarithmic transformation along F of index 2. Let f denote the multiple fibre such that F is homologous to $2f$. There exists a 2-sphere in $E(n)_2$ that intersects f at a single point (for a proof see Lemma 16). In particular, the complement of f in $E(n)_2$ is simply connected. The canonical class of $E(n)_2 = E(n)_{2,1}$ is given by $K = (2n-3)f$. We can assume that the torus f is symplectic (for example, by considering the logarithmic transformation to be done on the complex algebraic surface $E(n)$, resulting in the complex algebraic surface $E(n)_2$). Let K_1 be a fibred knot of genus $g_1 = 4km + k + 2$, and do knot surgery along f with K_1 as above. The result is a simply connected symplectic 4-manifold homeomorphic to $E(n)_2$. The canonical class is given by

$$\begin{aligned} K_{M_1} &= (2n-3)f + 2g_1 f \\ &= (4m-3+8km+2k+4)f = (8km+4m+2k+1)f \\ &= (4m+1)(2k+1)f. \end{aligned}$$

We now consider a fibred knot K_2 of genus $g_2 = 2k+1$ and do knot surgery along the rim torus R. We get a simply connected symplectic 4-manifold M homeomorphic to $E(n)_2$ with canonical class $K = (4m+1)(2k+1)f + 2(2k+1)R$. A similar argument as above shows that the divisibility of K is $d = 2k+1$. □

Lemma 16. *Let $p \geq 1$ be an integer, and let f be the multiple fibre in $E(n)_p$. Then there exists a sphere in $E(n)_p$ that intersects f transversely at one point.*

Proof. We can think of the logarithmic transformation as gluing $T^2 \times D^2$ into $E(n) \setminus \operatorname{int} \nu F$ by a certain diffeomorphism $\phi : T^2 \times S^1 \to \partial \nu F$. The fibre f corresponds to $T^2 \times \{0\}$. Consider a disk of the form $\{*\} \times D^2$. It intersects f once, and its boundary maps under ϕ to a certain simple closed curve on $\partial \nu F$. Since $E(n) \setminus \operatorname{int} \nu F$ is simply connected, this curve bounds a disk in $E(n) \setminus \operatorname{int} \nu F$. The union of this disk and the disk $\{*\} \times D^2$ is a sphere in $E(n)_p$ that intersects f transversely once. □

Remark 17. Under the assumptions of Theorem 15, it is possible to construct infinitely many homeomorphic but pairwise nondiffeomorphic symplectic homotopy elliptic surfaces $(M_r)_{r \in \mathbb{N}}$ with $\chi_h(M_r) = n$, whose canonical classes all have divisibility equal to d. This follows because we can vary in each case the knot K_1 and its genus g_1 without changing the divisibility of the canonical class. The claim then follows by the formula for the Seiberg–Witten invariants of knot surgery manifolds [Fintushel and Stern 1998].

5. Generalized knot surgery

Symplectic manifolds with $c_1^2 > 0$ and divisible canonical class can be constructed with a version of knot surgery for higher genus surfaces described in [Fintushel and Stern 2004]. Let $K = K_h$ denote the $(2h+1, -2)$-torus knot, which is a fibred knot of genus h. Consider the manifold $M_K \times S^1$ from the knot surgery construction of Section 3. This manifold has the structure of a Σ_h-bundle over T^2:

$$\begin{array}{ccc} M_K \times S^1 & \longleftarrow & \Sigma_h \\ \downarrow & & \\ T^2 & & \end{array}$$

We denote a fibre of this bundle by Σ_F. The fibration defines a trivialization of the normal bundle $\nu \Sigma_F$. We form g consecutive generalized fibre sums along the fibres Σ_F to get

$$Y_{g,h} = (M_K \times S^1)\,\#_{\Sigma_F=\Sigma_F}\,\#\cdots\#_{\Sigma_F=\Sigma_F}(M_K \times S^1).$$

We choose the gluing diffeomorphism so that it identifies the Σ_h fibres in the boundary of the tubular neighbourhoods. This implies that $Y_{g,h}$ is a Σ_h-bundle over Σ_g:

$$\begin{tikzcd} Y_{g,h} \arrow[d] & \Sigma_h \arrow[l] \\ \Sigma_g & \end{tikzcd}$$

We denote the fibre again by Σ_F. The fibre bundle has a section Σ_S sewed together from g torus sections of $M_K \times S^1$. Since the knot K is a fibred knot, the manifold $M_K \times S^1$ admits a symplectic structure such that the fibre and the section are symplectic. By the Gompf construction this is then also true for $Y_{g,h}$.

The invariants of the 4-manifold $Y_{g,h}$ can be calculated by standard formulas [Park 2002, Lemma 2.4]:

$$c_1^2(Y_{g,h}) = 8(g-1)(h-1), \quad e(Y_{g,h}) = 4(g-1)(h-1), \quad \sigma(Y_{g,h}) = 0.$$

By induction on g, one can show that the fundamental group $\pi_1(Y_{g,h})$ is normally generated by the image of $\pi_1(\Sigma_S)$ under inclusion [Fintushel and Stern 2004, Proposition 2]. This fact, together with the exact sequence

$$H_1(\Sigma_F) \to H_1(Y_{g,h}) \to H_1(\Sigma_g) \to 0$$

coming from the long exact homotopy sequence for the fibration $\Sigma_F \to Y_{g,h} \to \Sigma_g$ by abelianization, shows that the inclusion $\Sigma_S \to Y_{g,h}$ induces an isomorphism on H_1 and the inclusion $\Sigma_F \to Y_{g,h}$ induces the zero map. In particular, the homology group $H_1(Y_{g,h};\mathbb{Z})$ is free abelian of rank $b_1(Y_{g,h}) = gb_1(M_K \times S^1) = 2g$. This implies with the formula for the Euler characteristic above that

$$b_2(Y_{g,h}) = 4h(g-1) + 2.$$

The summand $4h(g-1)$ results from $2h$ *split classes* (or *vanishing classes*) together with $2h$ dual rim tori that are created in each fibre sum. The split classes are formed as follows: In each fibre sum, the interior of a tubular neighbourhood $\nu\Sigma_F$ of a fibre on each side of the sum is deleted and the boundaries $\partial\nu\Sigma_F$ glued together such that the fibres inside the boundary get identified pairwise. Since the inclusion of the fibre Σ_F into $M_K \times S^1$ induces the zero map on first homology, the $2h$ generators of $H_1(\Sigma_h)$, where Σ_h is considered as a fibre in $\partial\nu\Sigma_F$, bound surfaces in $M_K \times S^1$ minus the interior of the tubular neighbourhood $\nu\Sigma_F$. The split classes arise from sewing together surfaces bounding corresponding generators on each side of the fibre sum. Fintushel and Stern show that in the case above there exists a basis for the group of split classes consisting of $2h(g-1)$ disjoint surfaces of genus 2 and self-intersection 2. This implies

$$H^2(Y_{g,h};\mathbb{Z}) = 2h(g-1)\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

where the last summand is the intersection form on $(\mathbb{Z}\Sigma_S \oplus \mathbb{Z}\Sigma_F)$. They also show that the canonical class of $Y_{g,h}$ is given by $K_Y = (2h-2)\Sigma_S + (2g-2)\Sigma_F$, where Σ_S and Σ_F are oriented by the symplectic form.

Let M be a closed symplectic 4-manifold that contains a symplectic surface Σ_M of genus g and self-intersection 0, oriented by the symplectic form and representing an indivisible homology class. We can then form the symplectic generalized fibre sum $X = M \#_{\Sigma_M = \Sigma_S} Y_{g,h}$. If the manifolds M and $M \setminus \Sigma_M$ are simply connected, then X is again simply connected because the fundamental group of $Y_{g,h}$ is normally generated by the image of $\pi_1(\Sigma_S)$. Since the inclusion of the surface Σ_S in $Y_{g,h}$ induces an isomorphism on first homology, it follows by Theorem 9 and Lemma 10 that

$$H^2(X;\mathbb{Z}) = P(M) \oplus P(Y_{g,h}) \oplus (\mathbb{Z}B_X \oplus \mathbb{Z}\Sigma_X).$$

The surface B_X is sewed together from a surface B_M in M with $B_M\Sigma_M = 1$ and the fibre Σ_F in the manifold $Y_{g,h}$. Since $\Sigma_F^2 = 0$, the embedding $H^2(M;\mathbb{Z}) \to H^2(X;\mathbb{Z})$ given by Equation (1) preserves the intersection form. Therefore we can write

$$H^2(X;\mathbb{Z}) = H^2(M;\mathbb{Z}) \oplus P(Y_{g,h}) \tag{4}$$

with intersection form

$$Q_X = Q_M \oplus 2h(g-1)\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}.$$

The invariants of X are given by

$$c_1^2(X) = c_1^2(M) + 8h(g-1), \quad e(X) = e(M) + 4h(g-1), \quad \sigma(X) = \sigma(M).$$

The canonical class of X can be calculated by Theorem 11 to be

$$K_X = K_M + 2h\Sigma_M, \tag{5}$$

where the isomorphism in (4) is understood (this formula follows also from the calculation of Seiberg–Witten invariants in [Fintushel and Stern 2004]). Equation (5) is a generalization of Equation (3). In particular, we get:

Proposition 18. *Suppose that M is a closed, symplectic 4-manifold that contains a symplectic surface Σ_M of genus $g > 1$ and self-intersection 0. Suppose that $\pi_1(M) = \pi_1(M \setminus \Sigma_M) = 1$ and that the canonical class of M is divisible by d.*

- *If d is odd, there exists for every integer $t \geq 1$ a simply connected symplectic 4-manifold X with invariants*

$$c_1^2(X) = c_1^2(M) + 8td(g-1), \quad e(X) = e(M) + 4td(g-1), \quad \sigma(X) = \sigma(M)$$

and canonical class divisible by d.

- *If d is even, there exists for every integer $t \geq 1$ a simply connected symplectic 4-manifold X with invariants*

$$c_1^2(X) = c_1^2(M) + 4td(g-1), \quad e(X) = e(M) + 2td(g-1), \quad \sigma(X) = \sigma(M)$$

and canonical class divisible by d.

This follows from the construction above by taking the genus of the torus knot $h = td$ if d is odd and $h = \frac{1}{2}td$ if d is even. Hence if a symplectic surface Σ_M of genus $g > 1$ and self-intersection 0 exists in M, we can raise c_1^2 without changing the signature or the divisibility of the canonical class.

6. Spin symplectic 4-manifolds with $c_1^2 > 0$ and negative signature

We can apply the construction from Section 5 to the symplectic homotopy elliptic surfaces constructed in Theorem 15. In this section we consider the case of even divisibility d and in the following section the case of odd d.

Recall that in the first case in the proof of Theorem 15, we constructed a simply connected symplectic 4-manifold M from the elliptic surface $E(2m)$ by doing knot surgery along a general fibre F with a fibred knot K_1 of genus $g_1 = (k-1)m + 1$ and a further knot surgery along a rim torus R with a fibred knot K_2 of genus $g_2 = k$. Here $2m \geq 2$ and $d = 2k \geq 2$ are arbitrary even integers. The canonical class is given by

$$K_M = 2mkF + 2kR = mdF + dR.$$

The manifold M is still homeomorphic to $E(2m)$. There is an embedded 2-sphere S in $E(2m)$ of self-intersection -2 that intersects the rim torus R once. The sphere S is naturally Lagrangian [Auroux et al. 2005]. We can assume that S is disjoint from the fibre F and by a perturbation of the symplectic structure on $E(2m)$ that the regular fibre F, the rim torus R and the dual 2-sphere S are all symplectic and the symplectic form induces a positive volume form on each of them; see the proofs of [Fintushel and Stern 2001, Lemma 2.1] and [Vidussi 2007, Proposition 3.2].

The 2-sphere S minus a disk sews together with a Seifert surface for K_2 to give a symplectic surface C in M of genus k and self-intersection -2 that intersects the rim torus R once. By smoothing the double point we get a symplectic surface Σ_M in M of genus $g = k+1$ and self-intersection 0 that represents $C + R$.

The complement of Σ_M in M is simply connected, since we can assume $R \cup S$ in the elliptic surface $E(2m)$ is contained in an embedded nucleus $N(2)$; see

[Gompf and Mrowka 1993; Gompf and Stipsicz 1999] and Example 30. Inside the nucleus $N(2)$ there exists a cusp that is homologous to R and disjoint from it. The cusp is still contained in M and intersects the surface Σ_M once. Since M is simply connected and the cusp homeomorphic to S^2, the claim $\pi_1(M \setminus \Sigma_M) = 1$ follows.[1]

Let $t \geq 1$ be an arbitrary integer, and let K_3 be the $(2h+1, -2)$-torus knot of genus $h = tk$. Consider the generalized fibre sum $X = M \#_{\Sigma_M = \Sigma_S} Y_{g,h}$. Then X is a simply connected symplectic 4-manifold with invariants

$$c_1^2(X) = 8tk^2 = 2td^2, \quad e(X) = 24m + 4tk^2 = 24m + td^2, \quad \sigma(X) = -16m.$$

The canonical class is given by

$$K_X = K_M + 2tk\,\Sigma_M = d(mF + R + t\,\Sigma_M).$$

Hence K_X has divisibility d, since the class $mF + R + t\,\Sigma_M$ has intersection 1 with Σ_M. Therefore:

Theorem 19. *Let $d \geq 2$ be an even integer. Then for every pair m, t of positive integers, there exists a simply connected closed spin symplectic 4-manifold X with invariants*

$$c_1^2(X) = 2td^2, \quad e(X) = td^2 + 24m, \quad \sigma(X) = -16m,$$

such that the canonical class K_X has divisibility d.

Note that this solves by Lemma 1 and Rohlin's theorem the existence question for simply connected 4-manifolds with canonical class divisible by an even integer and negative signature. In particular (for $d = 2$), every possible lattice point with $c_1^2 > 0$ and $\sigma < 0$ can be realized by a simply connected spin symplectic 4-manifold with this construction; the existence of such 4-manifolds has been proved similarly in [Park and Szabó 2000].

Example 20 (spin homotopy Horikawa surfaces). To identify the homeomorphism type of some of the manifolds in Theorem 19, let $d = 2k$; hence

$$c_1^2(X) = 8tk^2 \quad \text{and} \quad \chi_h(X) = tk^2 + 2m.$$

We consider the case when the invariants are on the Noether line $c_1^2 = 2\chi_h - 6$. This happens if and only if $6tk^2 = 4m - 6$ and hence $2m = 3tk^2 + 3$, which has a solution if and only if both t and k are odd. Hence for every pair $t, k \geq 1$ of odd integers, there exists a simply connected symplectic 4-manifold X with invariants

$$c_1^2(X) = 8tk^2 \quad \text{and} \quad \chi_h(X) = 4tk^2 + 3$$

such that the divisibility of K_X is $2k$.

[1] This argument is similar to the argument showing that the complement of a section in $E(n)$ is simply connected; see [Gompf 1995, Example 5.2].

By a construction of Horikawa [1976a], there exists for every odd integer $r \geq 1$ a simply connected spin complex algebraic surface M on the Noether line with invariants

$$c_1^2(M) = 8r \quad \text{and} \quad \chi_h(M) = 4r + 3.$$

See also [Gompf and Stipsicz 1999, Theorem 7.4.20] where this surface is called $U(3, r+1)$.

By Freedman's theorem [1982] the symplectic 4-manifolds X constructed above for odd parameters t and k are homeomorphic to spin Horikawa surfaces with $r = tk^2$. If $k > 1$ and t is arbitrary, the canonical class of X has divisibility $2k > 2$. In this case the manifold X cannot be *diffeomorphic* to a Horikawa surface: It is known by [Horikawa 1976a] that all Horikawa surfaces M have a fibration in genus 2 curves; hence by Lemma 2 the divisibility of K_M is at most 2 and in the spin case is equal to 2. Since Horikawa surfaces are minimal complex surfaces of general type, the claim follows by Proposition 7.

7. Nonspin symplectic 4-manifolds with $c_1^2 > 0$ and negative signature

We now we construct some families of simply connected symplectic 4-manifolds with $c_1^2 > 0$ such that the divisibility of K is a given odd integer $d > 1$. However, we do not have a complete existence result as in Theorem 19.

We consider the case that the canonical class K_X is divisible by an odd integer d and the signature $\sigma(X)$ is divisible by 8.

Lemma 21. *Let X be a closed simply connected symplectic 4-manifold such that K_X is divisible by an odd integer $d \geq 1$ and $\sigma(X)$ is divisible by 8. Then $c_1^2(X)$ is divisible by $8d^2$.*

Proof. Suppose that $\sigma(X) = 8m$ for some integer $m \in \mathbb{Z}$. Then $b_2^-(X) = b_2^+(X) - 8m$ hence $b_2(X) = 2b_2^+(X) - 8m$. This implies $e(X) = 2b_2^+(X) + 2 - 8m$. Since X is symplectic, the integer $b_2^+(X)$ is odd, so we can write $b_2^+(X) = 2k + 1$ for some $k \geq 0$. This implies $e(X) = 4k + 4 - 8m$; hence $e(X)$ is divisible by 4. The equation $c_1^2(X) = 2e(X) + 3\sigma(X)$ shows that $c_1^2(X)$ is divisible by 8. Since $c_1^2(X)$ is also divisible by the odd integer d^2, the claim follows. □

The following theorem covers the case that K_X has odd divisibility and the signature is negative, divisible by 8 and no greater than -16:

Theorem 22. *Let $d \geq 1$ be an odd integer. Then for every pair n, t of positive integers with $n \geq 2$, there is a simply connected closed nonspin symplectic 4-manifold X with invariants*

$$c_1^2(X) = 8td^2, \quad e(X) = 4td^2 + 12n, \quad \sigma(X) = -8n$$

such that the canonical class K_X has divisibility d.

Proof. The proof is similar to the proof of Theorem 19. We can write $d = 2k+1$ with $k \geq 0$.

Case: $n = 2m+1$ is odd, where $m \geq 1$. In the proof of Theorem 15, a homotopy elliptic surface M with $\chi_h(M) = n$ was constructed from the elliptic surface $E(n)$ by doing knot surgery along a general fibre F with a fibred knot K_1 of genus $g_1 = 2km+k+1$ and a further knot surgery along a rim torus R with a fibred knot K_2 of genus $g_2 = 2k+1 = d$. The canonical class is given by

$$K_M = (2m+1)(2k+1)F + 2(2k+1)R = (2m+1)dF + 2dR.$$

There exists a symplectically embedded 2-sphere S in $E(n)$ of self-intersection -2 that sews together with a Seifert surface for K_2 to give a symplectic surface C in M of genus d and self-intersection -2 that intersects the rim torus R once. By smoothing the double point, we get a symplectic surface Σ_M in M of genus $g = d+1$ and self-intersection 0 that represents $C+R$. Using a cusp that intersects Σ_M once, it follows as above that the complement $M \setminus \Sigma_M$ is simply connected.

Let $t \geq 1$ be an arbitrary integer and K_3 the $(2h+1, -2)$-torus knot of genus $h = td$. We consider the generalized fibre sum $X = M \#_{\Sigma_M = \Sigma_S} Y_{g,h}$, where $g = d+1$. Then X is a simply connected symplectic 4-manifold with invariants

$$c_1^2(X) = 8td^2, \quad e(X) = 4td^2 + 12n, \quad \sigma(X) = -8n.$$

The canonical class is given by

$$K_X = K_M + 2td\Sigma_M = d((2m+1)F + 2R + 2t\Sigma_M).$$

Hence K_X has divisibility d, since the class $(2m+1)F + 2R + 2t\Sigma_M$ twice intersects Σ_M and has intersection $(2m+1)$ with a surface coming from a section of $E(n)$ and a Seifert surface for K_1.

Case: $n = 2m$ is even, where $m \geq 1$. This case can be proved similarly. By doing a logarithmic transform on the fibre F in $E(n)$ and two further knot surgeries with a fibred knot K_1 of genus $g_1 = 4km+k+2$ on the multiple fibre f and with a fibred knot K_2 of genus $g_2 = 2k+1 = d$ along a rim torus R, we get a homotopy elliptic surface M with $\chi_h(M) = n$ and canonical class $K_M = (4m+1)df + 2dR$. The same construction as above yields a simply connected symplectic 4-manifold X with invariants

$$c_1^2(X) = 8td^2, \quad e(X) = 4td^2 + 12n, \quad \sigma(X) = -8n.$$

The canonical class is given by

$$K_X = K_M + 2td\Sigma_M = d((4m+1)f + 2R + 2t\Sigma_M).$$

Hence K_X again has divisibility d. □

Example 23 (nonspin homotopy Horikawa surfaces). The invariants of the manifolds in Theorem 22 are

$$c_1^2(X) = 8td^2 \quad \text{and} \quad \chi_h(X) = td^2 + n.$$

Similarly to Example 20, this implies that for every pair $d, t \geq 1$ of positive integers with d odd and t arbitrary, there exists a nonspin symplectic homotopy Horikawa surface X on the Noether line $c_1^2 = 2\chi_h - 6$ with invariants

$$c_1^2(X) = 8td^2 \quad \text{and} \quad \chi_h(X) = 4td^2 + 3,$$

whose canonical class has divisibility d. Note that for every integer $s \geq 1$ there exists a nonspin complex Horikawa surface M [Horikawa 1976a] with invariants

$$c_1^2(M) = 8s \quad \text{and} \quad \chi_h(M) = 4s + 3.$$

If $d > 1$ and t is an arbitrary integer, we get nonspin homotopy Horikawa surfaces with $s = td^2$ whose canonical classes have divisibility d. By the argument from before, these 4-manifolds cannot be diffeomorphic to complex Horikawa surfaces.

With different constructions, it is possible to find examples of simply connected symplectic 4-manifolds with canonical class of odd divisibility, $c_1^2 > 0$ and signature not divisible by 8; see [Hamilton 2008, Section VI.2.3]. However, many cases remain uncovered. For example, we could not answer this:

Question 24. For a given odd integer $d > 1$, is there a simply connected symplectic 4-manifold M with $c_1^2(M) = d^2$ whose canonical class has divisibility d?

Note that there is a trivial example for $d = 3$, namely $\mathbb{C}P^2$.

8. Construction of inequivalent symplectic structures

In this section we prove a result similar to [Smith 2000, Theorem 1.5], which can be used to show that certain 4-manifolds X admit inequivalent symplectic structures, where equivalence is defined as follows (see [McMullen and Taubes 1999]).

Definition 25. Two symplectic forms on a closed oriented 4-manifold M are called equivalent if they can be made identical by a combination of deformations through symplectic forms and orientation-preserving self-diffeomorphisms of M.

The canonical classes of equivalent symplectic forms have the same (maximal) divisibility as elements of $H^2(M; \mathbb{Z})$. This follows because deformations do not change the canonical class and the application of an orientation preserving self-diffeomorphism does not change the divisibility.

Lemma 26. *Let (M, ω) be a symplectic 4-manifold with canonical class K. Then the symplectic structure $-\omega$ has canonical class $-K$.*

Proof. Let J be an almost complex structure on M compatible with ω. Then $-J$ is an almost complex structure compatible with $-\omega$. The complex vector bundle $(TM, -J)$ is the conjugate bundle to (TM, J). By [Milnor and Stasheff 1974], this implies that $c_1(TM, -J) = -c_1(TM, J)$. Since the canonical class is minus the first Chern class of the tangent bundle, the claim follows. $\square$

Let $M_K \times S^1$ be a 4-manifold used in knot surgery, where K is a fibred knot of genus h. Let T_K be a section of the fibre bundle

$$\begin{tikzcd} M_K \times S^1 \arrow[d] & \Sigma_h \arrow[l] \\ T^2 & \end{tikzcd}$$

and let B_K be a fibre. We fix an orientation on T_K and choose the orientation on B_K so that $T_K \cdot B_K = +1$. There exist symplectic structures on $M_K \times S^1$ such that both the fibre and the section are symplectic. We can choose a symplectic structure ω^+ that restricts to both T_K and B_K as a positive volume form with respect to the orientations. It has canonical class $K^+ = (2h-2)T_K$ by the adjunction formula. We also define the symplectic form $\omega^- = -\omega^+$. It restricts to a negative volume form on T_K and B_K. The canonical class of this symplectic structure is $K^- = -(2h-2)T_K$. Let X be a closed oriented 4-manifold with torsion-free cohomology that contains an embedded oriented torus T_X of self-intersection 0, representing an indivisible homology class. We form the oriented 4-manifold

$$X_K = X \#_{T_X = T_K} (M_K \times S^1),$$

by doing the generalized fibre sum along the pair (T_X, T_K) of oriented tori. Suppose that X has a symplectic structure ω_X such that T_X is symplectic. We consider two cases: If ω_X restricts to a positive volume form on T_X, we can glue ω_X to the symplectic form ω^+ on $M_K \times S^1$ to get a symplectic structure $\omega^+_{X_K}$ on X_K. The canonical class of this symplectic structure is $K^+_{X_K} = K_X + 2hT_X$, as seen above; see Equation (3).

Lemma 27. *Suppose that ω_X restricts to a negative volume form on T_X. We can glue ω_X to the symplectic form ω^- on $M_K \times S^1$ to get a symplectic structure $\omega^-_{X_K}$ on X_K. The canonical class of $\omega^-_{X_K}$ is $K^-_{X_K} = K_X - 2hT_X$.*

Proof. We use Lemma 26 twice: The symplectic form $-\omega_X$ restricts to a positive volume form on T_X. We can glue this symplectic form to the symplectic form ω^+ on $M_K \times S^1$, which also restricts to a positive volume form on T_K. By the standard formula (3), the canonical class of the resulting symplectic form on X_K is

$$K = -K_X + 2hT_X.$$

The symplectic form $\omega^{-}_{X_K}$ we want to consider is *minus* the symplectic form we have just constructed. Hence its canonical class is $K^{-}_{X_K} = K_X - 2hT_X$. □

Lemma 28. *Suppose (M, ω) is a closed symplectic 4-manifold with canonical class K_M. Suppose M contains pairwise disjoint embedded oriented Lagrangian surfaces $T_1, \dots, T_{r+1}$ (with $r \geq 1$) such that*

- *the classes of the surfaces $T_1, \dots, T_r$ are linearly independent in $H_2(M; \mathbb{R})$, and*
- *the surface T_{r+1} is homologous to $a_1T_1 + \cdots + a_rT_r$, where all coefficients $a_1, \dots, a_r$ are positive integers.*

Then for every nonempty subset $S \subset \{T_1, \dots, T_r\}$, there exists a symplectic form ω_S on M such that

- *all surfaces $T_1, \dots, T_{r+1}$ are symplectic, and*
- *the symplectic form ω_S induces on the surfaces in S and the surface T_{r+1} a positive volume form and on the remaining surfaces in $\{T_1, \dots, T_r\} \setminus S$ a negative volume form.*

Also, the canonical classes of the symplectic structures ω_S are all equal to K_M. We can also assume that any given closed oriented surface in M that is disjoint from the surfaces $T_1, \dots, T_{r+1}$ and is symplectic with respect to ω stays symplectic for ω_S with the same sign as the induced volume form.

Proof. The proof is similar to the proof of [Gompf 1995, Lemma 1.6]. We can assume that $S = \{T_{s+1}, \dots, T_r\}$ with $s + 1 \leq r$. Let

$$c = \sum_{i=1}^{s} a_i \quad \text{and} \quad c' = \sum_{i=s+1}^{r-1} a_i.$$

Since the classes of the surfaces $T_1, \dots, T_r$ are linearly independent in $H_2(M; \mathbb{R})$ and $H^2_{DR}(M)$ is the dual space of $H_2(M; \mathbb{R})$, there exists a closed 2-form η on M such that

$$\int_{T_i} \eta = \begin{cases} -1 & \text{for } i = 1, \dots, s, \\ +1 & \text{for } i = s+1, \dots, r-1, \\ c+1 & \text{for } i = r, \\ c'+1 & \text{for } i = r+1, \end{cases}$$

Note that we can choose the value of η on $T_1, \dots, T_r$ arbitrarily. The value on T_{r+1} is then determined by $T_{r+1} = a_1T_1 + \cdots + a_rT_r$. We can choose symplectic forms ω_i on each T_i such that

$$\int_{T_i} \omega_i = \int_{T_i} \eta \quad \text{for all } i = 1, \dots, r+1.$$

The symplectic form ω_i induces on T_i a negative volume form if $i \leq s$ and a positive volume form if $i \geq s+1$. The difference $\omega_i - j_i^*\eta$, where $j_i : T_i \to M$ is the embedding, has vanishing integral and hence is an exact 2-form on T_i of the form $d\alpha_i$. We can extend each α_i to a small tubular neighbourhood of T_i in M, cut it off differentiably in a slightly larger tubular neighbourhood and extend by 0 to all of M. We can do this such that the tubular neighbourhoods of $T_1, \ldots, T_{r+1}$ are pairwise disjoint. Define the closed 2-form $\eta' = \eta + \sum_{i=1}^{r+1} d\alpha_i$ on M. Then

$$j_i^*\eta' = j_i^*\eta + d\alpha_i = \omega_i.$$

The closed 2-form $\omega' = \omega + t\eta'$ is symplectic for small values of t. Since the surfaces T_i are Lagrangian, we have $j_i^*\omega = 0$ and hence $j_i^*\omega' = t\omega_i$. This implies that ω' is for small values $t > 0$ a symplectic form on M that induces a volume form on T_i of the same sign as ω_i for all $i = 1, \ldots, r+1$. The claim about the canonical class follows because the symplectic structures ω_S are constructed by a deformation of ω. We can also choose $t > 0$ small enough so that ω' still restricts to a symplectic form on any given symplectic surface disjoint from the tori without changing the sign of the induced volume form on this surface. $\square$

This construction will be used as follows: Suppose that (V_1, ω_1) and (V_2, ω_2) are symplectic 4-manifolds such that V_1 contains an embedded Lagrangian torus T_1 and V_2 contains an embedded symplectic torus T_2, both oriented and of self-intersection 0. Let W denote the smooth oriented 4-manifold $V_1 \#_{T_1=T_2} V_2$ obtained as a generalized fibre sum. By Lemma 28, there exist small perturbations of ω_1 to new symplectic forms ω_1^+ and ω_1^- on the manifold V_1 such that the torus T_1 becomes symplectic with positive and negative induced volume form, respectively. By the Gompf construction, it is then possible to define two symplectic forms on the same oriented 4-manifold W:

- The symplectic forms ω_1^+ and ω_2 determine a symplectic form on W.
- The symplectic forms ω_1^- and $-\omega_2$ determine a symplectic form on W.

Hence the symplectic forms on the first manifold differ only by a small perturbation, while on the second manifold they differ by the sign. Similarly, the canonical classes of both perturbed symplectic forms on V_1 are the same, while they differ by the sign on V_2. If additional tori exist and suitable fibre sums are performed, it is possible to end up with two or more inequivalent symplectic forms on the same 4-manifold, distinguished by the divisibilities of their canonical classes.

To define the configuration of tori we want to consider, recall that the nucleus $N(n)$ is the smooth manifold with boundary defined as a regular neighbourhood of a cusp fibre and a section in the simply connected elliptic surface $E(n)$; see [Gompf 1991]. It contains an embedded torus given by a regular fibre homologous to the cusp. It also contains two embedded disks of self-intersection -1 that bound vanishing cycles on the torus. The vanishing cycles are the simple closed loops given by the factors in $T^2 = S^1 \times S^1$.

Definition 29 (Lagrangian triple). Let (M, ω) be a symplectic 4-manifold. Given an integer $a \geq 1$, a *Lagrangian triple* consists of three pairwise disjoint oriented Lagrangian tori T_1, T_2 and R embedded in M with the following properties:

- All three tori have self-intersection zero and represent indivisible classes in integral homology.
- T_1 and T_2 are linearly independent over $\mathbb{Q}$ and R is homologous to $aT_1 + T_2$.
- There exists an embedded nucleus $N(2) \subset M$ that contains R, corresponding to a general fibre. Let S denote the 2-sphere in $N(2)$ of self-intersection -2, corresponding to a section. In addition to intersecting R, this sphere intersects T_2 transversely once. The torus T_2 is disjoint from the vanishing disks of R, coming from the cusp in $N(2)$.
- The torus T_1 is disjoint from the nucleus $N(2)$ above, and there exists an embedded 2-sphere S_1 in M, also disjoint from $N(2)$, that intersects T_1 transversely and positively once.

See Figure 1. The assumptions imply that $S_1T_2 = S_1(R - aT_1) = -a$.

Example 30. Let M be the elliptic surface $E(n)$ with $n \geq 2$. In this example we show that $E(n)$ contains $n-1$ disjoint Lagrangian triples (T_1^i, T_2^i, R^i) as above, where R^i is homologous to $a_iT_1^i + T_2^i$ for $i = 1, \dots, n-1$. The integers $a_i > 0$ can be chosen arbitrarily and for each triple independently. In this case both T_1^i and R^i are contained in disjoint embedded nuclei $N(2)$. Together with their dual 2-spheres they realize $2(n-1)$ H-summands in the intersection form of $E(n)$. In

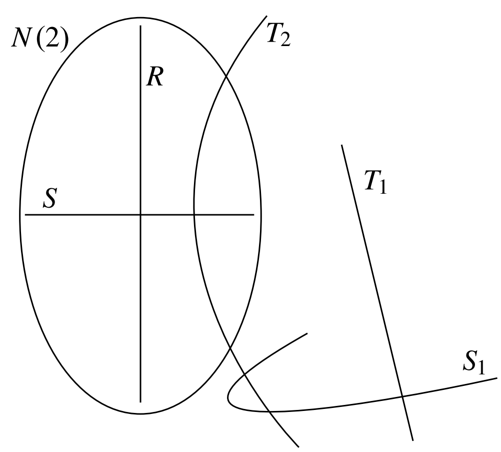

Figure 1. Lagrangian triple.

particular, the tori in different triples are linearly independent. We can also ensure that all Lagrangian tori and the 2-spheres that intersect them once are disjoint from the nucleus $N(n) \subset E(n)$, defined as a regular neighbourhood of a cusp fibre and a section in $E(n)$.

The construction is similar to [Gompf and Mrowka 1993, Section 2] and is done by induction. Suppose the Lagrangian triples are already constructed for $E(n)$ and consider a splitting of $E(n+1)$ as a fibre sum $E(n+1) = E(n) \#_{F=F} E(1)$ along general fibres F. We choose fibred tubular neighbourhoods for the general fibres in $E(n)$ and $E(1)$. The boundary of $E(1) \setminus \operatorname{int} \nu F$ is diffeomorphic to $F \times S^1$. Let γ_1 and γ_2 be two simple closed loops spanning the torus F, and let m be the meridian to F that spans the remaining S^1 factor. We consider the three tori

$$V_0 = \gamma_1 \times \gamma_2, \quad V_1 = \gamma_1 \times m, \quad V_2 = \gamma_2 \times m.$$

The tori are made disjoint by pushing them inside a collar of the boundary into the interior of $E(1) \setminus \operatorname{int} \nu F$ such that V_2 is the innermost and V_0 the outermost (closest to the boundary). The torus V_0 can be assumed symplectic, while V_1 and V_2 are rim tori that can be assumed Lagrangian. Similarly the boundary of $E(n) \setminus \operatorname{int} \nu F$ is diffeomorphic to $F \times S^1$, where F is spanned by the circles γ_1 and γ_2 and S^1 by the circle m and corresponding circles get identified in the gluing of the fibre sum. In the interior of $E(n) \setminus \operatorname{int} \nu F$ we consider three tori V_0, V_1, V_2 as above which get identified with the corresponding tori on the $E(1)$ side in the gluing. On the $E(n)$ side, the torus V_0 is the innermost and V_2 the outermost.

We can choose elliptic fibrations such that near the general fibre F there exist two cusp fibres in $E(1)$ and three cusp fibres in $E(n)$. This is possible because $E(m)$ has an elliptic fibration with $6m$ cusp fibres for all m; see [Gompf and Stipsicz 1999, Corollary 7.3.23]. The corresponding vanishing disks can be assumed pairwise disjoint. We can also choose three disjoint sections for the elliptic fibration on $E(1)$ and one section for $E(n)$.

The nuclei can now be defined as follows: The nucleus $N(n+1)$ containing V_0 has a dual $-(n+1)$-sphere sewed together from sections on each side of the fibre sum. The vanishing disks for V_0 come from the first cusp in $E(n)$. The nucleus $N(2)$ containing V_1 has a dual -2-sphere sewed together from two vanishing cycles parallel to γ_2 coming from the first cusp in $E(1)$ and the second cusp in $E(n)$. The vanishing disks for V_1 come from the second section of $E(1)$ and from the vanishing cycle parallel to γ_1 of the second cusp in $E(n)$. The nucleus $N(2)$ containing V_2 has a dual -2-sphere sewed together from two vanishing cycles parallel to γ_1 coming from the second cusp in $E(1)$ and the third cusp in $E(n)$. The vanishing disks for V_2 come from the third section of $E(1)$ and from the vanishing cycle parallel to γ_2 of the second cusp in $E(1)$.

To define the Lagrangian triple (T_1, T_2, R), let $T_1 = V_1$ and $R = V_2$. Denote by $c_a : S^1 \to F = S^1 \times S^1$ the embedded curve given by the $(-a, 1)$-torus knot, and let T_2 denote the Lagrangian rim torus $T_2 = c_a \times m$ in the collar above. Then T_2 represents the class $-aT_1 + R$; hence $R = aT_1 + T_2$. The torus T_2 has one positive transverse intersection with the sphere in the nucleus containing R and a negative transverse intersections with the sphere in the nucleus containing T_1.

Remark 31. To find more general examples of symplectic 4-manifolds containing Lagrangian triples, suppose that Y is an arbitrary closed symplectic 4-manifold that contains an embedded symplectic torus T_Y of self-intersection 0, representing an indivisible class. Then the symplectic generalized fibre sum $Y \#_{T_Y = F} E(n)$ also contains $n - 1$ Lagrangian triples.

Suppose (M, ω) is a simply connected symplectic 4-manifold that contains a Lagrangian triple T_1, T_2, R. Let K_1 and K_2 be fibred knots of genera h_1 and h_2 to be chosen later. Consider the associated oriented 4-manifolds $M_{K_i} \times S^1$ as in the knot surgery construction, and denote sections of the fibre bundles

$$\begin{tikzcd} M_{K_i} \times S^1 \arrow[d] & \Sigma_{h_i} \arrow[l] \\ T^2 & \end{tikzcd}$$

by T_{K_i}, which are tori of self-intersection 0. We choose an orientation on each torus T_{K_i}. Note that the Lagrangian tori T_1 and T_2 in M are oriented *a priori*.

We construct a smooth oriented 4-manifold X in three steps as follows: For an integer $m \geq 1$, consider the elliptic surface $E(m)$ and denote an oriented general fibre by F. Let M_0 denote the smooth generalized fibre sum $M_0 = E(m) \#_{F=R} M$. The gluing diffeomorphism is chosen as follows: The push-offs R' and F' into the boundary of the tubular neighbourhoods νR and νF each contain a pair of vanishing cycles. We choose the gluing so that the push-offs and the vanishing cycles get identified. The corresponding vanishing disks then sew together pairwise to give two embedded spheres of self-intersection -2 in M_0, which can be assumed disjoint by choosing two different push-offs given by the same trivializations.

Denote the torus in M_0 coming from the push-off R' by R_0. Consider the tori T_1 and T_2 in M_0 . Then R_0 is still homologous to $aT_1 + T_2$ in M_0, because the difference could only be a rim torus by [Hamilton 2008, Section V.3], which must have nonzero intersection with one of the two vanishing spheres in M_0. This is excluded by our assumptions on Lagrangian triples. In the second step of the construction, we do a knot surgery with the fibred knot K_1 along the torus T_1 in M_0 to get the oriented 4-manifold $M_1 = M_0 \#_{T_1 = T_{K_1}} (M_{K_1} \times S^1)$. The manifold M_1 contains a torus, which we still denote by T_2. We do a knot surgery with the fibred knot K_2 along the torus T_2 to get the oriented 4-manifold $X = M_1 \#_{T_2=T_{K_2}} (M_{K_2} \times S^1)$.

Lemma 32. *The closed oriented* 4*-manifold*

$$X = E(m) \#_{F=R} M \#_{T_1=T_{K_1}} (M_{K_1} \times S^1) \#_{T_2=T_{K_2}} (M_{K_2} \times S^1)$$

is simply connected.

Proof. The existence of the sphere S shows that $M \setminus R$ is simply connected. Since $E(m) \setminus F$ is simply connected, it follows that M_0 is simply connected.

The sphere S and a section for the elliptic fibration on $E(m)$ sew together to give an embedded sphere S_2 in M_0 of self-intersection $-(m+2)$. The sphere S_1 in M is disjoint from R and hence is still contained in M_0. These spheres have the following intersections:

- The sphere S_1 intersects T_1 transversely once, has intersection $-a$ with T_2, and is disjoint from R_0.
- The sphere S_2 intersects R_0 and T_2 transversely once and is disjoint from T_1.

The sphere S_1 shows that $M_0 \setminus T_1$ is simply connected and hence M_1 is simply connected. The sphere S_2 in M_0 is disjoint from T_1 and hence is still contained in M_1 and intersects T_2 once. By the same argument, this shows that the manifold X is simply connected. □

We define two symplectic forms ω_X^+ and ω_X^- on X: By Lemma 28 there exist two symplectic structures ω_+ and ω_- on M with the same canonical class K_M as ω such that

- the tori T_1, T_2 and R are symplectic with respect to both symplectic forms,
- the form ω_+ induces on T_1, T_2 and R a positive volume form, and
- the form ω_- induces on T_1 a negative volume form and on T_2 and R a positive volume form.

We can also choose the sphere S to be symplectic with positive volume form in both cases.

On the elliptic surface $E(m)$, we can choose a symplectic (Kähler) form ω_E that restricts to a positive volume form on the oriented fibre F. It has canonical class $K_E = (m-2)F$. We can glue both symplectic forms ω_+ and ω_- on M to the symplectic form ω_E on $E(m)$ to get symplectic forms ω_0^+ and ω_0^- on the 4-manifold M_0. The canonical class for both symplectic forms on M_0 is given by $K_{M_0} = K_M + mR_0$; see [Fintushel and Stern 2001, proof of Lemma 2.2]. Since rim tori exist in this fibre sum, Theorem 11 cannot be applied directly. However, the formula remains correct because rim tori do not contribute in this case; for details see [Hamilton 2008, Section V.6.1].

We want to extend the symplectic forms to the 4-manifold X: We choose in each fibre bundle $M_{K_i}\times S^1$ a fibre B_{K_i} and orient the surface B_{K_i} so that $T_{K_i}\cdot B_{K_i}=+1$ with the chosen orientation on T_{K_i}. There exist symplectic structures on the closed 4-manifolds $M_{K_i}\times S^1$ such that both the section and the fibre are symplectic. On $M_{K_1}\times S^1$, we choose two symplectic forms $\omega_1^{\pm}$: The form ω_1^{+} induces a positive volume form on both T_{K_1} and B_{K_1}. It has canonical class $K_1^{+}=(2h_1-2)T_{K_1}$. The form ω_1^{-} is given by $-\omega_1^{+}$. It induces a negative volume form on both T_{K_1} and B_{K_1} and has canonical class $K_1^{-}=-(2h_1-2)T_{K_1}$.

On the manifold $M_{K_2}\times S^1$ we only choose a symplectic form ω_2 that induces a positive volume form on T_{K_2} and B_{K_2}. The canonical class is $K_2=(2h_2-2)T_{K_2}$.

The oriented torus T_1 in M_0 is symplectic for both forms $\omega_0^{\pm}$ constructed as above so that ω_0^{+} induces a positive volume form and ω_0^{-} a negative volume form. Gluing ω_0^{+} to ω_1^{+} and ω_0^{-} to ω_1^{-}, Lemma 27 implies that the closed oriented 4-manifold M_1 has two symplectic structures with canonical classes

$$K_{M_1}^{+}=K_M+mR_0+2h_1T_1 \quad\text{and}\quad K_{M_1}^{-}=K_M+mR_0-2h_1T_1.$$

The torus T_2 can be considered as a symplectic torus in M_1 such that both symplectic structures induce positive volume forms, since we can assume that the symplectic forms on M_1 are still of the form ω_0^{+} and ω_0^{-} in a neighbourhood of T_2. Hence on the generalized fibre sum $X=M_1\,\#_{T_2=T_{K_2}}\,M_{K_2}\times S^1$, we can glue each of the two symplectic forms on M_1 to the symplectic form ω_2 on $M_{K_2}\times S^1$. We get two symplectic structures on X with canonical classes

$$\begin{aligned}K_X^{+}&=K_M+mR_0+2h_1T_1+2h_2T_2,\\ K_X^{-}&=K_M+mR_0-2h_1T_1+2h_2T_2.\end{aligned}$$

This can be written using $R_0=aT_1+T_2$ as

$$\begin{aligned}K_X^{+}&=K_M+(2h_1+am)T_1+(2h_2+m)T_2,\\ K_X^{-}&=K_M+(-2h_1+am)T_1+(2h_2+m)T_2.\end{aligned}$$

Theorem 33. *Suppose (M,ω) is a simply connected symplectic 4-manifold that contains a Lagrangian triple T_1,T_2,R such that R is homologous to aT_1+T_2. Let m be a positive integer, and let K_1 and K_2 be fibred knots of genus h_1 and h_2. Then the closed oriented 4-manifold*

$$X=E(m)\,\#_{F=R}\,M\,\#_{T_1=T_{K_1}}(M_{K_1}\times S^1)\,\#_{T_2=T_{K_2}}(M_{K_2}\times S^1)$$

is simply connected and admits symplectic structures ω_X^{+} and ω_X^{-} with canonical classes

$$\begin{aligned}K_X^{+}&=K_M+(2h_1+am)T_1+(2h_2+m)T_2,\\ K_X^{-}&=K_M+(-2h_1+am)T_1+(2h_2+m)T_2.\end{aligned}$$

Remark 34. Instead of doing the generalized fibre sum with $E(m)$ in the first step of the construction, we could also do a knot surgery with a fibred knot K_0 of genus $h_0 \geq 1$. This has the advantage that both c_1^2 and the signature do not change under the construction. However, the sphere S_2 in M_0 is then replaced by a surface of genus h_0 sewed together from the sphere S in M and a Seifert surface for K_0. Hence it is no longer clear that $M_1 \setminus T_2$ and X are simply connected.

The following two surfaces are useful for determining the divisibility of the canonical classes in Theorem 33.

Lemma 35. *There is an oriented surface C_2 in X that has intersection $C_2T_2 = 1$ and is disjoint from T_1.*

The surface C_2 is sewed together from the sphere S_2 and a Seifert surface for K_2.

Lemma 36. *There is an oriented surface C_1 in X that has intersection $C_1T_1 = 1$ and is disjoint from T_2.*

Proof. The surface C_1 can be constructed explicitly as follows: In the nucleus $N(2) \subset M$ containing R, we can find a surface of some genus homologous to aS and intersecting both R and T_2 in a positive transverse intersections. Tubing this surface to the sphere S_1, we get a surface A in M that has intersection number $AT_2 = 0$ and intersects T_1 transversely once. By increasing the genus we can make A disjoint from T_2. The surface A still intersects the torus R at a points. Sewing the surface A to a surface in $E(m)$ homologous to a times a section, we get a surface B in M_0 disjoint from T_2 and intersecting T_1 once. Sewing this surface to a Seifert surface for K_1 we get a surface C_1 in X with $C_1T_1 = 1$ disjoint from T_2. □

9. Examples of inequivalent symplectic structures

Definition 37 (the set Q). Let $N \geq 0$ and $d \geq 1$ be integers, and let $d_0, \ldots, d_N$ be positive integers dividing d, where $d = d_0$. If d is even, assume that all $d_1, \ldots, d_N$ are even. We define a set Q of positive integers as follows:

- If d is either odd or not divisible by 4, let Q be the set consisting of the greatest common divisors of all (nonempty) subsets of $\{d_0, \ldots, d_N\}$.
- If d is divisible by 4, we can assume by reordering that $d_1, \ldots, d_s$ are those elements such that d_i is divisible by 4, while $d_{s+1}, \ldots, d_N$ are those elements such that d_i is not divisible by 4, where $s \geq 0$ is some integer. Then Q is defined as the set of integers consisting of the greatest common divisors of all (nonempty) subsets of $\{d_0, \ldots, d_s, 2d_{s+1}, \ldots, 2d_N\}$.

We can now state the main theorem on the existence of inequivalent symplectic structures on homotopy elliptic surfaces.

Theorem 38. *Let N and $d \geq 1$ be integers, and let $d_0, \dots, d_N$ be positive integers dividing d as in Definition 37. Let Q be the associated set of greatest common divisors. Choose an integer $n \geq 3$ as follows*:

- *If d is odd, let n be an arbitrary integer with $n \geq 2N+1$.*
- *If d is even, let n be an even integer with $n \geq 3N+1$.*

Then there exists a homotopy elliptic surface W with $\chi_h(W) = n$ and the property that for each integer $q \in Q$, the manifold W admits a symplectic structure whose canonical class K has divisibility equal to q. Hence W admits at least $|Q|$ inequivalent symplectic structures.

Proof. The proof splits into three cases depending on the parity of d. In each case we follow the construction in Section 8, starting from the manifold $M = E(l)$, where l is an integer no less than $N+1$. By Example 30, $E(l)$ contains N pairwise disjoint Lagrangian triples T_1^i, T_2^i, R^i, where R^i is homologous to $a_i T_1^i + T_2^i$ for indices $i = 1, \dots, N$. The construction is done on each triple separately[2] and involves knot surgeries along T_1^i and T_2^i with fibred knots of respective genus h_i and h, as well as fibre summing with elliptic surfaces $E(m)$ along the tori R^i. The numbers a_i, h_i, h and m will be fixed in each case.

Case: d is odd. Then all divisors $d_1, \dots, d_N$ are odd. Consider the integers

$$\begin{aligned} m &= 1, & h &= \tfrac{1}{2}(d-1), \\ a_i &= d + d_i, & h_i &= \tfrac{1}{2}(d - d_i) \quad \text{for } 1 \leq i \leq N. \end{aligned}$$

Let l be an integer no less than $N+1$ and do the construction above, starting from the elliptic surface $E(l)$. We get a (simply connected) homotopy elliptic surface X with $\chi_h(X) = l+N$. By Theorem 33 the 4-manifold X has 2^N symplectic structures with canonical classes

$$\begin{aligned} K_X &= (l-2)F + \sum_{i=1}^{N} \big((\pm 2h_i + a_i)T_1^i + (2h+1)T_2^i\big) \\ &= (l-2)F + \sum_{i=1}^{N} \big((\pm(d-d_i) + d + d_i)T_1^i + dT_2^i\big). \end{aligned}$$

Here F denotes the torus in X coming from a general fibre in $E(l)$ and the $\pm$-signs in each summand can be varied independently. We can assume that F is symplectic with positive induced volume form for all 2^N symplectic structures on X. Consider the even integer $l(d-1)+2$, and let K be a fibred knot of genus $g = \frac{1}{2}(l(d-1)+2)$. We do knot surgery with K along the symplectic torus F to get a homotopy elliptic

[2]This is only a small generalization of Lemma 28, because the construction in the proof of this lemma changes the symplectic structure only in a small neighbourhood of the Lagrangian surfaces.

surface W with $\chi_h(W) = l + N$ having 2^N symplectic structures whose canonical classes are given by

$$\begin{aligned} K_W &= (l-2+2g)F + \sum_{i=1}^{N} \left((\pm(d-d_i)+d+d_i)T_1^i + dT_2^i\right) \\ &= dlF + \sum_{i=1}^{N} \left((\pm(d-d_i)+d+d_i)T_1^i + dT_2^i\right). \end{aligned}$$

Suppose that $q \in Q$ is the greatest common divisor of certain elements $\{d_i\}_{i\in I}$, where I is a nonempty subset of $\{0, \dots, N\}$. Let J be the complement of I in $\{0, \dots, N\}$. Choosing the minus sign for each i in I and the plus sign for each i in J defines a symplectic structure ω_I on W with canonical class given by

$$K_W = dlF + \sum_{i\in I}(2d_iT_1^i + dT_2^i) + \sum_{j\in J}(2dT_1^j + dT_2^j).$$

We claim that the divisibility of K_W is equal to q. Since q divides d and all integers d_i for $i \in I$, the class K_W is divisible by q. Considering separately the surfaces from Lemmas 35 and 36 for each Lagrangian triple implies that every number that divides K_W is odd (since it divides d) and a common divisor of all d_i with indices $i \in I$. This proves the claim in this case.

Case: d is even but not divisible by 4. We can write $d = 2k$ and $d_i = 2k_i$ for all $i = 1, \dots, N$. The assumption implies that all integers k, k_i are odd. Consider the integers defined by

$$\begin{aligned} m &= 2, & h &= k-1, \\ a_i &= \tfrac{1}{2}(k+k_i), & h_i &= \tfrac{1}{2}(k-k_i). \end{aligned}$$

Let l be an even integer no less than $N+1$ and consider the construction above, starting from $E(l)$. We get a homotopy elliptic surface X with $\chi_h(X) = l + 2N$. The 4-manifold X has 2^N symplectic structures with canonical classes

$$\begin{aligned} K_X &= (l-2)F + \sum_{i=1}^{N}\left((\pm 2h_i + 2a_i)T_1^i + (2h+2)T_2^i\right) \\ &= (l-2)F + \sum_{i=1}^{N}\left((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\right). \end{aligned}$$

Consider a fibred knot K of genus $g = \frac{1}{2}(l(d-1)+2)$, noting that l is even. Doing knot surgery with K along the symplectic torus F in X, we get a homotopy elliptic surface W with $\chi_h(W) = l+2N$ having 2^N symplectic structures whose canonical classes are

$$\begin{aligned}K_W &= (l-2+2g)F + \sum_{i=1}^{N}\left((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\right)\\ &= dlF + \sum_{i=1}^{N}\left((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\right).\end{aligned}$$

Let $q \in Q$ be the greatest common divisor of elements d_i, where $i \in I$ for some nonempty index set I with complement J in $\{0, \dots, N\}$. Choosing the plus and minus signs as before, we get a symplectic structure ω_I on W with canonical class

$$K_W = dlF + \sum_{i\in I}(d_iT_1^i + dT_2^i) + \sum_{j\in J}(dT_1^i + dT_2^i). \tag{6}$$

As above, it follows that the canonical class of ω_I has divisibility equal to q.

Case: d is divisible by 4. We can write $d=2k$ and $d_i=2k_i$ for all $i=1,\dots,N$. We can assume that the divisors are ordered as in Definition 37, that is, $d_1,\dots,d_s$ are those elements such that d_i is divisible by 4 while $d_{s+1},\dots,d_N$ are those elements such that d_i is not divisible by 4. This is equivalent to $k_1,\dots,k_s$ being even and $k_{s+1},\dots,k_N$ odd. Consider the integers defined by

$$\begin{aligned} &a_i = \tfrac{1}{2}(k+k_i) \quad \text{and} \quad h_i = \tfrac{1}{2}(k-k_i) \quad \text{for } i=1,\dots,s,\\ &a_i = \tfrac{1}{2}(k+2k_i) \quad \text{and} \quad h_i = \tfrac{1}{2}(k-2k_i) \quad \text{for } i=s+1,\dots,N.\end{aligned}$$

We also define $m=2$ and $h=k-1$. Let l be an even integer $\geq N+1$. We consider the same construction as above starting from $E(l)$ to get a homotopy elliptic surface X with $\chi_h(X)=l+2N$ that has 2^N symplectic structures with canonical classes given by the formula

$$\begin{aligned}K_X &= (l-2)F + \sum_{i=1}^{N}\left((\pm 2h_i + 2a_i)T_1^i + (2h+2)T_2^i\right)\\ &= (l-2)F + \sum_{i=1}^{s}\left((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\right)\\ &\qquad\qquad + \sum_{i=s+1}^{N}\left((\pm(k-2k_i)+k+2k_i)T_1^i + dT_2^i\right).\end{aligned}$$

We then do knot surgery with a fibred knot K of genus $g=\frac{1}{2}(l(d-1)+2)$ along the symplectic torus F in X to get a homotopy elliptic surface W with $\chi_h(W)=l+2N$

having 2^N symplectic structures whose canonical classes are

$$\begin{aligned} K_W &= (l-2+2g)F + \sum_{i=1}^{N}\big((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\big) \\ &= dlF + \sum_{i=1}^{s}\big((\pm(k-k_i)+k+k_i)T_1^i + dT_2^i\big) \\ &\qquad\qquad + \sum_{i=s+1}^{N}\big((\pm(k-2k_i)+k+2k_i)T_1^i + dT_2^i\big). \end{aligned} \tag{7}$$

Let q be an element in Q. Note that this time

$$\begin{aligned} (k-k_i)+(k_i+k) &= d \quad \text{and} \quad -(k-k_i)+(k_i+k) = d_i \quad \text{for } i \le s, \\ (k-2k_i)+(k+2k_i) &= d \quad \text{and} \quad -(k-2k_i)+(k+2k_i) = 2d_i \quad \text{for } i \ge s+1. \end{aligned}$$

Since q is the greatest common divisor of certain elements d_i for $i \le s$ and $2d_i$ for $i \ge s+1$, it follows as above that we can choose the plus and minus signs appropriately to get a symplectic structure ω_I on W whose canonical class has divisibility equal to q. ☐

Example 39. Suppose $d = 45$ and choose $d_0 = 45$, $d_1 = 15$, $d_2 = 9$, $d_3 = 5$. Then $Q = \{45, 15, 9, 5, 3, 1\}$, and for every integer $n \ge 7$ there exists a homotopy elliptic surfaces W with $\chi_h(W) = n$ that admits at least 6 inequivalent symplectic structures whose canonical classes have divisibilities given by the elements in Q. One can also find an infinite family of homeomorphic but nondiffeomorphic manifolds of this kind.

Corollary 40. *Let $m \ge 1$ be an arbitrary integer.*

- *There exist simply connected nonspin 4-manifolds W homeomorphic to the elliptic surfaces $E(2m+1)$ and $E(2m+2)_2$ that admit at least 2^m inequivalent symplectic structures.*
- *There exist simply connected spin 4-manifolds W homeomorphic to $E(6m-2)$ and $E(6m)$ that admit at least 2^{2m-1} inequivalent symplectic structures, and there are spin manifolds homeomorphic to $E(6m+2)$ that admit at least 2^{2m} inequivalent symplectic structures.*

Proof. Choose N pairwise different odd prime numbers $p_1, \dots, p_N$. Let $d = d_0 = p_1 \cdots p_N$, and consider the integers d_i obtained for $i = 1, \dots, N$ by dividing d by the prime p_i. Then the associated set Q of greatest common divisors consists of all products of the p_i where each prime occurs at most once: If such a product x does not contain precisely the primes $p_{i_1}, \dots, p_{i_r}$ then x is the greatest common divisor of $d_{i_1}, \dots, d_{i_r}$. The set Q has 2^N elements.

Let $m \geq 1$ be an arbitrary integer. If $N = m$, there exists by Theorem 38 for every integer $n \geq 2N+1 = 2m+1$ a homotopy elliptic surface W with $\chi_h(W) = n$ that has 2^m symplectic structures realizing all elements in Q as the divisibility of their canonical classes. Since d is odd, the 4-manifolds W are nonspin.

If $N = 2m-1$, there exists for every even integer $n \geq 3N+1 = 6m-2$ a homotopy elliptic surface W with $\chi_h(W) = n$ that has 2^{2m-1} symplectic structures realizing all elements in Q multiplied by 2 as the divisibility of their canonical classes. Since all divisibilities are even, the manifold W is spin. If $N = 2m$, we can choose $n = 6m+2$ to get a spin homotopy elliptic surface W with $\chi_h(W) = 6m+2$ and 2^{2m} inequivalent symplectic structures. □

We can extend construction in the proof of Theorem 38 to the spin manifolds in Theorem 19 with $c_1^2 > 0$:

Theorem 41. *Let $N \geq 1$ be an integer. Suppose that $d \geq 2$ is an even integer and $d_0, \ldots, d_N$ are positive even integers dividing d as in Definition 37. Let Q be the associated set of greatest common divisors. Let m be an integer such that $2m \geq 3N+2$, and let $t \geq 1$ be an arbitrary integer. Then there exists a simply connected closed spin 4-manifold W with invariants*

$$c_1^2(W) = 2td^2, \quad e(W) = td^2 + 24m, \quad \sigma(W) = -16m,$$

and the property that for each integer $q \in Q$, the manifold W admits a symplectic structure whose canonical class K has divisibility equal to q. Hence W admits at least $|Q|$ inequivalent symplectic structures.

Proof. Let $l = 2m - 2N$. By the construction of Theorem 19, there exists a simply connected symplectic spin 4-manifold X with invariants

$$c_1^2(X) = 2td^2, \quad e(X) = td^2 + 12l, \quad \sigma(X) = -8l, \quad K_X = d(\tfrac{1}{2}lF + R + t\Sigma_M).$$

In particular, the canonical class of X has divisibility d. In the construction of X starting from the elliptic surface $E(l)$, we have only used one Lagrangian rim torus. Hence $l-2$ of the $l-1$ triples of Lagrangian rim tori in $E(l)$ (see Example 30) remain unchanged. Note that $l-2 \geq N$ by our assumptions. Since the symplectic form on $E(l)$ in a neighbourhood of these tori does not change in the construction of X by the Gompf fibre sum, we can assume that X contains at least N triples of Lagrangian tori as in the proof of Theorem 38. We can now use the same construction as in this theorem on the N triples of Lagrangian tori in X to get a simply connected spin 4-manifold W with invariants

$$\begin{aligned} c_1^2(W) = 2td^2, \quad e(W) &= td^2 + 12l + 24N \qquad \sigma(W) = -8l - 16N = -16m, \\ &= td^2 + 24m, \end{aligned}$$

admitting 2^N symplectic structures. In particular, for each $q \in Q$ the manifold W admits a symplectic structure ω_I whose canonical class is given by (6) and (7) if the term dlF is replaced by $K_X = d(\frac{1}{2}lF + R + t\Sigma_M)$. It follows again that the canonical class of ω_I has divisibility precisely equal to q. □

Corollary 42. *Let $d \geq 6$ be an even integer, and let $t \geq 1$ and $m \geq 3$ be arbitrary integers. Then there exists a simply connected closed spin 4-manifold W with invariants*

$$c_1^2(W) = 2td^2, \quad e(W) = td^2 + 24m, \quad \sigma(W) = -16m$$

such that W admits at least two inequivalent symplectic structures.

This follows with $N = 1$ and choosing $d_0 = d$ and $d_1 = 2$, since in this case Q consists of two elements.

Example 43. We consider Corollary 42 for the spin homotopy Horikawa surfaces in Example 20. Let $t \geq 1$ and $k \geq 3$ be arbitrary odd integers, and define an integer m by $2m = 3tk^2 + 3$. Let $d = 2k$ and $d_1 = 2$. Since $d = 2k$ is not divisible by 4, the set Q is equal to $\{2k, 2\}$ by Definition 37. Hence there exists a spin homotopy Horikawa surface X on the Noether line with invariants $c_1^2(X) = 8tk^2$ and $\chi_h(X) = 4tk^2 + 3$, and admitting two inequivalent symplectic structures: the canonical class of the first symplectic structure has divisibility $2k$, while that of the second is divisible only by 2.

Similarly, we can extend the construction in Theorem 41 to the nonspin manifolds in Theorem 22 with $c_1^2 > 0$:

Theorem 44. *Let $N \geq 1$ be an integer. Suppose $d \geq 3$ is an odd integer, and let $d_0, \ldots, d_N$ be positive integers dividing d as in Definition 37. Let Q be the associated set of greatest common divisors. Let $m \geq 2N + 2$ and $t \geq 1$ be arbitrary integers. Then there exists a simply connected closed nonspin 4-manifold W with invariants*

$$c_1^2(W) = 8td^2, \quad e(W) = 4td^2 + 12m, \quad \sigma(W) = -8m$$

and the property that for each integer $q \in Q$, the manifold W admits a symplectic structure whose canonical class K has divisibility equal to q. Hence W admits at least $|Q|$ inequivalent symplectic structures.

Proof. The proof is analogous to the proof of Theorem 41. Let $l = m - N$. By the construction of Theorem 22, there exists a simply connected nonspin symplectic 4-manifold X with invariants

$$c_1^2(X) = 8td^2, \quad e(X) = 4td^2 + 12l, \quad \sigma(X) = -8l$$

whose canonical class K_X has divisibility d. The manifold X contains $l-2$ triples of Lagrangian tori. By our assumptions, $l-2 \geq N$. Hence we can perform the construction in Theorem 38 (for d odd) to get a simply connected nonspin 4-manifold W with invariants

$$c_1^2(W) = 8td^2, \quad e(W) = 4td^2 + 12l + 12N \qquad \sigma(X) = -8l - 8N = -8m.$$
$$= 4td^2 + 12m,$$

The 4-manifold W admits for every integer $q \in Q$ a symplectic structure whose canonical class has divisibility equal to q. □

Choosing $N = 1$, $d_0 = d$ and $d_1 = 1$, the set Q contains two elements.

Corollary 45. *Let $d \geq 3$ be an odd integer, and let $t \geq 1$ and $m \geq 4$ be integers. Then there exists a simply connected closed nonspin 4-manifold W with invariants*

$$c_1^2(W) = 8td^2, \quad e(W) = 4td^2 + 12m, \quad \sigma(W) = -8m$$

such that W admits at least two inequivalent symplectic structures.

10. Branched coverings

Let M^n be a closed, oriented smooth manifold, and let F^{n-2} be a closed, oriented submanifold of codimension 2. Suppose the fundamental class $[F] \in H_{n-2}(M; \mathbb{Z})$ is divisible by an integer $m > 1$ and choose a class $B \in H_{n-2}(M; \mathbb{Z})$ such that $[F] = mB$. The integer m together with B determine a branched covering of M.

Definition 46. We denote by $\phi : M(F, B, m) \to M$ the m-fold branched covering of M branched over F and determined by m and B.

For the construction of branched coverings, see [Hirzebruch 1969]. The smooth manifold $M(F, B, m)$ has the properties that

- over the complement $M' = M \setminus F$, the map $\phi : \phi^{-1}(M') \to M'$ is a standard m-fold cyclic covering;
- ϕ maps the submanifold $\overline{F} = \phi^{-1}(F)$ diffeomorphically onto F, and on tubular neighbourhoods of $\overline{F}$ and F, the map $\phi : \nu(\overline{F}) \to \nu(F)$ is locally of the form
$$U \times D^2 \to U \times D^2, \quad (x, z) \mapsto (x, z^m),$$
where D^2 is considered as the unit disk in $\mathbb{C}$.

Suppose M is a smooth complex algebraic surface, and let $D \subset M$ be a smooth connected complex curve. If $m > 0$ is an integer that divides $[D]$ and $B \in H_2(M; \mathbb{Z})$ is a homology class such that $[D] = mB$, then the branched covering $M(D, B, m)$ also admits the structure of an algebraic surface.

Proposition 47. *Let D be a smooth connected complex curve in a complex surface M such that $[D] = mB$. Let $\phi : M(D, B, m) \to M$ be the branched covering. Then the invariants of $N := M(D, B, m)$ are*

(a) $K_N = \phi^*(K_M + (m-1)B)$,

(b) $c_1^2(N) = m(K_M + (m-1)B)^2$,

(c) $e(N) = me(M) - (m-1)e(D)$,

where $e(D) = 2 - 2g(D) = -(K_M \cdot D + D^2)$ by the adjunction formula.

Proof. The formula for $e(N)$ follows by a well-known formula for the Euler characteristic of a topological space decomposed into two pieces and the formula for standard, unramified coverings. The formula for $c_1^2(N)$ then follows by the signature formula of Hirzebruch [1969]:

$$\sigma(N) = m\sigma(M) - \frac{m^2-1}{3m} D^2.$$

The formula for K_N can be found in [Barth et al. 1984, Chapter I, Lemma 17.1]. □

Suppose that the complex curve D is contained in the linear system $|nK_M|$ and hence represents in homology a multiple nK_M of the canonical class of M. Let $m > 0$ be an integer dividing n and write $n = ma$. Now set $[D] = nK_M$ and $B = aK_M$ in Proposition 47.

Corollary 48. *Let D be a smooth connected complex curve in a complex surface M with $[D] = nK_M$ and $\phi : M(D, aK_M, m) \to M$ the branched covering. Then the invariants of $N := M(D, aK_M, m)$ are*

(a) $K_N = (n+1-a)\phi^* K_M$,

(b) $c_1^2(N) = m(n+1-a)^2 c_1^2(M)$,

(c) $e(N) = me(M) + (m-1)n(n+1)c_1^2(M)$.

We consider again the general situation that M is a smooth, oriented manifold and F is an oriented submanifold of codimension 2. The fundamental group of M is related to the fundamental group of the complement $M' = M \setminus F$ by

$$\pi_1(M) \cong \pi_1(M')/N(\sigma), \tag{8}$$

where σ denotes the meridian to F, given by a circle fibre of $\partial\nu(F) \to F$, and $N(\sigma)$ denotes the normal subgroup in $\pi_1(M')$ generated by this element (a proof can be found in the appendix of [Hamilton 2008]). Using this formula, the fundamental group of a branched covering can be calculated in the following case.

Theorem 49. *Let M^n be a closed oriented manifold, and let F^{n-2} be a closed oriented submanifold. Suppose in addition that the fundamental group of M' is abelian. Then for all m and B with $[F] = mB$, there exists an isomorphism*

$$\pi_1(M(F, B, m)) \cong \pi_1(M).$$

Proof. Let $k > 0$ denote the maximal integer dividing $[F]$. Since m divides k, we can write $k = ma$ with $a > 0$. Let $\overline{M}'$ denote the complement to $\overline{F}$ in $M(F, B, m)$, and let $\bar{\sigma}$ be the meridian to $\overline{F}$. By Equation (8) we have

$$\pi_1(M(F, B, m)) \cong \pi_1(\overline{M}')/N(\bar{\sigma}).$$

There is an exact sequence $0 \to \pi_1(\overline{M}') \xrightarrow{\pi_*} \pi_1(M') \to \mathbb{Z}_m \to 0$ since $\pi : \overline{M}' \to M'$ is an m-fold cyclic covering. The assumption that $\pi_1(M')$ is abelian implies that $\pi_1(\overline{M}')$ is also abelian. Therefore the normal subgroups generated by the meridians are cyclic. The endpoints of the lifts of $0, \sigma, 2\sigma, \dots, (m-1)\sigma$, where σ is the meridian to F, realize all m points in the fibre over the basepoint. This implies that the induced map $\pi_* : \pi_1(\overline{M}') \longrightarrow \pi_1(M')/\langle\sigma\rangle$ is surjective. The kernel of this map is equal to $\langle\bar{\sigma}\rangle$, because only the multiples of $m\sigma = \pi_*\bar{\sigma}$ lift to loops in $\overline{M}'$; hence

$$\pi_1(\overline{M}')/\langle\bar{\sigma}\rangle \xrightarrow{\cong} \pi_1(M')/\langle\sigma\rangle.$$

Again by Equation (8), this implies $\pi_1(M(F, B, m)) \cong \pi_1(M)$. □

We want to apply this theorem in the case where M is a 4-manifold and F is an embedded surface. Even if M is simply connected, the complement M' does not have abelian fundamental group in general. However, in the complex case, we can use the following, which is [Nori 1983, Proposition 3.27].

Theorem 50. *Let M be a smooth complex algebraic surface, and let $D, E \subset M$ be smooth complex curves that intersect transversely. Assume that $D'^2 > 0$ for every connected component $D' \subset D$. Then the kernel of $\pi_1(M \setminus (D \cup E)) \to \pi_1(M \setminus E)$ is a finitely generated abelian group.*

If $E = \varnothing$, this implies that the kernel of $\pi_1(M') \to \pi_1(M)$ is a finitely generated abelian group if D is connected and $D^2 > 0$, where $M' = M \setminus D$. If M is simply connected, it follows that $\pi_1(M')$ is abelian. Thus with Theorem 49 we get this:

Corollary 51. *Let M be a simply connected, smooth complex algebraic surface, and let $D \subset M$ be a smooth connected complex curve with $D^2 > 0$. Let $\overline{M}$ be a cyclic ramified cover of M branched over D. Then $\overline{M}$ is also simply connected.*

Catanese [1984] has also used in a different situation restrictions on divisors to ensure that certain ramified coverings are simply connected.

11. Surfaces of general type and pluricanonical systems

We collect some results concerning the geography of simply connected surfaces of general type and the existence of smooth divisors in pluricanonical systems.

The following is the main geography result we use for our constructions.

Theorem 52 [Persson 1981, Proposition 3.23]. *Let x, y be positive integers such that $2x-6 \leq y \leq 4x-8$. Then there exists a simply connected minimal complex surface M of general type such that $\chi_h(M) = x$ and $c_1^2(M) = y$. Furthermore, M can be chosen as a genus 2 fibration.*

The smallest integer x for which an inequality can be realized with $y > 0$ is $x = 3$. Since $\chi_h(M) = p_g(M)+1$ for simply connected surfaces, this corresponds to surfaces with $p_g = 2$. Hence from Theorem 52, we get minimal simply connected complex surfaces M with $p_g = 2$ and $K^2 = 1, 2, 3, 4$. Similarly for $x = 4$ we get surfaces with $p_g = 3$ and $K^2 = 2, \ldots, 8$.

Proposition 53. *For $K^2 = 1$ and $K^2 = 2$, all possible values for p_g given by the Noether inequality $K^2 \geq 2p_g - 4$ can be realized by simply connected minimal complex surfaces of general type.*

Proof. By the Noether inequality, the only possible values for p_g are $p_g = 0, 1, 2$ if $K^2 = 1$ and $p_g = 0, 1, 2, 3$ if $K^2 = 2$. The cases $p_g = 2$ for $K^2 = 1$ and $p_g = 2, 3$ for $K^2 = 2$ are covered by Persson's theorem. In particular, the surface with $K^2 = 1$ and $p_g = 2$ and the surface with $K^2 = 2$ and $p_g = 3$ are Horikawa surfaces described in [Horikawa 1976a; 1976b]. The remaining cases can also be covered: The Barlow surface from [1985] is a simply connected numerical Godeaux surface, that is, a minimal complex surface of general type with $K^2 = 1$ and $p_g = 0$. Simply connected minimal surfaces of general type with $K^2 = 1, 2$ and $p_g = 1$ exist by constructions due to Enriques; see [Catanese 1979; Catanese and Debarre 1989; Chakiris 1980]. Finally, Lee and Park [2007] have constructed a simply connected minimal surface of general type with $K^2 = 2$ and $p_g = 0$. It is a numerical Campedelli surface. □

Suppose M is a minimal smooth complex algebraic surface of general type and consider the multiples $L = nK = K^{\otimes n}$ of the canonical line bundle of M. By a theorem of Bombieri [Bombieri 1973; Barth et al. 1984], all divisors in the linear system $|nK|$ are connected. If $|nK|$ has no fixed parts and is base point free, it determines an everywhere-defined holomorphic map to a projective space, and we can find a nonsingular divisor representing nK by taking the preimage of a generic hyperplane section.

Theorem 54. *Let M be a minimal smooth complex algebraic surface of general type. Then the pluricanonical system $|nK|$ determines an everywhere defined holomorphic map in the cases*

- *$n \geq 4$,*
- *$n = 3$ and $K^2 \geq 2$, and*
- *$n = 2$ and $K^2 \geq 5$ or $p_g \geq 1$.*

For proofs and references, see [Bombieri 1970; 1973; Catanese and Tovena 1992; Kodaira 1968; Mendes Lopes and Pardini 2002; Reider 1988].

Remark 55. In some of the remaining cases it is also known that pluricanonical systems define a holomorphic map. In particular, suppose that M is a numerical Godeaux surface. Then the map defined by $|3K|$ is holomorphic if $H_1(M;\mathbb{Z}) = 0$ or $\mathbb{Z}_2$, for example, if M is simply connected [Miyaoka 1976; Reid 1978]. This is also known for the map defined by $|2K|$ in the case of a simply connected surface M with $K^2 = 4$ and $p_g = 0$ by [Catanese and Tovena 1992; Kotschick 1994].

12. Branched covering construction of algebraic surfaces with divisible canonical class

Suppose that M is a simply connected minimal complex surface of general type. Let $m, d \geq 2$ be integers such that $m-1$ divides $d-1$ and define the integers $a = (d-1)/(m-1)$ and $n = ma$. Then $d = n + 1 - a$ and the assumptions imply that $n \geq 2$. We assume in addition that nK_M can be represented by a smooth complex connected curve D in M; see Theorem 54. Let $\overline{M} = M(D, aK_M, m)$ denote the associated m-fold branched cover over the curve D.

Theorem 56. *Let M be a simply connected minimal surface of general type, and let $m, d \geq 2$ be integers such that $d-1$ is divisible by $m-1$ with quotient a. Suppose that D is a smooth connected curve in the linear system $|nK_M|$, where $n = ma$. Then the m-fold cover of M, branched over D, is a simply connected complex surface $\overline{M}$ of general type with invariants*

$$\begin{aligned} K_{\overline{M}} &= d\phi^* K_M, & e(\overline{M}) &= m(e(M) + (d-1)(d+a)c_1^2(M)),\\ c_1^2(\overline{M}) &= md^2 c_1^2(M), & \chi_h(\overline{M}) &= m\chi_h(M) + \tfrac{1}{12}m(d-1)(2d+a+1)c_1^2(M),\\ & & \sigma(\overline{M}) &= -\tfrac{1}{3}m(2e(M) + (d(d-2) + 2a(d-1))c_1^2(M)). \end{aligned}$$

In particular, the canonical class $K_{\overline{M}}$ is divisible by d and $\overline{M}$ is minimal.

Proof. The invariants are given by Corollary 48. Since $D^2 = n^2 K_M^2 > 0$, the complex surface $\overline{M}$ is simply connected by Corollary 51. Also, $\overline{M}$ is of general type because $c_1^2(\overline{M}) > 0$ and $\overline{M}$ cannot be rational or ruled. Minimality follows from Lemma 2, since the divisibility of $K_{\overline{M}}$ is at least $d \geq 2$. □

Note that the signature $\sigma(\overline{M})$ is always negative; hence surfaces with positive signature cannot be constructed in this way.

The transformation

$$\Phi : (e(M), c_1^2(M)) \mapsto (e(\overline{M}), c_1^2(\overline{M}))$$

given by Theorem 56 is linear and can be written as

$$\begin{pmatrix} e(\overline{M}) \\ c_1^2(\overline{M}) \end{pmatrix} = m \begin{pmatrix} 1 & \Delta \\ 0 & d^2 \end{pmatrix} \begin{pmatrix} e(M) \\ c_1^2(M) \end{pmatrix},$$

with the abbreviation $\Delta = (d-1)(d+a)$. This map is invertible over $\mathbb{R}$ and maps the quadrant $\mathbb{R}^+ \times \mathbb{R}^+$ of positive coordinates in $\mathbb{R} \times \mathbb{R}$ into the same quadrant. The inverse of Φ is given by

$$\begin{pmatrix} e(M) \\ c_1^2(M) \end{pmatrix} = \frac{1}{m} \begin{pmatrix} 1 & -\Delta/d^2 \\ 0 & 1/d^2 \end{pmatrix} \begin{pmatrix} e(\overline{M}) \\ c_1^2(\overline{M}) \end{pmatrix}.$$

Definition 57. We call a point in $\mathbb{R}^+ \times \mathbb{R}^+$ *admissible* if $e(M)+c_1^2(M) \equiv 0 \bmod 12$. The coordinates $e(M)$ and $c_1^2(M)$ of a complex surface are always admissible by the Noether formula.

Lemma 58. *The image of the admissible points in $\mathbb{R}^+ \times \mathbb{R}^+$ under the map Φ consists of the points satisfying*

$$e(\overline{M}) \equiv 0 \bmod m, \quad c_1^2(\overline{M}) \equiv 0 \bmod md^2, \quad \tfrac{1}{m} e(\overline{M}) + \frac{1-\Delta}{md^2} c_1^2(\overline{M}) \equiv 0 \bmod 12.$$

The proof is immediate by the formula for the inverse of Φ. We want to calculate the image under Φ of the sector given by Theorem 52. First, we rewrite Persson's theorem in an equivalent form (we omit the proof):

Corollary 59. *Let e and c be positive integers with $c \geq 36-e$ and $e+c \equiv 0 \bmod 12$. If $\frac{1}{5}(e-36) \leq c \leq \frac{1}{2}(e-24)$, then there exists a simply connected minimal surface M of general type with invariants $e(M) = e$ and $c_1^2(M) = c$.*

In the next step, we calculate the image under Φ of the lines in the (e, c)-plane that appear in this corollary. A short calculation shows that the line $c = \frac{1}{5}(e-36)$ maps to

$$c_1^2(\overline{M}) = \frac{d^2}{5+\Delta}(e(\overline{M}) - 36m), \tag{9}$$

while the line $c = \frac{1}{2}(e-24)$ maps to

$$c_1^2(\overline{M}) = \frac{d^2}{2+\Delta}(e(\overline{M}) - 24m). \tag{10}$$

Similarly, the constraint $c \geq 36 - e$ maps to

$$c_1^2(\overline{M}) \leq \frac{d^2}{-1+\Delta}(e(\overline{M}) - 36m). \tag{11}$$

It follows that the image under Φ of the lattice points given by the constraints in Corollary 59 consists precisely of those points in the sector between the lines (9) and (10) that satisfy the constraint (11) and the constraints in Lemma 58.

The surfaces in Theorem 52 satisfy $p_g \geq 2$ and $K^2 \geq 1$. By Theorem 54, the linear system $|nK|$ for $n \geq 2$ on these surfaces defines a holomorphic map, except possibly in the case $p_g = 2$, $K^2 = 1$ and $n = 3$. Since $n = ma$ and $m \geq 2$, this occurs only for $m = 3$, $a = 1$ and $d = 3$. The corresponding image under Φ has invariants $(e, c_1^2) = (129, 27)$. This exception is implicitly understood in the following theorem. In all other cases we can consider the branched covering construction above. This can be summarized as follows: Consider integers m, a, d as above, with $m, d \geq 2$, $a \geq 1$ and $\Delta = (d-1)(d+a)$.

Theorem 60. *Let x and y be positive integers such that $y(1-\Delta) \geq 36 - x$ and $x + (1-\Delta)y \equiv 0 \bmod 12$. If*

$$\frac{1}{(5+\Delta)}(x-36) \leq y \leq \frac{1}{(2+\Delta)}(x-24),$$

then there is a simply connected minimal complex surface $\overline{M}$ of general type with invariants $e(\overline{M}) = mx$ and $c_1^2(\overline{M}) = md^2y$ such that the canonical class of $\overline{M}$ is divisible by d.

We calculate some explicit examples for the branched covering construction given by Theorem 60 and for some surfaces not covered by Persson's theorem. For any $d \geq 2$, we can choose $m = 2$ and $a = d-1$, corresponding to 2-fold covers branched over $(2d-2)K$. The formulas for the invariants simplify to

$$c_1^2(\overline{M}) = 2d^2c_1^2(M), \qquad e(\overline{M}) = 24\chi_h(M) + 2d(2d-3)c_1^2(M),$$
$$\chi_h(\overline{M}) = 2\chi_h(M) + \tfrac{1}{2}d(d-1)c_1^2(M).$$

The first two examples are double coverings with $m = 2$, whereas the third example uses coverings of higher degree. Because of their topological invariants, some of the surfaces are homeomorphic by Freedman's theorem to the simply connected symplectic 4-manifolds constructed in Sections 6 and 7.

Example 61. We consider the Horikawa surfaces [1976a] on the Noether line $c_1^2 = 2\chi_h - 6$, which exist for every $\chi_h \geq 4$ and are also given by Persson's Theorem 52. In this case $p_g \geq 3$ and $c_1^2 \geq 2$; hence by Theorem 54, the linear system $|nK|$ for $n \geq 2$ defines a holomorphic map on these surfaces.

Proposition 62. *Let M be a Horikawa surface on the Noether line $c_1^2 = 2\chi_h - 6$, where $\chi_h = 4 + l$ for $l \geq 0$. Then the 2-fold cover $\overline{M}$ of the surface M, branched* *over* $(2d-2)K_M$ *for an integer* $d \geq 2$, *has invariants*

$$c_1^2(\overline{M}) = 4d^2(l+1), \qquad \chi_h(\overline{M}) = 6 + (2+d(d-1))(l+1),$$
$$e(\overline{M}) = 72 + 4(l+1)(6+2d^2-3d), \quad \sigma(\overline{M}) = -48 - 4(l+1)(4+d^2-2d).$$

The canonical class $K_{\overline{M}}$ *is divisible by* d.

For d even, the integer $d^2-2d = d(d-2)$ is divisible by 4; hence σ is indeed divisible by 16, which is necessary by Rohlin's theorem. The invariants are on the line

$$c_1^2(\overline{M}) = \frac{4d^2}{2+d(d-1)}(\chi_h(\overline{M}) - 6),$$

which has inclination close to 4 for d very large.

Example 63. We calculate the invariants for the branched covers with $m = 2$ and integers $d \geq 3$ for the surfaces given by Proposition 53. Since $n = ma \geq 4$ in this case, Theorem 54 shows that the linear system $|nK|$ defines a holomorphic map and we can use the branched covering construction.

Proposition 64. *Let* M *be a minimal complex surface of general type with* K^2 *equal to* 1 *or* 2. *The* 2*-fold cover* $\overline{M}$ *of the surface* M, *branched over* $(2d-2)K_M$ *for an integer* $d \geq 3$, *has invariants as follows*:

If $K^2 = 1$ *and* $p_g = 0, 1, 2,$
$$c_1^2(\overline{M}) = 2d^2,$$
$$e(\overline{M}) = 24(p_g+1) + 2d(2d-3),$$
$$\sigma(\overline{M}) = -16(p_g+1) - 2d(d-2).$$

If $K^2 = 2$ *and* $p_g = 0, 1, 2, 3,$
$$c_1^2(\overline{M}) = 4d^2,$$
$$e(\overline{M}) = 24(p_g+1) + 4d(2d-3),$$
$$\sigma(\overline{M}) = -16(p_g+1) - 4d(d-2).$$

In both cases the canonical class $K_{\overline{M}}$ *is divisible by* d.

Example 65. Consider the Barlow surface M_{B} and the surface M_{LP} of Lee and Park mentioned in the proof of Proposition 53. The invariants are

$$c_1^2(M_{\mathrm{B}}) = 1, \quad \chi_h(M_{\mathrm{B}}) = 1, \quad e(M_{\mathrm{B}}) = 11;$$
$$c_1^2(M_{\mathrm{LP}}) = 2, \quad \chi_h(M_{\mathrm{LP}}) = 1, \quad e(M_{\mathrm{LP}}) = 10.$$

By Theorem 54, we can consider branched covers over both surfaces with $ma \geq 3$ (the Barlow surface is a simply connected numerical Godeaux surface, and hence $|3K|$ defines a holomorphic map by Remark 55). See Tables 1 and 2 for a calculation of the invariants of $\overline{M}$ for small values of d and m. There is an agreement between the 4-fold cover of M_{B} branched over $4K_M$ and the 2-fold cover of M_{LP} branched over $6K_M$: Both have the same Chern invariants and the same divisibility

d	m	ma	$(d-1)(d+a)$	$e(\overline{M_{\mathrm{B}}})$	$c_1^2(\overline{M_{\mathrm{B}}})$	$\chi_h(\overline{M_{\mathrm{B}}})$	$b_2^+(\overline{M_{\mathrm{B}}})$	$\sigma(\overline{M_{\mathrm{B}}})$
3	2	4	10	42	18	5	9	−22
3	3	3	8	57	27	7	13	−29
4	2	6	21	64	32	8	15	−32
4	4	4	15	104	64	14	27	−48
5	2	8	36	94	50	12	23	−46
5	3	6	28	117	75	16	31	−53
5	5	5	24	175	125	25	49	−75
6	2	10	55	132	72	17	33	−64
6	6	6	35	276	216	41	81	−112

Table 1. Ramified coverings of the Barlow surface M_{B} of degree m branched over maK.

d	m	ma	$(d-1)(d+a)$	$e(\overline{M_{\mathrm{LP}}})$	$c_1^2(\overline{M_{\mathrm{LP}}})$	$\chi_h(\overline{M_{\mathrm{LP}}})$	$b_2^+(\overline{M_{\mathrm{LP}}})$	$\sigma(\overline{M_{\mathrm{LP}}})$
3	2	4	10	60	36	8	15	−28
3	3	3	8	78	54	11	21	−34
4	2	6	21	104	64	14	27	−48
4	4	4	15	160	128	24	47	−64
5	2	8	36	164	100	22	43	−76
5	3	6	28	198	150	29	57	−82
5	5	5	24	290	250	45	89	−110
6	2	10	55	240	144	32	63	−112
6	6	6	35	480	432	76	151	−176

Table 2. Ramified coverings of the Lee–Park surface M_{LP} of degree m branched over maK.

$d = 4$ of the canonical class. Hence the manifolds are homeomorphic and by Theorem 4, both branched coverings have the same Seiberg–Witten invariants.

Remark 66. More general examples are possible by considering branched coverings over singular complex curves. The following example is described for instance in [Gompf and Stipsicz 1999, Chapter 7]: Let $B_{n,m}$ denote the singular complex curve in $\mathbb{C}P^1 \times \mathbb{C}P^1$ that is the union of $2n$ parallel copies of the first factor and $2m$ parallel copies of the second factor. The curve $B_{n,m}$ represents in cohomology the class $2nS_1 + 2mS_2$, where $S_1 = [\mathbb{C}P^1 \times \{*\}]$ and $S_2 = [\{*\} \times \mathbb{C}P^1]$. Let $X'(n,m)$ denote the double covering of $\mathbb{C}P^1 \times \mathbb{C}P^1$ branched over $B_{n,m}$. It is a singular complex surface that has a canonical resolution $X(n,m)$; see [Barth et al. 1984, Chapter III]. As a smooth 4-manifold, $X(n,m)$ is diffeomorphic to the double cover of $\mathbb{C}P^1 \times \mathbb{C}P^1$ branched over the smooth curve $\widetilde{B}_{n,m}$ given by smoothing the double points. Hence the topological invariants for $X = X(n, m)$ can be calculated with the formulas from Proposition 47 to be

$$c_1^2(X) = 4(n-2)(m-2), \quad e(X) = 6 + 2(2m-1)(2n-1), \quad \sigma(X) = -4mn.$$

Writing $X' = X'(n, m)$ and $M = \mathbb{C}P^1 \times \mathbb{C}P^1$, denote by $\phi : X' \to M$ the double covering, by $\pi : X \to X'$ the canonical resolution, and by $\psi = \phi \circ \pi$ the composition. Since all singularities of $B_{n,m}$ are ordinary double points, K_X can be calculated by a formula in [Barth et al. 1984, Theorem 7.2, Chapter III] as

$$\begin{aligned} K_X = \psi^*(K_M + \tfrac{1}{2} B_{m,n}) &= \psi^*(-2S_1 - 2S_2 + nS_1 + mS_2) \\ &= \psi^*((n-2)S_1 + (m-2)S_2). \end{aligned}$$

We interpret this formula as follows: The map $\psi : X \to \mathbb{C}P^1 \times \mathbb{C}P^1$ followed by the projection onto the first factor defines a fibration $X \to \mathbb{C}P^1$ whose fibres are the branched covers of the rational curves $\{p\} \times \mathbb{C}P^1$, where $p \in \mathbb{C}P^1$. The generic rational curve among them is disjoint from the $2m$ curves in $B_{n,m}$ parallel to $\{*\} \times \mathbb{C}P^1$ and intersects the $2n$ curves parallel to $\mathbb{C}P^1 \times \{*\}$ at $2n$ points. This implies that the generic fibre F_2 of the fibration is a double branched cover of $\mathbb{C}P^1$ at $2n$ distinct points and hence a smooth complex curve of genus $n-1$. This curve represents the class $\psi^* S_2$ in the surface X. Similarly, there is a fibration $X \to \mathbb{C}P^1$ in genus $m-1$ curves that represents $F_1 = \psi^* S_1$. Hence we can write $K_X = (n-2)F_1 + (m-2)F_2$. In particular, the divisibility of K_X is the greatest common divisor of $n-2$ and $m-2$.

Remark 67. In [Catanese 1984; 1986; Catanese and Wajnryb 2007], the authors constructed certain families of simply connected surfaces of general type with divisible canonical class, using branched coverings over singular curves. Some of these surfaces are diffeomorphic but not deformation equivalent, thus giving counterexamples to a well-known conjecture.

Acknowledgment

This article is part of my thesis, submitted in May 2008 at the University of Munich. I would like to thank my advisor D. Kotschick.

References

[Auroux et al. 2005] D. Auroux, V. Muñoz, and F. Presas, "Lagrangian submanifolds and Lefschetz pencils", *J. Symplectic Geom.* **3**:2 (2005), 171–219. MR 2006i:53118 Zbl 1093.53085

[Barlow 1985] R. Barlow, "A simply connected surface of general type with $p_g = 0$", *Invent. Math.* **79**:2 (1985), 293–301. MR 87a:14033 Zbl 0561.14015

[Barth et al. 1984] W. Barth, C. Peters, and A. Van de Ven, *Compact complex surfaces*, Ergebnisse series (3) **4**, Springer, Berlin, 1984. MR 86c:32026 Zbl 0718.14023

[Bombieri 1970] E. Bombieri, "The pluricanonical map of a complex surface", pp. 35–87 in *Several Complex Variables, I* (College Park, MD, 1970), edited by J. Horváth, Springer, Berlin, 1970. MR 43 #1975 Zbl 0213.47601

[Bombieri 1973] E. Bombieri, "Canonical models of surfaces of general type", *Inst. Hautes Études Sci. Publ. Math.* 42 (1973), 171–219. MR 47 #6710 Zbl 0259.14005

[Catanese 1979] F. Catanese, "Surfaces with $K^2 = p_g = 1$ and their period mapping", pp. 1–29 in *Algebraic geometry* (Copenhagen, 1978), edited by K. Lønsted, Lecture Notes in Math. **732**, Springer, Berlin, 1979. MR 81c:14018 Zbl 0423.14019

[Catanese 1984] F. Catanese, "On the moduli spaces of surfaces of general type", *J. Differential Geom.* **19**:2 (1984), 483–515. MR 86h:14031 Zbl 0549.14012

[Catanese 1986] F. Catanese, "Connected components of moduli spaces", *J. Differential Geom.* **24**:3 (1986), 395–399. MR 87m:14036 Zbl 0621.14014

[Catanese and Debarre 1989] F. Catanese and O. Debarre, "Surfaces with $K^2 = 2$, $p_g = 1$, $q = 0$", *J. Reine Angew. Math.* **395** (1989), 1–55. MR 89m:14017 Zbl 0658.14016

[Catanese and Tovena 1992] F. Catanese and F. Tovena, "Vector bundles, linear systems and extensions of π_1", pp. 51–70 in *Complex algebraic varieties* (Bayreuth, 1990), edited by K. Hulek et al., Lecture Notes in Math. **1507**, Springer, Berlin, 1992. MR 93h:14028 Zbl 0788.14013

[Catanese and Wajnryb 2007] F. Catanese and B. Wajnryb, "Diffeomorphism of simply connected algebraic surfaces", *J. Differential Geom.* **76**:2 (2007), 177–213. MR 2008c:14057 Zbl 1127.14039

[Chakiris 1980] K. N. Chakiris, "Counterexamples to global Torelli for certain simply connected surfaces", *Bull. Amer. Math. Soc.* (*N.S.*) **2**:2 (1980), 297–299. MR 81d:14021 Zbl 0432.32014

[Chen 1987] Z. J. Chen, "On the geography of surfaces: Simply connected minimal surfaces with positive index", *Math. Ann.* **277**:1 (1987), 141–164. MR 88c:14057 Zbl 0595.14027

[Fintushel and Stern 1994] R. Fintushel and R. J. Stern, "Surgery in cusp neighborhoods and the geography of irreducible 4-manifolds", *Invent. Math.* **117**:3 (1994), 455–523. MR 95f:57040 Zbl 0843.57021

[Fintushel and Stern 1997] R. Fintushel and R. J. Stern, "Rational blowdowns of smooth 4-manifolds", *J. Differential Geom.* **46**:2 (1997), 181–235. MR 98j:57047 Zbl 0896.57022

[Fintushel and Stern 1998] R. Fintushel and R. J. Stern, "Knots, links, and 4-manifolds", *Invent. Math.* **134**:2 (1998), 363–400. MR 99j:57033 Zbl 0914.57015

[Fintushel and Stern 2001] R. Fintushel and R. Stern, "The canonical class of a symplectic 4-manifold", *Turkish J. Math.* **25**:1 (2001), 137–145. MR 2002g:57053 Zbl 0977.57032

[Fintushel and Stern 2004] R. Fintushel and R. J. Stern, "Families of simply connected 4-manifolds with the same Seiberg–Witten invariants", *Topology* **43**:6 (2004), 1449–1467. MR 2005d:57044 Zbl 1064.57036

[Freedman 1982] M. H. Freedman, "The topology of four-dimensional manifolds", *J. Differential Geom.* **17**:3 (1982), 357–453. MR 84b:57006 Zbl 0528.57011

[Friedman and Morgan 1997] R. Friedman and J. W. Morgan, "Algebraic surfaces and Seiberg–Witten invariants", *J. Algebraic Geom.* **6**:3 (1997), 445–479. MR 99b:32045 Zbl 0896.14015

[Gompf 1991] R. E. Gompf, "Nuclei of elliptic surfaces", *Topology* **30**:3 (1991), 479–511. MR 92f:57042 Zbl 0732.57010

[Gompf 1995] R. E. Gompf, "A new construction of symplectic manifolds", *Ann. of Math.* (2) **142**:3 (1995), 527–595. MR 96j:57025 Zbl 0849.53027

[Gompf and Mrowka 1993] R. E. Gompf and T. S. Mrowka, "Irreducible 4-manifolds need not be complex", *Ann. of Math.* (2) **138**:1 (1993), 61–111. MR 95c:57053 Zbl 0805.57012

[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

[Hamilton 2008] M. J. D. Hamilton, *On symplectic 4-manifolds and contact 5-manifolds*, thesis, Ludwig-Maximilians-Universität München, 2008, available at http://nbn-resolving.de/urn:nbn:de:bvb:19-87797.

[Hamilton and Kotschick 2006] M. J. D. Hamilton and D. Kotschick, "Minimality and irreducibility of symplectic four-manifolds", *Int. Math. Res. Not.* (2006), Art. ID 35032. MR 2007i:57023 Zbl 1101.53052

[Hirzebruch 1969] F. Hirzebruch, "The signature of ramified coverings", pp. 253–265 in *Global Analysis (Papers in Honor of K. Kodaira)*, Univ. Tokyo Press, 1969. MR 41 #2707 Zbl 0208.51802

[Horikawa 1976a] E. Horikawa, "Algebraic surfaces of general type with small C_1^2, I", *Ann. of Math.* (2) **104**:2 (1976), 357–387. MR 54 #12789 Zbl 0339.14024

[Horikawa 1976b] E. Horikawa, "Algebraic surfaces of general type with small c_1^2, II", *Invent. Math.* **37**:2 (1976), 121–155. MR 57 #334 Zbl 0339.14025

[Ionel and Parker 2004] E.-N. Ionel and T. H. Parker, "The symplectic sum formula for Gromov–Witten invariants", *Ann. of Math.* (2) **159**:3 (2004), 935–1025. MR 2006b:53110 Zbl 1075.53092

[Kodaira 1968] K. Kodaira, "Pluricanonical systems on algebraic surfaces of general type", *J. Math. Soc. Japan* **20** (1968), 170–192. MR 37 #212 Zbl 0157.27704

[Kotschick 1994] D. Kotschick, "On the pluricanonical maps of Godeaux and Campedelli surfaces", *Internat. J. Math.* **5**:1 (1994), 53–60. MR 94m:14053 Zbl 0810.14016

[Kotschick 1997] D. Kotschick, "The Seiberg–Witten invariants of symplectic four-manifolds", pp. 195–220 in *Séminaire Bourbaki* 1995/96 (Exposé 812), Astérisque **241**, Société Mathématique de France, Paris, 1997. MR 98h:57057 Zbl 0882.57026

[Lee and Park 2007] Y. Lee and J. Park, "A simply connected surface of general type with $p_g = 0$ and $K^2 = 2$", *Invent. Math.* **170**:3 (2007), 483–505. MR 2008m:14076 Zbl 1126.14049

[Li and Liu 2001] T.-J. Li and A.-K. Liu, "Uniqueness of symplectic canonical class, surface cone and symplectic cone of 4-manifolds with $B^+ = 1$", *J. Differential Geom.* **58**:2 (2001), 331–370. MR 2003g:57042 Zbl 1051.57035

[Liu 1996] A.-K. Liu, "Some new applications of general wall crossing formula, Gompf's conjecture and its applications", *Math. Res. Lett.* **3**:5 (1996), 569–585. MR 97k:57038 Zbl 0872.57025

[McCarthy and Wolfson 1994] J. D. McCarthy and J. G. Wolfson, "Symplectic normal connect sum", *Topology* **33**:4 (1994), 729–764. MR 95h:57038 Zbl 0812.53033

[McMullen and Taubes 1999] C. T. McMullen and C. H. Taubes, "4-manifolds with inequivalent symplectic forms and 3-manifolds with inequivalent fibrations", *Math. Res. Lett.* **6**:5-6 (1999), 681–696. MR 2000m:57045 Zbl 0964.53051

[Mendes Lopes and Pardini 2002] M. Mendes Lopes and R. Pardini, "A survey on the bicanonical map of surfaces with $p_g = 0$ and $K^2 \geq 2$", pp. 277–287 in *Algebraic geometry*, edited by M. C. Beltrametti et al., de Gruyter, Berlin, 2002. MR 2003k:14051 Zbl 1046.14018

[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

[Miyaoka 1976] Y. Miyaoka, "Tricanonical maps of numerical Godeaux surfaces", *Invent. Math.* **34**:2 (1976), 99–111. MR 53 #13236 Zbl 0337.14010

[Morgan 1996] J. W. Morgan, *The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*, Mathematical Notes **44**, Princeton University Press, 1996. MR 97d:57042 Zbl 0846.57001

[Nori 1983] M. V. Nori, "Zariski's conjecture and related problems", *Ann. Sci. École Norm. Sup.* (4) **16**:2 (1983), 305–344. MR 86d:14027 Zbl 0527.14016

[Park 1998] J. Park, "The geography of irreducible 4-manifolds", *Proc. Amer. Math. Soc.* **126**:8 (1998), 2493–2503. MR 98j:57034 Zbl 0895.57005

[Park 2002] J. Park, "The geography of Spin symplectic 4-manifolds", *Math. Z.* **240**:2 (2002), 405–421. MR 2003c:57030 Zbl 1030.57032

[Park and Szabó 2000] B. D. Park and Z. Szabó, "The geography problem for irreducible spin four-manifolds", *Trans. Amer. Math. Soc.* **352**:8 (2000), 3639–3650. MR 2000m:57037 Zbl 0947.57023

[Persson 1981] U. Persson, "Chern invariants of surfaces of general type", *Compositio Math.* **43**:1 (1981), 3–58. MR 83b:14012 Zbl 0479.14018

[Persson et al. 1996] U. Persson, C. Peters, and G. Xiao, "Geography of spin surfaces", *Topology* **35**:4 (1996), 845–862. MR 98h:14046 Zbl 0874.14031

[Reid 1978] M. Reid, "Surfaces with $p_g = 0$, $K^2 = 1$", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **25**:1 (1978), 75–92. MR 80h:14018 Zbl 0399.14025

[Reider 1988] I. Reider, "Vector bundles of rank 2 and linear systems on algebraic surfaces", *Ann. of Math.* (2) **127**:2 (1988), 309–316. MR 89e:14038 Zbl 0663.14010

[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

[Smith 2000] I. Smith, "On moduli spaces of symplectic forms", *Math. Res. Lett.* **7**:5-6 (2000), 779–788. MR 2002a:53112 Zbl 1010.53063

[Taubes 1995a] C. H. Taubes, "More constraints on symplectic forms from Seiberg–Witten invariants", *Math. Res. Lett.* **2**:1 (1995), 9–13. MR 96a:57075 Zbl 0854.57019

[Taubes 1995b] C. H. Taubes, "The Seiberg–Witten and Gromov invariants", *Math. Res. Lett.* **2**:2 (1995), 221–238. MR 96a:57076 Zbl 0854.57020

[Thurston 1976] W. P. Thurston, "Some simple examples of symplectic manifolds", *Proc. Amer. Math. Soc.* **55**:2 (1976), 467–468. MR 53 #6578 Zbl 0324.53031

[Vidussi 2001] S. Vidussi, "Homotopy $K3$'s with several symplectic structures", *Geom. Topol.* **5** (2001), 267–285. MR 2002a:57045 Zbl 1067.57031

[Vidussi 2007] S. Vidussi, "On the number of components of the symplectic representatives of the canonical class", *J. Eur. Math. Soc.* **9**:4 (2007), 789–800. MR 2008g:57030 Zbl 1144.57023

[Witten 1994] E. Witten, "Monopoles and four-manifolds", *Math. Res. Lett.* **1**:6 (1994), 769–796. MR 96d:57035 Zbl 0867.57029

Received January 29, 2009.

MARK J. D. HAMILTON
MATHEMATISCHES INSTITUT
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
THERESIENSTRASSE 39
80333 MUNICH
GERMANY

mark.hamilton@math.lmu.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

BOUNDARY ASYMPTOTICAL BEHAVIOR OF LARGE SOLUTIONS TO HESSIAN EQUATIONS

YONG HUANG

We consider the exact asymptotic behavior of smooth solutions to boundary blow-up problems for the k-Hessian equation on Ω, where $\partial\Omega$ is strictly $(k-1)$-convex. Similar results were obtained by Cîrstea and Trombetti when $k=n$ (the Monge–Ampère equation) and by Bandle and Marcus for a semilinear equation.

1. Introduction and main results

We investigate the qualitative properties of solutions to the boundary blow-up problem for the k-Hessian equation of the form

$$\begin{cases} H_k[D^2u] = \sigma_k(\lambda_1, \dots, \lambda_n) = b(x) f(u), & x \in \Omega, \\ u(x) = \infty, & x \in \partial\Omega, \end{cases} \tag{1-1}$$

where $b(x)$ is a continuous weight function, $\lambda_1, \dots, \lambda_n$ are eigenvalues of D^2u, the Hessian matrix of a C^2-function u defined over Ω, and Ω is a bounded domain in $\mathbb{R}^n$. The boundary condition means $u(x) \to +\infty$ as $d(x) \triangleq \operatorname{dist}(x, \partial\Omega) \to 0_+$.

Following [Caffarelli et al. 1985; Trudinger 1995], σ_k is defined by

$$\sigma_k(\lambda_1, \dots, \lambda_n) = \sum_{1 \le i_1 < \cdots < i_k \le n} \lambda_{i_1} \cdots \lambda_{i_k}. \tag{1-2}$$

One can solve (1-1) in a class of k-convex functions by [Caffarelli et al. 1985; Jian 2006]. Recall that a function $u \in C^2(\Omega)$ is called k-convex (or strictly k-convex) if $(\lambda_1, \dots, \lambda_n) \in \overline{\Gamma}_k$ (or $(\lambda_1, \dots, \lambda_n) \in \Gamma_k$) for every $x \in \Omega$, where Γ_k is the convex cone with vertex at the origin given by

$$\Gamma_k = \{\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \mid \sigma_j(\lambda) > 0,\ j = 1, \dots, k\}.$$

Obviously,

$$\Gamma_1 \supset \Gamma_2 \supset \cdots \supset \Gamma_n = \{\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \mid \lambda_j > 0,\ j = 1, \dots, k\},$$

MSC2000: primary 35J60; secondary 35B40.
Keywords: boundary blow-up problem, Hessian equations.
The first author is supported by Tianyuan Fund of Mathematics (10826060).

where Γ_n is the positive cone, and $\sigma_k(\lambda_1, \dots, \lambda_n)$ is elliptic in the class of k-convex functions.

For an open bounded subset Ω of $\mathbb{R}^n$ with boundary of class C^2 and for every $x \in \partial\Omega$, we denote by $\rho_1(x), \dots, \rho_{n-1}(x)$ the principal curvatures of $\partial\Omega$ (relative to the interior normal). Recall that Ω is said to be l-convex if $(\rho_1(x), \dots, \rho_{n-1}(x)) \in \overline{\Gamma}_l$, and it is called strictly l-convex if $(\rho_1(x), \dots, \rho_{n-1}(x)) \in \Gamma_l$, for every $x \in \partial\Omega$. In particular, strictly $(n-1)$-convex is just strictly convex.

Using radial function methods and techniques of ordinary differential inequality, Jian [2006] constructed various barriers functions, then proved existence and nonexistence theorems using those barriers. Furthermore, generic boundary blow-up rates for the solution are derived for the k-Hessian equation with boundary blow-up problem. In this paper, we derive accurately the blow-up rate of solutions to boundary blow-up problems for Hessian equations.

Let $\mathfrak{K}_\ell$ denote the set of all positive nondecreasing C^1-functions m defined on $(0, \nu)$, for some $\nu > 0$, for which there exists

$$\text{(1-3)} \qquad \lim_{t\to 0^+} \frac{\int_0^t m(s)\,ds}{m(t)} = 0 \quad \text{and} \quad \lim_{t\to 0^+} \frac{d}{dt}\left(\frac{\int_0^t m(s)\,ds}{m(t)}\right) = \ell.$$

A complete characterization of $\mathfrak{K}_\ell$ (according to $\ell \neq 0$ or $\ell = 0$) is provided by [Cîrstea and Rădulescu 2006].

One has the following examples for special ℓ, where $p > 0$ is arbitrary:

(a) $m(t) = (-1/\ln t)^p$ with $\ell = 1$,

(b) $m(t) = t^p$ with $\ell = 1/(p+1)$,

(c) $m(t) = e^{-1/t^p}$ with $\ell = 0$.

Definition 1.1. A positive measurable function f defined on $[a, \infty)$, for some $a > 0$, is called regularly varying at infinity with index q, written $f \in \mathbb{RV}_q$, if for each $\lambda > 0$ and some $q \in \mathbb{R}$,

$$\text{(1-4)} \qquad \lim_{t\to\infty} \frac{f(\lambda t)}{f(t)} = \lambda^q.$$

The real number q is called the index of regular variation.

When $q = 0$, we have:

Definition 1.2. A positive measurable function L defined on $[a, \infty)$, for some $a > 0$, is called regularly varying at infinity, if for each $\lambda > 0$ and some $q \in \mathbb{R}$,

$$\text{(1-5)} \qquad \lim_{t\to\infty} \frac{L(\lambda t)}{L(t)} = 1.$$

By Definitions 1.1 and 1.2, if $f \in \mathbb{RV}_q$, it can be represented in the form

$$\text{(1-6)} \qquad f(t) = u^q L(t).$$

Notation. If H is a nondecreasing function on $\mathbb{R}$, then we denote by $H^{\leftarrow}$ the (left-continuous) inverse of H [Resnick 1987], that is,

$$H^{\leftarrow}(y) = \inf\{s : H(s) \geq y\}.$$

If $\alpha > 0$ is sufficiently large, we define

$$\mathcal{P}(u) = \sup\left\{\frac{f(y)}{y^k} : \alpha \leq y \leq u\right\}, \quad \text{for } u \geq \alpha. \tag{1-7}$$

Problem (1-1) is the Laplace operator when $k = 1$. There are many papers resolving existence, uniqueness and asymptotic behavior issues for blow-up solutions of semilinear/quasilinear elliptic equations: for instance [Osserman 1957; Resnick 1987; Véron 1992; Bandle and Marcus 1992; 1995; García-Melián et al. 2001; Chuaqui et al. 2004; Cîrstea and Rădulescu 2006; García-Melián 2006].

When $k = n$, problem (1-1) is the Monge–Ampère equation, for which Cîrstea and Trombetti [2008] obtained existence, uniqueness and asymptotic behavior; see also [Guan and Jian 2004; Mohammed 2007].

The boundary blow-up problem of the k-Hessian equation was considered in [Salani 1998; Colesanti et al. 2000; Jian 2006]. See also [Takimoto 2006] for recent results on boundary blow-up problems for k-curvature equations, where there is a considerable difference between the cases $1 \leq k \leq n-1$ and $k = n$. However, we can unify them by using techniques from [Colesanti et al. 2000; Cîrstea and Trombetti 2008] for k-Hessian equations.

Our asymptotic results are obtained in the case when $\partial\Omega$ is strictly $(k-1)$-convex, but for k-curvature equations in [Cîrstea and Trombetti 2008], the condition that $\partial\Omega$ is strictly convex is needed.

Theorem 1.3. *Let $n \geq 2$ and Ω be a smooth, strictly $(k-1)$-convex bounded domain in $\mathbb{R}^n$. Assume that $f \in \mathbb{RV}_q$ with $q > k$ and there exists $m \in \mathfrak{K}_\ell$ such that*

$$0 < \beta^- = \liminf_{d(x)\to 0} \frac{b(x)}{m^{k+1}(d(x))} \quad \textit{and} \quad \limsup_{d(x)\to 0} \frac{b(x)}{m^{k+1}(d(x))} = \beta^+ < \infty. \tag{1-8}$$

Then, every k-convex blow-up solution u_∞ of (1-1) *satisfies*

$$\xi^- \leq \liminf_{d(x)\to 0} \frac{u}{\phi(d(x))} \quad \textit{and} \quad \limsup_{d(x)\to 0} \frac{u}{\phi(d(x))} \leq \xi^+, \tag{1-9}$$

where ϕ is defined by

$$\phi(t) = \mathcal{P}^{\leftarrow}\left(\left(\int_0^t m(s)\,ds\right)^{-k-1}\right), \quad \textit{for } t > 0 \textit{ small}, \tag{1-10}$$

and $\xi^{\pm}$ are positive constants given by

$$\frac{(\xi^+)^{k-q}}{\beta^-}\max_{\partial\Omega}\sigma_{k-1}=\frac{(\xi^-)^{k-q}}{\beta^+}\min_{\partial\Omega}\sigma_{k-1}=\frac{\big((q-k)/(n+1)\big)^{k+1}}{1+\ell(q-k)/(k+1)}. \tag{1-11}$$

On the other hand, Colesanti et al. [2000] established asymptotic estimates for the behavior of the smallest viscosity solution near the boundary of Ω for the Hessian equation

$$\begin{cases} H_k[D^2u]=f(u), & x\in\Omega,\\ u(x)=\infty, & x\in\partial\Omega.\end{cases} \tag{1-12}$$

Theorem 1.3 may also been seen as a generalization of the asymptotic behavior for the viscosity solution in [Colesanti et al. 2000].

Remark 1.4. In the setting of Theorem 1.3, $\lim_{d(x)\to 0} u/\phi(d(x))$ exists provided that Ω is a ball and (1-8) holds with $\beta^-=\beta^+\in(0,\infty)$. The latter condition is equivalent to saying that

$$b(x)\sim (m(d(x)))^{k+1}\quad \text{as } d(x)\to 0,\quad \text{for some } m\in\mathfrak{K}_\ell. \tag{1-13}$$

More exactly, when Ω is a ball of radius $R>0$, Theorem 1.3 reads as follows.

Corollary 1.5. *Let $\Omega=B_R$ be a ball of radius $R>0$ and $f\in\mathbb{RV}_q$ with $q>k$. If* (1-13) *holds, then every strictly k-convex blow up solution u of* (1-1) *satisfies*

$$u(x)\sim\xi\phi(d(x))\quad \textit{as } d(x)\to 0, \tag{1-14}$$

where ϕ is defined by (1-10) *and ξ is given by*

$$\xi=\left(\frac{\big((q-k)/(k+1)\big)^{k+1}R^{k-1}}{1+\ell(q-k)/(k+1)}\right)^{1/(k-q)}. \tag{1-15}$$

Under slightly more restrictive conditions than those in Theorem 1.3, there is at most one strictly k-convex blow-up solution of (1-1).

Theorem 1.6. *Let Ω be a smooth, strictly $(k-1)$-convex, bounded domain in $\mathbb{R}^n$. Suppose $f\in\mathbb{RV}_q$ with $q>k$, and $f(u)/u^k$ is increasing on $(0,\infty)$. Then,* (1-1) *has at most one strictly k-convex blow-up solution, provided that either*

(i) *b is positive on $\overline{\Omega}$, or*

(ii) *b is zero on $\partial\Omega$, Ω is a ball of radius $R>0$ and* (1-13) *holds.*

Remark 1.7. When $k=n$ (the Monge–Ampère equation), Theorems 1.3 and 1.6 were obtained in [Cîrstea and Trombetti 2008].

2. Preliminaries

Proposition 2.1. *Let Ω be an open subset of $\mathbb{R}^n$ with $n \geq 2$. If $h \in C^2(\mathbb{R})$ and $g \in C^2(\Omega)$ then*

$$\sigma_k(D^2 h(g(x))) = \big(h'(g(x))\big)^{k-1} h''(x)\sigma_{k-1}(D^2 g|_{i,j}) g_i g_j + \big(h'(g(x))\big)^k \sigma_k(D^2 g), \quad \text{for all } x \in \Omega, \tag{2-1}$$

where $D^2 g|_{i,j}$ is the cofactor of the (i,j)-th entry of the symmetric matrix $D^2 g(x)$.

For $\mu > 0$, we set $\Gamma_\mu = \{x \in \overline{\Omega} : d(x, \partial\Omega) < \mu\}$.

Remark 2.2. If Ω is bounded and $\partial\Omega \in C^l$ for $l \geq 2$, then there exists a positive constant μ depending on Ω such that $d \in C^l(\Gamma_\mu)$. (See also Lemma 14.16 in [Gilbarg and Trudinger 1998].)

Corollary 2.3. *Let Ω be bounded with $\partial\Omega \in C^l$ for $l \geq 2$. Assume that $\mu > 0$ is small such that $d \in C^2(\Gamma_\mu)$ and h is a C^2-function on $(0, \mu)$. Let $x_0 \in \Gamma_\mu \setminus \partial\Omega$ and $y_0 \in \partial\Omega$ be such that $|x_0 - y_0| = d(x_0)$. Then, we have*

$$\sigma_k\big(D^2 h(d(x_0))\big) = \big(-h'(d(x_0))\big)^{k-1} h''(d(x_0))\sigma_{k-1}(\varepsilon_1, \dots, \varepsilon_{n-1}) + \big(-h'(d(x_0))\big)^k \sigma_k(\varepsilon_1, \dots, \varepsilon_{n-1}), \tag{2-2}$$

where $\rho_1(y_0), \dots, \rho_{n-1}(y_0)$ are the principal curvatures of $\partial\Omega$ at y_0 and $\varepsilon_i = \rho_i(y_0)/(1 - \rho_i(y_0)d(x_0))$, $i = 1, \dots, n-1$.

Proof. It is easy to calculate that the expression of the Hessian matrix of d at x_0 in terms of a principal coordinate system at y_0 (see also Lemma 14.17 in [Gilbarg and Trudinger 1998]), namely

$$\begin{aligned} D^2 d(x_0) &= \operatorname{diag}\left(\frac{-\rho_1(y_0)}{1 - \rho_1(y_0)d(x_0)}, \dots, \frac{-\rho_{n-1}(y_0)}{1 - \rho_{n-1}(y_0)d(x_0)}, 0\right), \\ Dd(x_0) &= (0, \dots, 0, 1). \end{aligned}$$

Thus by Proposition 2.1, we obtain

$$\begin{aligned} &\sigma_k\left(D^2 h(d(x_0))\right) \\ &= \big(-h'(d(x_0))\big)^{k-1} h''(d(x_0))\sigma_{k-1}\left(\begin{bmatrix} \frac{\rho_1(y_0)}{1-\rho_1(y_0)d(x_0)} & & \\ & \ddots & \\ & & \frac{\rho_{n-1}(y_0)}{1-\rho_{n-1}(y_0)d(x_0)} \end{bmatrix}\right) \\ &\quad + \big(-h'(d(x_0))\big)^k \sigma_k\left(\begin{bmatrix} \frac{\rho_1(y_0)}{1-\rho_1(y_0)d(x_0)} & & \\ & \ddots & \\ & & \frac{\rho_{n-1}(y_0)}{1-\rho_{n-1}(y_0)d(x_0)} \end{bmatrix}\right). \qquad \square \end{aligned}$$

We now give a brief account of the definitions and properties of regularly varying functions; see also [Resnick 1987; Cîrstea and Trombetti 2008].

Proposition 2.4 (Uniform convergence theorem). *If L is slowly varying,* $\dfrac{L(\lambda u)}{L(u)}$ *tends to* 1 *as $u \to \infty$, uniformly on each compact λ-set in* $(0, \infty)$.

Proposition 2.5. (See also Proposition 4.9 in [Cîrstea and Trombetti 2008].)

(i) *If $R \in \mathbb{RV}_q$, then* $\lim_{u\to\infty} \log R(u)/\log u = q$.

(ii) *If $R_1 \in \mathbb{RV}_{q_1}$ and $R_2 \in \mathbb{RV}_{q_2}$ with* $\lim_{u\to\infty} R_2(u) = \infty$, *then*
$$R_1 \circ R_2 \in \mathbb{RV}_{q_1 q_2}.$$

(iii) *Suppose R is nondecreasing and $R \in \mathbb{RV}_q$, $0 < q < \infty$. Then*
$$R^{\leftarrow} \in \mathbb{RV}_{q^{-1}}.$$

(iv) *Suppose R_1, R_2 are nondecreasing and q-varying with $q \in (0, \infty)$. Then, for $c \in (0, \infty)$, we have*
$$\lim_{u\to\infty} \frac{R_1(u)}{R_2(u)} = c \quad \text{if and only if} \quad \lim_{u\to\infty} \frac{R_1^{\leftarrow}(u)}{R_2^{\leftarrow}(u)} = c^{-1/q}.$$

Proposition 2.6. (See also Proposition 4.10 in [Cîrstea and Trombetti 2008]). *Let $R \in \mathbb{RV}_q$ and choose $B \geq 0$ so that R is locally bounded on $[B, \infty)$. If $q > 0$, then*

(a) $\sup\{R(y) : B \leq y \leq u\} \sim R(u)$ *as $u \to \infty$,*

(b) $\inf\{R(y) : y \geq u\} \sim R(u)$ *as $u \to \infty$.*

If $q < 0$, then

(c) $\inf\{R(y) : y \geq u\} \sim R(u)$ *as $u \to \infty$,*

(d) $\inf\{R(y) : B \leq y \leq u\} \sim R(u)$ *as $u \to \infty$.*

3. Asymptotic properties of ϕ

Using Karamata's theory of regular variation and its extensions, we now consider the asymptotic properties of the function ϕ defined in (1-10).

Lemma 3.1. *Let $m \in \mathfrak{K}_\ell$ and $f \in \mathbb{RV}_q$ with $q > k$. If ϕ is defined by* (1-10), *then there exists a function $\psi \in C^2(0, \tau)$ with $\tau > 0$ which satisfies* $\lim_{t\to0} \psi(t)/\phi(t) = 1$ *and*

$$\lim_{t\to0} \frac{\psi(t)\psi''(t)}{(\psi'(t))^2} = 1 + \frac{(q-k)\ell}{k+1}, \tag{3-1}$$

$$\lim_{t\to0} \frac{(-\psi'(t))^{k-1}\psi''(t)}{m^{k+1}(t) f(\psi(t))} = \left(\frac{k+1}{q-k}\right)^{k+1}\left(1 + \frac{(q-k)\ell}{k+1}\right). \tag{3-2}$$

where ℓ appears in (1-3).

Proof. To prove (3-1), denote $g(u) = f(u)/u^k$. Since $g \in \mathbb{RV}_{q-k}$ and $q > k$, by Proposition 2.6 we have $\lim_{u\to\infty} g(u)/\mathcal{P}(u) = 1$. By Remark 4.8 in [Cîrstea and Trombetti 2008] we infer that there exists a function $\hat{g} \in C^2(0,\tau)$ such that $\lim_{u\to\infty} \hat{g}(u)/g(u) = 1$ and

$$\lim_{u\to\infty} \frac{u\hat{g}'(u)}{\hat{g}(u)} = q-k, \quad \lim_{u\to\infty} \frac{u\hat{g}''(u)}{\hat{g}'(u)} = q-k-1, \tag{3-3}$$

where we have used $g \in \mathbb{RV}_{q-k}$.

We define ψ by

$$\hat{g}(\psi(t)) = \left(\int_0^t m(s)\,ds\right)^{-k-1}, \quad \text{for } t>0 \text{ small.} \tag{3-4}$$

Notice that

$$\phi(t) = \mathcal{P}^{\leftarrow}\left(\left(\int_0^t m(s)\,ds\right)^{-k-1}\right), \quad \text{for } t>0 \text{ small.} \tag{3-5}$$

Thus Proposition 2.5 gives

$$\lim_{t\to 0} \frac{\hat{g}^{\leftarrow}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)}{\mathcal{P}^{\leftarrow}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)} = \lim_{t\to 0} \frac{\hat{g}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)}{\mathcal{P}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)} = 1,$$

where we have used $\lim_{u\to\infty} g(u)/\mathcal{P}(u) = 1$ and $\lim_{u\to\infty} \hat{g}(u)/g(u) = 1$ in the last equality.

By the definition of the inverse of $\hat{g}$ we see that

$$\lim_{t\to 0} \frac{\psi(t)}{\phi(t)} = \lim_{t\to 0} \frac{\hat{g}^{\leftarrow}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)}{\mathcal{P}^{\leftarrow}\big(\big(\int_0^t m(s)\,ds\big)^{-k-1}\big)} = 1. \tag{3-6}$$

By differentiating (3-4) we obtain

$$\hat{g}'(\psi(t))\psi'(t) = -(k+1)\left(\int_0^t m(s)\,ds\right)^{-k-2} m(t), \quad \text{for } t>0 \text{ small.} \tag{3-7}$$

Then, by (3-3), (3-4) and (3-7),

$$\frac{\psi'(t)}{\psi(t)} \sim \frac{-(k+1)}{q-k} \frac{m(t)}{\int_0^t m(s)\,ds}, \quad \text{as } t\to 0. \tag{3-8}$$

We differentiate (3-7), then use (1-3) and (3-3) to deduce that as $t\to 0$

$$\begin{aligned}\hat{g}'(\psi(t))\frac{(\psi'(t))^2}{\psi(t)}&\left(q-k-1+\frac{\psi(t)\psi''(t)}{(\psi'(t))^2}\right)\\ &\sim (k+1)(k+1+\ell)m^2(s)\left(\int_0^t m(s)\,ds\right)^{-k-3}.\end{aligned} \tag{3-9}$$

Putting (3-7) and (3-8) into (3-9), we have

$$-(k+1)\left(\int_0^t m(s)\,ds\right)^{-k-2} m(t)\frac{-(k+1)}{q-k}\frac{m(t)}{\int_0^t m(s)\,ds}\left(q-k-1+\frac{\psi(t)\psi''(t)}{(\psi'(t))^2}\right)$$

$$= \frac{(k+1)^2}{q-k} m^2(t)\left(\int_0^t m(s)\,ds\right)^{-k-3}\left(q-k-1+\frac{\psi(t)\psi''(t)}{(\psi'(t))^2}\right) \tag{3-10}$$

$$\sim (k+1)(k+1+\ell) m^2(s)\left(\int_0^t m(s)\,ds\right)^{-k-3}.$$

Thus,

$$\frac{(k+1)}{q-k}\left(q-k-1+\frac{\psi(t)\psi''(t)}{(\psi'(t))^2}\right) \sim (k+1+\ell). \tag{3-11}$$

(3-1) now follows from (3-11).

From (3-4) and (3-8), we find

$$\lim_{t\to 0}\left(-\frac{\psi'(t)}{\psi(t)}\right)^{k+1}\frac{1}{m^{k+1}(t)\hat{g}(\psi(t))} = \left(\frac{k+1}{q-k}\right)^{k+1}. \tag{3-12}$$

This, combined with (3-1), proves (3-2). □

4. Proof of Theorem 1.3

Fix $\epsilon \in (0, 1/2)$ and choose $\delta > 0$ small enough such that:

(a) m is nondecreasing on $(0, 2\delta)$.

(b) $\beta^-(1-\epsilon)\big(m(d(x))\big)^{k+1} \le b(x) \le \beta^+(1+\epsilon)\big(m(d(x))\big)^{k+1}$, for every $x \in \Omega_{2\delta}$, where for $\lambda > 0$ we set

$$\Omega_\lambda = \{x \in \Omega : d(x) < \lambda\}.$$

(c) $d(x)$ is a C^2 function on $\Gamma_{2\delta} = \{x \in \overline{\Omega} : d(x) < 2\delta\}$.

(d) $0 < \psi$, $\psi' < 0$, and $\psi'' > 0$ on $(0, 2\delta)$, where ψ is as in Lemma 3.1.

(e) $\sigma_{k-1}(\operatorname{diag}(1-\rho_1(y)d(x), \dots, 1-\rho_{n-1}(y)d(x))) > 1-\varepsilon$, for every $x \in \Omega_{2\delta}$. Recall that $\rho_i(y)$, $i = 1, \dots, n-1$, denote the principal curvatures of $\partial\Omega$ at y, where $y \in \partial\Omega$ is such that $|x-y| = d(x)$.

Fix $\tau \in (0, \delta)$. With $\xi^\pm$ given by (1-11), we set

$$\eta^\pm = \big((1\mp\varepsilon)(1\mp 2\varepsilon)\big)^{1/(k-q)}\xi^\pm. \tag{4-1}$$

Define

$$\begin{cases} v_\tau^+ = \eta^+\psi((1-e^{-T(d(x)-\tau)})/T), & x \in \Omega_{2\delta}\setminus\overline{\Omega}_\tau, \\ v_\tau^- = \eta^-\psi((1-e^{-T(d(x)+\tau)})/T), & x \in \Omega_{2\delta-\tau}. \end{cases} \tag{4-2}$$

Step 1. *We prove that, near the boundary,* v_τ^+*(respectively,* v_τ^-*) is an upper (respectively, lower) solution of* (1-1), *that is,*

$$\begin{cases} H_k[D^2 v_\tau^+] \le b(x) f(v_\tau^+), & x \in \Omega_{2\delta} \setminus \overline{\Omega}_\tau, \\ H_k[D^2 v_\tau^-] \ge b(x) f(v_\tau^-), & x \in \Omega_{2\delta-\tau}. \end{cases} \tag{4-3}$$

We denote by

$$M^+ = \max_{y \in \partial\Omega} \sigma_{k-1}(y) \quad \text{and} \quad M^- = \min_{y \in \partial\Omega} \sigma_{k-1}(y). \tag{4-4}$$

After some computations we obtain, for a point $x \in \Omega_{2\delta} \setminus \overline{\Omega}_\tau$,

$$[v_\tau^+]_{ij} = \eta^+ e^{-T(d(x)-\tau)} \big(\psi' d_{ij} + d_i d_j (\psi'' e^{-T(d(x)-\tau)} - T\psi')\big).$$

Since $|Dd(x)| = 1$ in $x \in \Omega_{2\delta} \setminus \overline{\Omega}_\tau$, we can choose a coordinate system such that

$$\begin{aligned} Dd(x) &= (0, \dots, 0, 1), \\ D^2 d(x) &= \operatorname{diag}\big(d_{11}(x), \dots, d_{n-1,n-1}(x), 0\big), \end{aligned}$$

where $d_{ii}(x) = -\rho_i(y)/(1 - \rho_i(y) d(x))$, and $y \in \partial\Omega$ is such that $|x - y| = d(x)$ as in Corollary 2.3.

Hence

$$D^2 v_\tau^+ = \eta^+ e^{-T(d(x)-\tau)} \operatorname{diag}\big(\psi' d_{11}(x), \dots, \psi' d_{n-1,n-1}(x), \psi'' e^{-T(d(x)-\tau)} - T\psi'\big).$$

Using this and Corollary 2.3, we can easily compute the k-Hessian of v_τ^+:

$$\begin{aligned} H_k[D^2 v_\tau^+] &= (\eta^+)^k e^{-(k+1)T(d(x)-\tau)} [-\psi']^{k-1} \psi'' \sigma_{k-1}(-D^2 d(x)) \\ &\quad + (\eta^+)^k e^{-kT(d(x)-\tau)} [-\psi']^k \left(T \sigma_{k-1}(-D^2 d(x)) + \sigma_k(-D^2 d(x))\right). \end{aligned} \tag{4-5}$$

Now, if

$$T_1 \le -\frac{\max_{\Omega_{2\delta} \setminus \overline{\Omega}_\tau} |\sigma_k(D^2 d(x))|}{\min_{\Omega_{2\delta} \setminus \overline{\Omega}_\tau} \sigma_{k-1}(-D^2 d(x))},$$

then (4-5) and condition (e) yield for $T \le T_1$,

$$\begin{aligned} H_k[D^2 v_\tau^+] &\le (\eta^+)^k e^{-(k+1)T(d(x)-\tau)} [-\psi']^{k-1} \psi'' \sigma_{k-1}(-D^2 d(x)), \\ &\le \frac{(\eta^+)^k}{1-\varepsilon} M^+ e^{-(k+1)T(d(x)-\tau)} [-\psi']^{k-1} \psi'', \ x \in \Omega_{2\delta} \setminus \overline{\Omega}_\tau. \end{aligned}$$

Similarly, we have for T_2

$$T_2 \ge \frac{\max_{\Omega_{2\delta-\tau}} |\sigma_k(D^2 d(x))|}{\min_{\Omega_{2\delta-\tau}} \sigma_{k-1}(-D^2 d(x))},$$

for $T \geq T_2$,

$$H_k[D^2 v_\tau^-] \geq (\eta^-)^k e^{-(k+1)T(d(x)+\tau)} [-\psi']^{k-1} \psi'' \sigma_{k-1}(-D^2 d(x)),$$
$$\geq \frac{(\eta^-)^k}{1+\varepsilon} M^- e^{-(k+1)T(d(x)+\tau)} [-\psi']^{k-1} \psi'', \quad x \in \Omega_{2\delta-\tau}.$$

Therefore, to deduce (4-3) it is enough to establish that

$$\lim_{t\to 0} (\eta^\pm)^k \frac{M^\pm}{\beta^\mp} \frac{[-\psi'(t)]^{k-1}\psi''(t)}{m^{k+1}(t) f(\eta^\pm \psi(t))} = (1 \mp \varepsilon)(1 \mp \varepsilon). \tag{4-6}$$

Since $f \in \mathbb{RV}_q$, Lemma 3.1 and our choice of $\eta^\pm$ in (4-1),

$$\lim_{t\to 0} (\eta^\pm)^k \frac{M^\pm}{\beta^\mp} \frac{[-\psi'(t)]^{k-1}\psi''(t)}{m^{k+1}(t) f(\eta^\pm \psi(t))} = (\eta^\pm)^k \frac{M^\pm}{\beta^\mp} \Big(\frac{k+1}{q-k}\Big)^{k+1} \Big(1 + \frac{(q-k)\ell}{k+1}\Big) (\eta^\pm)^{-q}$$
$$= \big((1\mp\varepsilon)(1\mp 2\varepsilon)\big) \xi^{\pm(k-q)} \frac{M^\pm}{\beta^\mp} \Big(\frac{k+1}{q-k}\Big)^{k+1} \Big(1 + \frac{(q-k)\ell}{k+1}\Big) = (1\mp\varepsilon)(1\mp 2\varepsilon),$$

where we have used (1-11) in the last equality.

Step 2. *Every strictly k-convex blow-up solution u of* (1-1) *satisfies* (1-9).

Let $C = \max_{d(x)=\delta} u(x)$. Notice that

$$\begin{cases} v_\tau^+ + C = \infty > u(x), & x \in \Omega \text{ with } d(x) = \tau, \\ v_\tau^+ + C \geq u(x), & x \in \Omega \text{ with } d(x) = \delta. \end{cases} \tag{4-7}$$

Using (4-3) we deduce that for every $x \in \Omega_\delta \setminus \overline{\Omega}_\tau$,

$$H_k[D^2(v_\tau^+ + C)] = H_k[D^2 v_\tau^+] \leq b(x) f(v_\tau^+) \leq b(x) f(v_\tau^+ + C).$$

Since u is a solution to (1-1), by the comparison principle for k-Hessians [Jian 2006, Lemma 2.1] we find

$$v_\tau^+ + C \geq u(x), \quad \text{for all } x \in \Omega_\delta \setminus \overline{\Omega}_\tau. \tag{4-8}$$

We set $C' = \xi^- \psi(\delta)$. Hence, we have $C' \geq v_\tau^-(x)$ for every $x \in \Omega$ with $d(x) = \delta - \tau$. It follows that

$$u(x) + C' \geq v_\tau^-(x), \quad \text{for all } x \in \partial\Omega_{\delta-\tau}. \tag{4-9}$$

We see that, for every $x \in \Omega_{\delta-\tau}$,

$$H_k[u(x) + C'] = H_k[D^2 u(x)] \leq b(x) f(u(x)) \leq b(x) f(u(x) + C'),$$

while by (4-3) we have

$$H_k[D^2 v_\tau^-] \geq b(x) f(v_\tau^-), \quad x \in \Omega_{\delta-\tau}. \tag{4-10}$$

Using again the comparison principle for k-Hessian equations, we infer that

$$u(x)+C' \geq v_\tau^-(x), \quad \text{for all } x \in \Omega_{\delta-\tau}. \tag{4-11}$$

By (4-8) and (4-11), letting $\tau \to 0$ we obtain

$$\begin{cases} \big((1+\epsilon)(1+2\epsilon)\big)^{1/(k-q)}\xi^- \psi((1-e^{-T_2 d(x)})/T_2) - C' \leq u(x), \ x \in \Omega_\delta, \\ u(x) \leq \big((1-\epsilon)(1-2\epsilon)\big)^{1/(k-q)}\xi^+ \psi((1-e^{-T_1 d(x)})/T_1) + C. \end{cases} \tag{4-12}$$

Dividing by $\psi((1-e^{-T_i d(x)})/T_i)$ for $i=1,2$ and noticing that $\lim_{t\to 0} \psi(t)/\phi(t) = 1$, letting $d(x) \to 0$, we obtain

$$\begin{cases} \liminf_{d(x)\to 0} u/\phi(d(x)) \geq \big((1+\epsilon)(1+2\epsilon)\big)^{1/(k-q)}\xi^-, \\ \liminf_{d(x)\to 0} u/\phi(d(x)) \leq \big((1-\epsilon)(1-2\epsilon)\big)^{1/(k-q)}\xi^+. \end{cases} \tag{4-13}$$

Since $\epsilon > 0$ is arbitrary, we let $\epsilon \to 0$ and obtain (1-9). This completes the proof of Theorem 1.3.

5. Proof of Theorem 1.6

We follow the methods in [Cîrstea and Trombetti 2008] and divide the proof into two steps:

Step 1. *For all strictly k-convex blow-up solutions u_1, u_2 of* (1-1),

$$\lim_{d(x)\to 0} \frac{u_1(x)}{u_2(x)} = 1. \tag{5-1}$$

Step 2. *There is at most one strictly convex blow-up solution of* (1-1).

Proof of Step 1. The argument breaks into two cases.

Case (i): $b > 0$ on $\overline{\Omega}$. Since u_1 and u_2 are arbitrary, it suffices to show that

$$\liminf_{d(x)\to 0} \frac{u_1(x)}{u_2(x)} \geq 1. \tag{5-2}$$

Without loss of generality, we can assume that 0 belongs to Ω. Let $\varepsilon \in (0, 1)$ be fixed and let $\lambda > 1$ be close to 1.

We set

$$C_\lambda = \Big((1+\varepsilon)\lambda^{2k} \max_{x\in(1/\lambda)\overline{\Omega}} \frac{b(\lambda x)}{b(x)}\Big)^{1/(q-k)}, \tag{5-3}$$

where $(1/\lambda)\overline{\Omega} = \big\{(1/\lambda)x : x \in \overline{\Omega}\big\}$. Notice that $C_\lambda \to (1+\varepsilon)^{1/(q-k)}$ as $\lambda \to 1$.

Hence, by Proposition 2.4 and $\lim_{d(x)\to 0} u_1(x) = \infty$, we deduce that there exists $\delta = \delta(\varepsilon) > 0$, independent of λ, such that

$$C_\lambda^q \frac{f(u_1(x))}{f(C_\lambda u_1(x))} \leq 1+\varepsilon, \quad \text{for all } x \in \Omega_\delta \text{ and } \lambda \in (1, 1+\eta) \text{ for some } \eta. \tag{5-4}$$

We now define U_λ as

$$U_\lambda(x) = C_\lambda u_1(\lambda x), \quad \text{for all } x \in (1/\lambda)\Omega_\delta. \tag{5-5}$$

Notice by (5-3)–(5-5),

$$\begin{aligned} H_k[D^2 U_\lambda(x)] &= \lambda^{2k} C_\lambda^k b(\lambda x) f(u_1(\lambda x)) \\ &\le \lambda^{2k} C_\lambda^{k-q}(1+\varepsilon) b(\lambda x) f(C_\lambda u_1(\lambda x)) \\ &\le b(x) f(C_\lambda u_1(\lambda x)) = b(x) f(U_\lambda(x)), \quad x \in (1/\lambda)\Omega_\delta, \end{aligned} \tag{5-6}$$

which says that $U_\lambda(x)$ is a supersolution of (1-1) with domain $(1/\lambda)\Omega_\delta$.

Since f is increasing on $(0, \infty)$ and (5-6), for each constant $M > 0$,

$$\begin{aligned} H_k[D^2(U_\lambda(x) + M)] = H_k[D^2 U_\lambda(x)] &\le b(x) f(U_\lambda(x)) \\ &\le b(x) f(U_\lambda(x) + M), \quad \text{for all } x \in (1/\lambda)\Omega_\delta. \end{aligned} \tag{5-7}$$

Notice also that $U_\lambda(x) = \infty > u_2(x)$, for every $x \in (1/\lambda)\partial\Omega$. Moreover, $x \in (1/\lambda)\partial\Omega$ implies that $d(x) < \delta$ (as $\lambda > 1$ is close to 1).

Thus, if we choose $M > 0$ large enough (for example, $M = \max_{d(x)=\delta} u_2(x)$), then by the comparison principle for k-Hessian equations we obtain

$$U_\lambda(x) + M \ge u_2(x), \quad \text{for all } x \in \Omega_\delta \cap (1/\lambda)\Omega_\delta. \tag{5-8}$$

Letting $\lambda \to 1$ in (5-8), we find

$$(1+\varepsilon)^{1/(q-k)} u_1(x) + M \ge u_2(x), \quad \text{for all } x \in \Omega_\delta, \tag{5-9}$$

which implies that

$$\liminf_{d(x)\to 0} \frac{u_1(x)}{u_2(x)} \ge (1+\varepsilon)^{1/(k-q)}, \tag{5-10}$$

and then letting $\varepsilon \to 0$ we obtain (5-2).

Case (ii): $b \equiv 0$ on $\partial\Omega$, Ω is a ball of radius $R > 0$, and (1-13) holds. By Corollary 1.5, every strictly k-convex blow-up solution u of (1-1) satisfies

$$\lim_{d(x)\to 0} \frac{u}{\phi(d(x))} = \left(\frac{\big((q-k)/(k+1)\big)^{k+1} R^{k-1}}{1 + \ell(q-k)/(k+1)} \right)^{1/(k-q)}, \tag{5-11}$$

where ϕ is defined by (1-10) and ℓ appears in (1-3). □

Proof of Step 2. If u_1, u_2 are arbitrary strictly k-convex blow-up solutions of (1-1), it suffices to show that $u_1 \le u_2$ in Ω. Fix $\varepsilon > 0$. By Step 1 we infer that

$$\lim_{d(x)\to 0} \big(u_1(x) - (1+\varepsilon)u_2(x)\big) = -\infty. \tag{5-12}$$

Since $f(u)/u^k$ is increasing on $(0, \infty)$, we deduce that

$$\begin{aligned}H_k[D^2(1+\varepsilon)u_2(x)] = (1+\varepsilon)^k H_k[D^2u_2(x)] &\le (1+\varepsilon)^k b(x) f(u_2(x)) \\ &\le b(x) f((1+\varepsilon)u_2(x)), \quad \text{for all } x \in \Omega.\end{aligned} \tag{5-13}$$

By (5-12), (5-13) and the comparison principle for k-Hessian equations,

$$u_1 \le (1+\varepsilon)u_2, \quad \text{for all } x \in \Omega. \tag{5-14}$$

Letting $\varepsilon \to 0$, thus $u_1 \le u_2$ in Ω. This completes the proof of Step 2 and hence of Theorem 1.6. □

Acknowledgments

We thank Professor Huaiyu Jian for his encouragement and many suggestions in this project, and Doctor Lu Xu for helpful discussions.

References

[Bandle and Marcus 1992] C. Bandle and M. Marcus, "“Large” solutions of semilinear elliptic equations: existence, uniqueness and asymptotic behaviour", *J. Anal. Math.* **58** (1992), 9–24. MR 94c:35081 Zbl 0802.35038

[Bandle and Marcus 1995] C. Bandle and M. Marcus, "Asymptotic behaviour of solutions and their derivatives, for semilinear elliptic problems with blowup on the boundary", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **12**:2 (1995), 155–171. MR 96e:35038 Zbl 0840.35033

[Caffarelli et al. 1985] L. Caffarelli, L. Nirenberg, and J. Spruck, "The Dirichlet problem for nonlinear second-order elliptic equations. III. Functions of the eigenvalues of the Hessian", *Acta Math.* **155**:3-4 (1985), 261–301. MR 87f:35098 Zbl 0654.35031

[Chuaqui et al. 2004] M. Chuaqui, C. Cortazar, M. Elgueta, and J. Garcia-Melian, "Uniqueness and boundary behavior of large solutions to elliptic problems with singular weights", *Commun. Pure Appl. Anal.* **3**:4 (2004), 653–662. MR 2005h:35115 Zbl 02143210

[Cîrstea and Rădulescu 2006] F. C. Cîrstea and V. Rădulescu, "Nonlinear problems with boundary blow-up: a Karamata regular variation theory approach", *Asymptot. Anal.* **46**:3-4 (2006), 275–298. MR 2007a:35045 Zbl 05042405

[Cîrstea and Trombetti 2008] F. C. Cîrstea and C. Trombetti, "On the Monge–Ampère equation with boundary blow-up: existence, uniqueness and asymptotics", *Calc. Var. Partial Differential Equations* **31**:2 (2008), 167–186. MR 2008k:35178 Zbl 1148.35022

[Colesanti et al. 2000] A. Colesanti, P. Salani, and E. Francini, "Convexity and asymptotic estimates for large solutions of Hessian equations", *Differential Integral Equations* **13**:10-12 (2000), 1459–1472. MR 2001j:35075 Zbl 0977.35046

[García-Melián 2006] J. García-Melián, "Nondegeneracy and uniqueness for boundary blow-up elliptic problems", *J. Differential Equations* **223**:1 (2006), 208–227. MR 2007b:35113 Zbl 05019771

[García-Melián et al. 2001] J. García-Melián, R. Letelier-Albornoz, and J. Sabina de Lis, "Uniqueness and asymptotic behaviour for solutions of semilinear problems with boundary blow-up", *Proc. Amer. Math. Soc.* **129**:12 (2001), 3593–3602. MR 2002j:35117 Zbl 0989.35044

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Springer, Berlin, 1998. MR 2001k:35004 Zbl 1042.35002

[Guan and Jian 2004] B. Guan and H.-Y. Jian, "The Monge–Ampère equation with infinite boundary value", *Pacific J. Math.* **216**:1 (2004), 77–94. MR 2005f:35100 Zbl 1126.35318

[Jian 2006] H. Jian, "Hessian equations with infinite Dirichlet boundary value", *Indiana Univ. Math. J.* **55**:3 (2006), 1045–1062. MR 2008f:35120 Zbl 1126.35026

[Mohammed 2007] A. Mohammed, "On the existence of solutions to the Monge–Ampère equation with infinite boundary values", *Proc. Amer. Math. Soc.* **135**:1 (2007), 141–149. MR 2008f:35129 Zbl 05120216

[Osserman 1957] R. Osserman, "On the inequality $\Delta u \geq f(u)$", *Pacific J. Math.* **7** (1957), 1641–1647. MR 20 #4701 Zbl 0083.09402

[Resnick 1987] S. I. Resnick, *Extreme values, regular variation, and point processes*, Applied Probability **4**, Springer, New York, 1987. MR 89b:60241 Zbl 0633.60001

[Salani 1998] P. Salani, "Boundary blow-up problems for Hessian equations", *Manuscripta Math.* **96**:3 (1998), 281–294. MR 99e:35071 Zbl 0907.35052

[Takimoto 2006] K. Takimoto, "Solution to the boundary blowup problem for k-curvature equation", *Calc. Var. Partial Differential Equations* **26**:3 (2006), 357–377. MR 2007h:35115 Zbl 1105.35039

[Trudinger 1995] N. S. Trudinger, "On the Dirichlet problem for Hessian equations", *Acta Math.* **175**:2 (1995), 151–164. MR 96m:35113 Zbl 0887.35061

[Véron 1992] L. Véron, "Semilinear elliptic equations with uniform blow-up on the boundary", *J. Anal. Math.* **59** (1992), 231–250. MR 94k:35113 Zbl 0802.35042

Received September 18, 2008. Revised February 3, 2009.

YONG HUANG
WUHAN INSTITUTE OF PHYSICS AND MATHEMATICS
CHINESE ACADEMY OF SCIENCES
WUHAN 430071
CHINA

huangyong@wipm.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

EIGENVALUES OF THE STOKES OPERATOR VERSUS THE DIRICHLET LAPLACIAN IN THE PLANE

JAMES P. KELLIHER

We show that the k-th eigenvalue of the Dirichlet Laplacian is strictly less than the k-th eigenvalue of the classical Stokes operator (equivalently, of the clamped buckling plate problem) for a bounded domain in the plane having a locally Lipschitz boundary. For a C^2 boundary, we show that eigenvalues of the Stokes operator with Navier slip (friction) boundary conditions interpolate continuously between eigenvalues of the Dirichlet Laplacian and of the classical Stokes operator.

1. Introduction

Let Ω be a bounded domain in $\mathbb{R}^2$ with locally Lipschitz boundary Γ. Let σ_D be the spectrum of the negative Laplacian with homogeneous Dirichlet boundary conditions (which we refer to as the *Dirichlet Laplacian*) and let σ_S be the spectrum of the Stokes operator with homogeneous Dirichlet boundary conditions (which we refer to as the *classical Stokes operator*). Equivalently, σ_S is the set of eigenvalues of the clamped buckling plate problem [Payne 1955; 1967; Friedlander 2004]. Each spectrum is discrete with

$$\sigma_D = \{\lambda_j\}_{j=1}^{\infty}, \qquad \text{with } 0 < \lambda_1 < \lambda_2 \le \cdots, \tag{1-1}$$

$$\sigma_S = \{\nu_j\}_{j=1}^{\infty}, \qquad \text{with } 0 < \nu_1 \le \nu_2 \le \cdots, \tag{1-2}$$

each eigenvalue repeated according to its multiplicity.

Theorem 1.1. *For all positive integers k, we have $\lambda_k < \nu_k$.*

Further, let $\gamma_k(\theta)$ be the k-th eigenvalue of the Stokes operator with boundary conditions $(1-\theta)\omega(u) + \theta u \cdot \boldsymbol{\tau} = u \cdot \boldsymbol{n} = 0$, where $\omega(u) = \partial_1 u^2 - \partial_2 u^1$ is the vorticity of u, and $\boldsymbol{\tau}$ and $\boldsymbol{n}$ are the tangential and normal unit vectors; see Section 8 for details.

MSC2000: 35P99.
Keywords: Laplacian, Stokes operator, eigenvalues, clamped buckling plate.
This work was funded in part by NSF grant DMS-0705586. Part of the work was performed while the author was at Brown University.

Theorem 1.2. *If Γ is C^2 and has a finite number of components, for each positive integer k, the function γ_k is a strictly increasing continuous bijection from* $[0, 1]$ *onto* $[\lambda_k, \nu_k]$.

Theorem 1.1 is the analogue of the inequality $\mu_{k+1} < \lambda_k$ for $k = 1, 2, \ldots,$ proved by Filonov [2004]. Here, $\sigma_N = \{\mu_j\}_{j=1}^\infty$ is the spectrum of the negative Laplacian with homogeneous Neumann boundary conditions (which we refer to as the *Neumann Laplacian*). Then σ_N is also discrete with $0 = \mu_1 < \mu_2 \le \cdots$. Filonov's inequality applies in $\mathbb{R}^d$ for $d \ge 2$ and only requires that Ω have finite measure and that its boundary have sufficient regularity that the embedding of $W^1(\Omega)$ in $L^2(\Omega)$ is compact, which is slightly weaker than our assumption that Γ is locally Lipschitz. Because of the need to integrate by parts, however, we require the additional regularity.

Filonov's strict inequality is a strengthening of the partial inequality $\mu_{k+1} \le \lambda_k$ proved by L. Friedlander in [1991] using very different techniques.

A fairly direct variational argument shows that $\lambda_k \le \nu_k$; see Remark 5.3 or [Ashbaugh 2004, Equation (1.8)]. We are interested in the strict inequality.

For the unit disk, where one can calculate the eigenfunctions explicitly,

$$\sigma_D = \{j_{nk}^2 : n = 0, 1 \ldots, k = 1, 2, \ldots\},$$
$$\sigma_S = \{j_{nk}^2 : n = 1, 2 \ldots, k = 1, 2, \ldots\},$$

where j_{nk} is the k-th positive zero of the Bessel function J_n of the first kind of order n. Each eigenvalue has multiplicity 2 except for $\{j_{0k}^2 : k \in \mathbb{N}\} \subseteq \sigma_D$ and $\{j_{1k}^2 : k \in \mathbb{N}\} \subseteq \sigma_S$, which have multiplicity 1. This gives the ordering

$$0 < \lambda_1 < \lambda_2 = \lambda_3 = \nu_1 < \lambda_4 = \lambda_5 = \nu_2 = \nu_3 < \lambda_6 < \cdots.$$

In this case we have $\lambda_{k+1} \le \nu_k$ for all k, but $\lambda_{k+1} \not< \nu_k$ for $k = 1$. This leaves open the possibility that $\lambda_{k+1} \le \nu_k$ in full generality. This inequality was conjectured to hold by L. E. Payne many years ago, but has remained unproved.

To prove Theorem 1.1 we adapt Filonov's proof [2004] that $\mu_{k+1} < \lambda_k$, which is shockingly direct and simple. As we observed for a disk, $\lambda_{k+1} \not< \nu_k$, which shows that some aspect of Filonov's approach must fail if we attempt to adapt it to obtain Theorem 1.1. In fact, what fails is his use of a function of the form $f = e^{i\omega\cdot x}$ with $|\omega|^2 = \lambda$ for $\lambda > 0$, which has the properties that $\Delta f + \lambda f = 0$ and $|\nabla f|^2 = \lambda |f|$. This serves as an "extra" function that increases the dimension of a subspace of functions that he shows satisfy the bound in the variational formulation of the eigenvalue problem for the Neumann Laplacian. There can be no such function that will serve in general for us (else $\lambda_{k+1} < \nu_k$ would hold in general), but we describe the analogue of such a function in our setting in Section 7, show that given its existence we obtain $\lambda_{k+1} \le \nu_k$, and explain why it fails to give $\lambda_{k+1} < \nu_k$.

Our proof of $\lambda_k < \nu_k$ is largely a matter of transforming the eigenvalue problems so that the Stokes operator can play the role the Dirichlet Laplacian plays for Filonov and so that the Dirichlet Laplacian can play the role that the Neumann Laplacian plays for Filonov.

The approach of [Friedlander 1991] can also be adapted to prove Theorem 1.1, at least for C^1-boundaries.

In Section 8, we show that when Γ is C^2 and has a finite number of components, one can interpolate continuously between λ_j and ν_j using the eigenvalues of the negative Laplacian with Navier slip boundary conditions (Theorem 1.2). These boundary conditions, originally defined by Navier, have recently received considerable attention from fluid mechanics as a physically motivated replacement for Dirichlet boundary conditions, as they allow a thorough characterization of the boundary layer. See for instance [Clopeau et al. 1998; Lopes Filho et al. 2005; Kelliher 2006; Iftimie and Planas 2006; Iftimie and Sueur 2006]. We also discuss Neumann boundary conditions for the velocity and for the vorticity, and Robin boundary conditions for the vorticity.

This paper is organized as follows. We describe the necessary function spaces, trace operators, and related lemmas in Section 2. In Section 3, we define the classical Stokes operator and a variant of it using Lions boundary conditions (vanishing vorticity on the boundary). We show that the eigenvalue problem for the classical Stokes operator is equivalent to the eigenvalue problem for the clamped buckling plate problem. We also describe the strong forms of the associated eigenvalue problems in Section 3, giving the weak forms in Section 4. In Section 5 we describe the variational (min-max) formulations of the eigenvalue problems, using these formulations in Section 6 to prove Theorem 1.1. In Section 7, we describe the properties of the analogue of the function f used by Friedlander and Filonov and prove that its existence would imply the inequality $\lambda_{k+1} \leq \nu_k$. Finally, in Section 8, we discuss Navier boundary conditions and prove Theorem 1.2.

For a vector field u we define $u^\perp = (-u^2, u^1)$ and for a scalar field ψ we define $\nabla^\perp \psi = (-\partial_2 \psi, \partial_1 \psi)$. Observe that $(u^\perp)^\perp = -u$ and $(\nabla^\perp)^\perp \psi = -\nabla \psi$. By $\omega(u)$ we mean the vorticity (scalar curl) of u, that is, $\omega(u) = \partial_1 u^2 - \partial_2 u^1$. We make frequent use of the identities $\nabla^\perp \omega(u) = \Delta u$ and $\omega(u) = -\operatorname{div} u^\perp$, the former requiring that u be divergence-free.

Assumption. Unless specifically stated otherwise, we assume throughout that Ω is a bounded domain whose boundary Γ is locally Lipschitz.

2. Function spaces and related facts

Let $\boldsymbol{n}$ be the outward-directed unit vector normal to Γ, and let $\boldsymbol{\tau}$ be the unit tangent vector chosen so that $(\boldsymbol{n}, \boldsymbol{\tau})$ has the same orientation as the Cartesian unit vectors

$(\mathbf{e}_1, \mathbf{e}_2)$. These vectors are defined almost everywhere on Γ since Γ is locally Lipschitz.

The spaces $C^{k,\alpha}(\Omega)$, $C^{k,\alpha}(\overline{\Omega})$, and $W^s(\Omega)$ are the usual Hölder and L^2-based Sobolev spaces, with k an integer, $0 \le \alpha \le 1$, and s any real number. We need to say a few words about these spaces, which can be defined in various equivalent ways.

Define the norms

$$\|f\|_{C^k} = \sum_{j=0}^{m} \sup_{\Omega} \sup_{|\beta|=j} |D^\beta u|,$$

$$\|f\|_{C^{k,\alpha}} = \|f\|_{C^k} + \sup_{|\beta|=k} \sup_{x \neq y \in \Omega} \frac{|D^\beta f(x) - D^\beta f(y)|}{|x-y|^\alpha} \quad \text{for } 0 < \alpha \le 1.$$

Define $C^k(\Omega) = C^{k,0}(\Omega)$ and $C^{k,\alpha}(\Omega)$ to be the spaces of functions finite under their respective norms; $C^{k,\alpha}(\overline{\Omega})$ is defined similarly. Here β is a multiindex.

When $m \ge 0$ is an integer, $W^m(\Omega)$ is the completion of the space of all $C^\infty(\Omega)$ functions in the norm

$$\|f\|_{W^m} = \Big(\sum_{|\alpha| \le m} \|D^\alpha f\|^2_{L^2(\Omega)} \Big)^{1/2},$$

where α is a multiindex. Equivalently, $W^m(\Omega)$ is the space of all functions f such that $D^\alpha f$ is in $L^2(\Omega)$ for all $|\alpha| \le m$. $W^m_0(\Omega)$ is defined similarly as the closure of $C^\infty_0(\Omega)$ under the W^m norm. (See for instance [Adams 1975, Section 3.1].) $W^1_0(\Omega)$ can equivalently be defined as all functions in $W^1(\Omega)$ whose boundary trace is zero. $W^{-m}(\Omega)$ is the dual space of $W^m_0(\Omega)$. Fractional Sobolev spaces $W^s(\Omega)$ can be defined for instance as in [Adams 1975, Theorem 7.48].

On Ω, we will only need integer-order Hölder and Sobolev spaces, but on Γ we will need to use fractional spaces. Hölder spaces, however, can only be defined when the boundary has sufficient regularity.

We define a bounded domain Ω (or its boundary $\partial\Omega$) to be of class $C^{k,\alpha}$ for $k \ge 0$ an integer and $0 \le \alpha \le 1$ if locally there exists a $C^{k,\alpha}$ diffeomorphism ψ that maps Ω into the upper half-plane with $\partial\Omega$ being mapped to an open interval I. We say that φ is in $C^{k,\alpha}(\partial\Omega)$ if $\varphi \circ \psi^{-1}$ is in $C^{k,\alpha}(I)$. We also write C^k for $C^{k,0}$. If Ω is a $C^{k,\alpha}$ domain and φ lies in $C^{j,\beta}(\partial\Omega)$ for $j+\beta \le k+\alpha$, then there exists an extension of φ to $C^{j,\beta}(\overline{\Omega})$. See [Gilbarg and Trudinger 1977, Section 6.2] for more details. The inverse operation of restricting to the boundary gives an equivalent definition of $C^{k,\alpha}(\partial\Omega)$ as restrictions of functions in $C^{k,\alpha}(\overline{\Omega})$.

When Γ is locally Lipschitz, we will only have need for $W^s(\partial\Omega)$ for $s = \pm 1/2$ and 0. We define $W^{1/2}(\partial\Omega)$ to be the image (a subspace of $L^2(\partial\Omega)$) under the unique continuous extension to $W^1(\Omega)$ of the map that restricts the value of a

$C^\infty(\overline{\Omega})$ function to the boundary. The existence of this extension was proved by Gagliardo [1957] (or see [Grisvard 1985, Theorem 1.5.1.3]). Alternately, we could define $W^{1/2}(\Omega)$ intrinsically as in [Galdi 1994, Section II.3]. We define $W^{-1/2}(\partial\Omega)$ to be the space dual to $W^{1/2}(\partial\Omega)$ and let $W^0(\partial\Omega) = L^2(\partial\Omega)$.

For C^2 boundaries, we will need Corollary 2.2 and hence need to define $W^s(\partial\Omega)$ for all real s. We use the intrinsic definition of $W^s(\partial\Omega)$ due to J. L. Lions, which applies when the boundary is of class C^m for $m \geq 1$. This definition is similar to that for the Hölder spaces defined above, and requires for $s > 0$ that each $\varphi \circ \psi^{-1}$ be of class $W^s(I)$, where I is the domain of ψ^{-1}. (See [Adams 1975, pages 215–217] for details.) For $s < 0$ we define $W^s(\partial\Omega)$ to be the dual space of $W^{-s}(\partial\Omega)$ and let $W^0(\partial\Omega) = L^2(\partial\Omega)$ as above. It follows from [Adams 1975, Theorem 7.53] that the two definitions of these spaces are equivalent for $0 < s \leq m$ and hence for all real s. (Adams gives the proof only for $s = m - 1/2$, from which it follows immediately for all $s = j - 1/2$, where j is an integer with $1 \leq j \leq m$, since if $\partial\Omega$ is of class C^m it is of class C^k for all $1 \leq k \leq m$. We only need the equivalence for $m = 2$ and $s = 1/2$, so this will suffice.)

Lemma 2.1. *Let D be any bounded domain in $\mathbb{R}^n$ with C^∞ boundary. Let φ lie in $C^{k,\alpha}(\overline{D})$ and f lie in $W^s(D)$ for $s > 0$. Then φf lies in $W^s(D)$ as long as*

$$\begin{cases} k+\alpha \geq s & \text{if } s \text{ is an integer,} \\ k+\alpha > s & \text{if } s \text{ is not an integer.} \end{cases}$$

Let g lie in $W^{s'}(D)$. Then fg lies in $W^s(D)$ if $s' > s$ and $s' \geq n/2$ or if $s' \geq s$ and $s' > n/2$.

Proof. This follows from [Galdi 1994, Theorems 1.4.1.1 and 1.4.4.2]. □

Corollary 2.2. *Assume that Γ is of class $C^{k,\alpha}$. Let $\varphi \in C^{j,\beta}(\partial\Omega)$ for $j+\beta \leq k+\alpha$, and let $f \in W^s(\Gamma)$ for $s > 0$. Then $\varphi f \in W^s(\Gamma)$ as long as*

$$\begin{cases} j+\beta \geq s & \text{if } s \text{ is an integer,} \\ j+\beta > s & \text{if } s \text{ is not an integer.} \end{cases}$$

If $f \in W^s(\Gamma)$ and $\varphi \in W^{s+\epsilon}(\Gamma)$ with $\epsilon > 0$, then $\varphi f \in W^s(\Gamma)$ if $s \geq 1/2$.

Proof. Apply Lemma 2.1 to the functions $\varphi \circ \psi^{-1}$ and $f \circ \psi^{-1}$ with domain $D = I$. □

Corollary 2.3. *Assume that Γ is C^2. Then $g\boldsymbol{\tau}$ and $g\boldsymbol{n}$ are in $W^{1/2}(\Gamma)$ for any g in $W^{1/2}(\Gamma)$, and $u \cdot \boldsymbol{\tau}$ and $u \cdot \boldsymbol{n}$ are in $W^{1/2}(\Gamma)$ for any u in $(W^{1/2}(\Gamma))^2$.*

Proof. Because Γ is C^2, $\boldsymbol{\tau}$ and $\boldsymbol{n}$ are in $C^1 = C^{1,0}$. But $1+0 > 1/2$, so the second condition in Corollary 2.2 applies in each case to give the result. □

Let $\mathcal{V} = \{u \in (C_0^\infty(\Omega))^2 : \operatorname{div} u = 0\}$ be the space of complex vector-valued divergence-free test functions on Ω. We let H be the completion of $\mathcal{V}$ in $L^2(\Omega)$

and V be the completion of $\mathcal{V}$ in $W_0^1(\Omega)$. These definitions of H and V are valid for arbitrary domains. We will also find use for the space

$$E(\Omega) = \{v \in (L^2(\Omega))^2 : \ \operatorname{div} v \in L^2(\Omega)\} \tag{2-1}$$

with $\|u\|_{E(\Omega)} = \|u\|_{L^2(\Omega)} + \|\operatorname{div} u\|_{L^2(\Omega)}$.

We use $(\cdot\,,\cdot)$ to mean the inner product $(u, v) = \int_\Omega u\bar{v}$ in $L^2(\Omega)$ or sometimes to mean the pairing of v in a space Z with u in Z^* or of v in $\mathcal{D}(\Omega)$ with u in $\mathcal{D}'(\Omega)$. Which is meant is stated if it is not clear from context.

The integrations by parts we will make are justified by Lemma 2.4, which is [Temam 1984, Theorem 1.2, page 7] for locally Lipschitz domains. (Temam states the theorem for C^2 boundaries but the proof for locally Lipschitz boundaries is the same, using a trace operator for Lipschitz boundaries in place of that for C^2 boundaries: see [Galdi 1994, pages 117–119, specifically Theorem 2.1, page 119].)

Lemma 2.4. *There is an extension of the trace operator $\gamma_{\boldsymbol{n}} : (C_0^\infty(\overline{\Omega}))^2 \to C^\infty(\Gamma)$, $u \mapsto u \cdot \boldsymbol{n}$, on Γ to a continuous linear operator from $E(\Omega)$ onto $W^{-1/2}(\Gamma)$. The kernel of $\gamma_{\boldsymbol{n}}$ is the space $E_0(\Omega)$ — the completion of $C_0^\infty(\Omega)$ in the $E(\Omega)$ norm. For all u in $E(\Omega)$ and f in $W^1(\Omega)$,*

$$(u, \nabla f) + (\operatorname{div} u, f) = \int_\Gamma (u \cdot \boldsymbol{n})\,\bar{f}. \tag{2-2}$$

Remark 2.5. In (2-2) and in what follows we usually do not explicitly include the trace operators. On the right side of (2-2), for instance, $u \cdot \boldsymbol{n}$ is actually $\gamma_{\boldsymbol{n}} u$, which is thus in $W^{-1/2}(\Gamma)$, and f is actually $\gamma_0 f$, where γ_0 is the usual trace operator from $W^s(\Omega)$ to $W^{s-1/2}(\Gamma)$ for all $s > 1/2$. Also, the boundary integral should more properly be written as a pairing in the duality between $W^{-1/2}(\Gamma)$ and $W^{1/2}(\Gamma)$ of $u \cdot \boldsymbol{n}$ and f.

Lemma 2.6. *$W^s(\Omega)$ is compactly embedded in $W^r(\Omega)$ for all $s > r \geq 0$.*

Proof. This is an instance of the Rellich–Kondrachov theorem. That it holds for a bounded domain with locally Lipschitz boundary follows, for instance, from the comments on [Adams 1975, page 67 and Theorem 6.2, page 144]. □

We will use several times a basic result of elliptic regularity theory:

Lemma 2.7. *Let f lie in $W^{-1}(\Omega)$. There exists a unique ψ in $W_0^1(\Omega)$ that is a weak solution of $\Delta\psi = f$. Furthermore, $\|\psi\|_{W^1(\Omega)} \leq C\|f\|_{W^{-1}(\Omega)}$. When Γ is C^2 and f is in $L^2(\Omega)$,*

$$\|\psi\|_{W^2(\Omega)} \leq C\|\Delta\psi\|_{L^2(\Omega)}.$$

Proof. See for instance [Kesavan 1989, pages 118–121] for general bounded open domains and [Evans 1998, Theorem 4 and the remark following it on page 317] for C^2 boundaries. □

Poincaré's inequality holds in both its classical forms:

Lemma 2.8. *Let f lie in $W_0^1(\Omega)$ or else lie in $W^1(\Omega)$ with $\int_\Omega f = 0$. Then there exists a constant C such that $\|f\|_{L^2(\Omega)} \le C\|\nabla f\|_{L^2(\Omega)}$.*

Proof. See [Galdi 1994, Theorem 4.1 on page 49, and Theorem 4.3 on page 54]. □

Since Γ is locally Lipschitzian, we can define

$$\hat{H} = \{u \in (L^2(\Omega))^2 \ : \operatorname{div} u = 0 \text{ in } \Omega,\ \gamma_{\boldsymbol{n}} u = 0 \text{ on } \Gamma\},$$
$$\hat{V} = \{u \in (W^1(\Omega))^2 : \operatorname{div} u = 0 \text{ in } \Omega,\ \gamma_0 u = 0 \text{ on } \Gamma\}.$$

By the continuity of the trace operators $\gamma_{\boldsymbol{n}}$ and γ_0, it follows that $H \subseteq \hat{H}$ and $V \subseteq \hat{V}$. When Γ is a bounded domain with locally Lipschitz boundary, $H = \hat{H}$ and $V = \hat{V}$. For $H = \hat{H}$, see [Temam 1984, Theorem 1.4 in Chapter 1]. That $V = \hat{V}$ is proved in [Maslennikova and Bogovskiĭ 1983]; see the comments of [Galdi 1994, page 148] and [Adams 1975, page 67].

Lemma 2.9. *Assume that u is in $(\mathscr{D}'(\Omega))^2$ with $(u, v) = 0$ for all v in $\mathscr{V}$. Then $u = \nabla p$ for some p in $\mathscr{D}'(\Omega)$. If u is in $(L^2(\Omega))^2$ then p is in $W^1(\Omega)$; if u is in H, then p is in $W^1(\Omega)$ and $\Delta p = 0$.*

Proof. For u in $(\mathscr{D}'(\Omega))^2$, see [Temam 1984, Proposition 1.1, page 10]. For u in $(L^2(\Omega))^2$, the result is a combination of [Galdi 1994, Theorem 1.1, page 103, and Remark 4.1, page 54]; also see [Temam 1984, Remark 1.4, page 11]. □

We will also find a need for the spaces

$$\begin{aligned} Y = Y^1 &= H \cap W^1(\Omega), & X = X^1 &= \{u \in H : \omega(u) \in L^2(\Omega)\}, \\ Y^2 &= \{u \in Y : \omega \in W^1(\Omega)\}, & X^2 &= \{u \in H : \omega(u) \in W^1\}, \\ Y_0^2 &= \{u \in Y : \omega(u) \in W_0^1\}, & X_0^2 &= \{u \in H : \omega(u) \in W_0^1\}, \end{aligned}$$

with the obvious norms on each space. We give Y the $W^1(\Omega)$ norm, but place no norm on the other spaces. When Γ is C^2 and has a finite number of components, the X and Y spaces coincide as in Corollary 2.16.

The average value of any vector u in H — and hence in all of our spaces — is zero, as can be seen by integrating $u \cdot e_i$ over Ω, where $e_i = \nabla x_i$ is a coordinate vector, and applying Lemma 2.4. Thus, Poincaré's inequality holds for Y and V so we can, and will, use $\|u\|_Y = \|u\|_V = \|\nabla u\|_{L^2(\Omega)}$ in place of the $W^1(\Omega)$ norm for these two spaces.

Let $H_c = \{v \in H : \omega(v) = 0\}$ and, noting that H_c is a closed subspace of H, let H_0 be the orthogonal complement of H_c in H. Thus, $H = H_0 \oplus H_c$ is an orthogonal decomposition of H. Observe that $V \cap H_0 = V$, and when Ω is simply connected, $H = H_0$.

Lemma 2.10. *For any u in H_0 there exists a stream function ψ in $W^1(\Omega)$ for u, that is, $u = \nabla^\perp \psi$, and ψ is unique up to the addition of a constant. Moreover,*

$$H_0 = \{\nabla^\perp \psi : \psi \in W_0^1(\Omega)\} = \nabla^\perp W_0^1(\Omega).$$

If u is in $H_0 \cap Y$, then ψ can be taken to lie in $W_0^1(\Omega) \cap W^2(\Omega)$, and if u is in V, then ψ can be taken to lie in $W_0^2(\Omega)$.

Proof. Let u be in H_0, and let ψ in $W_0^1(\Omega)$ solve $\Delta\psi = \omega(u) \in W^{-1}(\Omega)$ as in Lemma 2.7. Letting $w = \nabla^\perp \psi \in L^2(\Omega)$, we have $\omega(w) = \Delta\psi = \omega(u)$, $\operatorname{div} w = 0$, and $w \cdot \boldsymbol{n} = 0$ on Γ, so w is in H. Thus, w is a vector in H with the same vorticity as u, meaning that $u - w$ is in H_c.

We claim that w is in H_0. To see this, let v be in H_c. Then

$$(w, v) = (\nabla^\perp \psi, v) = (-\nabla\psi, v^\perp) = (\psi, \operatorname{div} v^\perp) + \int_\Gamma (v^\perp \cdot \boldsymbol{n})\psi = 0.$$

The last equality follows from $\operatorname{div} v^\perp = \omega(v) = 0$ (showing also that $v^\perp$ is in $E(\Omega)$ and allowing integration by parts via Lemma 2.4) and $\psi = 0$ on Γ. Since this is true for all v in H_c, it follows that w is in H_0.

Thus, both u and w are in H_0, so $u - w$ is in H_0. But we already saw that $u - w$ is in H_c, so $u - w = 0$.

What we have shown is both the existence of a stream function and the expression for H_0, the uniqueness of the stream function up to a constant being then immediate. The additional regularity of ψ for u in $H_0 \cap Y$ or V follows simply because $\nabla\psi = -u^\perp$ is in $W^1(\Omega)$. For u in V it is also true that $\nabla\psi = 0$ on Γ, so ψ can be taken to lie in $W_0^2(\Omega)$. $\square$

Closely related to Lemma 2.10 is Lemma 2.11, a form of the Biot–Savart law.

Lemma 2.11. *The operator ω is a continuous linear bijection between the following pairs of spaces:*

$$H_0 \text{ and } W^{-1}(\Omega), \quad H_0 \cap X \text{ and } L^2(\Omega), \quad H_0 \cap X_0^2 \text{ and } W_0^1(\Omega).$$

Proof. That ω has the domains and ranges stated and that it is continuous follow directly from the definitions of the spaces.

For ω in $W^{-1}(\Omega)$, let ψ in $W_0^1(\Omega)$ solve $\Delta\psi = \omega$ on Ω as in Lemma 2.7, and let $u = \nabla^\perp \psi$. Then $\omega(u) = \omega$ and if $\omega(v) = \omega$ as well for v in H_0, then $\omega(u - v) = 0$, implying that $u - v$ is in H_c. But $u - v$ is also in H_0 so $u - v = 0$. Thus, $u = \omega^{-1}(\omega)$ with $\|u\|_H = \|\nabla\psi\|_{L^2} \le C\|\omega\|_{W^{-1}(\Omega)}$ by Lemma 2.7, showing that ω^{-1} is defined and bounded and hence continuous, since it is clearly linear.

For ω in $L^2(\Omega)$ or $W_0^1(\Omega)$ the same argument applies, though now we use either

$$\|u\|_X = \|\nabla\psi\|_{L^2} + \|\omega(u)\|_{L^2} \le C\|\omega\|_{L^2} + \|\omega\|_{L^2}$$
$$\text{or } \|u\|_{X_0^2} = \|\nabla\psi\|_{L^2} + \|\omega(u)\|_{W^1} \le C\|\omega\|_{L^2} + \|\omega\|_{W^1} \le C\|\omega\|_{W^1}$$

to demonstrate the continuity of ω^{-1}. □

Corollary 2.12. *X is dense and compactly embedded in H, and X_0^2 is dense and compactly embedded in X.*

Proof. Let $A = L^2(\Omega)$ and $B = W^{-1}(\Omega)$ or $A = W_0^1(\Omega)$ and $B = L^2(\Omega)$. In both cases, A is dense and compactly embedded in B. Density is transferred to the image spaces $\omega^{-1}(A)$ and $\omega^{-1}(B)$ by virtue of ω^{-1} being a continuous surjection. The property that the spaces are compactly embedded transfers to the image spaces by virtue of ω being bounded (since it is continuous linear) along with ω^{-1} being a continuous surjection. □

We also have the following useful decomposition of $L^2(\Omega)$, variously named after some combination of Leray, Helmholtz, and Weyl.

Lemma 2.13. *For any u in $(L^2(\Omega))^2$, there exists a unique v in H and p in $W^1(\Omega)$ such that $u = v + \nabla p$.*

Proof. This follows, for instance, from [Galdi 1994, Theorem 1.1, page 107], which holds for an arbitrary domain, along with Lemma 2.9. □

The mapping $u \mapsto v$, with u and v as in Lemma 2.13, defines the *Leray projector* $\mathcal{P}$ from $(L^2(\Omega))^2$ onto H.

A slight strengthening of Poincaré's inequality holds on Y (and so on V) when Ω is simply connected:

Lemma 2.14. *For any u in $H_0 \cap X$,*

$$\|u\|_{L^2(\Omega)} \le C\|\omega(u)\|_{L^2(\Omega)}, \tag{2-3}$$

and when Γ is C^2,

$$\|\nabla u\|_{L^2(\Omega)} \le C\|\omega(u)\|_{L^2(\Omega)}. \tag{2-4}$$

Proof. As in the proof of Lemma 2.10, $u = \nabla^\perp\psi$ for ψ in $W_0^1(\Omega)$ with $\Delta\psi = \omega(u)$ in $L^2(\Omega)$, and $\|\psi\|_{L^2(\Omega)} \le \|\psi\|_{W^1(\Omega)} \le C\|\omega(u)\|_{L^2(\Omega)}$ by Lemma 2.7. But $\nabla\psi$ is in $E(\Omega)$ and ψ is in $W^1(\Omega)$ so by Lemma 2.4 we can integrate by parts to give $(\omega(u), \psi) = (\Delta\psi, \psi) = -(\nabla\psi, \nabla\psi) = -\|u\|^2_{L^2(\Omega)}$. Hence by the Cauchy–Schwarz inequality,

$$\|u\|^2_{L^2(\Omega)} \le \|\psi\|_{L^2(\Omega)}\|\omega(u)\|_{L^2(\Omega)} \le C\|\omega(u)\|^2_{L^2(\Omega)},$$

giving Equation (2-3).

When Γ is C^2, using Lemma 2.7,

$$\|\nabla u\|_{L^2(\Omega)} = \|\nabla\nabla\psi\|_{L^2(\Omega)} \le \|\psi\|_{W^2(\Omega)} \le C\|\Delta\psi\|_{L^2(\Omega)} = C\|\omega(u)\|_{L^2(\Omega)},$$

giving Equation (2-4). □

Corollary 2.15. *If Γ is C^2 and has a finite number of components, then any u in H with $\omega(u)$ in $L^2(\Omega)$ is also in Y, and*

$$\|\nabla u\|_{L^2(\Omega)} \le C(\|\omega(u)\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)}).$$

Proof. This follows from the basic estimate of elliptic regularity theory. □

Corollary 2.16. *When Γ is C^2 and has a finite number of components,*

$$\begin{aligned} X = Y, \quad X^2 &= Y^2 = H \cap W^2(\Omega), \\ X_0^2 &= Y_0^2 = \{u \in H \cap W^2(\Omega) : \omega(u) = 0 \text{ on } \Gamma\}. \end{aligned}$$

Proof. The first identity follows from Corollary 2.15 and the second and third from the identity $\Delta u = \nabla^\perp\omega$ and Lemma 2.7. □

We will find a need for the trace operator of Proposition 2.17 in Section 8.

Proposition 2.17. *Assume that Γ is C^2 and has a finite number of components, and let*

$$U = \{\omega \in L^2(\Omega) : \Delta\omega \in L^2(\Omega)\}$$

endowed with the norm $\|\omega\|_U = \|\omega\|_{L^2(\Omega)} + \|\Delta\omega\|_{L^2(\Omega)}$. There exists a linear continuous trace operator $\gamma_\omega : U \to W^{-1/2}(\Gamma)$ such that $\gamma_\omega\omega$ is the restriction of ω to Γ for all ω in $C^\infty(\overline{\Omega})$. For any α in $W_0^1(\Omega) \cap W^2(\Omega)$,

$$(\gamma_\omega\omega, \nabla\alpha \cdot \boldsymbol{n})_{W^{-1/2}(\Gamma), W^{1/2}(\Gamma)} = (\Delta\alpha, \omega) - (\alpha, \Delta\omega). \tag{2-5}$$

Lemma 2.18. *For any f in $L^2(\Omega)$ and a in $(W^{1/2}(\Gamma))^2$ satisfying the compatibility condition*

$$\int_\Omega f = \int_\Gamma a \cdot \boldsymbol{n},$$

there exists a (nonunique) solution v in $W^1(\Omega)$ to $\operatorname{div} v = f$ in Ω and $v = a$ on Γ.

Proof. This follows from [Galdi 1994, Lemma 3.2 on pages 126–127, Remark 3.3 on pages 128–129, and Exercise 3.4 on page 131]. See also the comment of [Adams 1975, page 67]. □

Lemma 2.19. *Define $\gamma_{\boldsymbol{\tau}} : Y \to L^2(\Gamma)$ by $\gamma_{\boldsymbol{\tau}} v = \gamma_0 v \cdot \boldsymbol{\tau}$ for any v in Y. When Γ is C^2, $\gamma_{\boldsymbol{\tau}}$ maps Y onto $W^{1/2}(\Gamma)$. When Γ is C^2 and has a finite number of components, $\gamma_{\boldsymbol{\tau}}(H_0 \cap Y)$ is dense in $W^{1/2}(\Gamma)$.*

Proof. Assume that Γ is C^2 and let g lie in $W^{1/2}(\Gamma)$. Then since Γ is C^2, $g\boldsymbol{\tau}$ is also in $W^{1/2}(\Gamma)$ by Corollary 2.3, and by Lemma 2.18 there exists a vector field v in $W^1(\Omega)$ with $\operatorname{div} v = \int_\Gamma g\boldsymbol{\tau} \cdot \boldsymbol{n} = 0$ and $v = g\boldsymbol{\tau}$ on Γ. Thus, in fact, v lies in Y, which shows that $\gamma_{\boldsymbol{\tau}}(Y)$ maps onto $W^{1/2}(\Gamma)$. If Γ has a finite number of components, then $H_c \cap Y$ is finite-dimensional and so is its image under this map; hence the image of $H_0 \cap Y$ is dense in $W^{1/2}(\Gamma)$. □

Proof of Proposition 2.17. Assume first that $\omega \in C^\infty(\overline{\Omega})$, let $\alpha \in W_0^1(\Omega) \cap W^2(\Omega)$, and let $v = \nabla^\perp \alpha$, so that v lies in $H_0 \cap Y$ with $\Delta\alpha = \omega(v)$. Then

$$\begin{aligned}(\alpha, \Delta\omega) &= -(\nabla\alpha, \nabla\omega) + \int_\Gamma (\nabla\overline{\omega} \cdot \boldsymbol{n})\alpha = -(\nabla\alpha, \nabla\omega)\\ &= (\Delta\alpha, \omega) - \int_\Gamma (\nabla\overline{\alpha} \cdot \boldsymbol{n})\omega = (\Delta\alpha, \omega) - \int_\Gamma \omega\overline{v} \cdot \boldsymbol{\tau}.\end{aligned}$$

From this it follows that for any choice of v (equivalently, by Lemma 2.10, of α) with a given value of $\overline{v} \cdot \boldsymbol{\tau}$ on Γ, the value of $(\Delta\alpha, \omega) - (\alpha, \Delta\omega)$ is the same.

Now, because of Lemma 2.19, we can define $\gamma_\omega(\omega)$ to be that unique element of $W^{-1/2}(\Gamma)$ such that Equation (2-5) holds. This gives a linear mapping from U to $W^{-1/2}(\Gamma)$ whose restriction to $C^\infty(\overline{\Omega})$ is the classical trace.

To establish the continuity of this mapping, let a be any element of $W^{1/2}(\Gamma)$. If Ω is simply connected, then $a = v \cdot \boldsymbol{\tau} = \nabla^\perp\alpha \cdot \boldsymbol{\tau} = \nabla\alpha \cdot \boldsymbol{n}$ for some v in Y or equivalently for some α in $W_0^1(\Omega) \cap W^2(\Omega)$. Then

$$\begin{aligned}(\gamma_\omega\omega, a)_{W^{-1/2}(\Gamma), W^{1/2}(\Gamma)} &= |(\Delta\alpha, \omega) - (\alpha, \Delta\omega)| \le C\|\Delta\alpha\|_{L^2(\Omega)}\|\omega\|_U\\ &\le C\|\nabla\alpha\|_{W^1(\Omega)}\|\omega\|_U \le C\|\nabla\alpha\|_{W^{1/2}(\Gamma)}\|\omega\|_U\\ &= C\|\nabla\alpha \cdot \boldsymbol{n}\|_{W^{1/2}(\Gamma)}\|\omega\|_U = C\|a\|_{W^{1/2}(\Gamma)}\|\omega\|_U.\end{aligned}$$

Here, we Lemma 2.7 in the first and second inequalities and the continuity of the inverse of the standard trace operator in the third inequality. Also, the second-to-last equality holds because α has the constant value of zero on Γ, so $\nabla\alpha \cdot \boldsymbol{\tau} = 0$ and $|\nabla\alpha| = |\nabla\alpha \cdot \boldsymbol{n}|$. This shows that the mapping is bounded and hence continuous.

When Ω is multiply connected, the argument is the same except that we must employ a simple density argument using Lemma 2.19. □

3. Strong formulations of three eigenvalue problems

Assume for the moment that Γ is C^2. Then, given any u in $V \cap W^2(\Omega)$, the (classical) Stokes operator A_S applied to u is that unique element $A_S u$ of H such that $\Delta u + A_S u = \nabla p$ for some harmonic pressure field p. Equivalently, $A_S = -\mathcal{P}\Delta$, $\mathcal{P}$ being the Leray projector defined following Lemma 2.13. The operator A_S maps $V \cap W^2(\Omega)$ onto H (see for instance [Foias et al. 2001, pages 49–50] for more details), is strictly positive definite, self-adjoint, and as a map from V to V^*, the composition of A_S^{-1} with the inclusion map of V into V^* is compact. It follows that $\{u_j\}$ is complete in H (and in V) with corresponding eigenvalues $\{\nu_j\}$ satisfying $0 < \nu_1 \le \nu_2 \le \cdots$ and $\nu_j \to \infty$ as $j \to \infty$. Also, the eigenfunctions are orthogonal in both H and V.

When Γ is only locally Lipschitz, $-\mathcal{P}\Delta$ is only known to be symmetric on $V \cap W^2(\Omega)$, not self-adjoint. Thus, we define A_S to be the Friedrichs extension, as an operator on H, of $-\Delta$ defined on $V \cap C_0^\infty(\Omega)$. A concrete description of its domain, $D(A_S)$, in terms of more familiar spaces is not known, though $V \cap H^2(\Omega) \subseteq D(A_S) \subseteq V$. In three dimensions, tighter inclusions have been obtained; see for instance [Brown and Shen 1995]. In any case, basic properties of the Friedrich extension insure that A_S is strictly positive definite, self-adjoint, and maps $D(A_S)$ bijectively onto H.

Definition 3.1. A strong eigenfunction $u_j \in V \cap X^2$ of A_S with eigenvalue $\nu_j > 0$ satisfies, for some p_j in $W^1(\Omega)$,

$$\begin{cases} \Delta u_j + \nu_j u_j = \nabla p_j, \quad \Delta p_j = 0, \quad \operatorname{div} u_j = 0 & \text{in } \Omega, \\ u_j = 0 & \text{on } \Gamma. \end{cases} \tag{3-1}$$

Taking the curl of (3-1), we see that the vorticity $\omega_j = \omega(u_j)$ satisfies

$$\begin{cases} \Delta\omega_j + \nu_j \omega_j = 0 & \text{in } \Omega, \\ u_j = 0 & \text{on } \Gamma. \end{cases} \tag{3-2}$$

That is, ω_j is an eigenfunction of the negative Laplacian, but with boundary conditions on the velocity u_j.

Let ψ_j be the stream function for u_j given by Lemma 2.10, so $u_j = \nabla^\perp \psi_j$. Then $\omega_j = \Delta\psi_j$ and $\nabla\psi_j = -u_j^\perp = 0$ on Γ. Since ψ_j is determined only up to a constant, we can then assume that $\psi_j = 0$ on Γ. Thus, ψ_j satisfies

$$\begin{cases} \Delta\Delta\psi_j + \nu_j \Delta\psi_j = 0 & \text{in } \Omega, \\ \nabla\psi_j \cdot \boldsymbol{n} = \psi_j = 0 & \text{on } \Gamma. \end{cases} \tag{3-3}$$

This is the eigenvalue problem for the clamped buckling plate; see for instance [Payne 1967; Ashbaugh 2004].

Temam exploits the similar correspondence between the Stokes problem and the biharmonic problem in the proof of [Temam 1984, Proposition I.2.3] to get a relatively simple proof of the regularity of solutions to the Stokes problem in two dimensions with at least C^2 regularity of the boundary. Also, as pointed out in

[Ashbaugh 2004], there is a similar correspondence between the eigenvalue problems for the Dirichlet Laplacian and (3-3) with the boundary condition $\nabla \psi_j \cdot \boldsymbol{n} = 0$ replaced by $\Delta \psi_j = 0$. We use this correspondence in the proof of Theorem 1.1, though we view the correspondence as being that given in Lemma 2.11, instead.

What we have shown is that given u_j satisfying (3-1), the corresponding stream function ψ_j satisfies (3-3). Conversely, given ψ_j satisfying (3-3), the functions $\omega_j = \Delta \psi_j$ and $u_j = \nabla^\perp \psi_j$ satisfy (3-2) and one can show, at least for sufficiently smooth boundaries, that u_j satisfies (3-1). Thus, the eigenvalue problems for the Stokes operator and the clamped buckling plate are equivalent.

Returning to (3-1), if we use instead the boundary conditions employed by J.-L. Lions [1969, pages 87–98] and P.-L. Lions [1996, pages 129–131], namely

$$u_j \cdot \boldsymbol{n} = 0 \quad \text{and} \quad \omega_j = 0 \text{ on } \Gamma, \tag{3-4}$$

which we call *Lions boundary conditions*, we obtain the eigenvalue problem for the Dirichlet Laplacian of Definition 3.2.

Definition 3.2. A strong eigenfunction $\omega_j \in W_0^1(\Omega)$ of the Dirichlet Laplacian $-\Delta_D$ with eigenvalue $\lambda_j > 0$ satisfies

$$\begin{cases} \Delta\omega_j + \lambda_j \omega_j = 0 & \text{in } \Omega, \\ \qquad\quad \omega_j = 0 & \text{on } \Gamma. \end{cases} \tag{3-5}$$

Using Lemma 3.4, we can recover the divergence-free velocity u_j in X_0^2 uniquely from a vorticity in $W_0^1(\Omega)$ under the constraint that $u_j \cdot \boldsymbol{n} = 0$, leading to the eigenvalue problem in Definition 3.3 for an operator A_L, which we will call the Stokes operator with Lions boundary conditions. (We use λ_j^* in place of λ_j because of the presence of zero eigenvalues.)

Definition 3.3. A strong eigenfunction $u_j \in X_0^2$ of A_L with eigenvalue $\lambda_j^* > 0$ satisfies

$$\begin{cases} \Delta u_j + \lambda_j^* u_j = 0, & \operatorname{div} u_j = 0 \quad \text{in } \Omega, \\ \qquad\; u_j \cdot \boldsymbol{n} = 0, & \omega(u_j) = 0 \quad \text{on } \Gamma. \end{cases} \tag{3-6}$$

What we have done is to define the eigenvalue problem for the operator A_L before defining the operator itself. In fact, $A_L \colon X_0^2 \to H$ with $A_L u = -\Delta u$. That is, A_L is simply the negative Laplacian on X_0^2.

To see that A_L is well defined, observe that $\Delta u \cdot \boldsymbol{n} = \nabla^\perp \omega(u) \cdot \boldsymbol{n} = -\nabla \omega(u) \cdot \boldsymbol{\tau} = 0$ for any u in X_0^2, since $\omega(u)$ is constant (namely, zero) along Γ. (Another way of viewing this is that there is no need for a Leray projector in X_0^2, making the Stokes operator on X_0^2 akin to the Stokes operator on $H \cap W^2(\Omega)$ for a periodic domain, which of course has no boundary. This is one reason that the use of the boundary conditions of (3-4) in [Lions 1969] and [Lions 1996] is so effective.)

Lemma 3.4. *Given ω in $W_0^1(\Omega)$ that satisfies*

$$\begin{cases} \Delta\omega + \lambda\omega = 0 & \textit{in } \Omega, \\ \omega = 0 & \textit{on } \Gamma \end{cases}$$

with $\lambda > 0$, there exists a unique u in X_0^2 such that $\omega = \omega(u)$ and

$$\begin{cases} \Delta u + \lambda u = 0, \quad \operatorname{div} u = 0 & \textit{in } \Omega, \\ u \cdot \boldsymbol{n} = 0, \quad \omega(u) = 0 & \textit{on } \Gamma. \end{cases}$$

Proof. Let $v = \omega^{-1}(\omega)$, which lies in $H_0 \cap X_0^2$ by Lemma 2.11. Then $\Delta v = \nabla^\perp\omega$ is in $L^2(\Omega)$, so $w = \Delta v + \lambda v$ is a divergence-free vector field in $L^2(\Omega)$. Hence, by Lemma 2.13, $w = h + \nabla p$ for a unique vector field h in H and an harmonic scalar field p in $W^1(\Omega)$ satisfying $\nabla p \cdot \boldsymbol{n} = w \cdot \boldsymbol{n} = \Delta v \cdot \boldsymbol{n}$ on Γ. (Since $\operatorname{div} \Delta v = 0$, Δv is in $E(\Omega)$, so $\Delta v \cdot \boldsymbol{n}$ is in $W^{-1/2}(\Gamma)$ by Lemma 2.9.)

But $\Delta v \cdot \boldsymbol{n} = \nabla^\perp\omega(v) \cdot \boldsymbol{n} = \nabla^\perp\omega \cdot \boldsymbol{n} = -\nabla\omega \cdot \boldsymbol{\tau} = 0$ on Γ, where ω has the constant value of zero. Thus, $\Delta p = 0$ in Ω with $\nabla p \cdot \boldsymbol{n} = 0$ on Γ, so $\nabla p \equiv 0$, and thus $w = h$ and so lies in H. Also, $\omega(w) = \Delta\omega(v) + \lambda\omega(v) = \Delta\omega + \lambda\omega = 0$.

Then $u = v - (1/\lambda)w$ is in H and using $\Delta w = \nabla^\perp\omega(w) = 0$, we see that

$$\Delta u + \lambda u = \Delta v + \lambda v - w = w - w = 0,$$

which gives the boundary value problem for u in the statement of the lemma. □

4. Weak formulations of the eigenvalue problems

To establish in Proposition 4.10 the existence of the eigenfunctions in Section 3, we work with their weak formulation, then show that these weak formulations are equivalent to those of Section 3 (for A_S, though, only when the boundary or the eigenfunctions are sufficiently regular). The formulations for A_S and A_L are modeled along the lines of the formulation in Definition 4.2 for the Dirichlet Laplacian, which is classical; see for instance [Henrot 2006, Chapter 1].

Definition 4.1. The vector field u_j in V is a weak eigenfunction of A_S with eigenvalue $\nu_j > 0$ if

$$(\omega(u_j), \omega(v)) - \nu_j(u_j, v) = 0 \quad \text{for all } v \in V.$$

Definition 4.2. The scalar field ω_j in $W_0^1(\Omega)$ is a weak eigenfunction for the Dirichlet Laplacian with eigenvalue $\lambda_j > 0$ if

$$(\nabla\omega_j, \nabla\alpha) - \lambda_j(\omega_j, \alpha) = 0 \quad \text{for all } \alpha \in W_0^1(\Omega).$$

Definition 4.3. The vector field u_j in $H_0 \cap X$ is a weak eigenfunction for A_L for $\lambda_j^* > 0$ if

$$(\omega(u_j), \omega(v)) - \lambda_j^*(u_j, v) = 0 \quad \text{for all } v \in H_0 \cap X. \tag{4-1}$$

Any vector in H_c is an eigenfunction of A_L with zero eigenvalue.

Proposition 4.4. *In Definition 4.3, the eigenfunction u_j for $\lambda_j^* > 0$ and the test function v can be taken to lie in X.*

Proof. Suppose we change Definition 4.3 to assume that u_j and the test function v lie in X. Then in particular,

$$(\omega(u_j), \omega(v)) - \lambda_j^*(u_j, v) = -\lambda_j^*(u_j, v) = 0 \quad \text{for all } v \in H_c.$$

That is, u_j is normal to any vector in H_c and so lies in $H_0 \cap X$. But then knowing that u_j lies in $H_0 \cap X$, it follows that $(\omega(u_j), \omega(v)) - \lambda_j^*(u_j, v) = 0$ for any v in H_c; that is, one need only use test functions in $H_0 \cap X$. Thus, the more stringent requirement for being a weak eigenfunction of A_L reduces to the less stringent requirement, meaning that the two are equivalent. □

Proposition 4.5. *A strong eigenfunction of A_S is a weak eigenfunction of A_S; a weak eigenfunction of A_S lying in X^2 is a strong eigenfunction of A_S.*

Proof. If u_j is a strong eigenfunction of A_S as in Definition 3.1, then applying Corollary A.1, we have for all v in V

$$(\omega(u_j),\ \omega(v)) - \nu_j(u_j, v) = -(\Delta u_j + \nu_j u_j, v) = -(\nabla p_j, v) = 0. \tag{4-2}$$

Thus, u_j is a weak eigenfunction of A_S as in Definition 4.1.

Conversely, suppose u_j is a weak eigenfunction of A_S as in Definition 4.1 such that $\omega(u_j)$ lies in $W^1(\Omega)$. Letting v lie in V, we have $(\omega(u_j), \omega(v)) - \nu_j(u_j, v) = 0$, and u_j and v have sufficient regularity to apply Corollary A.1 as above to give $(\Delta u_j + \nu_j u, v) = 0$ for all v in V. From Lemma 2.9 we see that $\Delta u_j + \nu_j u = \nabla p_j$ for some harmonic pressure field p_j in $W^1(\Omega)$, since $\Delta u_j + \nu_j u$ is in $L^2(\Omega)$. This shows that u_j is a strong eigenfunction of A_S as in Definition 3.1. □

Proposition 4.6. *Definitions 3.2 and 4.2 are equivalent as, too, are Definitions 3.3 and 4.3. When Γ is C^2, Definitions 3.1 and 4.1 are equivalent.*

Proof. If u_j is a strong eigenfunction of A_L as in Definition 3.3, then by virtue of Corollary A.1, we have for all v in $W^1(\Omega)$

$$\begin{aligned}(\omega(u_j), \omega(v)) - \lambda_j^*(u_j, v) &= -(\Delta u_j, v) + \int_\Gamma \omega(u_j)\bar{v} \cdot \boldsymbol{\tau} - \lambda_j^*(u_j, v) \\ &= -(\Delta u_j + \lambda_j^* u_j, v) = 0.\end{aligned} \tag{4-3}$$

It follows that u_j is a weak eigenfunction of A_L as in Definition 4.3.

Now suppose that u_j is a weak eigenfunction of A_L as in Definition 4.3. Let ψ_j be the stream function for u_j lying in $W_0^1(\Omega)$ given by Lemma 2.10. Then for all v in X,

$$\begin{aligned}(u_j, v) &= (\nabla^\perp \psi_j, v) = -(\nabla \psi_j, v^\perp) = (\psi_j, \operatorname{div} v^\perp) - \int_\Gamma (v^\perp \cdot \boldsymbol{n}) \psi_j \\ &= -(\psi_j, \omega(v)).\end{aligned}$$

Hence, by virtue of Proposition 4.4, we have for all v in X

$$(\omega(u_j) + \lambda_j^* \psi_j, \omega(v)) = (\Delta \psi_j + \lambda_j^* \psi_j, \omega(v)) = 0.$$

Then $\Delta \psi_j + \lambda_j^* \psi_j = 0$ since by Lemma 2.11 $\omega(v)$ ranges over all of $L^2(\Omega)$, so $\omega_j = -\lambda_j^* \psi_j$ lies in $W_0^1(\Omega)$. Thus, $\Delta u_j = \nabla^\perp \omega_j$ is in $L^2(\Omega)$, so u_j is a strong eigenfunction of A_L as in Definition 3.3.

A strong eigenfunction of A_S is a weak eigenfunction of A_S by Proposition 4.5.

Suppose that u_j is a weak eigenfunction of A_S as in Definition 4.1 and that Γ is C^2. Let v lie in $\mathcal{V}$. Then

$$\begin{aligned}(\omega(u_j), \omega(v)) &= -(\omega(u_j), \operatorname{div} v^\perp) = (\nabla \omega(u_j), v^\perp) = -(\nabla^\perp \omega(u_j), v) \\ &= -(\Delta u_j, v).\end{aligned}$$

Hence $(\Delta u_j + \nu_j u_j, v) = 0$ for all $v \in \mathcal{V}$, so by Lemma 2.9

$$\Delta u_j + \nu_j u_j = \nabla p_j \quad \text{for some } p_j \text{ in } \mathcal{D}'(\Omega). \tag{4-4}$$

Now, by [Temam 1984, Proposition I.2.3], there exists w in $V \cap W^2(\Omega)$ and q in $W^1(\Omega)$ satisfying $\Delta w + \nu_j u_j = \nabla q$. (Only here do we require Γ to be C^2.)

Define the bilinear form a on $V \times V$ by $a(u, v) = (\omega(u), \omega(v))$. Then by Corollary A.3, $a(u, v) = (\nabla u, \nabla v)$, so $a(u, u) = \|u\|_V^2$, and we can apply the Lax–Milgram theorem to conclude that $w = u_j$. Hence, u_j is in $V \cap W^2(\Omega)$, showing that it is a strong eigenfunction of A_S.

That a strong eigenfunction of $-\Delta_D$ is weak is classical. It is also classical that for a weak eigenfunction, ω_j is in $C^\infty(\Omega)$, which is enough to conclude that $\Delta \omega_j$ is in $L^2(\Omega)$. □

Remark 4.7. When Γ is C^2, in fact the eigenfunctions of A_L and A_S lie in $W^2(\Omega)$, as can seen for A_L by the proof of Proposition 4.6 and for A_S by, for instance, [Temam 1984, Proposition I.2.3].

Proposition 4.8. *There exists a bijection between the strong eigenfunctions of A_L having positive eigenvalues and the weak eigenfunctions of the Dirichlet Laplacian, with a corresponding bijection between the eigenvalues.*

Proof. By Lemma 2.11 for any u in $H_0 \cap X_0^2$, there exists $\omega(u)$ in $W_0^1(\Omega)$, and this gives a bijection between the spaces. Also by Lemma 2.11 and its proof, for any v

in $H_0 \cap X_0^2$ there exists $\omega(v)$ in $W_0^1(\Omega)$, and associated to v is its stream function ψ in $W_0^1(\Omega)$ with $\Delta\psi = \omega(v)$. With u, v, and ψ as above,

$$\begin{aligned}\frac{(\nabla\omega, \nabla\psi)}{(\omega, \psi)} &= \frac{-(\omega, \Delta\psi) + \int_\Gamma (\nabla\psi \cdot \boldsymbol{n})\omega}{-(\operatorname{div} u^\perp, \psi)} \\ &= \frac{-(\omega(u), \omega(v))}{(u^\perp, \nabla\psi) - \int_\Gamma (u^\perp \cdot \boldsymbol{n})\psi} = \frac{-(\omega(u), \omega(v))}{-(u, \nabla^\perp\psi)} = \frac{(\omega(u), \omega(v))}{(u, v)}.\end{aligned}$$

We applied Lemma 2.4 twice, the first time using ω in $W_0^1(\Omega)$ with $\nabla\psi$ in $E(\Omega)$ and the second time using ψ in $W_0^1(\Omega)$ with $u^\perp$ in $E(\Omega)$.

By the bijections above, this shows that if ω is a weak eigenfunction of $-\Delta_D$, then $u = \omega^{-1}(\omega)$ is a weak eigenfunction of A_L (also using Corollary 2.12) that lies in X_0^2, and hence is a strong eigenfunction of the A_L by Proposition 4.6. The converse follows from the same equality. □

Corollary 4.9. *There exists a bijection between the weak eigenfunctions of A_L having positive eigenvalues and the weak eigenfunctions of the Dirichlet Laplacian, with a corresponding bijection between the eigenvalues: $\lambda_k^* = \lambda_k$ for all k.*

Proof. Combine Propositions 4.6 and 4.8. □

Proposition 4.10. *There exists a sequence of weak eigenfunctions for each of our three eigenvalue problems with spectra increasing to infinity as in Equation* (1-1) *for $-\Delta_D$ and A_S and with*

$$\sigma_L = \{\lambda_j\}_{j=1}^\infty, \quad \textit{where } 0 < \lambda_1 < \lambda_2 \le \cdots.$$

If Ω is multiply connected, σ_L will also include 0. *The eigenfunctions of $-\Delta_D$ form an orthonormal basis of both $L^2(\Omega)$ and $W_0^1(\Omega)$, while those of A_S form an orthogonal basis of both H and V. The eigenfunctions of A_L lie in $C^\infty(\Omega) \cap X_0^2$ and form an orthogonal basis of both H and X. The eigenfunctions of $-\Delta_D$ are in $C^\infty(\Omega) \cap W^2(\Omega)$.*

Proof. To prove the existence of eigenfunctions of A_S, let G be the inverse of A_S. Let u and v be in H. Since A_S is a bijection from $D(A_S)$ onto H, there exists w in $D(A_S)$ such that $v = A_S w$ and $w = Gv$. Then because A_S is self-adjoint,

$$(Gu, v) = (Gu, A_S w) = (A_S Gu, w) = (u, w) = (u, Gv),$$

showing that G is symmetric and hence, being defined on all of H, self-adjoint. The calculation above also shows that $(Gu, u) = (A_S Gu, Gu) = \|\nabla Gu\|_{L^2(\Omega)}^2$, which is positive for all nonzero u in H.

But V is compactly embedded in H by Lemma 2.6, so G, viewed as a map from H to H, is compact. Therefore, G is a compact, positive, self-adjoint operator. The spectral theorem thus gives a complete set of eigenfunctions in H and a discrete set of eigenvalues decreasing to zero; applying G to these eigenfunctions and using the reciprocal of the eigenvalues gives the eigenfunctions and eigenvalues of A_S in the usual way.

The results for $-\Delta_D$ are classical; those for A_L then follow from Corollary 4.9 or they can be proved directly using an argument similar to that above. □

Remark 4.11. Because the strong form $\Delta u_j + \lambda_j^* u_j = \nabla p_j$ of the eigenvalue problem for A_S has a nonzero pressure, the classical interior regularity argument for $-\Delta_D$ cannot be made for A_S. To obtain further regularity, one must assume a more regular boundary.

5. Min-max formulations of the eigenvalue problems

Proposition 5.1. *Let*

$$\begin{aligned} S_k &= \text{the span of the first } k \text{ eigenfunctions of } A_S, \\ L_k &= \text{the span of the first } k \text{ eigenfunctions of } A_L, \\ D_k &= \text{the span of the first } k \text{ eigenfunctions of } -\Delta_D, \end{aligned}$$

with $S_0 = L_0 = D_0 = \{0\}$. *Then*

$$\begin{aligned} \nu_k &= \min\{R_S(u) \colon u \in S_{k-1}^\perp \cap V \setminus \{0\}\}, \\ \lambda_k &= \min\{R_D(\omega) \colon \omega \in D_{k-1}^\perp \cap W_0^1(\Omega) \setminus \{0\}\} \\ &= \min\{R_L(u) \colon u \in L_{k-1}^\perp \cap H_0 \cap X \setminus \{0\}\} \\ &= \min\{R_L(u) \colon u \in L_{k-1}^\perp \cap H_0 \cap X_0^2 \setminus \{0\}\}, \end{aligned}$$

where the Rayleigh quotients are

$$R_S(u) = R_L(u) = \|\omega(u)\|^2_{L^2(\Omega)} / \|u\|^2_{L^2(\Omega)}, \quad R_D(\omega) = \|\nabla\omega\|^2_{L^2(\Omega)} / \|\omega\|^2_{L^2(\Omega)}.$$

Proof. The form of the Rayleigh coefficient for ν_k and the form in the first two expressions for λ_k come from the weak formulations of the eigenvalue problems in Definitions 4.1–4.3. The third expression for λ_k follows from the bijection in Lemma 2.11 and by noting that if u is any element of X_0^2, then $R_L(u) = R_D(\omega(u))$, as in the proof of Proposition 4.8. □

Defining four functions mapping $\mathbb{R}$ to $\mathbb{Z}$ by

$$\begin{aligned} N_S(\lambda) &= \#\{j \in \mathbb{N} \colon \nu_j < \lambda\}, \quad & N_L(\lambda) &= \#\{j \in \mathbb{N} \colon \lambda_j < \lambda\}, \\ \overline{N}_S(\lambda) &= \#\{j \in \mathbb{N} \colon \nu_j \le \lambda\}, \quad & \overline{N}_L(\lambda) &= \#\{j \in \mathbb{N} \colon \lambda_j \le \lambda\}, \end{aligned}$$

we have an immediate corollary of Proposition 5.1:

Corollary 5.2. $\overline{N}_S(\lambda) = \max_{Z \subseteq V}\{\dim Z : R_S(u) \le \lambda \textit{ for all } u \in Z\}$,

$$\begin{aligned}\overline{N}_L(\lambda) &= \max_{Z \subseteq H_0 \cap X_0^2}\{\dim Z : R_L(u) \le \lambda \textit{ for all } u \in Z\}\\ &= \max_{Z \subseteq H_0 \cap X}\{\dim Z : R_L(u) \le \lambda \textit{ for all } u \in Z\}.\end{aligned}$$

Remark 5.3. By Corollary A.3, $R_S(u) = \|\nabla u\|^2_{L^2(\Omega)}/\|u\|^2_{L^2(\Omega)}$, so $\lambda_k \le \nu_k$ follows from Corollary 5.2. Strict inequality, however, is not so immediate.

6. Proof of Theorem 1.1

Lemma 6.1 is the analogue of the (only) lemma in [Filonov 2004] and, in fact, follows from it. For completeness we give the full proof.

Lemma 6.1. *For all λ in $\mathbb{R}$,*

$$V \cap \ker\{A_L - \lambda\} \cap X_0^2 = \{0\}.$$

Proof. Let u be in $V \cap \ker\{A_L - \lambda\} \cap X_0^2 = \ker\{A_S - \lambda\} \cap X_0^2$, where we used Proposition 4.5. Then

$$\begin{cases} \Delta u + \lambda u = \nabla p, \quad \operatorname{div} u = 0, \quad \Delta\omega + \lambda\omega = 0 & \text{in } \Omega,\\ \qquad u = 0, \qquad\qquad \omega = 0, & \text{on } \Gamma.\end{cases}$$

Because $\omega = 0$ on Γ, $\nabla p = 0$ on Ω by Lemma 3.4. Hence, $\nabla\omega = -(\Delta u)^\perp = \lambda u^\perp = 0$ on Γ. Thus, ω extended by 0 to all of $\mathbb{R}^2$ lies in $W^1(\mathbb{R}^2)$. Then for all ψ in $\mathcal{S}(\mathbb{R}^2)$,

$$\begin{aligned}(-\Delta\omega, \psi)_{\mathcal{S}'(\mathbb{R}^2),\mathcal{S}(\mathbb{R}^2)} = (\nabla\omega, \nabla\psi)_{\mathcal{S}'(\mathbb{R}^2),\mathcal{S}(\mathbb{R}^2)} &= \int_{\mathbb{R}^2} \nabla\omega \cdot \nabla\overline{\psi}\\ = \int_\Omega \nabla\omega\cdot\nabla\overline{\psi} &= -\int_\Omega \Delta\omega\overline{\psi} + \int_\Gamma (\nabla\omega\cdot\boldsymbol{n})\overline{\psi}\\ = \lambda\int_\Omega \omega\overline{\psi} &= \lambda\int_{\mathbb{R}^2}\omega\overline{\psi} = (\lambda\omega, \psi)_{\mathcal{S}'(\mathbb{R}^2),\mathcal{S}(\mathbb{R}^2)},\end{aligned}$$

which shows that $\Delta\omega = -\lambda\omega$ as distributions. But ω is in $W^1(\mathbb{R}^2)$ so, in fact, $\Delta\omega$ is in $W^1(\mathbb{R}^2)$ and $\Delta\omega + \lambda\omega = 0$ on $\mathbb{R}^2$. Moreover, ω vanishes outside of Ω. But the Laplacian is hypoelliptic so ω is real analytic and hence vanishes on all of $\mathbb{R}^2$.

Now, were Ω simply connected it would follow immediately that $u \equiv 0$. In any case, observe that $\omega \equiv 0$ implies $\Delta u = \nabla^\perp\omega \equiv 0$. But $\Delta u = -\lambda u$, so $u \equiv 0$. □

Proof of Theorem 1.1. Let $\lambda > 0$ and choose a subspace F of V of dimension $\overline{N}_S(\lambda)$ with

$$\|\omega(u)\|^2_{L^2(\Omega)} \le \lambda\|u\|^2_{L^2(\Omega)} \quad \text{for all } u \in F. \tag{6-1}$$

This is possible by the variational formulation of the eigenvalue problem for A_S in Corollary 5.2. By Lemma 6.1,

$$G = F \oplus (\ker\{A_L - \lambda\} \cap X_0^2)$$

is a direct sum and so has dimension $\overline{N}_S(\lambda) + \dim\ker\{-\Delta_D - \lambda\}$, where we used Propositions 4.5 and 4.8. (Either of the vector spaces above could contain only 0.)

For any $u \in F$ and $v \in \ker\{A_L - \lambda\} \cap X_0^2$,

$$\begin{aligned}\|\omega(u+v)\|^2_{L^2(\Omega)} &= \|\omega(u)\|^2_{L^2(\Omega)} + \|\omega(v)\|^2_{L^2(\Omega)} + 2\operatorname{Re}(\omega(u), \omega(v)) \\ &= \|\omega(u)\|^2_{L^2(\Omega)} + \|\omega(v)\|^2_{L^2(\Omega)} + 2\lambda \operatorname{Re}(u, v),\end{aligned}$$

because $(\omega(u), \omega(v)) = \lambda(u, v)$ by Definition 4.3.

Also by Definition 4.3,

$$\|\omega(v)\|^2_{L^2(\Omega)} = \lambda\|v\|^2_{L^2(\Omega)},$$

and combined with Equation (6-1) this gives

$$\|\omega(u+v)\|^2_{L^2(\Omega)} \le \lambda\|u\|^2_{L^2(\Omega)} + \lambda\|v\|^2_{L^2(\Omega)} + 2\lambda \operatorname{Re}(u, v) = \lambda\|u+v\|^2_{L^2(\Omega)}.$$

Then it follows by the variational formulation of the eigenvalue problem for A_L in Corollary 5.2 that

$$\overline{N}_L(\lambda) \ge \dim G = \overline{N}_S(\lambda) + \dim\ker\{-\Delta_D - \lambda\},$$

so

$$N_L(\lambda) = \overline{N}_L(\lambda) - \dim\ker\{-\Delta_D - \lambda\} \ge \overline{N}_S(\lambda).$$

Setting $\lambda = \nu_k$ gives $N_L(\nu_k) \ge \overline{N}_S(\nu_k) \ge k$. In words, there are at least k eigenvalues in σ_D (counted according to multiplicity) strictly less than ν_k; that is, $\lambda_k < \nu_k$. □

7. Toward the inequality $\boldsymbol{\lambda_{k+1} \le \nu_k}$

Theorem 7.1. *For each k in $\mathbb{N}$, define $U_R^k = (\nu_k, x)$, where x is the smallest element of $(\sigma_S \cup \sigma_D) \cap (\nu_k, \infty)$, and define $U_L^k = (y, \lambda_k)$, where y is the largest element of $(\sigma_S \cup \sigma_D) \cap (-\infty, \lambda_k)$. (Let $y = -\infty$ if $k = 1$.) Suppose that for some λ in U_R^k there exists a nonzero vector field w in X^2 and a scalar field q in $W^1(\Omega)$ satisfying the underdetermined problem*

$$\begin{cases} \Delta w + \lambda w = \nabla q, \quad \operatorname{div} w = 0 & \text{on } \Omega, \\ \qquad w \cdot \boldsymbol{n} = 0 & \text{on } \Gamma, \end{cases} \tag{7-1}$$

but with the constraint

$$\int_\Gamma \omega(w)\overline{w} \cdot \boldsymbol{\tau} = \|\omega(w)\|^2_{L^2(\Omega)} - \lambda\|w\|^2_{L^2(\Omega)} \le 0. \tag{7-2}$$

Then $\lambda_{k+1} \le \nu_k$. If for each k there exist λ in U_L^k a nonzero vector field w in X^2 and a scalar field q in $W^1(\Omega)$ satisfying (7-1) *and* (7-2), *then $\lambda_{k+1} \le \nu_k$ for all k.*

Proof. Observe first that $\int_\Gamma \omega(w)\overline{w}\cdot\boldsymbol{\tau} = \|\omega(w)\|^2_{L^2(\Omega)} - \lambda\|w\|^2_{L^2(\Omega)}$ follows from Corollary A.1.

Assume that λ in U_R^k and w and q are as in (7-1) and (7-2). Let the set F be defined as in the proof of Lemma 6.1, but let $G = F \oplus \operatorname{span}\{w\}$. This is a direct sum since otherwise w would be in span F, meaning that it would vanish on Γ and so would actually be an eigenfunction of A_S; but this is impossible since λ is not in σ_S by assumption. The dimension of G is $\overline{N}_S(\lambda)+1$.

Then for any u in F and c in $\mathbb{C}$,

$$\|\omega(u+cw)\|^2_{L^2} = \|\omega(u)\|^2_{L^2} + \|\omega(cw)\|^2_{L^2} + 2\operatorname{Re}(\omega(u), \omega(cw)).$$

But by Corollary A.1,

$$(\omega(u), \omega(w)) = -(\Delta w, u) = (\lambda w, u) - (\nabla q, u) = \lambda(u, w)$$

and $\|\omega(w)\|^2_{L^2} \le \lambda\|w\|^2_{L^2}$ by (7-2). Also, $\|\omega(u)\|^2_{L^2} \le \lambda\|u\|^2_{L^2}$, so we can conclude that

$$\|\omega(u+cw)\|^2_{L^2} \le \lambda\|u\|^2_{L^2} + \lambda\|cw\|^2_{L^2} + 2\lambda\operatorname{Re}(u, cw) = \lambda\|u+cw\|^2_{L^2}.$$

Then it follows by the variational formulation of the eigenvalue problem for A_L in Corollary 5.2 that $\overline{N}_L(\lambda) \ge \dim G = \overline{N}_S(\lambda)+1$.

Because λ is larger than ν_k but smaller than any eigenvalue in $(\sigma_D \cup \sigma_S) \cap (\lambda, \infty)$, $N_L(\lambda) = \overline{N}_L(\nu_k)$ and $\overline{N}_S(\lambda) = \overline{N}_S(\nu_k)$, so $\overline{N}_L(\nu_k) \ge \overline{N}_S(\nu_k)+1 \ge k+1$. In other words, there are at least $k+1$ eigenvalues in σ_D (counted according to multiplicity) less than or equal to ν_k; that is, $\lambda_{k+1} \le \nu_k$. This establishes the result for λ in U_R^k.

Now assume that for all k there exists a λ in U_L^k with w and q as in (7-1) and (7-2). Given j in $\mathbb{N}$, let δ be the lowest eigenvalue greater than ν_j in $\sigma_S \cup \sigma_D$. If δ is in σ_S, then $\delta = \nu_n$ for some $n > j$, and if $\lambda_{n+1} \le \nu_n$ then it will follow that $\lambda_{j+1} \le \nu_j$ since there are no eigenvalues in σ_D between ν_j and ν_n (though ν_j, ν_n, or both might also be in σ_D). We can continue this line of reasoning until eventually we reach a value of j such that the next lowest eigenvalue δ in $\sigma_S \cup \sigma_D$ is in σ_D (δ might also be in σ_S, but this will not affect our argument). Then $\delta = \lambda_n$ for some n in $\mathbb{N}$.

Then by assumption there is some λ in U_L^n with w and q as in (7-1) and (7-2). But this λ is also in U_R^j, so we conclude that $\lambda_{j+1} \le \nu_j$, and from our argument above, this inequality holds, then, for all j in $\mathbb{N}$. □

Remark 7.2. For λ in σ_D, even if a w exists satisfying the conditions in (7-1) and (7-2), w might be an eigenfunction of A_L and so lie in $\ker\{A_L - \lambda\}$. This means that we cannot extend the argument along the lines in the proof of Theorem 1.1, since span$\{w\}$ might not be linearly independent of the set G in the proof of that theorem. This prevents us from concluding that $\lambda_{k+1} < \nu_k$ for all k, which is in any case not true in general.

The difficulty with applying Theorem 7.1 is that it is relatively easy to find vector fields w satisfying the given conditions in a left neighborhood of ν_k, or perhaps in a right neighborhood of λ_k, but hard to find ones in the required neighborhoods. We give an example in Section 8.

8. Proof of Theorem 1.2 and related issues

Navier slip boundary conditions for the Stokes operator provide a physically justifiable alternative to the classical no-slip boundary conditions used to define A_S. To the extent possible, we will work with these boundary conditions with a locally Lipschitz boundary, but we will find that they are really only of use when the boundary is C^2 and has a finite number of components. (Observe that under this assumption, by Corollary 2.16, the distinctions we have been making between the X spaces and the Y spaces disappear.)

To define Navier boundary conditions in the classical sense, we must assume that Γ is C^2. (Here, as elsewhere in this paper, $C^{1,1}$ would suffice, but introduces added complexities we wish to avoid.) The Navier conditions can be written in the form

$$\omega(u) = (2\kappa - \alpha)u \cdot \boldsymbol{\tau} \quad \text{on } \Gamma, \tag{8-1}$$

where κ is the curvature of the boundary and α is any function in $L^\infty(\Gamma)$.

If u in $H \cap W^2(\Omega)$ satisfies Equation (8-1) then by Corollary A.1,

$$(-\Delta u, v) = (\omega(u), \omega(v)) - \int_\Gamma (2\kappa - \alpha)u \cdot \bar{v} \quad \text{for any } v \text{ in } X.$$

Let $H_V = \{u \in H \cap W^2(\Omega) : \omega(u) = (2\kappa - \alpha)u \cdot \boldsymbol{\tau} \text{ on } \Gamma\}$, endowed with the same norm as Y. We define the operator $A_V : Y \to H$ by requiring that

$$(A_V u, v) = (\omega(u), \omega(v)) + \int_\Gamma (\alpha - 2\kappa)u \cdot \bar{v} = (\nabla u, \nabla v) + \int_\Gamma (\alpha - \kappa)u \cdot \bar{v}, \tag{8-2}$$

for all v in Y. The second equality (which gives the form of the operator A defined on [Kelliher 2006, page 218]) follows from Lemma A.2, Lemma A.4, and the density of $(C^1(\Omega))^2$ in Y.

Now assume that Ω is bounded and Γ is locally Lipschitz. Then the curvature is no longer defined, so we replace the function $\alpha - 2\kappa$ with a function f lying in $L^\infty(\Gamma)$, though we lose in this way the physical meaning. In place of (8-1), we have

$$\omega(u) + fu \cdot \boldsymbol{\tau} = 0 \quad \text{on } \Gamma, \tag{8-3}$$

$$(A_V u, v) = (\omega(u), \omega(v)) + \int_\Gamma fu \cdot \bar{v}. \tag{8-4}$$

Observe that the second expression for A_V in (8-2) now has insufficient regularity, so it no longer applies.

Definition 8.1. A vector field $u_j \in X^2$ is a strong eigenfunction of A_V with eigenvalue γ_j if

$$\begin{cases} \Delta u_j + \gamma_j u_j = \nabla p_j, \quad \Delta p_j = 0, \quad \operatorname{div} u_j = 0 \quad \text{in } \Omega, \\ \qquad u_j \cdot \boldsymbol{n} = 0, \qquad \omega(u_j) + f u_j \cdot \boldsymbol{\tau} = 0 \quad \text{on } \Gamma. \end{cases}$$

Definition 8.2. The vector field u_j in X is a weak eigenfunction of A_V with eigenvalue γ_j if

$$(\omega(u_j),\ \omega(v)) + \int_\Gamma f u_j \cdot \bar{v} - \gamma_j (u_j, v) = 0 \quad \text{for all } v \in X.$$

Proposition 8.3. *If u_j is a strong eigenfunction of A_V, then it is a weak eigenfunction of A_V. If u_j is a weak eigenfunction of A_V that happens to be in X^2 and satisfy $\omega(u_j) + f u_j \cdot \boldsymbol{\tau} = 0$ on Γ, then u_j is a strong eigenfunction of A_V.*

Proof. Strong implies weak follows by the integration by parts performed above. For the reverse implication, assume that u_j is a weak eigenfunction of A_V lying in X^2. Then choosing v to lie in V, it follows that

$$(\omega(u_j),\ \omega(v)) - \gamma_j (u_j, v) = 0 \quad \text{for all } v \in V.$$

Applying Corollary A.1 gives $(\Delta u_j + \gamma_j u_j, v) = 0$ for all $v \in V$, and we conclude that $\Delta u_j + \gamma_j u_j = \nabla p_j$ for some harmonic field p in $W^1(\Omega)$ by Lemma 2.9. □

When Γ is C^2 and has a finite number of components, we can consider the special case $\alpha = \kappa$, which gives $\omega(u_j) = \kappa u_j \cdot \boldsymbol{\tau}$. It follows from Lemma A.5 that $\nabla u_j \boldsymbol{n} \cdot \bar{v} = 0$ for any v in X. More simply, we can write this as $\nabla u_j \boldsymbol{n} \cdot \boldsymbol{\tau} = 0$. These boundary conditions imply that $(-\Delta u_j, v) = (\nabla u_j, \nabla v)$ for all v in X, (or we can take advantage of the second form of $(A_V u, u)$ in (8-2)), and we can explicitly define such eigenfunctions as follows, though we need no longer assume that the boundary is C^2:

Definition 8.4. A vector field $u_j \in X^2$ is a strong eigenfunction of A_N if

$$\begin{cases} \Delta u_j + \beta_j u_j = \nabla p_j, \qquad \Delta p_j = 0, \quad \operatorname{div} u_j = 0 \quad \text{in } \Omega, \\ \qquad u_j \cdot \boldsymbol{n} = 0, \quad \nabla u_j \boldsymbol{n} \cdot \boldsymbol{\tau} = 0 \qquad\qquad \text{on } \Gamma. \end{cases}$$

Definition 8.5. A vector field u_j in X is a weak eigenfunction of A_N if

$$(\nabla u_j, \nabla v) - \beta_j(u_j, v) = 0 \quad \text{for all } v \in X.$$

We also have the following min-max formulations for the eigenvalues of A_V and the special case of A_N.

Proposition 8.6. *Let*

$$V_k = \textit{the span of the first keigenfunctions of } A_V,$$
$$N_k = \textit{the span of the first keigenfunctions of } A_N,$$

with $V_0 = N_0 = \{0\}$. *Then*

$$\gamma_k = \min\{R_V(u) \colon u \in V_{k-1}^{\perp} \cap X \setminus \{0\}\},$$
$$\beta_k = \min\{R_N(u) \colon u \in N_{k-1}^{\perp} \cap X \setminus \{0\}\},$$

where

$$R_V(u) = \frac{\|\omega(u)\|_{L^2(\Omega)}^2 + \int_\Gamma f|u|^2}{\|u\|_{L^2(\Omega)}^2}, \quad R_N(u) = \frac{\|\nabla u\|_{L^2(\Omega)}^2}{\|u\|_{L^2(\Omega)}^2}.$$

The eigenvalues are real with $0 = \beta_1 \le \beta_2 \le \cdots$ *and, when* f *is nonnegative,* $0 < \gamma_1 \le \gamma_2 \le \cdots$ *with* $\gamma_k \to \infty$.

Proof. Define the operator $T \colon X \to X$ by $T = (iI + A_V)^{-1} \circ j$, where I is the identity map, j is the inclusion map from X to X^* (which is compact by Corollary 2.12), and $i = \sqrt{-1}$. Then since $(iI + A_V)^{-1}$ is bounded (its norm can be no greater than 1) T is compact, and the spectral theorem provides us with eigenvalues of T accumulating at zero. To each eigenvalue λ of T there corresponds an eigenvalue $\gamma = \mu^{-1} - i$ of A_V. But A_V is self-adjoint, so γ is real. And when f is nonnegative, since $R_V(u)$ is nonnegative, $0 < \gamma_1 \le \gamma_2 \le \cdots$ with $\gamma_k \to \infty$. $\square$

Defining two functions mapping $\mathbb{R}$ to $\mathbb{Z}$ by

$$\overline{N}_V(\lambda) = \#\{j \in \mathbb{N} \colon \gamma_j \le \lambda\} \quad \text{and} \quad \overline{N}_N(\lambda) = \#\{j \in \mathbb{N} \colon \beta_j \le \lambda\},$$

we have an immediate corollary of Proposition 8.6:

Corollary 8.7. $\overline{N}_V(\lambda) = \max_{Z \subseteq X}\{\dim Z \colon R_V(u) \le \lambda \textit{ for all } u \in Z\},$

$$\overline{N}_N(\lambda) = \max_{Z \subseteq X}\{\dim Z \colon R_N(u) \le \lambda \textit{ for all } u \in Z\}.$$

Proposition 8.8. *Assume* Γ *is* C^2 *and has a finite number of components and*

$$f \in C^{1/2+\epsilon}(\Gamma) + W^{1/2+\epsilon}(\Gamma). \tag{8-5}$$

A weak eigenfunction of A_V *is a strong eigenfunction of* A_V. *In particular, a weak eigenfunction* u_j *of* A_V *satisfies* $\omega(u_j) + f u_j \cdot \boldsymbol{\tau} = 0$ *on* Γ.

Proof. Suppose that u is a weak eigenfunction of A_V as in Definition 8.2 with $\omega = \omega(u)$. Then for any v in $\mathcal{V}$ integration by parts gives $(\Delta u + \lambda u, v) = 0$, so $\Delta u + \lambda u = \nabla p$ by Lemma 2.9, equality holding in terms of distributions. Taking the curl, it follows that $\Delta\omega = -\lambda\omega$, so ω is in U of Proposition 2.17, since ω is in L^2. Thus, by Proposition 2.17, ω is well defined on Γ as an element of $W^{-1/2}(\Gamma)$.

Let v be any vector in $H_0 \cap Y$, and let α be its associated stream function lying in $W_0^1(\Omega) \cap W^2(\Omega)$ given by Lemma 2.10, so that $\Delta\alpha = \omega(v)$ is in $L^2(\Omega)$. Thus, again by Proposition 2.17, since $\nabla\alpha \cdot \boldsymbol{n} = -v \cdot \boldsymbol{\tau}$,

$$\begin{aligned}(\gamma_\omega \omega, v \cdot \boldsymbol{\tau})_{W^{-1/2}(\Gamma), W^{1/2}(\Gamma)} &= (\alpha, \Delta\omega) - (\omega(v), \omega)\\ &= -\lambda(\alpha, \omega) - (\omega(v), \omega) = \lambda(u, v) - (\omega(v), \omega).\end{aligned}$$

Here we used

$$(\alpha, \omega) = -(\alpha, \operatorname{div} u^\perp) = (\nabla\alpha, u^\perp) + \int_\Gamma (u^\perp \cdot \boldsymbol{n})\alpha = -(v, u),$$

noting that we had enough regularity to apply Corollary A.1.

But because u is a weak eigenfunction of A_V, also

$$(f u \cdot \boldsymbol{\tau}, v \cdot \boldsymbol{\tau})_{W^{-1/2}(\Gamma), W^{1/2}(\Gamma)} = \lambda(u, v) - (\omega(v), \omega).$$

Thus, the two boundary integrals are equal, and because of Lemma 2.19, we can conclude that $\omega = -f u \cdot \boldsymbol{\tau}$ on Γ, and in particular that ω is in $W^{1/2}(\Gamma)$. (By Corollaries 2.2 and 2.3 and (8-5) we know $f u \cdot \boldsymbol{\tau}$ is in $W^{1/2}(\Gamma)$.) From this gain of regularity on the boundary, along with $\Delta\omega = -\lambda\omega \in L^2(\Omega)$, we conclude ω is in $W^1(\Omega)$, from which it follows that u is a strong solution to A_V as in Definition 8.1.

The origin of this proof was the proof of [Clopeau et al. 1998, Lemma 2.2]. □

We have the following simple extension of Lemma 6.1:

Lemma 8.9. *If Γ is C^2 and has a finite number of components and* (8-5) *holds, then $V \cap \ker\{A_V - \lambda\} = \{0\}$ for all λ in $\mathbb{R}$.*

Proof. By Proposition 8.8, u is a strong eigenfunction of A_V and hence satisfies $\omega(u) = -f u \cdot \boldsymbol{\tau} = 0$ on Γ, and so is a strong eigenfunction of A_L. But then $u = 0$ by Lemma 6.1. □

Restricting our attention to the case where f is nonnegative and constant on Γ (in which case (8-5) holds), we can write the boundary conditions in Definition 8.1 as $(1 - \theta)\omega(u_j) + \theta u_j \cdot \boldsymbol{\tau} = 0$ on Γ, where θ lies in [0, 1]. When $\theta = 0$, we have the special case of Lions boundary conditions and when $\theta = 1$ we have Dirichlet boundary conditions on the velocity. In Definition 8.2, $f = \theta/(1-\theta)$ for θ in [0, 1). With this parameterization, we can view γ_j as a function of θ. That is, $\gamma_j(\theta)$ is the j-th eigenvalue of A_V (or A_L or A_S) so, for instance, to each eigenvalue $\gamma_j(\theta)$ of multiplicity k there will be exactly k values of n for which $\gamma_n(\theta) = \gamma_j(\theta)$.

Because f is constant on Γ, it is certainly in $C^1(\Gamma)$, which is a requirement of Proposition 8.8.

Proposition 8.10. *Assume that Γ is C^2 and has a finite number of components. For all j in $\mathbb{N}$, the function $\gamma_j \colon [0,1) \to [\lambda_j, \nu_j)$ and is strictly increasing and continuous.*

Proof. To show that $\gamma_j(\theta) < \nu_j$ for θ in $[0,1)$ we repeat the proof of Theorem 1.1 using $G = F \oplus \ker\{A_V - \lambda\}$ in place of $F \oplus \ker(\{A_L - \lambda\} \cap X_0^2)$. Let $u \in F$ and $v \in \ker\{A_V - \lambda\}$. Then because v is a weak eigenfunction of A_V as in Definition 8.2 and u is zero on the boundary, letting $z = f = \theta/(1-\theta)$, we have

$$(\omega(u), \omega(v)) = \lambda(u, v) - z \int_\Gamma v \cdot \bar{u} = \lambda(u, v).$$

Thus,

$$\begin{aligned}\|\omega(u+v)\|^2_{L^2(\Omega)} &= \|\omega(u)\|^2_{L^2(\Omega)} + \|\omega(v)\|^2_{L^2(\Omega)} + 2\operatorname{Re}(\omega(u), \omega(v)) \\ &= \|\omega(u)\|^2_{L^2(\Omega)} + \|\omega(v)\|^2_{L^2(\Omega)} + 2\lambda \operatorname{Re}(u, v),\end{aligned}$$

as was the case for A_L. Now, however,

$$\|\omega(v)\|^2_{L^2(\Omega)} = \lambda \|v\|^2_{L^2(\Omega)} - z \int_\Gamma |v|^2 = \lambda \|v\|^2_{L^2(\Omega)} - z \int_\Gamma |u+v|^2,$$

and combined with (6-1) this gives

$$\begin{aligned}\|\omega(u+v)\|^2_{L^2(\Omega)} &\le \lambda \|u\|^2_{L^2(\Omega)} + \lambda \|v\|^2_{L^2(\Omega)} + 2\lambda \operatorname{Re}(u, v) - z \int_\Gamma |u+v|^2 \\ &= \lambda \|u+v\|^2_{L^2(\Omega)} - z \int_\Gamma |u+v|^2.\end{aligned}$$

Thus, $R_V(u+v) \le \lambda$, and the proof of $\gamma_j(\theta) < \nu_j$ is completed as in the proof of Theorem 1.1.

The argument that γ_j is strictly increasing on $[0,1)$ is more direct, because the variational formulations in Corollary 8.7 for different values of θ all involve maximums over subspaces of the same space Y. (That γ_j is nondecreasing on $[0,1)$ follows immediately from the principle of monotonicity, as in [Weinstein and Stenger 1972, Theorem 2.5.1, page 21].)

For θ in $[0,1)$, write A_V^θ for the operator A_V and similarly for R_V^θ and $\overline{N}_V^\theta$. In particular, $A_L = A_V^0$. Let $f(\theta) = \theta/(1-\theta)$, which we note is an increasing function of θ on $[0,1)$.

Now suppose that θ and θ' are in $[0,1)$ with $\theta < \theta'$. Let $\lambda > 0$ and choose a subspace F of Y of dimension $\overline{N}_V^{\theta'}(\lambda)$ with $R_V^{\theta'} \le \lambda$; that is,

$$\|\omega(u)\|^2_{L^2(\Omega)} + \int_\Gamma f(\theta')|u|^2 \le \lambda \|u\|^2_{L^2(\Omega)} \quad \text{for all } u \in F, \tag{8-6}$$

which is possible by Corollary 8.7. Let $G = F \oplus \ker\{A_V^\theta - \lambda\}$. This is, in fact, a direct sum, since if a nonzero u lies in both F and $\ker\{A_V^\theta - \lambda\}$, then from (8-6) and Definition 8.2 it follows that

$$\int_\Gamma (f(\theta') - f(\theta))|u|^2 \le 0.$$

But $f(\theta') - f(\theta)$ is a positive constant on Γ, so in fact $u = 0$ on Γ and hence lies in V. It follows from Lemma 8.9 that u is identically zero.

This shows that G has at least one more element than F when $\lambda = \gamma_j(\theta)$. But then setting $Z = F$ in the definition of $\overline{N}_V^\theta(\gamma_j(\theta))$ in Corollary 8.7, we see because $R_V^\theta \le R_V^{\theta'}$ that $\overline{N}_V^\theta(\gamma_j(\theta)) \ge \dim G > \dim F = \overline{N}_V^{\theta'}(\gamma_j(\theta))$, which means that $\gamma_j(\theta) < \gamma_j(\theta')$.

This shows that γ_j is strictly increasing. To show that it is continuous, fix θ in $[0, 1)$, and let Z be any subspace of Y that achieves the maximum in the expression for $k = \overline{N}_V^\theta(\gamma_k(\theta))$ in Corollary 8.7. Here we assume that if λ_k is a multiple eigenvalue, k is the largest such index.

Choose a basis $(v_1, \dots, v_k)$ for Z and observe that because $R_V(u) = R_V(cu)$ for any nonzero constant c,

$$\sup_{u \in Z} R_V^{\theta'}(u) = \max_{u \in Z'} R_V^{\theta'}(u) \quad \text{for any } \theta' \text{ in } [0, 1),$$

where

$$Z' = \{c_1 v_1 + \dots + c_k v_k : c_1, \dots, c_k \in \mathbb{C},\ |c_1|^2 + \dots + |c_k|^2 = 1\}.$$

Now, the map from the complex k-sphere to $\mathbb{R}$ defined by $(c_1, \dots, c_k) \mapsto \|c_1 v_1 + \dots + c_k v_k\|_{L^2(\Omega)}$ is continuous and so achieves its minimum a, which is the same as the minimum of $\|u\|_{L^2(\Omega)}$ on Z'. Because $(v_1, \dots, v_k)$ is independent, a must be positive. Similarly, $\|u\|_Y$ achieves its maximum $b > 0$ on Z'.

Thus, on Z' and so on Z, for any $\theta' > \theta$,

$$\begin{aligned} R_V^{\theta'}(u) - R_V^\theta(u) = \frac{(f(\theta') - f(\theta)) \int_\Gamma |u|^2}{\|u\|_{L^2(\Omega)}^2} &\le C a^{-2} \|u\|_Y^2 (f(\theta') - f(\theta)) \\ &\le C a^{-2} b^2 (f(\theta') - f(\theta)), \end{aligned}$$

where we used the standard trace inequality $\|u\|_{L^2(\Gamma)} \le C \|u\|_{L^2(\Omega)}^{1/2} \|\nabla u\|_{L^2(\Omega)}^{1/2}$ for u in Y, followed by Poincaré's inequality. But this shows that

$$\overline{N}_V^{\theta'}(\lambda) \ge \overline{N}_V^\theta(\gamma_k(\theta)) \quad \text{for } \lambda = \gamma_k(\theta) + C a^{-2} b^2 (f(\theta') - f(\theta)).$$

Since we already know that $\gamma_k(\theta') > \gamma_k(\theta)$ it follows that

$$|\gamma_k(\theta') - \gamma_k(\theta)| \le C a^{-2} b^2 (f(\theta') - f(\theta)),$$

meaning that γ_k is continuous on [0, 1). □

The first part of Theorem 8.11 is Theorem 1.2.

Theorem 8.11. *Assume that* Γ *is* C^2 *and has a finite number of components. For all* j *in* $\mathbb{N}$, *the function* $\gamma_j : [0, 1] \to [\lambda_j, \nu_j]$ *is a strictly increasing continuous bijection. Also*, (7-2) *holds for any eigenfunction of* A_V.

Proof. For any value of θ in (0, 1), we let $w = w(\theta)$ be any eigenfunction of A_V with eigenvalue $\gamma_j(\theta)$, normalized so that $\|w\|_H = \|w\|_{L^2(\Omega)} = 1$. We know from Proposition 8.10 that $\gamma_j(\theta)$ strictly increases continuously from λ_j at $\theta = 0$ and remains bounded by ν_j. Formally, as $\theta \to 1$, w becomes an eigenfunction of A_S, since w must approach zero on the boundary so that $\omega(w) = (\theta/(1-\theta)) w \cdot \boldsymbol{\tau}$ can remain finite. We now make this formal argument rigorous.

Letting $z = f = \theta/(1-\theta)$, we have

$$\|w\|^2_{L^2(\Gamma)} = \int_\Gamma (w \cdot \boldsymbol{\tau})(\overline{w} \cdot \boldsymbol{\tau}) = -z^{-1} \int_\Gamma \omega(w) \overline{w} \cdot \boldsymbol{\tau},$$

the boundary integral being well defined because of Proposition 8.8. Then

$$\int_\Gamma \omega(w) \overline{w} \cdot \boldsymbol{\tau} = -z \|w\|^2_{L^2(\Gamma)} \le 0,$$

so (7-2) holds.

Moreover, from Definition 8.2,

$$\|\omega(w)\|^2_{L^2(\Omega)} + z\|w\|^2_{L^2(\Gamma)} = \gamma_j(\theta)\|w\|^2_{L^2(\Omega)} = \gamma_j(\theta).$$

From this we conclude two things. First, that

$$\|w\|^2_{L^2(\Gamma)} = \frac{\gamma_j(\theta) - \|\omega(w)\|^2_{L^2(\Omega)}}{z} \le \frac{\nu_j}{z}, \tag{8-7}$$

since $\gamma_j(\theta) < \nu_j$. Second, that $\|\omega(w)\|_{L^2(\Omega)} \le \gamma_j(\theta)^{1/2}$ and hence that $\|w\|_Y \le C$ because $\gamma_j(\theta) < \nu_j$ and by virtue of Corollary 2.15.

Now letting the parameter θ vary over the set $\{1 - 1/n : n \in \mathbb{N}\}$, we get a sequence (u^n) of eigenfunctions $u^n = w(1-1/n)$ of A_V, with eigenvalues $\gamma^n = \gamma_j(1-1/n)$. By the observations above, (u^n) is a bounded sequence in Y. But Y is compactly embedded in H by Lemma 2.6 (or by Corollaries 2.12 and 2.16), so there exists a subsequence of (u^n) that converges strongly in H. Since this subsequence is bounded in Y, which is a separable, reflexive Banach space, taking a further subsequence, and relabeling it (u^n), we conclude that $u^n \to u$ strongly in H and weak* in Y to some vector field u in Y with $\|u\|_H = 1$ (so u is nonzero).

Furthermore, $\|u^n\|_{W^{1/2}(\Gamma)} \le C\|u^n\|_Y \le C$, so (u^n) is bounded in $W^{1/2}(\Gamma)$, which is compactly embedded in $L^2(\Gamma)$, and hence extracting a subsequence and relabeling once more, we conclude that also $u^n \to u$ strongly in $L^2(\Gamma)$. But since $z \to \infty$ as $n \to \infty$, we have $u^n \to u = 0$ in $L^2(\Gamma)$ by (8-7).

Then by Definition 8.2, $(\omega(u^n), \omega(v)) - \gamma^n(u^n, v) = 0$ for any v in V. Letting $\gamma = \lim_{n\to\infty} \gamma^n$ (the limit exists because γ^n is a bounded increasing sequence of real numbers), we have $(\omega(u^n), \omega(v)) - \gamma(u^n, v) = (\gamma^n - \gamma)(u^n, v)$. Since $|(u^n, v)| \le \|u^n\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} \le C$, the right side converges to zero. Since $u^n \to u$ strongly in $L^2(\Omega)$, $(u^n, v) \to (u, v)$. Since $u^n \to u$ weak* in Y,

$$(\omega(u_n), \omega(v)) = (\nabla u_n, \nabla v) \to (\nabla u, \nabla v) = (\omega(u), \omega(v)),$$

where we used Corollary A.3. We conclude that $(\omega(u), \omega(v)) - \gamma(u, v) = 0$ and thus that u is a weak eigenfunction of A_S with eigenvalue $\gamma \le \nu_j$.

What we have shown is that $\gamma_j : [0, 1] \to [\lambda_j, \nu_k]$ for some $k \le j$ and that γ_j is strictly increasing and continuous on all of $[0, 1]$. To show that $k = j$, we first observe that if $\gamma_k(1) = \gamma_m(1) = \nu_j$ for some $k \ne m$, then the eigenvalue ν_j has multiplicity at least 2. To see this, we repeat the compactness argument above, this time choosing the original sequence of eigenvectors $(u^{k,n})_{n=1}^\infty$ and $(u^{m,n})_{n=1}^\infty$ such that $u^{k,n}$ is orthogonal in $L^2(\Omega)$ to $u^{m,n}$, which we can always do even if they lie in the same eigenspace. We showed above that $u^{k,n} \to u$ and $u^{m,n} \to w$ in $L^2(\Omega)$ for some u and w that are eigenvectors of A_S. It is elementary to see, then, that $(u, w) = 0$, which shows that ν_j has multiplicity at least two.

Similarly, the multiplicity of the eigenvalue ν_j is at least as high as the number of distinct values of k for which $\gamma_k(1) = \nu_j$. This means that the total number of eigenvalues of A_S including multiplicity reached by $\gamma_j(1)$ for some j with $1 \le j \le k$ is at least k. But it can be no more than k since $\gamma_j(1) = \nu_m$ for some $m \le j \le k$. Thus, the first k eigenvalues of A_L according to multiplicity are mapped via γ_j for $j = 1, \dots, k$ into the first k eigenvalues of A_S, showing that $\gamma_j : [0, 1] \to [\lambda_j, \nu_j]$ for all $j = 1, \dots, k$ and hence for all j in $\mathbb{N}$, since k was arbitrary. □

To round out the picture of how the eigenvalues for different boundary conditions compare, we consider the eigenfunctions of the negative Laplacian with Robin boundary conditions on the vorticity. For simplicity, we restrict our attention to constant coefficients, writing the boundary conditions in terms of a parameter θ lying in $[0, 1]$, and stating only the strong form:

Definition 8.12. An eigenfunction $\omega_j \in W_0^1(\Omega)$ of the Dirichlet Laplacian with Robin boundary conditions satisfies

$$\begin{cases} \Delta\omega_j + \eta_j\omega_j = 0 & \text{in } \Omega, \\ (1-\theta)\nabla\omega_j \cdot \boldsymbol{n} + \theta\omega_j = 0 & \text{on } \Gamma. \end{cases}$$

The analogue for divergence-free vector fields leads to the eigenvalue problem for a Stokes operator A_R with Robin boundary conditions:

Definition 8.13. An eigenfunction $u_j \in X^2$ of A_R satisfies $A_R u_j = \lambda_j^* u_j$ or, equivalently,

$$\begin{cases} \Delta u_j + \eta_j^* u_j = \nabla p_j, & \operatorname{div} u_j = 0 \quad \text{in } \Omega, \\ \quad u_j \cdot \boldsymbol{n} = 0, \quad (1-\theta)\nabla\omega_j \cdot \boldsymbol{n} + \theta\omega_j = 0 & \text{on } \Gamma. \end{cases}$$

A value of $\theta = 1$ gives the operator A_L, and $\theta = 0$ gives Neumann boundary conditions on the vorticity.

Taking the vorticity of u_j in Definition 8.13 shows that a strong eigenfunction of A_R corresponds to a strong eigenfunction of the Dirichlet Laplacian with Robin boundary conditions. Also, the equivalent of Lemma 3.4 for Robin boundary conditions on ω shows that to each strong eigenfunction of the Dirichlet Laplacian with Robin boundary conditions there corresponds a strong eigenfunctions of A_R. Thus, there is a bijection between the eigenfunctions and eigenvalues; that is, $\eta_j^* = \eta_j$. Moreover, η_j is continuous on [0, 1), because the bilinear form associated to Definition 8.12 is continuous with θ; see [Filonov 2004].

Proposition 8.14. *For all j in $\mathbb{N}$, the function $\eta_j \colon [0, 1) \to [\mu_j, \lambda_j)$ and is strictly increasing.*

Proof. The proof goes like that of Proposition 8.10, making adaptations of Filonov's proof of his theorem that parallel those in the proof of Proposition 8.10. □

Theorem 8.15. *For all j in $\mathbb{N}$, the function $\eta_j : [0, 1] \to [\mu_j, \lambda_j]$ is continuous and strictly increasing.*

Proof. The proof parallels that of Theorem 8.11. □

The addendum of [Filonov 2004] considers Robin boundary conditions as in Definition 8.12 with, in effect, θ negative. In that case, $\eta_{j+1}(\theta) < \lambda_j$ for all j in $\mathbb{N}$.

For any θ,

$$\begin{aligned} &\|\nabla p_j\|^2_{L^2(\Omega)} - \int_\Gamma (\nabla\omega_j \cdot \boldsymbol{n})\omega_j \\ &\qquad = \|\Delta u_j\|^2_{L^2(\Omega)} - \eta_j(\theta)\|u_j\|^2_{L^2(\Omega)} - \int_\Omega \Delta\omega(u_j)\omega(u_j) - \|\nabla\omega(u_j)\|^2_{L^2(\Omega)} \\ &\qquad = \eta_j(\theta)\big(\|\omega(u_j)\|^2_{L^2(\Omega)} - \eta_j(\theta)\|u_j\|^2_{L^2(\Omega)}\big). \end{aligned}$$

Thus, (7-2) holds for an eigenfunction of A_L ($\theta = 1$), where $\nabla p_j \equiv 0$ and $\omega_j = 0$ on Γ, and fails for an eigenfunction of the Stokes operator with Neumann boundary conditions on the vorticity (A_R for $\theta = 0$), where $\nabla p_j \not\equiv 0$ and $\nabla\omega_j \cdot \boldsymbol{n} = 0$ on Γ. For θ in (0, 1), it is not clear whether (7-2) holds or not, leaving open the possibility that the inequality $\lambda_{j+1} \le \nu_j$ could be proved by showing that (7-2) holds for all θ in some left neighborhood of 1 for each λ_j.

In any case, for all j we have the inequalities

$$\mu_j < \eta_j(\theta) < \lambda_j < \gamma_j(\theta') < \nu_j \quad \text{for all } \theta, \theta' \text{ in } (0,1),$$
$$\mu_{j+1} < \lambda_j < \beta_j < \nu_j.$$

Appendix A. Various lemmas

Corollary A.1 is a corollary of Lemma 2.4 and is the main tool we use to prove the equivalence of the weak and strong formulations of our eigenvalue problems. The conditions in this corollary for equality to hold are the weakest possible to insure that each factor lies in the correct space for each term to be finite.

Corollary A.1. *Assume that Ω is a bounded domain with locally Lipschitz boundary. For any divergence-free distribution u for which $\omega(u)$ is in $W^1(\Omega)$ and any v in $L^2(\Omega)$ with $\omega(v)$ in $L^2(\Omega)$,*

$$(\omega(u), \omega(v)) = -(\Delta u, v) + \int_\Gamma \omega(u)\bar{v} \cdot \boldsymbol{\tau}.$$

Proof. The vector field v is in $E(\Omega)$ because $v^\perp$ is in $L^2(\Omega)$ and $\operatorname{div} v^\perp = -\omega(v)$ is in $L^2(\Omega)$. Thus, $\omega(u)$ lying in $W^1(\Omega)$, we can apply Lemma 2.4 to obtain

$$(\omega(u), \omega(v)) = -(\omega(u), \operatorname{div} v^\perp) = (\nabla\omega(u), v^\perp) - \int_\Gamma \omega(u)(\bar{v}^\perp \cdot \boldsymbol{n}).$$

But $(\nabla\omega(u), v^\perp) = -(\nabla^\perp\omega(u), v) = (-\Delta u, v)$ and $(\bar{v}^\perp \cdot \boldsymbol{n}) = -\bar{v} \cdot \boldsymbol{\tau}$, from which the result follows. □

Lemma A.2. *Assume that Ω is a bounded domain with locally Lipschitz boundary. If u is in $W^1(\Omega)$ with $\operatorname{div} u = 0$ and v is in $(C^1(\Omega))^2$, then*

$$(\omega(u), \omega(v)) = (\nabla u, \nabla v) - \int_\Gamma (\nabla u \bar{v}) \cdot \boldsymbol{n}.$$

Proof. We have

$$\begin{aligned}
\omega(u)\omega(\bar{v}) &= (\partial_1 u^2 - \partial_2 u^1)(\partial_1 \bar{v}^2 - \partial_2 \bar{v}^1) \\
&= \partial_1 u^2 \partial_1 \bar{v}^2 + \partial_2 u^1 \partial_2 \bar{v}^1 - (\partial_1 u^2 \partial_2 \bar{v}^1 + \partial_2 u^1 \partial_1 \bar{v}^2) \\
&= \partial_1 u^2 \partial_1 \bar{v}^2 + \partial_2 u^1 \partial_2 \bar{v}^1 + \partial_1 u^1 \partial_1 \bar{v}^1 + \partial_2 u^2 \partial_2 \bar{v}^2 \\
&\qquad - (\partial_1 u^2 \partial_2 \bar{v}^1 + \partial_2 u^1 \partial_1 \bar{v}^2 + \partial_1 u^1 \partial_1 \bar{v}^1 + \partial_2 u^2 \partial_2 \bar{v}^2) \\
&= \partial_i u^j \partial_i \bar{v}^j - \partial_i u^j \partial_j \bar{v}^i = \nabla u \cdot \nabla \bar{v} - (\nabla u)^T \cdot \nabla \bar{v}.
\end{aligned}$$

Since $\operatorname{div} u = 0$, we have $(\nabla u)^T \cdot \nabla \bar{v} = \partial_i u^j \partial_j \bar{v}^i = \partial_j(\partial_i u^j \bar{v}^i) = \operatorname{div}(\nabla u \bar{v})$. But $\nabla u \bar{v}$ is in $L^2(\Omega)$ and $\|\operatorname{div}(\nabla u \bar{v})\|_{L^2(\Omega)} \le \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^\infty(\Omega)}$ is finite, so $\nabla u \bar{v}$

is in $E(\Omega)$ and we can apply Lemma 2.4 to give

$$(\omega(u),\ \omega(v)) = (\nabla u,\ \nabla v) - \int_\Omega \operatorname{div}(\nabla u \bar{v}) = (\nabla u,\ \nabla v) - \int_\Gamma (\nabla u \bar{v}) \cdot \boldsymbol{n}. \qquad \square$$

Corollary A.3. *Assume that Ω is a bounded domain with locally Lipschitz boundary. For all u in $W^1(\Omega)$ with* $\operatorname{div} u = 0$ *and all v in V,*

$$(\omega(u), \omega(v)) = (\nabla u,\ \nabla v).$$

Proof. This follows from Lemma A.2 and the density of $C^1(\Omega)$ in $W^1(\Omega)$. $\square$

Lemma A.4. *Assume that Γ is C^2. For all u in $H \cap W^2(\Omega)$ and v in Y, we have*

$$\nabla u v \cdot \boldsymbol{n} = -\kappa u \cdot v.$$

Proof. Because $u \cdot \boldsymbol{n}$ has a constant value (of zero) along Γ,

$$0 = \frac{\partial}{\partial \boldsymbol{\tau}}(u \cdot \boldsymbol{n}) = \frac{\partial u}{\partial \boldsymbol{\tau}} \cdot \boldsymbol{n} + u \cdot \frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\tau}} = \nabla u \boldsymbol{\tau} \cdot \boldsymbol{n} + \kappa u \cdot \boldsymbol{\tau}.$$

But $v = (v \cdot \boldsymbol{\tau})\boldsymbol{\tau}$, so multiplying both sides of the above inequality by $v \cdot \boldsymbol{\tau}$ completes the proof. $\square$

Lemma A.5. *Assume that Γ is C^2. For all u in $H \cap W^2(\Omega)$ and v in Y, we have*

$$\nabla u \boldsymbol{n} \cdot v = \omega(u) v \cdot \boldsymbol{\tau} - \kappa u \cdot v.$$

Proof. Writing

$$\boldsymbol{n} = \begin{pmatrix} n^1 \\ n^2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\tau} = \begin{pmatrix} -n^2 \\ n^1 \end{pmatrix}$$

with $(n^1)^2 + (n^2)^2 = 1$, we have

$$\begin{aligned}
&\nabla u \boldsymbol{n} \cdot \boldsymbol{\tau} - \nabla u \boldsymbol{\tau} \cdot \boldsymbol{n} \\
&\quad = \left(\begin{pmatrix} \partial_1 u^1 & \partial_2 u^1 \\ \partial_1 u^2 & \partial_2 u^2 \end{pmatrix} \begin{pmatrix} n^1 \\ n^2 \end{pmatrix} \right) \cdot \begin{pmatrix} -n^2 \\ n^1 \end{pmatrix} - \left(\begin{pmatrix} \partial_1 u^1 & \partial_2 u^1 \\ \partial_1 u^2 & \partial_2 u^2 \end{pmatrix} \begin{pmatrix} -n^2 \\ n^1 \end{pmatrix} \right) \cdot \begin{pmatrix} n^1 \\ n^2 \end{pmatrix} \\
&\quad = \begin{pmatrix} \partial_1 u^1 n^1 + \partial_2 u^1 n^2 \\ \partial_1 u^2 n^1 + \partial_2 u^2 n^2 \end{pmatrix} \cdot \begin{pmatrix} -n^2 \\ n^1 \end{pmatrix} - \begin{pmatrix} -\partial_1 u^1 n^2 + \partial_2 u^1 n^1 \\ -\partial_1 u^2 n^2 + \partial_2 u^2 n^1 \end{pmatrix} \cdot \begin{pmatrix} n^1 \\ n^2 \end{pmatrix} \\
&\quad = -\partial_1 u^1 n^1 n^2 - \partial_2 u^1 (n^2)^2 + \partial_1 u^2 (n^1)^2 + \partial_2 u^2 n^1 n^2 \\
&\qquad\qquad + \partial_1 u^1 n^1 n^2 - \partial_2 u^1 (n^1)^2 + \partial_1 u^2 (n^2)^2 - \partial_2 u^2 n^1 n^2 \\
&\quad = ((n^1)^2 + (n^2))(\partial_1 u^2 - \partial_2 u) = \omega(u).
\end{aligned}$$

Thus by Lemma A.4,

$$\nabla u \boldsymbol{n} \cdot \boldsymbol{\tau} = \omega(u) + \nabla u \boldsymbol{\tau} \cdot \boldsymbol{n} = \omega(u) - \kappa u \cdot \boldsymbol{\tau},$$

and multiplying both sides by $v \cdot \boldsymbol{\tau}$ completes the proof. $\square$

Acknowledgments

I would like to thank Leonid Friedlander and Mark Ashbaugh for helpful comments, and James Stafney for a number of illuminating discussions.

References

[Adams 1975] R. A. Adams, *Sobolev spaces*, Pure and Applied Mathematics **65**, Academic Press, New York-London, 1975. MR 56 #9247 Zbl 0314.46030

[Ashbaugh 2004] M. S. Ashbaugh, "On universal inequalities for the low eigenvalues of the buckling problem", pp. 13–31 in *Partial differential equations and inverse problems*, edited by C. Conca et al., Contemp. Math. **362**, Amer. Math. Soc., Providence, RI, 2004. MR 2006f:35198 Zbl 1062.35050

[Brown and Shen 1995] R. M. Brown and Z. Shen, "Estimates for the Stokes operator in Lipschitz domains", *Indiana Univ. Math. J.* **44**:4 (1995), 1183–1206. MR 97c:35152 Zbl 0858.35098

[Clopeau et al. 1998] T. Clopeau, A. Mikelić, and R. Robert, "On the vanishing viscosity limit for the 2D incompressible Navier–Stokes equations with the friction type boundary conditions", *Nonlinearity* **11**:6 (1998), 1625–1636. MR 99g:35102 Zbl 0911.76014

[Evans 1998] L. C. Evans, *Partial differential equations*, Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 1998. MR 99e:35001 Zbl 0902.35002

[Filonov 2004] N. Filonov, "On an inequality for the eigenvalues of the Dirichlet and Neumann problems for the Laplace operator", *Algebra i Analiz* **16**:2 (2004), 172–176. In Russian; translated in *St. Petersburg Math. J.* **16**:2 (2005), 413–416. MR 2005f:35228

[Foias et al. 2001] C. Foias, O. Manley, R. Rosa, and R. Temam, *Navier–Stokes equations and turbulence*, Encyclopedia of Mathematics and its Applications **83**, Cambridge University Press, 2001. MR 2003a:76001 Zbl 0994.35002

[Friedlander 1991] L. Friedlander, "Some inequalities between Dirichlet and Neumann eigenvalues", *Arch. Rational Mech. Anal.* **116**:2 (1991), 153–160. MR 93h:35146 Zbl 0789.35124

[Friedlander 2004] L. Friedlander, "Remarks on the membrane and buckling eigenvalues for planar domains", *Mosc. Math. J.* **4**:2 (2004), 369–375, 535. MR 2005m:35204 Zbl 1072.35134

[Gagliardo 1957] E. Gagliardo, "Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in *n* variabili", *Rend. Sem. Mat. Univ. Padova* **27** (1957), 284–305. MR 21 #1525 Zbl 0087.10902

[Galdi 1994] G. P. Galdi, *An introduction to the mathematical theory of the Navier–Stokes equations, I: Linearized steady problems*, Springer Tracts in Natural Philosophy **38**, Springer, New York, 1994. MR 95i:35216a Zbl 0949.35004

[Gilbarg and Trudinger 1977] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der Mathematischen Wissenschaften **224**, Springer, Berlin, 1977. MR 57 #13109 Zbl 0361.35003

[Grisvard 1985] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics **24**, Pitman, Boston, MA, 1985. MR 86m:35044 Zbl 0695.35060

[Henrot 2006] A. Henrot, *Extremum problems for eigenvalues of elliptic operators*, Birkhäuser, Basel, 2006. MR 2007h:35242 Zbl 1109.35081

[Iftimie and Planas 2006] D. Iftimie and G. Planas, "Inviscid limits for the Navier–Stokes equations with Navier friction boundary conditions", *Nonlinearity* **19**:4 (2006), 899–918. MR 2007c:35130 Zbl 1169.35365

[Iftimie and Sueur 2006] D. Iftimie and F. Sueur, "Viscous boundary layers for the Navier–Stokes equations with the Navier slip conditions", preprint, Centre Interfacultaire Bernoulli, 2006, Available at http://infoscience.epfl.ch/search.py?recid=126262.

[Kelliher 2006] J. P. Kelliher, "Navier–Stokes equations with Navier boundary conditions for a bounded domain in the plane", *SIAM J. Math. Anal.* **38**:1 (2006), 210–232. MR 2007a:35118 Zbl 05029418

[Kesavan 1989] S. Kesavan, *Topics in functional analysis and applications*, Wiley, New York, 1989. MR 90m:46002 Zbl 0666.46001

[Lions 1969] J.-L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod, Paris, 1969. MR 41 #4326 Zbl 0189.40603

[Lions 1996] P.-L. Lions, *Mathematical topics in fluid mechanics, I: Incompressible models*, Oxford Lecture Series in Mathematics and its Applications **3**, Oxford University Press, New York, 1996. MR 98b:76001 Zbl 0866.76002

[Lopes Filho et al. 2005] M. C. Lopes Filho, H. J. Nussenzveig Lopes, and G. Planas, "On the inviscid limit for two-dimensional incompressible flow with Navier friction condition", *SIAM J. Math. Anal.* **36**:4 (2005), 1130–1141. MR 2005k:76026 Zbl 1084.35060

[Maslennikova and Bogovskiĭ 1983] V. N. Maslennikova and M. E. Bogovskiĭ, "Approximation of potential and solenoidal vector fields", *Sibirsk. Mat. Zh.* **24**:5 (1983), 149–171. In Russian; translated in *Sib. Math. J.* **24**, 768–787. MR 85j:46052 Zbl 0599.46045

[Payne 1955] L. E. Payne, "Inequalities for eigenvalues of membranes and plates", *J. Rational Mech. Anal.* **4** (1955), 517–529. MR 17,42a Zbl 0064.34802

[Payne 1967] L. E. Payne, "Isoperimetric inequalities and their applications", *SIAM Rev.* **9** (1967), 453–488. MR 36 #2058 Zbl 0154.12602

[Temam 1984] R. Temam, *Navier–Stokes equations: Theory and numerical analysis*, 3rd ed., Studies in Mathematics and its Applications **2**, North-Holland, Amsterdam, 1984. MR 86m:76003 Zbl 0568.35002

[Weinstein and Stenger 1972] A. Weinstein and W. Stenger, *Methods of intermediate problems for eigenvalues: Theory and ramifications*, Mathematics in Science and Engineering **89**, Academic Press, New York, 1972. MR 57 #17469 Zbl 0291.49034

Received December 5, 2008. Revised July 24, 2009.

JAMES P. KELLIHER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
900 UNIVERSITY AVENUE
RIVERSIDE, CA 92521
UNITED STATES

kelliher@math.ucr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

TORUS ACTIONS ON SMALL BLOWUPS OF $\mathbb{CP}^2$

LIAT KESSLER

A manifold obtained by k simultaneous symplectic blowups of $\mathbb{CP}^2$ of equal sizes ϵ (where the size of $\mathbb{CP}^1 \subset \mathbb{CP}^2$ is one) admits an effective two dimensional torus action if $k \leq 3$. We show that it does not admit such an action if $k \geq 4$ and $\epsilon \leq 1/(3k2^{2k})$. For the proof, we show a correspondence between the geometry of a symplectic toric four-manifold and the combinatorics of its moment map image. We also use techniques from the theory of J-holomorphic curves.

1. Introduction

Let a torus $\mathbb{T}^\ell = (S^1)^\ell$ act effectively on a symplectic $2n$-dimensional manifold (M, ω). The action is called *Hamiltonian* if there exists a *moment map*, that is, a map

$$\Phi : M \to (\mathfrak{t}^\ell)^* = \mathbb{R}^\ell$$

that satisfies

$$d\Phi_i = -\iota(\xi_i)\omega$$

for $i = 1, \ldots, \ell$, where $\xi_1, \ldots, \xi_\ell$ are the vector fields that generate the $\mathbb{T}^\ell$-action. Unless said otherwise, we assume that M is compact and connected. The image of the moment map,

$$\Delta := \Phi(M),$$

is then a convex polytope [Guillemin and Sternberg 1982].

If $\dim \mathbb{T}^\ell = \frac{1}{2} \dim M$, the triple (M, ω, Φ) is a *symplectic toric manifold*, and the torus action is called *toric*. The moment map image is a *Delzant polytope*; this means that the edges emanating from each vertex are generated by vectors $v_1, \ldots, v_n$ that span the lattice $\mathbb{Z}^n$. By the Delzant theorem, (M, ω, Φ) is determined by Δ up to an equivariant symplectomorphism. Conversely, given a Delzant polytope Δ in $\mathbb{R}^n$, Delzant [1988] constructs a symplectic toric manifold $(M_\Delta, \omega_\Delta, \Phi_\Delta)$ whose moment map image is Δ.

MSC2000: primary 53D20, 53D35, 53D45; secondary 57S15.
Keywords: torus action, symplectic manifold, dimension four, symplectic blowup, J-holomorphic curve, moment map, Delzant polygon.

As a result of Delzant's theorem and a combinatorial analysis of Delzant polygons, any symplectic toric four-manifold is obtained from either a standard $\mathbb{CP}^2$ or a Hirzebruch surface by a sequence of equivariant symplectic blowups. (See Lemma 2.9.) However, it may be difficult to determine whether a given symplectic four-dimensional manifold is symplectomorphic to such a manifold.

For instance, let (M_k, ω_ϵ) be a symplectic manifold obtained from $(\mathbb{CP}^2, \omega_{\mathrm{FS}})$ by k simultaneous symplectic blowups of equal sizes $\epsilon > 0$. (Our normalization convention for the Fubini–Study form ω_{FS} is that the size of $\mathbb{CP}^1 \subset \mathbb{CP}^n$, $(1/2\pi)\int_{\mathbb{CP}^1} \omega_{\mathrm{FS}}$, is equal to one.) If $k \geq 4$, this manifold does not admit a toric action that is *consistent with the blowups*, that is, the blowups cannot be performed equivariantly. (See Lemma 2.8.) Does it admit any other toric action?

In [Karshon and Kessler 2007] we show that the answer is "no" when ϵ is $1/n$ for an integer n. In this paper we show that the answer is "no" for $\epsilon \leq 1/(3k2^{2k})$, as a corollary of the following theorem.

Theorem 1.1. *If (M_k, ω_ϵ) is symplectomorphic to $(M_\Delta, \omega_\Delta)$, for a Delzant polygon Δ, and*

$$\epsilon \leq \frac{1}{3k2^{2k}},$$

then $(M_\Delta, \omega_\Delta, \Phi_\Delta)$ can be obtained from $(\mathbb{CP}^2, \omega_{\mathrm{FS}})$ by k equivariant symplectic blowups of equal size ϵ.

The theorem becomes false if we do not restrict ϵ; for $\epsilon > \frac{1}{2}$, there is a toric action on (M_1, ω_ϵ) that is not consistent with the ϵ-blowup; see Remark 5.5. Theorem 1.1 can be strengthen to the case $\epsilon \leq \frac{1}{3}$; see [Pinsonnault 2008, Corollary 3.14; Kessler 2004, Theorem 3]. However, here we use different methods in the proof; in particular, our arguments illustrate explicitly the behavior of J_T-holomorphic curves and their moment map images. (J_T denotes a $\mathbb{T}^2$-invariant complex structure on the manifold that is compatible with the symplectic form.) These novel arguments might be useful in other studies of torus actions on symplectic manifolds.

In proving Theorem 1.1, we apply Gromov's compactness theorem for J-holomorphic curves to show the existence of J_T-curves in the homology classes of exceptional divisors obtained by the symplectic ϵ-blowups. In the case presented here, (as opposed to the case $\epsilon = \frac{1}{n}$ for an integer n), a priori these might be nonsmooth cusp curves. We claim that in one of these homology classes there is a smooth J_T-holomorphic sphere. To prove this claim, we represent J_T-holomorphic spheres and cusp curves on the boundary of the moment map image, and reduce the claim to a combinatorial claim on the moment map polygon. A key ingredient is Lemma 4.3, saying that a J_T-holomorphic sphere whose moment map image avoids a neighbourhood of a vertex in the moment map polygon Δ can be pushed, by a gradient flow, to a connected union of preimages of a chain of edges of Δ.

The geometry-combinatorics correspondence is established in Section 2 and Section 4. The relevant results from Gromov's theory of J-holomorphic curves are recalled in Section 3.

To complete the proof of Theorem 1.1 by recursion, we need uniqueness of symplectic blowdowns: symplectic blowdowns along homologous curves result in symplectomorphic manifolds. This is shown in the appendix.

2. Reading geometric data from the moment map polygon

2.1. An important model for a Hamiltonian action is $\mathbb{C}^n$ with the standard symplectic form, the standard $\mathbb{T}^n$-action given by rotations of the coordinates, and the moment map

$$(z_1, \dots, z_n) \mapsto \frac{1}{2}(|z_1|^2, \dots, |z_n|^2).$$

The image of this moment map is the positive orthant,

$$\mathbb{R}^n_+ = \{(s_1, \dots, s_n) \mid s_j \geq 0 \text{ for all } j\ \}.$$

A Delzant polytope can be obtained by gluing open subsets of $\mathbb{R}^n_+$ by means of elements of $\mathrm{AGL}(n, \mathbb{Z})$. ($\mathrm{AGL}(n, \mathbb{Z})$ is the group of affine transformations of $\mathbb{R}^n$ that have the form $x \mapsto Ax + \alpha$ with $A \in \mathrm{GL}(n, \mathbb{Z})$ and $\alpha \in \mathbb{R}^n$.) Similarly, a symplectic toric manifold can be obtained by gluing open $\mathbb{T}^n$-invariant subsets of $\mathbb{C}^n$ by means of equivariant symplectomorphisms and reparametrizations of $\mathbb{T}^n$.

2.2. The *rational length* of an interval d of rational slope in $\mathbb{R}^n$ is the unique number $\ell = |d|$ such that the interval is $\mathrm{AGL}(n, \mathbb{Z})$-congruent to an interval of length ℓ on a coordinate axis. In what follows, intervals are always measured by rational length.

2.3. An almost complex structure on a $2n$-dimensional manifold M is an automorphism of the tangent bundle, $J : TM \to TM$, such that $J^2 = -\mathrm{Id}$. It is *compatible* with a symplectic form ω if $\langle u, v\rangle = \omega(u, Jv)$ is symmetric and positive definite. The *first Chern class* of the symplectic manifold (M, ω) is defined to be the first Chern class of the complex vector bundle (TM, J) and is denoted $c_1(TM)$. This class is independent of the choice of compatible almost complex structure J [McDuff and Salamon 1998, Section 2.6].

Lemma 2.4. *Let (M, ω) be a compact connected symplectic four-manifold. Let $\Phi : M \to \mathbb{R}^2$ be a moment map for a toric action, and let Δ be its image.*

(1) *The moment map preimage of a vertex of Δ is a fixed point for the torus action, and the moment map image of a fixed point is a vertex of Δ.*

(2) *Let d be an edge of Δ of rational length ℓ. Then its preimage*, $\Phi^{-1}(d)$, *is a symplectically embedded* 2*-sphere in M of symplectic area*

$$\int_{\Phi^{-1}(d)} \omega = 2\pi\ell.$$

(3) *The* (*rational*) *perimeter of Δ is*

$$\text{perimeter } \Delta = \frac{1}{2\pi}\int_M \omega \wedge c_1(TM).$$

(4) *The area of Δ is*

$$\frac{1}{(2\pi)^2}\int_M \frac{1}{2!}\omega \wedge \omega.$$

For proof, see [Karshon et al. 2007, Lemma 2.2 and Lemma 2.10].

Example 2.5. Figure 1 shows examples of Delzant polygons with three and four edges. On the left there is a *Delzant triangle*,

$$\Delta_\lambda = \{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq \lambda\}.$$

This is the moment map image of the standard toric action $(a, b) \cdot [z_0 : z_1 : z_2] = [z_0 : az_1 : bz_2]$ on $\mathbb{CP}^2$, with the Fubini–Study symplectic form normalized so that the symplectic area of $\mathbb{CP}^1 \subset \mathbb{CP}^2$ is $2\pi\lambda$. The rational lengths of all its edges is λ.

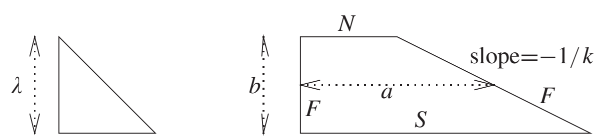

Figure 1. A Delzant triangle, Δ_λ, and a Hirzebruch trapezoid, $\text{Hirz}_{a,b,k}$.

On the right there is a *Hirzebruch trapezoid*,

$$\text{Hirz}_{a,b,k} = \left\{(x_1, x_2) \;\middle|\; -\frac{b}{2} \leq x_2 \leq \frac{b}{2}, 0 \leq x_1 \leq a - kx_2\right\},$$

where b is the height of the trapezoid, a is its average width, and k is a nonnegative integer such that the east edge has slope $-1/k$ or is vertical if $k = 0$. We assume that $a \geq b$ and that $a - k\frac{b}{2} > 0$. This trapezoid is a moment map image of a standard toric action on a Hirzebruch surface. The rational lengths of its west and east edges are b; the rational lengths of its north and south edges are $a \pm kb/2$.

2.6. Let Δ be a Delzant polytope in $\mathbb{R}^n$, let v be a vertex of Δ, and let $\delta > 0$ be smaller than the rational lengths of the edges emanating from v. The edges of Δ emanating from v have the form $\{v + s\alpha_j \mid 0 \leq s \leq \ell_j\}$ where the vectors $\alpha_1, \ldots, \alpha_n$

generate the lattice $\mathbb{Z}^n$ and $\delta < \ell_j$ for all j. The *corner chopping of size* δ of Δ at v is the polytope $\widetilde{\Delta}$ obtained from Δ by intersecting with the half-space

$$\{ v + s_1\alpha_1 + \cdots + s_n\alpha_n \mid s_1 + \cdots + s_n \geq \delta \}.$$

See, for example, the chopping of the top right corner in Figure 2. The resulting polytope $\widetilde{\Delta}$ is again a Delzant polytope. The corner chopping operation commutes with $\mathrm{AGL}(n, \mathbb{Z})$-congruence: if $\widetilde{\Delta}$ is obtained from Δ by a corner chopping of size $\delta > 0$ at a vertex $v \in \Delta$ then, for any $g \in \mathrm{AGL}(n, \mathbb{Z})$, the polytope $g(\widetilde{\Delta})$ is obtained from the polytope $g(\Delta)$ by a corner chopping of size δ at the vertex $g(v)$.

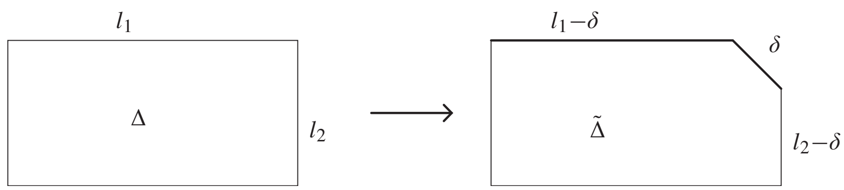

Figure 2. Corner chopping.

2.7. Recall that a blowup of *size* $\epsilon = r^2/2$ of a $2n$-dimensional symplectic manifold (M, ω) is a new symplectic manifold $(\widetilde{M}, \tilde{\omega})$ that is constructed in the following way. Let $\Omega \subset \mathbb{C}^n$ be an open subset that contains a ball about the origin of radius greater than r, and let $i : \Omega \to M$ be a symplectomorphism onto an open subset of M. (We consider $\mathbb{C}^n$ with the standard symplectic form.) The standard symplectic blowup of Ω of size $r^2/2$ is obtained by removing the open ball $B^{2n}(r)$ of radius r about the origin and collapsing its boundary along the Hopf fibration $\partial B^{2n}(r) \to \mathbb{CP}^{n-1}$; the resulting space is naturally a smooth symplectic manifold [McDuff and Salamon 1998, Section 7.1]. This blowup transports to M through i. The resulting copy of $(\mathbb{CP}^{n-1}, \epsilon\omega_{\mathrm{FS}})$ in $\widetilde{M}$ is called the *exceptional divisor*.

If M admits an action of a torus $\mathbb{T}^\ell$, and $i : \Omega \to M$ is $\mathbb{T}^\ell$-equivariant, where $\mathbb{T}^\ell$ acts on Ω through some homomorphism $\mathbb{T}^\ell \to U(n)$, then the torus action naturally extends to the symplectic blowup of M obtained from i, and the blowup is *equivariant*. If the action on M is Hamiltonian, its moment map naturally extends to the blowup; in the case $\ell = n$ we call this a *toric blowup*.

The moment map image of the standard symplectic blowup of $\mathbb{C}^n$ of size ϵ is obtained from the moment map image $\mathbb{R}^n_+$ of $\mathbb{C}^n$ by corner chopping of size ϵ. See Figure 3 for $n = 2$.

A toric blowup of size ϵ of a symplectic toric manifold (M, ω, Φ) at a fixed point p amounts to chopping off a corner of size ϵ of its moment map image Δ at the vertex $v = \Phi(p)$ to get a new polytope $\widetilde{\Delta}$. The preimage of the resulting new facet in $\widetilde{\Delta}$ is the exceptional divisor in $\widetilde{M}$.

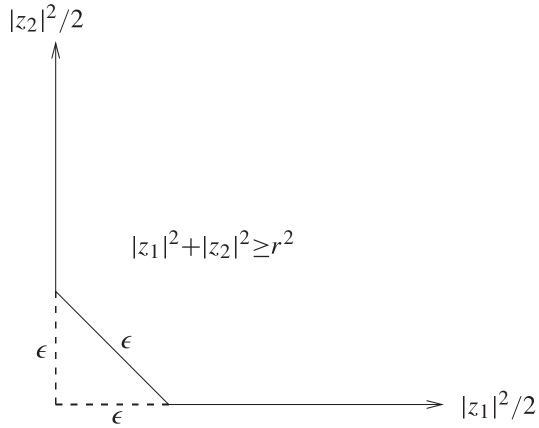

Figure 3. Blowup of $\mathbb{C}^2$ of size $\epsilon = r^2/2$.

We restrict our attention to symplectic toric manifolds of dimension 4. Chopping off a corner of size ϵ of a polygon Δ can be done if and only if there exist two adjacent edges in Δ whose rational lengths are both strictly greater than ϵ. As a result, starting from a Delzant triangle of size 1 we can perform one corner chopping of size $\epsilon > 0$ if and only if $\epsilon < 1$, two or three corner choppings of size $\epsilon > 0$ if and only if $\epsilon < \frac{1}{2}$, and no more than three corner choppings of the same size. Therefore:

Lemma 2.8. $(\mathbb{CP}^2, \omega_{\text{FS}})$ *admits a toric blowup of size* $\epsilon > 0$ *if and only if* $\epsilon < 1$. $(\mathbb{CP}^2, \omega_{\text{FS}})$ *admits two or three toric blowups of size* $\epsilon > 0$ *if and only if* $\epsilon < \frac{1}{2}$. $(\mathbb{CP}^2, \omega_{\text{FS}})$ *does not admit four or more toric blowups of equal sizes.*

For a detailed proof, see [Karshon and Kessler 2007, Lemma 3.1].

In $\mathbb{R}^2$, all Delzant polygons can be obtained by a simple recursive recipe:

Lemma 2.9. (1) *Let* Δ *be a Delzant polygon with three edges. Then there exists a unique* $\lambda > 0$ *such that* Δ *is* $\mathrm{AGL}(2, \mathbb{Z})$*-congruent to the Delzant triangle* Δ_λ. (See Example 2.5.)

(2) *Let* Δ *be a Delzant polygon with four or more edges. Let* s *be the nonnegative integer such that the number of edges is* $4 + s$. *Then there exist positive numbers* $a \geq b > 0$, *an integer* $0 \leq k \leq 2a/b$, *and positive numbers* $\delta_1, \ldots, \delta_s$, *such that* Δ *is* $\mathrm{AGL}(2, \mathbb{Z})$*-congruent to a Delzant polygon that is obtained from the Hirzebruch trapezoid* $\mathrm{Hirz}_{a,b,k}$ (see Example 2.5) *by a sequence of corner choppings of sizes* $\delta_1, \ldots, \delta_s$.

Proof. See [Fulton 1993, Section 2.5 and Notes to Chapter 2]. $\square$

2.10. For any Delzant polygon Δ, consider the free Abelian group generated by its edges:

$$\mathbb{Z}[\text{edges of } \Delta].$$

The "length functional"

$$\mathbb{Z}[\text{edges of } \Delta] \to \mathbb{R}$$

is the homomorphism that associates to each basis element its rational length. If Δ_{i+1} is obtained from Δ_i by a corner chopping, we consider the injective homomorphism

$$\mathbb{Z}[\text{edges of } \Delta_i] \hookrightarrow \mathbb{Z}[\text{edges of } \Delta_{i+1}] \tag{1}$$

whose restriction to the generators is defined in the following way. If d is an edge of Δ that does not touch the corner that was chopped, then d is mapped to the edge of Δ_{i+1} with the same outward normal vector. If d is an edge of Δ_i that touches the corner that was chopped, then d is mapped to $d+e$ where e is the new edge of Δ_{i+1}, created in the chopping.

The definition of corner chopping in 2.6 implies that the homomorphism (1) respects the length homomorphisms.

By induction and the definition of corner chopping we get the following lemma.

Lemma 2.11. *Let*

$$\Delta_0, \Delta_1, \ldots, \Delta_s$$

be a sequence of Delzant polygons such that, for each i, the polygon Δ_i is obtained from the polygon Δ_{i-1} by a corner chopping of size δ_i.

(1) *The image of an edge d of Δ_j by $s-j$ iterations of homomorphism* (1) *is a linear combination $\sum_{i=0}^{\ell} m_i c_i$, such that $c_0, \ldots, c_\ell$ are edges of Δ_s whose union U_d is connected, $\ell \leq (s-j)$, and for $0 \leq i \leq \ell$, the coefficient m_i is a nonnegative integer that is less than or equal to 2^{s-j}; we say that d is* given *by the chain U_d with multiplicities $m_0, \ldots, m_\ell$.*
(2) area $\Delta_s =$ area $\Delta_0 - \frac{1}{2}\delta_1^2 - \cdots - \frac{1}{2}\delta_s^2$.
(3) perimeter $\Delta_s =$ perimeter $\Delta_0 - \delta_1 - \cdots - \delta_s$.

Lemma 2.12. *Let (M, ω, Φ) be a four-dimensional symplectic toric manifold, with moment-map polygon Δ of n edges. Then there are $n-2$ edges of Δ whose union is connected, such that the classes of their Φ-preimages form a basis to $H_2(M;\mathbb{Z})$. Moreover, for any $n-2$ edges of Δ whose union is connected, the classes of their preimages form a basis to $H_2(M;\mathbb{Z})$.*

Proof. By Lemma 2.9, we can prove this by induction. In the induction step, suppose that $(\widetilde{M}, \tilde{\omega}, \tilde{\Phi})$ with moment map polygon $\widetilde{\Delta}$ of $n+1$ edges is obtained by a toric blowup of (M, ω, Φ) with moment map polygon Δ. Let B_Δ be a set of $n-2$ edges of Δ whose union is connected, such that the classes of their Φ-preimages form a basis to $H_2(M; \mathbb{Z})$. If B_Δ consists of an edge that touches the corner that was chopped, set $B_{\widetilde{\Delta}}$ to be the edges of $\widetilde{\Delta}$ with the same outward normal vector as the edges in B_Δ plus the new edge e of $\widetilde{\Delta}$, created in the chopping. If none of the edges in B_Δ touches the corner that was chopped, set $B_{\widetilde{\Delta}}$ to be the edges of $\widetilde{\Delta}$ with the same outward normal vector as the edges in B_Δ plus one of the edges adjacent to e in $\widetilde{\Delta}$. $\square$

Corollary 2.13. *Let (M, ω, Φ) be a four-dimensional symplectic toric manifold, with moment-map polygon Δ. The number of edges of Δ is equal to the second Betti number* $\dim H_2(M)$ *plus two.*

By the Delzant theorem, every toric action on $\mathbb{CP}^2$ is obtained from a symplectomorphism of $\mathbb{CP}^2$ with a symplectic toric manifold M_Δ that is associated to a Delzant polygon Δ. By Corollary 2.13, Δ must be a triangle. By part (1) of Lemma 2.9, Δ is $\mathrm{AGL}(2, \mathbb{Z})$-congruent to a Delzant triangle Δ_λ. (See Example 2.5.) By our normalization convention for the Fubini–Study form, $\lambda = 1$. It follows that:

Lemma 2.14. *Every toric $\mathbb{T}^2$-action on $\mathbb{CP}^2$ is equivariantly symplectomorphic to the standard action.*

3. *J*-holomorphic spheres in symplectic 4-manifolds

In this section we will highlight results from the theory of J-holomorphic curves that we will use for the proof of Lemma 4.3, and to show uniqueness of symplectic blowdowns in the appendix.

Let (M, ω) be a compact symplectic manifold. Let $\mathcal{J} = \mathcal{J}(M, \omega)$ be the space of almost complex structures on M that are compatible with ω. The space $\mathcal{J}$ is contractible [McDuff and Salamon 1998]. Given $J \in \mathcal{J}$, a *parametrized J-holomorphic sphere* is a map $u : \mathbb{CP}^1 \to M$, such that $du : T\mathbb{CP}^1 \to TM$ satisfies the Cauchy–Riemann equation $du \circ i = J \circ du$. Such a u represents a homology class in $H_2(M; \mathbb{Z})$ that we denote $[u]$. A J-holomorphic sphere is called *simple* if it cannot be factored through a branched covering of $\mathbb{CP}^1$. One similarly defines a holomorphic curve in (M, J) whose domain is a Riemann surface other than $\mathbb{CP}^1$.

For any class $A \in H_2(M; \mathbb{Z})$, consider the universal moduli space of simple parametrized holomorphic spheres in the class A,

$$\mathcal{M}(A, \mathcal{J}) = \{(u, J) \mid J \in \mathcal{J}, u : \mathbb{CP}^1 \to M \text{ is simple } J\text{-holomorphic, and } [u] = A\},$$

and the projection map

$$p_A : \mathcal{M}(A, \mathcal{J}) \to \mathcal{J}.$$

For a fixed $J \in \mathcal{J}$, we denote by $\mathcal{M}(A, J)$ the space $p_A^{-1}(J)$.

The automorphism group $\mathrm{PSL}(2, \mathbb{C})$ of $\mathbb{CP}^1$ acts on $\mathcal{M}(A, \mathcal{J})$ by reparametrizations. The quotient $\mathcal{M}(A, \mathcal{J}) / \mathrm{PSL}(2, \mathbb{C})$ is the space of unparametrized J-holomorphic spheres representing $A \in H_2(M)$.

Lemma 3.1. *Let $0 \neq A \in H_2(M; \mathbb{Z})$. The action of $G = \mathrm{PSL}(2, \mathbb{C})$ on $\mathcal{M}(A, \mathcal{J})$ is free and proper.*

Proof. For any sphere $u \in \mathcal{M}(A, \mathcal{J})$, the stabilizer $G_u = \{\psi \in G \mid u \circ \psi = u\}$ is trivial, since u is simple; this proves that the action is free.

We now need to show that the action map $(u, \psi) \mapsto (u, u \circ \psi)$ is proper. Let $K \subset \mathcal{M}(A, \mathcal{J}) \times \mathcal{M}(A, \mathcal{J})$ be a compact subset. Without loss of generality $K = K_1 \times K_2$, where K_1 and K_2 are compact in $\mathcal{M}(A, \mathcal{J})$. Because $\mathcal{M}(A, \mathcal{J})$ is Hausdorff and first countable, it is enough to show that for every sequence $\{(u_n, \psi_n)\}$ in the preimage of $K_1 \times K_2$ there exists a subsequence such that $\{\psi_n\}$ converges uniformly and $\{u_n\}$ converges in the C^∞ topology. Take such a sequence $\{u_n, \psi_n\}$. Because $u_n \in K_1$ and K_1 is compact, after passing to a subsequence we may assume that $\{u_n\}$ C^∞-converges.

By [McDuff and Salamon 2004, Lemma D.1.2], if the sequence ψ_n does not have a uniformly convergent subsequence, then there exist points $x, y \in \mathbb{CP}^1$ and a subsequence ψ_μ which converges to the point y uniformly in compact subsets of $\mathbb{CP}^1 \setminus \{x\}$. In particular ψ_μ converges to a point on a half sphere, hence $u_\mu \circ \psi_\mu$, restricted to a half sphere, converge to a constant map. However, the sequence of holomorphic spheres $\{u_n \circ \psi_n\}$, (as a sequence in the compact subset K_2 of $\mathcal{M}(A, \mathcal{J})$), has a C^∞-convergent (hence u.c.s.-convergent) subsequence whose limit is in the nontrivial homology class A, and we get a contradiction. $\square$

Gromov [1985] introduced a notion of weak convergence of a sequence of holomorphic curves. This notion is preserved under reparametrization of the curve, and it implies convergence in homology. *Gromov's compactness theorem* guarantees that, given a converging sequence of almost complex structures, a corresponding sequence of holomorphic curves with bounded symplectic area has a weakly converging subsequence. The limit under weak convergence might not be a curve; it might be a cusp curve, that is, a connected union of holomorphic curves. As a result of Gromov's compactness, we have the following lemma.

Lemma 3.2. *Let $\{J_n\} \subset \mathcal{J}$ be a sequence of almost complex structures that converges in the C^∞ topology to an almost complex structure $J_\infty \in \mathcal{J}$. For each n, let $f_n : \mathbb{CP}^1 \to M$ be a parametrized J_n-holomorphic sphere. Suppose that the set of areas $\omega([f_n])$ is bounded. Then one of the following two possibilities occurs.*

(1) *There exist a J_∞-holomorphic sphere $u : \mathbb{CP}^1 \to M$ and elements $A_n \in \mathrm{PSL}(2, \mathbb{C})$ such that a subsequence of the $f_n \circ A_n$'s converges to u in the C^∞ topology. In particular, there exist infinitely many n's for which $[f_n] = [u]$.*

(2) *There exist two or more J_∞-holomorphic spheres $u_\ell : \mathbb{CP}^1 \to M$ that are non-constant and simple and positive integers m_ℓ, for $\ell = 1, \dots, L$, and infinitely many n's for which*

$$[f_n] = \sum_{\ell=1}^{L} m_\ell [u_\ell] \quad \text{in } H_2(M).$$

For details, see [Karshon et al. 2007, Lemma A.3].

In the proof of Lemma 4.3, we will use the following Lemma.

Lemma 3.3. *Let (M, ω) be a closed symplectic four-manifold. Let $E \in H_2(M; \mathbb{Z})$ be a homology class that can be represented by an embedded symplectic sphere and such that $c_1(TM)(E) = 1$. Then for every $J \in \mathcal{J}$ there exists a J-holomorphic cusp curve in the class E.*

To deduce the lemma from Gromov's compactness we need the existence of a dense set $U \subset \mathcal{J}$ such that for any $J \in U$, the class E is represented by an embedded J-holomorphic sphere.

For any positive number K, let

$$\mathcal{N}_K = \{A \in H_2(M; \mathbb{Z}) \mid A \neq 0, c_1(TM)(A) \leq 0, \text{ and } \omega(A) < K\}.$$

The importance of this set lies in the fact that if a homology class E with $\omega(E) \leq K$ and $c_1(TM)(E) \leq 1$ is represented by a J-holomorphic cusp curve with two or more components, then at least one of these components must lie in a homology class in $\mathcal{N}_K$; see Lemma A.5 in [Karshon et al. 2007]. Let

$$U_K = \mathcal{J} \setminus \bigcup_{A \in \mathcal{N}_K} \text{image } p_A.$$

Let (M, ω) be a compact symplectic four-manifold. Then the subset $U_K \subset \mathcal{J}$ is open, dense, and path connected. This is proved in [McDuff 1990, Lemma 3.1; 1991, Section 3] and presented in [Karshon et al. 2007, Lemma A.8 and Lemma A.10]. The following is also shown in [Karshon et al. 2007, Lemma A.12].

Lemma 3.4. *Let (M, ω) be a compact symplectic four-manifold. Let $E \in H_2(M)$ be a homology class that can be represented by an embedded symplectic sphere and such that $c_1(TM)(E) = 1$.*

(1) *The projection map $p_E : \mathcal{M}(E, \mathcal{J}) \to \mathcal{J}$ is open.*

(2) *Let $K \geq \omega(E)$. Then, for any $J \in U_K$, the class E is represented by an embedded J-holomorphic sphere.*

Lemma 3.3 now follows.

For the proof of Theorem 1.1, we also need the following lemmas.

Lemma 3.5. *Let (M, ω) be a compact symplectic four-manifold. Let $A \in H_2(M; \mathbb{Z})$ be a homology class which is represented by an embedded symplectic sphere C.*

(1) *There exists an almost complex structure $J_0 \in \mathcal{J}$ for which C is a J_0-holomorphic sphere.*

(2) *For any $J \in \mathcal{J}$ and any simple parametrized J-holomorphic sphere $f : \mathbb{CP}^1 \to M$ in the class A, the map f is an embedding.*

plus 1pt plus 1pt The lemma is a consequence of the adjunction formula. For details and references see, for example, [Karshon and Kessler 2007, Lemma 5.3].

Lemma 3.6. *Let (M, ω) be a compact symplectic four-manifold. Let $A \in H_2(M; \mathbb{Z})$ be a homology class that is represented by an embedded symplectic sphere, and such that $c_1(TM)(A) = 1$. Let $J \in \operatorname{image} p_A$, and $(u, J) \in \mathcal{M}(A, J)$.*

If $A = \sum_{i=1}^n m_i[u_i]$, where each component u_i is a simple parametrized J-holomorphic sphere and $m_i \in \mathbb{N}$, then all the components but one must be constants, and the nonconstant component differs from u by reparametrization of $\mathbb{CP}^1$.

Proof. By Lemma 3.5, u is an embedding, so the adjunction equality

$$0 = 2 + A \cdot A - c_1(TM)(A)$$

holds; since $c_1(TM)(A) = 1$ this implies $A \cdot A = -1$. If $n > 1$ and there is more than one nonconstant component, then for $1 \le i \le n$, $\omega([u]) > \omega([u_i])$ so $u \neq u_i$, hence by positivity of intersections of J-holomorphic spheres in a four-manifold [McDuff and Salamon 2004, Theorem 2.6.3], $[u_i] \cdot [u] \ge 0$. Thus $0 \le \sum_{i=1}^n m_i([u_i] \cdot [u]) = A \cdot A$, in contradiction to $A \cdot A = -1$.

Thus, all the components but one must be constants. By a similar argument, the nonconstant component differs from u at most by reparametrization of $\mathbb{CP}^1$. □

Lemma 3.7. *Let (M, ω) be a closed symplectic four-manifold. Let $E \in H_2(M; \mathbb{Z})$ be a homology class that can be represented by an embedded symplectic sphere and such that $c_1(TM)(E) = 1$. Let*

$$U_E = \operatorname{image} p_E.$$

(1) *$U_E \subset \mathcal{J}$ is open, dense, and path connected. Between any two elements in U_E there is a path in U_E that is transversal to p_E.*

(2) *The map*

$$\widetilde{p_E} : \mathcal{M}(E, \mathcal{J}) / \operatorname{PSL}(2, \mathbb{C}) \to U_E$$

induced from the projection map p_E is proper.

(3) *For $J_0, J_1 \in U_E$, the sets $\mathcal{M}(E, J_0)/\operatorname{PSL}(2, \mathbb{C})$ and $\mathcal{M}(E, J_1)/\operatorname{PSL}(2, \mathbb{C})$ consist each of a single point, and there exists a path $\{J_t\}_{0 \le t \le 1}$ such that*

$$\mathcal{W}(E; \{J_t\}) = \{(u_t, J_t) \mid u_t \in \mathcal{M}(E, J_t)\} / \operatorname{PSL}(2, \mathbb{C})$$

is a compact one-dimensional manifold, and each u_t is an embedding.

Proof. (1) Since p_E is an open map by Lemma 3.4(1), its image U_E is an open set in $\mathcal{J}$. Set $K = \omega(E)$. By part (2) of Lemma 3.4, $U_K \subseteq U_E$. Since U_K is dense in $\mathcal{J}$, so is U_E. Since U_E is open, $\mathcal{J}$ locally path connected, and U_K is dense in U_E and path connected, we get that U_E is path connected.

By the regularity criterion of Hofer–Lizan–Sikorav [1997], any element in U_E is a regular value for p_E. A path between regular values for p_E can be perturbed to a path with the same endpoints that is transversal to p_E; see [McDuff and Salamon 2004, Theorem 3.1.7(ii); Karshon et al. 2007, Lemma A.9(d)].

(2) This follows from Gromov's compactness in the following way. Let $D \subset U_E$ be a compact subset. We need to show that $p_E^{-1}(D)/\,\mathrm{PSL}(2, \mathbb{C})$ is compact. Because $\mathcal{M}(E, \mathcal{J})$ is Hausdorff and first countable, it is enough to show that for every sequence $\{(f_n, J_n)\}$ in $p_E^{-1}(D)$ there exists a subsequence that, after reparametrization, has a limit in $p_E^{-1}(D)$ in the C^∞ topology.

Take such a sequence, $\{(f_n, J_n)\}$. Because $J_n \in D$ and D is compact and contained in U_E, after passing to a subsequence we may assume that $\{J_n\}$ converges to $J_\infty \in U_E$. Each f_n is a J_n-holomorphic sphere in the class E. Suppose that there exists a subsequence that, after reparametrization, converges to some $u : \mathbb{CP}^1 \to M$ in the C^∞ topology. Then u must be in the class E and it must be J_∞-holomorphic. If u is not simple, we get a contradiction to Lemma 3.6. Then the pair (u, J_∞) is in the moduli space $\mathcal{M}(E, \mathcal{J})$, and since $J_\infty \in D$, this pair is in $p_E^{-1}(D)$.

Now suppose that there does not exist such a subsequence. Then there exist two or more nonconstant simple J_∞-holomorphic spheres $u_\ell : \mathbb{CP}^1 \to M$ and positive integers m_ℓ such that $\sum m_\ell [u_\ell] = E$, by Lemma 3.2. This contradicts Lemma 3.6.

(3) For $J \in U_E = \text{image}\, p_E$, the set $\mathcal{M}(E, J) = p_E^{-1}(J) \neq \varnothing$. Hence, by Lemma 3.6, the set $\mathcal{M}(E, J)/\,\mathrm{PSL}(2, \mathbb{C})$ consists of a single point. For $J_0, J_1 \in U_E$, by part (1), there is a path $\{J_t\}$ in U_E from J_0 to J_1, that is transversal to p_E. Hence, by [McDuff and Salamon 2004, Theorem 3.1.7], $\mathcal{W}^*(E; \{J_t\}) = \{(u_t, J_t) \mid u_t \in \mathcal{M}(E, J_t)\}$ is a manifold of dimension $1 + 6 = 1 + \operatorname{index} p_E$. By Lemma 3.1, the action of $\mathrm{PSL}(2, \mathbb{C})$ on $\mathcal{W}^*(E; \{J_t\})$ is free and proper, thus

$$\mathcal{W}(E; \{J_t\}) = \{(u_t, J_t) \mid u_t \in \mathcal{M}(E, J_t)\}/\,\mathrm{PSL}(2, \mathbb{C})$$

is a manifold of dimension one. $\mathcal{W}(E; \{J_t\})$ is the inverse image of the path $\{J_t\}$ under the map $\widetilde{p_E}$, hence, by part (2), it is compact.

By Lemma 3.5, each u_t is an embedding. □

4. Representing J_T-holomorphic curves on the moment map polygon

Notation. For (M, ω, Φ), let J_T denote a $\mathbb{T}^n$-invariant complex structure on M that is compatible with ω. By Delzant's construction [1988], such a structure exists.

Claim 4.1. *Let (M, ω, Φ) be a four-dimensional symplectic toric manifold, with moment-map polygon Δ. The preimage under Φ of an edge d of Δ is an embedded J_T-holomorphic sphere.*

Proof. By part (2) of Lemma 2.4, $Y = \Phi^{-1}(d)$ is a symplectically embedded 2-sphere in M. Being a connected component of a fixed point set of a holomorphic S^1-action, $TY = J_T TY$. As an almost complex manifold of real dimension two, $(Y, J_T|_{TY})$ is a complex manifold. Thus the embedded sphere Y is an embedded holomorphic sphere in the complex manifold (M, J_T). □

Lemma 4.2. *Let (M, ω, Φ) be a four-dimensional symplectic toric manifold, with moment-map polygon Δ.*

- *Any J_T-holomorphic sphere is homologous in $H_2(M; \mathbb{Z})$ to a linear combination with coefficients in $\mathbb{N}$ of inverse images under Φ of edges of Δ.*
- *For any set S of $n-2$ edges whose union is connected, any simple J_T-holomorphic sphere C that is not the preimage of an edge of Δ is homologous to a linear combination with coefficients in $\mathbb{N}$ of preimages of edges of Δ whose union is connected and that are contained in S; if the intersection of C with each of the two edges of Δ that are not in S is positive, then all the $n-2$ edges of S appear with positive coefficients in this linear combination.*

Proof.

- Let Ψ be an S^1-moment map obtained by composing Φ with projection in a rational direction along which there is not any edge of Δ. Denote by $v_{\min}$ ($v_{\max}$) the vertex of minimal (maximal) value of that projection. Let $D_1, \dots, D_m$ be a chain of Φ-preimages of edges between $v_{\min}$ and $v_{\max}$. Let $D'_1, \dots, D'_{m'}$ be the other chain of Φ-preimages of edges between $v_{\min}$ and $v_{\max}$.

  Without loss of generality we assume that C is a simple J_T-holomorphic sphere that is not the Φ-preimage of an edge of Δ. By Lemma 2.12, in $H_2(M; \mathbb{Z})$

  $$[C] = \sum_{i=1}^{m} a_i D_i + \sum_{j=1}^{m'} b_j D'_j, \quad \text{with } a_1 = b_1 = 0.$$

  Adapting the proof of Lemma C.6 in [Karshon 1999] we get that

  $$a_{i+1}/k_{i+1} \geq a_i/k_i \geq 0, \quad \text{for } 1 \leq i < m \quad (1 \leq i < m'), \tag{2}$$

  where k_i is the order of the stabilizer of the i-th sphere in a chain.

  Notice that (2) implies that

  $$a_\ell > 0 \Rightarrow a_i > 0 \text{ for all } \ell \leq i \leq m,$$
  $$b_\ell > 0 \Rightarrow b_j > 0 \text{ for all } \ell \leq j \leq m'.$$

Hence, C is homologous in $H_2(M;\mathbb{Z})$ to a linear combination with coefficients in $\mathbb{N}$ of inverse images under Φ of, at most $n-2$, edges of Δ whose union is connected.

- It is enough to observe that for any set S of $n-2$ edges whose union is connected, there is an S^1-moment map Ψ, obtained by composing Φ with projection in a rational direction along which there is not any edge of Δ, such that the vertex $v_{\min}$ is the vertex between the two edges of Δ that are not in S. Then the previous proof gives the required. □

Lemma 4.3. *Let (M,ω,Φ) be a four-dimensional symplectic toric manifold with moment-map polygon Δ. Let J_T be a $\mathbb{T}^2$-invariant ω-compatible complex structure on M, and g_T be the Riemannian metric defined by (ω, J_T). Let $\mathfrak{i}^*$ be a projection in a rational direction along which there is not any edge of Δ. Let $v_{\min}$ be the vertex of Δ of minimal value of that projection.*

Let C be a J_T-holomorphic sphere such that $\Phi(C)$ avoids the vertex $v_{\min}$. Let α and β be the points of $\Phi(C)$ on the boundary of Δ, that are closest to $v_{\min}$ from left and right. Let v_α and v_β be the vertices following α and β. Then the gradient flow of $\Psi=\mathfrak{i}^*\circ\Phi$ with respect to g_T carries C to a family of J_T-holomorphic spheres; this family weakly converges to a connected union of preimages of edges of Δ (maybe with multiplicities). These edges form a chain that we denote L_C. The vertices of L_C closest to $v_{\min}$ from left and right are v_α and v_β.

Proof. The function $\Psi=\mathfrak{i}^*\circ\Phi : M\to\mathbb{R}$ is a moment map associated with a Hamiltonian action on (M,ω) of S^1 embedded in $\mathbb{T}^2$ by $\mathfrak{i}: S^1\hookrightarrow\mathbb{T}^2$.

Let ξ_M be the vector field generating the S^1-action. The gradient flow η_t of Ψ with respect to the invariant metric g_T is generated by $-J_T\xi_M$. This flow is equivariant with respect to the action, that is, for each t, the diffeomorphism $\eta_t: M\to M$ is $\mathbb{T}^2$-equivariant. Consequently, it sends a set that is a Φ-preimage of a vertex or a Φ-preimage of an edge to itself.

Set L to be the chain of edges of Δ that do not touch $v_{\min}$. Let

$$B=\{p\in M : i^*\circ\Phi(p)>r\}$$

for some $i^*(v_{\min})<r<\min\{i^*(v'), i^*(v'')\}$, where v' (v'') is the vertex following $v_{\min}$ immediately from the left (right). Then $\bigcap_{t>0}(\eta_t(B))\supseteq\Phi^{-1}L$. On the other hand, a point $p\in B$ that is not in $\Phi^{-1}(L)$, is sent to $v_{\min}$ by the gradient flow η_t as $t\to-\infty$, that is, for t' big enough, $q=\eta_{-t'}(p)$ is not in B. Since $\eta_{t'}$ is a diffeomorphism, there cannot be $b\in B$ such that $\eta_{t'}(b)=\eta_{t'}(q)=\eta_0(p)=p$, in particular, p is not in the intersection $\bigcap_{t>0}(\eta_t(B))$. So

$$\bigcap_{t>0}(\eta_t(B))=\Phi^{-1}(L).$$

We choose B big enough such that, for some complex coordinates, the complexified toric action on $M - B$ is the standard action of the complex torus on an open subset of $\mathbb{C}^2$. In particular, for t_1, t_2 close to 0, if $t_1 > t_2 > 0$, $\eta_{t_1}(M - B) \supset \eta_{t_2}(M - B)$, hence $\eta_{t_1}(B) \subset \eta_{t_2}(B)$. Since η_t is a flow, (that is, a homomorphism from $(\mathbb{R}, +)$ to $(\mathrm{Diff}, \circ)$), this implies that for any $t_1 > t_2 > 0$, $\eta_{t_1}(B) \subset \eta_{t_2}(B)$, that is, η_t is monotonic on B.

Now, choose B such that, in addition to the above, its image contains $\Phi(C)$. Consider a sequence $\{C_i\}$, where $C_i = \eta_i(C)$, with discrete $i \to \infty$. Each C_i is a J_T-holomorphic sphere in the homology class $[C]$. By Gromov's compactness theorem, there is a subsequence $\{C_\mu\}$ that weakly converges to a J_T-holomorphic (maybe nonsmooth) cusp curve C' in $[C]$. In particular, each point in the limit C' is the limit of a sequence of points in $\{C_\mu\}$, hence, since $C_\mu = \eta_\mu(C) \subset \eta_\mu(B)$, and η_t is monotonic on B, we get that $C' \subset \cap_\mu(\eta_\mu(B)) \subset \Phi^{-1}(L)$. Thus, since each edge preimage is an irreducible J_T-holomorphic sphere in the complex manifold (M, J_T) (by Claim 4.1), the irreducible components of C' are preimages of edges in L. We conclude that the cusp curve C' is a connected union of preimages of the edges of a subchain L_C of L, with positive multiplicities.

Let p_α (p_β) be the preimage of v_α (v_β) in M. The chain L_C includes v_α and v_β, as the limits of $\eta_\mu(p_\alpha)$ and $\eta_\mu(p_\beta)$. Assume a vertex v on L_C is closer to $v_{\min}$ from the left than v_α. Let e_v be the edge that touches v from below. Then L_C intersects e_v at v, hence $\Phi^{-1}(L_C)$ intersects $\Phi^{-1}(e_v)$ at the point $\Phi^{-1}(v)$, maybe with multiplicities. However $C \cap \Phi^{-1}(e_v) = \varnothing$, in contradiction to $[\Phi^{-1}(L_C)] = [C]$. Similarly, the vertex on L_C closest to $v_{\min}$ from the right is v_β. □

Claim 4.4. *Let (M, ω, Φ) be a four-dimensional symplectic toric manifold with moment-map polygon Δ.*

Every J_T-cusp curve C is homologous in $H_2(M; \mathbb{Z})$ to a linear combination with coefficients in $\mathbb{N}$ of preimages of edges of Δ whose union is connected. In particular, C is homologous to a $\mathbb{T}^2$-invariant J_T-cusp curve.

We already know that a J_T-cusp curve C is homologous to a linear combination with coefficients in $\mathbb{N}$ of preimages of edges of Δ (by applying the first part of Lemma 4.2 to the components of the cusp curve). However, the union of these edges might not be connected. The "connected" part that we add here plays an important role in the proof of Theorem 1.1.

Proof. Let i* be a projection in a rational direction along which there is not any edge of Δ. Let $v_{\min}$ be the vertex of Δ of minimal value of i*. If for any component of C that is not a Φ-preimage of an edge of Δ, the moment map image avoids a neighbourhood of $v_{\min}$, then the claim follows from Lemma 4.3 (and the fact that C is connected). Otherwise, there is such a component D; by positivity of intersections, the intersection number of D with the preimage of each of the two edges adjacent to $v_{\min}$ is positive. Thus, by the second part of Lemma 4.2, D is homologous to a linear combination with coefficients in $\mathbb{N}$ of Φ-preimages of all the edges of Δ but the two adjacent to $v_{\min}$. By the first part of Lemma 4.2, each component of C is homologous to a linear combination of Φ-preimages of edges of Δ with coefficients in $\mathbb{N}$. Combining such representatives of D and the other components of C gives the claim. □

Lemma 4.5. *Let (M, ω, Φ) be a symplectic toric four-manifold with moment map polygon Δ. Let C be an embedded symplectic sphere in (M, ω) which satisfies $c_1(TM)(C) = 1$.*

Then C is homologous in $H_2(M; \mathbb{Z})$ to a linear combination with coefficients in $\mathbb{N}$ of preimages of edges of Δ whose union is connected.

Proof. By Lemma 3.3 there exists a J_T-holomorphic cusp curve in the class $[C]$. Now apply Claim 4.4. □

5. No toric action on (M_k, ω_ϵ) for $k > 3$ and small ϵ

For $\epsilon > 0$, denote by

$$(M_k, \omega_\epsilon)$$

a symplectic manifold that is obtained from $(\mathbb{CP}^2, \omega_{\mathrm{FS}})$ by k simultaneous symplectic blowups of equal sizes ϵ. For description of symplectic blowup, see 2.7. The k *simultaneous blowups* are obtained from embeddings $i_1 : \Omega_1 \to M, \dots, i_k : \Omega_k \to M$ whose images are disjoint. We denote by $E_1, \dots, E_k$ the homology classes in $H_2(M_k; \mathbb{Z})$ of the exceptional divisors obtained by the blowups, and by L the homology class of a line $\mathbb{CP}^1 \subset M_k$.

5.1. By McDuff and Polterovich [1994], for $k \leq 8$ there exists a symplectic blowup of $\mathbb{CP}^2$ k times by size ϵ if and only if ϵ satisfies the following conditions. If $k = 2, 3, 4$: $\epsilon < \frac{1}{2}$. If $k = 5, 6$: $\epsilon < \frac{2}{5}$. If $k = 7$: $\epsilon < \frac{3}{8}$. If $k = 8$: $\epsilon < \frac{6}{17}$. According to Biran [1997], for $k \geq 9$, there exist k symplectic blowups of equal sizes ϵ if and only if ϵ satisfies the volume constraint, that is, $\epsilon < 1/\sqrt{k}$.

Assume that (M_k, ω_ϵ) admits a toric action with moment map polygon Δ. By Lemma 4.5, each E_i can be represented by a linear combination with coefficients in $\mathbb{N}$ of preimages of edges of Δ. We call the union of these edges, with the $\mathbb{N}$-multiplicities, a Δ-*representative* of E_i. If this union is connected, we call it a *connected* Δ-*representative*. We observe the following properties of Δ-representatives of $E_1, \dots, E_k$.

Claim 5.2. *Assume that (M_k, ω_ϵ) admits a toric action with moment map image Δ. Choose Δ-representatives for $E_1, \dots, E_k$. For $m \leq k$, the number of edges in the union of the Δ-representatives of m different E_i's is $> m$, unless each of these Δ-representatives is a single edge with multiplicity one.*

Proof. Assume that the union of the chosen Δ-representatives of $E_1, \dots, E_m$ (without loss of generality) is a subset of the set of edges $C_1, \dots, C_m$, that is, in $H_2(M_k; \mathbb{Z})$, for $1 \le i \le m$,

$$E_i = \sum_{j=1}^{m} a^i_j [\Phi^{-1} C_j], \quad a^i_j \in 0 \cup \mathbb{N}. \tag{3}$$

Denote by A the $m \times m$ matrix of the coefficients a^i_j. Since the homology classes $E_1, \dots, E_m$ are independent, the matrix A is invertible (over $\mathbb{R}$). We get that

$$([\Phi^{-1} C_1], \dots, [\Phi^{-1} C_m])^t = A^{-1} (E_1, \dots, E_m)^t. \tag{4}$$

The homology classes $L, E_1, \dots, E_k$ form a basis of $H_2(M_k; \mathbb{Z})$, therefore each $[\Phi^{-1} C_j] = d_j L + \sum_i b^j_i E_i$, with unique integers as coefficients. The coefficients do not change if we write $[\Phi^{-1} C_j]$ as a linear combination of L, E_i in $H_2(M_k; \mathbb{R})$. By this and (4), all the entries of A^{-1} are in $\mathbb{Z}$, so in $H_2(M_k; \mathbb{Z})$,

$$[\Phi^{-1} C_j] = \sum_{i=1}^{m} b^j_i E_i, \quad b^j_i \in \mathbb{Z}.$$

Since the size of each E_i is ϵ we deduce that the length $|C_j|$ of each C_j is an integer multiple of ϵ. Since $|C_j| > 0$, it must be a multiple of ϵ by $N_j \in \mathbb{N}$. However, by (3), for $1 \le i \le m$,

$$\epsilon = \sum_{j=1}^{m} a^i_j |C_j|, \quad a^i_j \in 0 \cup \mathbb{N}.$$

Thus

$$\epsilon = \sum_{j=1}^{m} a^i_j N_j \epsilon, \quad a^i_j \in 0 \cup \mathbb{N}, \; N_j \in \mathbb{N}.$$

We get that in each line (and each column) of (the invertible matrix) A there is 1 in one entry and 0 in each of the other entries, that is, each of the Δ-representatives is a single edge with multiplicity one. □

Claim 5.3. *Assume that (M_k, ω_ϵ) admits a toric action with moment map image Δ. Choose connected Δ-representatives for $E_1, \dots, E_k$. Denote their union by U. If none of the chosen connected Δ-representatives is a single edge of Δ with multiplicity one, then U is connected and consists of at least $k+1$ edges.*

Proof. By Claim 5.2, U consists of more than k edges. Assume that U is disconnected. Then it consists of at most $k+1$ edges, hence it consists of exactly $k+1$ edges out of the $k+3$ edges of Δ. Since none of the Δ-representatives is a single edge, Claim 5.2 implies that the m_j edges of a connected component j support at most $m_j - 1$ of the E_i's. Thus the nonconnected $k+1$ edges support at most

$\sum_{j=1}^{c}(m_j - 1) = k + 1 - c < k$ of these classes, where $c > 1$ is the number of connected components, and we get a contradiction. $\square$

For a convex polygon Δ in $\mathbb{R}^2$, we denote by

$$(M_\Delta, \omega_\Delta, \Phi_\Delta)$$

a symplectic toric manifold whose moment map image is Δ.

The main ingredient of the proof of Theorem 1.1 is:

Claim 5.4. *If (M_k, ω_ϵ) is symplectomorphic to $(M_\Delta, \omega_\Delta)$, and*

$$\epsilon \leq \frac{1}{3k2^{2k}},$$

then one of the classes $E_1, \dots, E_k$ is realized by an embedded $\mathbb{T}^2$-invariant symplectic exceptional sphere; equivariantly blowing down along it yields $(M_{k-1}, \omega_\epsilon)$ with a toric action.

Proof. If $k \geq 1$, the moment map image Δ is a Delzant polygon of $k + 3 \geq 4$ edges, so by Lemma 2.9, up to $\mathrm{AGL}(2, \mathbb{Z})$-congruence, it is obtained by $(k - 1)$ corner-choppings of sizes $(\delta_1, \dots, \delta_{k-1})$ from a standard Hirzebruch trapezoid Σ with west and east edges F_w, F_e, south edge S, north edge N, and slope $-1/d$.

By part (1) of Lemma 2.11,

$$|S| + |N| < 2^k \text{ perimeter } \Delta, \tag{5}$$

and F_w and F_e are given by two disjoint connected unions of edges of Δ with multiplicities $\leq 2^k$.

For each class E_i, we choose a connected Δ-representative, that is a connected union of edges (with multiplicities in $\mathbb{N}$) whose preimage is in E_i. Assume that none of these Δ-representatives is a single edge of Δ with multiplicity one. By Claim 5.3, the union U of these Δ-representatives is connected and consists of at least $k + 1$ edges of the $k + 3$ edges of Δ. Then, (at least) one of the two chains of edges giving F_w and F_e as above is contained in U : the connected at most two edges that are not in U can overlap at most one chain giving F_w or F_e, since the two chains are separated at each end by an edge. Thus

$$|F| = |F_w| = |F_e| \leq 2^k k\epsilon. \tag{6}$$

Then

$$\begin{aligned}\tfrac{1}{2}(1 - k\epsilon^2) = \text{area } \Delta &= \tfrac{1}{2}(|S| + |N|)|F| - \sum_{i=1}^{k-1} \tfrac{1}{2}{\delta_i}^2 \leq \tfrac{1}{2}2^k |F| \text{ perimeter } \Delta - \sum_{i=1}^{k-1} \tfrac{1}{2}{\delta_i}^2 \\ &= \tfrac{1}{2}2^k(3 - k\epsilon)|F| - \sum_{i=1}^{k-1} \tfrac{1}{2}{\delta_i}^2 \leq \tfrac{1}{2}2^k(3 - k\epsilon)2^k k\epsilon - \sum_{i=1}^{k-1} \tfrac{1}{2}{\delta_i}^2.\end{aligned}$$

The first (in)equality holds by part (4) of Lemma 2.4, the second holds by part (2) of Lemma 2.11, the third inequality holds by Equation (5), the fourth follows from part (3) of Lemma 2.4 and the fact that the Poincare dual to $c_1(TM_k)$ equals $3L - \sum_{i=1}^{k} E_i$, and the last holds by Equation (6).

We get that

$$1 - k\epsilon^2 \le 2^{2k}(3 - k\epsilon)(k\epsilon) \le 2^{2k}(3k\epsilon - k\epsilon^2).$$

So, $1 \le 2^{2k}3k\epsilon - k\epsilon^2(2^{2k} - 1)$, thus $1 < 3k2^{2k}\epsilon$, in contradiction to the assumption on ϵ. Therefore, (at least) one of the classes $E_1, \ldots, E_k$ is represented by the inverse image under the moment map of a single edge of Δ with multiplicity one. By Claim 4.1, such a representative C_T is an embedded J_T-holomorphic sphere. It is $\mathbb{T}^2$-invariant: let $a \in \mathbb{T}^2$; because $\mathbb{T}^2$ is connected, $[aC_T] = [C_T]$; by positivity of intersections and since $E_i \cdot E_i = -1$, aC_T and C_T must coincide. Because C_T is an embedded J_T-sphere and J_T is compatible with ω_ϵ, C_T is symplectic.

Without loss of generality, the class E_1 is represented by such a J_T-holomorphic sphere C_T. Set J_0 to be an almost complex structure on (M_k, ω_ϵ) for which the exceptional divisors obtained by the symplectic blowups are disjoint embedded J_0-holomorphic spheres $S_1, \ldots, S_k$ that represent the classes $E_1, \ldots, E_k$. (Such a structure exists by Lemma 3.5.) By Lemma A.1 in the appendix, the symplectic manifold resulting from (M_k, ω_ϵ) by blowing down along C_T is symplectomorphic to the symplectic manifold obtained by blowing down along S_1, which is $(M_{k-1}, \omega_\epsilon)$. □

Proof of Theorem 1.1. Assume that (M_k, ω_ϵ) is symplectomorphic to $(M_\Delta, \omega_\Delta)$ and $\epsilon \le 1/(3k2^{2k})$. After k iterations of Claim 5.4, we get $\mathbb{CP}^2$ with a toric action. By Lemma 2.14, this manifold is equivariantly symplectomorphic to $\mathbb{CP}^2$ with its standard toric action. By reversing our steps we get $\mathbb{CP}^2$ blown up equivariantly k times by equal sizes ϵ. □

Remark 5.5. Theorem 1.1 becomes false if we do not restrict ϵ. For $\epsilon > \frac{1}{2}$, let $(M_1, \omega_\epsilon, \Phi_1)$ be $\mathbb{CP}^2$ blown up equivariantly by size ϵ. The moment map image is obtained by chopping off a corner of size ϵ from a Delzant triangle of edge-size 1, to get a trapezoid $\mathrm{Hirz}_{(1+\epsilon)/2,1-\epsilon,1}$, that is, of height $(1-\epsilon)$, average width $(1+\epsilon)/2$, and slope -1. Let (N, ω_2, Φ_2) be a Hirzebruch surface whose image is a trapezoid $\mathrm{Hirz}_{(1+\epsilon)/2,1-\epsilon,3}$ (Notice that the north edge is then of size $2\epsilon - 1$, which is > 0 if and only if $\epsilon > \frac{1}{2}$.) See Figure 4. Since these Hirzebruch trapezoids have the same average width and height and the inverse of their slopes differ by 2, the corresponding manifolds are isomorphic as symplectic manifolds with Hamiltonian S^1-action (by [Karshon 2003, Lemma 3]), however they are not isomorphic as symplectic toric manifolds (their moment map polygons are not equivalent).

Theorem 1.1 and Lemma 2.8 yield the following corollary.

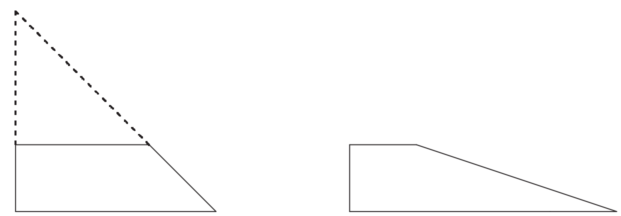

Figure 4. Symplectomorphic but not equivariantly symplectomorphic symplectic toric manifolds.

Corollary 5.6. *(M_k, ω_ϵ) with $\epsilon \leq 1/(3k2^{2k})$ admits a toric action if and only if $k \leq 3$.*

By the sharpness of the constrains listed in 5.1, when $\epsilon \leq 1/(3k2^{2k})$ there exists a symplectic blowup of $\mathbb{CP}^2$ k times by size ϵ.

Since $H^1(M_k, \mathbb{R}) = \{0\}$, any effective $(S^1)^2$-action on (M_k, ω_ϵ) is toric.

Corollary 5.7. *(M_k, ω_ϵ) with $\epsilon \leq 1/(3k2^{2k})$ admits an effective $(S^1)^2$-action if and only if $k \leq 3$.*

Appendix: Uniqueness of blowdown

Lemma A.1. *Let (M, ω) be a compact four-dimensional symplectic manifold. Let $J_0, J_1 \in \mathcal{J}$. Let A be a class in $H_2(M; \mathbb{Z})$ such that $c_1(TM)(A) = 1$ and $\omega(A) > 0$. Assume that A is represented by an embedded J_0-holomorphic sphere C_0 and by an embedded J_1-holomorphic sphere C_1.*

Then for $i = 0, 1$, there are neighbourhoods U_i of C_i, each symplectomorphic to a tubular neighbourhood of $\mathbb{CP}^1$, and a symplectomorphism ϕ of (M, ω), that sends (U_0, C_0) to (U_1, C_1), and induces the identity map on $H_2(M; \mathbb{Z})$.

Proof. By part (3) of Lemma 3.7, there is a smooth family (with parameter $0 \leq t \leq 1$) of J_t-holomorphic embeddings ρ_t from $\mathbb{CP}^1$ to the manifold. Their images are all in the homology class A. Notice that the pullbacks of ω to $\mathbb{CP}^1$ by the homotopic maps are all in the same cohomology class. Hence, by Moser, there is a family of diffeomorphisms $\phi_t : \mathbb{CP}^1 \to \mathbb{CP}^1$, starting at the identity map, that satisfy $\phi_t^*(\rho_0^*(\omega)) = \rho_t^*(\omega)$. Hence we may assume that ρ_0 is a symplectic embedding of the standard $\mathbb{CP}^1$ and compose the embeddings $\{\rho_t\}$ on the family $\{\phi_t\}$ to get a one-parameter family of symplectic embeddings of the standard $\mathbb{CP}^1$ into M. Moreover, using a parametrized version of Weinstein's tubular neighbourhood theorem, this family can be extended to a one-parameter family of symplectic embeddings σ_t of a neighbourhood of $\mathbb{CP}^1$ (as the zero-section) in the tautological bundle with a symplectic form, into M; denote the image of σ_t by U_t.

We get a "partial flow" that moves along the neighbourhoods U_t. Differentiating it by t gives vector fields X_t, defined at U_t. The Lie derivative $\mathrm{Lie}_{X_t}\,\omega$ is 0. By Cartan's formula,

$$\mathrm{Lie}_{X_t}\,\omega = d(\iota_{X_t}\omega) + (\iota_{X_t})d\omega = d\iota_{X_t}\omega,$$

where the last equality holds since ω is closed. Thus the one form $\iota_{X_t}\omega$ on U_t is closed. Therefore, and since $\mathbb{CP}^1$ is simply connected, when we consider X_t as a vector field defined at a neighbourhood of $\mathbb{CP}^1 \times [0, 1] \subseteq M \times [0, 1]$, we get a function h defined on a (maybe smaller) neighbourhood of $\mathbb{CP}^1 \times [0, 1] \subseteq M \times [0, 1]$, such that $\iota_{X_t}\omega = dh_t$. Using partition of unity in $M \times [0, 1]$, we expand h to a smooth function $H : M \times [0, 1] \to \mathbb{R}$, whose restriction to a small neighborhood of image ρ_t coincides with h_t.

This gives a Hamiltonian flow on M, thus a family of symplectomorphisms $\{\alpha_t\}_{0\leq t\leq 1}$, starting from the identity map. Take α_1 to be ϕ. $\square$

Acknowledgments

I am grateful to Yael Karshon and Leonid Polterovich for giving the right answers and asking the right questions, and to Dusa McDuff for helpful discussions.

References

[Biran 1997] P. Biran, "Symplectic packing in dimension 4", *Geom. Funct. Anal.* **7**:3 (1997), 420–437. MR 98i:57057 Zbl 0892.53022

[Delzant 1988] T. Delzant, "Hamiltoniens périodiques et images convexes de l'application moment", *Bull. Soc. Math. France* **116**:3 (1988), 315–339. MR 90b:58069 Zbl 0676.58029

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Annals of Math. Studies **131**, Princeton Univ. Press, 1993. The William H. Roever Lectures in Geometry. MR 94g:14028 Zbl 0813.14039

[Gromov 1985] M. Gromov, "Pseudoholomorphic curves in symplectic manifolds", *Invent. Math.* **82**:2 (1985), 307–347. MR 87j:53053 Zbl 0592.53025

[Guillemin and Sternberg 1982] V. Guillemin and S. Sternberg, "Convexity properties of the moment mapping", *Invent. Math.* **67**:3 (1982), 491–513. MR 83m:58037 Zbl 0503.58017

[Hofer et al. 1997] H. Hofer, V. Lizan, and J.-C. Sikorav, "On genericity for holomorphic curves in four-dimensional almost-complex manifolds", *J. Geom. Anal.* **7** (1997), 149–159. MR 2000d:32045 Zbl 0911.53014

[Karshon 1999] Y. Karshon, *Periodic Hamiltonian flows on four-dimensional manifolds*, Mem. Amer. Math. Soc. **141**, Amer. Math. Soc., Providence, R.I., 1999. MR 2000c:53113 ZBL 0982.70011

[Karshon 2003] Y. Karshon, "Maximal tori in the symplectomorphism groups of Hirzebruch surfaces", *Math. Res. Lett.* **10**:1 (2003), 125–132. MR 2004f:53101 Zbl 1036.53063

[Karshon and Kessler 2007] Y. Karshon and L. Kessler, "Circle and torus actions on equal symplectic blow-ups of $\mathbb{CP}^2$", *Math. Res. Lett.* **14**:5 (2007), 807–823. MR 2009c:53121 Zbl 1139.53041

[Karshon et al. 2007] Y. Karshon, L. Kessler, and M. Pinsonnault, “A compact symplectic four-manifold admits only finitely many inequivalent toric actions”, *J. Symplectic Geom.* **5**:2 (2007), 139–166. MR 2008m:53199 Zbl 1136.53060

[Kessler 2004] L. Kessler, *Torus actions on small blow ups of* $\mathbb{CP}^2$, thesis, The Hebrew University of Jerusalem, 2004.

[McDuff 1990] D. McDuff, “The structure of rational and ruled symplectic 4-manifolds”, *J. Amer. Math. Soc.* **3**:3 (1990), 679–712. MR 91k:58042 Zbl 0723.53019

[McDuff 1991] D. McDuff, “Blow ups and symplectic embeddings in dimension 4”, *Topology* **30**:3 (1991), 409–421. MR 92m:57039 Zbl 0731.53035

[McDuff and Polterovich 1994] D. McDuff and L. Polterovich, “Symplectic packings and algebraic geometry”, *Invent. Math.* **115**:3 (1994), 405–434. With an appendix by Y. Karshon. MR 95a:58042 Zbl 0833.53028

[McDuff and Salamon 1998] D. McDuff and D. Salamon, *Introduction to symplectic topology*, 2nd ed., Oxford Math. Monogr., The Clarendon Press, Oxford Univ. Press, New York, 1998. MR 2000g:53098 Zbl 1066.53137

[McDuff and Salamon 2004] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, Amer. Math. Soc. Colloquium Publ. **52**, Amer. Math. Soc., Providence, RI, 2004. MR 2004m:53154 Zbl 1064.53051

[Pinsonnault 2008] M. Pinsonnault, “Maximal compact tori in the Hamiltonian group of 4-dimensional symplectic manifolds”, *J. Mod. Dyn.* **2**:3 (2008), 431–455. MR MR2417479 Zbl 1154.57023

Received September 8, 2008.

LIAT KESSLER
MASSACHUSSETS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF MATHEMATICS
CAMBRIDGE, MA 02139
UNITED STATES

kessler@math.mit.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

A QUOTIENT OF THE BRAID GROUP RELATED TO PSEUDOSYMMETRIC BRAIDED CATEGORIES

FLORIN PANAITE AND MIHAI D. STAIC

Motivated by the recent concept of a pseudosymmetric braided monoidal category, we define the pseudosymmetric group PS_n to be the quotient of the braid group B_n by the relations $\sigma_i\sigma_{i+1}^{-1}\sigma_i = \sigma_{i+1}\sigma_i^{-1}\sigma_{i+1}$ with $1 \le i \le n-2$. It turns out that PS_n is isomorphic to the quotient of B_n by the commutator subgroup $[P_n, P_n]$ of the pure braid group P_n (which amounts to saying that $[P_n, P_n]$ coincides with the normal subgroup of B_n generated by the elements $[\sigma_i^2, \sigma_{i+1}^2]$ with $1 \le i \le n-2$), and that PS_n is a linear group.

Introduction

A symmetric category consists of a monoidal category $\mathscr{C}$ equipped with a family of natural isomorphisms $c_{X,Y} : X \otimes Y \to Y \otimes X$ satisfying natural "bilinearity" conditions together with the symmetry relation $c_{Y,X} \circ c_{X,Y} = \mathrm{id}_{X\otimes Y}$ for all $X, Y \in \mathscr{C}$. This concept was generalized by Joyal and Street [1993] by dropping this symmetry relation from the axioms and arriving thus at the concept of braided category, of central importance in quantum group theory; see [Kassel 1995; Majid 1995].

Inspired by recently introduced categorical concepts of pure-braided structures [Staic 2004] and twines [Bruguières 2006], Panaite, Staic and Van Oystaeyen [Panaite et al. 2009] defined the concept of pseudosymmetric braiding to generalize symmetric braidings. A braiding c on a strict monoidal category $\mathscr{C}$ is pseudo-symmetric if it satisfies the modified braid relation

$$(c_{Y,Z}\otimes\mathrm{id}_X)\circ(\mathrm{id}_Y\otimes c^{-1}_{Z,X})\circ(c_{X,Y}\otimes\mathrm{id}_Z) = (\mathrm{id}_Z\otimes c_{X,Y})\circ(c^{-1}_{Z,X}\otimes\mathrm{id}_Y)\circ(\mathrm{id}_X\otimes c_{Y,Z})$$

for all $X, Y, Z \in \mathscr{C}$. The main result in [Panaite et al. 2009] asserts that, if H is a Hopf algebra with bijective antipode, then the canonical braiding of the Yetter–Drinfeld category ${}_H\mathscr{YD}^H$ is pseudosymmetric if and only if H is commutative and cocommutative.

MSC2000: primary 20F36; secondary 18D10.
Keywords: braid group, symmetric group, braided categories, pseudosymmetric braidings.
Research partially supported by the CNCSIS project "Hopf algebras, cyclic homology and monoidal categories", contract number 560/2009.

It is well known that, at several levels, braided categories correspond to the braid groups B_n, while symmetric categories correspond to the symmetric groups S_n. It is natural to expect that there exist some groups corresponding, in the same way, to pseudosymmetric braided categories. Indeed, it is clear that these groups, denoted by PS_n and called (naturally) the pseudosymmetric groups, should be the quotients of the braid groups B_n by the relations $\sigma_i\sigma_{i+1}^{-1}\sigma_i = \sigma_{i+1}\sigma_i^{-1}\sigma_{i+1}$. Our aim is to study and find more explicitly the structure of these groups. We prove first that the kernel of the canonical group morphism $\mathrm{PS}_n \to S_n$ is abelian, and consequently PS_n is isomorphic to the quotient of B_n by the commutator subgroup $[P_n, P_n]$ of the pure braid group P_n. (This amounts to saying that $[P_n, P_n]$ coincides with the normal subgroup of B_n generated by the elements $[\sigma_i^2, \sigma_{i+1}^2]$ with $1 \le i \le n-2$.)

There exist similarities, but also differences, between braid groups and pseudosymmetric groups. Bigelow [2001] and Krammer [2002] proved that braid groups are linear, and we show that so are pseudosymmetric groups. More precisely, we prove that the Lawrence–Krammer representation of B_n induces a representation of PS_n if the parameter q is chosen to be 1, and that this representation of PS_n is faithful over $\mathbb{R}[t^{\pm 1}]$. On the other hand, although PS_n is an infinite group, like B_n, it does have nontrivial elements of finite order, unlike B_n.

1. Preliminaries

Definition 1.1 [Panaite et al. 2007]. Let $\mathscr{C}$ be a strict monoidal category and let $T_{X,Y} : X \otimes Y \to X \otimes Y$ be a family of natural isomorphisms in $\mathscr{C}$. We call T a *strong twine* if, for all $X, Y, Z \in \mathscr{C}$,

$$T_{I,I} = \mathrm{id}_I, \qquad (T_{X,Y} \otimes \mathrm{id}_Z) \circ T_{X\otimes Y,Z} = (\mathrm{id}_X \otimes T_{Y,Z}) \circ T_{X,Y\otimes Z},$$
$$(T_{X,Y} \otimes \mathrm{id}_Z) \circ (\mathrm{id}_X \otimes T_{Y,Z}) = (\mathrm{id}_X \otimes T_{Y,Z}) \circ (T_{X,Y} \otimes \mathrm{id}_Z).$$

Definition 1.2 [Panaite et al. 2009]. Let $\mathscr{C}$ be a strict monoidal category and c a braiding on $\mathscr{C}$. We say that c is *pseudosymmetric* if, for all $X, Y, Z \in \mathscr{C}$,

$$(1) \quad (c_{Y,Z} \otimes \mathrm{id}_X) \circ (\mathrm{id}_Y \otimes c_{Z,X}^{-1}) \circ (c_{X,Y} \otimes \mathrm{id}_Z) = (\mathrm{id}_Z \otimes c_{X,Y}) \circ (c_{Z,X}^{-1} \otimes \mathrm{id}_Y) \circ (\mathrm{id}_X \otimes c_{Y,Z}).$$

In this case we say that $\mathscr{C}$ is a *pseudosymmetric braided category*.

The next proposition, a key result in [Panaite et al. 2009], led to the introduction of the concept of pseudosymmetric braiding. Here, it will serve as a source of inspiration for a certain key result for braids, Proposition 2.1.

Proposition 1.3 [Panaite et al. 2009]. *Let $\mathscr{C}$ be a strict monoidal category and c a braiding on $\mathscr{C}$. Then the double braiding $T_{X,Y} := c_{Y,X} \circ c_{X,Y}$ is a strong twine if and only if c is pseudosymmetric.*

2. Defining relations for PS_n

Let $n \geq 3$ be a natural number. We denote by B_n the braid group on n strands, with its usual presentation by generators σ_i with $1 \leq i \leq n-1$ and relations

$$\sigma_i \sigma_j = \sigma_j \sigma_i \qquad \text{if } |i-j| \geq 2, \tag{2}$$

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \quad \text{if } 1 \leq i \leq n-2. \tag{3}$$

We begin with the analogue for braids of Proposition 1.3:

Proposition 2.1. *For all* $1 \leq i \leq n-2$, *the relations*

$$\sigma_i \sigma_{i+1}^{-1} \sigma_i = \sigma_{i+1} \sigma_i^{-1} \sigma_{i+1}, \tag{4}$$

$$\sigma_i^2 \sigma_{i+1}^2 = \sigma_{i+1}^2 \sigma_i^2 \tag{5}$$

are equivalent in B_n.

Proof. We show first that (4) implies (5):

$$\begin{aligned}
\sigma_i^2 \sigma_{i+1}^2 &= \sigma_i \sigma_{i+1}^{-1} \sigma_{i+1} \sigma_i \sigma_{i+1} \sigma_{i+1} \\
&\overset{(3)}{=} \sigma_i \sigma_{i+1}^{-1} \sigma_i \sigma_{i+1} \sigma_i \sigma_{i+1} \overset{(3),(4)}{=} \sigma_{i+1} \sigma_i^{-1} \sigma_{i+1} \sigma_i \sigma_{i+1} \sigma_i \\
&\qquad\qquad\qquad\qquad\qquad\quad \overset{(3)}{=} \sigma_{i+1} \sigma_i^{-1} \sigma_i \sigma_{i+1} \sigma_i \sigma_i = \sigma_{i+1}^2 \sigma_i^2.
\end{aligned}$$

Conversely, we prove that (5) implies (4):

$$\begin{aligned}
\sigma_i \sigma_{i+1}^{-1} \sigma_i &= \sigma_i \sigma_{i+1}^{-2} \sigma_i^{-1} \sigma_i \sigma_{i+1} \sigma_i \\
&\overset{(3)}{=} \sigma_i \sigma_{i+1}^{-2} \sigma_i^{-1} \sigma_{i+1} \sigma_i \sigma_{i+1} \\
&= \sigma_i \sigma_{i+1}^{-2} \sigma_i^{-2} \sigma_i \sigma_{i+1} \sigma_i \sigma_{i+1} \overset{(3),(5)}{=} \sigma_i \sigma_i^{-2} \sigma_{i+1}^{-2} \sigma_{i+1} \sigma_i \sigma_{i+1}^2 \\
&\qquad\qquad\qquad\qquad\qquad\quad = \sigma_i^{-1} \sigma_{i+1}^{-1} \sigma_i \sigma_{i+1}^2 \\
&\qquad\qquad\qquad\qquad\qquad\quad = \sigma_{i+1} \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_{i+1}^{-1} \sigma_i \sigma_{i+1}^2 \\
&\qquad\qquad\qquad\qquad\qquad\quad \overset{(3)}{=} \sigma_{i+1} \sigma_i^{-1} \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_i \sigma_{i+1}^2 \\
&\qquad\qquad\qquad\qquad\qquad\quad = \sigma_{i+1} \sigma_i^{-1} \sigma_{i+1}. \qquad \square
\end{aligned}$$

Definition 2.2. For a natural number $n \geq 3$, we define the *pseudosymmetric group* PS_n as the group with generators σ_i for $1 \leq i \leq n-1$, and relations (2), (3) and (4), or equivalently (2), (3) and (5).

Proposition 2.3. *For* $1 \leq i \leq n-2$, *consider the elements*

$$p_i := \sigma_i \sigma_{i+1}^{-1} \quad \textit{and} \quad q_i := \sigma_i^{-1} \sigma_{i+1} \tag{6}$$

in PS_n. *Then, in* PS_n, *we have*

$$p_i^3 = q_i^3 = (p_i q_i)^3 = 1 \quad \textit{for all } 1 \leq i \leq n-2. \tag{7}$$

Proof. The relations $p_i^3 = 1$ and $q_i^3 = 1$ follow immediately from (4); actually each of them is equivalent to (4). Now we compute

$$\begin{aligned}(p_i q_i)^2 &= (\sigma_i \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_{i+1})^2 \\ &= \sigma_i \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_{i+1} \sigma_i \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_{i+1} \\ &= \sigma_i \sigma_{i+1}^{-1} \sigma_i^{-1} \sigma_{i+1} \sigma_i \sigma_{i+1} \sigma_{i+1}^{-2} \sigma_i^{-1} \sigma_{i+1} \\ &\overset{(3)}{=} \sigma_i^2 \sigma_{i+1}^{-2} \sigma_i^{-1} \sigma_{i+1} \overset{(5)}{=} \sigma_{i+1}^{-2} \sigma_i \sigma_{i+1} \\ &\qquad\qquad\qquad\qquad = \sigma_{i+1}^{-2} \sigma_i \sigma_{i+1} \sigma_i \sigma_i^{-1} \\ &\qquad\qquad\qquad\qquad \overset{(3)}{=} \sigma_{i+1}^{-1} \sigma_i \sigma_{i+1} \sigma_i^{-1} = (p_i q_i)^{-1},\end{aligned}$$

and so $(p_i q_i)^3 = 1$. □

Consider now the symmetric group S_n with its usual presentation by generators s_i with $1 \le i \le n-1$ and relations (2), (3) and $s_i^2 = 1$ for all $1 \le i \le n-1$. We denote by $\pi : B_n \to S_n$, $\beta : B_n \to \mathrm{PS}_n$ and $\alpha : \mathrm{PS}_n \to S_n$ the canonical surjective group homomorphisms given by $\pi(\sigma_i) = s_i$, $\alpha(\sigma_i) = s_i$ and $\beta(\sigma_i) = \sigma_i$ for all $1 \le i \le n-1$. Obviously we have $\pi = \alpha \circ \beta$; hence in particular we obtain $\mathrm{Ker}(\alpha) = \beta(\mathrm{Ker}(\pi))$. We denote as usual $\mathrm{Ker}(\pi) = P_n$, the pure braid group on n strands. It is well known (see [Kassel and Turaev 2008, page 21]) that P_n is generated by the elements

$$a_{ij} := \sigma_{j-1} \sigma_{j-2} \cdots \sigma_{i+1} \sigma_i^2 \sigma_{i+1}^{-1} \cdots \sigma_{j-2}^{-1} \sigma_{j-1}^{-1} \quad \text{for } 1 \le i < j \le n \tag{8}$$

that satisfy certain relations, of which we will use only one, namely, that for $1 \le i < j \le n$ and $1 \le r < s \le n$,

$$a_{ij} a_{rs} = a_{rs} a_{ij} \quad \text{if } s < i \ \text{ or } \ i < r < s < j. \tag{9}$$

Alternatively, P_n is generated by the elements

$$b_{ij} := \sigma_{j-1}^{-1} \sigma_{j-2}^{-1} \cdots \sigma_{i+1}^{-1} \sigma_i^2 \sigma_{i+1} \cdots \sigma_{j-2} \sigma_{j-1} \quad \text{for } 1 \le i < j \le n. \tag{10}$$

It is easy to see that in B_n we have

$$\sigma_{i+1} \sigma_i^2 \sigma_{i+1}^{-1} = \sigma_i^{-1} \sigma_{i+1}^2 \sigma_i \quad \text{and} \quad \sigma_{i+1}^{-1} \sigma_i^2 \sigma_{i+1} = \sigma_i \sigma_{i+1}^2 \sigma_i^{-1}, \tag{11}$$

and by using repeatedly these relations we obtain an equivalent description of the elements a_{ij} and b_{ij}:

$$a_{ij} = \sigma_i^{-1} \sigma_{i+1}^{-1} \cdots \sigma_{j-2}^{-1} \sigma_{j-1}^2 \sigma_{j-2} \cdots \sigma_{i+1} \sigma_i \quad \text{for } 1 \le i < j \le n, \tag{12}$$

$$b_{ij} = \sigma_i \sigma_{i+1} \cdots \sigma_{j-2} \sigma_{j-1}^2 \sigma_{j-2}^{-1} \cdots \sigma_{i+1}^{-1} \sigma_i^{-1} \quad \text{for } 1 \le i < j \le n. \tag{13}$$

Now, for all $1 \le i < j \le n$, we define $A_{i,j}$ and $B_{i,j}$ as the elements in PS_n given by $A_{i,j} := \beta(a_{ij})$ and $B_{i,j} := \beta(b_{ij})$. From the discussion above it follows that $\mathrm{Ker}(\alpha)$ is generated by $\{A_{i,j}\}_{1\le i<j\le n}$ and also by $\{B_{i,j}\}_{1\le i<j\le n}$.

Lemma 2.4. *The following relations hold in* PS_n *for* $1 \le i < j < n$:

$$A_{i,j+1} = \sigma_j A_{i,j} \sigma_j^{-1}, \tag{14}$$

$$B_{i,j+1} = \sigma_j^{-1} B_{i,j} \sigma_j. \tag{15}$$

Proof. These relations are consequences of corresponding relations in B_n for the a_{ij} and b_{ij}, which in turn follow immediately from (8) and (10). □

Lemma 2.5. *For all* $i, j \in \{1, 2, \dots, n\}$ *with* $i+1 < j$, *we have in* PS_n

$$A_{i,j} = \sigma_i A_{i+1,j} \sigma_i^{-1}, \tag{16}$$

$$B_{i,j} = \sigma_i^{-1} B_{i+1,j} \sigma_i. \tag{17}$$

Proof. We prove (16), while (17) is similar and left to the reader. Note that in PS_n we have $\sigma_{i+1}^{-1}\sigma_i^2\sigma_{i+1} = \sigma_{i+1}\sigma_i^2\sigma_{i+1}^{-1}$, which together with the second of (11) implies $\sigma_i\sigma_{i+1}^2\sigma_i^{-1} = \sigma_{i+1}\sigma_i^2\sigma_{i+1}^{-1}$; hence

$$\begin{aligned} A_{i,j} &= \sigma_{j-1}\sigma_{j-2}\cdots(\sigma_{i+1}\sigma_i^2\sigma_{i+1}^{-1})\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1} \\ &= \sigma_{j-1}\sigma_{j-2}\cdots(\sigma_i\sigma_{i+1}^2\sigma_i^{-1})\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1} \\ &= \sigma_i\sigma_{j-1}\sigma_{j-2}\cdots\sigma_{i+1}^2\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1}\sigma_i^{-1} = \sigma_i A_{i+1,j}\sigma_i^{-1}. \end{aligned}$$ □

Proposition 2.6. *For all* $1 \le i < j \le n$, *we have* $A_{i,j} = B_{i,j}$ *in* PS_n.

Proof. We use (16) repeatedly:

$$\begin{aligned} A_{i,j} = \sigma_i A_{i+1,j}\sigma_i^{-1} &= \sigma_i\sigma_{i+1}A_{i+2,j}\sigma_{i+1}^{-1}\sigma_i^{-1} \\ &\cdots \\ &= \sigma_i\sigma_{i+1}\cdots\sigma_{j-2}A_{j-1,j}\sigma_{j-2}^{-1}\cdots\sigma_{i+1}^{-1}\sigma_i^{-1} \\ &= \sigma_i\sigma_{i+1}\cdots\sigma_{j-2}\sigma_{j-1}^2\sigma_{j-2}^{-1}\cdots\sigma_{i+1}^{-1}\sigma_i^{-1} \overset{(13)}{=} B_{i,j}. \end{aligned}$$ □

Lemma 2.7. *For all* $1 \le i < j \le n$ *and* $1 \le h \le k < n$, *we have in* PS_n

$$A_{i,j}\sigma_i^2 = \sigma_i^2 A_{i,j}, \tag{18}$$

$$A_{h,k+1}\sigma_k^2 = \sigma_k^2 A_{h,k+1}. \tag{19}$$

Proof. Note first that (18) is obvious for $j = i+1$. Assume that $i+1 < j$; using the fact that $A_{r,s} = B_{r,s}$ for all r, s, we compute

$$A_{i,j}\sigma_i^2 \overset{(16)}{=} \sigma_i A_{i+1,j}\sigma_i = \sigma_i B_{i+1,j}\sigma_i \overset{(17)}{=} \sigma_i^2 B_{i,j} = \sigma_i^2 A_{i,j}.$$

Note also that (19) is obvious for $h=k$. Assume that $h<k$; using again $A_{r,s}=B_{r,s}$ for all r, s, we compute

$$A_{h,k+1}\sigma_k^2 \overset{(14)}{=} \sigma_k A_{h,k}\sigma_k = \sigma_k B_{h,k}\sigma_k \overset{(15)}{=} \sigma_k^2 B_{h,k+1} = \sigma_k^2 A_{h,k+1}. \qquad \square$$

3. The structure of PS$_n$

We denote by $\mathfrak{P}_n$ the kernel of the morphism $\alpha : \mathrm{PS}_n \to S_n$ defined above.

Proposition 3.1. *$\mathfrak{P}_n$ is an abelian group.*

Proof. It is enough to prove that any two elements $A_{i,j}$ and $A_{k,l}$ commute in PS$_n$. We only have to analyze the following seven cases for the numbers i, j, k, l:

(i) $i<j<k<l$. This is an obvious consequence of (9).

(ii) $i<j=k<l$. We write

$$A_{i,j} = \sigma_i^{-1}\sigma_{i+1}^{-1}\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^2\sigma_{j-2}\cdots\sigma_{i+1}\sigma_i,$$
$$A_{j,l} = \sigma_{l-1}\sigma_{l-2}\cdots\sigma_{j+1}\sigma_j^2\sigma_{j+1}^{-1}\cdots\sigma_{l-2}^{-1}\sigma_{l-1}^{-1},$$

and we obtain $A_{i,j}A_{j,l}=A_{j,l}A_{i,j}$ by using (2) and the fact that σ_{j-1}^2 and σ_j^2 commute in PS$_n$.

(iii) $i<k<j<l$. This follows since $A_{k,l}=B_{k,l}$ in PS$_n$ (Proposition 2.6), and a_{ij} and b_{kl} commute in P_n if $i<k<j<l$, which is easily seen geometrically.

(iv) $i=k<j=l$. This is trivial.

(v) $i<k<l<j$. This is an obvious consequence of (9).

(vi) $i=k<j<l$. In case $j=i+1$, we have $A_{i,j}=\sigma_i^2$ and so we obtain $A_{i,j}A_{i,l}=A_{i,l}A_{i,j}$ by using (18); assuming now $i+1<j$, by using repeatedly (16) we can compute

$$\begin{aligned}A_{i,j}A_{i,l} &= \sigma_i A_{i+1,j}A_{i+1,l}\sigma_i^{-1}\\ &= \sigma_i\sigma_{i+1}A_{i+2,j}A_{i+2,l}\sigma_{i+1}^{-1}\sigma_i^{-1}\\ &\cdots\\ &= \sigma_i\sigma_{i+1}\cdots\sigma_{j-2}A_{j-1,j}A_{j-1,l}\sigma_{j-2}^{-1}\cdots\sigma_{i+1}^{-1}\sigma_i^{-1},\end{aligned}$$

and similarly

$$A_{i,l}A_{i,j} = \sigma_i\sigma_{i+1}\cdots\sigma_{j-2}A_{j-1,l}A_{j-1,j}\sigma_{j-2}^{-1}\cdots\sigma_{i+1}^{-1}\sigma_i^{-1};$$

these are equal since $A_{j-1,j}=\sigma_{j-1}^2$ and by (18), $\sigma_{j-1}^2A_{j-1,l}=A_{j-1,l}\sigma_{j-1}^2$.

(vii) $i < k < j = l$. In case $j = k+1$, we have $A_{k,j} = \sigma_k^2$ and so we obtain $A_{i,j}A_{k,j} = A_{k,j}A_{i,j}$ by using (19); assuming now $k+1 < j$, by repeatedly using (14) we can compute

$$\begin{aligned}
A_{i,j}A_{k,j} &= \sigma_{j-1}A_{i,j-1}A_{k,j-1}\sigma_{j-1}^{-1}\\
&= \sigma_{j-1}\sigma_{j-2}A_{i,j-2}A_{k,j-2}\sigma_{j-2}^{-1}\sigma_{j-1}^{-1}\\
&\dots\\
&= \sigma_{j-1}\sigma_{j-2}\cdots\sigma_{k+1}A_{i,k+1}A_{k,k+1}\sigma_{k+1}^{-1}\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1},
\end{aligned}$$

and similarly

$$A_{k,j}A_{i,j} = \sigma_{j-1}\sigma_{j-2}\cdots\sigma_{k+1}A_{k,k+1}A_{i,k+1}\sigma_{k+1}^{-1}\cdots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1};$$

these are equal since $A_{k,k+1} = \sigma_k^2$ and by (19), $A_{i,k+1}\sigma_k^2 = \sigma_k^2 A_{i,k+1}$. □

Let G be a group. If $x, y \in G$ we denote by $[x, y] := x^{-1}y^{-1}xy$ the commutator of x and y, and by G' the commutator subgroup of G (the subgroup of G generated by all commutators $[x, y]$), which is the smallest normal subgroup N of G with the property that G/N is abelian. Moreover, G' is a characteristic subgroup of G, that is, $\theta(G') = G'$ for all $\theta \in \mathrm{Aut}(G)$.

Proposition 3.2. $\mathfrak{P}_n \simeq P_n/P_n' \simeq \mathbb{Z}^{n(n-1)/2}$.

Proof. For $1 \le i \le n-2$ we define $t_i \in P_n$ by $t_i := [\sigma_i^2, \sigma_{i+1}^2] = [a_{i,i+1}, a_{i+1,i+2}]$. These elements are the relators added to the ones of B_n in order to obtain PS_n; therefore, as a particular case of a general fact about groups given by generators and relations (see for instance [Coxeter and Moser 1972, page 2]), the kernel of the map $\beta : B_n \to \mathrm{PS}_n$ defined above coincides with the normal subgroup of B_n generated by $\{t_i\}_{1\le i\le n-2}$, which will be denoted by L_n. We obviously have $L_n \subseteq P_n$, and if we consider the map β restricted to P_n, we have a surjective morphism $P_n \to \mathfrak{P}_n$ with kernel L_n, so $\mathfrak{P}_n \simeq P_n/L_n$. By Proposition 3.1 we know that $\mathfrak{P}_n$ is abelian, so we obtain $P_n' \subseteq L_n$. On the other hand, since P_n' is characteristic in P_n and P_n is normal in B_n, it follows (see [Suzuki 1982, Proposition 6.14]) that P_n' is normal in B_n, and since $t_1, \dots, t_{n-2} \in P_n'$ and L_n is the normal subgroup of B_n generated by $\{t_i\}_{1\le i\le n-2}$, we obtain $L_n \subseteq P_n'$. Thus, we have obtained $L_n = P_n'$ and so $\mathfrak{P}_n \simeq P_n/P_n'$. On the other hand, it is well known that $P_n/P_n' \simeq \mathbb{Z}^{n(n-1)/2}$; see for instance [Kassel and Turaev 2008, Corollary 1.20]. □

As a consequence of the equality $L_n = P_n'$, we obtain B_n/P_n':

Corollary 3.3. $\mathrm{PS}_n \simeq B_n/P_n'$.

The extension with abelian kernel $1 \to \mathfrak{P}_n \to \mathrm{PS}_n \to S_n \to 1$ induces an action of S_n on $\mathfrak{P}_n$, given by $\sigma \cdot a = \tilde\sigma a \tilde\sigma^{-1}$ for $\sigma \in S_n$ and $a \in \mathfrak{P}_n$, where $\tilde\sigma$ is an element of PS_n with $\alpha(\tilde\sigma) = \sigma$. In particular, on generators we have $s_k \cdot A_{i,j} = \sigma_k A_{i,j}\sigma_k^{-1}$, for $1 \leq k \leq n-1$ and $1 \leq i < j \leq n$. By using some of the formulas given above, one can describe explicitly this action as

(20a) $s_k \cdot A_{i,j} = A_{i,j}$ if $k < i-1$,

(20b) $s_{i-1} \cdot A_{i,j} = A_{i-1,j}$,

(20c) $s_i \cdot A_{i,j} = A_{i+1,j}$ if $j-i > 1$ and $s_i \cdot A_{i,i+1} = A_{i,i+1}$,

(20d) $s_k \cdot A_{i,j} = A_{i,j}$ if $i < k < j-1$,

(20e) $s_{j-1} \cdot A_{i,j} = A_{i,j-1}$ if $j-i > 1$ and $s_{j-1} \cdot A_{j-1,j} = A_{j-1,j}$,

(20f) $s_j \cdot A_{ij} = A_{i,j+1}$ for $1 \leq i < j < n$,

(20g) $s_k \cdot A_{i,j} = A_{i,j}$ if $j < k$.

Note that the first equality in (20c) follows by using (17) together with the fact that $A_{i,j} = B_{i,j}$ (Proposition 2.6), and the first equality in (20e) follows by an easy computation using also the fact that $A_{i,j} = B_{i,j}$. Also, one can easily see that these formulas may be expressed more compactly as follows: If $\sigma \in \{s_1, \dots, s_{n-1}\}$ and $1 \leq i < j \leq n$, then $\sigma \cdot A_{i,j} = A_{\sigma(i),\sigma(j)}$, where we made the convention $A_{r,t} := A_{t,r}$ for $t < r$. Since $s_1, \dots, s_{n-1}$ generate S_n, we have found the action of S_n on $A_{i,j}$:

Proposition 3.4. *For any $\sigma \in S_n$ and $1 \leq i < j \leq n$, the action of σ on $A_{i,j}$ is given by $\sigma \cdot A_{i,j} = A_{\sigma(i),\sigma(j)}$, with the convention $A_{r,t} := A_{t,r}$ for $t < r$.*

Lemma 3.5. *Let F be a free $\mathbb{Z}$-module of rank m, and let $\{X_1, \dots, X_m\}$ be a generating system for F over $\mathbb{Z}$. Then $\{X_1, \dots, X_m\}$ is a basis of F over $\mathbb{Z}$.*

Proof. Assume $X_1, \dots, X_m$ are linearly dependent over $\mathbb{Z}$ and take $\sum_{i=1}^{m} \alpha_i X_i = 0$ a nontrivial linear combination over $\mathbb{Z}$. Choose a prime number p such that $|\alpha_i| < p$ for all $1 \leq i \leq m$, and consider $\overline{F} := F/pF$, a linear space over the field $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$, and $\overline{X}_i$, the images of the elements X_i in $\overline{F}$. These elements generate $\overline{F}$ over $\mathbb{Z}_p$, and since the dimension of $\overline{F}$ over $\mathbb{Z}_p$ is m, it follows that $\{\overline{X}_1, \dots, \overline{X}_m\}$ is a basis of $\overline{F}$ over $\mathbb{Z}_p$. Thus, it follows that $\alpha_i \equiv 0 \pmod{p}$ for all $1 \leq i \leq m$, which is a contradiction because we have chosen p so that $|\alpha_i| < p$ for all $1 \leq i \leq m$. $\square$

Proposition 3.6. *In* PS_n, *there is no element of order* 2 *whose image in S_n is the transposition $s_1 = (1, 2)$. Consequently, the extension $1 \to \mathfrak{P}_n \to \mathrm{PS}_n \to S_n \to 1$ is* not *split.*

Proof. Take $x \in \mathrm{PS}_n$ such that $\alpha(x) = s_1$. Since $\alpha(\sigma_1) = s_1$, we obtain that $x\sigma_1^{-1} \in \mathrm{Ker}(\alpha) = \mathfrak{P}_n$. By Proposition 3.2 and Lemma 3.5, it follows that the abelian group $\mathfrak{P}_n$ is freely generated by $\{A_{i,j}\}_{1 \leq i < j \leq n}$, so we can write uniquely

$x = \prod_{1\le i<j\le n} A_{i,j}^{m_{ij}} \sigma_1$, with $m_{ij} \in \mathbb{Z}$. We compute

$$\begin{aligned} x^2 &= \Big(\prod_{1\le i<j\le n} A_{i,j}^{m_{ij}}\sigma_1\Big)\Big(\prod_{1\le i<j\le n} A_{i,j}^{m_{ij}}\sigma_1\Big) \\ &= \Big(\prod_{1\le i<j\le n} A_{i,j}^{m_{ij}}\Big)\Big(\sigma_1 \prod_{1\le i<j\le n} A_{i,j}^{m_{ij}}\sigma_1^{-1}\Big)\sigma_1^2 \\ &= \Big(\prod_{1\le i<j\le n} A_{i,j}^{m_{ij}}\Big)\Big(\prod_{1\le i<j\le n} \sigma_1 A_{i,j}^{m_{ij}}\sigma_1^{-1}\Big)A_{1,2} \\ &= A_{1,2}^{2m_{12}+1}\Big(\prod_{3\le j\le n} A_{1,j}^{m_{1j}+m_{2j}} A_{2,j}^{m_{1j}+m_{2j}}\Big)\Big(\prod_{3\le i<j\le n} A_{i,j}^{2m_{ij}}\Big), \end{aligned}$$

and this element cannot be trivial because $2m_{12}+1$ cannot be 0. Note that for the last equality we used the commutation relations

$$\begin{aligned} \sigma_1 A_{1,2}\sigma_1^{-1} &= A_{1,2}, \\ \sigma_1 A_{1,j}\sigma_1^{-1} &= A_{2,j} \quad \text{for all } j \ge 3, \\ \sigma_1 A_{2,j}\sigma_1^{-1} &= A_{1,j} \quad \text{for all } j \ge 3, \\ \sigma_1 A_{i,j}\sigma_1^{-1} &= A_{i,j} \quad \text{for all } 3 \le i < j, \end{aligned}$$

which can be easily proved by using some of the formulas given above. □

Remark 3.7. As is well known [Brown 1982], any extension with abelian kernel corresponds to a 2-cocycle. Specifically, the extension $1 \to \mathfrak{P}_n \to \mathrm{PS}_n \to S_n \to 1$ corresponds to an element in $H^2(S_n, \mathbb{Z}^{n(n-1)/2})$. We illustrate this by computing explicitly the corresponding 2-cocycle for $n = 3$. We consider the set-theoretical section $f : S_3 \to \mathrm{PS}_3$ defined by $f(1) = 1$, $f(s_2) = \sigma_2$, $f(s_1) = \sigma_1$, $f(s_1s_2) = \sigma_1\sigma_2$, $f(s_2s_1) = \sigma_2\sigma_1$ and $f(s_2s_1s_2) = \sigma_2\sigma_1\sigma_2$. The 2-cocycle afforded by this section is defined by $u : S_3 \times S_3 \to \mathfrak{P}_3$, $(x, y) \mapsto f(x)f(y)f(xy)^{-1}$, and a direct computation gives its explicit formula as in Table 1, where we have chosen an additive notation for the abelian group $\mathfrak{P}_3 \simeq \mathbb{Z}^3$.

	1	s_2	s_1	s_1s_2	s_2s_1	$s_2s_1s_2$
1	0	0	0	0	0	0
s_2	0	$A_{2,3}$	0	0	$A_{2,3}$	$A_{2,3}$
s_1	0	0	$A_{1,2}$	$A_{1,2}$	0	$A_{1,2}$
s_1s_2	0	$A_{1,3}$	0	$A_{1,2}$	$A_{1,2}+A_{1,3}$	$A_{1,2}+A_{1,3}$
s_2s_1	0	0	$A_{1,3}$	$A_{1,3}+A_{2,3}$	$A_{2,3}$	$A_{1,3}+A_{2,3}$
$s_2s_1s_2$	0	$A_{1,2}$	$A_{2,3}$	$A_{1,3}+A_{2,3}$	$A_{1,2}+A_{1,3}$	$A_{1,2}+A_{1,3}+A_{2,3}$

Table 1. The 2-cocycle for $n = 3$ associated to the section f.

4. PS_n is linear

Bigelow [2001] and Krammer [2002] proved that the braid group B_n is linear. More precisely, let R be a commutative ring, let q and t be two invertible elements in R, and let V be a free R-module of rank $n(n-1)/2$ with a basis $\{x_{i,j}\}_{1\le i<j\le n}$. Then the map $\rho : B_n \to \mathrm{GL}(V)$, defined by

$$\begin{aligned}
\sigma_k x_{k,k+1} &= tq^2 x_{k,k+1}, \\
\sigma_k x_{i,k} &= (1-q)x_{i,k} + qx_{i,k+1} && \text{for } i<k, \\
\sigma_k x_{i,k+1} &= x_{i,k} + tq^{k-i+1}(q-1)x_{k,k+1} && \text{for } i<k, \\
\sigma_k x_{k,j} &= tq(q-1)x_{k,k+1} + qx_{k+1,j} && \text{for } k+1<j, \\
\sigma_k x_{k+1,j} &= x_{k,j} + (1-q)x_{k+1,j} && \text{for } k+1<j, \\
\sigma_k x_{i,j} &= x_{i,j} && \text{for } i<j<k \text{ or } k+1<i<j, \\
\sigma_k x_{i,j} &= x_{i,j} + tq^{k-i}(q-1)^2 x_{k,k+1} && \text{for } i<k<k+1<j,
\end{aligned}$$

and $\rho(x)(v) = xv$ for $x \in B_n$ and $v \in V$, gives a representation of B_n, and if also $R = \mathbb{R}[t^{\pm 1}]$ and $q \in \mathbb{R} \subseteq R$ with $0<q<1$, then the representation is faithful; see [Krammer 2002].

We consider now the general formula for ρ, in which we take $q=1$:

$$\begin{aligned}
\sigma_k x_{k,k+1} &= tx_{k,k+1}, \\
\sigma_k x_{i,k} &= x_{i,k+1} && \text{for } i<k, \\
\sigma_k x_{i,k+1} &= x_{i,k} && \text{for } i<k, \\
\sigma_k x_{k,j} &= x_{k+1,j} && \text{for } k+1<j, \\
\sigma_k x_{k+1,j} &= x_{k,j} && \text{for } k+1<j, \\
\sigma_k x_{i,j} &= x_{i,j} && \text{for } i<j<k \text{ or } k+1<i<j, \\
\sigma_k x_{i,j} &= x_{i,j} && \text{for } i<k<k+1<j.
\end{aligned}$$

One can easily see that these formulas imply

$$\sigma_k^2 x_{k,k+1} = t^2 x_{k,k+1} \qquad \text{and} \qquad \sigma_k^2 x_{i,j} = x_{i,j} \quad \text{if } (i,j) \neq (k,k+1).$$

One can then check that $\rho(\sigma_k^2)$ commutes with $\rho(\sigma_{k+1}^2)$ for all $1 \le k \le n-2$, and so for $q=1$ it turns out that ρ is a representation of PS_n.

Theorem 4.1. *This representation of* PS_n *is faithful if* $R = \mathbb{R}[t^{\pm 1}]$. *Therefore*, PS_n *is linear.*

Proof. We first prove that $A_{i,j}x_{i,j} = t^2 x_{i,j}$, and $A_{i,j}x_{k,l} = x_{k,l}$ if $(i,j) \neq (k,l)$. We do it by induction over $|j-i|$. If $|j-i|=1$, the relations follow from the fact that $A_{i,i+1} = \sigma_i^2$. Assume the relations hold for $|j-i| = s-1$. We want to prove them for $|j-i|=s$. We recall that $A_{i,j}=\sigma_{j-1}A_{i,j-1}\sigma_{j-1}^{-1}$; see (14). We compute

$$\begin{aligned} A_{i,j}x_{i,j}=\sigma_{j-1}A_{i,j-1}\sigma_{j-1}^{-1}x_{i,j} &= \sigma_{j-1}A_{i,j-1}x_{i,j-1} \\ &= \sigma_{j-1}t^2x_{i,j-1} \quad \text{(by induction)} \\ &= t^2x_{i,j}. \end{aligned}$$

On the other hand, if $(i,j)\neq(k,l)$ then $\sigma_{j-1}^{-1}x_{k,l}=x_{u,v}$ with $(i,j-1)\neq(u,v)$, and so

$$\begin{aligned} A_{i,j}x_{k,l}=\sigma_{j-1}A_{i,j-1}\sigma_{j-1}^{-1}x_{k,l} &= \sigma_{j-1}A_{i,j-1}x_{u,v} \\ &= \sigma_{j-1}x_{u,v} \quad \text{(by induction)} \\ &= \sigma_{j-1}\sigma_{j-1}^{-1}x_{k,l}=x_{k,l}, \end{aligned}$$

as desired.

To show that the representation is faithful, take $b\in \mathrm{PS}_n$ such that $\rho(b)=\mathrm{id}_V$ and consider $\alpha(b)$, the image of b in S_n. From the way ρ is defined it follows that

$$bx_{i,j}=t^px_{\alpha(b)(i),\alpha(b)(j)} \quad \text{for all } 1\le i<j\le n,$$

with $p\in\mathbb{Z}$, where we made the convention $x_{r,s}:=x_{s,r}$ if $1\le s<r\le n$. Since $x_{i,j}$ is a basis in V and we assumed $\rho(b)=\mathrm{id}_V$, we find that the permutation $\alpha(b)\in S_n$ has the property that if $1\le i<j\le n$, then either $\alpha(b)(i)=i$ and $\alpha(b)(j)=j$ or $\alpha(b)(i)=j$ and $\alpha(b)(j)=i$. Since we assumed $n\ge 3$, the only such permutation is the trivial one. Thus, we have obtained that $b\in\mathrm{Ker}(\alpha)=\mathfrak{P}_n$ and so we can write $b=\prod_{1\le i<j\le n}A_{i,j}^{m_{i,j}}$, with $m_{i,j}\in\mathbb{Z}$. By using the formulas given above for the action of $A_{i,j}$ on $x_{k,l}$ we immediately obtain $bx_{k,l}=t^{2m_{k,l}}x_{k,l}$ for all $1\le k<l\le n$. Using again the assumption $\rho(b)=\mathrm{id}_V$, we obtain $t^{2m_{k,l}}=1$ and hence $m_{k,l}=0$ for all $1\le k<l\le n$, that is $b=1$, finishing the proof. □

5. Pseudosymmetric groups and pseudosymmetric braidings

We recall from [Kassel 1995, XIII.2] that to braid groups one can associate the so-called *braid category* $\mathcal{B}$, a universal braided monoidal category. Similarly, we can construct a pseudosymmetric braided category $\mathcal{PS}$ associated to pseudo-symmetric groups. Namely, the objects of $\mathcal{PS}$ are natural numbers $n\in\mathbb{N}$. The set of morphisms from m to n is empty if $m\neq n$ and is PS_n if $m=n$. The monoidal structure of $\mathcal{PS}$ is defined as the one for $\mathcal{B}$, and so is the braiding, namely

$$c_{n,m}:n\otimes m\to m\otimes n,$$

$$c_{0,n}=\mathrm{id}_n=c_{n,0},$$

$$c_{n,m}=(\sigma_m\sigma_{m-1}\cdots\sigma_1)(\sigma_{m+1}\sigma_m\cdots\sigma_2)\cdots(\sigma_{m+n-1}\sigma_{m+n-2}\cdots\sigma_n) \quad \text{if } m,n>0.$$

We denote by $t_{m,n} = c_{n,m} \circ c_{m,n}$ the double braiding. In view of Proposition 1.3, to prove that c is pseudosymmetric it is enough to check that, for all $m, n, p \in \mathbb{N}$,

$$(t_{m,n} \otimes \mathrm{id}_p) \circ (\mathrm{id}_m \otimes t_{n,p}) = (\mathrm{id}_m \otimes t_{n,p}) \circ (t_{m,n} \otimes \mathrm{id}_p). \tag{21}$$

Note that $t_{m,n} \otimes \mathrm{id}_p$ and $\mathrm{id}_m \otimes t_{n,p}$ are elements in $\mathfrak{P}_{m+n+p}$, which is an abelian group, and the composition $\circ$ between $t_{m,n} \otimes \mathrm{id}_p$ and $\mathrm{id}_m \otimes t_{n,p}$ is just the multiplication in the group $\mathfrak{P}_{m+n+p}$, so (21) is obviously true.

Let $\mathscr{C}$ be a strict braided monoidal category with braiding c, let n be a natural number and let $V \in \mathscr{C}$. Consider the automorphisms $c_1, \dots, c_{n-1}$ of $V^{\otimes n}$ defined by $c_i = \mathrm{id}_{V^{\otimes(i-1)}} \otimes c_{V,V} \otimes \mathrm{id}_{V^{\otimes(n-i-1)}}$. It is well known (see [Kassel 1995, XV.4]) that there exists a unique group morphism $\rho_n^c : B_n \to \mathrm{Aut}(V^{\otimes n})$ such that $\rho_n^c(\sigma_i) = c_i$ for all $1 \le i \le n-1$. It is clear that, if c is pseudosymmetric, then ρ_n^c factorizes to a group morphism $\mathrm{PS}_n \to \mathrm{Aut}(V^{\otimes n})$. Thus, pseudosymmetric braided categories provide representations of pseudosymmetric groups.

References

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

[Brown 1982] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1982. MR 83k:20002 Zbl 0584.20036

[Bruguières 2006] A. Bruguières, "Double braidings, twists and tangle invariants", *J. Pure Appl. Algebra* **204**:1 (2006), 170–194. MR 2006k:57030 Zbl 1091.57007

[Coxeter and Moser 1972] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, 3rd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete **14**, Springer, New York, 1972. MR 50 #2313 Zbl 0239.20040

[Joyal and Street 1993] A. Joyal and R. Street, "Braided tensor categories", *Adv. Math.* **102**:1 (1993), 20–78. MR 94m:18008 Zbl 0817.18007

[Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR 96e:17041 Zbl 0808.17003

[Kassel and Turaev 2008] C. Kassel and V. Turaev, *Braid groups*, Graduate Texts in Mathematics **247**, Springer, New York, 2008. MR 2009e:20082 Zbl 05268073

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

[Majid 1995] S. Majid, *Foundations of quantum group theory*, Cambridge University Press, 1995. MR 97g:17016 Zbl 0857.17009

[Panaite et al. 2007] F. Panaite, M. D. Staic, and F. Van Oystaeyen, "On some classes of lazy cocycles and categorical structures", *J. Pure Appl. Algebra* **209**:3 (2007), 687–701. MR 2007k:16074 Zbl 1155.16030

[Panaite et al. 2009] F. Panaite, M. D. Staic, and F. Van Oystaeyen, "Pseudosymmetric braidings, twines and twisted algebras", *J. Pure Appl. Algebra* (2009). arXiv 0801.2055

[Staic 2004] M. D. Staic, "Pure-braided Hopf algebras and knot invariants", *J. Knot Theory Ramifications* **13**:3 (2004), 385–400. MR 2005d:16069 Zbl 1106.16042

[Suzuki 1982] M. Suzuki, *Group theory, I*, Grundlehren der Mathematischen Wissenschaften **247**, Springer, Berlin, 1982. MR 82k:20001c Zbl 0472.20001

Received December 17, 2008. Revised June 26, 2009.

FLORIN PANAITE
INSTITUTE OF MATHEMATICS OF THE ROMANIAN ACADEMY
P.O. BOX 1-764
014700 BUCHAREST
ROMANIA
Florin.Panaite@imar.ro
http://www.imar.ro/~fpanaite/

MIHAI D. STAIC
INSTITUTE OF MATHEMATICS OF THE ROMANIAN ACADEMY
P.O. BOX 1-764
014700 BUCHAREST
ROMANIA
Current address:
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
RAWLES HALL
BLOOMINGTON, IN 47405
UNITED STATES
mstaic@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

KNOTS YIELDING HOMEOMORPHIC LENS SPACES BY DEHN SURGERY

TOSHIO SAITO AND MASAKAZU TERAGAITO

We show that there exist infinitely many pairs of distinct knots in the 3-sphere such that each pair can yield homeomorphic lens spaces by the same Dehn surgery. Moreover, each knot of the pair can be chosen to be a torus knot, a satellite knot or a hyperbolic knot, except that both cannot be satellite knots simultaneously. This exception is shown to be unavoidable by the classical theory of binary quadratic forms.

1. Introduction

For a knot K in the 3-sphere S^3, let $K(m/n)$ denote the closed oriented 3-manifold obtained by m/n-Dehn surgery on K, that is, $K(m/n)$ is the union of the knot exterior $E(K)=S^3-\operatorname{int}N(K)$ and a solid torus V in such a way that the meridian of V is attached to a loop on $\partial E(K)$ with slope m/n. In this paper, all 3-manifolds are oriented, and two knots in S^3 are said to be *equivalent* if there is an orientation-preserving homeomorphism of S^3 sending one to the other.

For a fixed slope m/n, m/n-surgery can be regarded as a map from the set of the equivalence classes of knots to that of 3-manifolds. There are many results on the injectivity of this map. Lickorish [1976] gave two nonequivalent knots on which (-1)-surgeries yield the same homology sphere. Brakes [1980] showed that for any integer $n\geq 2$, there exist n distinct knots on which 1-surgeries yield the same 3-manifold. See also [Kawauchi 1996; Livingston 1982; Teragaito 1994]. Finally, Osoinach [1998; 2006] showed the existence of 3-manifolds, in fact, a hyperbolic 3-manifold and a toroidal manifold, which can be obtained from infinitely many hyperbolic knots by 0-surgery. By using Osoinach's construction, Teragaito [2007] gave a Seifert fibered manifold over the 2-sphere with three exceptional fibers that can be obtained from infinitely many hyperbolic knots by 4-surgery. Thus it is natural to ask whether there exists a lens space that can be obtained from infinitely

MSC2000: primary 57M25; secondary 11B39, 11E16.
Keywords: Dehn surgery, lens space, knot, Fibonacci number, binary quadratic form.
Saito was partially supported by the Japan Society for the Promotion of Science, Research Fellowships for Young Scientists. Teragaito was partially supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (C), 19540089.

many knots by the same Dehn surgery. Although we do not know the answer yet, we feel that it is negative based on our computer experiments. In fact, as far as we know, at most two knots can yield homeomorphic lens spaces by the same Dehn surgery.

We should note that Berge's table [1990s] shows that among the lens spaces with fundamental groups of order up to 500, there are 32 that admit two knots yielding S^3 by Dehn surgery. This strongly suggests that many lens spaces can be obtained from nonequivalent knots in S^3 by the same Dehn surgery. In this paper, we study whether a pair of nonequivalent knots can yield homeomorphic lens spaces, ignoring orientations, by the same Dehn surgery. We should be attentive to this orientation convention. Let U be the unknot and K a knot in S^3. By using Floer homology for Seiberg–Witten monopoles, it is proved in [Kronheimer et al. 2007] that if there exists an orientation-preserving homeomorphism between $K(m/n)$ and $U(m/n)$, then K is trivial. In other words, if $K(m/n)$ is homeomorphic to the lens space $L(m,n)$ under an orientation-preserving homeomorphism, then K is trivial. Here, the preservation of orientation is important, because 5-surgery on the right-handed trefoil yields $L(5,4) = L(5,-1)$. From our point of view, the right-handed trefoil and the unknot yield homeomorphic lens spaces under the same 5-surgery.

As a consequence of the cyclic surgery theorem [Culler et al. 1987], any nontrivial amphicheiral knot has no Dehn surgery yielding a lens space, and the pair of a knot and its mirror image cannot yield homeomorphic lens spaces by the same Dehn surgery. Also, only torus knots admit nonintegral lens space surgeries.

Our first result is the following. We recall that all knots are classified into three families: torus knots, satellite knots, and hyperbolic knots.

Theorem 1.1. *There exist infinitely many pairs $\{K_1, K_2\}$ of nonequivalent knots in S^3 such that m-surgeries on them yield homeomorphic lens spaces for some integer m. Also, K_i can be chosen to be a torus knot, a satellite knot or a hyperbolic knot, except that K_1 and K_2 cannot be satellite knots simultaneously.*

The exceptional case in Theorem 1.1 is unavoidable as shown in Corollary 1.3, which is obtained as a consequence of the next theorem.

Theorem 1.2. (1) *There exist infinitely many pairs of nonequivalent torus knots in S^3 such that some half-integral surgeries on them yield homeomorphic lens spaces.*

(2) *Let K_1 and K_2 be nonequivalent torus knots. Suppose a slope r corresponds to a lens space surgery for both K_1 and K_2. If the slope r runs at least three times in the longitudinal direction, then r-surgeries on K_1 and K_2 cannot yield homeomorphic lens spaces.*

Corollary 1.3. *Nonequivalent satellite knots cannot yield homeomorphic lens spaces by the same Dehn surgery.*

Question 1.4. Is there a lens space that can be obtained from three nonequivalent knots in S^3 by the same Dehn surgery?

Based on a computer experiment, we conjecture that the answer is negative.

The paper is organized as follows. In Section 2, we give infinitely many pairs of torus knots that yield homeomorphic lens spaces. After establishing one result concerning a divisibility of integers by using the classical theory of integral binary quadratic forms in Section 3, we prove Theorem 1.2 and Corollary 1.3 in Section 4. In Section 5, we review one special class of doubly primitive knots. In Section 6, we construct by using tangles infinitely many pairs of hyperbolic knots that yield homeomorphic lens spaces. Finally, Section 7 completes the proof of Theorem 1.1 by treating the case where the knots of a pair belong to different classes.

2. Torus knots

Here, we give infinitely many pairs of torus knots that yield homeomorphic lens spaces by the same integral Dehn surgery.

Recall that the Fibonacci numbers are defined by the recurrence equation

$$F_{n+2} = F_{n+1} + F_n \quad \text{with } F_0 = 0 \text{ and } F_1 = 1.$$

We make use of Cassini's identity (see [Graham et al. 1994])

$$F_{k-1}F_{k+1} - F_k^2 = (-1)^k \quad \text{for } k > 0.$$

Let $a_n = F_{n+2}$ and $b_n = F_{n+3} + F_{n+1}$ for $n \geq 1$.

Lemma 2.1. *For any $n \geq 1$,*

$$a_{n+1}b_n + (-1)^{n+1} = a_n b_{n+1} + (-1)^n.$$

Proof. By using Cassini's identity,

$$\begin{aligned} a_{n+1}b_n + (-1)^{n+1} &= F_{n+3}(F_{n+3} + F_{n+1}) + (-1)^{n+1} \\ &= F_{n+3}^2 + F_{n+3}F_{n+1} + (-1)^{n+1} = F_{n+3}^2 + F_{n+2}^2. \end{aligned}$$

Similarly,

$$\begin{aligned} a_n b_{n+1} + (-1)^n &= F_{n+2}(F_{n+4} + F_{n+2}) + (-1)^n \\ &= F_{n+2}F_{n+4} + F_{n+2}^2 + (-1)^n = F_{n+3}^2 + F_{n+2}^2. \qquad \square \end{aligned}$$

As seen from Cassini's identity, two successive Fibonacci numbers are relatively prime. Then it is easy to see that $\gcd(a_{n+1}, b_n) = \gcd(a_n, b_{n+1}) = 1$.

Proposition 2.2. *For $n \geq 1$, let K be the torus knot of type (a_{n+1}, b_n), and K' the torus knot of type (a_n, b_{n+1}). Let $m = a_{n+1}b_n + (-1)^{n+1}$ $(= a_n b_{n+1} + (-1)^n)$. Then K and K' are not equivalent, and m-surgery on K and K' yields homeomorphic lens spaces.*

Proof. K and K' are not equivalent since $a_n < a_{n+1} < b_n < b_{n+1}$. By [Moser 1971], m-surgery on K and K' yields the lens spaces $L(a_{n+1}b_n + (-1)^{n+1}, a_{n+1}^2)$ and $L(a_n b_{n+1} + (-1)^n, a_n^2)$, respectively. Since $a_n^2 + a_{n+1}^2 = F_{n+2}^2 + F_{n+3}^2 = m$ as seen in the proof of Lemma 2.1, $a_n^2 + a_{n+1}^2 \equiv 0 \pmod{m}$. Thus these lens spaces are homeomorphic. □

3. Binary quadratic form

In this section, we prove Proposition 3.1, which will be used in Section 4. For its proof, we quickly review the classical theory of integral binary quadratic forms. See [Flath 1989], for example.

Let $f(x, y) = Ax^2 + Bxy + Cy^2$ be an integral binary quadratic form with discriminant $\Delta = B^2 - 4AC$. For our purposes, it is enough to assume that Δ is a positive nonsquare. Let m be a nonzero integer. Then there is a finite algorithm to find all integral solutions $(x, y) \in \mathbb{Z}^2$ of $f(x, y) = m$, as described below.

Let $\mathcal{S} = \{(x, y) \in \mathbb{Z}^2 \mid f(x, y) = m\}$ be the set of integral solutions of $f(x, y) = m$. Set

$$\rho = \begin{cases} \frac{1}{2}\sqrt{\Delta} & \text{if } \Delta \equiv 0 \pmod 4, \\ \frac{1}{2}(1 + \sqrt{\Delta}) & \text{if } \Delta \equiv 1 \pmod 4. \end{cases}$$

Let us consider the ring $\mathbb{O}_\Delta = \{x + y\rho \mid x, y \in \mathbb{Z}\}$. Let $\mathbb{O}_\Delta^\times$ be the group of units of $\mathbb{O}_\Delta$, and let $\mathbb{O}_{\Delta,1}^\times = \{\alpha \in \mathbb{O}_\Delta^\times \mid N(\alpha) = 1\}$ be the subgroup of units for norm 1. Note that the norm $N(\alpha)$ of $\alpha = x + y\rho$ is given

$$N(\alpha) = \begin{cases} x^2 - \frac{1}{4}\Delta y^2 & \text{if } \Delta \equiv 0 \pmod 4, \\ x^2 + xy - \frac{1}{4}(\Delta - 1)y^2 & \text{if } \Delta \equiv 1 \pmod 4. \end{cases}$$

In fact, $\mathbb{O}_{\Delta,1}^\times$ corresponds to the solution set of the Pell equation $N(\alpha) = 1$. Then $\mathbb{O}_{\Delta,1}^\times$ acts on the set $\mathcal{S}$. It is well known that the number of $\mathbb{O}_{\Delta,1}^\times$-orbits in $\mathcal{S}$ is finite. Since $\mathbb{O}_{\Delta,1}^\times$ is infinite, the orbit of each solution is infinite, so $\mathcal{S}$ is infinite unless $\mathcal{S} = \varnothing$. The action is explicitly given by the formulas

$$(x', y') = \begin{cases} (x, y)\begin{pmatrix} u - \frac{1}{2}Bv & Av \\ -Cv & u + \frac{1}{2}Bv \end{pmatrix} & \text{if } \Delta \equiv 0 \pmod 4, \\ (x, y)\begin{pmatrix} u + \frac{1}{2}(1 - B)v & Av \\ -Cv & u + \frac{1}{2}(1 + B)v \end{pmatrix} & \text{if } \Delta \equiv 1 \pmod 4, \end{cases}$$

for $u + v\rho \in \mathbb{O}_{\Delta,1}^\times$ and $(x, y) \in \mathcal{S}$.

Let τ be the smallest unit of $\mathbb{O}_{\Delta,1}^{\times}$ that is greater than 1. Then every $\mathbb{O}_{\Delta,1}^{\times}$-orbit of integral solutions of $f(x, y) = m$ contains a solution $(x, y) \in \mathbb{Z}^2$ such that

$$0 \le y \le U = \begin{cases} |(Am/\Delta)(\tau + \bar{\tau} - 2)|^{1/2} & \text{if } Am > 0, \\ |(Am/\Delta)(\tau + \bar{\tau} + 2)|^{1/2} & \text{if } Am < 0, \end{cases}$$

where $\bar{\tau}$ is the conjugate of τ. Also, two distinct solutions (x_1, y_1), $(x_2, y_2) \in \mathbb{Z}^2$ of $f(x, y) = m$ such that $0 \le y_i \le U$ belong to the same $\mathbb{O}_{\Delta,1}^{\times}$-orbit if and only if $y_1 = y_2 = 0$ or $y_1 = y_2 = U$.

Proposition 3.1. *Let $n \ge 3$ be an integer. Let a, b and c be positive integers such that $a > 1$ and $\gcd(a, b) = \gcd(a, c) = 1$. Then $b^2 \pm c^2$ is not divisible by $nabc \pm 1$.*

Proof. Without loss of generality, we may assume that $b > c$. Let $\varepsilon \in \{1, -1\}$. If $b^2 + c^2$ is divisible by $nabc + \varepsilon$, then

$$b^2 + c^2 = Q(nabc + \varepsilon) \tag{3-1}$$

for some integer $Q \ge 1$. Consider an integral binary quadratic form $f(x, y) = x^2 - Qnaxy + y^2$. Then Equation (3-1) means that the equation $f(x, y) = \varepsilon Q$ has a solution (b, c).

Similarly, if $b^2 - c^2$ is divisible by $nabc + \varepsilon$, then for a binary quadratic form $g(x, y) = x^2 - Qnaxy - y^2$, the equation $g(x, y) = \varepsilon Q$ has a solution (b, c). We remark that the discriminants $\Delta_f = (Qna)^2 - 4$ of f and $\Delta_g = (Qna)^2 + 4$ of g are positive and nonsquare.

First, we list all solutions in positive integers of the equation $f(x, y) = Q$. For simplicity, let $\Delta = \Delta_f$. Let $\mathcal{S} = \{(x, y) \in \mathbb{Z}^2 \mid f(x, y) = Q\}$ be the set of all integral solutions of the equation $f(x, y) = Q$. Then the action of $\mathbb{O}_{\Delta,1}^{\times}$ on the set $\mathcal{S}$ is given by the formula

$$(x', y') = \begin{cases} (x, y)\begin{pmatrix} u + \frac{1}{2}Qnav & v \\ -v & u - \frac{1}{2}Qnav \end{pmatrix} & \text{if } \Delta \equiv 0 \pmod 4, \\ (x, y)\begin{pmatrix} u + \frac{1}{2}(1 + Qna)v & v \\ -v & u + \frac{1}{2}(1 - Qna)v \end{pmatrix} & \text{if } \Delta \equiv 1 \pmod 4, \end{cases} \tag{3-2}$$

for $u + v\rho \in \mathbb{O}_{\Delta,1}^{\times}$ and $(x, y) \in \mathcal{S}$.

Let τ be the smallest unit of $\mathbb{O}_{\Delta,1}^{\times}$ that is greater than 1. In fact, we see that

$$\tau = \begin{cases} \frac{1}{2}Qna + \rho & \text{if } \Delta \equiv 0 \pmod 4, \\ \frac{1}{2}(Qna - 1) + \rho & \text{if } \Delta \equiv 1 \pmod 4. \end{cases}$$

Then every orbit contains a solution $(x, y) \in \mathbb{Z}^2$ such that

$$0 \le y \le U = |(Q/\Delta)(\tau + \bar{\tau} - 2)|^{1/2}.$$

In our case, $U < 1$, and so $\mathcal{S}$ consists of a single $\mathbb{O}^{\times}_{\Delta,1}$-orbit. Furthermore, Q must be a square in order that $\mathcal{S} \neq \varnothing$. We start a solution $(\sqrt{Q}, 0) \in \mathcal{S}$. By (3-2),

$$\tau \cdot (\sqrt{Q}, 0) = (\sqrt{Q}, 0)\begin{pmatrix} Qna & 1 \\ -1 & 0 \end{pmatrix} = (Q^{3/2}na, \sqrt{Q}).$$

Since

$$(x, y)\begin{pmatrix} Qna & 1 \\ -1 & 0 \end{pmatrix} = (Qnax - y, x),$$

every solution in positive integers has a coordinate that is divisible by a. Thus $f(x, y) = Q$ cannot have the solution (b, c), because $\gcd(a, b) = \gcd(a, c) = 1$.

For the equation $f(x, y) = -Q$, we have $U = |(-Q/\Delta)(\tau + \bar{\tau} + 2)|^{1/2} < 1$ again. However, $f(x, 0) = x^2$ implies that the set of solutions of the equation $f(x, y) = -Q$ is empty.

Next, consider the equation $g(x, y) = Q$. Let $\mathcal{T} = \{(x, y) \in \mathbb{Z}^2 \mid g(x, y) = Q\}$. Put $\Delta = \Delta_g$. Then $\mathbb{O}_\Delta$, $\mathbb{O}^{\times}_{\Delta,1}$ are defined in the same way, but the action of $\mathbb{O}^{\times}_{\Delta,1}$ on the set $\mathcal{T}$ is given by the formula

(3-3)
$$(x', y') = \begin{cases} (x, y)\begin{pmatrix} u + \frac{1}{2}Qnav & v \\ v & u - \frac{1}{2}Qnav \end{pmatrix} & \text{if } \Delta \equiv 0 \pmod 4, \\ (x, y)\begin{pmatrix} u + \frac{1}{2}(1 + Qna)v & v \\ v & u + \frac{1}{2}(1 - Qna)v \end{pmatrix} & \text{if } \Delta \equiv 1 \pmod 4, \end{cases}$$

for $u + v\rho \in \mathbb{O}^{\times}_{\Delta,1}$ and $(x, y) \in \mathcal{T}$. Also, the smallest unit τ of $\mathbb{O}^{\times}_{\Delta,1}$ that is greater than 1 is given by

$$\tau = \begin{cases} \left(\frac{1}{2}Qna + \rho\right)^2 = \frac{1}{2}(Qna)^2 + 1 + Qna\rho & \text{if } \Delta \equiv 0 \pmod 4, \\ \left(\frac{1}{2}(Qna - 1) + \rho\right)^2 = \frac{1}{2}((Qna)^2 - Qna) + 1 + Qna\rho & \text{if } \Delta \equiv 1 \pmod 4. \end{cases}$$

As before, we can evaluate $U = |(Q/\Delta)(\tau + \bar{\tau} - 2)|^{1/2} < \sqrt{Q}$.

On the other hand, if $(x, y) \in \mathcal{T}$, then $\Delta y^2 + 4Q = (2x - Qnay)^2$. That is, $\Delta y^2 + 4Q$ must be a square. If $0 < y < \sqrt{Q}$, then

$$Qnay < \sqrt{\Delta y^2 + 4Q} < Qnay + 1.$$

Hence $y = 0$, and so $\mathcal{T}$ consists of a single $\mathbb{O}^{\times}_{\Delta,1}$-orbit. Thus Q must be a square in order that $\mathcal{T} \neq \varnothing$. Starting a solution $(\sqrt{Q}, 0) \in \mathcal{T}$, we have

$$\tau \cdot (\sqrt{Q}, 0) = (\sqrt{Q}, 0)\begin{Bmatrix} (Qna)^2 + 1 & Qna \\ Qna & 1 \end{Bmatrix} = (Q^{5/2}n^2a^2 + \sqrt{Q}, Q^{3/2}na)$$

by the formulas (3-3). Thus for every solution in positive integers, the second coordinate is divisible by a.

Finally, for the equation $g(x, y) = -Q$, we have $U = |(-Q/\Delta)(\tau + \bar{\tau} + 2)|^{1/2}$, which is less than or equal to $\sqrt{Q}$ when $\Delta \equiv 0 \pmod 4$ and less than $\sqrt{Q}$ when $\Delta \equiv 1 \pmod 4$.

If $g(x, y) = -Q$, then $\Delta y^2 - 4Q = (2x - Qnay)^2$. Thus $y \neq 0$. Furthermore, if $y < \sqrt{Q}$, then

$$Qnay - 1 < \sqrt{\Delta y^2 - 4Q} < Qnay.$$

Therefore, $y = \sqrt{Q}$ is the only possibility, and so Q must be a square. As before, the set of solutions of the equation $g(x, y) = -Q$ consists of a single $\mathbb{O}^\times_{\Delta,1}$-orbit, whose representative is $(0, \sqrt{Q})$. Then

$$\tau \cdot (0, \sqrt{Q}) = (0, \sqrt{Q}) \begin{pmatrix} (Qna)^2 + 1 & Qna \\ Qna & 1 \end{pmatrix} = (Q^{3/2}na, \sqrt{Q}).$$

Hence the first coordinate is divisible by a for any solution in positive integers. $\square$

Remark 3.2. The requirement $a > 1$ in Proposition 3.1 is necessary. For example, let $a = 1,\ b = 3,\ c = 8$ and $n = 3$. Then $b^2 + c^2 = 73$ is divisible by $nabc + 1 = 73$.

4. Nonintegral surgery on torus knots

In this section, we prove Theorem 1.2.

Let $\{a_n\}$ and $\{b_n\}$ be the sequences of positive integers defined by

$$a_{n+1} = a_n + b_n \quad \text{and} \quad b_{n+1} = a_{n+1} + a_n \tag{4-1}$$

with $a_1 = 2$ and $b_1 = 3$.

Lemma 4.1. *For any $n \geq 1$,*

(1) $2a_n b_{n+1} + (-1)^{n+1} = 2a_{n+1} b_n + (-1)^n$,

(2) $4a_{n+1}^2 b_{n+1}^2 + 1 = (2a_{n+1} b_{n+2} + (-1)^{n+2})(2a_n b_{n+1} + (-1)^{n+1})$.

Proof. By (4-1), we have $a_{n+1} = 2a_n + a_{n-1}$. Then

$$\begin{aligned} 2a_n b_{n+1} - 2a_{n+1} b_n &= 2a_n(a_{n+1} + a_n) - 2a_{n+1}(a_{n+1} - a_n) \\ &= 2(a_n^2 - a_{n+1}^2 + 2a_n a_{n+1}) \\ &= -2(a_{n-1}^2 - a_n + 2a_{n-1} a_n) \\ &\ \ \vdots \\ &= (-1)^{n-1} 2(a_1^2 - a_2^2 + 2a_1 a_2) = (-1)^n 2 = (-1)^n - (-1)^{n+1}. \end{aligned}$$

This proves (1).

To prove (2), we observe that $2b_{n+1} = a_{n+2} + a_n$ by (4-1). Also, as shown above, $2a_n b_{n+1} - 2a_{n+1} b_n = (-1)^n 2$. Thus, $a_n b_{n+1} - a_{n+1} b_n = (-1)^n$. From (4-1),

$a_n(a_{n+1}+a_n)-a_{n+1}(a_{n+1}-a_n)=(-1)^n$. Then $a_n^2+2a_na_{n+1}-a_{n+1}^2=(-1)^n$. Thus,

$$\begin{aligned}
&(2a_{n+1}b_{n+2}+(-1)^{n+2})(2a_nb_{n+1}+(-1)^{n+1})\\
&=(2a_{n+2}b_{n+1}+(-1)^{n+1})(2a_nb_{n+1}+(-1)^{n+1})\\
&=4a_na_{n+2}b_{n+1}^2+(-1)^{n+1}2b_{n+1}(a_n+a_{n+2})+1\\
&=4b_{n+1}^2(a_na_{n+2}+(-1)^{n+1})+1\\
&=4b_{n+1}^2(a_n(a_{n+1}+b_{n+1})+(-1)^{n+1})+1\\
&=4b_{n+1}^2(a_na_{n+1}+a_n(a_{n+1}+a_n)+(-1)^{n+1})+1\\
&=4b_{n+1}^2(2a_na_{n+1}+a_n^2+(-1)^{n+1})+1=4b_{n+1}^2a_{n+1}^2+1. \qquad \square
\end{aligned}$$

From (1), we have that $\gcd(a_n, b_{n+1})=\gcd(b_n, a_{n+1})=1$.

Proof of Theorem 1.2(1). Let K_1 be the torus knot of type (a_n, b_{n+1}), and let K_2 be the torus knot of type (b_n, a_{n+1}). Since $a_n < b_n < a_{n+1} < b_{n+1}$ for any $n \geq 1$, K_1 and K_2 are not equivalent. Then $\frac{1}{2}(2a_nb_{n+1}+(-1)^{n+1})$-surgery on K_1 and $\frac{1}{2}(2a_{n+1}b_n+(-1)^n)$-surgery on K_2 yield the lens spaces

$$L(2a_nb_{n+1}+(-1)^{n+1}, 2b_{n+1}^2) \quad \text{and} \quad L(2a_{n+1}b_n+(-1)^n, 2a_{n+1}^2),$$

respectively. By Lemma 4.1, the surgery coefficients are the same, and the two lens spaces are homeomorphic. $\square$

In the rest of this section, we prove Theorem 1.2(2) and Corollary 1.3.

Let K_1 be the torus knot of type (p,q); let K_2 be the torus knot of type (r,s). Suppose $n \geq 3$. If m/n-surgery on K_1 yields a lens space, then $\Delta(pq/1, m/n) = |npq-m| = 1$, so $m = npq \pm 1$. Hence if m/n-surgery on K_1 and K_2 yields homeomorphic lens spaces, then $npq+\varepsilon = nrs+\varepsilon'$ for some $\varepsilon, \varepsilon' \in \{1,-1\}$. Since we consider nontrivial torus knots, we can assume that p, q, r and s are positive by taking mirror images, if necessary. Moreover, we may assume that $2 \leq q < p$, $2 \leq s < r$ and $r < p$. Now $\varepsilon = \varepsilon'$ because $n \geq 3$, and so $pq = rs$. By [Moser 1971], m/n-surgery on K_1 and K_2 yields $L(m, nq^2)$ and $L(m, ns^2)$, respectively.

Theorem 1.2(2) follows directly from the following.

Proposition 4.2. *The two lens spaces $L(m, nq^2)$ and $L(m, ns^2)$ are not homeomorphic.*

Proof. The two lens spaces are homeomorphic if and only if

(4-2) $$nq^2 \equiv \pm ns^2 \pmod{m} \text{ or}$$

(4-3) $$n^2q^2s^2 \equiv \pm 1 \pmod{m}.$$

Since $npq+\varepsilon = nrs+\varepsilon$, we have $npq = nrs$. Thus $q < s < r < p$.

First, $nq^2 \not\equiv ns^2 \pmod m$ because $0 < n(s^2 - q^2) < ns^2 < nrs - 1 \le m$. Since $nq^2 + ns^2 < n(pq + rs) - 2 = 2npq - 2 \le 2m$, the equation $nq^2 \equiv -ns^2 \pmod m$ is possible only when $nq^2 + ns^2 = m$. However, this is impossible because m is not divisible by n. The impossibility of Equation (4-3) is shown in Proposition 4.3. □

Proposition 4.3. $n^2q^2s^2 \not\equiv \pm 1 \pmod m$.

Proof. Suppose $n^2q^2s^2 \equiv 1 \pmod m$. Then $n^2q^2s^2 - 1 = km$ for some integer $k \ge 1$. Recall that $m = npq + \varepsilon$, so $-1 \equiv k\varepsilon \pmod n$ and thus $k \equiv -\varepsilon \pmod n$. Put $k = n\ell - \varepsilon$ with $\ell \ge 1$. (If $\ell = 0$, then $k = -\varepsilon = -1$, so $n^2q^2s^2 - 1 = m$. This implies that q divides p, a contradiction.) Then $n^2q^2s^2 - 1 = (n\ell - \varepsilon)(npq + \varepsilon)$ implies

$$q(nqs^2 - p(n\ell - \varepsilon)) = \varepsilon\ell. \tag{4-4}$$

Thus q divides ℓ, and $\gcd(p, s)$ divides ℓ/q. For simplicity, we denote $\gcd(x, y)$ by (x, y).

Hence

$$\begin{aligned} p(n\ell - \varepsilon) &= nqs^2 - \varepsilon\ell/q = nq(p, s)^2(q, s)^2 - \varepsilon\ell/q \\ &= (p, s)\left(\frac{nq}{(q, s)}(q, s)^3(p, s) - \frac{\varepsilon\ell}{q(p, s)}\right) \end{aligned} \tag{4-5}$$

Here we put $a = q(p, s)/(q, s)$, $b = (q, s)$ and $c = \ell/(q(p, s))$. Then $abc = \ell$.

Claim 4.4. $a > 1$.

Proof. Assume $a = 1$. Then $(p, s) = 1$ and $q = (q, s)$. Since $s = (p, s)(q, s)$, $s = (q, s)$. Thus $s = q$, so $p = r$, a contradiction. □

Claim 4.5. $(a, b) = (a, c) = 1$.

Proof. First, (p, s) and (q, s) are coprime. Also, $q/(q, s)$ and (q, s) are coprime, for otherwise $(r, s) > 1$. Thus $(a, b) = 1$.

Next, assume $(a, c) > 1$. Let d be a prime factor of (a, c). From Equation (4-4),

$$nqs^2 - p(n\ell - \varepsilon) = \varepsilon\ell/q.$$

Dividing this by (p, s) gives

$$nqs\frac{s}{(p, s)} - \frac{p}{(p, s)}(n\ell - \varepsilon) = \varepsilon c. \tag{4-6}$$

Since d divides a, it divides q or s. Similarly, d divides ℓ, since d divides c. Thus Equation (4-6) gives

$$\frac{p}{(p, s)}\varepsilon \equiv 0 \pmod d.$$

However, this is impossible, because (p, s) and $p/(p, s)$ are coprime. □

On the other hand, Equation (4-5) yields $p(nabc-\varepsilon)=(p,s)(nab^3-\varepsilon c)$, which, since (p,s) divides p, means that $nab^3-\varepsilon c$ is divisible by $nabc-\varepsilon$. Furthermore

$$\frac{nab^3-\varepsilon c}{nabc-\varepsilon}=c+\frac{nab(b^2-c^2)}{nabc-\varepsilon}$$

implies that b^2-c^2 is divisible by $nabc-\varepsilon$, since nab and $nabc-\varepsilon$ are coprime.

Similarly, if $n^2q^2s^2\equiv -1 \pmod{m}$, then b^2+c^2 is divisible by $nabc-\varepsilon$. However, these are impossible by Proposition 3.1. □

Proof of Corollary 1.3. Among satellite knots, only the $(2, 2pq+\varepsilon)$-cable K of the (p,q)-torus knot admits a lens space surgery for $\varepsilon=\pm1$. Then the slope is $4pq+\varepsilon$, and $L(4pq+\varepsilon, 4q^2)$ arises. This surgery on K is equivalent to $(4pq+\varepsilon)/4$-surgery on its companion torus knot. Thus the result follows from Theorem 1.2(2). □

5. Doubly primitive knot

In this section, we study a special class of doubly primitive knots $k^+(a,b)$ defined by Berge [1990s]. In particular, two infinite sequences of $k^+(a,b)$ are proved to be hyperbolic via dual knots in lens spaces. As far as we know, whether $k^+(a,b)$ is hyperbolic is still an open question.

For a pair (a,b) of coprime positive integers, let $k^+(a,b)$ denote the doubly primitive knot defined by Berge [1990s], which lies on a genus one fiber surface of the left-handed trefoil as shown in Figure 1(1). Then (a^2+ab+b^2)-surgery on $k^+(a,b)$ yields the lens space $L(a^2+ab+b^2, (a/b)^2)$, where a/b is calculated in $\mathbb{Z}_{a^2+ab+b^2}$. (We adopt the notation of [Yamada 2005], but there the orientation of lens spaces is opposite to ours). We note that $k^+(a,b)$ and $k^+(b,a)$ are equivalent by the symmetry of the fiber surface. For example, $k^+(1,3)$ is the $(3,4)$-torus knot whose 13-surgery yields $L(13,9)$, and $k^+(2,3)$, as shown in Figure 1(2), is the $(-2,3,7)$-pretzel knot whose 19-surgery yields $L(19,7)$.

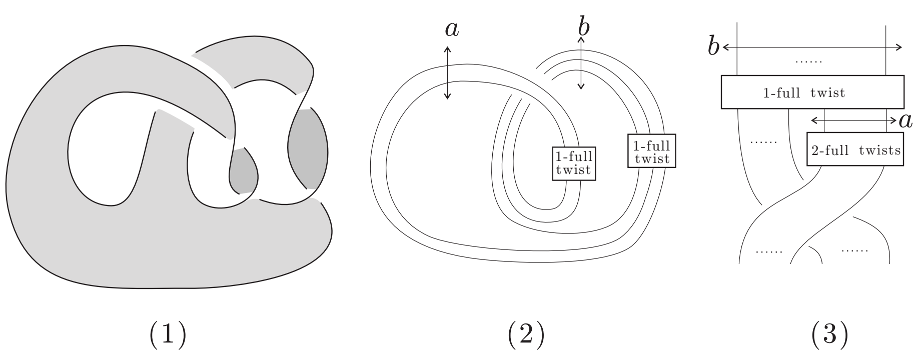

Figure 1. The knot $k^+(a,b)$. Check all longer captions.

Lemma 5.1. *$k^+(a,b)$ is a fibered knot with genus $\frac{1}{2}((a+b-1)^2-ab)$.*

Proof. It is easy to see that $k^+(a,b)$ has a form of the closure of a positive braid as shown in Figure 1(3). By [Stallings 1978], Seifert's algorithm gives a fiber surface. The braid has b strings and a^2+ab+b^2-2a-b crossings, so the fiber has the given genus. (See also [Yamada 2005, Corollary 3] or [Hill and Murasugi 2000].) □

In general, let K be a knot in S^3 whose p-surgery yields $L(p,q)$ with $p > q > 0$. Then the core K^* of the attached solid torus of $K(p)$ is called the *dual knot* of K (with respect to p-surgery). Berge [1990s] shows that if K is a doubly primitive knot whose surface slope is p, then K^* is a $(1,1)$-knot in $L(p,q)$ and has a canonical form parametrized by a single integer k with $0 < k < p$ (see [Saito 2007; 2008b]). Following [Saito 2007], we denote it by $K(L(p,q);k)$. It is known that $K(L(p,q);k)$ is isotopic to $K(L(p,q);\, p-k)$.

For $n = 1, 2, \dots, p-1$, let ϕ_n be an integer such that $\phi_n \equiv nq \pmod p$ and $0 < \phi_n < p$. We call this finite sequence $\{\phi_n\}$ the *basic sequence* for (p,q). Because of $\gcd(p,q) = 1$, the ϕ_n are mutually distinct. In particular, k appears in the basic sequence. Let h be the position of k, that is, $\phi_h = k$. Here, set

$$s = \sharp\{i \mid i < k \text{ and } i \text{ appears before } k \text{ in the basic sequence}\},$$
$$\ell = \sharp\{i \mid i > k \text{ and } i \text{ appears before } k \text{ in the basic sequence}\},$$
$$s' = \sharp\{i \mid i < k \text{ and } i \text{ appears after } k \text{ in the basic sequence}\},$$
$$\ell' = \sharp\{i \mid i > k \text{ and } i \text{ appears after } k \text{ in the basic sequence}\}.$$

Let $\Phi = \min\{s, s', \ell, \ell'\}$. This is determined for the triplet (p,q,k) and so also for the dual knot $K(L(p,q);k)$. However, the main result of [Saito 2008a] says that Φ depends only on the original knot K and a lens space surgery slope p, and that K is hyperbolic if and only if $\Phi \geq 2$ or equivalently each of s, s', ℓ, ℓ' is at least two.

For $k^+(a,b)$, let $p = a^2+ab+b^2$. Then p-surgery yields a lens space $L(p,q)$ where $q \equiv (b/(a+b))^2$; note that $(a/b)^2 \equiv (b/(a+b))^2 \pmod p$. By [Saito 2007], the dual knot is represented as $K(L(p,q);k)$ with $k \equiv -b/(a+b) \pmod p$. (By definition, the parameter k is chosen so that $0 < k < p$.)

Lemma 5.2. *Let p, q and k be defined as above.*

(1) $k+q+1 \equiv 0 \pmod p$.

(2) $k \equiv q^2 \pmod p$.

(3) $kq \equiv 1 \pmod p$.

Proof. (1) We compute

$$k+q+1 \equiv -\frac{b}{a+b}+\frac{b^2}{(a+b)^2}+1=\frac{-b(a+b)+b^2+(a+b)^2}{(a+b)^2}$$
$$=\frac{(a+b)^2-ab}{(a+b)^2}=\frac{p}{(a+b)^2}\equiv 0 \pmod{p}.$$

(2) By (1), $q^2-k \equiv (-k-1)^2-k=k^2+k+1 \equiv q+k+1 \equiv 0 \pmod{p}$.

(3) Similarly, $kq \equiv k(-k-1)=-k^2-k \equiv -q-k \equiv 1 \pmod{p}$ by (1). □

5.3. ***The knot $k^+(3n+1, 3n+4)$.*** For $k^+(3n+1, 3n+4)$, let $p=27n^2+45n+21$. Then p-surgery yields a lens space $L(p,q)$ with $q=(3n+2)^2$, and the dual knot is $K(L(p,q);k)$ with $k \equiv -(3n+2)^2-1 \pmod{p}$.

Lemma 5.4. *The knot $k^+(3n+1, 3n+4)$ is hyperbolic for $n \geq 1$.*

Proof. Let $a=3n+1$, $b=3n+4$ and $k_0=p-q-1$. Then direct calculations show that $3q<p<4q$, $k_0 \equiv k \pmod{p}$ and $2q-1<k_0<3q-a$. Thus we can use the triplet (p,q,k_0) to calculate the invariant Φ.

Let $\{\phi_i\}$ be the basic sequence. (Recall that any term ϕ_i of the basic sequence is chosen so that $0<\phi_i<p$.) Since $q^2 \equiv k \equiv k_0 \pmod{p}$ by Lemma 5.2, $\phi_q=k_0$.

First, we study the four consecutive terms $\phi_{a+b}, \phi_{a+b+1}, \phi_{a+b+2}, \phi_{a+b+3}$, which appear before k_0. Since $(a+b)q \equiv p-a \pmod{p}$, we have $\phi_{a+b}=p-a$. Then

$$q-a<2q-a<k_0<3q-a<p-a,$$

so

$$\phi_{a+b+1}=q-a, \quad \phi_{a+b+2}=2q-a, \quad \phi_{a+b+3}=3q-a.$$

Hence

$$\phi_{a+b}>k_0, \quad \phi_{a+b+1}<k_0, \quad \phi_{a+b+2}<k_0, \quad \phi_{a+b+3}>k_0.$$

Similarly, we study the four consecutive terms right after k_0. (Since $q+4<p-1$, there are more than four terms after k_0.) Since $k_0+q=p-1$, we have $\phi_{q+1}=p-1$. Then

$$\phi_{q+2}=q-1, \quad \phi_{q+3}=2q-1, \quad \phi_{q+4}=3q-1.$$

Hence

$$\phi_{q+1}>k_0, \quad \phi_{q+2}<k_0, \quad \phi_{q+3}<k_0, \quad \phi_{q+4}>k_0.$$

Thus $\Phi \geq 2$, showing that the dual knot (and the original knot) is hyperbolic. □

5.5. ***The knot $k^+(F_{n+2}, F_n)$.*** For Fibonacci numbers, see Section 2. Let $p=F_n^2+F_nF_{n+2}+F_{n+2}^2$. Then p-surgery on $k^+(F_{n+2}, F_n)$ yields a lens space $L(p,q)$ with $q \equiv (F_n/(F_{n+2}+F_n))^2 \pmod{p}$. Denote the dual knot by $K(L(p,q);k)$.

Lemma 5.6. *We have $p=4F_nF_{n+2}+(-1)^n$, $q \equiv (-1)^{n+1}4F_n^2 \pmod{p}$, and $k \equiv (-1)^n 4F_n(F_n+F_{n+2}) \pmod{p}$.*

Proof. By Cassini's identity $F_nF_{n+2} - F_{n+1}^2 = (-1)^{n+1}$,

$$\begin{aligned}
4F_nF_{n+2} + (-1)^n &= 4F_{n+1}^2 + 3(-1)^{n+1} \\
&= 3(F_{n+1}^2 + (-1)^{n+1}) + F_{n+1}^2 \\
&= 3F_nF_{n+2} + F_{n+1}^2 \\
&= F_nF_{n+2} + 2F_n(F_n + F_{n+1}) + F_{n+1}^2 \\
&= F_nF_{n+2} + 2F_n^2 + 2F_nF_{n+1} + F_{n+1}^2 \\
&= F_nF_{n+2} + F_n^2 + (F_n + F_{n+1})^2 = F_nF_{n+2} + F_n^2 + F_{n+2}^2 = p.
\end{aligned}$$

Thus $4F_nF_{n+2} + (-1)^n \equiv 0 \pmod p$. To show $q \equiv (-1)^{n+1}4F_n^2 \pmod p$, it suffices to show $(-1)^{n+1}4(F_n + F_{n+2})^2 \equiv 1 \pmod p$. This follows from the equation $(F_n + F_{n+2})^2 \equiv F_nF_{n+2} \pmod p$.

Finally,

$$(-1)^n 4F_n(F_n + F_{n+2}) = (-1)^n 4(F_n^2 + F_nF_{n+2}) \equiv (-1)^{n+1}4F_{n+2}^2 \pmod p.$$

Then $(-1)^{n+1}4F_{n+2}^2 q \equiv (4F_{n+2}F_n)^2 \equiv 1 \pmod p$. This shows by Lemma 5.2(3) that $(-1)^{n+1}4F_n(F_n + F_{n+2}) \equiv 1/q \equiv k \pmod p$. □

Lemma 5.7. *For $n \geq 3$, the knot $k^+(F_{n+2}, F_n)$ is hyperbolic.*

Proof. As mentioned above, p-surgery on $k^+(F_{n+2}, F_n)$ yields $L(p, q)$. Consider the dual knot $K(L(p, q); k)$ in $L(p, q)$.

First, we assume that n is odd. Then $p = 4F_nF_{n+2} - 1$, $q \equiv 4F_n^2 \pmod p$, and $k \equiv -4F_n(F_n + F_{n+2}) \pmod p$ by Lemma 5.6.

To simplify calculation of the invariant Φ, put $q_0 = p - 4F_n^2$ and $k_0 = p - q_0 + 1$. Then $0 < q_0 < p$ and $0 < k_0 < p$, and $q_0 \equiv -q \pmod p$ and $k_0 \equiv -k \pmod p$.

Claim 5.8. *$3q_0/2 < p < 2q_0$ and $2q_0 - p < k_0 < q_0$.*

Proof. $2q_0 - p = p - 8F_n^2 = 4F_nF_{n+2} - 1 - 8F_n^2 = 4F_n(F_n + F_{n+1}) - 8F_n^2 - 1 = 4F_n(F_{n+1} - F_n) - 1 \geq 7$. Since $3F_n > F_{n+2}$, we have $2p - 3q_0 = 12F_n^2 - p = 4F_n(3F_n - F_{n+2}) + 1 \geq 9$.

Next, $k_0 - 2q_0 + p = 2p - 3q_0 + 1 \geq 10$. Finally, $q_0 - k_0 = 2q_0 - p - 1 \geq 6$. □

For (p, q_0), let $\{\phi_i\}$ be the basic sequence; let $h = p - q_0$. As $hq_0 \equiv k_0 \pmod p$, the number k_0 appears as the h-th term in the sequence. Note that $h > 4$, because $2p - 3q_0 \geq 9$.

To evaluate Φ, we investigate some specific terms in the basic sequence. We have $\phi_1 = q_0 > k_0$ and $\phi_2 = 2q_0 - p < k_0$. Also, $\phi_{h-1} = k_0 - q_0 + p > k_0$ and $\phi_{h-2} = k_0 - 2q_0 + p < k_0$. Since $h > 4$, these four terms $\phi_1, \phi_2, \phi_{h-2}, \phi_{h-1}$ are distinct. Next, $\phi_{p-1} = p - q_0 < k_0$ and $\phi_{p-2} = 2p - 2q_0 > k_0$. Since $2h < p$, $h < p - h + 1$. Thus ϕ_{p-h+1} and ϕ_{p-h+2}, which are distinct from ϕ_{p-1} and ϕ_{p-2},

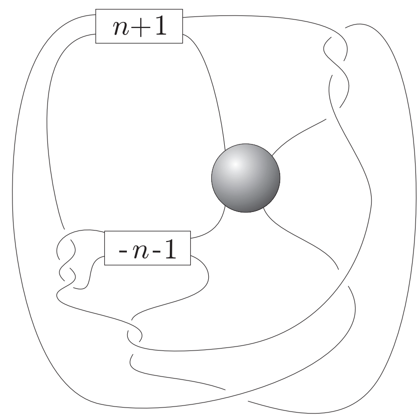

Figure 2. The tangle B_n.

appear after k_0 in the basic sequence. Since $(p-h+1)q_0 \equiv q_0 - k_0 \pmod p$, we have $\phi_{p-h+1} = q_0 - k_0$. Then $k_0 - (q_0 - k_0) = 2k_0 - q_0 = 2p - 3q_0 + 2 > 0$ implies $\phi_{p-h+1} < k_0$. Finally, $\phi_{p-h+2} = 2q_0 - k_0 > k_0$. Again, the fact $h > 4$ means that the four terms $\phi_{p-h+1}, \phi_{p-h+2}, \phi_{p-2}, \phi_{p-1}$ are distinct. Hence $\Phi \geq 2$.

Second, assume that n is even. Then $p = 4F_nF_{n+2} + 1$, $q \equiv -4F_n^2 \pmod p$, and $k \equiv 4F_n(F_n + F_{n+2}) \pmod p$ by Lemma 5.6. In this case, put $q_0 = p - 4F_n^2 + 1$ and $k_0 = p - q_0 + 1$. Then $0 < q_0 < p$ and $0 < k_0 < p$. It is easy to check that Claim 5.8 holds without any change.

By Lemma 5.2, $q_0 q \equiv (q+1)q \equiv k + q \equiv -1 \pmod p$ and $k_0 \equiv -q \pmod p$. Under a (orientation-reversing) homeomorphism from $L(p,q)$ to $L(p,q_0)$, the dual knot $K(L(p,q);k)$ is mapped to $K(L(p,q_0);k_0)$; see [Saito 2008a]. Thus we can use (p, q_0, k_0) instead of (p,q,k) to evaluate Φ.

By Lemma 5.2(2), $q^2 \equiv k \pmod p$. Thus $16F_n^4 \equiv 4F_n(F_n + F_{n+2})$. Hence $q_0^2 + k_0 \equiv (1 - 4F_n^2)^2 + 4F_n^2 \equiv 16F_n^4 - 4F_n^2 + 1 \equiv 4F_n(F_n + F_{n+2}) - 4F_n^2 + 1 \equiv 4F_n^2F_{n+2}^2 + 1 \equiv 0 \pmod p$. This means that $(p - q_0)q_0 \equiv k_0 \pmod p$. Let $h = p - q_0$. Then, k_0 appears in the basic sequence for (p, q_0) as the h-term. Since $h > 4$, the argument in the case where n is odd works verbatim, so we have $\Phi \geq 2$. □

6. Hyperbolic knots

We say a Seifert-fibered manifold is *of type* $X(p_1, p_2, \ldots, p_n)$ if it admits a Seifert fibration over the surface X with n exceptional fibers of indices $p_1, p_2, \ldots, p_n$. In this paper, X will be either the 2-sphere S^2 or the disk D^2.

For $n \geq 1$, let B_n be the tangle illustrated in Figure 2, in which a rectangle denotes horizontal half-twists. If the number is positive, the twist is right-handed; otherwise, it is left-handed.

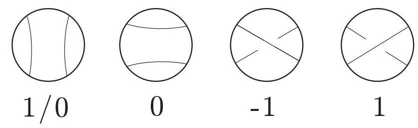

Figure 3. Some rational tangles.

Given $\alpha \in \mathbb{Q} \cup \{1/0\}$, we denote by $B_n(\alpha)$ the knot or link in S^3 obtained by inserting the rational tangle of slope α into the central puncture of B_n. Also, $\widetilde{B}_n$ is the double branched cover of S^3 branched over $B_n(\alpha)$. In fact, we need only four rational tangles as shown in Figure 3.

Lemma 6.1. (1) $\widetilde{B}_n(1/0) = S^3$.

(2) $\widetilde{B}_n(0) = L(27n^2 + 45n + 21, -9n^2 - 12n - 5)$.

(3) $\widetilde{B}_n(1)$ *is a Seifert fibered manifold of type* $S^2(2, n+2, 15n+11)$.

(4) $\widetilde{B}_n(-1)$ *is a non-Seifert toroidal manifold* $D^2(2, n) \cup D^2(2, 3n+1)$, *which contains a unique incompressible torus if* $n \geq 2$, *or a Seifert fibered manifold of type* $S^2(2, 3, 4)$ *if* $n = 1$.

Proof. It is straightforward to check that $B_n(1/0)$ is the unknot and that $B_n(0)$ is the 2-bridge knot corresponding to $-(9n^2 + 12n + 5)/(27n^2 + 45n + 21)$.

Figure 4 shows that $B_n(1)$ is a Montesinos link or knot of length three. Thus $\widetilde{B}_n(1)$ is a Seifert fibered manifold of type $S^2(2, n+2, 15n+11)$.

Figure 5 shows that $B_n(-1)$ is decomposed along a tangle sphere P into two tangles. If $n > 1$, then each side of P is a Montesinos tangle. Thus $\widetilde{B}_n(1)$ is decomposed along a torus into two Seifert fibered manifolds over the disk with two exceptional fibers. Since Seifert fibers on both sides intersect once on the torus, $\widetilde{B}_n(-1)$ is not Seifert. It is well known that such a 3-manifold contains a

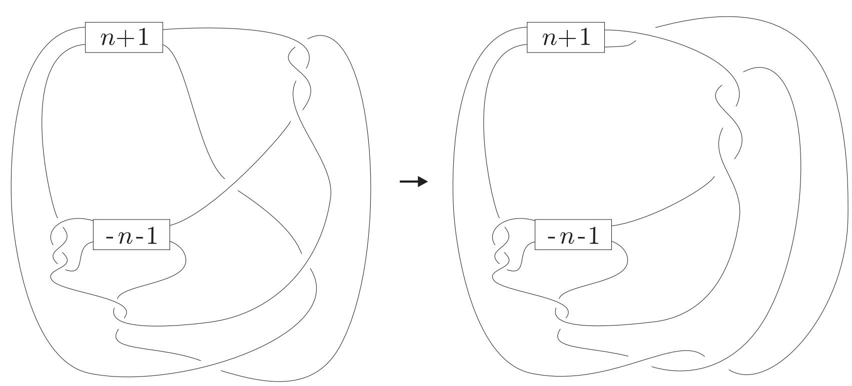

Figure 4. $B_n(1)$, a Montesinos link.

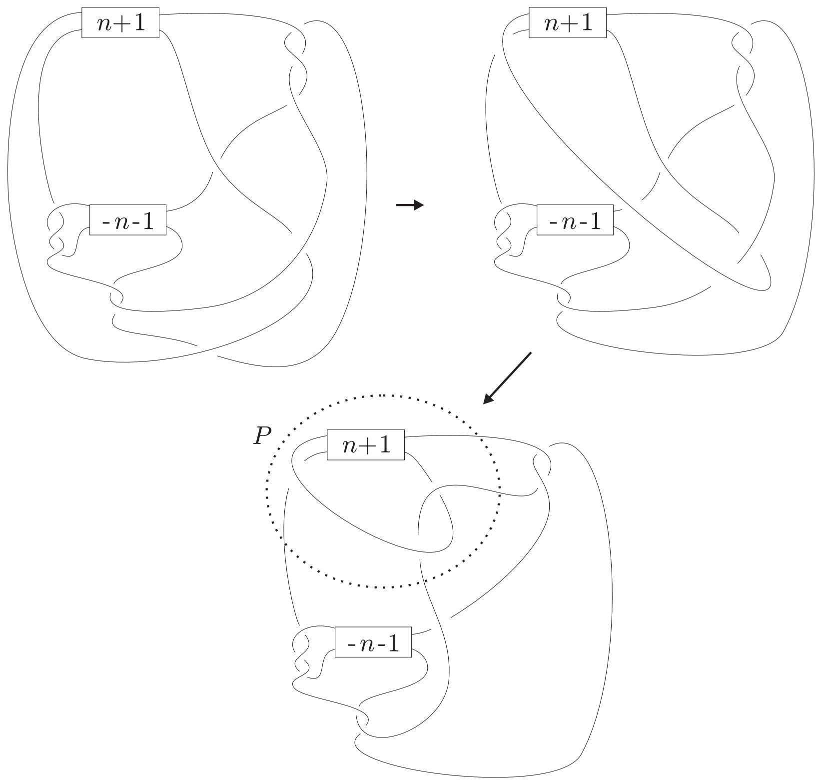

Figure 5. $B_n(-1)$.

unique incompressible torus. When $n = 1$, $B_n(-1)$ is a Montesinos link of length three. Hence $\widetilde{B}_n(-1)$ is a Seifert fibered manifold over the 2-sphere with three exceptional fibers. □

By Lemma 6.1(1), the lift of B_n in $\widetilde{B}_n(1/0)$ gives the knot exterior of some knot K_n in S^3, which is uniquely determined by Gordon and Luecke's theorem [1989]. Furthermore, K_n admits integral Dehn surgeries yielding a lens space, a Seifert fibered manifold, and a toroidal manifold (unless $n = 1$) by Lemma 6.1.

The following criterion of hyperbolicity is used also in Section 7.

Lemma 6.2. *If a knot K in S^3 admits an integral lens space surgery m, and neither $K(m-1)$ nor $K(m+1)$ has a lens space summand, then K is hyperbolic.*

Proof. Assume the contrary. Then K is either a torus knot or a satellite knot. For the (nontrivial) (p, q)-torus knot, the only integral lens space surgery slopes are $pq - 1$ and $pq + 1$, and pq-surgery yields the connected sum of two lens spaces by [Moser 1971]. Thus K is not a torus knot.

Assume K is a satellite knot. Since K has a lens space surgery, we know by [Bleiler and Litherland 1989; Wang 1989; Wu 1990] that K is the $(2, 2pq+\varepsilon)$-cable of the (p, q)-torus knot where $\varepsilon \in \{1, -1\}$. Then the lens space surgery is $4pq+\varepsilon$. However, the adjacent slope $4pq+2\varepsilon$ is equal to the cabling slope, and so $K(4pq+2\varepsilon)$ has a lens space summand, a contradiction. Thus K is hyperbolic. □

Lemma 6.3. *K_n is hyperbolic.*

Proof. This immediately follows from Lemmas 6.1 and 6.2. □

Lemma 6.4. *The knot K_n defined above satisfies the following.*

(1) *The genus of K_n is $(27n^2 + 33n + 10)/2$.*

(2) *Let $m = 27n^2 + 45n + 21$. Then m-surgery on K yields the lens space $L(m, -9n^2 - 12n - 5)$.*

Proof. Insert the $1/0$-tangle to B_n, and put a band b as shown in Figure 6 to keep track of framing. Isotope the unknot $B_n(1/0)$ to a standard diagram as shown in Figure 8 (in which the cases $n = 5$ and $n = 4$ are drawn), and take the double branched cover along it. Then (the core of) the lift of b gives K_n, and its framing corresponds to the 0-tangle filling downstairs. (In Figures 6, 7 and 8, we draw b in a line for simplicity during the deformation.) From Figure 8, we see that K_n is the closure of a braid with $3n+2$ strings. Moreover, there are $27n^2+41n+10$ positive crossings and $5n-1$ negative crossings. After canceling the negative crossings by positive crossings, K_n becomes the closure of a positive braid with $3n+2$ strings and $27n^2 + 36n + 11$ crossings. By [Stallings 1978], K_n is fibered and Seifert's algorithm gives a fiber surface, whose genus is equal to the genus $g(K_n)$ of K_n. Now (1) follows because $1 - 2g(K_n) = (3n + 2) - (27n^2 + 36n + 11)$.

The framing of the lift of b can be calculated to equal m. This proves (2). □

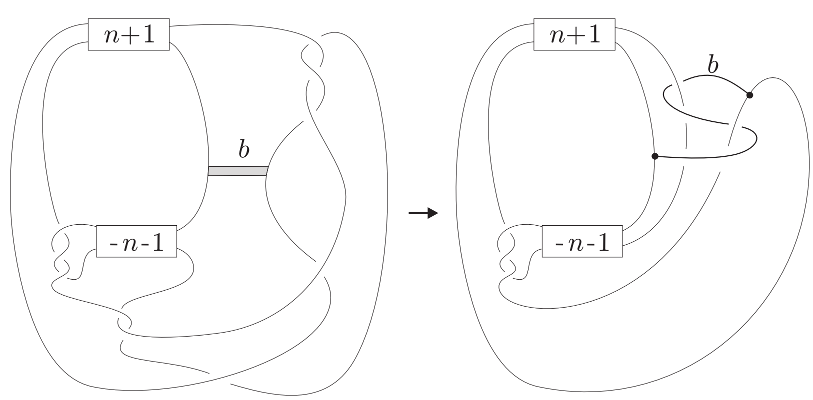

Figure 6. $B_n(1/0)$ and the band b.

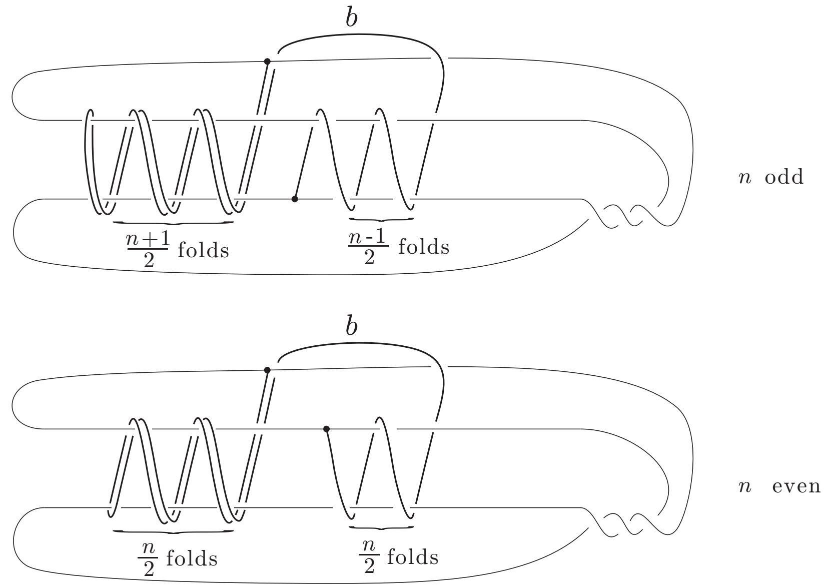

Figure 7. $B_n(1/0)$ with b.

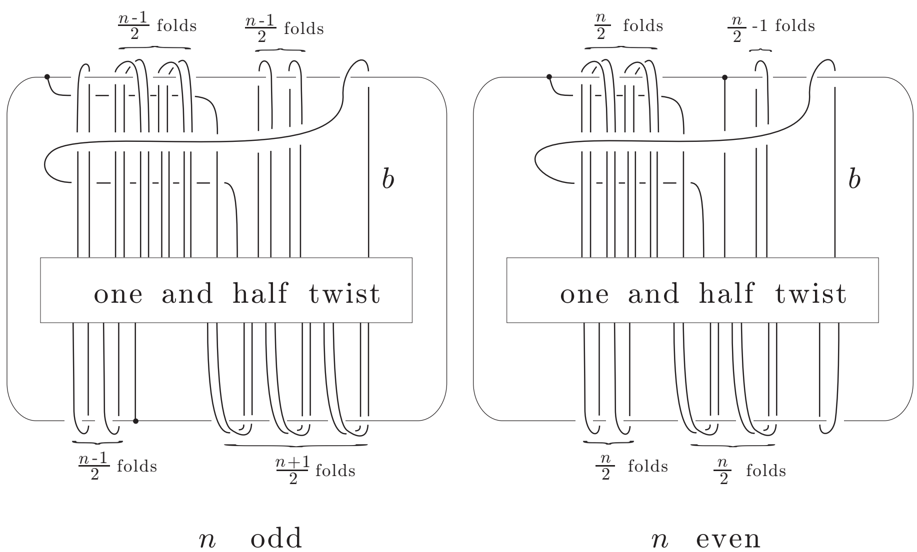

Figure 8. The standard diagram of $B_n(1/0)$ with b.

Recall that $k^+(3n+1, 3n+4)$ is hyperbolic for $n \geq 1$ by Lemma 5.4.

Proposition 6.5. *For $n \geq 1$, let K be the hyperbolic knot K_n defined above, and let K' be $k^+(3n+1, 3n+4)$. Let $m = 27n^2 + 45n + 21$. Then K and K' are not equivalent, and m-surgery on K and K' yield homeomorphic lens spaces.*

Proof. By Lemma 6.4(1), K has genus $(27n^2+33n+10)/2$, while K' has genus $(27n^2+33n+12)/2$ by Lemma 5.1. Thus they are not equivalent.

Also, by Lemma 6.4(2), m-surgery on K yields $L(m, -9n^2-12n-5) = L(m, 18n^2+33n+16)$. As we stated in Section 5, m-surgery on K' yields $L(m, ((3n+1)/(3n+4))^2)$. Those lens spaces are homeomorphic since

$$\left(\frac{3n+1}{3n+4}\right)^2(18n^2+33n+16) \equiv 1 \pmod{m}. \qquad \square$$

7. Different classes

In this last section, we give pairs of knots, each of which yields homeomorphic lens spaces by the same integral surgery, and consist of knots belonging to different classes of hyperbolic, satellite, torus knots.

7.1. ***Torus knot and satellite knot.*** Let $C(a,b)$ be the $(2, 2ab+1)$-cable of the torus knot of type (a,b).

Proposition 7.2. *For $n \geq 1$, let K be the torus knot of type $(2n+1, 4n+4)$, and let $K' = C(n+1, 2n+1)$. Let $m = 8n^2+12n+5$. Then m-surgery on K and K' yields homeomorphic lens spaces.*

Proof. By [Moser 1971], m-surgery on K yields the lens space $L(m,(2n+1)^2)$. Also, m-surgery on K' yields $L(m, 4(n+1)^2)$ by [Fintushel and Stern 1980]. Since $(2n+1)^2+4(n+1)^2 = m$, these lens spaces are homeomorphic. $\square$

7.3. ***Satellite knot and hyperbolic knot.***

Lemma 7.4. *For $n \geq 0$,*

$$4F_n^4 + (-1)^n F_{n+2}^2 = (4F_nF_{n+2} + (-1)^n)(F_{n+2}^2 - 4F_nF_{n+1}).$$

Proof. First,

$$\begin{aligned} 4F_n^4 &+ (-1)^n F_{n+2}^2 - (4F_nF_{n+2} + (-1)^n)(F_{n+2}^2 - 4F_nF_{n+1}) \\ &= 4F_n(F_n^3 - F_{n+2}^3 + 4F_nF_{n+1}F_{n+2} - (-1)^{n+1}F_{n+1}). \end{aligned}$$

From Cassini's identity,

$$\begin{aligned} &F_n^3 - F_{n+2}^3 + 4F_nF_{n+1}F_{n+2} - (-1)^{n+1}F_{n+1} \\ &= F_n^3 - F_{n+2}^3 + 3F_nF_{n+1}F_{n+2} + F_{n+1}^3 \\ &= F_n^3 - (F_n+F_{n+1})^3 + 3F_nF_{n+1}F_{n+2} + F_{n+1}^3 \\ &= -3F_nF_{n+1}(F_n+F_{n+1}-F_{n+2}) = 0. \end{aligned} \qquad \square$$

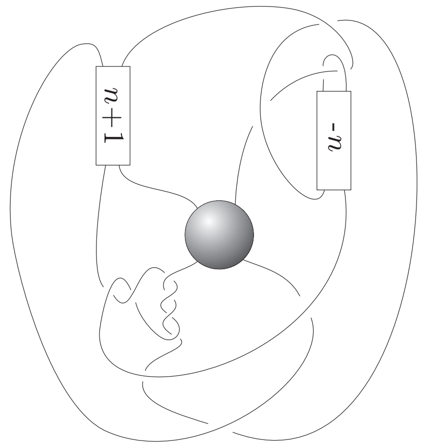

Figure 9. The tangle B_n.

Since $\gcd(F_n, F_{n+2}) = \gcd(F_n, F_{n+1}) = 1$, the Fibonacci numbers F_n and F_{n+2} are coprime. By Lemma 5.7, $k^+(F_{n+2}, F_n)$ is hyperbolic for $n \geq 3$.

Proposition 7.5. *For $n \geq 3$, let K be the satellite knot $C(F_n, F_{n+2})$, and let K' be the hyperbolic knot $k^+(F_{n+2}, F_n)$. Let $m = 4F_nF_{n+2} + (-1)^n$. Then m-surgery on K and K' yields homeomorphic lens spaces.*

Proof. By [Fintushel and Stern 1980], m-surgery on K yields the lens space $L(m, 4F_n^2)$. From Lemma 5.6, m-surgery on K' yields $L(m, (F_n/F_{n+2})^2)$. Then

$$4F_n^2\left(\frac{F_n}{F_{n+2}}\right)^2 \equiv (-1)^{n+1} \pmod{m}$$

by Lemma 7.4, Thus the two lens spaces are homeomorphic. □

7.6. ***Torus knot and hyperbolic knot.*** For $n \geq 1$, let B_n be the tangle as shown in Figure 9, where a vertical box denotes right-handed vertical half-twists.

Given $\alpha \in \mathbb{Q} \cup \{1/0\}$, we denote by $B_n(\alpha)$ the knot or link in S^3 obtained by inserting the rational tangle of slope α into the central puncture of B_n. Also, $\widetilde{B}_n$ is the double branched cover of S^3 branched over $B_n(\alpha)$.

Lemma 7.7. (1) $\widetilde{B}_n(1/0) = S^3$.

(2) $\widetilde{B}_n(0) = L(18n^2 + 33n + 15, 18n + 19)$.

(3) $\widetilde{B}_n(-1)$ *is a non-Seifert toroidal manifold* $D^2(2, n+2) \cup D^2(4, 2n+1)$.

(4) $\widetilde{B}_n(1)$ *is a non-Seifert toroidal manifold* $D^2(2, n) \cup D^2(5, 2n+3)$ *if* $n \geq 2$, *and a Seifert fibered manifold of type* $S^2(3, 5, 5)$ *if* $n = 1$.

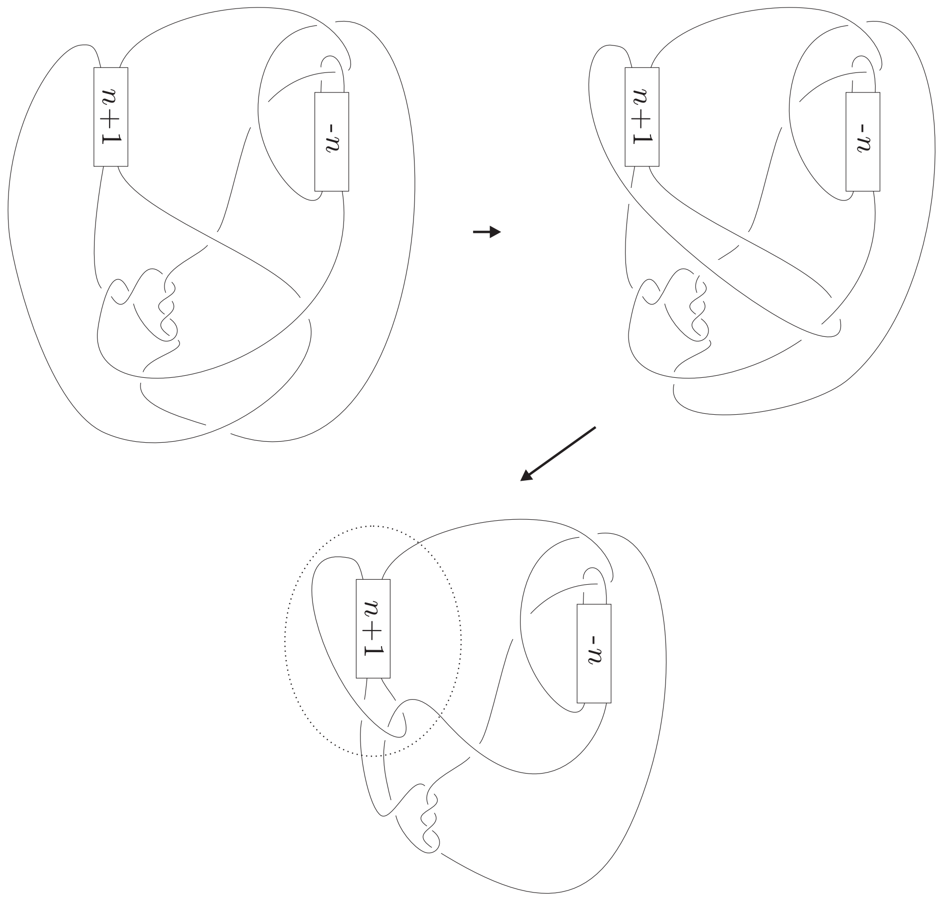

Figure 10. $B_n(-1)$.

Proof. It is straightforward to see that $B(1/0)$ is the unknot and $B(0)$ is the 2-bridge link corresponding to $(18n^2+33n+15)/(18n+19)$. For $B(-1)$ and $B(1)$, see Figures 10 and 11, respectively. □

By Lemma 7.7(1), the lift of B_n in $\widetilde{B}_n(1/0)$ gives the knot exterior of some knot K_n in S^3, which is uniquely determined by Gordon and Luecke's theorem [1989].

Lemma 7.8. *K_n is hyperbolic.*

Proof. This immediately follows from Lemmas 6.2 and 7.7. □

Lemma 7.9. *Let $m = 18n^2+33n+15$. Then m-surgery on K_n yields the lens space $L(m, -18n-19)$.*

Proof. The argument is similar to the proof of Lemma 6.4. We omit it. □

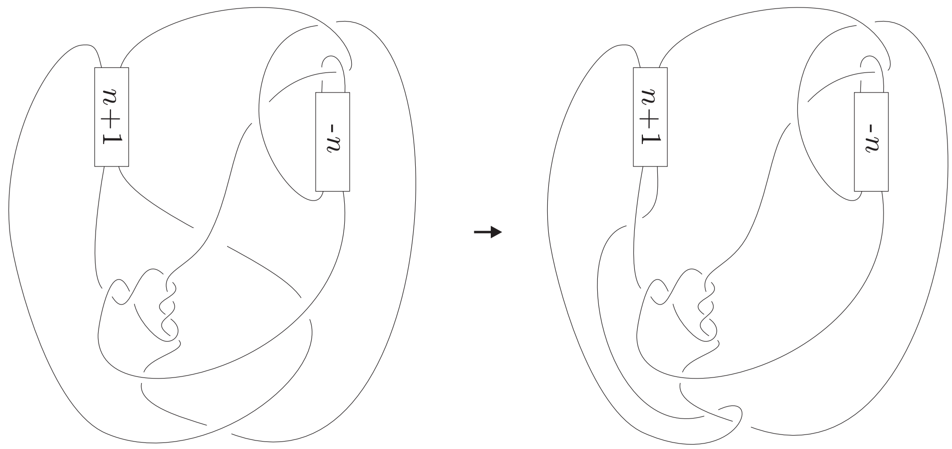

Figure 11. $B_n(1)$.

Proposition 7.10. *For $n \geq 1$, let K be the torus knot of type $(3n+2, 6n+7)$, and let K' be the knot K_n defined above. Let $m = 18n^2+33n+15$. Then m-surgery on K and K' yields homeomorphic lens spaces.*

Proof. By [Moser 1971], m-surgery on K yields $L(m, 9n^2+12n+4)$. Then by Lemma 7.9, m-surgery on K' yields $L(m, 18n+19)$. Since

$$(9n^2+12n+4)(18n+19) \equiv 1 \pmod{m},$$

two lens spaces are homeomorphic. □

Theorem 1.1 now follows from Propositions 2.2, 6.5, 7.2, 7.5 and 7.10. □

Acknowledgment

We would like to thank Kazuhiro Kawasaki for performing computer experiments.

References

[Berge 1990s] J. Berge, "Some knots with surgeries yielding lens spaces", unpublished manuscript, 1990s.

[Bleiler and Litherland 1989] S. A. Bleiler and R. A. Litherland, "Lens spaces and Dehn surgery", *Proc. Amer. Math. Soc.* **107**:4 (1989), 1127–1131. MR 90e:57031 Zbl 0686.57007

[Brakes 1980] W. R. Brakes, "Manifolds with multiple knot-surgery descriptions", *Math. Proc. Cambridge Philos. Soc.* **87**:3 (1980), 443–448. MR 81b:57002 Zbl 0443.57010

[Culler et al. 1987] M. Culler, C. M. Gordon, J. Luecke, and P. B. Shalen, "Dehn surgery on knots", *Ann. of Math.* (2) **125**:2 (1987), 237–300. MR 88a:57026 Zbl 0633.57006

[Fintushel and Stern 1980] R. Fintushel and R. J. Stern, "Constructing lens spaces by surgery on knots", *Math. Z.* **175**:1 (1980), 33–51. MR 82i:57009a Zbl 0425.57001

[Flath 1989] D. E. Flath, *Introduction to number theory*, Wiley, New York, 1989. MR 90c:11001 Zbl 0651.10001

[Gordon and Luecke 1989] C. M. Gordon and J. Luecke, "Knots are determined by their complements", *J. Amer. Math. Soc.* **2**:2 (1989), 371–415. MR 90a:57006a Zbl 0678.57005

[Graham et al. 1994] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: A foundation for computer science*, 2nd ed., Addison-Wesley, Reading, MA, 1994. MR 97d:68003 Zbl 0836.00001

[Hill and Murasugi 2000] P. Hill and K. Murasugi, "On double-torus knots, II", *J. Knot Theory Ramifications* **9**:5 (2000), 617–667. MR 2001f:57006 Zbl 0991.57011

[Kawauchi 1996] A. Kawauchi, "Mutative hyperbolic homology 3-spheres with the same Floer homology", *Geom. Dedicata* **61**:2 (1996), 205–217. MR 97g:57020 Zbl 0857.57012

[Kronheimer et al. 2007] P. Kronheimer, T. Mrowka, P. Ozsváth, and Z. Szabó, "Monopoles and lens space surgeries", *Ann. of Math.* (2) **165**:2 (2007), 457–546. MR 2008b:57037 Zbl 05180741

[Lickorish 1976] W. B. R. Lickorish, "Surgery on knots", *Proc. Amer. Math. Soc.* **60** (1976), 296–298. MR 58 #7632 Zbl 0345.57002

[Livingston 1982] C. Livingston, "More 3-manifolds with multiple knot-surgery and branched-cover descriptions", *Math. Proc. Cambridge Philos. Soc.* **91**:3 (1982), 473–475. MR 83j:57002

[Moser 1971] L. Moser, "Elementary surgery along a torus knot", *Pacific J. Math.* **38** (1971), 737–745. MR 52 #4287 Zbl 0202.54701

[Osoinach 1998] J. K. Osoinach, Jr., *Manifolds obtained by Dehn surgery on infinitely many distinct knots in* S^3, thesis, University of Texas at Austion, 1998.

[Osoinach 2006] J. K. Osoinach, Jr., "Manifolds obtained by surgery on an infinite number of knots in S^3", *Topology* **45**:4 (2006), 725–733. MR 2007m:57010 Zbl 1097.57018

[Saito 2007] T. Saito, "Dehn surgery and (1, 1)-knots in lens spaces", *Topology Appl.* **154**:7 (2007), 1502–1515. MR 2008f:57011 Zbl 1115.57005

[Saito 2008a] T. Saito, "The dual knots of doubly primitive knots", *Osaka J. Math.* **45**:2 (2008), 403–421. MR 2009e:57014 Zbl 1146.57012

[Saito 2008b] T. Saito, "Knots in lens spaces with the 3-sphere surgery", *Algebr. Geom. Topol.* **8**:1 (2008), 53–79. MR 2009d:57014 Zbl 05236552

[Stallings 1978] J. R. Stallings, "Constructions of fibred knots and links", pp. 55–60 in *Algebraic and geometric topology* (Stanford, CA, 1976), vol. 2, edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, R.I., 1978. MR 80e:57004 Zbl 0394.57007

[Teragaito 1994] M. Teragaito, "Homology handles with multiple knot-surgery descriptions", *Topology Appl.* **56**:3 (1994), 249–257. MR 95b:57013 Zbl 0811.57020

[Teragaito 2007] M. Teragaito, "A Seifert fibered manifold with infinitely many knot-surgery descriptions", *Int. Math. Res. Not.* **2007**:9 (2007), Art. ID rnm 028. MR 2008g:57008 Zbl 1138.57012

[Wang 1989] S. C. Wang, "Cyclic surgery on knots", *Proc. Amer. Math. Soc.* **107**:4 (1989), 1091–1094. MR 90e:57030 Zbl 0688.57008

[Wu 1990] Y. Q. Wu, "Cyclic surgery and satellite knots", *Topology Appl.* **36**:3 (1990), 205–208. MR 91k:57009 Zbl 0715.57002

[Yamada 2005] Y. Yamada, "Berge's knots in the fiber surfaces of genus one, lens space and framed links", *J. Knot Theory Ramifications* **14**:2 (2005), 177–188. MR 2005m:57016 Zbl 1086.57010

Received August 20, 2008.

TOSHIO SAITO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106
UNITED STATES

tsaito@math.ucsb.edu

MASAKAZU TERAGAITO
DEPARTMENT OF MATHEMATICS AND MATHEMATICS EDUCATION
HIROSHIMA UNIVERSITY
1-1-1 KAGAMIYAMA
HIGASHI-HIROSHIMA 739-8524
JAPAN

teragai@hiroshima-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 244, No. 1, 2010

A TOPOLOGICAL SPHERE THEOREM FOR ARBITRARY-DIMENSIONAL MANIFOLDS

YULIANG WEN

We study manifolds with bounded volume, sectional curvature, and injectivity radius. We obtain a topological sphere theorem.

Sphere theorems are common in differential geometry; one often asks whether a manifold is homeomorphic to a sphere under certain topological or geometric restrictions; see for instance [Grove and Shiohama 1977; Perelman 1995; Shen 1989; Shiohama 1983; Suyama 1991; Wu 1989]. Coghlan and Itokawa [1991] proved a sphere theorem that says that if an even-dimensional, simply connected Riemannian manifold $\mathcal{M}$ has sectional curvature $K_{\mathcal{M}} \in (0, 1]$, volume $V_{\mathcal{M}} \leq \frac{3}{2} V_{S^n}$ with V_{S^n} the volume of the standard n-dimensional unit sphere S^n in $\mathbb{R}^{n+1}$, then $\mathcal{M}$ must be homeomorphic to S^n. In [Wen 2004], we improved this result by relaxing the upper bound on $V_{\mathcal{M}}$ to a bound larger than $\frac{3}{2} V_{S^n}$. In both of these papers, the hypotheses of simple connectivity and even dimension were merely used to deduce that the injectivity radius $i_{\mathcal{M}}$ is no less than π. Here we find that we can weaken the assumptions on $K_{\mathcal{M}}$ and $i_{\mathcal{M}}$. If the simple connectivity condition is removed, the conclusion holds in *any dimension*.

Before stating our result, we introduce some notation. Let $(\mathcal{M}, g)$ be a compact, connected n-dimensional Riemannian manifold with metric g. We denote by $K_{\mathcal{M}}$ the sectional curvature of $\mathcal{M}$, by $i_{\mathcal{M}}$ its injectivity radius, and by $V_{\mathcal{M}}$ its volume. For any points $P, Q \in \mathcal{M}$, we denote by $\gamma_{P,Q}$ the shortest geodesic on $\mathcal{M}$ from P to Q.

Theorem 1. *Given $k > 0$, there exists an $\varepsilon_0 > 0$ such that if a compact connected n-dimensional Riemannian manifold $(\mathcal{M}, g)$ satisfies*

$$-k^2 \leq K_{\mathcal{M}} \leq 1, \quad i_{\mathcal{M}} \geq \pi - \varepsilon_0, \quad V_{\mathcal{M}} \leq \tfrac{3}{2} V_{s^n} + \varepsilon_0,$$

then $\mathcal{M}$ is homeomorphic to S^n.

The examples of real projective spaces $\mathbb{R}P^n$ for $n \geq 2$ and product manifolds $S^n \times S^m$ for $m, n \geq 1$ show that the hypotheses on the lower bound on $i_{\mathcal{M}}$ or the upper bound on $V_{\mathcal{M}}$ cannot be removed.

MSC2000: 53C20.
Keywords: sphere theorem.
Supported in part by National Science Foundation of China (number 10871069) and Shanghai Leading Academic Discipline Project (project number B407).

In what follows, we denote by $\mathbb{B}(P,r)$ the open geodesic ball in $\mathcal{M}$ with center P and radius r, and by $\overline{\mathbb{B}}(P,r)$ its closure. Also, we denote by $\mathcal{B}_r$ the open geodesic ball in S^n with radius r. Instead of proving Theorem 1 directly, we will prove a more precise version.

Proposition 1. *Let $k > 0$. There exist $\delta, \sigma > 0$ satisfying $\sigma + \delta < \pi$ such that if a compact connected n-dimensional Riemannian manifold $(\mathcal{M}, g)$ satisfies*

$$(1)\qquad -k^2 \le K_{\mathcal{M}} \le 1, \quad i_{\mathcal{M}} \ge \pi - \sigma, \quad V_{\mathcal{M}} \le 3V(\mathcal{B}_{\pi/2-\sigma/2}) + V(\mathcal{B}_{\delta/2}),$$

then $\mathcal{M}$ is homeomorphic to S^n.

Remark 1. The choice of σ or δ here is of course not optimal. We conjecture that $\sigma < \pi/2$ is optimal.

Proof of Proposition 1. We proceed by way of contradiction. Suppose there exists a manifold $\mathcal{M}$ satisfying (1) that is not homeomorphic to S^n. Take points p, q in $\mathcal{M}$ such that $d(p,q) = d_{\mathcal{M}}$, the diameter $d_{\mathcal{M}}$ of $\mathcal{M}$. Then by a well-known topological fact (see for instance [Brown 1960]), there is a point $x_0 \in \mathcal{M} - \mathbb{B}(p, i_{\mathcal{M}}) \cup \mathbb{B}(q, i_{\mathcal{M}})$. Without loss of generality, let $d(q, x_0) \ge d(p, x_0) = l_0$. Therefore $l_0 \ge i_{\mathcal{M}} \ge \pi - \sigma$. First we show an explicit upper bound on $d_{\mathcal{M}}$.

Lemma 1. *$d_{\mathcal{M}} \le \pi - \sigma + \delta$.*

Proof. We argue by contradiction. If $d_{\mathcal{M}} > \pi - \sigma + \delta$, then we consider the balls $\mathbb{B}(p, \pi/2 - \sigma/2 + \delta/2)$, $\mathbb{B}(q, \pi/2 - \sigma/2 + \delta/2)$ and $\mathbb{B}(x_0, l_0 - \pi/2 + \sigma/2 - \delta/2)$. They are obviously pairwise disjoint. Therefore since $K_{\mathcal{M}} \le 1$, Günther's volume comparison theorem gives

$$(2)\qquad V_{\mathcal{M}} \ge 2V(\mathcal{B}_{\pi/2-\sigma/2+\delta/2}) + V(\mathcal{B}_{l_0-\pi/2+\sigma/2-\delta/2}).$$

In what follows, we check that

$$(3)\qquad 2V(\mathcal{B}_{\pi/2-\sigma/2+\delta/2}) + V(\mathcal{B}_{l_0-\pi/2+\sigma/2-\delta/2}) > 3V(\mathcal{B}_{\pi/2-\sigma/2}) + V(\mathcal{B}_{\delta/2}).$$

Noting that $l_0 - \pi/2 + \sigma/2 - \delta/2 \ge \pi/2 - \sigma/2 - \delta/2 > 0$, we have

$$V(\mathcal{B}_{l_0-\pi/2+\sigma/2-\delta/2}) \ge V(\mathcal{B}_{\pi/2-\sigma/2-\delta/2}).$$

By the definition of S^n, we have $V(\mathcal{B}_r) = \omega_{n-1}\int_0^r (\sin t)^{n-1}\,dt$ for any $r > 0$, where ω_{n-1} is the volume of the standard unit $(n-1)$-sphere S^{n-1}. Since $\sin t$ is increasing in $(0, \pi/2)$, we have

$$\begin{aligned}
&\frac{1}{\omega_{n-1}}\big[2V(\mathcal{B}_{\pi/2-\sigma/2+\delta/2}) + V(\mathcal{B}_{l_0-\pi/2+\sigma/2-\delta/2}) - 3V(\mathcal{B}_{\pi/2-\sigma/2}) - V(\mathcal{B}_{\delta/2})\big] \\
&\quad \geq 2\int_0^{\pi/2-\sigma/2+\delta/2} (\sin t)^{n-1}\mathrm{d}t + \int_0^{\pi/2-\sigma/2-\delta/2} (\sin t)^{n-1}\mathrm{d}t \\
&\qquad\qquad - 3\int_0^{\pi/2-\sigma/2} (\sin t)^{n-1}\mathrm{d}t - \int_0^{\delta/2} (\sin t)^{n-1}\mathrm{d}t \\
&\quad = \int_{\pi/2-\sigma/2}^{\pi/2-\sigma/2+\delta/2} (\sin t)^{n-1}\mathrm{d}t - \int_{\pi/2-\sigma/2-\delta/2}^{\pi/2-\sigma/2} (\sin t)^{n-1}\mathrm{d}t \\
&\qquad\qquad + \int_{\pi/2-\sigma/2}^{\pi/2-\sigma/2+\delta/2} (\sin t)^{n-1}\mathrm{d}t - \int_0^{\delta/2} (\sin t)^{n-1}\mathrm{d}t \\
&\quad > \int_{\pi/2-\sigma/2}^{\pi/2-\sigma/2+\delta/2} (\sin t)^{n-1}\mathrm{d}t - \int_{\pi/2-\sigma/2-\delta/2}^{\pi/2-\sigma/2} (\sin t)^{n-1}\mathrm{d}t > 0.
\end{aligned}$$

Clearly, the estimates (2) and (3) contradict the assumptions (1). □

Lemma 2. *If $\delta > 0$ and $\sigma = 2/3\int_0^{\delta/2}(\sin t)^{n-1}\mathrm{d}t$ satisfy $\sigma + \delta < \pi$, then*

$$V(\mathcal{B}_{\delta/2}) + V(\mathcal{B}_{\pi/2-\sigma/2}) > \tfrac{3}{2}V_{S^n}. \tag{4}$$

Proof. In fact, since $|\sin t| \leq 1$,

$$\begin{aligned}
V(\mathcal{B}_{\delta/2}) &= \omega_{n-1}\int_0^{\delta/2}(\sin t)^{n-1}\mathrm{d}t = \tfrac{3}{2}\omega_{n-1}\sigma \\
&> 3\omega_{n-1}\int_{\pi/2-\sigma/2}^{\pi/2}(\sin t)^{n-1}\mathrm{d}t \\
&= 3V(\mathcal{B}_{\pi/2}) - V(\mathcal{B}_{\pi/2-\sigma/2}) = \tfrac{3}{2}V_{S^n} - V(\mathcal{B}_{\pi/2-\sigma/2}).
\end{aligned}$$

□

Lemma 3. *There exists a point E on $\partial\mathbb{B}(p, \pi/2-\sigma/2)$, that is, the boundary of $\overline{\mathbb{B}}(P, \pi/2-\sigma/2)$, such that*

$$d(E, q) \leq \pi/2 - \sigma/2 + \delta \quad \textit{and} \quad d(E, x_0) \leq l_0 - \pi/2 + \sigma/2 + \delta. \tag{5}$$

Proof. Since $i_{\mathcal{M}} \geq \pi - \sigma$, the boundary $\partial\mathbb{B}(p, \pi/2-\sigma/2)$ is arc-connected in $\mathcal{M}$. Let $W = \gamma_{p,x_0} \cap \partial\mathbb{B}(p, \pi/2-\sigma/2)$ and $T = \gamma_{p,q} \cap \partial\mathbb{B}(p, \pi/2-\sigma/2)$. Take a continuous curve $f(t)$ $(t \in 0, 1])$ on $\partial\mathbb{B}(p, \pi/2-\sigma/2)$ such that $W = f(0)$ and $T = f(1)$. Let Γ be the image curve of f, and let

$$\begin{aligned}
\Gamma_1 &= \{x \in \Gamma \mid d(x, q) \leq \pi/2 - \sigma/2 + \delta\}, \\
\Gamma_2 &= \{x \in \Gamma \mid d(x, x_0) \leq l_0 - \pi/2 + \sigma/2 + \delta\}.
\end{aligned}$$

It is clear that Γ_1 and Γ_2 both are nonempty closed since $T \in \Gamma_1$ and $W \in \Gamma_2$. We will prove that there exists a point E on Γ satisfying (5). For this, we need only to verify that $\Gamma_1 \cap \Gamma_2 \neq \varnothing$. First we shall exclude the case that there exists a point E in Γ such that

$$(6) \qquad d(E, q) > \pi/2 - \sigma/2 + \delta \quad \text{and} \quad d(E, x_0) > l_0 - \pi/2 + \sigma/2 + \delta.$$

In fact, if (6) occurs, there must exist a point F in the shortest geodesic $\overline{\gamma}_p$ issuing from p and passing through E, such that $d(F, p) = \pi/2 - \sigma/2 + \delta/2$. By the triangle inequality, we have

$$(7) \qquad \begin{aligned} d(F, q) &\geq d(E, q) - d(E, F) \;\; > \pi/2 - \sigma/2 + \delta/2, \\ d(F, x_0) &\geq d(E, x_0) - d(E, F) > l_0 - \pi/2 + \sigma/2 + \delta/2. \end{aligned}$$

Therefore the four balls $\mathbb{B}(p, \pi/2 - \sigma/2)$, $\mathbb{B}(q, \pi/2 - \sigma/2)$, $\mathbb{B}(x_0, l_0 - \pi/2 + \sigma/2)$ and $\mathbb{B}(F, \delta/2)$ are pairwise disjoint. Applying again Günther's volume comparison theorem, we get

$$\begin{aligned} V_{\mathcal{M}} &> V(\mathbb{B}(p, \pi/2 - \sigma/2)) + V(\mathbb{B}(q, \pi/2 - \sigma/2)) \\ &\qquad\qquad + V(\mathbb{B}(x_0, l_0 - \pi/2 + \sigma/2)) + V(\mathbb{B}(F, \delta/2)) \\ &\geq 2V(\mathcal{B}_{\pi/2 - \sigma/2}) + V(\mathcal{B}_{\pi/2 - \sigma/2}) + V(\mathcal{B}_{\delta/2}) \\ &= 3V(\mathcal{B}_{\pi/2 - \sigma/2}) + V(\mathcal{B}_{\delta/2}), \end{aligned}$$

which contradicts the assumption on $V_{\mathcal{M}}$. Thus (6) cannot hold, which means $\Gamma = \Gamma_1 \cup \Gamma_2$. Since Γ is connected, we get a point $E \in \Gamma_1 \cap \Gamma_2 \neq \varnothing$; this point clearly satisfies (5). □

Lemma 1 and the triangle inequalities easily imply another result:

Corollary 1. *The point E obtained in Lemma 3 satisfies the inequalities*

$$(8) \qquad \begin{aligned} &\pi/2 - \delta/6 < d(E, p) = \pi/2 - \sigma/2, \\ &\pi/2 - \delta/6 \leq d(E, q) \leq \pi/2 - \sigma/2 + \delta, \\ &\pi/2 - \delta/6 \leq d(E, x_0) \leq l_0 - \pi/2 + \sigma/2 + \delta. \end{aligned}$$

On the other hand,

$$(9) \qquad d(p, q) \leq \pi - \sigma + \delta \quad \textit{and} \quad \pi - \sigma \leq l_0 = d(p, x_0) \leq \pi - \sigma + \delta.$$

Take $E \in \partial\mathbb{B}(p, \pi/2 - \sigma/2)$ satisfying (5). We consider a geodesic triangle $(\gamma_{E,p}, \gamma_{E,x_0}, \gamma_{p,x_0})$ in $\mathcal{M}$. Since $K_{\mathcal{M}} \geq -k^2$, Toponogov's comparison theorem gives

$$(10) \qquad \begin{aligned} &\cosh[kd(p, x_0)] \\ &\quad \leq \cosh[kd(E, p)] \cosh[kd(E, x_0)] - \sinh[kd(E, p)] \sinh[kd(E, x_0)] \cos\alpha \\ &\quad = \cosh[k(d(E, p) + d(E, x_0))] - \sinh[kd(E, p)] \sinh[kd(E, x_0)](1 + \cos\alpha), \end{aligned}$$

where the angle α is defined by $\alpha = \angle(\dot\gamma_{E,p}, \dot\gamma_{E,x_0})|_E$. By Corollary 1, we have

$$\begin{aligned} 1+\cos\alpha &\le \frac{\cosh(k(d(E,p)+d(E,x_0))) - \cosh(kd(p,x_0))}{\sinh(kd(E,p))\sinh(kd(E,x_0))} \\ &\le \frac{\cosh(k(l_0+\delta)) - \cosh(kl_0)}{\sinh^2(k(\pi/2-\delta/6))}. \end{aligned} \tag{11}$$

Clearly $t \mapsto \cosh(k(t+c)) - \cosh(kt)$ is increasing in $[0,\infty)$ for $c>0$, so we get

$$\begin{aligned} 1+\cos\alpha &< \frac{\cosh(k(\pi+2\delta)) - \cosh(k(\pi+\delta))}{\sinh^2(k(\pi/2-\delta/6))} \\ &< \frac{\cosh(k(\pi+3\delta)) - \cosh(k(\pi+\delta))}{\sinh^2(k(\pi/2-\delta/6))}. \end{aligned} \tag{12}$$

Similarly, if we consider the geodesic triangle $(\gamma_{E,p}, \gamma_{E,q}, \gamma_{p,q})$ and the angle $\beta = \angle(\dot\gamma_{E,p}, \dot\gamma_{E,q})|_E$, we have

$$\begin{aligned} 1+\cos\beta &\le \frac{\cosh(k(d(E,p)+d(E,q))) - \cosh(kd(p,q))}{\sinh(kd(E,p))\sinh(kd(E,q))} \\ &\le \frac{\cosh(k(\pi-\sigma+\delta)) - \cosh(kl_0)}{\sinh^2(k(\pi/2-\delta/6))} \\ &\le \frac{\cosh(k(\pi-\sigma+2\delta)) - \cosh(k(\pi-\sigma))}{\sinh^2(k(\pi/2-\delta/6))} \\ &< \frac{\cosh(k(\pi+3\delta)) - \cosh(k(\pi+\delta))}{\sinh^2(k(\pi/2-\delta/6))}. \end{aligned} \tag{13}$$

Likewise, if we think of the geodesic triangle $(\gamma_{E,q}, \gamma_{E,x_0}, \gamma_{q,x_0})$ and the angle $\gamma = \angle(\dot\gamma_{E,q}, \dot\gamma_{E,x_0})|_E$, then, noting that $d(q,x_0) \ge l_0 \ge \pi - \varepsilon_0$, we have

$$\begin{aligned} 1+\cos\gamma &\le \frac{\cosh(k(d(E,q)+d(E,x_0))) - \cosh(kd(q,x_0))}{\sinh(kd(E,q))\sinh(kd(E,x_0))} \\ &< \frac{\cosh(k(\pi+3\delta)) - \cosh(k(\pi+\delta))}{\sinh^2(k(\pi/2-\delta/6))}. \end{aligned} \tag{14}$$

Now we will conclude the proof of Proposition 1 using the following lemma, whose proof will be postponed.

Lemma 4. *For $k>0$, there exists a positive number $\delta_0 \in (0, 3\pi/5)$ such that δ_0 is a solution of*

$$\cosh(k(\pi+3t)) - \cosh(k(\pi+t)) - (1-\sqrt{3}/2)\sinh^2(k(\pi/2-t/6)) = 0. \tag{15}$$

Take $\delta = \delta_0$ in Lemma 4, take the σ from Lemma 2, and let E be the point given by Lemma 3. Obviously, $\sigma < \delta/3$, hence $\sigma + \delta < 4\delta/3 < \pi$. Applying (12)–(14), one immediately deduces

$$\cos\alpha < -\sqrt{3}/2, \quad \cos\beta < -\sqrt{3}/2, \quad \cos\gamma < -\sqrt{3}/2.$$

That is,

$$\alpha > 2\pi/3, \quad \beta > 2\pi/3, \quad \gamma > 2\pi/3.$$

However, since $0 \le \gamma \le 2\pi - (\alpha+\beta)$, we get a contradiction. Thus our hypothesis on $\mathcal{M}$ was wrong, so $\mathcal{M}$ must be homeomorphic to S^n. $\square$

In Theorem 1 or Proposition 1, we require that the sectional curvature $K_{\mathcal{M}}$ is in the interval $[-k^2, 1]$ for some $k > 0$. Trivially the result holds if $K_{\mathcal{M}} \in (0, 1]$. In the situation $0 \le K_{\mathcal{M}} \le 1$, we can simplify our proof by comparing against Euclidean space; however the estimates (12)–(14) would need to be changed for the case $k = 0$.

Theorem 2. *Suppose* $(\mathcal{M}, g)$ *is a compact connected n-dimensional Riemannian manifold with sectional curvature* $0 \le K_{\mathcal{M}} \le 1$. *Let* $\delta > 0$, *and let*

$$\sigma = \frac{2}{3}\int_0^{\delta/2} (\sin t)^{n-1}\mathrm{d}t \quad \textit{such that } (2-\sqrt{3})(\pi-\sigma)^2 - 16\delta(\pi-\sigma+2\delta) \ge 0. \tag{16}$$

Assume also that $i_{\mathcal{M}} \ge \pi - \sigma$ *and* $0 < V_{\mathcal{M}} \le 3V(\mathcal{B}_{\pi/2-\sigma/2}) + V(\mathcal{B}_{\delta/2})$. *Then* $\mathcal{M}$ *is homeomorphic to* S^n.

Proof. We prove this result by contradiction. If some manifold $\mathcal{M}$ satisfies the assumptions of Theorem 2 and is not homeomorphic to S^n, there is a point $x_0 \in \mathcal{M}$ such that $x_0 \in \mathcal{M} - \mathbb{B}(p, i_{\mathcal{M}}) \cup \mathbb{B}(q, i_{\mathcal{M}})$, with $d(p,q) = d_{\mathcal{M}}$. Assume that $d(q, x_0) \ge d(p, x_0) = l_0 \ge i_{\mathcal{M}}$. By Lemma 3, there exists a point $E \in \partial\mathbb{B}(p, \pi/2 - \sigma/2)$ satisfying (5). By triangle inequality, we get because $K_{\mathcal{M}} \ge 0$ that

$$d(E, q) \ge \pi/2 - \sigma/2 \quad \text{and} \quad d(E, x_0) \ge \pi/2 - \sigma/2. \tag{17}$$

Now consider the geodesic triangle $(\gamma_{p,E}, \gamma_{x_0,E}, \gamma_{p,x_0})$; let $\alpha = \angle(\dot\gamma_{E,p}, \dot\gamma_{E,x_0})|_E$. By Toponogov's comparison theorem,

$$d^2(p, x_0) \le d^2(E, p) + d^2(E, x_0) - 2d(E, p)d(E, x_0)\cos\alpha,$$

so

$$\begin{aligned} 1 + \cos\alpha &\le \frac{(d(E,p) + d(E,x_0))^2 - d^2(p,x_0)}{2d(E,p)d(E,x_0)} \\ &\le \frac{(l_0+\delta)^2 - l_0^2}{2(\pi/2-\sigma/2)^2} < \frac{2\delta(\pi-\sigma+2\delta)}{(\pi/2-\sigma/2)^2}. \end{aligned} \tag{18}$$

Similarly, consider the triangle $(\gamma_{E,p}, \gamma_{E,q}, \gamma_{p,q})$, with $\beta = \angle(\dot\gamma_{E,p}, \dot\gamma_{E,q})|_E$ and the triangle $(\gamma_{E,q}, \gamma_{E,x_0}, \gamma_{q,x_0}))$, with $\gamma = \angle(\dot\gamma_{E,q}, \dot\gamma_{E,x_0})|_E$. Then

$$\begin{aligned} 1+\cos\beta &\le \frac{(d(E,p)+d(E,q))^2-d^2(p,q)}{2d(E,p)d(E,q)} \\ &\le \frac{(\pi-\sigma+\delta)^2-(\pi-\sigma)^2}{2(\pi/2-\sigma/2)^2} < \frac{2\delta(\pi-\sigma+2\delta)}{(\pi/2-\sigma/2)^2}, \\ 1+\cos\gamma &\le \frac{(d(E,q)+d(E,x_0))^2-d^2(q,x_0)}{2d(E,q)d(E,x_0)} \\ &\le \frac{2\delta(l_0+\delta)}{(\pi/2-\sigma/2)^2} < \frac{2\delta(\pi-\sigma+2\delta)}{(\pi/2-\sigma/2)^2}. \end{aligned} \tag{19}$$

Let δ and σ satisfy (16). Then from (18) and (19), one can infer again that

$$\alpha > 2\pi/3, \quad \beta > 2\pi/3, \quad \gamma > 2\pi/3,$$

which is impossible as above. □

Proof of Lemma 4. First, we will show that the Equation (15) indeed contains a positive solution δ_0. Define

$$F(t,k) = \cosh(k(\pi+3t)) - \cosh(k(\pi+t)) - (1-\sqrt{3}/2)\sinh^2(k(\pi/2-t/6)).$$

For fixed $k > 0$ and for $t \in [0, 3\pi]$,

$$\frac{dF}{dt} = k\Big\{3\sinh(k(3t+\pi)) - \sinh(k(t+\pi)) + \frac{2-\sqrt{3}}{12}\sinh(k(\pi-t/3))\Big\} > 0,$$

which implies that $F(t,k)$ is increasing with respect to t in $[0, 3\pi]$. Moreover, $F(0,k) < 0$ and $F(3\pi,k) > 0$. So (15) has a unique solution $\delta_0 \in (0, 3\pi)$ for any $k > 0$. Consider the function $k \mapsto F(3\pi/5, k)$. Then

$$\frac{dF}{dk}\Big(\frac{3\pi}{5}, k\Big) = \frac{14\pi}{5}\sinh\Big(\frac{14k\pi}{5}\Big) - \frac{8\pi}{5}\sinh\Big(\frac{8k\pi}{5}\Big) - \frac{(2-\sqrt{3})\pi}{5}\sinh\Big(\frac{4k\pi}{5}\Big).$$

We can check that

$$\frac{14\pi}{5}\sinh\Big(\frac{14k\pi}{5}\Big) - \frac{8\pi}{5}\sinh\Big(\frac{8k\pi}{5}\Big) > \frac{4\pi}{5}e^{8\pi/5} > \frac{(2-\sqrt{3})\pi}{5}\sinh\Big(\frac{4k\pi}{5}\Big),$$

which implies that $F(3\pi/5, k)$ is increasing in $[0, \infty)$. Note that $F(3\pi/5, 0) = 0$; thus $F(3\pi/5, k) > 0$ for $k > 0$. This shows there is a solution in $0 < \delta_0 < 3\pi/5$. □

Acknowledgments

I thank Professor Dong Ye, whose comments simplified many proofs of this paper. Also, I thank the referee for suggestions.

References

[Brown 1960] M. Brown, "A proof of the generalized Schoenflies theorem", *Bull. Amer. Math. Soc.* **66** (1960), 74–76. MR 22 #8470b Zbl 0132.20002

[Coghlan and Itokawa 1991] L. Coghlan and Y. Itokawa, "A sphere theorem for reverse volume pinching on even-dimensional manifolds", *Proc. Amer. Math. Soc.* **111**:3 (1991), 815–819. MR 91f:53033 Zbl 0719.53019

[Grove and Shiohama 1977] K. Grove and K. Shiohama, "A generalized sphere theorem", *Ann. Math.* (2) **106**:2 (1977), 201–211. MR 58 #18268 Zbl 0341.53029

[Perelman 1995] G. Perelman, "A diameter sphere theorem for manifolds of positive Ricci curvature", *Math. Z.* **218**:4 (1995), 595–596. MR 96f:53056 Zbl 0831.53033

[Shen 1989] Z. M. Shen, "A sphere theorem for manifolds of positive Ricci curvature", *Indiana Univ. Math. J.* **38**:1 (1989), 229–233. MR 90h:53044 Zbl 0678.53032

[Shiohama 1983] K. Shiohama, "A sphere theorem for manifolds of positive Ricci curvature", *Trans. Amer. Math. Soc.* **275**:2 (1983), 811–819. MR 84c:53041 Zbl 0518.53047

[Suyama 1991] Y. Suyama, "Differentiable sphere theorem by curvature pinching", *J. Math. Soc. Japan* **43**:3 (1991), 527–553. MR 92g:53033 Zbl 0790.53041

[Wen 2004] Y. Wen, "A note on pinching sphere theorem", *C. R. Math. Acad. Sci. Paris* **338**:3 (2004), 229–234. MR 2005k:53040 Zbl 1046.53019

[Wu 1989] J. Y. Wu, "A diameter pinching sphere theorem for positive Ricci curvature", *Proc. Amer. Math. Soc.* **107**:3 (1989), 797–802. MR 90h:53045 Zbl 0676.53045

Received October 5, 2008. Revised April 9, 2009.

YULIANG WEN
DEPARTMENT OF MATHEMATICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI 20062
CHINA

ylwen@math.ecnu.edu.cn

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 web site 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 244 No. 1 January 2010

0030-8730(201001)244:1;1-J