Pacific Journal of Mathematics

Volume 258 No. 2 August 2012

PACIFIC JOURNAL OF MATHEMATICS

http://pacificmath.org

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

See inside back cover or pacificmath.org for submission instructions.

The subscription price for 2012 is US $420/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

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

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

UNIQUENESS THEOREMS FOR CR AND CONFORMAL MAPPINGS

YOUNG-JUN CHOI AND JAE-CHEON JOO

We provide a uniqueness theorem for CR and conformal mappings that generate compact sequences of iteration.

1. Introduction

The primary aim of this paper is to prove a version of uniqueness theorem for CR and conformal mappings.

Let M be a C^∞-smooth manifold and let $\mathscr{C}$ be a class of smooth mappings from M into itself containing the identity map. We say that the pair $(M, \mathscr{C})$ satisfies *the Cartan uniqueness property* (or simply *Cartan uniqueness*) at $p \in M$ if an element $f \in \mathscr{C}$ coincides with the identity map whenever $f(p) = p$, $df_p = \mathrm{Id}_{T_pM}$ and

$$\{f^k : k \in \mathbb{Z}_+\}$$

is compact, where Id_{T_pM} is the identity transform of tangent space T_pM of M at p, $\mathbb{Z}_+$ is the set of nonnegative integers, and f^k is the k-time composition of f, namely,

$$f^k = \underbrace{f \circ \cdots \circ f}_{k}.$$

In this definition, the compactness of a subclass $\mathscr{C}'$ of $\mathscr{C}$ means that every sequence in $\mathscr{C}'$ contains a subsequence that is uniformly Cauchy on every compact subset of M. (In other words, $\mathscr{C}'$ is relatively compact with respect to the compact-open topology.) One of the most important examples of this property in complex analysis is the uniqueness theorem of H. Cartan.

Theorem 1.1 (The Cartan uniqueness theorem). *Let Ω be a domain in $\mathbb{C}^n$ and let $\mathscr{H}$ be the class of holomorphic mappings from Ω into itself. Then $(\Omega, \mathscr{H})$ satisfies the Cartan uniqueness property at every point of Ω. In particular, if Ω is bounded, then $f \in \mathscr{H}$ coincides with the identity map if $f(p) = p$ and $df_p = \mathrm{Id}_{T_p\Omega}$ for some $p \in \Omega$, since the sequence $\{f^k : k \in \mathbb{Z}_+\}$ is automatically compact by the Montel theorem.*

MSC2010: 32V10, 32V40, 53A30.
Keywords: Cartan uniqueness, CR mappings, conformal mappings.

It is not trivial to determine under what conditions a similar uniqueness theorem holds if the reference point p lies on the domain boundary. For biholomorphic mappings we have the following theorems:

Theorem 1.2 [Krantz 1987]. *Suppose that Ω is a strongly pseudoconvex domain that is not biholomorphic to the ball. Let $f : \Omega \to \Omega$ be a biholomorphic mapping and let $p \in \partial\Omega$. If $f(z) = z + o(|z - p|)$, then $f \equiv \mathrm{Id}$.*

Theorem 1.3 [Huang 1993]. *Suppose that Ω is a bounded pseudoconvex domain satisfying Condition R. Let $f : \Omega \to \Omega$ be a biholomorphic map and let $p \in \partial\Omega$. If $f(z) = z + o(|z - p|)$ and $\{f^k : k \in \mathbb{Z}\}$ is compact, then $f \equiv \mathrm{Id}$.*

A crucial part in the proof of the Cartan uniqueness theorem are the Cauchy estimates for holomorphic mappings, which are consequences of the ellipticity of $\bar\partial$. In contrast, the operator $\bar\partial$ does not enjoy ellipticity or even subellipticity on the boundary of a domain. Huang [1993] exploited Bell's theorem [1987] on the boundary behavior of biholomorphic mappings in the C^∞ smooth sense, which is obtained from an analysis of the transformation formula of the Bergman kernel function.

In this paper, we prove a CR version of the uniqueness theorem:

Theorem 1.4 (CR case). *Let M be either a real hypersurface in $\mathbb{C}^{n+1}$ that does not contain any analytic hypersurface or a compact real hypersurface that bounds a domain. Let $\mathcal{H}_b$ be the class of all CR mappings from M into itself. Then $(M, \mathcal{H}_b)$ satisfies the Cartan uniqueness property at every strongly pseudoconvex point.*

The main interest of this theorem is that we assume neither a global type condition on M nor global injectivity of the mappings in the class $\mathcal{H}_b$. Therefore, we may regard this theorem as a generalization of Theorem 1.2 in the CR case. The proof of Theorem 1.4 is based on the method of derivative estimates of CR diffeomorphisms by the local solvability of the CR Yamabe equation, which was mainly developed in [Schoen 1995; Fischer-Colbrie and Schoen 1980].

By considering the conformal Yamabe equation instead of the CR Yamabe equation, we can have a conformal version of Theorem 1.4. For given two Riemannian manifolds (M, g) and (N, h), a diffeomorphism f from M to N is said to be *conformal* if $f^*h = u\, g$ for some positive function u on M. We define a little bit wider class of mappings as follows.

Definition 1.5. Let (M, g) and (N, h) be Riemannian manifolds. A smooth map $f : M \to N$ is said to be *semiconformal* if $f^*h = \lambda g$ for some smooth function λ. In this definition, we assume neither that f is 1-1 nor $\lambda > 0$.

Theorem 1.6 (Conformal case). *Let (M, g) be a Riemannian manifold of dimension $n > 2$, and let $\mathcal{S}$ be the class of all semiconformal mappings from M into itself. Then $(M, \mathcal{S})$ satisfies the Cartan uniqueness property at every point of M.*

We present the proof of Theorem 1.6 in Section 2 and then prove Theorem 1.4 in Section 3. Each section contains fundamental definitions of corresponding geometric objects — Yamabe equation, CR and pseudohermitian structures, CR mappings and so on.

2. Proof of Theorem 1.6

We start this section by recalling the Yamabe equation and problem. Let (M, g) be a Riemannian manifold of real dimension $n \geq 3$. Let $\tilde{g} = e^{2f} g$ be a conformal change of a Riemannian metric g, where f is smooth real-valued function on M. If we denote by S and $\widetilde{S}$ the scalar curvatures of g and $\tilde{g}$, respectively, it turns out that they satisfy the transformation law

$$\widetilde{S} = e^{-2f}\big(S - 2(n-1)\Delta_g f - (n-1)(n-2)|\nabla_g f|^2\big),$$

where $\Delta_g f$ denotes the Laplacian — the trace of the second covariant derivative — of f and $\nabla_g f$ its covariant derivative for the metric g. Let ϕ be the positive function satisfying $e^{2f} = \phi^{p_n-2}$, where $p_n = 2n/(n-2)$. Then the equation above turns into the following nonlinear equation for ϕ:

$$-a_n \Delta_g \phi + S\phi = \widetilde{S}\phi^{p_n-1}, \tag{2-1}$$

where $a_n = 4(n-1)/(n-2)$. This is called the *Yamabe equation* and the linear operator $L_g = -a_n \Delta_g + S$ is called the *conformal Laplacian* for g. When we mention the *Yamabe problem*, we mean the problem of finding a positive solution ϕ of (2-1) that makes $\widetilde{S}$ constant. This problem was first introduced in [Yamabe 1960], and its solvability has been intensively investigated there and elsewhere [Trudinger 1968; Aubin 1976a; 1976b; Schoen 1984; 1995].

For our purposes, a local scalar flattening argument is needed rather than the global solvability of the Yamabe problem:

Theorem 2.1 [Fischer-Colbrie and Schoen 1980]. *Let (M, g) be a Riemannian manifold and let Q be a smooth function on M. For $x \in M$ and $R > 0$, we denote by $B_R(x)$ the geodesic ball centered at x of radius R. If the minimum eigenvalue*

$$\lambda(B_R(x)) = \inf\left\{\int_{B_R(x)} (|\nabla f|^2 + Qf^2)\, dV : \mathrm{Support}(f) \subset B_R(x),\ \int_{B_R(x)} f^2 = 1\right\}$$

of $\Delta_g - Q$ on $B_R(x)$ is positive, then there exists a positive function ϕ on M such that

$$(\Delta_g - Q)\phi = 0$$

on $B_R(x)$.

Let us return back to the situation of Theorem 1.6. Let p be a fixed point of M and let $f : M \to M$ be a semiconformal map satisfying that $f(p) = p$, $df_p = \mathrm{Id}_{T_pM}$ and that $\{f^k : k \in \mathbb{Z}_+\}$ is compact. Then we can choose a neighborhood $U = B_R(p)$ of p such that

(i) f is one-to-one on U.

Choosing R small enough, we may also assume that the minimum eigenvalue of $-L_g$ is positive on U by the Poincaré inequality (see [Gilbarg and Trudinger 1983], for example), and that

$$\|u\|_q \le CR\,\|\nabla u\|_q$$

for every $u \in C_0^\infty(U)$ and $1 \le q < \infty$, where C is a constant depending only on the dimension n. Therefore, there exists a positive function ϕ on M such that $L_g\,\phi = 0$ on U by Theorem 2.1. Replacing the metric g by $\phi^{4/(n-2)}g$, then we may assume that

(ii) the metric g is scalar flat on U

by the Yamabe equation (2-1).

Thanks to the assumption that $\{f^k : k \in \mathbb{Z}_+\}$ is compact, there exists a neighborhood V of p that is relatively compact in U, such that $f^k(V) \subset U$ for every $k = 1, 2, \dots$. By (i), f^k is a conformal transformation from V to $f^k(V) \subset U$. Therefore, there exists a positive function u_k on V such that

$$(f^k)^*g = (u_k)^{4/(n-2)}g$$

for every $k = 1, 2, \dots$. We denote u_1 by u.

Since g and $(f^k)^*g$ are scalar flat on V, u_k satisfies the homogeneous Yamabe equation

$$\Delta_g u_k = 0$$

on V by (2-1).

Since f^k is a one-to-one map from V into U, it follows that

$$\int_V u_k^{2n/(n-2)}\,dV_g = \mathrm{Vol}_{(f^k)^*g}V = \mathrm{Vol}_g f^k(V) \le \mathrm{Vol}_g U < \infty.$$

By the elliptic mean value inequality, there exists $C > 0$ such that $u_k < C$ for every k on a neighborhood V' of p that is relatively compact in V.

Let V'' be a neighborhood of p that is relatively compact in V'. By the elliptic estimate for Δ_g [Gilbarg and Trudinger 1983], there is a C_j independent of k such that

$$\|u_k\|_{C^j(V'')} < C_j. \tag{2-2}$$

Note that

$$u_k(x) = u(f^{k-1}(x)) \cdot \dots \cdot u(f(x)) \cdot u(x) \tag{2-3}$$

for every positive integer k. Let $x = (x^1, \dots, x^n)$ be local coordinates on V centered at p and let

$$u(x) = 1 + h_j(x) + O(|x|^{j+1}),$$

where h_j is a j-th degree homogeneous polynomial. Since $f(x) = x + o(|x|)$ by the hypothesis, we see that

$$u_k(x) = 1 + kh_j(x) + O(|x|^{j+1})$$

from (2-3). Therefore, if h_j does not vanish, then the j-th order differential of u_k at $p = 0$ diverges as $k \to \infty$, which contradicts the inequality (2-2). This means that h_j vanishes identically on V'', hence $v = u - 1$ vanishes at p up to infinite order. By the unique continuation principle [Garofalo and Lin 1987; Kazdan 1988], we have $v = 0$ on V'', namely, $u = 1$ on V''. This implies that f is an isometry on V''. Therefore, f coincides with the identity map on V'', since every local geodesic passing through p should be preserved by f.

Let $F = \{x \in M : f(x) = x\}$ and let F° the interior of F. Since $V'' \subset F^\circ$, we see that F° is nonempty. If x_0 is a limit point of F°, then obviously f satisfies that $f(x_0) = x_0$ and $df_{x_0} = \mathrm{Id}_{T_{x_0}M}$. Repeating all the arguments above, we conclude that $x_0 \in F^\circ$. Thus F° is closed. This completes the proof of Theorem 1.6. □

Corollary 2.2. *Let f be a conformal transformation of M such that $f(p) = p$ and $df_p = \mathrm{Id}_{T_pM}$. Then either $f \equiv \mathrm{Id}$ or M is conformally equivalent to the unit sphere S^n and f can be transformed into the conformal transformation ϕ_a of S^n that fixes $p_0 = (1, 0, \dots, 0) \in \mathbb{R}^{n+1}$ for some $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, where $\phi_a = \Phi_a/|\Phi_a|$ and Φ_a is the affine transformation of $\mathbb{R}^{n+1}$ defined by*

$$y_0 = \frac{|a|^2}{4} + \Big(1 - \frac{|a|^2}{4}\Big)x^0 + \frac{1}{\sqrt{2}}\sum_{j=1}^n a_j x^j,$$
$$y^j = x^j + \frac{a_j}{\sqrt{2}}(1 - x^0)$$

for $j = 1, \dots, n$.

Proof. By Theorem 1.6, we may assume that $\{f^k : k \in \mathbb{Z}\}$ is noncompact if f is not the identity transform. Then Schoen's theorem [1995] implies that M is conformally equivalent to either $\mathbb{R}^n$ or S^n. Since $\mathbb{R}^n$ is conformally equivalent to $S^n \setminus \{p_\infty = (-1, 0, \dots, 0)\}$, we may also assume that f is a conformal transformation of S^n fixing $p_0 = (1, 0, \dots, 0)$. By Obata's theorem [1970], either f has two fixed points or it has only one fixed point. Moreover, if f has two fixed points, then each fixed point is a contracting fixed point of f or f^{-1}. This contradicts the hypothesis that $df_{p_0} = \mathrm{Id}_{T_{p_0}M}$. In particular, M is conformally equivalent to S^n, since if M were conformally equivalent to $\mathbb{R}^n$, then f should fix two points p_0 and p_∞ as a conformal transform of S^n. If f has only one fixed point p_0 and $df_{p_0} = \mathrm{Id}$, then $f = \Phi_a/|\Phi_a|$ for some $a \in \mathbb{R}^n$ by the algebraic characterization of the conformal transformation group of S^n via projectivization. See [Obata 1970] for more details. □

3. Proof of Theorem 1.4

We first review some definitions related to CR manifolds and CR mappings.

Let M be a smooth manifold of real dimension $2n+1$ and let H be a subbundle of TM with fiber dimension $2n$. Let J be a smooth section of the endomorphism bundle of H satisfying $J^2 = -\mathrm{Id}_H$. Then the triple (M, H, J) is called a *CR manifold* if it satisfies the integrability condition

$$[\Gamma(H_{1,0}), \Gamma(H_{1,0})] \subset \Gamma(H_{1,0}),$$

where $H_{1,0}$ is the subbundle of $\mathbb{C} \otimes H$ on which $J = i$, and $\Gamma(H_{1,0})$ is the space of smooth sections of $H_{1,0}$. A typical example of a CR manifold is a real hypersurface of a complex manifold. Let (M, H, J) and (M', H', J') be two CR manifolds. A smooth map $f : M \to M'$ is called a *CR mapping* if $df(H_{1,0}) \subset {H'}_{1,0}$. For a CR manifold (M, H, J), let θ be a nonvanishing real 1-form on M that vanishes on H. *The Levi form* L_θ is the symmetric bilinear form on H defined by

$$(L_\theta)_x(X, Y) = d\theta_x(X, JY)$$

for every $x \in M$ and $X, Y \in H_x$.

A CR manifold (M, H, J) is said to be *strongly pseudoconvex* if the Levi form L_θ for some θ is positive definite on H. In this case, the 1-form becomes a contact form. The quadruple (M, H, J, θ) is called a *pseudohermitian manifold;* see [Webster 1978]. We abbreviate this by (M, θ).

For a pseudohermitian manifold (M, θ), let ξ be the vector field on M defined by $\theta(\xi) = 1$ and $\xi \lrcorner d\theta = 0$. Then for every $x \in M$, $T_x M = [\xi_x] \oplus H_x$, where $[\xi_x]$ denotes the space generated by ξ_x. This decomposition defines a natural projection $\pi : T_x M \to H_x$. We define a Riemannian metric g on M by

$$g_x(X, Y) = \theta_x(X)\,\theta_x(Y) + (L_\theta)_x(\pi(X), \pi(Y)) \tag{3-1}$$

for every $X, Y \in T_x M$. The *Tanaka–Webster connection* ∇ on (M, θ) is an affine connection for which g, V and J are parallel. This connection is determined uniquely under suitable conditions on the torsion tensor. See [Tanaka 1975; Webster 1978]. By differentiating the Tanaka–Webster connection form, the *pseudohermitian* *curvature tensor* can be defined. The trace of this tensor is the Ricci tensor, and the trace of the Ricci tensor is the *Webster scalar curvature* R_θ. Obviously, the Webster scalar curvature depends the choice of the contact form θ.

Let $\tilde{\theta} = u^{2/n}\theta$ be another choice of contact form, where u is a positive smooth function, and let $\widetilde{R}$ and R be the Webster scalar curvatures for $\tilde{\theta}$ and θ, respectively. Then it is known that

$$-b_n \Delta_\theta u + Ru = \widetilde{R} u^{p-1}, \tag{3-2}$$

where

$$b_n = \frac{2(2n+1)}{n+1}, \quad p = 2 + \frac{2}{n},$$

and Δ_θ is the sublaplacian for θ. Equation (3-2) is called the *CR Yamabe equation.* The *CR Yamabe problem* is to find a positive solution of (3-2) that makes $\widetilde{R}$ constant. One may refer to [Jerison and Lee 1987; 1989; Lee 1986] for the properties of the CR Yamabe equation and the solvability of the CR Yamabe problem.

Now let us consider the situation of Theorem 1.4. Let M be a smooth real hypersurface in $\mathbb{C}^{n+1}$, $p \in M$ be a strongly pseudoconvex point and let $f : M \to M$ be a CR mapping satisfying that $f(p) = p$, $df_p = \mathrm{Id}_{T_p M}$ and that the iteration sequence $\{f^k : k \in \mathbb{Z}_+\}$ is compact. Let Γ be the connected component of the set of strongly pseudoconvex points in M that contains p. One should notice that Theorem 2.1 is still valid for subelliptic cases, since the proof depends only on the Fredholm alternative theorem. Therefore, by a similar argument as in the conformal case, we can choose a neighborhood U of p in Γ and a contact 1-form θ on Γ such that

- f is one-to-one, and
- the Webster scalar curvature for θ vanishes on U.

Take a relatively compact neighborhood V of p in U such that $f^k(V) \subset U$ for every $k = 1, 2, \ldots$. Then an iteration argument as in the conformal case and the subellipticity of the sublaplacian yield that $u - 1$ vanishes at p up to infinite order, where u is the positive function on V defined by

$$f^*\theta = u^{2/n}\theta. \tag{3-3}$$

Remark 3.1. Although $\Delta_\theta u = 0$ on V, we cannot conclude that $u \equiv 1$ at this stage, since the unique continuation principle for subelliptic operator has not been completely solved. In fact, an example in [Bahouri 1986] shows that a continuation theorem of Garofalo–Lin–Kazdan type cannot hold in the 3-dimensional Heisenberg group. Some partial results on the unique continuation principle for the sublaplacian were obtained in [Garofalo and Lanconelli 1990] for the Heisenberg group and in [Niu and Wang 2010] for more general nilpotent groups.

Let g be the Riemannian metric on Γ defined by (3-1). Then (3-3) yields that

$$\tilde{g} := f^*g = \lambda^2\theta\otimes\theta + \lambda L_\theta,$$

where $\lambda = u^{2/n}$. Since $u-1$ vanishes at p up to infinite order, so does $\lambda - 1$.

Let $(x^0, \dots, x^{2n})$ be local coordinates centered at p such that $g_{ij}(0) = \delta_{ij}$. Since $\lambda - 1$ vanishes at $p = 0$ up to infinite order, the Taylor coefficients of g_{ij} and $\tilde{g}_{ij}$ at $p = 0$ coincide. Since derivatives of f at 0 of order ≥ 2 are completely determined by differences between the Taylor coefficients of g_{ij} and $\tilde{g}_{ij}$, we see that f coincides with the identity at 0 up to infinite order. Note that f is a CR diffeomorphism from V onto $f(V)$. Since every local CR diffeomorphism on a strongly pseudoconvex CR manifold is uniquely determined by its finite order jet at the fixed point [Chern and Moser 1974; Kim and Zaitsev 2005], we conclude that $f \equiv \mathrm{Id}$ on V.

If M is a compact real hypersurface that bounds a domain D, then the CR mapping f extends continuously to a holomorphic map F on D by the Bochner–Hartogs extension theorem. Since $F - \mathrm{Id}_D$ vanishes on an open piece V of the boundary $\partial D = M$, we see that F and hence f coincide with the identity map. Now suppose that M is a real hypersurface in $\mathbb{C}^{n+1}$ containing no analytic hypersurface. Let

$$F = \{x \in M : f(x) = x\}$$

and let F° be the interior of F, which is nonempty by the argument above. Let x_0 be a limit point of F°. By Trépreau's theorem [1986], there exists a neighborhood Ω of x_0 in $\mathbb{C}^{n+1}$ such that $\Omega \setminus M = \Omega_+ \cup \Omega_-$ and such that f extends continuously to a holomorphic map F defined on Ω_+. Since $F \equiv \mathrm{Id}$ on a nonempty open piece $F^\circ \cap \Omega$ in $\partial\Omega_+$, we see that $F \equiv \mathrm{Id}$ on Ω_+. Therefore, $x_0 \in F^\circ$. This yields the conclusion. □

Corollary 3.2. *Let M be a real hypersurface in $\mathbb{C}^{n+1}$ containing no analytic hypersurface. Let p be a strongly pseudoconvex point of M. If $f : M \to M$ is a CR automorphism that $f(p) = p$ and $df_p = \mathrm{Id}$, then either $f = \mathrm{Id}$ or M is CR equivalent to the sphere S^{2n+1}, and f can be transformed into the CR transformation $\phi_{a,r}$ of S^{2n+1} fixing $p_0 = (1, 0, \dots, 0) \in \mathbb{C}^{n+1}$ for some $a = (a_1, \dots, a_n) \in \mathbb{C}^n$ and $r \in \mathbb{R}$, where $\phi_{a,r} = \Phi_{a,r}/\Psi_{a,r}$ and $\Phi_{a,r}$ is the affine transformation of $\mathbb{C}^{n+1}$ defined by*

$$\begin{aligned} w^0 &= \bigl(1 - \tfrac{1}{2}(|a|^2 + ir)\bigr)z^0 - i\sum_{j=1}^{n} \bar{a}^j z^j + \tfrac{1}{2}(|a|^2 + ir), \\ w^j &= z^j + ia^j(1 - z^0) \end{aligned}$$

for $j = 1, \dots, n$, and $\Psi_{a,r}$ is the $\mathbb{C}$-valued function defined by

$$\Psi_{a,r}(z) = \tfrac{1}{2}\bigl(|a|^2 + 2 + ir\bigr) - i\sum_{j=1}^{n} \bar{a}^j z^j - \tfrac{1}{2}\bigl(|a|^2 + ir\bigr)z^0.$$

Proof. By Theorem 1.4, we may assume that $\{f^k : k \in \mathbb{Z}\}$ is noncompact if f is not the identity map. Let Γ be the connected component of the set of strongly pseudoconvex points that contains p. Since f is a CR diffeomorphism of M onto M, it preserves Γ. Moreover, the group $\{f^k_{|\Gamma}\}$ is also noncompact, since otherwise, $f_{|\Gamma} \equiv \mathrm{Id}$ by Theorem 1.4 and this implies that $f \equiv \mathrm{Id}$ on M. Therefore, Schoen's theorem [1995] implies that Γ is CR equivalent to either the Heisenberg group $\mathbf{H}^n$ or the standard unit sphere in $\mathbb{C}^{n+1}$. Since S^{2n+1} is the one point compactification of $\mathbf{H}^n$, we also may assume that f is a CR transformation of S^{2n+1} fixing $p_0 = (1, 0, \dots, 0)$. By a result in [Webster 1977], either f has two fixed points or it has only one fixed point. If f has two fixed points, then each fixed point is a contracting fixed point of f or f^{-1}. This contradicts the hypothesis that $df_{p_0} = \mathrm{Id}$. Hence f has only one fixed point p_0. In particular, Γ cannot be equivalent to the Heisenberg group. Since S^{2n+1} is a boundary-free compact manifold, we can conclude that $M = \Gamma$ and that M is CR equivalent to the unit sphere S^{2n+1}.

To obtain explicit formulas for CR transformations of S^{2n+1} it is useful to imbed S^{2n+1} into the complex projective $(n+1)$-space $\mathbb{C}\mathrm{P}^{n+1}$ in the following manner: Let $\mathbb{C}^{n+1}$ be a complex Euclidean $(n+1)$-space with a coordinate system $(z^0, \dots, z^n)$, and let S^{2n+1} be given by the equation

$$|z^0|^2 + |z^1|^2 + \cdots + |z^n|^2 = 1\,.$$

We also let $\mathbb{C}^{n+2}$ be a complex Euclidean $(n+2)$-space with a coordinate system $(Z^0, Z^1, \dots, Z^{n+1})$, and let $\mathbb{C}\mathrm{P}^{n+1}$ be the projective $(n+1)$-space with a homogeneous coordinate system $[Z^0, Z^1, \dots, Z^{n+1}]$. We define a holomorphic embedding of $\mathbb{C}^{n+1}$ into $\mathbb{C}\mathrm{P}^{n+1}$ by the equations

$$Z^0 = 1 + z^0, \quad Z^j = z^j \ (j = 1, \dots, n), \quad Z^{n+1} = i(1 - z^0), \tag{3-4}$$

and the image of S^{2n+1} in $\mathbb{C}\mathrm{P}^{n+1}$ under this embedding is the real hypersurface Q that is defined by

$$|Z^1|^2 + \cdots + |Z^n|^2 + \frac{i}{2}\big(Z^{n+1}\overline{Z}^0 - Z^0\overline{Z}^{n+1}\big) = 0.$$

The special unitary group $\mathrm{SU}(n+1, 1)$ is the group of the linear transformations of $\mathbb{C}^{n+2}$ leaving the quadratic form

$$|Z^1|^2 + \cdots + |Z^n|^2 + \frac{i}{2}\big(Z^{n+1}\overline{Z}^0 - Z^0\overline{Z}^{n+1}\big)$$

invariant, and whose determinant has absolute value 1. We can regard the CR transformation group of S^{2n+1} as $\mathrm{SU}(n+1, 1)$. The Lie algebra of $\mathrm{SU}(n+1, 1)$

consists of $(n+2)\times(n+2)$ matrices of the form

$$\begin{pmatrix} \lambda & -2i\,{}^t\bar{a} & r \\ -\frac{1}{2i}\bar{b} & B & a \\ q & {}^t b & -\bar{\lambda} \end{pmatrix},$$

where B is a skew-hermitian $n\times n$ matrix, a and b are column n-vectors with complex entries, and r, q are real numbers.

In particular, the Lie algebra of the isotropy group $\mathrm{SU}_{p_0}(n+1, 1)$ at the point p_0 consists of the matrices of the form

$$\begin{pmatrix} \lambda & -2i\,{}^t\bar{a} & r \\ 0 & B & a \\ 0 & 0 & -\bar{\lambda} \end{pmatrix}.$$

If λ is not purely imaginary, then f has two fixed points, by [Webster 1977]. This contradicts our hypothesis. Write $\lambda = i\theta$ for some $\theta \in \mathbb{R}$. The isotropy group $\mathrm{SU}_{p_0}(n+1, 1)$ itself, as a subgroup of $\mathrm{SU}(n+1, 1)$, consists of the matrices of the form

$$\begin{pmatrix} e^{i\theta} & -2ie^{i\theta}({}^t\bar{a}) & -ie^{i\theta}|a|^2+re^{i\theta} \\ 0 & T & Ta \\ 0 & 0 & e^{i\theta} \end{pmatrix}, \tag{3-5}$$

where $T \in \mathrm{SU}(n)$ and r is a real number. Hence we can consider f as a linear transformation of the form (3-5). Since $df_{p_0}|_{T_{p_0}S^{2n+1}} = \mathrm{Id}$, T is the identity map. So f is represented by the matrix

$$\begin{pmatrix} e^{i\theta} & -2ie^{i\theta}({}^t\bar{a}) & -ie^{i\theta}|a|^2+re^{i\theta} \\ 0 & \mathrm{Id} & a \\ 0 & 0 & e^{i\theta} \end{pmatrix}. \tag{3-6}$$

By Webster [1977], the CR transformation represented by (3-6) has only one fixed point.

Since $[Z^0, \dots, Z^{n+1}]$ is the homogeneous coordinate of $\mathbb{C}\mathrm{P}^{n+1}$, the inverse map of (3-4) is given by

$$z^0 = \frac{iZ^0 - Z^{n+1}}{iZ^0 + Z^{n+1}} \quad \text{and} \quad z^j = \frac{2iZ^j}{iZ^0 + Z^{n+1}} \tag{3-7}$$

for $j = 1, \dots, n$. If $e^{i\theta}$ is not 1, then the map represented by the matrix (3-6) does not satisfy that $df_{p_0}|_{T_{p_0}S^{2n+1}} = \mathrm{Id}$. This implies that $e^{i\theta} = 1$. Using (3-6) and (3-7) we can conclude that $f = \Phi_{a,r}/\Psi_{a,r}$ for some $a \in \mathbb{C}^n$ and $r \in \mathbb{R}$. □

References

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

[Aubin 1976b] T. Aubin, "Problèmes isopérimétriques et espaces de Sobolev", *J. Differential Geom.* **11**:4 (1976), 573–598. MR 56 #6711 Zbl 0371.46011

[Bahouri 1986] H. Bahouri, "Non prolongement unique des solutions d'opérateurs "somme de carrés"", *Ann. Inst. Fourier* (*Grenoble*) **36**:4 (1986), 137–155. MR 88c:35027 Zbl 0603.35008

[Bell 1987] S. Bell, "Compactness of families of holomorphic mappings up to the boundary", pp. 29–42 in *Complex analysis* (University Park, PA, 1986), edited by S. G. Krantz, Lecture Notes in Math. **1268**, Springer, Berlin, 1987. MR 88k:32066 Zbl 0633.32020

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

[Fischer-Colbrie and Schoen 1980] D. Fischer-Colbrie and R. Schoen, "The structure of complete stable minimal surfaces in 3-manifolds of nonnegative scalar curvature", *Comm. Pure Appl. Math.* **33**:2 (1980), 199–211. MR 81i:53044 Zbl 0439.53060

[Garofalo and Lanconelli 1990] N. Garofalo and E. Lanconelli, "Frequency functions on the Heisenberg group, the uncertainty principle and unique continuation", *Ann. Inst. Fourier* (*Grenoble*) **40**:2 (1990), 313–356. MR 91i:22014 Zbl 0694.22003

[Garofalo and Lin 1987] N. Garofalo and F.-H. Lin, "Unique continuation for elliptic operators: a geometric-variational approach", *Comm. Pure Appl. Math.* **40**:3 (1987), 347–366. MR 88j:35046 Zbl 0674.35007

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **224**, Springer, Berlin, 1983. MR 86c:35035 Zbl 0562.35001

[Huang 1993] X. J. Huang, "Some applications of Bell's theorem to weakly pseudoconvex domains", *Pacific J. Math.* **158**:2 (1993), 305–315. MR 93m:32032 Zbl 0807.32016

[Jerison and Lee 1987] D. Jerison and J. M. Lee, "The Yamabe problem on CR manifolds", *J. Differential Geom.* **25**:2 (1987), 167–197. MR 88i:58162 Zbl 0661.32026

[Jerison and Lee 1989] D. Jerison and J. M. Lee, "Intrinsic CR normal coordinates and the CR Yamabe problem", *J. Differential Geom.* **29**:2 (1989), 303–343. MR 90h:58083 Zbl 0671.32016

[Kazdan 1988] J. L. Kazdan, "Unique continuation in geometry", *Comm. Pure Appl. Math.* **41**:5 (1988), 667–681. MR 89k:35039 Zbl 0632.35015

[Kim and Zaitsev 2005] S.-Y. Kim and D. Zaitsev, "Equivalence and embedding problems for CR-structures of any codimension", *Topology* **44**:3 (2005), 557–584. MR 2005j:32037 Zbl 1079.32022

[Krantz 1987] S. G. Krantz, "A new compactness principle in complex analysis", *Universidad Autonoma de Madrid Seminarios* **3** (1987), 171–194.

[Lee 1986] J. M. Lee, "The Fefferman metric and pseudo-Hermitian invariants", *Trans. Amer. Math. Soc.* **296**:1 (1986), 411–429. MR 87j:32063 Zbl 0595.32026

[Niu and Wang 2010] P. Niu and J. Wang, "On unique continuation properties for the sub-Laplacian on Carnot groups", *Acta Math. Sci. Ser. B Engl. Ed.* **30**:5 (2010), 1776–1784. MR 2012c:35076 Zbl 05951858

[Obata 1970] M. Obata, "Conformal transformations of Riemannian manifolds", *J. Differential Geom.* **4** (1970), 311–333. MR 42 #2387 Zbl 0205.52003

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

[Schoen 1995] R. Schoen, "On the conformal and CR automorphism groups", *Geom. Funct. Anal.* **5**:2 (1995), 464–481. MR 96h:53047 Zbl 0835.53015

[Tanaka 1975] N. Tanaka, *A differential geometric study on strongly pseudo-convex manifolds*, Lectures in Mathematics **9**, Kinokuniya, Tokyo, 1975. MR 53 #3361 Zbl 0331.53025

[Trépreau 1986] J.-M. Trépreau, "Sur le prolongement holomorphe des fonctions C-R défines sur une hypersurface réelle de classe C^2 dans $\mathbb{C}^n$", *Invent. Math.* **83**:3 (1986), 583–592. MR 87f:32035 Zbl 0586.32016

[Trudinger 1968] N. S. Trudinger, "Remarks concerning the conformal deformation of Riemannian structures on compact manifolds", *Ann. Scuola Norm. Sup. Pisa* (3) **22** (1968), 265–274. MR 39 #2093 Zbl 0159.23801

[Webster 1977] S. M. Webster, "On the transformation group of a real hypersurface", *Trans. Amer. Math. Soc.* **231**:1 (1977), 179–190. MR 58 #1231 Zbl 0368.57013

[Webster 1978] S. M. Webster, "Pseudo-Hermitian structures on a real hypersurface", *J. Differential Geom.* **13**:1 (1978), 25–41. MR 80e:32015 Zbl 0379.53016

[Yamabe 1960] H. Yamabe, "On a deformation of Riemannian structures on compact manifolds", *Osaka Math. J.* **12** (1960), 21–37. MR 23 #A2847 Zbl 0096.37201

Received November 8, 2011.

YOUNG-JUN CHOI
SCHOOL OF MATHEMATICS
KOREA INSTITUTE FOR ADVENCED STUDY
85 HOEGIRO
DONGDAEMON-GU
SEOUL 130-722
SOUTH KOREA
choiyj@kias.re.kr

JAE-CHEON JOO
DEPARTMENT OF MATHEMATICS
PUSAN NATIONAL UNIVERSITY
GEUMJEONG-GU
BUSAN 609-735
SOUTH KOREA
jcjoo91@pusan.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.269

SOME FINITE PROPERTIES FOR VERTEX OPERATOR SUPERALGEBRAS

CHONGYING DONG AND JIANZHI HAN

Vertex operator superalgebras are studied and various results on rational vertex operator superalgebras are obtained. In particular, the vertex operator super subalgebras generated by the weight $\frac{1}{2}$ and weight 1 subspaces are determined. It is also established that if the even part $V_{\bar{0}}$ of a vertex operator superalgebra V is rational, so is V.

1. Introduction

Vertex operator superalgebras, which are natural generalizations of vertex operator algebras, have been studied extensively in [Dong and Zhao 2005; 2006; Kac and Wang 1994; Li 1996b; 1996a; Xu 1998]. In this paper, we study certain finite properties of vertex operator superalgebras following [Dong et al. 1998b; Dong and Mason 2004; 2006; Mason 2011].

A vertex operator superalgebra $V = V_{\bar{0}} \oplus V_{\bar{1}}$ has even part $V_{\bar{0}}$ and odd part $V_{\bar{1}}$, where $V_{\bar{0}}$ consists of vectors of integral weights and $V_{\bar{1}}$ consists of vectors whose weights are half integers but not integers. So there is a canonical automorphism σ of V acting on $V_{\bar{i}}$ as $(-1)^i$ and $V_{\bar{0}}$ is a vertex operator algebra which is also a fixed point subalgebra of V. So a better understanding of the relationship between representation theories of V and $V_{\bar{0}}$ is definitely useful for the study of orbifold theory; see [Dijkgraaf et al. 1989; Dong et al. 2000]. Even the orbifold theory for vertex operator algebras with order 2 automorphism has not been understood fully.

Rationality, which is an analogue of semisimplicity of associative and Lie algebras, is probably the most important concept in the representation theory of vertex operator superalgebra. We first establish that if $V_{\bar{0}}$ is rational, V is rational, although we believe that the rationalities of V and $V_{\bar{0}}$ are equivalent from the orbifold theory. The main tool consists of the associative algebras $A_{g,n}(V)$ for $n \in \frac{1}{2}\mathbb{Z}_+$, which are generalizations of $A_{g,n}(V)$ as introduced and studied in [Dong et al. 1998b] (also see [Zhu 1996; Kac and Wang 1994; Dong et al. 1998a; Dong et al.

Chongying Dong was supported by NSF grants and a faculty research fund from the University of California at Santa Cruz.
MSC2010: 17B65, 17B69.
Keywords: vertex operator superalgebra.

1998c]), where g is an automorphism of V of finite order. It is established that V is g-rational if and only if $A_{g,n}$ is a finite dimensional semisimple associative algebra for large n. This is the key result to prove the rationality of V from the rationality of $V_{\bar{0}}$. Another characterization of rationality is given through the Ext functor.

Our investigation next centers around the vertex operator super subalgebras of V generated by homogeneous subspaces of small weights. The vertex operator subalgebra generated by $V_{\frac{1}{2}}$ is a holomorphic vertex operator superalgebra U associated to an infinite dimensional Clifford algebra built from a finite dimensional vector space with a nondegenerate symmetric bilinear form. This enables us to decompose V as a tensor product $U \otimes U^c$ where U^c, whose weight $\frac{1}{2}$ subspace is zero, is the commutant of U in V [Frenkel and Zhu 1992; Lepowsky and Li 2004]. Moreover, the module categories of V and U^c are equivalent. To study V_1, we first need to understand the algebraic structure of V_1. Under the assumption that V is rational or σ-rational together with C_2-cofiniteness, we are able to show that V_1 is a reductive Lie algebra, using the modular invariance results from [Dong and Zhao 2005; Zhu 1996], and the fact that $E_2(\tau)$ is not modular. Also, the rank of V_1 and the dimension of $V_{\frac{1}{2}}$ are controlled by the effective central charge. Furthermore, for any simple Lie subalgebra $\mathfrak{g}$ of V_1, the vertex operator subalgebra generated by $\mathfrak{g}$ is isomorphic to the vertex operator algebra $L(k, 0)$, which is the integrable highest weight module for the affine Kac–Moody algebra $\hat{\mathfrak{g}}$. We also give a rational vertex operator subalgebra, which is a tensor product of affine vertex operator algebras and a lattice vertex operator algebra, and whose weight one subspace is exactly V_1.

We should point out that most of the results in this paper have been obtained in the case where V is a vertex operator algebra in [Dong et al. 1998b; Dong and Mason 2004; 2006; Mason 2011]. So the results of this paper can be regarded as a “super” analogues of results presented in [Dong et al. 1998b; Dong and Mason 2004; 2006; Mason 2011]. The main ideas and the broad outlines also follow from these papers. A lot of arguments are omitted if they are the same as in the case of vertex operator algebras. On the other hand, there is a new phenomenon in the super case. Namely, either rationality together with C_2-cofiniteness or σ-rationality together with C_2-cofiniteness implies that V_1 is reductive. This gives strong evidence that rationality, σ-rationality of V, and rationality of $V_{\bar{0}}$ are equivalent. But we have no idea how to establish this.

This paper is organized as follows. In Section 2, we recall various notions of twisted modules for a vertex operator superalgebra and g-rationality for any automorphism of finite order from [Frenkel et al. 1988; Zhu 1996; Dong et al. 1998a; Dong and Zhao 2006]. In Section 3, we define a series of associative algebras $A_{g,n}(V)$ for a vertex operator superalgebra V and $n \in \mathbb{Z}_+$. We exhibit how to use $A_n(V)$ to prove rationality of V from the rationality of $V_{\bar{0}}$. It is also shown that if V is C_2-cofinite or rational, V is finitely generated and the automorphism group Aut(V) is an algebraic group. Section 4 is devoted to the study of vertex operator super subalgebras generated by $V_{\frac{1}{2}}$. In Section 5 we show that if V is rational or σ-rational together with C_2-cofiniteness then the weight one subspace V_1 is a reductive Lie algebra whose rank is bounded above by the effective central charge $\tilde{c}$. Consequently, dim $V_{\frac{1}{2}}$ is bounded above by $2\tilde{c}+1$. Section 6 deals with the vertex operator subalgebra of V generated by V_1.

We make the assumption that the reader is familiar with the theory of vertex operator algebras as presented in [Borcherds 1986; Dong and Lepowsky 1993; Frenkel et al. 1988; Lepowsky and Li 2004].

2. Basics

In this section we give the definition of a vertex operator superalgebra and several notions of modules; cf. [Dong et al. 1997; Dong and Zhao 2006; Feingold et al. 1991; Frenkel et al. 1988; Li 1996b; Zhu 1996].

We first recall that a super vector space is a $\mathbb{Z}_2$-graded vector space $V=V_{\bar{0}}\oplus V_{\bar{1}}$. The elements in $V_{\bar{0}}$ (respectively $V_{\bar{1}}$) are called even (respectively odd). Let $\tilde{v}$ be 0 if $v\in V_{\bar{0}}$, and 1 if $v\in V_{\bar{1}}$.

Definition 2.1. A vertex operator superalgebra (VOSA) is a $\frac{1}{2}\mathbb{Z}$-graded vector space

$$V=\bigoplus_{n\in\frac{1}{2}\mathbb{Z}} V_n = V_{\bar{0}}\oplus V_{\bar{1}}$$

with $V_{\bar{0}}=\sum_{n\in\mathbb{Z}} V_n$ and $V_{\bar{1}}=\sum_{n\in\mathbb{Z}} V_{n+\frac{1}{2}}$ satisfying all the axioms in the definition of vertex operator algebra except that the Jacobi identity is replaced by

$$z_0^{-1}\delta\left(\frac{z_1-z_2}{z_0}\right)Y(u,z_1)Y(v,z_2)-(-1)^{\tilde{u}\tilde{v}}z_0^{-1}\delta\left(\frac{-z_2+z_1}{z_0}\right)Y(v,z_2)Y(u,z_1)$$
$$=z_2^{-1}\delta\left(\frac{z_1-z_0}{z_2}\right)Y(Y(u,z_0)v,z_2).$$

Throughout the paper we always assume that V is a vertex operator superalgebra unless otherwise stated.

Definition 2.2. An automorphism g of a VOSA V is a linear automorphism of V preserving the vacuum vector $\mathbf{1}$ and the conformal vector ω such that the actions of g and $Y(v,z)$ on V are compatible in the sense that

$$gY(v,z)g^{-1}=Y(gv,z)$$

for $v\in V$.

Denote by Aut(V) the set consisting of all automorphisms of V. Observe that any automorphism of V commutes with $L(0)$ and hence preserves each homogeneous subspace V_n. As a consequence, any automorphism preserves both $V_{\bar{0}}$ and

$V_{\bar{1}}$. There is a canonical automorphism σ of V with $\sigma \mid V_{\bar{i}} = (-1)^i$ associated to the $\mathbb{Z}_2$-grading of V.

Let $g \in \mathrm{Aut}(V)$ with finite order T. Then we can decompose V into eigenspaces of g:

$$V = \bigoplus_{r=0}^{T-1} V^r,$$

where $V^r = \{v \in V \mid gv = e^{-2\pi ir/T} v\}$.

Definition 2.3. A weak g-twisted V-module M is a $\mathbb{Z}_2$-graded vector space equipped with a linear map

$$Y_M : V \to (\mathrm{End}\, M)[[z, z^{-1}]], \tag{2-1}$$

$$v \mapsto Y_M(v, z) = \sum_{n \in (1/T)\mathbb{Z}} v_n z^{-n-1}, \tag{2-2}$$

such that, for all $u \in V^r (0 \le r \le T-1)$, $v \in V$, and $w \in W$, the following conditions hold:

$$Y_M(u, z) = \sum_{n \in r/T+\mathbb{Z}} u_n z^{-n-1}, \quad u_n w = 0 \ \text{ for } n \gg 0,$$

$$Y_M(1, z) = \mathrm{Id}_M,$$

and

$$z_0^{-1}\delta\left(\frac{z_1 - z_2}{z_0}\right) Y_M(u, z_1) Y_M(v, z_2) - (-1)^{\tilde{u}\tilde{v}} z_0^{-1}\delta\left(\frac{-z_2 + z_1}{z_0}\right) Y_M(v, z_2) Y_M(u, z_1)$$
$$= z_2^{-1}\delta\left(\frac{z_1 - z_0}{z_2}\right)^{-r/T} \delta\left(\frac{z_1 - z_0}{z_2}\right) Y_M(Y(u, z_0)v, z_2).$$

Definition 2.4. An admissible g-twisted V-module is a weak g-twisted V-module M which carries a $(1/T)\mathbb{Z}_+$-grading

$$M = \bigoplus_{n \in (1/T)\mathbb{Z}_+} M(n)$$

satisfying

$$v_m M(n) \subseteq M(n + \mathrm{wt}\, v - m - 1)$$

for homogeneous $v \in V$ and $m \in \frac{1}{T}\mathbb{Z}$.

Definition 2.5. An ordinary g-twisted V-module is a weak g-twisted V-module

$$M = \bigoplus_{\lambda \in \mathbb{C}} M_\lambda$$

such that $\dim M_\lambda$ is finite and for fixed λ, $M_{n/T+\lambda} = 0$ for all small enough integers n, where $M_\lambda = \{w \in M \mid L(0)w = \lambda w\}$ and $Y_M(\omega, z) = \sum_{n \in \mathbb{Z}} L(n) z^{-n-2}$.

We say V is *g-rational* if every admissible g-twisted V-module is completely reducible, that is, a direct sum of simple admissible g-twisted V-modules. V is *g-regular* if the category of weak g-twisted V-modules is semisimple, namely, every weak g-twisted V-module is a direct sum of irreducible weak g-twisted V-modules. If $g = 1$, we have the definitions of rationality and regularity for vertex operator superalgebras.

The following definitions are given for vertex operator algebras in [Dong and Mason 2006; Zhu 1996] and we extend these to vertex operator superalgebras here.

A vertex operator superalgebra V is said to be of *CFT type* if the $L(0)$-grading on V has no negative weights and the degree-zero homogeneous subspace V_0 is one-dimensional: in symbols, $V = \bigoplus_{n\in\frac{1}{2}\mathbb{Z}_+} V_n$ and $V_0 = \mathbb{C}\mathbf{1}$. We say V is of *strong* CFT type if V satisfies the further condition $L(1)V_1 = 0$. V is said to be C_2*-cofinite* in the case where $C_2(V)$ has finite codimension in V, where $C_2(V)$ is the subspace of V linearly spanned by all elements of the form $u_{-2}v$ for $u, v \in V$.

For convenience, let us introduce the term *strongly g-rational* for a simple vertex operator superalgebra V which satisfies the following conditions:

(1) V is of strong CFT type.

(2) V is C_2-cofinite.

(3) V is g-rational.

Definition 2.6. A bilinear form $(\cdot\,,\cdot)$ on a V-module M is said to be invariant [Frenkel et al. 1993] if it satisfies the condition

$$(Y(a,z)u, v) = \big(u, Y(e^{zL(1)}(e^{\pi i}z^{-2})^{L(0)}a, z^{-1})v\big) \quad \text{for } a \in V \text{ and } u, v \in M.$$

It is proved in [Li 1994; Xu 1998] that there exists a linear isomorphism from the space of invariant bilinear forms on V to $\mathrm{Hom}_{\mathbb{C}}(V_0/L(1)V_1, \mathbb{C})$. This implies that there is a unique, up to multiplication by a nonzero scalar, nondegenerate symmetric invariant bilinear form on V if V is simple and of strong CFT type.

3. Rationality

In this section we give a characterization of the rationality of a vertex operator superalgebra V in terms of the rationality of a vertex operator subalgebra $V_{\bar{0}}$. We will show that if $V_{\bar{0}}$ is rational, V is rational. We certainly believe that the converse is also true, that is, if V is rational, $V_{\bar{0}}$ is also rational. This is similar to a well-known conjecture in orbifold theory: Let V be a rational vertex operator algebra, and g is an order 2 automorphism of V. Then the fixed point vertex operator subalgebra is also rational. We will establish some other results on rationality. We also discuss the generators of V.

The tool we use to prove the main result is the associative algebras $A_n(V)$, which is defined in [Dong et al. 1998c] for vertex operator algebra. Let V be a vertex operator superalgebra. Let $O_n(V)$ be the subspace of V linearly spanned by all $L(-1)u + L(0)u$ and $u \circ_n v$ where, for homogeneous $u \in V$ and $v \in V$,

$$u \circ_n v = \begin{cases} \mathrm{Res}_z \dfrac{(1+z)^{\mathrm{wt}\,u+n}}{z^{2n+2}} Y(u,z)v, & \text{if } u \in V_{\bar{0}}, \\ \mathrm{Res}_z \dfrac{(1+z)^{\mathrm{wt}\,u+n-\frac{1}{2}}}{z^{2n+1}} Y(u,z)v, & \text{if } u \in V_{\bar{1}}. \end{cases}$$

Define another operation $*_n$ on V by

$$u *_n v = \begin{cases} \displaystyle\sum_{m=0}^{n} (-1)^m \binom{m+n}{n} \mathrm{Res}_z \frac{(1+z)^{\mathrm{wt}\,u+n}}{z^{n+m+1}} Y(u,z)v, & \text{if } u, v \in V_{\bar{0}}, \\ 0, & \text{if } u \in V_{\bar{1}} \text{ or } v \in V_{\bar{1}}. \end{cases}$$

Set $A_n(V) = V/O_n(V)$. Then $A_0(V)$ is the $A(V)$ studied in [Kac and Wang 1994]. Let M be a weak V-module. Define the "n-th lowest weight vector" subspace of M as

$$\Omega_n(M) = \{w \in M \mid u_{\mathrm{wt}\,u+n+i} w = 0, u \in V, i \geq 0\}.$$

As in [Dong et al. 1998c] we have the following results.

Theorem 3.1. (1) *Suppose that M is a weak V-module. Then $\Omega_n(M)$ is an $A_n(V)$-module such that a acts as $o(a)$ for $a \in V_{\bar{0}}$, where $o(a)$ is defined to be $a_{\mathrm{wt}\,a-1}$ for homogeneous $a \in V_{\bar{0}}$ and extends it linearly.*

(2) *Suppose $M = \bigoplus_{i \in \frac{1}{2}\mathbb{Z}_+} M(i)$ is an admissible V-module. Then*

(a) $\Omega_n(M) \supset \bigoplus_{i \leq n} M(i)$;

(b) *assuming M is simple*, $\Omega_n(M) = \bigoplus_{i \leq n} M(i)$, *and each $M(i)$ is a simple $A_n(V)$-module for $i = 0, \frac{1}{2}, \ldots, n$.*

(3) *$M \mapsto M(0)$ gives a bijection between irreducible admissible V-modules and simple $A(V)$-modules.*

(4) *The identity map induces an epimorphism from $A_n(V)$ to $A_m(V)$ for any $n \geq m$.*

(5) *If V is g-rational, there are only finitely many irreducible admissible g-twisted V-modules up to isomorphism, and each irreducible admissible g-twisted V-module is ordinary.*

Note that part (3) of the theorem was obtained in [Kac and Wang 1994].

The next lemma will be used as a characterization of the rationality of V in terms of semisimplicity of $A_n(V)$ for large enough n.

Lemma 3.2. *Suppose that $A(V)$ is finite dimensional. Then any admissible V-module is a direct sum of generalized eigenspaces for $L(0)$.*

Proof. Let $M = \bigoplus_{i\in\frac{1}{2}\mathbb{Z}_+} M(i)$ be an admissible V-module with $M(0) \neq 0$. Let W be a maximal subspace of M which is a direct sum of generalized eigenspaces with respect to $L(0)$. Then it is not hard to see that W is a submodule of M. Consider the $A(V)$-module $M(0)$. By our assumption on finite dimension of $A(V)$, we see that there exists a nonzero simple $A(V)$-submodule of $M(0)$, on which $L(0)$ acts as a scalar by Schur's lemma. This shows that $W \neq 0$. We shall show $W = M$. Suppose $M/W \neq 0$. Choose the minimal $n \in \frac{1}{2}\mathbb{Z}_+$ such that $M(n)/W(n) \neq 0$, where $W(n) = W \cap M(n)$. Then, by similar argument as above, we see that $M(n)/W(n)$ contains a nonzero simple $A(V)$-submodule, say $\mathcal{W}(n)/W(n) \neq 0$, where $\mathcal{W}(n)$ is a subspace of $M(n)$. Since both $\mathcal{W}(n)/W(n)$ and $W(n)$ are a direct sum of generalized eigenspaces for $L(0)$, so is $\mathcal{W}(n)$. Thus $\mathcal{W}(n) \subset W$ and $\mathcal{W}(n) = W(n)$, a contradiction. □

Assume that $A(V)$ is finite dimensional. Let

$$f(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_r) \in \mathbb{C}[x]$$

be the monic polynomial of least degree such that $f([w]) = 0$ in $A(V)$. Then, on any given simple $A(V)$-module, $L(0)$ must act as a constant λ_i for some i. Note from Theorem 3.1 that V has exactly r inequivalent irreducible admissible modules $M^i = \sum_{n\in\frac{1}{2}\mathbb{Z}_+} M^i_{\lambda_i+n}$ for $i = 1, \cdots, r$. Then there exists $m_i > 0$ such that $M^i_{\lambda_i+n} \neq 0$ for all $n \geq m_i$. Let N be a positive integer greater than $|\lambda_i - \lambda_j|$, $|\lambda_i| + 1$, and m_i for $i, j = 1, \ldots, r$.

Note that the rationality is defined from the representation theory. It is always believed that such a property, which is analogous to the semisimplicity of Lie and associative algebras, should have its own internal characterization. The following result can be regraded as an internal characterization of rationality.

Theorem 3.3. *V is rational if and only if $A_n(V)$ is finite dimensional and semisimple for some $n \geq N$.*

Proof. The proof of [Dong et al. 1998c, Theorem 4.10] shows that if V is rational, $A_n(V)$ is semisimple and finite dimensional for all n. Now we assume that $A_n(V)$ is semisimple for some $n \geq N$. By Theorem 3.1, $A_m(V)$ is semisimple for all $m \leq n$. Let $M = \bigoplus_{i\in\frac{1}{2}\mathbb{Z}_+} M(i)$ be an admissible V-module with $M(0) \neq 0$. By Lemma 3.2, we can write

$$M = \sum_{\lambda\in\{\lambda_1,\ldots,\lambda_r\}} \bigoplus_{n\in\frac{1}{2}\mathbb{Z}_+} M_{\lambda+n},$$

where $M_{\lambda+n}$ is the generalized eigenspace for $L(0)$ with eigenvalue $\lambda + n$. Note that, for each $\lambda \in \{\lambda_1, \ldots, \lambda_r\}$, the subspace $\bigoplus_{n\in\frac{1}{2}\mathbb{Z}_+} M_{\lambda+n}$ is an admissible submodule of M. Without loss of generality, we may assume that $M = \bigoplus_{n\in\frac{1}{2}\mathbb{Z}_+} M(n)$ for some $\lambda \in \{\lambda_1, \ldots, \lambda_r\}$, where $M(n) = M_{\lambda+n}$.

We assert that the submodule W generated by $\bigoplus_{n\le N,\, n\in\frac{1}{2}\mathbb{Z}_+} M(n)$ is equal to the entire M. Otherwise, $0\ne M/W=\bigoplus_{n>N,\, n\in\frac{1}{2}\mathbb{Z}_+} M(n)/W(n)$, where $W(n)=W\cap M(n)$. Let $n_0\in\frac{1}{2}\mathbb{Z}_+$ be minimal such that $M(n_0)/W(n_0)\ne 0$. Then $n_0>N$ and $M(n_0)/W(n_0)$ is an $A(V)$-module by Theorem 3.1. Since $A(V)$ is semisimple, there exists a nonzero simple $A(V)$-submodule of $M(n_0)/W(n_0)$ on which $L(0)$ acts as the constant $\lambda+n_0\in\{\lambda_1,\dots,\lambda_r\}$, which implies $|\lambda-\lambda_j|=n_0$ for some j. But this is impossible by our choice on N. Thus we must have $W=M$.

We next show that if X is a simple $A(V)$-submodule, X generates an irreducible V-module U. Denote by $J=\bigoplus_{n\in\frac{1}{2}\mathbb{Z}_+} J(n)$ the maximal submodule of U such that $J(0)=0$, where $J(n)=J\cap U(n)$. Then the quotient $W=U/J$ is irreducible and $W(0)=X$. Since $\bigoplus_{0\le n\le N} U(n)$ is a semisimple $A_N(V)$-module we can regard each $W(n)$ as an $A_N(V)$-submodule of $U(n)$ for $n\le N$. From the choice of N, we know that $W(N)\ne 0$. Then the admissible V-submodule of U generated by $W(N)$ contains $W(0)=X$. Thus $W(N)=U(N)$, so $J(N)=0$. By our choice of N, again we see that J must be trivial. This implies that $U=W$ is irreducible.

It follows that the admissible V-submodule $\mathcal{W}$ of M generated by $M(0)$ is completely reducible. Note that $M(1)=\mathcal{W}(1)\oplus P$, where P is a semisimple $A(V)$-module. Again the admissible submodule of M generated by P is completely reducible. Continuing in this way completes the proof. □

Remark 3.4. Even in the case where V is a vertex operator algebra, Theorem 3.3 strengthens [Dong et al. 1998c, Theorem 4.11], where we require that $A_n(V)$ is semisimple for all n.

Remark 3.5. There is a twisted analogue $A_{g,n}(V)$ (cf. [Dong et al. 1998b]) of $A_n(V)$. One can similarly define the positive integer N_g. Then Theorem 3.3 still holds, that is, V is g-rational if and only if $A_{g,n}(V)$ is finite dimensional and semisimple for some $n\ge N_g$.

We now use Theorem 3.3 to prove the following result.

Proposition 3.6. *Let $V=V_{\bar{0}}\oplus V_{\bar{1}}$ be a VOSA. If $V_{\bar{0}}$ is rational, V is rational.*

Proof. Suppose $V_{\bar{0}}$ is rational. Then, by Theorem 3.3, $A_n(V_{\bar{0}})$ is a finite dimensional semisimple associative algebra if n is sufficiently large. This implies that $A_n(V)$ is semisimple as $A_n(V)$ is a quotient of $A_n(V_{\bar{0}})$. Applying Theorem 3.3 again yields that V is rational. □

We remark that we do not know how to prove the rationality of V from the rationality of $V_{\bar{0}}$ without using $A_n(V)$. It is certainly a very interesting problem to find a different approach that does not use $A_n(V)$. Although we can not show the converse of Proposition 3.6, we strongly believe that rationalities of V and $V_{\bar{0}}$ are equivalent.

In the rest of this section we use the extension functor to consider the rationality of a vertex operator superalgebra V. This approach has been studied in [Abe 2005] for vertex operator algebra, but our rationality result is different from that given in [ibid.].

First let us describe the set $\mathrm{Ext}^1_V(M^2, M^1)$ for any weak V-module M^1 and M^2. We call a weak V-module M an *extension of* M^2 *by* M^1 if there is a short exact sequence $0 \to M^1 \to M \to M^2 \to 0$. Two extensions M and N of M^2 by M^1 are said to be equivalent if there exists a V-homomorphism $f : M \to N$ such that the following diagram commutes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M^1 & \xrightarrow{\phi} & M & \xrightarrow{\varphi} & M^2 & \longrightarrow & 0 \quad \text{(exact)} \\ & & \| & & \downarrow f & & \| & & \\ 0 & \longrightarrow & M^1 & \xrightarrow{\phi'} & N & \xrightarrow{\varphi'} & M^2 & \longrightarrow & 0 \quad \text{(exact)}. \end{array}$$

Define $\mathrm{Ext}^1_V(M^2, M^1)$ to be the set of all equivalent classes of M^2 by M^1. It is well known that $\mathrm{Ext}^1_V(M^2, M^1)$ carries the structure of an abelian group such that the equivalent class of $M^1 \oplus M^2$ is the zero element.

Here is another equivalent condition of rationality.

Proposition 3.7. *Let V be a vertex operator superalgebra. Then V is rational if and only if the following two conditions hold.*

(a) *Every admissible V-module contains a nontrivial irreducible admissible submodule.*

(b) *For any irreducible V-modules M and N,* $\mathrm{Ext}^1_V(M, N) = 0$.

Proof. It is clear that rationality implies both (a) and (b). Now we assume that (a) and (b) hold. Let $M = \bigoplus_{n \in \frac{1}{2}\mathbb{Z}_+} M(n)$ be a nonzero admissible V-module. Let W be the sum of irreducible admission V-submodules of M. Then $W = \bigoplus_{i \in I} W^i$, where each W^i is an irreducible admissible V-module. By condition (a), $W \neq 0$. We assert that $W = M$. Otherwise consider the quotient module M/W. It follows from condition (a) again that there exists a weak V-submodule M' such that $M' \supsetneq W$ and M'/W is an irreducible admissible V-module. Then by condition (b) and the properties of Ext, we have

$$\mathrm{Ext}^1_V(M'/W, W) = \bigoplus_{i \in I} \mathrm{Ext}_V(M'/W, W^i) = 0,$$

that is, $M' = M'/W \oplus W$ as V-modules, contradicting the maximality of W. So the assertion is true and M is a direct sum of irreducible admissible V-modules. □

We now turn our attention to the generators of vertex operator superalgebras.

Proposition 3.8. *Let V be a vertex operator superalgebra.*

(a) *If V is rational or C_2-cofinite, V is finitely generated.*

(b) *If V is finitely generated,* $\mathrm{Aut}(V)$ *is an algebraic group.*

These results were obtained in the case of vertex operator algebras in [Dong and Zhang 2008; Karel and Li 1999]; see also [Gaberdiel and Neitzke 2003; Dong and Griess 2002]. The same proof works here.

4. Vertex operator subalgebra generated by $V_{\frac{1}{2}}$

In this section we study the vertex operator super subalgebra U of V generated by $V_{\frac{1}{2}}$, and decompose V as a tensor product $U \otimes U^c$, where U is holomorphic in the sense that U is the only irreducible module for itself and U^c, whose weight $\frac{1}{2}$ subspace is 0, is the commutant of U in V. This decomposition reduces the study of vertex operator superalgebras to the study of vertex operator superalgebras whose weight $\frac{1}{2}$ subspaces are 0.

Let V be a simple vertex operator superalgebra of strong CFT type. Then there is a unique invariant, symmetric, and nondegenerate bilinear form $(\cdot\,,\cdot)$ such that

$$(\mathbf{1}, \mathbf{1}) = \sqrt{-1}; \tag{4-1}$$

see [Li 1994; Xu 1998]. Then, for $u, v \in V_{\frac{1}{2}}$, one has

$$u_0 v = (u, v)\mathbf{1} \tag{4-2}$$

and

$$[u(m), v(n)]_+ = (u, v)\delta_{m+n+1,0}. \tag{4-3}$$

Note that the restriction of $(\cdot\,,\cdot)$ to $V_{\frac{1}{2}}$ is still nondegenerate. Let $\{a^1, a^2, \dots, a^l\}$ be an orthonormal basis of $V_{\frac{1}{2}}$ with respect to the form $(\cdot\,,\cdot)$, where $l = \dim V_{\frac{1}{2}}$.

Let U be the vertex super subalgebra of V generated by $V_{\frac{1}{2}}$. Then, using (4-3), we see that

$$U = \mathrm{Span}\left\{u^1_{-n_1} u^2_{-n_2} \cdots u^r_{-n_r}\mathbf{1} \;\middle|\; u^i \in V_{\frac{1}{2}},\ n_1 \geq n_2 \geq \cdots \geq n_r > 0 \text{ and } r \in \mathbb{Z}_+\right\}.$$

In fact, U carries the structure of a vertex operator superalgebra with conformal vector

$$\omega' = \frac{1}{2}\sum_{i=1}^{l} a^i_{-2} a^i_{-1}\mathbf{1}.$$

Define operators $L'(n)$ for $n \in \mathbb{Z}$ by

$$Y(\omega', z) = \sum_{n\in\mathbb{Z}} w'_n z^{-n-1} = \sum_{n\in\mathbb{Z}} L'(n) z^{-n-2}.$$

Then the weight n subspace U_n for $L'(0)$ is given by

$$U_n = \Big\langle u^1_{-n_1} u^2_{-n_2} \cdots u^r_{-n_r} \mathbf{1} \;\Big|\; u^i \in V_{\frac{1}{2}},\ n_1 \geq n_2 \geq \cdots \geq n_r > 0,\ r \in \mathbb{Z}_+, \text{ and } n_1 + n_2 \cdots + n_r = n + \frac{r}{2} \Big\rangle.$$

It is well known (cf. [Kac and Wang 1994]) that the vertex operator algebra U generated by $V_{\frac{1}{2}}$ is holomorphic. So for any admissible V-module M, we can decompose M into irreducible U-modules as follows

$$M = U \otimes \overline{M},$$

where $\overline{M} = \{w \in M \mid u_n w = 0 \text{ for all } u \in U \text{ and } n \in \mathbb{Z}_+\}$ is the multiplicity space of U in M. If $M = V$, the multiplicity space $\overline{M}$ is denoted by U^c and is called the commutant of U in V. In particular, $V = U \otimes U^c$. The U^c is a vertex operator superalgebra (see [Frenkel and Zhu 1992; Lepowsky and Li 2004]) with $\omega - \omega'$ as its conformal vector and $U^c_{\frac{1}{2}} = 0$.

Let $\mathrm{Irr}(V)$ and $\mathrm{Irr}(U^c)$ denote the sets of the isomorphism classes of admissible irreducible V-modules and U^c-modules, respectively. The following result is straightforward.

Proposition 4.1. *Let V be a vertex operator superalgebra.*

(a) *For any admissible V-module M. $\overline{M}$ is an admissible U^c-module. Moreover, M is irreducible if and only if $\overline{M}$ is irreducible.*

(b) *The map $U \otimes *:\ \mathrm{Irr}(U^c) \to \mathrm{Irr}(V)$ is a bijection.*

(c) *V is rational if and only if U^c is rational.*

5. The structure of weight 1 subspace

In this section we will investigate the Lie algebra structure of weight 1 subspace V_1 and show that V_1 is a reductive Lie algebra if V is σ-rational, using the modular invariance results obtained in [Dong and Zhao 2006]. We also find an upper bound for the rank of V_1 in terms of effective central charge. Similar results for vertex operator algebras were given previously in [Dong and Mason 2004], and the proof presented here is a modification of that used in [ibid.]. We also apply these results to estimate the dimension of weight $\frac{1}{2}$ subspace $V_{\frac{1}{2}}$ of V.

First, we need to discuss vertex operator superalgebras on the torus [Zhu 1996; Dong and Zhao 2005], vector-valued modular forms [Knopp and Mason 2003], and the modular invariance of trace functions [Zhu 1996; Dong and Zhao 2005].

Let V be a vertex operator superalgebra. The vertex operator superalgebra

$$(V, Y[v, z], \mathbf{1}, \tilde{\omega})$$

on a torus (see [Zhu 1996; Dong and Zhao 2005]) is defined as follows:

$$Y[v, z] = Y(v, e^z - 1)e^{wt v} = \sum_{n\in\mathbb{Z}} v[n]z^{-n-1},$$

$$Y[\tilde{w}, z] = \sum_{n\in\mathbb{Z}} L[n]z^{-n-2}$$

for homogeneous v and $\tilde{\omega} = \omega - c/24$.

We denote the eigenspace of $L[0]$ with eigenvalue $n \in \frac{1}{2}\mathbb{Z}$ by $V_{[n]}$. If $v \in V_{[n]}$, we write $\mathrm{wt}[v] = n$.

A holomorphic *vector-valued modular form of weight* k (where k is any real number) on the modular group $\Gamma = \mathrm{SL}(2, \mathbb{Z})$ may be described as follows: for any integer $p \geq 1$ it is a tuple $(T_1(\tau), \ldots, T_p(\tau))$ of functions holomorphic in the complex upper half-plane together with a p-dimensional complex representation $\rho : \Gamma \to \mathrm{GL}(p, \mathbb{C})$ satisfying the following conditions.

(a) For all $\gamma \in \Gamma$ we have

$$(T_1, \ldots, T_p)^t \mid_k \gamma(\tau) = \rho(\gamma)(T_1(\tau), \ldots, T_p(\tau))^t$$

(where t refers to the transpose of vectors and matrices).

(b) Each function $T_j(\tau)$ has a convergent q-expansion holomorphic at infinity:

$$T_j(\tau) = \sum_{n\geq 0} a_n(j)q^{n/N_j}$$

for positive integer N_j. (Here and below, $q = \exp 2\pi i\tau$).

The following result plays an important role in this section.

Proposition 5.1 [Knopp and Mason 2003]. *Let $(T_1, \ldots, T_p)$ be a holomorphic vector-valued modular form of weight k associated to a representation ρ of Γ. Then there is a nonnegative constant α depending only on ρ such that the Fourier coefficients $a_n(j)$ satisfy the polynomial growth condition $a_n(j) = O(n^{k+2\alpha})$ for every $1 \leq j \leq p$.*

Fix automorphisms g, h of V of finite orders. Let M be a simple $g\sigma$-twisted V-module. Then

$$M = \bigoplus_{n=0}^{\infty} M_{\lambda+n/T'}$$

for some λ called the conformal weight of M ($M_\lambda \neq 0$), where T' is the order of $g\sigma$. Suppose M is σh-stable, which is equivalent to the existence of a linearly isomorphic map $\phi(\sigma h) : M \to M$ such that

$$\phi(\sigma h)Y_M(v, z)\phi(\sigma h)^{-1} = Y_M((\sigma h)v, z)$$

for all $v \in V$. From now on we assume that V is C_2-cofinite. Then a function F_M which is linear in $v \in V$ is defined for homogeneous $v \in V$ as follows:

$$F_M(v,\tau)=q^{\lambda-c/24}\sum_{n=0}^{\infty}\mathrm{tr}_{M_{\lambda+n/T}}o(v)\phi(\sigma h)q^{n/T}=\mathrm{tr}_M o(v)\phi(\sigma h)q^{L(0)-c/24}, \tag{5-1}$$

which is a holomorphic function in the upper half-plane [Dong and Zhao 2005]. Here and below we write $F_M(\tau)$ rather than $F_M(\mathbf{1},\tau)$ for simplicity. Then for any u, v in V such that $gv = hv = v$, we have

$$\begin{aligned}&\mathrm{tr}_M o(u)o(v)\phi(\sigma h)q^{L(0)-c/24}\\ &\qquad = F_M(u[-1]v,\tau)-\sum_{k\geq 1}E_{2k}(q)F_M(u[2k-1]v,\tau);\end{aligned} \tag{5-2}$$

see [Dong and Zhao 2005; Zhu 1996]. The functions $E_{2k}(\tau)$ are the Eisenstein series of weight $2k$:

$$E_{2k}(q)=\frac{-B_{2k}}{2k!}+\frac{2}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)q^n,$$

where $\sigma_k(n)$ is the sum of the k-powers of the divisors of n, and B_{2k} is a Bernoulli number. The $E_2(\tau)$ enjoys an exceptional transformation law. Namely, its transformation with respect to the matrix

$$S=\begin{pmatrix}0&-1\\1&0\end{pmatrix},$$

has the form

$$E_2\Big(\frac{-1}{\tau}\Big)=\tau^2E_2(\tau)-\frac{\tau}{2\pi i}. \tag{5-3}$$

We also need results on 1-point functions on the torus [Dong and Zhao 2005]. Let g, h be automorphisms of V of finite order. The space of (g,h) 1-point functions $\mathscr{C}(g,h)$ is the $\mathbb{C}$-linear space consisting of functions

$$S:\ V\times\mathbb{H}\to\mathbb{C}$$

(where $\mathbb{H}$ is the upper half-plane) satisfying certain conditions; see [Dong and Zhao 2005] for details. The following results can be found in [ibid.].

Theorem 5.2. *Let V be C_2-cofinite and $g, h \in \mathrm{Aut}(V)$ of finite orders.*

(1) *For $S \in \mathscr{C}(g,h)$ and*

$$\gamma=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\Gamma,$$

we define

$$S|_\gamma(v,\tau)=S|_k(v,\tau)=(c\tau+d)^{-k}S(v,\tau)$$

for $v \in V_{[k]}$, and extend linearly. Then $S|_\gamma \in \mathscr{C}((g,h)\gamma))$, where $(g,h)\gamma = (g^a h^c, g^b h^d)$.

(2) *Let M be a simple $g\sigma$-twisted V-module such that M is h and σ-stable. Then $F_M(v,\tau) \in \mathscr{C}(g,h)$.*

(3) *Suppose that V is $g\sigma$-rational and $M^1, \dots, M^m$ are the inequivalent, simple $g\sigma$-twisted V-module such that M^i is h and σ-stable. Let $F_1, \dots, F_m$ be the corresponding trace functions defined by* (5-1). *Then $F_1, \dots, F_m$ form a basis of $\mathscr{C}(g,h)$.*

We now assume that V is of strong CFT type. Recall from [Frenkel et al. 1993] that the weight 1 subspace V_1 of V carries a natural Lie algebra structure, the Lie bracket being given by $[u,v] = u_0 v$ for $u, v \in V_1$. Then any weak V-module is automatically a V_1-module such that $v \in V_1$ acts as v_0. Note that there is a unique nondegenerate symmetric invariant bilinear form $\langle\,\cdot\,,\cdot\,\rangle$ such that $\langle \mathbf{1}, \mathbf{1}\rangle = -1$ and the restriction of $\langle\,\cdot\,,\cdot\,\rangle$ to V_1 endows V_1 with a nondegenerate, symmetric, invariant bilinear form such that $u_1 v = \langle u, v\rangle \mathbf{1}$ for $u, v \in V_1$.

The following two theorems are extensions of similar results from vertex operator algebras [Dong and Mason 2004] to vertex operator superalgebras.

Theorem 5.3. *Let V be strongly rational or strongly σ-rational. Then the Lie algebra V_1 is reductive.*

Proof. We first deal with the case where V is σ-rational. We have to show that the nilpotent radical N of the Lie algebra V_1 is zero. Suppose not, and take any nonzero element $u \in N$. Each V_i for $i \in \frac{1}{2}\mathbb{Z}$ is a finite dimensional V_1-module and has a composition series $0 = W^0 \subset W^1 \subset W^2 \subset W^3 \subset \cdots \subset$ such that u_0 acts trivially on each composite factor $W^i/W^{i-1} (i = 1, 2, \cdots)$. Note that we can take $\phi(\sigma) = \sigma$ on V. Thus V is σ-stable. In fact, any irreducible V-module is σ-stable; see [Dong and Zhao 2005, Lemma 6.1]. As a result, $tr_{V_i} o(u)o(v)\sigma = 0$ for all $v \in V_1$ and $i \in \frac{1}{2}\mathbb{Z}$. It follows from (5-2) that

$$F_V(u[-1]v,\tau) = \sum_{k\geq 1} E_{2k}(\tau) F_V(u[2k-1]v,\tau), \tag{5-4}$$

where $(g,h) = (\sigma, 1)$ and $F_V \in \mathscr{C}(\sigma, 1)$, by Theorem 5.2.

Note that if $k > 1$ is an integer, the element $u[2k-1]v$ has $L[0]$-weight $2-2k < 0$ and hence is 0. The nondegeneracy of the bilinear form $\langle\,\cdot\,,\cdot\,\rangle$ guarantees that there exists $v \in V_1$ such that $\langle u, v\rangle = 1$. With this choice of v, (5-4) simplifies to read

$$F_V(u[-1]v,\tau) = E_2(\tau) F_V(\tau). \tag{5-5}$$

By Theorem 3.1, V has finitely many irreducible σ-twisted V-modules up to isomorphism. We denote these modules by $M^1, \dots, M^m$. Note from Theorem 5.2

that the

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \Gamma$$

maps $\mathscr{C}(\sigma, 1)$ to $\mathscr{C}(1, \sigma)$. By Theorem 5.2 again we see that

$$F_V\Big(u[-1]v, -\frac{1}{\tau}\Big) = \tau^2 \sum_{i=1}^{m} s_i F_{M^i}(u[-1]v, \tau)$$

and

$$F_V\Big(-\frac{1}{\tau}\Big) = \sum_{i=1}^{m} s_i F_{M^i}(\tau)$$

for some $s_i \in \mathbb{C}$. Similar to equality (5-5), we also have

$$F_{M^i}(u[-1]v, \tau) = E_2(\tau) F_{M^i}(\tau) \quad \text{for } i = 1, \dots, m.$$

Thus

$$\begin{aligned}
\tau^2 \sum_{i=1}^{m} s_i F_{M^i}(u[-1]v, \tau) &= F_V\Big(u[-1]v, \frac{-1}{\tau}\Big) \\
&= E_2\Big(\frac{-1}{\tau}\Big) F_V\Big(\frac{-1}{\tau}\Big) \\
&= \Big(\tau^2 E_2(\tau) - \frac{\tau}{2\pi i}\Big) \sum_{i=1}^{m} s_i F_{M^i}(\tau) \\
&= \tau^2 \sum_{i=1}^{m} s_i F_{M^i}(u[-1]v, \tau) - \frac{\tau}{2\pi i} \sum_{i=1}^{m} s_i F_{M^i}(\tau).
\end{aligned}$$

Of course, the equality (5-3) is involved in the calculations above. Canceling the term $\tau^2 \sum_{i=1}^{m} s_i F_{M^i}(u[-1]v, \tau)$ gives rise to the identity $\sum_{i=1}^{m} s_i F_{M^i}(\tau) = 0$, which in turn implies $F_V(-1/\tau) = 0$. But this is clearly not true, since

$$F_V\Big(\frac{-1}{\tau}\Big) = q^{-c/24}\Big(\sum_{n \in \mathbb{Z}} (\dim V_n) q^n - \sum_{n \in \frac{1}{2} + \mathbb{Z}} (\dim V_n) q^n\Big) \neq 0.$$

So $N = 0$ and V_1 is reductive.

Now we assume that V is rational. As before we need to show that the nilpotent radical N of the Lie algebra V_1 is zero. This time we use $\mathscr{C}(\sigma, \sigma)$ instead of $\mathscr{C}(\sigma, 1)$ and $\mathscr{C}(1, \sigma)$. In this case, the $S \in \Gamma$ maps $\mathscr{C}(\sigma, \sigma)$ to itself. A similar argument applies. □

Remark 5.4. It is proved in [Dong and Mason 2004] that if a vertex operator algebra is strongly rational, weight one subspace is reductive. If one can prove the rationality of $V_{\bar{0}}$ from the rationality and σ-rationality of V, Theorem 5.3 follows immediately. Unfortunately, none of these results have been established.

The following result will be used in the next section.

Lemma 5.5. *Let V be a vertex operator superalgebra.*

(a) *If V is strongly rational, any admissible V-module is a completely reducible V_1-module. This is also equivalent to saying the action of any Cartan subalgebra of the Lie algebra V_1 is semisimple on any admissible V-module.*

(b) *If V is strongly σ-rational, any admissible σ-twisted V-module is a completely reducible V_1-module.*

(c) *If V is either strongly rational or strongly σ-rational, any irreducible σ^i-twisted V-module is a completely reducible V_1-module for $i = 0, 1$.*

Proof. Since the proof of (b) is similar to that of (a), we only show (a) and (c) for strongly rational vertex operator superalgebra V. Let H be a Cartan subalgebra of V_1. It is enough to show that H acts semisimply on any irreducible σ^i-twisted V-module for $i = 0, 1$. Since the homogeneous subspaces of an irreducible σ^i-twisted V-module are always finite dimensional, there is a common eigenvector of H on the irreducible module. So it is enough to show that H acts on V semisimply.

First we show that for any nonzero $u \in H$, h_0 is not nilpotent. Note that the restriction of the bilinear form $\langle \cdot, \cdot \rangle$ to H is nondegenerate. If u_0 is nilpotent for some nonzero $u \in H$, we can take $v \in H$ such that $\langle u, v \rangle = 1$. The proof of Theorem 5.3 then gives a contradiction.

We now prove that u_0 is semisimple on V. Since $\mathrm{Aut}(V)$ is an algebraic group by Proposition 3.8, and $\{e^{tu_0} \mid t \in \mathbb{C}\}$ is a one-dimensional algebraic subgroup of $\mathrm{Aut}(V)$, we immediately see that $\{e^{tu_0} \mid t \in \mathbb{C}\}$ is isomorphic to the one-dimensional multiplicative algebraic group $\mathbb{C}_m$ as u_0 is not nilpotent; cf. [Mason 2011]. □

Now that V_1 is reductive, there are two extreme cases: V_1 is a semisimple Lie algebra, and V_1 is abelian. The vertex operator subalgebra generated by V_1 will be extensively investigated in Section 6. We study the rank of V_1 in the rest of this section. Let l be the rank of V_1, that is, l is the dimension of a Cartan subalgebra H of V_1. Similar to the case of vertex operator algebras in [Dong and Mason 2004], l is closely related to the *effective central charge* $\tilde{c}$, which is defined as follows: Let $\{M^1, \ldots, M^m\}$ be the irreducible σ-twisted V-modules up to isomorphism. Then there exist $\lambda_i \in \mathbb{C}$ such that $M^i = \sum_{n \in \frac{1}{2}\mathbb{Z}_+} M^i_{\lambda_i + n}$ with $M^i_{\lambda_i} \neq 0$. The λ_i is called the conformal weight of M^i. By [Dong and Zhao 2005, Theorem 8.9], λ_i and the central charge c of V are rational numbers for all i. Define $\lambda_{\min}$ to be the minimum of the conformal weights λ_i, and set

$$\tilde{c} = c - 24\lambda_{\min}, \quad \tilde{\lambda}_i = \lambda_i - \lambda_{\min}.$$

Theorem 5.6. *Let V be strongly σ-rational. Then $l \leq \tilde{c}$.*

Proof. Let H be a Cartan subalgebra of V_1. Note that the component operators of the vertex operators $Y(u, z)$ on V for $u \in H$ form a Heisenberg Lie algebra. This amounts to saying that for $u, v \in H$ the following relations hold:

$$[u_m, v_n] = m\delta_{m,-n}\langle u, v\rangle. \tag{5-6}$$

In fact, these relations also hold true on any σ-twisted V-module M.

Consider $(g, h) = (1, 1)$. Let $F_i = F_{M^i}$ be as defined in (5-1). Then $F_i \in \mathscr{C}(1, 1)$. Recall that

$$\eta(\tau) = q^{1/24}\prod_{n=1}^{\infty}(1 - q^n)$$

is a modular form of weight $\frac{1}{2}$. Then

$$\eta(\tau)^{\tilde{c}} F_i(\tau) = q^{\tilde{\lambda}_i}\prod_{n=1}^{\infty}(1 - q^n)^{\tilde{c}}\sum_{n=0}^{\infty}\mathrm{tr}_{M_{\lambda_i+n/2}}\phi(\sigma)q^{n/2}$$

is holomorphic in $\mathbb{H} \cup \{i\infty\}$. Now it follows from the transformation law for $\eta(\tau)$ and Theorem 5.2 that the m-tuple

$$\big(\eta(\tau)^{\tilde{c}} F_1(\tau), \ldots, \eta(\tau)^{\tilde{c}} F_m(\tau)\big)$$

is a holomorphic vector-valued modular form of weight $\tilde{c}/2$. So the Fourier coefficients of $\eta(\tau)^{\tilde{c}} F_i(\tau)$ have polynomial growth by Proposition 5.1.

The Stone–von Neumann theorem provides us a somewhat different way to look more closely at $F_i(\tau)$. Namely, M^i has the following tensor decomposition:

$$M^i = M(1) \otimes_{\mathbb{C}} \Omega_{M^i}, \tag{5-7}$$

where $M(1) = \mathbb{C}[u_m \mid u \in H,\ m > 0]$ is the Heisenberg vertex operator algebra of rank l generated by H and $\Omega_{M^i} = \{w \in M^i \mid u_n w = 0 \text{ for } u \in H \text{ and } n > 0\}$. Then the trace function $F_i(\tau)$ corresponding to the decomposition (5-7) is equal to

$$q^{(l-c)/24}\eta(\tau)^{-l}\mathrm{tr}_{\Omega_i}\phi(\sigma)q^{L(0)},$$

as $\mathrm{tr}_{M(1)}\phi(\sigma)q^{L(0)} = q^{l/24}\eta(\tau)^{-l}$. Thus

$$\eta(\tau)^{\tilde{c}} F_i(\tau) = q^{(l-c)/24}\eta(\tau)^{\tilde{c}-l}\mathrm{tr}_{\Omega_i}\phi(\sigma)q^{L(0)}. \tag{5-8}$$

We know the Fourier coefficients of the left side of (5-8) have polynomial growth. This forces the same to be true on $\eta(\tau)^{\tilde{c}-l}$. Then one has $\tilde{c} - l \geq 0$, as $\eta(\tau)^s$ has exponential growth of Fourier coefficients whenever $s < 0$; cf. [Knopp 1970]. □

We now use Theorem 5.6 to do an estimation on the dimension of $V_{\frac{1}{2}}$.

Corollary 5.7. *Let V be strongly σ-rational. Then* $\dim V_{\frac{1}{2}} \leq 2\tilde{c} + 1$.

Proof. Let d be a nonnegative integer such that $2d \le \dim V_{\frac{1}{2}} \le 2d+1$. Then there exists a unique (up to a constant) nondegenerate bilinear form satisfying (4-1). We point out that the restriction of $(\cdot\,,\cdot)$ to $V_{\frac{1}{2}}$ is still nondegenerate. So we can choose elements $b^i, b^{i*} \in V_{\frac{1}{2}}$ such that $(b^i, b^{j*}) = \delta_{ij}$ and $(b^i, b^j) = 0 = (b^{i*}, b^{j*})$ for all $1 \le i, j \le d$. Set $h^i = b^i_{-1}(b^i)^*_{-1}\mathbf{1}$ for $i = 1, \dots, d$. Then $h^i \in V_1$ and $h^i_1 h^j = \delta_{i,j}$. $h^i_0 h^j = 0$ for $i, j \in \{1, \cdots, d\}$. As a result, $\sum_{i=1}^d \mathbb{C}h^i \subset V_1$ is contained in a Cartan subalgebra of V_1. By Corollary 5.7, $d \le l \le \tilde{c}$, and the proof is complete. □

6. C_2-cofiniteness and integrability

We continue our discussion on the weight 1 subspace V_1. We will determine the vertex operator subalgebra $\langle V_1\rangle$ of V generated by V_1 following the approach in [Dong and Mason 2006]. It turns out that $\langle V_1\rangle$ is isomorphic to

$$L_{\mathfrak{g}_1}(k_1, 0) \otimes \cdots \otimes L_{\mathfrak{g}_s}(k_s, 0) \otimes M(1),$$

where $V_1 = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s \oplus Z(V_1)$, $\mathfrak{g}_i$ are simple, $k_i \ge 1$ are integers, and $M(1)$ is the Heisenberg vertex operator algebra built up from $Z(V_1)$ (see below for the definition of $L_{\mathfrak{g}}(k, 0)$). Moreover, $\langle V_1\rangle$ is contained in the rational vertex operator subalgebra $L_{\mathfrak{g}_1}(k_1, 0) \otimes \cdots \otimes L_{\mathfrak{g}_s}(k_s, 0) \otimes V_L$ for some positive definite lattice $L \subset Z(V_1)$ satisfying $\operatorname{rank}(L) = \dim Z(V_1)$.

Here we need to review the construction of untwisted affine Kac–Moody Lie algebras $\hat{\mathfrak{g}}$ associated with simple Lie algebras $\mathfrak{g}$ and relevant results from [Kac 1990]. Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$ and Φ the corresponding root system. Fix a nondegenerate symmetric invariant bilinear form $(\cdot\,,\cdot)$ on $\hat{\mathfrak{g}}$ such that the square length of a long root is 2, where we have identified $\mathfrak{h}$ with its dual via the bilinear form. Then the affine Kac–Moody algebra associated to $\mathfrak{g}$ is given by

$$\hat{\mathfrak{g}} = \mathfrak{g} \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}K$$

with the bracket relations

$$[u(m), v(n)] = [u, v](m+n) + m(u, v)\delta_{m+n,0}K \quad \text{and} \quad [K, \hat{\mathfrak{g}}] = 0 \tag{6-1}$$

for $u, v \in \mathfrak{g}$ and $m, n \in \mathbb{Z}$, where $u(m) = u \otimes t^m$. Let $L(\lambda)$ be the irreducible $\mathfrak{g}$-module with highest weight $\lambda \in \mathfrak{h}$. Consider $L(\lambda)$ as a $\mathfrak{g} \otimes \mathbb{C}[t]$-module with $\mathfrak{g} \otimes \mathbb{C}t[t]$ acting trivially and with K acting as the scalar $k \in \mathbb{C}$. Then the generalized Verma module

$$V(k, \lambda) = \operatorname{Ind}_{\mathfrak{g}}^{\hat{\mathfrak{g}}} L(\lambda) = U(\hat{\mathfrak{g}}) \otimes_{U(\mathfrak{g}\otimes\mathbb{C}[t]\oplus\mathbb{C}K)} L(\lambda)$$

has the unique irreducible quotient $L(k, \lambda)$. It is well known that $L(k, \lambda)$ is integrable if and only if k is a nonnegative integer and λ is a dominant integral weight such that $(\lambda, \theta) \le k$, where $\theta \in \Phi$ is the maximal root.

Let V be a VOSA of strong CFT type and let $\langle\cdot,\cdot\rangle$ be the unique nondegenerate bilinear form satisfying $\langle\mathbf{1},\mathbf{1}\rangle=-1$. Suppose that $\mathfrak{g}\subset V_1$ is a simple subalgebra. Then both bilinear forms $(\cdot,\cdot)$ and $\langle\cdot,\cdot\rangle$ on $\mathfrak{g}$ are symmetric and invariant, so they must be proportional, that is,

$$\langle\cdot,\cdot\rangle=k(\cdot,\cdot)\quad\text{for some } k\in\mathbb{C}. \tag{6-2}$$

Then, for any $u, v\in V_1$ and integers m, n, one has

$$[u_m, v_n]=[u,v]_{m+n}+mu_1v\delta_{m+n,0}.$$

Comparing this with (6-1) shows that the map

$$u(m)\to u_m\quad\text{for } u\in\mathfrak{g} \text{ and } m\in\mathbb{Z}$$

together with $K\to k$ gives rise to a representation of $\hat{\mathfrak{g}}$ of level k.

Now we are going to state our main result related to C_2-integrability, which has already been proved to be true in [Dong and Mason 2006] for vertex operator algebras satisfying C_2-cofiniteness. But given a vertex operator superalgebra $V=V_{\bar{0}}\oplus V_{\bar{1}}$ which satisfies the C_2-cofinite condition, generally, we can not prove that the even part $V_{\bar{0}}$ also has such a property. So in this sense, the following result sharpens [Dong and Mason 2006, Theorem 3.1], although the idea is similar.

Theorem 6.1. *Let V be a simple vertex operator superalgebra which is C_2-cofinite of strong CFT type, with $\mathfrak{g}\subset V_1$ a simple Lie subalgebra, k the level of V as $\hat{\mathfrak{g}}$-module, and the vertex operator subalgebra U of V generated by $\mathfrak{g}$. Then*:

(a) *The restriction of $\langle\cdot,\cdot\rangle$ to $\mathfrak{g}$ is nondegenerate.*

(b) $U\cong L(k,0)$.

(c) *k is a positive integer.*

(d) *V is an integrable $\hat{\mathfrak{g}}$-module.*

Proof. Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$, and $\mathfrak{g}=\mathfrak{h}\oplus\sum_{\alpha\in\Phi}\mathfrak{g}_\alpha$ the corresponding Cartan decomposition of $\mathfrak{g}$. Since $\mathfrak{g}$ is generated by subalgebras isomorphic to $\mathrm{sl}(2,\mathbb{C})$, it is good enough to show the theorem for $\mathfrak{g}=\mathrm{sl}(2,\mathbb{C})$. Let $\{h,x,y\}$ be the standard basis of $\mathfrak{g}$. Then $(\alpha,\alpha)=2$ and $k=\langle\alpha,\alpha\rangle/2$ from this and Equation (6-2).

Clearly, $U=\langle\mathfrak{g}\rangle$ is a quotient of $V(k,0)$. So U is a $\hat{\mathfrak{g}}$-integrable module if and only if $\mathcal{U}=L(k,0)$ for some $k\in\mathbb{Z}_+$. This is also equivalent to the existence of a positive integer r such that

$$(x_{-1})^r\mathbf{1}=0. \tag{6-3}$$

The proof of (6-3) is similar to the same result in [Dong and Mason 2006] and we omit the proof. (b) then immediately follows. Also note that $\mathfrak{g}\subset U$, so U can not be a one-dimensional trivial module. Thus $k\neq 0$ and k must be a positive integer, proving (c) and (a). Since $L(k,0)$ is rational (cf. [Dong et al. 1997]), V is a direct sum of irreducible $L(k,0)$-modules, each of which is integrable as $\hat{\mathfrak{g}}$-module. Hence V is an integrable $\hat{\mathfrak{g}}$-module. This proves (d). □

Next we consider a toral subalgebra of V_1. Let V be strongly rational or strongly σ-rational, and let $\mathfrak{h} \subset V_1$ be a toral subalgebra such that the restriction of $\langle\cdot,\cdot\rangle$ to $\mathfrak{h}$ remains nondegenerate. Notably, any Cartan subalgebra of V_1 automatically satisfies such a condition.

Theorem 6.2. *Suppose V is strongly rational or strongly σ-rational. Let $\mathfrak{h} \subset V_1$ be a toral subalgebra such that the restriction of $\langle\cdot,\cdot\rangle$ to $\mathfrak{h}$ is nondegenerate. Then there exists a positive-definite even lattice $L \subset \mathfrak{h}$ with rank* $\dim\mathfrak{h}$ *and a vertex operator super subalgebra U of V such that $\mathfrak{h} \subset U \cong V_L$.*

This theorem has been proved in [Dong and Mason 2006] for vertex operator algebras; see also [Mason 2011]. The same argument using Lemma 5.5 is also valid for vertex operator superalgebras.

We now assume that

$$V_1 = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s \oplus \mathfrak{h},$$

where $\mathfrak{g}_i$ are simple Lie algebras and $Z(V_1) = \mathfrak{h}$. By Theorems 5.3, 6.1, and 6.2 we have the following; see [Dong and Mason 2006; Mason 2011].

Corollary 6.3. *The V contains a strongly rational vertex operator subalgebra*

$$U = L_{\mathfrak{g}_1}(k_1,0) \otimes \cdots \otimes L_{\mathfrak{g}_s}(k_s,0) \otimes V_L,$$

where the commutant U^c of U in V is a vertex operator superalgebra such that $U_1^c = 0$.

References

[Abe 2005] T. Abe, "Rationality of the vertex operator algebra V_L^+ for a positive definite even lattice L", *Math. Z.* **249**:2 (2005), 455–484. MR 2005k:17033 Zbl 1136.17021

[Borcherds 1986] R. E. Borcherds, "Vertex algebras, Kac–Moody algebras, and the monster", *Proc. Nat. Acad. Sci. U.S.A.* **83**:10 (1986), 3068–3071. MR 87m:17033 Zbl 0613.17012

[Dijkgraaf et al. 1989] R. Dijkgraaf, C. Vafa, E. Verlinde, and H. Verlinde, "The operator algebra of orbifold models", *Comm. Math. Phys.* **123**:3 (1989), 485–526. MR 91c:81132 Zbl 0674.46051

[Dong and Griess 2002] C. Dong and R. L. Griess, Jr., "Automorphism groups and derivation algebras of finitely generated vertex operator algebras", *Michigan Math. J.* **50**:2 (2002), 227–239. MR 2003e:17043 Zbl 1038.17020

[Dong and Lepowsky 1993] C. Dong and J. Lepowsky, *Generalized vertex algebras and relative vertex operators*, Progress in Mathematics **112**, Birkhäuser, Boston, MA, 1993. MR 95b:17032 Zbl 0803.17009

[Dong and Mason 2004] C. Dong and G. Mason, "Rational vertex operator algebras and the effective central charge", *Int. Math. Res. Not.* **2004**:56 (2004), 2989–3008. MR 2005k:17034 Zbl 1106.17032

[Dong and Mason 2006] C. Dong and G. Mason, "Integrability of C_2-cofinite vertex operator algebras", *Int. Math. Res. Not.* **2006** (2006), Art. ID 80468. MR 2007c:17036 Zbl 1106.17033

[Dong and Zhang 2008] C. Dong and W. Zhang, "Rational vertex operator algebras are finitely generated", *J. Algebra* **320**:6 (2008), 2610–2614. MR 2009e:17053 Zbl 1233.17023

[Dong and Zhao 2005] C. Dong and Z. Zhao, "Modularity in orbifold theory for vertex operator superalgebras", *Comm. Math. Phys.* **260**:1 (2005), 227–256. MR 2007a:17041 Zbl 1133.17016

[Dong and Zhao 2006] C. Dong and Z. Zhao, "Twisted representations of vertex operator superalgebras", *Commun. Contemp. Math.* **8**:1 (2006), 101–121. MR 2007b:17038 Zbl 1130.17012

[Dong et al. 1997] C. Dong, H. Li, and G. Mason, "Regularity of rational vertex operator algebras", *Adv. Math.* **132**:1 (1997), 148–166. MR 98m:17037 Zbl 0902.17014

[Dong et al. 1998a] C. Dong, H. Li, and G. Mason, "Twisted representations of vertex operator algebras", *Math. Ann.* **310**:3 (1998), 571–600. MR 99d:17030 Zbl 0890.17029

[Dong et al. 1998b] C. Dong, H. Li, and G. Mason, "Twisted representations of vertex operator algebras and associative algebras", *Int. Math. Res. Not.* **1998**:8 (1998), 389–397. MR 99f:17032 Zbl 0916.17023

[Dong et al. 1998c] C. Dong, H. Li, and G. Mason, "Vertex operator algebras and associative algebras", *J. Algebra* **206**:1 (1998), 67–96. MR 99i:17029 Zbl 0911.17017

[Dong et al. 2000] C. Dong, H. Li, and G. Mason, "Modular-invariance of trace functions in orbifold theory and generalized moonshine", *Comm. Math. Phys.* **214**:1 (2000), 1–56. MR 2001k:17043 Zbl 1061.17025

[Feingold et al. 1991] A. J. Feingold, I. B. Frenkel, and J. F. X. Ries, *Spinor construction of vertex operator algebras, triality, and $E_8^{(1)}$*, Contemporary Mathematics **121**, American Mathematical Society, Providence, RI, 1991. MR 92k:17041 Zbl 0743.17029

[Frenkel and Zhu 1992] I. B. Frenkel and Y. Zhu, "Vertex operator algebras associated to representations of affine and Virasoro algebras", *Duke Math. J.* **66**:1 (1992), 123–168. MR 93g:17045 Zbl 0848.17032

[Frenkel et al. 1988] I. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the monster*, Pure and Applied Mathematics **134**, Academic, Boston, MA, 1988. MR 90h:17026 Zbl 0674.17001

[Frenkel et al. 1993] I. B. Frenkel, Y.-Z. Huang, and J. Lepowsky, *On axiomatic approaches to vertex operator algebras and modules*, Mem. Amer. Math. Soc. **104**, American Mathematical Society, Providence, RI, 1993. MR 94a:17007 Zbl 0789.17022

[Gaberdiel and Neitzke 2003] M. R. Gaberdiel and A. Neitzke, "Rationality, quasirationality and finite W-algebras", *Comm. Math. Phys.* **238**:1-2 (2003), 305–331. MR 2004g:81084 Zbl 1042.17025

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, 1990. MR 92k:17038 Zbl 0716.17022

[Kac and Wang 1994] V. Kac and W. Wang, "Vertex operator superalgebras and their representations", pp. 161–191 in *Mathematical aspects of conformal and topological field theories and quantum groups* (South Hadley, MA, 1992), edited by P. Sally, Jr. et al., Contemp. Math. **175**, American Mathematical Society, Providence, RI, 1994. MR 95k:17040 Zbl 0838.17035

[Karel and Li 1999] M. Karel and H. Li, "Certain generating subspaces for vertex operator algebras", *J. Algebra* **217**:2 (1999), 393–421. MR 2001i:17034 Zbl 0951.17012

[Knopp 1970] M. I. Knopp, *Modular functions in analytic number theory*, Markham, Chicago, IL, 1970. MR 42 #198 Zbl 0259.10001

[Knopp and Mason 2003] M. Knopp and G. Mason, "On vector-valued modular forms and their Fourier coefficients", *Acta Arith.* **110**:2 (2003), 117–124. MR 2004j:11043 Zbl 1044.11020

[Lepowsky and Li 2004] J. Lepowsky and H. Li, *Introduction to vertex operator algebras and their representations*, Progress in Mathematics **227**, Birkhäuser, Boston, MA, 2004. MR 2004k:17050 Zbl 1055.17001

[Li 1994] H. S. Li, "Symmetric invariant bilinear forms on vertex operator algebras", *J. Pure Appl. Algebra* **96**:3 (1994), 279–297. MR 96e:17063 Zbl 0813.17020

[Li 1996a] H.-S. Li, "Local systems of twisted vertex operators, vertex operator superalgebras and twisted modules", pp. 203–236 in *Moonshine, the monster, and related topics* (South Hadley, MA, 1994), Contemp. Math. **193**, American Mathematical Society, Providence, RI, 1996. MR 1372724 (96m:17050) Zbl 0844.17022

[Li 1996b] H.-S. Li, "Local systems of vertex operators, vertex superalgebras and modules", *J. Pure Appl. Algebra* **109**:2 (1996), 143–195. MR 97d:17016 Zbl 0854.17035

[Mason 2011] G. Mason, "Lattice subalgebras of strongly regular vertex operator algebras", preprint, 2011. arXiv 1110.0544

[Xu 1998] X. Xu, *Introduction to vertex operator superalgebras and their modules*, Mathematics and its Applications **456**, Kluwer, Dordrecht, 1998. MR 2000h:17019 Zbl 0929.17030

[Zhu 1996] Y. Zhu, "Modular invariance of characters of vertex operator algebras", *J. Amer. Math. Soc.* **9**:1 (1996), 237–302. MR 96c:17042 Zbl 0854.17034

Received November 23, 2011. Revised January 25, 2012.

CHONGYING DONG
DEPARTMENT OF MATHEMATICS
UC SANTA CRUZ
194 BASKIN ENGINEERING
SANTA CRUZ, CALIFORNIA 95064
UNITED STATES

dong@ucsc.edu

JIANZHI HAN
WU WEN-TSUN KEY LABORATORY OF MATHEMATICS
USTC
CHINESE ACADEMY OF SCIENCES, HEFEI 230026
CHINA

and

DEPARTMENT OF MATHEMATICS
SICHUAN UNIVERSITY
CHENGDU, 610064
CHINA

jzzhan@mail.ustc.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

ON THE GEOMETRIC FLOWS SOLVING KÄHLERIAN INVERSE σ_k EQUATIONS

HAO FANG AND MIJIA LAI

Here we extend our previous work on the inverse σ_k problem. The inverse σ_k problem is a fully nonlinear geometric PDE on compact Kähler manifolds. Given a proper geometric condition, we prove that a large family of nonlinear geometric flows converges to the desired solution of the given PDE.

1. Introduction

We study general flows for the inverse σ_k-curvature problem in Kähler geometry. This is a continuation of our previous work [Fang et al. 2011].

Geometric curvature flow has been a central topic in the recent development of geometric analysis. The σ_k-curvature problems and inverse σ_k-curvature problems, fully nonlinear in nature, have appeared in several geometric settings. Andrews [1994; 2007] studies the curvature flow of embedded convex hypersurfaces in the Euclidean space. Several authors study the σ_k-equation in conformal geometry; see, for example, [Viaclovsky 2000; Chang et al. 2002; Guan and Wang 2003; Brendle 2005] and references therein. It is thus interesting to explore the corresponding problem in Kähler geometry.

In Kähler geometry, special cases of the σ_k-problem have appeared in earlier literature. Among them, one important example is Yau's seminal work on the complex Monge–Ampère equations in the Calabi conjecture. The general case has been studied recently in [Hou 2009; Hou et al. 2010]. There exist, however, some analytical difficulties in completely solving this problem for $k < n$.

Another important example is Donaldson's J-flow [1999], which gives rise to an inverse σ_1-type equation. J-flow is fully studied in [Chen 2000; 2004; Weinkove 2004; 2006; Song and Weinkove 2008]. The general case is described and treated in [Fang et al. 2011], via a specific geometric flow. In contrast to the σ_k-problem, we can pose nice geometric conditions to overcome the analytical difficulties in the inverse σ_k-problem. Here we construct more general geometric flows to solve this problem.

Hao Fang is supported in part by National Science Foundation grant DMS-1008249.
MSC2010: 35K55, 53C44, 53C55.
Keywords: fully nonlinear geometric flows, inverse σ_k equation.

We now describe the problem in more detail.

Let (M, ω) be a compact Kähler manifold without boundary. Let χ be a Kähler metric in the class $[\chi]$ other than $[\omega]$. For a fixed integer $1 \le k \le n$, we define

$$\sigma_k(\chi) = \binom{n}{k} \frac{\chi^k \wedge \omega^{n-k}}{\omega^n}.$$

It is easy to see that $\sigma_k(\chi)$ is a global defined function on M, and pointwise it is the k-th elementary symmetric polynomial on the eigenvalues of χ with respect to ω. Define

$$c_k := \frac{\int_M \sigma_{n-k}(\chi)}{\int_M \sigma_n(\chi)} = \binom{n}{k} \frac{[\chi]^{n-k} \cdot [\omega]^k}{[\chi]^n},$$

a topological constant depending only on cohomology classes $[\chi]$ and $[\omega]$.

Problem [Fang et al. 2011]. Let (M, ω), χ and c_k be given as above. Is there a metric $\tilde{\chi} \in [\chi]$ satisfying

$$c_k \tilde{\chi}^n = \binom{n}{k} \tilde{\chi}^{n-k} \wedge \omega^k? \tag{1-1}$$

To tackle this problem, we consider the geometric flow

$$\begin{cases} \dfrac{\partial}{\partial t}\varphi = c_k^{1/k} - \Big(\dfrac{\sigma_{n-k}(\chi_\varphi)}{\sigma_n(\chi_\varphi)}\Big)^{1/k}, \\ \varphi(0) = 0 \end{cases} \tag{1-2}$$

in the space of Kähler potentials of χ:

$$\mathcal{P}_\chi := \Big\{\varphi \in C^\infty(M) \,\big|\, \chi_\varphi := \chi + \frac{\sqrt{-1}}{2}\partial\bar{\partial}\varphi > 0\Big\}.$$

It is easy to see that the stationary point of the flow corresponds to the solution of (1-1).

When $k = 1$, Equation (1-2) is Donaldson's J-flow [1999], defined in the setting of the moment map; see [Chen 2000]. In this case, Song and Weinkove [2008] provide a necessary and sufficient condition for the flow to converge to the critical metric. For general k, this problem is solved in [Fang et al. 2011] with an analogous condition, which we now describe.

We define $\mathscr{C}_k(\omega)$ to be

$$\mathscr{C}_k(\omega) = \Big\{[\chi] > 0 \,\Big|\, \text{there exists } \chi' \in [\chi] \text{ such that } n c_k \chi'^{n-1} - \binom{n}{k}(n-k)\chi'^{n-k-1} \wedge \omega^k > 0\Big\}. \tag{1-3}$$

Here the inequality indicates that the left-hand side is a positive $(n-1, n-1)$ form.

For $k = n$, condition (1-3) holds for any Kähler class. Hence $\mathscr{C}_n(\omega)$ is the entire Kähler cone of M.

The need for the cone condition (1-3) is easy to see once we write (1-1) locally as

$$\frac{\sigma_{n-k}(\chi)}{\sigma_n(\chi)} = \sigma_k(\chi^{-1}) = c_k.$$

Here χ^{-1} denotes the inverse matrix of χ under local coordinates. Since $\chi^{-1} > 0$, we necessarily have, for all i,

$$\sigma_k(\chi^{-1} \mid i) < c_k.$$

This condition is equivalent to the cone condition (1-3). See [Fang et al. 2011, Proposition 2.4].

In this note, we generalize the following result:

Theorem 1.1 [Fang et al. 2011]. *Let (M, ω) be a compact Kähler manifold. Let k be a fixed integer $1 \le k \le n$. Assume $\chi \in [\chi]$ is another Kähler form and $[\chi] \in \mathscr{C}_K(\omega)$; then the flow*

$$\frac{\partial}{\partial t}\varphi = c_k^{1/k} - \left(\frac{\sigma_{n-k}(\chi_\varphi)}{\sigma_n(\chi_\varphi)}\right)^{1/k}, \tag{1-4}$$

with any initial value $\chi_0 \in [\chi]$, has long-time existence and converges to a unique smooth metric $\tilde{\chi} \in [\chi]$ satisfying

$$c_k \tilde{\chi}^n = \binom{n}{k} \tilde{\chi}^{n-k} \wedge \omega^k. \tag{1-5}$$

Specifically, we study an abstract flow on M of the form

$$\begin{cases} \dfrac{\partial}{\partial t}\varphi = F(\chi_\varphi) - C, \\ \varphi(0) = 0, \end{cases} \tag{1-6}$$

where, for $f \in C^\infty(\mathbb{R}_{>0}, \mathbb{R})$,

$$F(\chi_\varphi) = f\left[\frac{\sigma_{n-k}(\chi_\varphi)}{\sigma_n(\chi_\varphi)}\right], \quad C = f(c_k).$$

Note that (1-2) is a special case of (1-6) for $f(x) = -x^{1/k}$.

Abusing notation, we also regard F as a symmetric function on

$$\Gamma_n := \{\chi \in \mathbb{R}^n \mid \chi_1 > 0, \ \chi_2 > 0, \ \dots, \ \chi_n > 0\}$$

by writing $F(\chi_\varphi) = F(\chi_1, \dots, \chi_n)$, where (χ_i) are eigenvalues of χ_φ with respect to ω. Then by carefully examining the proof of Theorem 1.1 in [Fang et al. 2011], we observed that the following structure conditions on F are necessary:

- Ellipticity: $F_i > 0$.
- Concavity: $F_{ij} \le 0$.
- Strong concavity: $F_{ij} + (F_i/\chi_j)\delta_{ij} \le 0$.

Here $F_i = \partial F/\partial \chi_i$ and $F_{ij} = \partial^2 F/\partial \chi_i \partial \chi_j$. Concavity of F follows from strong concavity and ellipticity of F.

It is easy to check that $F(\chi_1, \dots, \chi_n) := -\big(\sigma_{n-k}(\chi)/\sigma_n(\chi)\big)^{1/k}$ satisfies these conditions.

We prove the following:

Theorem 1.2 (Main theorem). *Let (M, ω) be a compact Kähler manifold and let k be a fixed integer*, $1 \le k \le n$. *Let χ be another Kähler metric such that $[\chi] \in \mathscr{C}_k$. Assume that $f \in C^\infty(\mathbb{R}_{>0}, \mathbb{R})$ satisfies the conditions*

$$f' < 0, \quad f'' \ge 0, \quad f'' + \frac{f'}{x} \le 0. \tag{1-7}$$

Then the flow (1-6) *with any initial value $\chi_0 \in [\chi]$ has long-time existence and the metric χ_φ converges in C^∞-norm to the critical metric $\tilde{\chi} \in [\chi]$ that is the unique solution of* (1-1).

Remark 1.3. The novelty of our theorem is that there exists a large family of nonlinear geometric flows that yields the convergence towards the solution of the inverse σ_k problem (1-1). For example, the function f can be chosen as $f(x) = -\ln x$ or $f(x) = -x^p$, for $0 < p \le 1$. For the special case $f(x) = -\ln x$ and $k = n$, we get an analogue of the Kähler–Ricci flow. For $f(x) = -x$ and $k = n$, a similar flow was studied in [Cao and Keller 2011].

Remark 1.4. Theorem 1.2 is inspired by, and can be viewed as a Kähler analogue of, Andrews' result [2007] on pinching estimates of evolutions of convex hypersurfaces. In fact, our structure conditions are very similar to his.

This paper is organized as follows: in Section 2, we discuss the conditions on f and strong concavity of F; in Section 3, we give the proof of the main result.

2. Strong concavity

Here we explore concavity properties for functions involving the quotient of elementary symmetric polynomials.

Proposition 2.1. *Let $\chi \in \Gamma_n$ and $f : \mathbb{R}_{>0} \to \mathbb{R}$, define*

$$\rho(\chi_1, \dots, \chi_n) = f\big(\sigma_{n-k}(\chi)/\sigma_n(\chi)\big),$$

and suppose f satisfies the conditions

$$f' < 0, \quad f'' \ge 0, \quad f'' + \frac{f'}{x} \le 0. \tag{2-1}$$

Then ρ satisfies:

- *Ellipticity*: $\rho_i > 0$ *for all* i.
- *Concavity*: $\rho_{ij} \le 0$.
- *Strong concavity*: $\rho_{ij} + (\rho_i/\chi_j)\delta_{ij} \le 0$.

We refer to the conditions in (2-1) as the structure conditions on f.

The proof is based on the following two propositions:

Proposition 2.2. *Let* $g(\chi_1, \dots, \chi_n) = \log \sigma_k(\chi)$ *and* $\chi \in \Gamma_n$. *Then*

- $g_i > 0$,
- $g_{ij} \le 0$, *and*
- $g_{ij} + (g_i/\chi_j)\delta_{ij} \ge 0$.

Proposition 2.3. *Let* $h(\chi_1, \dots, \chi_n) := -g(1/\chi_1, \dots, 1/\chi_n) = -\log \sigma_k(\chi^{-1})$ *and* $\chi \in \Gamma_n$. *Then*

- $h_i > 0$,
- $h_{ij} \le 0$, *and*
- $h_{ij} + (h_i/\chi_j)\delta_{ij} \le 0$.

We refer the reader to the appendix of [Fang et al. 2011] for a detailed proof of Propositions 2.2 and 2.3.

Proof of Proposition 2.1. Direct computation shows

$$\rho_i = -f'\sigma_{k-1}(\chi^{-1} \mid i)\frac{1}{\chi_i^2} > 0.$$

Concavity of ρ follows from strong concavity and $\rho_i > 0$, and hence it suffices to show that

$$\rho_{ij} + \frac{\rho_i}{\chi_j}\delta_{ij} \le 0.$$

Direct computation yields

$$\rho_{ij} + \frac{\rho_i}{\chi_j}\delta_{ij} = f''\sigma_{k-1}(\chi^{-1} \mid i)\sigma_{k-1}(\chi^{-1} \mid j)\frac{1}{\chi_i^2}\frac{1}{\chi_j^2} + f'\sigma_{k-2}(\chi^{-1} \mid i, j)\frac{1}{\chi_i^2}\frac{1}{\chi_j^2}(1-\delta_{ij}) + \sigma_{k-1}(\chi^{-1} \mid i)\frac{1}{\chi_i^3}\delta_{ij}. \tag{2-2}$$

Since $f'' + f'/x \le 0$ and $f'' \ge 0$, we have

$$\rho_{ij} + \frac{\rho_i}{\chi_j}\delta_{ij} \le f''\left\{\frac{\sigma_{k-1}(\chi^{-1} \mid i)\sigma_{k-1}(\chi^{-1} \mid j)}{\chi_i^2\chi_j^2} - \sigma_k(\chi^{-1})\left[\frac{\sigma_{k-2}(\chi^{-1} \mid i, j)}{\chi_i^2\chi_j^2}(1-\delta_{ij}) + \frac{\sigma_{k-1}(\chi^{-1} \mid i)}{\chi_i^3}\delta_{ij}\right]\right\} \le 0. \tag{2-3}$$

The last inequality follows from Proposition 2.3 and the equality

$$h_{ij} + \frac{h_i}{\chi_j}\delta_{ij} = \frac{1}{\sigma_k(\chi^{-1})^2}\bigg\{\frac{\sigma_{k-1}(\chi^{-1} \mid i)\sigma_{k-1}(\chi^{-1} \mid j)}{\chi_i^2\chi_j^2} - \sigma_k(\chi^{-1})\bigg[\frac{\sigma_{k-2}(\chi^{-1} \mid i, j)}{\chi_i^2\chi_j^2}(1-\delta_{ij}) + \frac{\sigma_{k-1}(\chi^{-1} \mid i)}{\chi_i^3}\delta_{ij}\bigg]\bigg\}. \quad \square \tag{2-4}$$

For a hermitian matrix $A = (a_{i\bar{j}})$, let its eigenvalues be $\chi = (\chi_1, \dots, \chi_n)$. For $f \in C^\infty(\mathbb{R}_{>0}, \mathbb{R})$, we define

$$F(A) := \rho(\chi_1, \dots, \chi_n) = f\Big(\frac{\sigma_{n-k}(\chi)}{\sigma_n(\chi)}\Big).$$

Define

$$F^{i\bar{j}} := \frac{\partial F}{\partial a_{i\bar{j}}}, \quad F^{i\bar{j},k\bar{l}} := \frac{\partial^2 F}{\partial a_{i\bar{j}} a_{k\bar{l}}}.$$

It is a classical result that the properties of $F(A)$ follow from those of $\rho(\chi)$; see, for example, [Spruck 2005, Theorem 1.4]. In particular, Proposition 2.1 leads to the following:

Proposition 2.4. *Let $F(A)$ be defined as above, and let $f \in C^\infty(\mathbb{R}_{>0}, \mathbb{R})$ satisfy* (2-1). *Then F satisfies*:

- *Ellipticity*: $F^{i\bar{j}} > 0$.
- *Concavity*: $F^{i\bar{j},k\bar{l}} \le 0$.
- *Strong concavity*: *at* $A = \operatorname{diag}(\chi_1, \dots, \chi_n)$, *we have* $F^{i\bar{i},j\bar{j}} + (F^{i\bar{i}}/\chi_j)\delta_{ij} \le 0$.

3. Proof of the main theorem

Long-time existence. Differentiating the flow (1-6), we get

$$\frac{\partial}{\partial t}\Big(\frac{\partial\varphi}{\partial t}\Big) = F^{i\bar{j}}(\chi)\partial_i\partial_{\bar{j}}\Big(\frac{\partial\varphi}{\partial t}\Big).$$

By Proposition 2.4, $\partial\varphi/\partial t$ satisfies a parabolic equation. By the maximum principle, we have

$$\min_{t=0}\frac{\partial\varphi}{\partial t} \le \frac{\partial\varphi}{\partial t} \le \max_{t=0}\frac{\partial\varphi}{\partial t},$$

and thus

$$\min F(\chi_0) \le F(\chi_\varphi) = f(\sigma_k(\chi_\varphi^{-1})) \le \max F(\chi_0).$$

By the monotonicity of f, there exist two universal positive constants λ_1 and λ_2 such that

$$\lambda_1 \le \sigma_k(\chi_\varphi^{-1}) \le \lambda_2. \tag{3-1}$$

This implies that χ_φ remains Kähler; that is, $\chi_\varphi > 0$. Also, with the bound (3-1), regarding the estimate aspect, f, f', and f'' are all bounded.

Concerning the behavior of the flow (1-6) for arbitrary triple data (M, ω, χ), we have:

Theorem 3.1. *Let (M, ω, χ) be given as above; the general inverse σ_k flow (1-6) has long-time existence.*

Proof. Following [Chen 2004], we derive time-dependent C^2-estimates for the potential φ. Since $\chi_\varphi > 0$, it suffices to derive an upper bound for $G := tr_\omega \chi_\varphi = g^{p\bar{q}} \chi_{p\bar{q}}$. By a straightforward computation, we get

$$\begin{aligned}\frac{\partial G}{\partial t} &= g^{p\bar{q}} F^{i\bar{j},k\bar{l}} \chi_{i\bar{j},p} \chi_{k\bar{l},\bar{q}} + g^{p\bar{q}} F^{i\bar{j}} \chi_{i\bar{j},p\bar{q}} \\ &= F^{i\bar{j}} (g^{p\bar{q}} \chi_{p\bar{q}})_{i\bar{j}} + g^{p\bar{q}} F^{i\bar{j},k\bar{l}} \chi_{i\bar{j},p} \chi_{k\bar{l},\bar{q}} + g^{p\bar{q}} F^{i\bar{j}} (\chi_{m\bar{q}} R^m_{pi\bar{j}} - \chi_{m\bar{j}} R^m_{pi\bar{q}}).\end{aligned} \tag{3-2}$$

The second term is nonpositive by the concavity of F. For the last term, by choosing normal coordinates, it is easy to see that

$$g^{p\bar{q}} F^{i\bar{j}} (\chi_{m\bar{q}} R^m_{pi\bar{j}} - \chi_{m\bar{j}} R^m_{pi\bar{q}}) \le C_3 + C_4 G, \tag{3-3}$$

for two universal positive constants.

Now the upper bound of G follows from the standard maximum principle. Consequently, we have long-time existence for the flow (1-6). □

In what follows, we give the proof of the main theorem. Following [Fang et al. 2011], we first derive a partial C^2-estimate for the potential φ depending on the C^0-norm of φ when the condition $[\chi] \in \mathscr{C}_k(\omega)$ holds. Then we follow the method developed in [Song and Weinkove 2008] to get a uniform C^0-estimate and the convergence of the flow.

Partial C^2-estimate. Without loss of generality, we can assume the initial metric χ_0 is the metric χ' in $[\chi]$ satisfying cone condition (1-3). Since different initial data differ by a fixed potential function, the same estimates carry over. Again, since $\chi_\varphi > 0$, it suffices to bound χ_φ from above. Take $G(x, t, \xi) := \log(\chi_{i\bar{j}} \xi^i \xi^{\bar{j}}) - A\varphi$, for $x \in M$ and $\xi \in \mathbf{T}^{(1,0)}_x M$ with $g_{i\bar{j}} \xi^i \xi^{\bar{j}} = 1$. A is a constant to be determined. Assume G attains its maximum at $(x_0, t_0) \in M \times [0, t]$, along the direction ξ_0. Choose normal coordinates of ω at x_0, such that $\xi_0 = \partial/\partial z_1$ and $(\chi_{i\bar{j}})$ is diagonal at x_0. By the definition of G, it is easy to see that $\chi_{1\bar{1}} = \chi_1$ is the largest eigenvalue of $\{\chi_{i\bar{j}}\}$ at x_0. We can assume $t_0 > 0$; otherwise we would be done. Thus, locally, we consider $H := \log \chi_{1\bar{1}} - A\varphi$ instead, which also achieves its maximum at (x_0, t_0).

For simplicity, we write $\chi = \chi_\varphi$. At x_0, assume that $\chi = \operatorname{diag}(\chi_1, \dots, \chi_n)$ with $\chi_1 \ge \chi_2 \cdots \ge \chi_n > 0$. We use χ to denote the hermitian matrix $(\chi_{i\bar{j}})$ or the set of the eigenvalues of χ_φ interchangeably when no confusion arises.

We compute the evolution of H:

$$\frac{\partial H}{\partial t}=\frac{\chi_{1\bar 1,t}}{\chi_{1\bar 1}}-A\frac{\partial\varphi}{\partial t}=\frac{F^{i\bar j}\chi_{i\bar j,1\bar 1}+F^{i\bar j,k\bar l}\chi_{i\bar j,1}\chi_{k\bar l,\bar 1}}{\chi_{1\bar 1}}-A\frac{\partial\varphi}{\partial t},$$
$$H_{i\bar i}=\frac{\chi_{1\bar 1,i\bar i}}{\chi_{1\bar 1}}-\frac{|\chi_{1\bar 1,i}|^2}{\chi_{1\bar 1}^2}-A\varphi_{i\bar i}.$$

By the maximum principle, at (x_0, t_0) we have

$$0\le\frac{\partial H}{\partial t}-\sum_{i=1}^n F^{i\bar i}H_{i\bar i}=\frac{1}{\chi_{1\bar 1}}F^{i\bar i}(\chi_{i\bar i,1\bar 1}-\chi_{1\bar 1,i\bar i})-A\frac{\partial\varphi}{\partial t}+AF^{i\bar i}\varphi_{i\bar i}+B, \tag{3-4}$$

where

$$B=\frac{1}{\chi_{1\bar 1}}\sum_{1\le i,j,k,l\le n}F^{i\bar j,k\bar l}\chi_{i\bar j,1}\chi_{k\bar l,\bar 1}+\sum_{i=1}^n F^{i\bar i}\frac{|\chi_{1\bar 1,i}|^2}{\chi_{1\bar 1}^2}$$

is the collection of all terms involving third-order derivatives.

We claim that $B\le 0$; the proof is presented at the end of this section. Assuming that, (3-4) leads to

$$\frac{1}{\chi_{1\bar 1}}F^{i\bar i}(\chi_{i\bar i,1\bar 1}-\chi_{1\bar 1,i\bar i})\ge A\frac{\partial\varphi}{\partial t}-AF^{i\bar i}\varphi_{i\bar i}. \tag{3-5}$$

We simplify the left-hand side of (3-5) by the Ricci identity:

$$\begin{aligned}\text{LHS}&=\frac{1}{\chi_{1\bar 1}}\sum_{i=1}^n F^{i\bar i}(\chi_{i\bar i}R_{i\bar i1\bar 1}-\chi_{1\bar 1}R_{1\bar 1i\bar i})\\&\le\frac{C_1\sum_{i=1}^n F^{i\bar i}\chi_i}{\chi_{1\bar 1}}-\sum_{i=1}^n F^{i\bar i}R_{1\bar 1i\bar i}\le\frac{C_0}{\chi_{1\bar 1}}+C_2\sum_{i=1}^n F^{i\bar i}.\end{aligned}\tag{3-6}$$

For the bound on $\sum_{i=1}^n F^{i\bar i}\chi_i$, we used (3-1) and the following computation:

$$\begin{aligned}\sum_{i=1}^n F^{i\bar i}\chi_i&=-f'\sum_{i=1}^n\sigma_{k-1}(\chi^{-1}|i)\frac{1}{\chi_i^2}\chi_i\\&=-f'\sum_{i=1}^n\sigma_{k-1}(\chi^{-1}|i)\frac{1}{\chi_i}=-kf'\sigma_k(\chi^{-1})\le C.\end{aligned}\tag{3-7}$$

To deal with the right-hand side of (3-5), we divide into two cases:

Case 1: $k<n$. In this case, we have the following technical lemma due to the cone condition.

Lemma 3.2. *For $k<n$, assume that $\chi_0=\chi'\in[\chi]$ is a Kähler form satisfying the cone condition* (1-3), *and that $C_1\le\sigma_k(\chi^{-1})\le C_2$ for two universal constants.*

Then there exists a universal constant N such that if $\chi_1/\chi_n \geq N$, then there exists a universal constant $\theta > 0$ such that

$$\sigma_k^{1/k}\Big(\frac{\chi_{0i\bar{i}}}{\chi_i^2}\Big) \geq (1+\theta)c_k^{-1/k}\sigma_k^{2/k}(\chi^{-1}). \tag{3-8}$$

We refer the reader to [Fang et al. 2011, Theorem 2.8] for a proof.

Case 1a: $\chi_1/\chi_n \geq N$, where N is given in Lemma 3.2. Applying Lemma 3.2, we claim that there exists a universal constant $\epsilon > 0$ such that

$$\frac{\partial\varphi}{\partial t} - F^{i\bar{i}}\chi_{i\bar{i}} + (1-\epsilon)F^{i\bar{i}}\chi_{0i\bar{i}} \geq 0. \tag{3-9}$$

Indeed, by direct computation, we have

$$\begin{aligned}\sum_{i=1}^n F^{i\bar{i}}\chi_{0i\bar{i}} &= -f'\sum_{i=1}^n \sigma_{k-1}(\chi^{-1}\mid i)\frac{\chi_{0i\bar{i}}}{\chi_i^2}\\ &\geq -kf'\sigma_k^{1-1/k}(\chi^{-1})\sigma_k^{1/k}\Big(\frac{\chi_{0i\bar{i}}}{\chi_i^2}\Big)\\ &\geq -kf'\sigma_k^{1-1/k}(\chi^{-1})(1+\theta)c_k^{-1/k}\sigma_k^{2/k}(\chi^{-1}).\end{aligned} \tag{3-10}$$

The first inequality follows from Gårding's inequality.

Therefore, by taking ϵ such that $(1-\epsilon)(1+\theta) = 1$, Equation (3-9) is reduced to

$$\frac{\partial\varphi}{\partial t} - F^{i\bar{i}}\chi_{i\bar{i}} - kf'\sigma_k^{1+1/k}(\chi^{-1})c_k^{-1/k} \geq 0. \tag{3-11}$$

By scaling, we can assume $c_k = 1$, and modifying f by adding a constant, we can further assume that $f(1) = 0$. Plugging in $F^{i\bar{i}}$ and letting $x = \sigma_k(\chi^{-1})$, (3-11) is equivalent to

$$f(x) + kf'(x)x - kf'(x)x^{1+1/k} \geq 0. \tag{3-12}$$

The inequality above holds provided $f'' + f'/x \leq 0$ and $f(1) = 0$.

Combining (3-5), (3-6) and (3-9), we have

$$A\epsilon\sum_{i=1}^n F^{i\bar{i}}\chi_{0i\bar{i}} \leq \frac{C_1}{\chi_1} + C_2\sum_{i=1}^n F^{i\bar{i}}. \tag{3-13}$$

Since χ_0 is a fixed form, there exists a universal constant $\lambda > 0$ such that

$$A\lambda\sum_{i=1}^n F^{i\bar{i}} \leq A\epsilon\sum_{i=1}^n F^{i\bar{i}}\chi_{0i\bar{i}}.$$

Hence, in (3-13), taking A such that $A\lambda - C_2 = 1$, an upper bound for χ_1 will follow once we have shown $\sum_{i=1}^n F^{i\bar{i}}$ is bounded from below. For that we have

$$\sum_{i=1}^n F^{i\bar{i}} = -f' \sum \sigma_{k-1}(\chi^{-1} \mid i)\frac{1}{\chi_i^2} \geq -kf'\sigma_k^{1-1/k}(\chi^{-1})\sigma_k^{1/k}\Big(\frac{1}{\chi_i^2}\Big) \geq \tilde{C}\sigma_k^{1+1/k}(\chi^{-1}) \geq C. \tag{3-14}$$

Case 1b: $\chi_1/\chi_n \leq N$. In this case, the upper bound for χ_1 follows directly from the lower bound (3-1) on $\sigma_k(\chi^{-1})$. Since

$$\lambda_1 \leq \sigma_k(\chi^{-1}) \leq \binom{n}{k}\frac{1}{\chi_n^k}, \tag{3-15}$$

we get an upper bound for χ_n, and thus an upper bound for χ_1, because $\chi_1 \leq N\chi_n$.

Case 2: $k = n$. In this case, we continue on (3-5) directly. Since we are only concerned with f on the closed interval $[\lambda_1, \lambda_2]$, we can assume that f is positive by adding a constant. By (3-6), we have that

$$\text{LHS of (3-5)} \leq \frac{C_0}{\chi_1} + C_2\sum_{i=1}^n F^{i\bar{i}} \leq C_3\sum_{i=1}^n \frac{1}{\chi_i}. \tag{3-16}$$

For the right-hand side, we have

$$\text{RHS of (3-5)} \geq A\big(-f(c_k) + nf'\sigma_n(\chi^{-1})\big) + A\epsilon C_4\sum_{i=1}^n \frac{1}{\chi_i}. \tag{3-17}$$

Combining (3-16) and (3-17) and taking A such that $A\epsilon C_4 - C_3 = 1$, we find there exists a universal constant C such that

$$\sum_{i=1}^n \frac{1}{\chi_i} \leq C. \tag{3-18}$$

Consequently, we have a lower bound on χ_i for all i, and thus an upper bound for χ_1 by (3-1).

Thus we have proved that there exists a universal constant C such that

$$\chi_1 \leq C.$$

This leads to:

Theorem 3.3. *Let the notation be as above*; *we have*

$$|\partial\bar{\partial}\varphi|_{C^0} \leq Ce^{A\varphi - \inf_{M\times[0,t]}\varphi}$$

for two universal constants A and C and any time interval $[0, t]$.

Finally, we prove the claim that

$$B = \frac{1}{\chi_{1\bar{1}}} \sum_{1\le i,j,k,l\le n} F^{i\bar{j},k\bar{l}} \chi_{i\bar{j},1} \chi_{k\bar{l},\bar{1}} + \sum_{i=1}^{n} F^{i\bar{i}} \frac{|\chi_{1\bar{1},i}|^2}{\chi_{1\bar{1}}^2} \le 0.$$

We divide B into three groups:

$$X = \frac{1}{\chi_{1\bar{1}}} \sum_{1\le i,j\le n} F^{i\bar{i},j\bar{j}} \chi_{i\bar{i},1} \chi_{j\bar{j},\bar{1}} + F^{1\bar{1}} \frac{|\chi_{1\bar{1},1}|^2}{\chi_{1\bar{1}}^2}.$$

That X is nonpositive follows from the strong concavity of F in Proposition 2.4.

$$Y = \frac{1}{\chi_{1\bar{1}}} \sum_{i=2}^{n} F^{i\bar{1},1\bar{i}} \chi_{i\bar{1},1} \chi_{1\bar{i},\bar{1}} + \sum_{i=2}^{n} F^{i\bar{i}} \frac{|\chi_{1\bar{1},i}|^2}{\chi_{1\bar{1}}^2}.$$

One sees by direct computation that $F^{i\bar{1},1\bar{i}} + F^{i\bar{i}}/\chi_1 \le 0$ for all i, and thus $Y \le 0$.

$$Z = \frac{1}{\chi_{1\bar{1}}} \sum_{i\ne j,\, j>1,\, k\ne l,\, k>1} F^{i\bar{j},k\bar{l}} \chi_{i\bar{j},1} \chi_{k\bar{l},\bar{1}}.$$

Again by direct computation, each term is nonpositive. We have thus finished the proof of the claim.

C^0-estimate and convergence of the flow. Following the method in [Song and Weinkove 2008], we introduce two functionals. The monotonic behavior of these functionals along the flow (1-6) yields the C^0-estimate and convergence of the flow. Define functionals in $\mathcal{P}_{\chi_0}$ by

$$\mathcal{F}_{k,\chi_0}(\phi) = \mathcal{F}_k(\phi) = \int_0^1 \int_M \dot{\phi}_t \chi_{\phi_t}^k \wedge \omega^{n-k}\, dt, \tag{3-19}$$

where ϕ_t is an arbitrary smooth path in $\mathcal{P}_{\chi_0}$ connecting 0 and ϕ, and $\dot{\phi}_t$ denotes a time derivative. One can readily check that this definition is independent of the choice of the path φ_t. Moreover, define

$$\mathcal{F}_{k,n}(\phi) = \binom{n}{k} \mathcal{F}_k(\phi) - c_{n-k} \mathcal{F}_n(\phi). \tag{3-20}$$

The first variation of $\mathcal{F}_{n-k,n}$ is

$$\frac{d}{dt} \mathcal{F}_{n-k,n}(\phi) = \int_M \dot{\phi}_t \left(\binom{n}{k} \chi_{\phi_t}^{n-k} \wedge \omega^k - c_k \chi_{\phi_t}^n \right).$$

It follows that the Euler–Lagrange equation of $\mathcal{F}_{n-k,n}$ is precisely the critical equation (1-1):

$$c_k \chi_\phi^n = \binom{n}{k} \chi_\phi^{n-k} \wedge \omega^k.$$

We have the following properties, the first of which is shown in [Fang et al. 2011, Theorem 4.1].

Proposition 3.4 (uniqueness). *The solution to the critical equation* (1-1) *is unique up to a constant.*

Proposition 3.5 (monotonicity of $\mathcal{F}_{n-k,n}$). *The functional $\mathcal{F}_{n-k,n}$ is decreasing along the flow* (1-6).

Proof. By direct computation, we have

$$\begin{aligned}\frac{d}{dt}\mathcal{F}_{n-k,n}(\varphi_t) &= \int_M \dot{\varphi}_t\left(\binom{n}{k}\chi_\varphi^{n-k}\wedge\omega^k - c_k\chi_\varphi^n\right)\\ &= \int_M \big(f(\sigma_k(\chi_\varphi^{-1})) - f(c_k)\big)(\sigma_k(\chi_\varphi^{-1}) - c_k)\chi_\varphi^n < 0.\end{aligned} \tag{3-21}$$

The integrand is of the form $(f(a) - f(b))(a - b)$, which is negative because $f' < 0$. □

Proposition 3.6 (monotonicity of $\mathcal{F}_{n-k}$). *The functional $\mathcal{F}_{n-k}$ is nonincreasing along the flow* (1-6).

Proof. First define $g(x) = f(1/x)$. It follows that g is concave if and only if $f'' + f'/x \le 0$. Then by Jensen's inequality, we have

$$\begin{aligned}\frac{1}{\int_M \chi^{n-k}\wedge\omega^k}\int_M f(\sigma_k(\chi^{-1}))\chi^{n-k}\wedge\omega^k &= \frac{1}{\int_M \chi^{n-k}\wedge\omega^k}\int_M g\left(\frac{\sigma_n(\chi)}{\sigma_{n-k}(\chi)}\right)\chi^{n-k}\wedge\omega^k\\ &\le g\left(\frac{1}{\int_M \chi^{n-k}\wedge\omega^k}\int_M \frac{\sigma_n(\chi)}{\sigma_{n-k}(\chi)}\chi^{n-k}\wedge\omega^k\right)\\ &= g\left(\frac{1}{c_k}\right) = f(c_k).\end{aligned} \tag{3-22}$$

Hence

$$\frac{\partial}{\partial t}\mathcal{F}_{n-k} = \int_M \big(f(\sigma_k(\chi_\varphi^{-1})) - f(c_k)\big)\chi_\varphi^{n-k}\wedge\omega^k \le 0. \tag{3-23}$$ □

Finally, we single out the essential steps for the rest of the proof. By [Fang et al. 2011, Theorem 4.5], we have uniform bounds for the oscillation of φ_t, that is,

$$\|\sup\varphi_t - \inf\varphi_t\| \le C.$$

Then using the functional $\mathcal{F}_{n-k}$, we obtain a suitable normalization $\hat{\varphi}_t$ of φ_t, for which we can get uniform C^0-estimates, and thus uniform C^2-estimates by Theorem 3.3. Higher-order estimates follow from the Evans–Krylov and Schauder estimates. The corresponding metric thus converges to the critical metric solving the inverse σ_k problem (1-1).

References

[Andrews 1994] B. Andrews, "Contraction of convex hypersurfaces in Euclidean space", *Calc. Var. Partial Differential Equations* **2**:2 (1994), 151–171. MR 97b:53012 Zbl 0805.35048

[Andrews 2007] B. Andrews, "Pinching estimates and motion of hypersurfaces by curvature functions", *J. Reine Angew. Math.* **608** (2007), 17–33. MR 2008i:53087 Zbl 1129.53044

[Brendle 2005] S. Brendle, "Convergence of the Yamabe flow for arbitrary initial energy", *J. Differential Geom.* **69**:2 (2005), 217–278. MR 2006e:53119 Zbl 1085.53028

[Cao and Keller 2011] H.-D. Cao and J. Keller, "About the Calabi problem: a finite dimensional approach", preprint, 2011. arXiv 1102.1097

[Chang et al. 2002] S.-Y. A. Chang, M. J. Gursky, and P. C. Yang, "An equation of Monge–Ampère type in conformal geometry, and four-manifolds of positive Ricci curvature", *Ann. of Math.* (2) **155**:3 (2002), 709–787. MR 2003j:53048 Zbl 1031.53062

[Chen 2000] X.-X. Chen, "On the lower bound of the Mabuchi energy and its application", *Internat. Math. Res. Notices* **12** (2000), 607–623. MR 2001f:32042 Zbl 0980.58007

[Chen 2004] X.-X. Chen, "A new parabolic flow in Kähler manifolds", *Comm. Anal. Geom.* **12**:4 (2004), 837–852. MR 2005h:53116 Zbl 1073.53089

[Donaldson 1999] S. K. Donaldson, "Moment maps and diffeomorphisms", *Asian J. Math.* **3**:1 (1999), 1–16. MR 2001a:53122 Zbl 0999.53053

[Fang et al. 2011] H. Fang, M. Lai, and X.-N. Ma, "On a class of fully nonlinear flows in Kähler geometry", *J. Reine Angew. Math.* **653** (2011), 189–220. MR 2012g:53134 Zbl 1222.53070

[Guan and Wang 2003] P. Guan and G. Wang, "Local estimates for a class of fully nonlinear equations arising from conformal geometry", *Internat. Math. Res. Notices* **2003**:26 (2003), 1413–1432. MR 2003m:53055 Zbl 1042.53021

[Hou 2009] Z. Hou, "Complex Hessian equation on Kähler manifold", *Internat. Math. Res. Notices* **2009**:16 (2009), 3098–3111. MR 2010m:32026 Zbl 1177.32013

[Hou et al. 2010] Z. Hou, X.-N. Ma, and D. Wu, "A second order estimate for complex Hessian equations on a compact Kähler manifold", *Math. Res. Lett.* **17**:3 (2010), 547–561. MR 2011j:32030 Zbl 1225.32026

[Song and Weinkove 2008] J. Song and B. Weinkove, "On the convergence and singularities of the J-flow with applications to the Mabuchi energy", *Comm. Pure Appl. Math.* **61**:2 (2008), 210–229. MR 2009a:32038 Zbl 1135.53047

[Spruck 2005] J. Spruck, "Geometric aspects of the theory of fully nonlinear elliptic equations", pp. 283–309 in *Global theory of minimal surfaces*, edited by D. Hoffman, Clay Math. Proc. **2**, Amer. Math. Soc., Providence, RI, 2005. MR 2006j:53051 Zbl 1151.53345

[Viaclovsky 2000] J. A. Viaclovsky, "Conformal geometry, contact geometry, and the calculus of variations", *Duke Math. J.* **101**:2 (2000), 283–316. MR 2001b:53038 Zbl 0990.53035

[Weinkove 2004] B. Weinkove, "Convergence of the J-flow on Kähler surfaces", *Comm. Anal. Geom.* **12**:4 (2004), 949–965. MR 2005g:32027 Zbl 1107.53048

[Weinkove 2006] B. Weinkove, “On the J-flow in higher dimensions and the lower boundedness of the Mabuchi energy.”, *J. Differ. Geom.* **73**:2 (2006), 351–358. MR 2007a:32026 Zbl 1107.53048

Received September 22, 2011.

HAO FANG
14 MACLEAN HALL
UNIVERSITY OF IOWA
IOWA CITY, IA 52242
UNITED STATES
haofang@math.uiowa.edu

MIJIA LAI
915 HYLAN BUILDING
UNIVERSITY OF ROCHESTER
RC BOX 270138
ROCHESTER, NY 14627
UNITED STATES
lai@math.rochester.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.305

AN OPTIMAL ANISOTROPIC POINCARÉ INEQUALITY FOR CONVEX DOMAINS

GUOFANG WANG AND CHAO XIA

In this paper, we prove a sharp lower bound of the first (nonzero) eigenvalue of the anisotropic Laplacian with the Neumann boundary condition. Equivalently, we prove an optimal anisotropic Poincaré inequality for convex domains, which generalizes the classical result of Payne and Weinberger. A lower bound of the first (nonzero) eigenvalue of the anisotropic Laplacian with the Dirichlet boundary condition is also proved.

1. Introduction

In this paper we are interested in studying the eigenvalues of the anisotropic Laplacian Q, which is a natural generalization of the ordinary Laplacian Δ. We say that F is a *norm* on $\mathbb{R}^n$ if $F : \mathbb{R}^n \to [0, +\infty)$ is a convex function of class $C^1(\mathbb{R}^n\backslash\{0\})$, which is even and positively 1-homogeneous, that is,

$$F(t\xi) = |t|F(\xi) \quad \text{for any } t \in \mathbb{R},\ \xi \in \mathbb{R}^n,$$

and

$$F(\xi) > 0 \quad \text{for any } \xi \neq 0.$$

A typical norm on $\mathbb{R}^n$ is $F(\xi) = (\sum_{i=1}^n |\xi_i|^q)^{1/q}$ for $q \in (1, \infty)$. The *anisotropic Laplacian* (or *Finsler–Laplacian*) of $u : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$Qu(x) := \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\big(F(\nabla u(x))F_{\xi_i}(\nabla u(x))\big), \quad x \in \mathbb{R}^n, \tag{1}$$

where

$$F_{\xi_i}(\xi) = \frac{\partial F}{\partial \xi_i}(\xi) \quad \text{and} \quad \nabla u(x) = \Big(\frac{\partial u}{\partial x_1}(x), \dots, \frac{\partial u}{\partial x_n}(x)\Big).$$

When $F(\xi) = |\xi| = (\sum_{i=1}^n |\xi_i|^2)^{1/2}$, the anisotropic Laplacian $Q = \Delta$, the usual Laplacian. Note that, in this paper, we use $\xi \in \mathbb{R}^n$ for F and $x \in \mathbb{R}^n$ for functions u.

Wang is partly supported by SFB/TR71 "Geometric partial differential equations" of DFG. Xia is supported by the China Scholarship Council.
MSC2010: primary 35P15; secondary 35J62, 35P30.
Keywords: anisotropic Laplacian, first eigenvalue, gradient estimate, optimal Poincaré inequality.

The anisotropic Laplacian has been studied by many mathematicians, in the context of both Finsler geometry (see, for example, [Amar and Bellettini 1994; Ge and Shen 2001; Ohta 2009; Ohta and Sturm 2011; Shen 2001]) and quasilinear PDE (see, for example, [Alvino et al. 1997; Belloni et al. 2003; Ferone and Kawohl 2009; Wang and Xia 2011b; 2011a; 2012]). Particularly, many problems related to the first eigenvalue of the anisotropic Laplacian have already been considered in [Belloni et al. 2003; Ge and Shen 2001; Kawohl 2011; Ohta 2009; Wang and Xia 2011a]. In this paper we investigate the estimates of the first eigenvalue of the anisotropic Laplacian.

Let Ω be a smooth bounded domain in $\mathbb{R}^n$ and ν the outward normal unit vector of its boundary $\partial\Omega$. The first (nonzero) eigenvalue λ_1 of the anisotropic Laplacian Q is defined by the smallest positive constant such that there exists a nonconstant function u satisfying

$$-Qu = \lambda_1 u \quad \text{in } \Omega \tag{2}$$

with the Dirichlet boundary condition

$$u = 0 \quad \text{on } \partial\Omega \tag{3}$$

or the Neumann boundary condition

$$\langle F_\xi(\nabla u), \nu\rangle = 0 \quad \text{on } \partial\Omega. \tag{4}$$

We call λ_1 the *first Dirichlet eigenvalue* (respectively the *first Neumann eigenvalue*) and denote it by λ_1^D (respectively λ_1^N). Here $\langle F_\xi(\nabla u), \nu\rangle = \sum_{i=1}^n F_{\xi_i}(\nabla u)\nu^i$ and $\nu = (\nu^1, \dots, \nu^n)$. Equation (4) is a natural Neumann boundary condition for the anisotropic Laplacian. When $F(\xi) = |\xi|$, $\langle F_\xi(\nabla u), \nu\rangle = \partial u/\partial\nu$.

The first (nonzero) Dirichlet (respectively Neumann) eigenvalue can be formulated as a variational problem by

$$\lambda_1^D(\Omega) = \inf\left\{ \frac{\int_\Omega F^2(\nabla u)\,dx}{\int_\Omega u^2\,dx} \,\middle|\, 0 \neq u \in W_0^{1,2}(\Omega) \right\}. \tag{5}$$

$$\lambda_1^N(\Omega) = \inf\left\{ \frac{\int_\Omega F^2(\nabla u)\,dx}{\int_\Omega u^2\,dx} \,\middle|\, 0 \neq u \in W^{1,2}(\Omega),\ \int_\Omega u\,dx = 0 \right\}. \tag{6}$$

Therefore obtaining a sharp estimate of first eigenvalue is equivalent to obtaining the best constant in Poincaré type inequalities.

We remark that Equation (2) should be understood in a weak sense, that is,

$$\int_\Omega \sum_{i=1}^n \frac{\partial}{\partial \xi_i}\left(\tfrac{1}{2}F^2\right)(\nabla u)\varphi_i\,dx = \int_\Omega \lambda_1 u\varphi\,dx \quad \text{for any } \varphi \in C_0^\infty(\Omega).$$

Finding a lower bound for the first eigenvalue is always an interesting problem. In [Belloni et al. 2003; Ge and Shen 2001], the authors proved the Faber–Krahn type inequality for the first Dirichlet eigenvalue of the anisotropic Laplacian. A Cheeger type estimate for the first eigenvalue of the anisotropic Laplacian involving the isoperimetric constant was also obtained there. In this paper, we are interested in the Payne–Weinberger type sharp estimate [Payne and Weinberger 1960] of the first eigenvalue in terms of some geometric quantity, such as the diameter with respect to F.

Before stating our main result, we need to introduce some concepts and definitions. We say that $\partial\Omega$ is *weakly convex* if the second fundamental form of $\partial\Omega$ with respect to the inward normal is nonnegative definite. We say that $\partial\Omega$ is *F-mean convex* if the F-mean curvature H_F is nonnegative. For the definition of F-mean curvature, see Section 2.

There is another convex function F^0 related to F, which is defined to be the support function of $K := \{x \in \mathbb{R}^n : F(x) < 1\}$, namely,

$$F^0(x) := \sup_{\xi\in K}\langle x, \xi\rangle.$$

It is easy to verify that $F^0 : \mathbb{R}^n \mapsto [0, +\infty)$ is also a convex, even, 1-positively, homogeneous function. Actually F^0 is dual to F (see, for instance, [Alvino et al. 1997]) in the sense that

$$F^0(x) = \sup_{\xi\neq 0}\frac{\langle x, \xi\rangle}{F(\xi)} \quad \text{and} \quad F(x) = \sup_{\xi\neq 0}\frac{\langle x, \xi\rangle}{F^0(\xi)}.$$

Hence the Cauchy–Schwarz inequality holds in the sense that

$$\langle \xi, \eta\rangle_{\mathbb{R}^n} \leq F(\xi)F^0(\eta). \tag{7}$$

We call $\mathcal{W}_r(x_0) := \{x \in \mathbb{R}^n \mid F^0(x - x_0) \leq r\}$ a *Wulff ball* of radius r with center at x_0. We say $\gamma : [0, 1] \to \Omega$ a *minimal geodesic* from x_1 to x_2 if

$$d_F(x_1, x_2) := \int_0^1 F^0(\dot\gamma(t))\,dt = \inf \int_0^1 F^0(\dot{\tilde\gamma}(t))\,dt,$$

where the infimum takes on all C^1 curves $\tilde\gamma(t)$ in Ω from x_1 to x_2. In fact γ is a straight line and $d_F(x_1, x_2) = F^0(x_2 - x_1)$. We call $d_F(x_1, x_2)$ the *F-distance* between x_1 and x_2.

Now we can define the *diameter d_F* of Ω with respect to the norm F on $\mathbb{R}^n$ as

$$d_F := \sup_{x_1,x_2\in\overline{\Omega}} d_F(x_1, x_2).$$

In the same spirit we define the *inscribed radius i_F* of Ω with respect to the norm F on $\mathbb{R}^n$ as the radius of the biggest Wulff ball that can be enclosed in $\overline{\Omega}$.

Our main result is the following.

Theorem 1.1. *Let Ω be a smooth bounded domain in $\mathbb{R}^n$ and $F \in C^1(\mathbb{R}^n \setminus \{0\})$ a norm on $\mathbb{R}^n$. Let λ_1^N be the first Neumann eigenvalue of the anisotropic Laplacian* (1). *Assume that $\partial\Omega$ is weakly convex. Then λ_1^N satisfies*

$$\lambda_1^N \geq \frac{\pi^2}{d_F^2}. \tag{8}$$

Moreover, *equality holds in* (8) *if and only if $n = 1$, and hence Ω is a segment.*

Estimate (8) for the Neumann boundary problem is optimal. This is in fact a generalization of the classical result of Payne and Weinberger [1960] on an optimal estimate of the first Neumann eigenvalue of the ordinary Laplacian. See also [Bebendorf 2003]. There are many interesting generalizations. Here we just mention its generalization to Riemannian manifolds, since we will use the methods developed there. It should also be interesting to ask if the methods of [Payne and Weinberger 1960] and [Bebendorf 2003] work to reprove our result, since there are lots of motivations in computational mathematics.

For a smooth compact n-dimensional Riemannian manifold (M, g) with nonnegative Ricci curvature and diameter d, possibly with boundary, the first Neumann eigenvalue λ_1 of the Laplace operator Δ is defined to be the smallest positive constant such that there is a nonconstant function u satisfying

$$-\Delta u = \lambda_1 u \quad \text{in } M$$

with

$$\frac{\partial u}{\partial \nu} = 0 \quad \text{on } \partial M,$$

if ∂M is not empty, where ν denotes the outward normal of ∂M. The fundamental work in [Li 1979; Li and Yau 1980; Zhong and Yang 1984] gives us the following optimal estimate

$$\lambda_1 \geq \frac{\pi^2}{d^2}, \tag{9}$$

where d is the diameter of M with respect to g. Li and Yau [1980] derived a gradient estimate for the eigenfunction u and proved that $\lambda_1 \geq \pi^2/(4d^2)$, and Li [1979] used another auxiliary function to obtain a better estimate $\lambda_1 \geq \pi^2/(2d^2)$. Finally, Zhong and Yang [1984] were able to use a more precise auxiliary function to get the sharp estimate $\lambda_1 \geq \pi^2/d^2$, which is optimal in the sense that the lower bound is achieved by a circle or a segment. Recently Hang and Wang [2007] proved that equality (9) holds if and only if M is a circle or a segment. For related work see [Kröger 1992; Chen and Wang 1997; Bakry and Qian 2000]. These results were generalized to the p-Laplacian in [Valtorta 2012] and to the Laplacian on Alexandrov spaces in [Qian et al. 2012].

For the Dirichlet problem we have the following.

Theorem 1.2. *Let Ω be a smooth bounded domain in $\mathbb{R}^n$ and $F \in C^1(\mathbb{R}^n \setminus \{0\})$ a norm on $\mathbb{R}^n$. Assume that λ_1^D is the first Dirichlet eigenvalue of the anisotropic Laplacian* (1). *Assume further that $\partial\Omega$ is F-mean convex. Then λ_1^N satisfies*

$$\lambda_1^D \geq \frac{\pi^2}{4i_F^2}. \tag{10}$$

Estimate (10) is by no means optimal.

Our idea to prove the result on the Dirichlet eigenvalue is based on the gradient estimate technique for eigenfunctions from [Li 1979; Li and Yau 1980]. This idea also works for the first Neumann eigenvalue to get a rough estimate, say $\lambda_1^N \geq \pi^2/(2d_F^2)$. However, for getting the sharp estimate of the first Neumann eigenvalue (8), the method of Zhong and Yang seems hard to apply. Instead, we adopt the technique based on gradient comparison with a one dimensional model function, which was developed in [Kröger 1992] and improved in [Chen and Wang 1997; Bakry and Qian 2000]. Surprisingly, we find that the one dimensional model coincides with that for the Laplacian case. In fact, this must be the case because when we consider F in $\mathbb{R}$, it can only be $F(x) = c|x|$ with $c > 0$, a multiple of the standard Euclidean norm. In order to get the gradient comparison theorem, we need a Bochner type formula (13), a Kato type inequality (14), and a refined inequality (15), which was referred to as the "extended curvature-dimension inequality" in the context of [Bakry and Qian 2000]. Interestingly, the proof of these inequalities sounds more "natural" than the proof of their counterpart for the usual Laplace operator. These inequalities may have their own interest. Another difficulty we encounter is handling the boundary maximum due to the different representation of the Neumann boundary condition (4). We find a suitable vector field V (see its explicit construction in Section 3) to avoid this difficulty. With the gradient comparison theorem, we are able to follow step by step the argument in [Bakry and Qian 2000] to get the sharp estimate. The proof for the rigidity part of Theorem 1.1 closely follows [Hang and Wang 2007]. Here we need to pay more attention to the points with vanishing $|\nabla u|$.

A natural question arises of whether one can generalize Theorem 1.1 to manifolds. The anisotropic Laplacian with the norm F does not have a direct generalization to Riemannian manifolds. However, it has a (natural) generalization to Finsler manifolds. In fact, $\mathbb{R}^n$ with F can be viewed as a special Finsler manifold. On a general Finsler manifold, there is a generalized anisotropic Laplacian; see for instance [Ge and Shen 2001; Ohta 2009; Shen 2001]. A Lichnerowicz type result for the first eigenvalue of this Laplacian was obtained in [Ohta 2009] under a condition on some kind of new Ricci curvature Ric_N, $N \in [n, \infty]$. A Li–Yau–Zhong–Yang type sharp estimate, that is, a generalization of Theorem 1.1 for this generalized Laplacian on Finsler manifolds would be a challenging problem. We will study this problem in a forthcoming paper.

The paper is organized as follows. In Section 2, we give some preliminary results on 1-homogeneous convex functions and the F-mean curvature, and prove useful inequalities. In Section 3 we prove the sharp estimate for the first Neumann eigenvalue and classify the equality case. We handle the first Dirichlet eigenvalue in Section 4.

2. Preliminary

Without loss of generality, we may assume that $F \in C^3(\mathbb{R}^n \setminus \{0\})$ and F is a *strongly convex norm* on $\mathbb{R}^n$, that is, F satisfies

$$\mathrm{Hess}(F^2) \text{ is positive definite in } \mathbb{R}^n \setminus \{0\}.$$

In fact, for any norm $F \in C^1(\mathbb{R}^n \setminus \{0\})$, there exists a sequence $F_\varepsilon \in C^3(\mathbb{R}^n \setminus \{0\})$ such that the strongly convex norm $\widetilde{F}_\varepsilon := \sqrt{F_\varepsilon^2 + \varepsilon |x|^2}$ converges to F uniformly in $C^1_{\mathrm{loc}}(\mathbb{R}^n \setminus \{0\})$. Then the corresponding first eigenvalue λ_1^ε of the anisotropic Laplacian with respect to $\widetilde{F}_\varepsilon$ converges to λ_1 as well. Here $|\cdot|$ denotes the Euclidean norm. Therefore, in the following sections, we assume that $F \in C^3(\mathbb{R}^n \setminus \{0\})$ and F is a strongly convex norm on $\mathbb{R}^n$. Thus (2) is degenerate elliptic among Ω and uniformly elliptic in $\Omega \setminus \mathscr{C}$, where $\mathscr{C} := \{x \in \Omega \mid \nabla u(x) = 0\}$ denotes the set of degenerate points. The standard regularity theory for degenerate elliptic equations (see, for example, [Belloni et al. 2003; Tolksdorf 1984]) implies that $u \in C^{1,\alpha}(\Omega) \cap C^{2,\alpha}(\Omega \setminus \mathscr{C})$.

The following property is an obvious consequence of the 1-homogeneity of F.

Proposition 2.1. *Let $F : \mathbb{R}^n \to [0, +\infty)$ be a 1-homogeneous function. Then the following holds*:

(i) $\sum_{i=1}^n F_{\xi_i}(\xi)\xi_i = F(\xi)$;

(ii) $\sum_{j=1}^n F_{\xi_i \xi_j}(\xi)\xi_j = 0$ *for any* $i = 1, 2, \ldots, n$. □

For simplicity, from now on we will follow the summation convention and frequently use the notations $F = F(\nabla u)$, $F_i = F_{\xi_i}(\nabla u)$, $u_i = \partial u/\partial x_i$, $u_{ij} = \partial^2 u/(\partial x_i \partial x_j)$, and so on. Denote

$$\begin{aligned} a_{ij}(\nabla u)(x) &:= \frac{\partial^2}{\partial \xi_i \partial \xi_j}\left(\tfrac{1}{2}F^2\right)(\nabla u(x)) = (F_i F_j + F F_{ij})(\nabla u(x)), \\ a_{ijk}(\nabla u)(x) &:= \frac{\partial^3}{\partial \xi_i \partial \xi_j \partial \xi_k}\left(\tfrac{1}{2}F^2\right)(\nabla u(x)). \end{aligned} \tag{11}$$

In the following we simply write a_{ij} and a_{ijk} if no confusion appears. With these notations, we can rewrite the anisotropic Laplacian (1) as

$$Qu = a_{ij} u_{ij}. \tag{12}$$

For the function $\frac{1}{2}F^2(\nabla u)$ we have a Bochner type formula.

Lemma 2.1 (Bochner formula). *At a point where* $\nabla u \neq 0$, *we have*

$$\begin{aligned} &a_{ij}\big(\tfrac{1}{2}F^2(\nabla u)\big)_{ij} \\ &\qquad = a_{ij}a_{kl}u_{ik}u_{jl} + (Qu)_k \frac{\partial}{\partial \xi_k}\big(\tfrac{1}{2}F^2\big)(\nabla u) - a_{ijl}\frac{\partial}{\partial x_l}\big(\tfrac{1}{2}F^2(\nabla u)\big)u_{ij}. \end{aligned} \tag{13}$$

Proof. The formula is derived from a direct computation.

$$\begin{aligned} a_{ij}(\nabla u)\big(\tfrac{1}{2}F^2(\nabla u)\big)_{ij} &= a_{ij}\frac{\partial}{\partial x_j}\Big(\frac{\partial}{\partial \xi_k}\big(\tfrac{1}{2}F^2\big)(\nabla u) u_{ik}\Big) \\ &= a_{ij}\frac{\partial^2}{\partial \xi_k \partial \xi_l}\big(\tfrac{1}{2}F^2\big)(\nabla u)u_{ik}u_{jl} + a_{ij}\frac{\partial}{\partial \xi_k}\big(\tfrac{1}{2}F^2\big)(\nabla u)u_{ijk} \\ &= a_{ij}a_{kl}u_{ik}u_{jl} + \frac{\partial}{\partial \xi_k}\big(\tfrac{1}{2}F^2\big)(\nabla u)\Big(\frac{\partial}{\partial x_k}(a_{ij}u_{ij}) - \Big(\frac{\partial}{\partial x_k}a_{ij}\Big)u_{ij}\Big). \end{aligned}$$

Taking into account (12) and

$$\frac{\partial}{\partial \xi_k}\big(\tfrac{1}{2}F^2\big)\frac{\partial}{\partial x_k}a_{ij} = a_{ijl}\frac{\partial}{\partial x_l}\big(\tfrac{1}{2}F^2(\nabla u)\big),$$

we get (13). □

When $F(\xi) = |\xi|$, (13) is just the usual Bochner formula

$$\tfrac{1}{2}\Delta(|\nabla u|^2) = |D^2 u|^2 + \langle \nabla u, \nabla(\Delta u)\rangle.$$

We have a Kato type inequality for the square of the "anisotropic" norm of the Hessian.

Lemma 2.2 (Kato inequality). *At a point where* $\nabla u \neq 0$, *we have*

$$a_{ij}a_{kl}u_{ik}u_{jl} \geq a_{ij}F_k F_l u_{ik}u_{jl}. \tag{14}$$

Proof. It is clear that

$$a_{ij}a_{kl}u_{ik}u_{jl} - a_{ij}F_kF_lu_{ik}u_{jl} = a_{ij}FF_{kl}u_{ik}u_{jl} = FF_iF_jF_{kl}u_{ik}u_{jl} + F^2F_{ij}F_{kl}u_{ik}u_{jl}.$$

Since (F_{ij}) is positive definite, we know the first term

$$FF_iF_jF_{kl}u_{ik}u_{jl} = FF_{kl}(F_iu_{ik})(F_ju_{jl}) \geq 0.$$

The second term $F_{ij}F_{kl}u_{ik}u_{jl}$ is nonnegative as well. Indeed, we can write the matrix $(F_{kl})_{k,l} = O^T \Lambda O$ for some orthogonal matrix O and diagonal matrix $\Lambda = \operatorname{diag}(\mu_1, \mu_2, \dots, \mu_n)$ with $\mu_i \geq 0$ for any $i = 1, 2, \dots, n$. Set $U = (u_{ij})_{i,j}$ and $\widetilde{U} = OUO^T = (\tilde{u}_{ij})_{i,j}$. Then we have

$$\begin{aligned} F_{ij}F_{kl}u_{lj}u_{ki} &= \operatorname{tr}(O^T \Lambda OUO^T \Lambda OU) = \operatorname{tr}(\Lambda OUO^T \Lambda OUO^T) \\ &= \operatorname{tr}(\Lambda \widetilde{U} \Lambda \widetilde{U}) = \mu_i \mu_j \tilde{u}_{ij}^2 \geq 0, \end{aligned}$$ □

When $F(\xi) = |\xi|$, (14) is the usual Kato inequality

$$|\nabla^2 u|^2 \geq |\nabla|\nabla u||^2.$$

The following inequality is crucial to apply the gradient comparison argument in Section 3.

Lemma 2.3. *At a point where* $\nabla u \neq 0$, *we have*

$$a_{ij}a_{kl}u_{ik}u_{jl} \geq \frac{(a_{ij}u_{ij})^2}{n} + \frac{n}{n-1}\left(\frac{a_{ij}u_{ij}}{n} - F_iF_ju_{ij}\right)^2 \tag{15}$$

Proof. Let

$$A = F_iF_ju_{ij} \quad \text{and} \quad B = FF_{ij}u_{ij}.$$

The right hand side of (15) equals

$$\frac{(A+B)^2}{n} + \frac{n}{n-1}\left(\frac{B}{n} - \frac{n-1}{n}A\right)^2 = A^2 + \frac{1}{n-1}B^2.$$

The left hand side of (15) is

$$A^2 + 2FF_iF_jF_{kl}u_{ik}u_{jl} + F^2F_{ij}F_{kl}u_{ik}u_{jl}.$$

Since (F_{ij}) is semipositively definite, we know

$$FF_iF_jF_{kl}u_{ik}u_{jl} = FF_{kl}(F_iu_{ik})(F_ju_{jl}) \geq 0.$$

Using the same notations as in the proof of Lemma 2.2, we have

$$F^2F_{ij}F_{kl}u_{ik}u_{jl} = F^2\mu_i\mu_j\tilde{u}_{ij}^2 = F^2\mu_i^2\tilde{u}_{ii}^2 + F^2\sum_{i\neq k}\mu_i\mu_k\tilde{u}_{ik}^2 \geq F^2\mu_i^2\tilde{u}_{ii}^2,$$

$$B = FF_{ij}u_{ij} = \operatorname{tr}(O^T\Lambda OU) = \operatorname{tr}(\Lambda OUO^T) = \mu_i\tilde{u}_{ii}.$$

We claim that (F_{ij}) is a matrix of rank $n-1$, that is, one of μ_i is zero. Firstly, $F_{ij}u_j = 0$. Secondly, for any nonzero $V \perp F_\xi(\nabla u)$, $F_{ij}V^iV^j = a_{ij}V^iV^j > 0$. The claim follows easily. Thus the Hölder inequality gives

$$F^2\mu_i^2\tilde{u}_{ii}^2 \geq \frac{1}{n-1}F^2(\mu_i\tilde{u}_{ii})^2 = \frac{1}{n-1}B^2.$$ □

When $F(\xi) = |\xi|$, (15) is

$$|\nabla^2 u|^2 \geq \frac{(\Delta u)^2}{n} + \frac{n}{n-1}\left(\frac{\Delta u}{n} - \frac{u_i u_j u_{ij}}{|\nabla u|^2}\right)^2.$$

We now recall the definition of F-mean curvature. Let $\Omega \subset \mathbb{R}^n$ be a smooth bounded domain whose boundary $\partial\Omega$ is an $(n-1)$-dimensional, oriented, compact submanifold without boundary in $\mathbb{R}^n$. We denote by ν and $d\sigma$ the outward normal of $\partial\Omega$ and the area element, respectively. Let $\{e_\alpha\}_{\alpha=1}^{n-1}$ be a basis of the tangent space $T_p(\partial\Omega)$, and let $g_{\alpha\beta} = g(e_\alpha, e_\beta)$ and $h_{\alpha\beta}$ be the first and second fundamental forms, respectively. $\partial\Omega$ is called weakly convex if $(h_{\alpha\beta})$ is nonnegative definite. Moreover, let $(g^{\alpha\beta})$ be the inverse matrix of $(g_{\alpha\beta})$ and $\overline{\nabla}$ the covariant derivative in $\mathbb{R}^n$. The F-second fundamental form $h^F_{\alpha\beta}$ and the F-mean curvature H_F are defined by

$$h^F_{\alpha\beta} := \langle F_{\xi\xi} \circ \overline{\nabla}_{e_\alpha} \nu, e_\beta \rangle \quad \text{and} \quad H_F = \sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta} h^F_{\alpha\beta},$$

respectively. We call $\overrightarrow{H_F} = -H_F \nu$ the F-mean curvature vector (it is easy to check that all definitions are independent of the choice of coordinates). $\partial\Omega$ is called weakly F-convex (respectively F-mean convex) if $(h^F_{\alpha\beta})$ is nonnegative definite (respectively $H_F \geq 0$). It is well known that when we consider a variation of $\partial\Omega$ with variation vector field $\varphi \in C_0^\infty(\partial\Omega, \mathbb{R}^n)$, the first variation of the F-area functional $\mathcal{F}(X) := \int_{\partial\Omega} F(\nu) d\sigma$ reads as

$$\delta_\varphi \mathcal{F}(X) = -\int_{\partial\Omega} \langle \overrightarrow{H_F}, \varphi \rangle d\sigma.$$

It is easy to see from the convexity of F that $h^F_{\alpha\beta}$ being nonnegative definite is equivalent to the ordinary second fundamental form $h_{\alpha\beta}$ being nonnegative definite. In other words, there is no difference between weakly F-convex and weakly convex. However, F-mean convex is different from mean convex. For more properties of H_F, we refer to [Wang and Xia 2011b] and the references therein. Here we will use the following lemma, which interprets the relation between the anisotropic Laplacian and the F-mean curvature of level sets of functions.

Lemma 2.4 [Wang and Xia 2011b, Theorem 3]. *Let u be a C^2 function with a regular level set $S_t := \{x \in \overline{\Omega} \mid u = t\}$. Let $H_F(S_t)$ be the F-mean curvature of the level set S_t. We then have*

$$Qu(x) = -F H_F(S_t) + F_i F_j u_{ij} = -F H_F(S_t) + \frac{\partial^2 u}{\partial \nu_F^2}$$

for x with $u(x) = t$, where $\nu_F := F_\xi(\nu) = -F_\xi(\nabla u)$.

We point out that we have used the inward normal in [Wang and Xia 2011b] and there is an sign error in formula (5) there. Hence the term $FH_F(S_t)$ in formula (9) there should be read as $-FH_F(S_t)$.

3. Sharp estimate of the first Neumann eigenvalue

It is well-known that the existence of the first Neumann eigenfunction can be obtained from the direct method in the calculus of variations. We note that the first Neumann eigenfunction must change sign, for its average vanishes.

In this Section we first prove the following gradient comparison theorem, which is the most crucial part for the proof of the sharp estimate. For simplicity, we write λ_1 instead of λ_1^N throughout this section.

Theorem 3.1. *Let Ω, u, λ_1 be as in Theorem 1.1. Let v be a solution of the* 1-*dimensional model problem on some interval* (a, b):

$$v'' - Tv' = -\lambda_1 v, \quad v'(a) = v'(b) = 0, \quad v' > 0 \tag{16}$$

with $T(t) = -(n-1)/t$ or 0. *Assume that* $[\min u, \max u] \subset [\min v, \max v]$. *Then*

$$F(\nabla u)(x) \le v'\big(v^{-1}(u(x))\big). \tag{17}$$

Proof. First, since $\int u = 0$, we know that $\min u < 0$ while $\max u > 0$. We may assume that $[\min u, \max u] \subset (\min v, \max v)$ by multiplying u by a constant $0 < c < 1$. If we prove the result for this u, then, letting $c \to 1$, we have (17).

Under the condition $[\min u, \max u] \subset (\min v, \max v)$, v^{-1} is smooth on a neighborhood U of $[\min u, \max u]$.

Consider $P := \psi(u)(\frac{1}{2}F(\nabla u)^2 - \phi(u))$, where $\psi, \phi \in C^\infty(U)$ are two positive smooth functions to be determined later. We first assume that P attains its maximum at $x_0 \in \Omega$. Then we consider the case where $x_0 \in \partial\Omega$. If $\nabla u(x_0) = 0$, $P \le 0$ is obvious. Hence we assume $\nabla u(x_0) \ne 0$. From now on we compute at x_0. As in Section 2, we use the notation (11). Since x_0 is the maximum of P, we have

$$P_i(x_0) = 0, \tag{18}$$

$$a_{ij}(x_0)P_{ij}(x_0) \le 0. \tag{19}$$

Equality (18) gives

$$\frac{\partial}{\partial x_i}\big(\tfrac{1}{2}F^2(\nabla u) - \phi(u)\big) = -\frac{\psi(u))_i}{\psi^2}P, \quad F_iF_ju_{ij} = \phi' - \frac{\psi'}{\psi^2}P. \tag{20}$$

Then we compute $a_{ij}P_{ij}$.

$$a_{ij}P_{ij} = \frac{P}{\psi}a_{ij}(\psi(u))_{ij} + \psi a_{ij}\frac{\partial}{\partial x_i x_j}\big(\tfrac{1}{2}F^2(\nabla u) - (\phi(u))\big) + 2a_{ij}(\psi(u))_i\frac{\partial}{\partial x_j}\big(\tfrac{1}{2}F^2(\nabla u) - \phi(u)\big).$$

It is easy to see from Proposition 2.1 that

$$\frac{\partial}{\partial \xi_i}\big(\tfrac{1}{2}F^2\big)(\nabla u)u_i = F^2(\nabla u), \quad a_{ij}u_iu_j = F^2(\nabla u), \quad a_{ijk}u_k = 0. \tag{21}$$

By using (20), (21), the Bochner formula (13), and eigenvalue equation (2), we get

$$\begin{aligned}(22)\quad & a_{ij}P_{ij} \\ &= \left(-\lambda_1 u\frac{\psi'}{\psi}+F^2\frac{\psi''}{\psi}-2F^2\frac{\psi'^2}{\psi^2}\right)P+\psi(a_{ij}a_{kl}u_{ik}u_{jl}-\lambda_1F^2)+\psi(\lambda_1u\phi'-F^2\phi'').\end{aligned}$$

Applying Lemma 2.3 to (22), replacing F^2 by $2P/\psi+\phi$, and using (20), (2), and (19), we deduce

$$\begin{aligned}(23)\quad 0 &\geq a_{ij}P_{ij} \\ &\geq \left(-\lambda_1 u\frac{\psi'}{\psi}+F^2\frac{\psi''}{\psi}-2F^2\frac{\psi'^2}{\psi^2}\right)P+\psi(\lambda_1u\phi'-F^2\phi'') \\ &\qquad +\psi\left(\frac{(a_{ij}u_{ij})^2}{n}+\frac{n}{n-1}\left(\frac{a_{ij}u_{ij}}{n}-F_iF_ju_{ij}\right)^2-\lambda_1F^2\right) \\ &= \frac{1}{\psi}\left[2\frac{\psi''}{\psi}-(4-\frac{n}{n-1})\frac{\psi'^2}{\psi^2}\right]P^2 \\ &\quad +\left[2\phi\left(\frac{\psi''}{\psi}-2\frac{\psi'^2}{\psi^2}\right)-\frac{n+1}{n-1}\frac{\psi'}{\psi}\lambda_1u-\frac{2n}{n-1}\frac{\psi'}{\psi}\phi'-2\lambda_1-2\phi''\right]P \\ &\qquad +\psi\left[\frac{1}{n-1}\lambda_1^2u^2+\frac{n+1}{n-1}\lambda_1u\phi'+\frac{n}{n-1}\phi'^2-2\lambda_1\phi-2\phi\phi''\right] \\ &:= a_1P^2+a_2P+a_3.\end{aligned}$$

We are lucky to observe that the coefficients a_i, $i=1,2,3$, coincide with those appearing in the ordinary Laplacian case; see, for example, [Bakry and Qian 2000, Lemma 1]. The next step is to choose suitable positive functions ψ and ϕ such that $a_1, a_2 > 0$ everywhere and $a_3 = 0$, which has already be done in [Bakry and Qian 2000]. For completeness, we sketch the main idea here.

Choose $\phi(u)=\frac{1}{2}v'(v^{-1}(u))^2$, where v is a solution of the 1-dimensional problem (16). One can compute that

$$\phi'(u)=v''(v^{-1}(u)), \quad \phi''(u)=\frac{v'''}{v'}(v^{-1}(u)).$$

Setting $t=v^{-1}(u)$ and $u=v(t)$, we have

$$\begin{aligned}\frac{a_3(t)}{\psi} &= \frac{1}{n-1}\lambda_1^2v^2+\frac{n+1}{n-1}\lambda_1vv''+\frac{n}{n-1}v''^2-\lambda_1v'^2-v'v''' \\ &= -v'(v''-Tv'+\lambda_1v)'+\frac{1}{n-1}(v''-Tv'+\lambda_1v)(nv''+Tv'+\lambda_1v)=0.\end{aligned}$$

Here we have used the fact that T satisfies $T' = T^2/(n-1)$. For a_1, a_2, we introduce

$$X(t) = \lambda_1 \frac{v(t)}{v'(t)}, \quad \psi(u) = \exp\left(\int h(v(t))\right), \quad f(t) = -h(v(t))v'(t).$$

With these notations, we have

$$\begin{aligned} f' &= -h'v'^2 + f(T - X), \\ v'|_{v^{-1}}^2 a_1 \psi &= 2f(T-X) - \frac{n-2}{n-1} f^2 - 2f' := 2(Q_1(f) - f'), \\ a_2 = f\Big(\frac{3n-1}{n-1}T - 2X\Big) &- 2T\Big(\frac{n}{n-1}T - X\Big) - f^2 - f' := Q_2(f) - f'. \end{aligned}$$

We may now use [Bakry and Qian 2000, Corollary 3], which says that there exists a bounded function f on $[\min u, \max u] \subset (\min v, \max v)$ such that $f' < \min\{Q_1(f), Q_2(f)\}$.

In view of (23), we know that, by our choice of ψ and ϕ, $P(x_0) \le 0$, and hence $P(x) \le 0$ for every $x \in \Omega$, which leads to (17).

Now we consider the case $x_0 \in \partial\Omega$. Suppose that P attains its maximum at $x_0 \in \partial\Omega$. We introduce a new vector field $V(x) = (V^i(x))_{i=1}^n$ defined on $\partial\Omega$ by

$$V^i(x) = \sum_{j=1}^n a_{ij}(\nabla u(x))\nu^j(x).$$

Because (a_{ij}) is positive, $V(x)$ must point outward. Hence

$$\frac{\partial P}{\partial V}(x_0) \ge 0.$$

On the other hand, we see, from the Neumann boundary condition and homogeneity of F, that

$$\frac{\partial u}{\partial V}(x_0) = u_i a_{ij}(\nabla u(x))\nu^j = F F_j \nu^j = 0.$$

Thus we have

$$0 \le \frac{\partial P}{\partial V}(x_0) = \psi F F_i u_{ij} a_{jk} \nu^k. \tag{24}$$

We now choose a local coordinate $\{e_i\}_{i=1,\dots,n}$ around x_0 such that $e_n = \nu$ and $\{e_\alpha\}_{\alpha=1,\dots,n-1}$ is the orthonormal basis of the tangent space of $\partial\Omega$. Denote by $h_{\alpha\beta}$ the second fundamental form of $\partial\Omega$. By the assumption that $\partial\Omega$ is weakly convex, we know the matrix $(h_{\alpha\beta}) \ge 0$.

The Neumann boundary condition implies that

$$F_i \nu^i(x_0) = F_n(x_0) = 0. \tag{25}$$

By taking the tangential derivative of (25), we get

$$D_{e_\beta}\left(\sum_{i=1}^{n} F_i \nu^i\right)(x_0) = 0$$

for any $\beta = 1, \ldots, n-1$. Computing $D_{e_\beta}(\sum_{i=1}^{n} F_i \nu^i)(x_0)$ explicitly, we have

$$\begin{aligned} 0 = D_{e_\beta}\left(\sum_{i=1}^{n} F_i \nu^i\right)(x_0) &= \sum_{i,j=1}^{n} F_{ij} u_{j\beta} \nu^i + \sum_{i=1}^{n} F_i \nu^i_\beta \\ &= \sum_{i,j=1}^{n} F_{ij} u_{j\beta} \nu^i + \sum_{i=1}^{n}\sum_{\gamma=1}^{n-1} F_i h_{\beta\gamma} e^i_\gamma \\ &= \sum_{j=1}^{n} F_{nj} u_{j\beta} + \sum_{\gamma=1}^{n-1} F_\gamma h_{\beta\gamma}. \end{aligned} \tag{26}$$

In the last equality, we used $\nu_n = 1$, and $\nu_\beta = 0$ for $\beta = 1, \ldots, n-1$ in the chosen coordinate.

Combining (24), (25), and (26), we obtain

$$\begin{aligned} 0 \le \frac{\partial P}{\partial V}(x_0) &= \sum_{i,j,k=1}^{n} \psi F F_i u_{ij} a_{jk} \nu^k = \psi F \sum_{\alpha=1}^{n-1}\sum_{j=1}^{n} F_\alpha u_{\alpha j} a_{jn} \\ &= \psi F \sum_{\alpha=1}^{n-1}\sum_{j=1}^{n} F_\alpha u_{\alpha j} F_{jn} = -\psi F \sum_{\alpha,\gamma=1}^{n-1} F_\alpha F_\gamma h_{\alpha\gamma} \le 0. \end{aligned}$$

Therefore we obtain that $(\partial P/\partial V)(x_0) = 0$. Since the tangent derivatives of P also vanish, we have $\nabla P(x_0) = 0$. It is also the case that (19) holds. Thus the previous proof for an interior maximum also works in this case. This finishes the proof of Theorem 3.1. □

Following the idea from [Bakry and Qian 2000], besides the gradient comparison with the 1-dimensional models, in order to prove the sharp estimate on the first eigenvalue of the anisotropic Laplacian, we need to study many properties of the 1-dimensional models, such as the difference $\delta(a) = b(a) - a$ as a function of $a \in [0, +\infty]$, where $b(a)$ is the first number for which $v'(b(a)) = 0$ (Note that $v' > 0$ in $(a, b(a))$). As we already saw in Theorem 3.1, the 1-dimensional model (16) appears the same as that in the Laplacian case. Therefore, we can directly use the results of [Bakry and Qian 2000] on the properties of 1-dimensional models. Here we use some simpler statement from [Valtorta 2012].

We define $\delta(a)$ as a function of $a \in [0, +\infty]$ as follows. On the one hand, we denote $\delta(\infty) = \pi/\sqrt{\lambda_1}$. This number comes from the 1-dimensional model (16) with $T = 0$. In fact, it is easy to see that solutions of (16) with $T = 0$ can be explicitly written as

$$v(t) = \sin \sqrt{\lambda_1} t$$

up to dilations. Hence in this case, $b(a) - a = \pi/\sqrt{\lambda_1}$ for any $a \in \mathbb{R}$. On the other hand, we denote $\delta(a) = b(a) - a$ as a function of $a \in [0, +\infty)$ relative to the 1-dimensional model (16) with $T = -(n-1)/x$.

We have the following property of $\delta(a)$.

Lemma 3.1 [Bakry and Qian 2000; Valtorta 2012, Theorem 5.3, Corollary 5.4]. *The function* $\delta(a) : [0, \infty] \to \mathbb{R}^+$ *is a continuous function such that*

$$\delta(a) > \frac{\pi}{\sqrt{\lambda_1}} \quad \textit{and} \quad \delta(\infty) = \frac{\pi}{\sqrt{\lambda_1}}.$$

$m(a) := v(b(a)) < 1$, $\lim_{a\to\infty} m(a) = 1$, *and* $m(a) = 1$ *if and only if* $a = \infty$.

In order to prove the main result, we also need the following comparison theorem on the maximum values of eigenfunctions. This theorem is obtained as a consequence of a standard property of the volume of small balls with respect to some invariant measure; see [Bakry and Qian 2000, Section 6].

Lemma 3.2. *Let* Ω, u, λ_1 *be as in Theorem 1.1. Let* v *be a solution of the* 1-*dimensional model problem on some interval* $(0, \infty)$:

$$v'' = -\frac{n-1}{t} v' - \lambda_1 v, \quad v(0) = -1, \quad v'(0) = 0.$$

Let b *be the first number after* 0 *with* $v'(b) = 0$ *and denote* $m = v(b)$. *Then* $\max u \geq m$.

The proof of Lemma 3.2 is similar to that of [Bakry and Qian 2000, Theorem 11]. The essential part is the gradient comparison theorem 3.1. We omit it here.

Now we are in a position to prove Theorem 1.1.

Proof of Theorem 1.1. Let u be an eigenfunction with eigenvalue λ_1. Since $\int u = 0$, we may assume $\min u = -1$ and $0 \leq k = \max u \leq 1$. Given a solution v to (16), denote $m(a) = v(b(a))$ with $b(a)$ the first number with $v'(b(a)) = 0$ after a.

Lemmas 3.1 and 3.2 imply that for any eigenfunction u, there exists a solution v to (16) such that $\min v = \min u = -1$ and $\max v = \max u = k \leq 1$.

We now get the expected estimate by using Theorem 3.1. Choosing $x_1, x_2 \in \overline{\Omega}$ with $u(x_1) = \min u = -1$, $u(x_2) = \max u = k$ and $\gamma(t) : [0, 1] \to \overline{\Omega}$ the minimal geodesic from x_1 to x_2. Consider the subset I of [0,1] such that $(d/dt)u(\gamma(t)) \geq 0$.

By the gradient comparison estimate (17) and Lemma 3.1, we have

$$\begin{aligned} d_F &\geq \int_0^1 F^0(\dot\gamma(t))\,dt \geq \int_I F^0(\dot\gamma(t))\,dt \\ &\geq \int_0^1 \frac{1}{F(\nabla u)}\langle \nabla u, \dot\gamma(t)\rangle\,dt = \int_{-1}^k \frac{1}{F(\nabla u)}\,du \\ &\geq \int_{-1}^k \frac{1}{v'(v^{-1}(u))}\,du = \int_a^{b(a)} dt = \delta(a) \geq \frac{\pi}{\sqrt{\lambda_1}}, \end{aligned}$$

which leads to

$$\lambda_1 \geq \frac{\pi^2}{d_F^2}.$$

It remains to prove the equality case. The idea of the proof follows from [Hang and Wang 2007]. Here we must pay more attention to the points with vanishing ∇u.

Assume that $\lambda_1 = \pi^2/d_F^2$. It can be easily seen from the proof of Theorem 1.1 that $a = \infty$, which leads to $\max u = \max v = 1$ by Lemma 3.1. We will prove that Ω is in fact a segment in $\mathbb{R}$. We divide the proof into several steps.

Step 1. $S := \{x \in \overline{\Omega} \mid u(x) = \pm 1\} \subset \partial\Omega$.

Let $\mathcal{P} = F(\nabla u)^2 + \lambda_1 u^2$. After a simple calculation using the Bochner formula (13) and the Kato inequality (14), we obtain

$$\begin{aligned} \tfrac{1}{2} a_{ij}\mathcal{P}_{ij} &= a_{ij}a_{kl}u_{ik}u_{jl} - \tfrac{1}{2}a_{ijl}u_{ij}\mathcal{P}_l - \lambda_1^2 u^2 \\ &\geq a_{ij}F_kF_l u_{ik}u_{jl} - \tfrac{1}{2}a_{ijl}u_{ij}\mathcal{P}_l - \lambda_1^2 u^2 \\ &= -\tfrac{1}{2}a_{ijl}u_{ij}\mathcal{P}_l + \frac{1}{4F^2}(a_{ij}\mathcal{P}_i\mathcal{P}_j - 4\lambda_1 u u_i \mathcal{P}_i) \end{aligned}$$

on $\Omega \setminus \mathcal{C}$. Namely,

$$\tfrac{1}{2} a_{ij}\mathcal{P}_{ij} + b_i\mathcal{P}_i \geq 0 \tag{27}$$

on $\Omega \setminus \mathcal{C}$ for some $b_i \in C^0(\Omega)$. If $\mathcal{P}$ attains its maximum on $x_0 \in \partial\Omega$, then arguing as in Theorem 3.1, we have $\nabla\mathcal{P}(x_0) = 0$. However, from the Hopf Theorem, $\nabla\mathcal{P}(x_0) \neq 0$, a contradiction. Hence $\mathcal{P}$ attains its maximum at $\mathcal{C}$, and therefore,

$$\mathcal{P} \leq \lambda_1. \tag{28}$$

Take any two points $x_1, x_2 \in S$ with $u(x_1) = -1$, $u(x_2) = 1$. Let

$$\gamma(t) = \Big(1 - \frac{t}{F^0(x_2 - x_1)}\Big)x_1 + \frac{t}{F^0(x_2 - x_1)}x_2 : [0, l] \to \overline{\Omega}$$

be the straight line from x_1 to x_2, where $l := F^0(x_2 - x_1)$ is the distance from x_1 to x_2 with respect to F. Denote $f(t) := u(\gamma(t))$. It is easy to see $F^0(\dot\gamma(t)) = 1$. It follows from (28) and the Cauchy–Schwarz inequality (7) that

$$|f'(t)| = |\nabla u(\gamma(t)) \cdot \dot{\gamma}(t)| \le F(\nabla u)(\gamma(t)) \le \sqrt{\lambda_1(1-f(t)^2)}. \tag{29}$$

Here we have used the Cauchy–Schwarz inequality (7) again. Hence

$$\begin{aligned} d_F \ge l \ge \int_{\{0\le t\le l,\, f'(t)>0\}} dt &\ge \int_0^l \frac{1}{\sqrt{\lambda_1}} \frac{f'(t)}{\sqrt{1-f(t)^2}}\, dt \\ &= \frac{1}{\sqrt{\lambda_1}} \int_{-1}^1 \frac{1}{\sqrt{1-x^2}}\, dx = \frac{\pi}{\sqrt{\lambda_1}}. \end{aligned} \tag{30}$$

Since $d_F = \pi/\sqrt{\lambda_1}$, we must have $d_F = l$, which means $S \subset \partial\Omega$.

Step 2. $\mathcal{P} = F^2(\nabla u) + \lambda_1 u^2 \equiv \lambda_1$ in $\overline{\Omega}$. Hence $S \equiv \mathcal{C}$.

From Step 1, we know that $\Omega^* := \overline{\Omega} \setminus S$ is connected. Let $E := \{x \in \Omega^* : \mathcal{P} = \lambda_1\}$. It is clear that E is closed. In view of (27), thanks to the strong maximum principle, we know that E is also open. We now show that E is nonempty. Indeed, from the fact that all inequalities in (29) and (30) are equalities, we obtain $f(t) = u(\gamma(t)) = -\cos\sqrt{\lambda_1}t$ for $t \in (0, l)$. Hence

$$\mathcal{P}(\gamma(t)) = f'(t)^2 + \lambda_1 f(t)^2 = \lambda_1.$$

Thus E is nonempty, open, and closed in Ω^*. Therefore, we obtain $\mathcal{P} \equiv \lambda_1$ in $\overline{\Omega}$ (for $x \in S$, $\mathcal{P} = \lambda_1$ is obvious).

Step 3: Define $X = \nabla u / F(\nabla u)$ in Ω^* and X^* the cotangent vector given by $X^*(Y) = \langle X, Y\rangle$ for any tangent vector Y. Then, in Ω^*, we claim that

$$D^2 u = -\lambda_1 u X^* \otimes X^*, \tag{31}$$

and, moreover, $X = \vec{c}$ for some constant vector $\vec{c}$.

First, taking the derivative of $F^2(\nabla u) + \lambda_1 u^2 \equiv \lambda_1$ gives

$$F_i F_j u_{ij} = -\lambda_1 u. \tag{32}$$

On the other hand, since $\mathcal{P} \equiv \lambda_1$, the proof of (27) leads to

$$a_{ij} a_{kl} u_{ik} u_{jl} = \lambda_1^2 u^2 = (F_i F_j u_{ij})^2. \tag{33}$$

Equation (33) in fact gives that

$$F_{ij} F_{kl} u_{ik} u_{jl} = 0. \tag{34}$$

Set $X^\perp := \{V \in \mathbb{R}^n \mid V \perp X\}$. $X^\perp$ is an $(n-1)$-dimensional vector subspace. Note that (F_{ij}) is exactly a matrix of rank $n-1$ (see the proof of Lemma 2.3) and $F_{ij} X^j = 0$. It follows from this fact and (34) that

$$u_{ij} V^i V^j = 0 \quad \text{for any } V \in X^\perp. \tag{35}$$

Equations (32) and (35) imply (31), which in turn implies

$$u_{ij} = \frac{-\lambda_1 u u_i u_j}{F^2(\nabla u)}. \tag{36}$$

By differentiating X, we obtain from (36) that

$$\nabla_i X^j = \frac{u_{ij}}{F(\nabla u)} - \frac{u_j}{F^2(\nabla u)} F_k u_{ki} = 0.$$

Thus $X = \vec{c}$ in Ω^*.

Step 4: The maximum point and the minimum point are unique.

We already knew that $f(t) = u(\gamma(t)) = -\cos\sqrt{\lambda_1}t$ and $\nabla u(\gamma(t)) \neq 0$ for $t \in (0, l)$. Hence u is C^2 along $\gamma(t)$ for $t \in (0, l)$, and it follows that

$$D^2 u(\dot\gamma(t), \dot\gamma(t))\,|_{\gamma(t)} = \lambda_1 \cos t \quad \text{for any } t \in (0, l). \tag{37}$$

On the other hand, we deduce from (31) that

$$D^2 u(\dot\gamma(t), \dot\gamma(t))\,|_{\gamma(t)} = -\lambda_1 u(\gamma(t)) \langle X, \dot\gamma(t)\rangle^2. \tag{38}$$

Combining (37) and (38), and taking $t \to 0$, we get

$$|\langle X, \dot\gamma(t)\rangle| = 1 = F(X) F^0(\dot\gamma(t)),$$

which means equality in the Cauchy–Schwarz inequality (7) holds. Hence $X = \pm F^0_\xi(\dot\gamma(t))$. Noting that $\dot\gamma(t) = x_2 - x_1/F^0(x_2 - x_1)$, we have

$$X = F^0_\xi(x_2 - x_1).$$

Suppose there is some point x_3 with $u(x_3) = 1$. Using the same argument, we obtain $X = F^0_\xi(x_3 - x_1)$. In view of $F^0(x_3 - x_1) = F^0(x_2 - x_1)$, we conclude that $x_3 = x_2$. Therefore, there is only one maximum point and only one minimum point.

Step 5: $n = 1$ and Ω is a segment.

From Step 4, we have $\nabla u \neq 0$ for most points of $\partial\Omega$, and at these points $X = \nabla u / F(\nabla u)$ lies in the tangent spaces due to the Neumann boundary condition. This is impossible unless $n = 1$, because X is a constant vector. This completes the proof. □

4. Estimate of the first Dirichlet eigenvalue

As in Section 3, for simplicity, we write λ_1 instead of λ_1^D throughout this section.

It is well-known that the existence of first Dirichlet eigenfunction can be easily proved by using the direct method in the calculus of variations. Moreover, by the assumption that F is even, the first Dirichlet eigenfunction u does not change sign; see [Belloni et al. 2003, Theorem 3.1]. We may assume u is nonnegative. By multiplying u by a constant, we can also assume that $\sup_\Omega u = 1$ and $\inf_\Omega u = 0$ without loss of generality.

For any $\alpha, \beta \in \mathbb{R}$ with $\alpha > 0$, $\beta^2 > \sup(\alpha+u)^2$, consider the function

$$P(x) = \frac{F^2(\nabla u)}{2(\beta^2 - (\alpha+u)^2)}.$$

Suppose that $P(x)$ attains its maximum at $x_0 \in \overline{\Omega}$.

With the assumption that Ω is F-mean convex, we first exclude the possibility $x_0 \in \partial\Omega$ with $\nabla u(x_0) \neq 0$. Indeed, suppose we have $x_0 \in \partial\Omega$ with $\nabla u(x_0) \neq 0$. Define

$$\nu_F := F_\xi(\nu)$$

on $\partial\Omega = \{x \in \overline{\Omega} \mid u(x) = 0\}$. In view of $\langle \nu_F, \nu \rangle = F(\nu) > 0$, ν_F must point outward. From the Dirichlet boundary condition, we know that

$$\nu = -\nabla u / |\nabla u|$$

for $\nabla u \neq 0$. Hence $\nu_F = -F_\xi(\nabla u)$. Since P attains its maximum at x_0, we have

$$0 \le \frac{\partial P}{\partial \nu_F}(x_0) = \frac{F F_i u_{ij} \nu_F^j}{\beta^2 - (\alpha+u)^2} + F^2 \frac{\alpha(\partial u / \partial \nu_F)}{(\beta^2 - (\alpha+u)^2)^2}$$

Hence

$$-\frac{\partial^2 u}{\partial \nu_F^2} + \frac{F\alpha(\partial u/\partial \nu_F)}{\beta^2 - \alpha^2} \ge 0.$$

Note that $\partial u / \partial \nu_F = -F(\nabla u)$. Since $\partial\Omega$ itself is a level set of u, we can apply Lemma 2.4 to obtain

$$\frac{\partial^2 u}{\partial \nu_F^2} = Qu + F H_F.$$

In view of $Qu(x_0) = -\lambda_1 u(x_0) = 0$, we obtain that

$$-F H_F - F^2 \frac{\alpha}{\beta^2 - \alpha^2} \ge 0.$$

This contradicts the fact that $H_F(\partial\Omega) \ge 0$.

On the other hand, if $\nabla u(x_0) = 0$, $F(\nabla u)(x_0) = 0$ and $P(x_0) = 0$, which implies $F(\nabla u) = 0$, that is, u is constant, a contradiction.

Therefore we may assume $x_0 \in \Omega$ and $\nabla u(x_0) \neq 0$. Since a_{ij} is positively definite on $\overline{\Omega} \setminus \mathscr{C}$, where $\mathscr{C} := \{x \mid \nabla u(x) = 0\}$, it follows from the maximum principle that

$$P_i(x_0) = 0, \tag{39}$$

$$a_{ij}(x_0) P_{ij}(x_0) \le 0. \tag{40}$$

From now on we will compute at the point x_0. Equality (39) gives

$$\frac{\partial}{\partial x_i}\left(\tfrac{1}{2}F^2(\nabla u)\right) = -\frac{F^2(\nabla u)(\alpha+u)u_i}{\beta^2-(\alpha+u)^2}. \tag{41}$$

Then we compute $a_{ij}(x_0)P_{ij}(x_0)$.

$$\begin{aligned}
a_{ij}(x_0)P_{ij}(x_0) = {} & \frac{1}{\beta^2-(\alpha+u)^2}a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}\left(\tfrac{1}{2}F^2(\nabla u)\right) \\
& + 2a_{ij}\frac{\partial}{\partial x_i}\left(\tfrac{1}{2}F^2(\nabla u)\right)\frac{\partial}{\partial x_j}\left(\frac{1}{\beta^2-(\alpha+u)^2}\right) \\
& + a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}\left(\frac{1}{\beta^2-(\alpha+u)^2}\right)\tfrac{1}{2}F^2(\nabla u) \\
= {} & I + II + III.
\end{aligned}$$

By using (41), (21), the Bochner formula (13), and Equation (2), we obtain

$$I = \frac{1}{\beta^2-(\alpha+u)^2}[a_{ij}a_{kl}u_{ik}u_{jl} - \lambda_1 F^2], \tag{42}$$

$$II = -\frac{4F^4(\alpha+u)^2}{(\beta^2-(\alpha+u)^2)^3}, \tag{43}$$

$$III = \frac{F^4}{(\beta^2-(\alpha+u)^2)^2} + \frac{4F^4(\alpha+u)^2}{(\beta^2-(\alpha+u)^2)^3} - \frac{\lambda_1 F^2 u(\alpha+u)}{(\beta^2-(\alpha+u)^2)^2}. \tag{44}$$

We now apply Lemma 2.2 to (42) and obtain

$$\begin{aligned}
a_{ij}a_{kl}u_{ik}u_{jl} &\geq a_{ij}F_kF_lu_{ik}u_{jl} \\
&= \frac{1}{F^2}a_{ij}\frac{\partial}{\partial x_i}\left(\tfrac{1}{2}F^2(\nabla u)\right)\frac{\partial}{\partial x_j}\left(\tfrac{1}{2}F^2(\nabla u)\right) \\
&= \frac{F^4(\alpha+u)^2}{(\beta^2-(\alpha+u)^2)^2}.
\end{aligned}$$

Here we have used (41) and (21) again in the last equality. Therefore, we have

$$I \geq \frac{F^4(\alpha+u)^2}{(\beta^2-(\alpha+u)^2)^3} - \frac{\lambda_1 F^2}{\beta^2-(\alpha+u)^2}. \tag{45}$$

Combining (40), (43), (44), and (45), we obtain

$$0 \geq a_{ij}P_{ij} \geq \frac{F^4\beta^2}{(\beta^2-(\alpha+u)^2)^3} - \frac{\lambda_1 F^2}{\beta^2-(\alpha+u)^2} - \frac{\lambda_1 F^2 u(\alpha+u)}{(\beta^2-(\alpha+u)^2)^2}.$$

It follows that

$$\frac{F^2(\nabla u)}{\beta^2-(\alpha+u)^2}(x_0) \leq \frac{\lambda_1}{\beta^2}(\beta^2-\alpha(\alpha+u)). \tag{46}$$

Noting that $\sup_\Omega u = 1$, we choose $\alpha > 0$ and $\beta = \alpha + 1$. Then estimate (46) becomes

$$\frac{F^2(\nabla u)}{(\alpha+1)^2-(\alpha+u)^2}(x_0) \le \lambda_1\Big(1-\frac{\alpha(\alpha+u)}{(\alpha+1)^2}\Big) \le \lambda_1.$$

Hence we conclude that

$$\frac{F^2(\nabla u)}{(\alpha+1)^2-(\alpha+u)^2} \le \lambda_1. \tag{47}$$

for any $x \in \overline{\Omega}$.

Choose $x_1 \in \Omega$ with $u(x_1) = \sup u = 1$ and $x_2 \in \partial\Omega$ with

$$d_F(x_1, x_2) = d_F(x_1, \partial\Omega) \le i_F$$

and $\gamma(t) : [0, 1] \to \overline{\Omega}$ the minimal geodesic connecting x_1 with x_2. Using the gradient estimates (47), we have

$$\begin{aligned}\frac{\pi}{2} - \arcsin\Big(\frac{\alpha}{\alpha+1}\Big) &= \int_0^1 \frac{1}{\sqrt{(\alpha+1)^2-(\alpha+u)^2}}\,du \le \sqrt{\lambda_1}\int_0^1 \frac{1}{F(\nabla u)}\,du\\ &\le \sqrt{\lambda_1}\int_0^1 \frac{1}{F(\nabla u(\gamma(t)))}\langle \nabla u(\gamma(t)), \dot\gamma(t)\rangle\,dt\\ &\le \sqrt{\lambda_1}\int_0^1 F^0(\dot\gamma(t))\,dt \le \sqrt{\lambda_1}\,i_F.\end{aligned}$$

Here we have used the Cauchy–Schwarz inequality (7). Letting $\alpha \to 0$, we obtain

$$\lambda_1 \ge \frac{\pi^2}{4i_F^2}.$$

Thus we finish the proof of Theorem 1.2.

References

[Alvino et al. 1997] A. Alvino, V. Ferone, G. Trombetti, and P.-L. Lions, "Convex symmetrization and applications", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **14**:2 (1997), 275–293. MR 98d:35068 Zbl 0877.35040

[Amar and Bellettini 1994] M. Amar and G. Bellettini, "A notion of total variation depending on a metric with discontinuous coefficients", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **11**:1 (1994), 91–133. MR 97a:49057 Zbl 0842.49016

[Bakry and Qian 2000] D. Bakry and Z. Qian, "Some new results on eigenvectors via dimension, diameter, and Ricci curvature", *Adv. Math.* **155**:1 (2000), 98–153. MR 2002g:58048 Zbl 0980.58020

[Bebendorf 2003] M. Bebendorf, "A note on the Poincaré inequality for convex domains", *Z. Anal. Anwendungen* **22**:4 (2003), 751–756. MR 2004k:26025 Zbl 1057.26011

[Belloni et al. 2003] M. Belloni, V. Ferone, and B. Kawohl, "Isoperimetric inequalities, Wulff shape and related questions for strongly nonlinear elliptic operators", *Z. Angew. Math. Phys.* **54**:5 (2003), 771–783. MR 2005e:35177 Zbl 1099.35509

[Chen and Wang 1997] M. Chen and F. Wang, "General formula for lower bound of the first eigenvalue on Riemannian manifolds", *Sci. China Ser. A* **40**:4 (1997), 384–394. MR 98d:58189 Zbl 0895.58056

[Ferone and Kawohl 2009] V. Ferone and B. Kawohl, "Remarks on a Finsler–Laplacian", *Proc. Amer. Math. Soc.* **137**:1 (2009), 247–253. MR 2009m:35135 Zbl 1161.35017

[Ge and Shen 2001] Y. Ge and Z. Shen, "Eigenvalues and eigenfunctions of metric measure manifolds", *Proc. London Math. Soc.* (3) **82**:3 (2001), 725–746. MR 2002g:58059 Zbl 1018.58005

[Hang and Wang 2007] F. Hang and X. Wang, "A remark on Zhong–Yang's eigenvalue estimate", *Int. Math. Res. Not.* **2007**:18 (2007), Art. ID rnm064. MR 2008m:53083 Zbl 1129.53021

[Kawohl 2011] B. Kawohl, "Variations on the p-Laplacian", pp. 35–46 in *Nonlinear elliptic partial differential equations*, edited by D. Bonheure et al., Contemp. Math. **540**, Amer. Math. Soc., Providence, RI, 2011. MR 2807407 Zbl 05954310

[Kröger 1992] P. Kröger, "On the spectral gap for compact manifolds", *J. Differential Geom.* **36**:2 (1992), 315–330. MR 94g:58236 Zbl 0738.58048

[Li 1979] P. Li, "A lower bound for the first eigenvalue of the Laplacian on a compact manifold", *Indiana Univ. Math. J.* **28**:6 (1979), 1013–1019. MR 81j:58084 Zbl 0429.35054

[Li and Yau 1980] P. Li and S. T. Yau, "Estimates of eigenvalues of a compact Riemannian manifold", pp. 205–239 in *Geometry of the Laplace operator* (Honolulu, HI, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, RI, 1980. MR 81i:58050 Zbl 0441.58014

[Ohta 2009] S.-I. Ohta, "Finsler interpolation inequalities", *Calc. Var. Partial Differential Equations* **36**:2 (2009), 211–249. MR 2011m:58027 Zbl 1175.49044

[Ohta and Sturm 2011] S.-I. Ohta and K.-T. Sturm, "Bochner–Weitzenboeck formula and Li–Yau estimates on Finsler manifolds", preprint, 2011. arXiv 1105.0983

[Payne and Weinberger 1960] L. E. Payne and H. F. Weinberger, "An optimal Poincaré inequality for convex domains", *Arch. Rational Mech. Anal.* **5** (1960), 286–292. MR 22 #8198 Zbl 0099.08402

[Qian et al. 2012] Z. Qian, H.-C. Zhang, and X.-P. Zhu, "Sharp spectral gap and Li–Yau's estimate on Alexandrov spaces", preprint, 2012. To appear in *Math. Z.* arXiv 1102.4159

[Shen 2001] Z. Shen, *Lectures on Finsler geometry*, World Scientific, Singapore, 2001. MR 2002f:53032 Zbl 0974.53002

[Tolksdorf 1984] P. Tolksdorf, "Regularity for a more general class of quasilinear elliptic equations", *J. Differential Equations* **51**:1 (1984), 126–150. MR 85g:35047 Zbl 0488.35017

[Valtorta 2012] D. Valtorta, "Sharp estimate on the first eigenvalue of the p-Laplacian", *Nonlinear Anal.* **75**:13 (2012), 4974–4994. MR 2927560 arXiv 1102.0539

[Wang and Xia 2011a] G. Wang and C. Xia, "A Brunn–Minkowski inequality for a Finsler–Laplacian", *Analysis* (*Munich*) **31**:2 (2011), 103–115. MR 2012g:52013 Zbl 1235.52017

[Wang and Xia 2011b] G. Wang and C. Xia, "A characterization of the Wulff shape by an overdetermined anisotropic PDE", *Arch. Rational Mech. Anal.* **199**:1 (2011), 99–115. MR 2012c:35303 Zbl 1232.35103

[Wang and Xia 2012] G. Wang and C. Xia, "Blow-up analysis of a Finsler–Liouville equation in two dimensions", *J. Differential Equations* **252**:2 (2012), 1668–1700. MR 2853556 Zbl 1233.35053

[Zhong and Yang 1984] J. Q. Zhong and H. C. Yang, "On the estimate of the first eigenvalue of a compact Riemannian manifold", *Sci. Sinica Ser. A* **27**:12 (1984), 1265–1273. MR 87a:58162 Zbl 0561.53046

Received October 4, 2011. Revised January 26, 2012.

GUOFANG WANG
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
MATHEMATISCHES INSTITUT
ECKERSTRASSE 1
D-79104 FREIBURG
GERMANY
guofang.wang@math.uni-freiburg.de

CHAO XIA
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
MATHEMATISCHES INSTITUT
ECKERSTRASSE 1
D-79104 FREIBURG
GERMANY
chao.xia@math.uni-freiburg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.327

EINSTEIN METRICS AND EXOTIC SMOOTH STRUCTURES

MASASHI ISHIDA

We prove new existence theorems of 4-manifolds admitting infinitely many distinct smooth structures for which no Einstein metric exists.

1. Introduction

A Riemannian metric g is called Einstein if its Ricci curvature, considered as a function on the unit tangent bundle, is constant. It is known that any closed oriented Einstein 4-manifold X satisfies

$$2\chi(X) \geq 3|\tau(X)|, \tag{1}$$

where $\chi(X)$ and $\tau(X)$ denote respectively the Euler characteristic and signature of X. This is called the Hitchin–Thorpe inequality [Hitchin 1974; Thorpe 1969; Besse 1987]. Hitchin [1974] proved that any closed oriented Einstein 4-manifold satisfying $2\chi(X) = 3|\tau(X)|$ is finitely covered by either a K3 surface or the 4-torus.

On the other hand, by using Seiberg–Witten invariants [Witten 1994], LeBrun [1996] constructed the first example of a simply connected closed 4-manifold X without Einstein metrics that nonetheless satisfies the strict Hitchin–Thorpe inequality $2\chi(X) > 3|\tau(X)|$. It is now well-known [LeBrun 1995a; 1995b; 2001; 2009] that the existence of monopole classes (see Definition 2 below) gives rise to obstructions to the existence of Einstein metrics on 4-manifolds. In particular, any Einstein 4-manifold X with a nonzero special monopole class (see Section 2 below) must satisfy the inequality

$$\chi(X) \geq 3\tau(X). \tag{2}$$

This equality occurs only if X is a compact quotient of the complex hyperbolic plane equipped with a constant multiple of its standard Kähler–Einstein metric. For Kähler surfaces, this inequality reduces to the celebrated Miyaoka–Yau inequality. We shall call (2) the Miyaoka–Yau–LeBrun inequality. Moreover, an obstruction found in [LeBrun 1996; 2001] provided the first means of exhibiting

This work is supported in part by the Grant-in-Aid for Scientific Research (C), Japan Society for the Promotion of Science, number 20540090.
MSC2010: primary 53C25; secondary 57R57, 57R55.
Keywords: Einstein metrics, smooth structures, Seiberg–Witten invariants.

the dependence of Einstein metrics on smooth structures of underlying topological 4-manifolds [Kotschick 1998]. In particular, we know that there exist infinitely many topological 4-manifolds which often admit infinitely many smooth structures for which Einstein metrics do not exist, but nevertheless satisfy the Hitchin–Thorpe inequality. For instance, see [LeBrun 2001; 2003; Ishida and LeBrun 2002; 2003; Brunnbauer et al. 2009].

In this article, we shall prove the following general existence theorem of 4-manifolds without Einstein metrics, which nicely highlights how much there is to be said about the subject beyond the Hitchin–Thorpe inequality (1) and the Miyaoka–Yau–LeBrun inequality (2).

Theorem A. *For any pair of integers* (m, n) *satisfying* $m + n \equiv 0 \pmod 2$, *there exist infinitely many nonhomeomorphic topological nonspin* 4*-manifolds with*

$$(\chi, \tau) = (m, n),$$

and all such topological nonspin 4*-manifolds admit infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist. In particular*:

(1) *If* $2m > 3|n|$, *there are nonspin* 4*-manifolds admitting infinitely many distinct smooth structures for which no Einstein metric exists*, *but that nevertheless satisfy the strict Hitchin–Thorpe inequality.*

(2) *If* $m > 3n$, *there are nonspin* 4*-manifolds admitting infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist*, *but nevertheless satisfy the strict Miyaoka–Yau–LeBrun inequality.*

Notice that any closed 4-manifold X always satisfies $\chi(X)+\tau(X) \equiv 0 \pmod 2$. Therefore, $m+n \equiv 0 \pmod 2$ in the above theorem is the best possible. Theorem A follows from Theorem 11 proved in Section 4. Theorem 11 provides us a way to construct new examples of 4-manifolds without Einstein metrics. See also Remark 12 at the end of Section 4.

One of motivations for Theorem A comes from an interesting result due to Sambusetti [1998]. Using a remarkable inequality of Besson, Courtois, and Gallot [Besson et al. 1995] concerning the volume entropy, Sambusetti [1998] proved a topological obstruction to the existence of Einstein metrics on a 4-manifold admitting a nonzero degree map onto compact real or complex hyperbolic 4-manifolds. By applying the obstruction, Sambusetti proved an interesting existence result for 4-manifolds without Einstein metrics.

Theorem 1 [Sambusetti 1998, Theorem 4.4]. *Any pair of integers* (m, n) *satisfying* $m + n \equiv 0 \pmod 2$ *can be realized as the Euler characteristic* χ *and signature* τ

of infinitely many nonhomeomorphic closed smooth 4*-manifolds without Einstein metrics.*

We notice that Sambusetti's obstruction actually depends only on the homotopy type of the manifold, and therefore in principle applies to all smooth structures. However, it appears to be unknown if most of the examples considered by Sambusetti actually admit exotic smooth structures. For instance, Theorem 1 is proved by considering the following connected sum $Y^p_{m,n}$ [Sambusetti 1998, Remarks 4.5]:

$$(|m|+|n|+p)M\,\#\,(|m|+|n|-n+p)\overline{\mathbb{C}P^2}\,\#\,\Big(|m|+|n|-\frac{m+n}{2}+1+p\Big)Y,$$

where p is any nonnegative integer, M is a Mumford fake projective plane [1979] and $Y := S^2\times T^2$. We also remark that there is a vanishing theorem of Witten [1994] which asserts that all the Seiberg–Witten invariants of a connected sum $X_1\#X_2$ of 4-manifolds with $b_2^+(X_1)\geq 1$ and $b_2^+(X_2)\geq 1$ vanish, where $b^+(X)$ is the dimension of a maximal linear subspace of $H^2(X,\mathbb{R})$ on which the cup product pairing is positive definite. Hence, all the Seiberg–Witten invariants of the connected sum $Y^p_{m,n}$ vanish in general. At least, the present author does not know how to detect the existence or nonexistence of monopole classes of $Y^p_{m,n}$, and, to the best of our knowledge, it is also unknown whether the underlying topological manifold of $Y^p_{m,n}$ admits infinitely many smooth structures for which no Einstein metric exists. Hence, Theorem A can be seen as a natural generalization of Sambusetti's result and actually contains several new aspects which were not covered by Sambusetti's result. Moreover, our method of proof is totally different from that of Theorem 1. In particular, we use the Seiberg–Witten monopole equations [Witten 1994] to prove Theorem A.

We mention that a Seiberg–Witten refinement of Theorem 1 was first proved by Del Rio Guerra [2002, Theorem D], who showed the existence of non-Einstein 4-manifolds with free fundamental group. Our Theorem A can be seen as a natural generalization of that result, because our method of proof implies the existence of topological 4-manifolds with free fundamental group and admitting infinitely many distinct smooth structures for which Einstein metrics do not exist. (See also Remark 12 below.)

Theorem A is a result on the nonspin case. The second main result of the present article tells us that a similar result still holds in the spin case:

Theorem B. *For any pair of integers* (m,n) *satisfying* $m+n\equiv 0 \pmod 2$ *and* $n\equiv 0 \pmod{16}$, *there exist infinitely many nonhomeomorphic topological spin* 4*-manifolds with* $(\chi,\tau)=(m,n)$ *and all such topological spin* 4*-manifolds admit infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist. In particular*:

(1) *If* $2m > 3|n|$, *there are spin* 4-*manifolds admitting infinitely many distinct smooth structures for which no Einstein metric exists*, *but nevertheless satisfy the strict Hitchin–Thorpe inequality.*

(2) *If* $m > 3n$, *there are spin* 4-*manifolds admitting infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist*, *but nevertheless satisfy the strict Miyaoka–Yau–LeBrun inequality.*

By Rohlin's theorem, any spin 4-manifold X must satisfy $\tau(X) \equiv 0 \pmod{16}$. Therefore, we cannot remove the condition $n \equiv 0 \pmod{16}$ from Theorem B.

2. Obstruction to the existence of Einstein metrics

By using several nice results proved in [LeBrun 2009], we shall prove an obstruction to the existence of Einstein metrics on 4-manifold; see Theorem 7 below. We shall use the obstruction to prove the main results.

Let X be a closed oriented Riemannian 4-manifold with $b^+(X) \geq 2$. Recall that a spin^c-structure Γ_X on a smooth Riemannian 4-manifold X induces a pair of spinor bundles $S^{\pm}_{\Gamma_X}$ which are Hermitian vector bundles of rank 2 over X. A Riemannian metric on X and a unitary connection A on the determinant line bundle $\mathscr{L}_{\Gamma_X} := \det(S^+_{\Gamma_X})$ induce the twisted Dirac operator $\mathfrak{D}_A : \Gamma(S^+_{\Gamma_X}) \to \Gamma(S^-_{\Gamma_X})$. Seiberg–Witten monopole equations over X are the following nonlinear partial differential equations for a unitary connection A of the complex line bundle $\mathscr{L}_{\Gamma_X}$ and a spinor $\phi \in \Gamma(S^+_{\Gamma_X})$:

$$\mathfrak{D}_A\phi = 0, \quad F^+_A = iq(\phi), \tag{3}$$

where F^+_A is the self-dual part of the curvature of A and $q : S^+_{\Gamma_X} \to \wedge^+$ is a certain natural real-quadratic map satisfying

$$|q(\phi)| = \frac{1}{2\sqrt{2}}|\phi|^2,$$

where $\wedge^+$ is the bundle of self-dual 2-forms. We recall some background.

Definition 2 [Kronheimer 1999; Ishida and LeBrun 2003; LeBrun 2009]. Let X be a closed oriented smooth 4-manifold with $b^+(X) \geq 2$. An element

$$\mathfrak{a} \in H^2(X, \mathbb{Z})/\text{torsion} \subset H^2(X, \mathbb{R})$$

is called the monopole class of X if there exists a spin^c-structure Γ_X with

$$c_1^{\mathbb{R}}(\mathscr{L}_{\Gamma_X}) = \mathfrak{a},$$

which has the property that the corresponding Seiberg–Witten monopole equations (3) have a solution for every Riemannian metric on X. Here $c_1^{\mathbb{R}}(\mathscr{L}_{\Gamma_X})$ is the image of the first Chern class $c_1(\mathscr{L}_{\Gamma_X})$ of the complex line bundle $\mathscr{L}_{\Gamma_X}$ in $H^2(X, \mathbb{R})$.

In what follows, we shall usually denote $c_1^{\mathbb{R}}(\mathscr{L}_{\Gamma_X})$ by $c_1(\mathscr{L}_{\Gamma_X})$ for short. A monopole class $c_1(\mathscr{L}_{\Gamma_X})$ of X is called special [Kotschick 2004] if $c_1^2(\mathscr{L}_{\Gamma_X}) \geq 2\chi(X) + 3\tau(X)$ holds. We shall also denote the set of all monopole classes on X by $\mathfrak{C}(X)$. Then we have the following fundamental result on $\mathfrak{C}(X)$.

Proposition 3 [Ishida and LeBrun 2003, Proposition 3]. *Let X be a closed oriented smooth 4-manifold with $b^+(X) \geq 2$. Then the set $\mathfrak{C}(X)$ is a finite set.*

Now recall that, for any subset W of a real vector space V, one can consider the convex hull $\mathrm{Hull}(W) \subset V$, meaning the smallest convex subset of V containing W. Finiteness of $\mathfrak{C}(X)$ implies that the convex hull

$$\mathrm{Hull}(\mathfrak{C}(X)) \subset H^2(X, \mathbb{R})$$

is compact. Moreover, it is known that the convex hull $\mathrm{Hull}(\mathfrak{C}(X))$ is symmetric, that is, $\mathrm{Hull}(\mathfrak{C}(X)) = -\mathrm{Hull}(\mathfrak{C}(X))$. See [LeBrun 2009] for more details.

Since $\mathfrak{C}(X)$ is a finite set, we are able to write $\mathfrak{C}(X) = \{\mathfrak{a}_1, \mathfrak{a}_2, \dots, \mathfrak{a}_n\}$. The convex hull $\mathrm{Hull}(\mathfrak{C}(X))$ is then expressed as

$$\mathrm{Hull}(\mathfrak{C}(X)) = \left\{ \sum_{i=1}^{n} t_i \mathfrak{a}_i \mid t_i \in [0, 1], \ \sum_{i=1}^{n} t_i = 1 \right\}. \tag{4}$$

Notice that the symmetric property tells us that $\mathrm{Hull}(\mathfrak{C}(X))$ contains the zero element. Let us consider the self-intersection function

$$\mathfrak{Q} : H^2(X, \mathbb{R}) \to \mathbb{R},$$

which is defined by $x \mapsto x^2 := \langle x \cup x, [X] \rangle$, where $[X]$ is the fundamental class of X. Since the function $\mathfrak{Q}$ is a polynomial function, it is a continuous function on $H^2(X, \mathbb{R})$. Therefore, the restriction $\mathfrak{Q}|_{\boldsymbol{H}}$ to the compact subset $\boldsymbol{H} := \mathrm{Hull}(\mathfrak{C}(X))$ of $H^2(X, \mathbb{R})$ achieves its maximum.

Definition 4 [LeBrun 2009]. Suppose X is a closed oriented smooth 4-manifold with $b^+(X) \geq 2$. Let $\mathrm{Hull}(\mathfrak{C}(X)) \subset H^2(X, \mathbb{R})$ be the convex hull of the set $\mathfrak{C}(X)$ of all monopole classes on X. If $\mathfrak{C}(X) \neq \varnothing$, define

$$\beta^2(X) := \max\{\mathfrak{Q}(x) := x^2 \mid x \in \mathrm{Hull}(\mathfrak{C}(X))\}.$$

If instead $\mathfrak{C}(X) = \varnothing$, simply define $\beta^2(X) := 0$.

We are now in a position to recall the following theorem.

Theorem 5 [LeBrun 2009]. *Suppose that X is a closed oriented smooth 4-manifold with $b^+(X) \geq 2$. If X admits an Einstein metric g,*

$$2\chi(X) + 3\tau(X) \geq \tfrac{2}{3}\beta^2(X)$$

with equality only if both sides vanish, in which case g must be a hyper-Kähler metric, and X must be diffeomorphic to K3 or T^4.

There are several ways to detect the existence of monopole classes. For any closed oriented smooth 4-manifold X with $b^+(X) \geq 2$, one can define the integer valued Seiberg–Witten invariant $SW_X(\Gamma_X) \in \mathbb{Z}$ for any spin^c-structure Γ_X by integrating a cohomology class on the moduli space of solutions of the Seiberg–Witten monopole equations associated with Γ_X:

$$SW_X : \mathrm{Spin}(X) \to \mathbb{Z},$$

where $\mathrm{Spin}(X)$ is the set of all spin^c-structures on X. For more details, see [Witten 1994; Morgan 1996]. We call the first Chern class $c_1(\mathscr{L}_{\Gamma_X})$ a Seiberg–Witten basic class of X if $SW_X(\Gamma_X) \neq 0$ for a spin^c-structure Γ_X. Notice that Seiberg–Witten basic classes are monopole classes.

On the other hand, there is a sophisticated refinement of the idea of the construction of the Seiberg–Witten invariant due to Bauer and Furuta [2004] (see also [Bauer 2004a; 2004b]). We call it the stable cohomotopy Seiberg–Witten invariant and denote it by BF_X. This invariant detects the presence of a monopole class via an element of a certain complicated stable cohomotopy group $\pi^0_{S^1,\mathcal{U}}(\mathcal{Q})$:

$$BF_X(\Gamma_X) \in \pi^0_{S^1,\mathcal{U}}(\mathcal{Q}).$$

(See [Bauer 2004a] for the construction of the stable cohomotopy group.) Under the assumption that $b^+(X) \geq 2$, it is also known that there is the homeomorphism

$$t^{BF} : \pi^0_{S^1,\mathcal{U}}(\mathcal{Q}) \to \mathbb{Z}, \tag{5}$$

which maps $BF_X(\Gamma_X)$ to $SW_X(\Gamma_X)$ [Bauer 2004a, Theorem 4.1 and Proposition 4.4]. In particular, this map tells us that, if $BF_X(\Gamma_X) = 0$ for some spin^c-structure Γ_X, we have $SW_X(\Gamma_X) = 0$. At the same time, it is known that the nontriviality of $BF_X(\Gamma_X)$ implies that there are solutions of the following perturbed equations associated with Γ_X for all metrics and all self-dual 2-forms η:

$$\mathcal{D}_A\phi = 0, \quad F^+_A = iq(\phi) + i\eta.$$

Namely, $c_1(\mathscr{L}_{\Gamma_X})$ is a generic monopole class in the sense of [Kotschick 2004, Definition 7]. Then, by the standard argument of gauge theory, we know that $c_1(\mathscr{L}_{\Gamma_X})$ becomes a special monopole class [Kotschick 2004, Lemma 8]. Hence the nontriviality of $BF_X(\Gamma_X)$ implies the existence of a special monopole class $c_1(\mathscr{L}_{\Gamma_X})$. We shall use this fact to prove the main results.

By using BF_X, we are able to prove the following result:

Proposition 6. *Let X be a closed oriented smooth 4-manifold with $b^+(X) \geq 2$. Suppose that $SW_X(\Gamma_X) \neq 0$ holds for a spin^c-structure Γ_X. Let N be a closed oriented smooth 4-manifold with $b^+(N) = 0$. Then a connected sum $M := X \# N$ has monopole classes and satisfies the bound*

$$\beta^2(M) \geq 2\chi(X) + 3\tau(X). \tag{6}$$

Proof. As was already mentioned, there is a comparison map (5) between BF_X and SW_X, where we used the assumption that $b^+(X) \geq 2$. In particular, if $SW_X(\Gamma_X) \neq 0$ for some spin^c-structure Γ_X, then $BF_X(\Gamma_X) \neq 0$. Then the proofs of [Ishida and LeBrun 2003, Proposition 6 and Corollary 8] (see also [Bauer 2004a, Theorem 8.8]) tell us that

$$\pm c_1(\mathscr{L}_{\Gamma_X}) + \sum_{i=1}^{k} \pm E_i \tag{7}$$

are monopole classes of the connected sum $M := X \# N$, where $c_1(\mathscr{L}_{\Gamma_X})$ is the first Chern class of the complex line bundle $\mathscr{L}_{\Gamma_X}$ associated with Γ_X. Additionally $E_1, E_2, \ldots, E_k$ is a set of generators for $H^2(N, \mathbb{Z})/\text{torsion}$ relative to which the intersection form is diagonal and the $\pm$ signs are arbitrary and independent of one another. In particular, by (7), we have the following two monopole classes of M:

$$\mathfrak{b}_1 := c_1(\mathscr{L}_{\Gamma_X}) + \sum_{i=1}^{k} E_i, \quad \mathfrak{b}_2 := c_1(\mathscr{L}_{\Gamma_X}) - \sum_{i=1}^{k} E_i.$$

By (4), we obtain

$$c_1(X) = \tfrac{1}{2}\mathfrak{b}_1 + \tfrac{1}{2}\mathfrak{b}_2 \in \mathrm{Hull}(\mathfrak{C}(M)).$$

We therefore get the following bound (see also Definition 4):

$$\beta^2(M) \geq c_1^2(\mathscr{L}_{\Gamma_X}). \tag{8}$$

On the other hand, the assumption that $SW_X(\Gamma_X) \neq 0$ forces the dimension d of Seiberg–Witten monopole moduli space to be nonnegative; that is,

$$d = \tfrac{1}{4}(c_1^2(\mathscr{L}_{\Gamma_X}) - 2\chi(X) - 3\tau(X)) \geq 0.$$

Equivalently, we have

$$c_1^2(\mathscr{L}_{\Gamma_X}) \geq 2\chi(X) + 3\tau(X). \tag{9}$$

It is clear that (8) and (9) imply the desired bound (6). □

Theorem 5 and Proposition 6 imply the next result, a particular case of which, for $N = k\overline{\mathbb{C}P^2}\,\#\,\ell(S^1 \times S^3)$, was proved in [LeBrun 2001, Theorem 3.3].

Theorem 7. *Let X be a closed oriented 4-manifold with $b^+(X) \geq 2$. Suppose that $SW_X(\Gamma_X) \neq 0$ holds for a spin^c-structure Γ_X on X. Let N be a closed oriented smooth 4-manifold with $b^+(N) = 0$. Then a connected sum $M := X \# N$ cannot admit any Einstein metric if*

$$4b_1(N) + b_2(N) > \tfrac{1}{3}(2\chi(X) + 3\tau(X)). \tag{10}$$

In particular, suppose that N is not an integral homology 4-sphere whose fundamental group has no nontrivial finite quotient. Then a connected sum $M := X \# N$ cannot admit any Einstein metric if

$$4b_1(N) + b_2(N) \geq \tfrac{1}{3}(2\chi(X) + 3\tau(X)). \tag{11}$$

Proof. Suppose that the connected sum $M := X \# N$ admits an Einstein metric. Then Theorem 5 tells us that

$$2\chi(M) + 3\tau(M) \geq \tfrac{2}{3}\beta^2(M).$$

This bound with (6) implies

$$2\chi(M) + 3\tau(M) \geq \tfrac{2}{3}(2\chi(X) + 3\tau(X)). \tag{12}$$

On the other hand, a direct computation tells us that

$$2\chi(M) + 3\tau(M) = 2\chi(X) + 3\tau(X) - (4b_1(N) + b_2(N)). \tag{13}$$

By the bounds (12) and (13), we have

$$2\chi(X) + 3\tau(X) - (4b_1(N) + b_2(N)) \geq \tfrac{2}{3}(2\chi(X) + 3\tau(X)).$$

Equivalently,

$$4b_1(N) + b_2(N) \leq \tfrac{1}{3}(2\chi(X) + 3\tau(X)). \tag{14}$$

By contraposition, we are able to conclude that M cannot admit any Einstein metric if (10) holds.

Now suppose that N is not an integral homology 4-sphere whose fundamental group has no nontrivial finite quotient. Then the equality cannot occur in (14). We shall prove this as follows. First of all, notice that Theorem 5 tells us that the equality can occur only in the case where the connected sum $M := X \# N$ is diffeomorphic to K3 or T^4. Both K3 and T^4 are minimal Kähler surfaces. At the same time, [Kotschick 1997, Theorem 5.4] tells us that if a minimal Kähler surface with $b^+ > 1$ admits the connected sum decomposition $X \# N$, then N must be an integral homology 4-sphere whose fundamental group has no nontrivial finite quotient. Hence equality never occurs in (14) as desired. Therefore we conclude that, in this case, M cannot admit any Einstein metric if (11) holds. □

3. Smooth structures and the geography of spin 4-manifolds

The main result of this section is Proposition 10 below. We start by recalling a nice result of Park [2002] on the geography of spin symplectic 4-manifolds.

Let X be a simply connected closed 4-manifold. We define the quantities

$$\chi_h(X) := \frac{\chi(X)+\tau(X)}{4} = \frac{b^+(X)+1}{2}, \tag{15}$$

$$c(X) := c_1^2(X) = 2\chi(X)+3\tau(X) = 4+5b^+(X)-b^-(X). \tag{16}$$

Now suppose that X is a spin symplectic 4-manifold with $b^+ > 1$. Then [Taubes 1994] tells us that X must satisfy

$$c(X) \geq 0. \tag{17}$$

Moreover, Rohlin's theorem forces $\tau(X) \equiv 0 \pmod{16}$. As mentioned in [Park 2002, Lemma 2.1], this is equivalent to

$$c(X) \equiv 8\chi_h(X) \pmod{16}. \tag{18}$$

The above facts tell us that the only lattice points (χ_h, c) satisfying both (17) and (18) can possibly be realized as $(\chi_h(X), c(X))$ of a simply connected spin symplectic 4-manifold X. Such pairs (χ_h, c) of integers are called allowed lattice points.

We are now in a position to recall the following result on the geography of the spin symplectic 4-manifolds:

Theorem 8 [Park 2002, Theorem 1.1]. *There is a line $c = f(\chi_h)$ with a slope > 8.76 in the (χ_h, c)-plane such that any allowed lattice point satisfying $c \leq f(\chi_h)$ in the first quadrant can be realized as (χ_h, c_1^2) of a simply connected spin non-complex symplectic 4-manifold which admits infinitely many distinct smooth structures, all of which admit a symplectic form. In particular, all allowed lattice points (χ_h, c) except finitely many lying in the region $0 \leq c \leq 8.76\chi_h$ satisfy $c \leq f(\chi_h)$.*

On the other hand, let K_g be a fibered knot in S^3 with a punctured genus g surface as fiber. Let M_{K_g} be the 3-manifold obtained by performing 0-framed surgery on K_g. Let m be a meridional circle to K_g. Then the meridional circle m can be seen as a circle in M_{K_g}. The 3-manifold M_{K_g} can be consider as a fiber bundle over the circle m with a closed Riemann surface Σ_g as a fiber. In $M_{K_g} \times S^1$, there is a smoothly embedded torus $T_m = m \times S^1$ of self-intersection 0. A famous result of Thurston [1976] tells us that the 4-manifold $M_{K_g} \times S^1$ admits a symplectic structure with symplectic section T_m. For any symplectic 4-manifold

X with a symplectically embedded torus T of self-intersection 0, we can consider the symplectic fiber sum $X_{M_{K_g}}$ of X with $M_{K_g} \times S^1$ as follows:

$$X_{K_g} := X \#_{T=T_m} (M_{K_g} \times S^1) = [X - (T \times D^2)] \cup [(M_{K_g} \times S^1) - (T_m \times D^2)],$$

where $T \times D^2$ is a tubular neighborhood of the torus T in the manifold X. Under a certain condition on X, Fintushel and Stern proved that X_{K_g} is homeomorphic to X and provided a way to compute the Seiberg–Witten invariants of X_{K_g}:

Theorem 9 [Fintushel and Stern 1998, Theorem 1.5]. *Let X be a simply connected symplectic 4-manifold that contains a symplectically embedded torus T of self-intersection 0 in a cusp neighborhood with $\pi_1(X - T) = 1$ and representing a nontrivial homology class $[T]$. If K_g is a fibered knot, $X_{K_g} := X \#_{T=T_m} (M_{K_g} \times S^1)$ is a symplectic 4-manifold which is homeomorphic to X and whose Seiberg–Witten polynomial is given by*

$$\mathcal{SW}_{K_g} = \mathcal{SW}_X \cdot \Delta_{X_{K_g}}(t),$$

where $\Delta_{X_{K_g}}(t)$ is the Alexander polynomial of K_g and $t = \exp(2[T])$.

The polynomial $\mathcal{SW}_X$ is defined as follows. Let $\{\pm\beta_1, \pm\beta_2, \dots, \pm\beta_n\}$ be the set of nonzero Seiberg–Witten basic classes of X. Then we set $b_0 = SW_X(0)$, $b_j = SW_X(\beta_j)$, and $t_{\beta_j} = \exp(\beta_j)$. Then we define the Seiberg–Witten polynomial $\mathcal{SW}_X$ as follows:

$$\mathcal{SW}_X = b_0 + \sum_{j=1}^{n} b_j \bigl(t_{\beta_j} + (-1)^{(\chi(X)+\tau(X))/4} t_{\beta_j}^{-1}\bigr).$$

Since any two smooth 4-manifolds which have different Seiberg–Witten polynomials are nondiffeomorphic, one can apply Theorem 9 to construct 4-manifolds admitting infinitely many distinct smooth structures. Indeed, Park [2002] proved that every symplectic 4-manifold W in Theorem 8 admits infinitely many distinct smooth structures by showing that the 4-manifold W contains a symplectically embedded torus T of self-intersection 0 in a cusp neighborhood with $\pi_1(X-T)=1$ and representing a nontrivial homology class $[T]$, that is, the 4-manifold W satisfies the assumption in Theorem 9. See [Park 2002, Claim 1] for more details.

To prove the main results of the present article, we need to refine Park's result on the existence of exotic smooth structures:

Proposition 10. *Let W be any symplectic 4-manifold in Theorem 8 and let N be any closed smooth 4-manifold with $b^+(N) = 0$. Then the underlying topological 4-manifold of $W \# N$ admits infinitely many distinct smooth structures.*

Proof. As was already proved in [Park 2002, Claim 1], every symplectic 4-manifold W in Theorem 8 contains a symplectically embedded torus T of self-intersection 0 in a cusp neighborhood with $\pi_1(X-T)=1$ and representing a nontrivial homology class $[T]$. On the other hand, for any fibered knot K_g in S^3 with a punctured genus g surface as fiber, consider the 3-manifold M_{K_g} obtained by performing 0-framed surgery on K_g. As was already mentioned above, we are able to consider the symplectic fiber sum

$$W_{K_g} = W \#_{T=T_m} (M_{K_g} \times S^1).$$

Then, Theorem 9 tells us that W_{K_g} is homeomorphic to W. Now, let N be any closed smooth 4-manifold with $b^+(N) = 0$. Then we consider a connected sum $W_{K_g} \# N$. Of course, $W_{K_g} \# N$ is homeomorphic to $W \# N$. Notice that the connected sum $W_{K_g} \# N$ is not necessarily symplectic in general.

Next, we show that there are monopole classes of $W_{K_g} \# N$. In fact, [Bauer 2004a, Proposition 5.4] tells us that the comparison map (5), that is,

$$t^{BF} : \pi^0_{S^1,\mathcal{U}}(\mathcal{Q}) \to \mathbb{Z},$$

becomes an isomorphism for any symplectic 4-manifold M with $b^+(M) > 1$. This fact and a result of Taubes [1994] on the nontriviality of Seiberg–Witten invariants of any symplectic 4-manifold M with $b^+(M) > 1$ imply the nontriviality of the stable cohomotopy Seiberg–Witten invariants of M. In particular, we can conclude that the symplectic 4-manifold W_{K_g} has

$$\pm c_1(\mathcal{K}_{W_{K_g}})$$

as its monopole classes, where $\mathcal{K}_{W_{K_g}}$ is the canonical line bundle of W_{K_g}. Since the nontriviality of the stable cohomotopy Seiberg–Witten invariants does not change under the connected sum with N [Ishida and LeBrun 2003, Proposition 6], we can conclude that the connected sum $W_{K_g} \# N$ also has monopole classes [Ishida and LeBrun 2003, Proposition 6 and Corollary 8], that is, the cohomology classes

$$\pm c_1(\mathcal{K}_{W_{K_g}}) + \sum_{i=1}^{k} \pm E_i \tag{19}$$

become monopole classes of the connected sum $W_{K_g} \# N$, where $E_1, E_2, \dots, E_k$ is a set of generators for $H^2(N, \mathbb{Z})$/torsion relative to which the intersection form is diagonal and the $\pm$ signs are arbitrary and independent of one another.

On the other hand, following the argument at the beginning of the proof of [Fintushel and Stern 1998, Corollary 1.7], we are able to express $c_1(\mathcal{K}_{W_{K_g}})$ more explicitly. For the reader, let us explain it here. First of all, notice that the homology $H_2(M_{K_g} \times S^1)$ is generated by the classes of the symplectic curves T_m and Σ_g. This tells us that the first Chern class $c_1(\mathcal{K}_{M_{K_g}\times S^1})$ of the canonical line bundle $\mathcal{K}_{M_{K_g}\times S^1}$ of the symplectic 4-manifold $M_{K_g} \times S^1$ has the form

$$c_1(\mathcal{K}_{K_g\times S^1}) = \alpha[T_m] + \beta[\Sigma_g].$$

Since we have $[T_m]^2 = [\Sigma_g]^2 = 0$ and $[T_m]\cdot[\Sigma_g] = 1$,

$$c_1(\mathcal{K}_{K_g\times S^1})\cdot[\Sigma_g] = \alpha, \quad c_1(\mathcal{K}_{W_{K_g}})\cdot[T_m] = \beta.$$

These facts and the adjunction formula tell us that $\alpha = 2g-2$ and $\beta = 0$ hold. Therefore, we conclude that

$$c_1(\mathcal{K}_{K_g\times S^1}) = (2g-2)[T_m] = (2g-2)[T].$$

On the other hand, $c_1(\mathcal{K}_{W_{K_g}}) = c_1(\mathcal{K}_W) + c_1(\mathcal{K}_{K_g\times S^1}) + 2[T]$ holds by the construction of W_{K_g}. Therefore,

$$c_1(\mathcal{K}_{W_{K_g}}) = c_1(\mathcal{K}_W) + c_1(\mathcal{K}_{K_g\times S^1}) + 2[T] = c_1(\mathcal{K}_W) + 2g[T]. \tag{20}$$

By (19) and (20), we conclude that

$$\pm(c_1(\mathcal{K}_W) + 2g[T]) + \sum_{i=1}^{k} \pm E_i$$

are monopole classes of $W_{K_g} \# N$.

By considering an infinite sequence $\{K_{g_\ell}\}_{\ell\in\mathbb{N}}$ of fibered knots with $g_\ell \geq 1$, where the genus g_ℓ is strictly increasing with respect to ℓ, that is, $g_\ell \to \infty$ when $\ell \to \infty$, we have an infinite sequence $\{W_{K_{g_\ell}} \# N\}_{\ell\in\mathbb{N}}$ of smooth 4-manifolds which are homeomorphic to $W \# N$ and, for each ℓ,

$$\pm(c_1(\mathcal{K}_W) + 2g_\ell[T]) + \sum_{i=1}^{k} \pm E_i \tag{21}$$

are monopole classes of $W_{K_{g_\ell}} \# N$. Suppose now that the sequence

$$\{W_{K_{g_\ell}} \# N\}_{\ell\in\mathbb{N}}$$

contains only finitely many diffeomorphism types. Specifically, suppose that there exists a positive integer ℓ_0 such that $W_{K_{g_{\ell_0}}} \# N$ is diffeomorphic to $W_{K_{g_{\ell_i}}} \# N$ for any $\ell_i \geq \ell_0$. Then, by the expression (21) of the monopole classes and taking $\ell_i \to \infty$, we conclude that the set of monopole classes of 4-manifold $W_{K_{g_{\ell_0}}} \# N$ is unbounded. However, this is a contradiction because the set of monopole classes of any given smooth 4-manifold with $b^+ > 1$ must be finite by Proposition 3. Therefore, the sequence $\{W_{K_{g_\ell}} \# N\}_{\ell\in\mathbb{N}}$ actually contains infinitely many diffeomorphism types. As was already mentioned above, since each 4-manifold $W_{K_{g_\ell}} \# N$ is homeomorphic to $W \# N$, the underlying topological 4-manifold of $W \# N$ admits infinitely many distinct smooth structures as desired. □

4. Proof of Theorem A

Theorem 11. *Let Z be a closed oriented smooth 4-manifold satisfying $b^+(Z) = 0$, $b_1(Z) \neq 0$, and*

$$0 < \frac{b_2(Z)}{b_1(Z)} < \frac{19}{50}, \tag{22}$$

where $b_1(Z)$ and $b_2(Z)$ denote the first and second Betti numbers of Z, respectively. Let M be any closed oriented smooth 4-manifold with $b_1(M) = 0$, $b^+(M) = 0$, and $b_2(M) = 1$.

For such 4-manifolds as Z and M, and for any pair of integers (m, n) satisfying $m + n \equiv 0 \pmod 2$, there exist infinitely many nonhomeomorphic topological nonspin 4-manifolds $X^{m,n}_{Z,M}$ with $(\chi, \tau) = (m, n)$, and all such topological nonspin 4-manifolds admit infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist.

Proof. For any pair of integers (m, n) satisfying $m + n \equiv 0 \pmod 2$, it is easy to see that there exist infinitely many pairs (k, ℓ) of sufficiently large positive integers satisfying the following three conditions:

$$(x, y) := \left(\frac{m + n + 2b_1(Z)\ell}{4},\ 2m + 3n + k + (4b_1(Z) + b_2(Z))\ell\right) \in \mathfrak{D},$$

$$n + k + b_2(Z)\ell \equiv 0 \pmod{16}, \tag{23}$$

$$k + (4b_1(Z) + b_2(Z))\ell > \tfrac{1}{3}y, \tag{24}$$

where $\mathfrak{D}$ is the set of all pairs of integers satisfying the conditions of Theorem 8. Notice that the condition (23) is equivalent to

$$y \equiv 8x \pmod{16}.$$

Moreover, the condition (22) was already used, that is,

$$8.76 > \frac{(4b_1(Z) + b_2(Z))\ell}{(2b_1(Z)\ell)/4} = 8 + 2\frac{b_2(Z)}{b_1(Z)},$$

or, equivalently,

$$\frac{b_2(Z)}{b_1(Z)} < 0.38 = \frac{19}{50}.$$

By Theorem 8, for each (x, y) above, there is a simply connected spin noncomplex symplectic 4-manifold W with

$$\chi_h(W) := \frac{\chi(W) + \tau(W)}{4} = x = \frac{m + n + 2b_1(Z)\ell}{4},$$

$$c_1^2(W) := 2\chi(W) + 3\tau(W) = y = 2m + 3n + k + (4b_1(Z) + b_2(Z))\ell. \tag{25}$$

Hence we obtain

$$m+n=\chi(W)+\tau(W)-2b_1(Z)\ell, \tag{26}$$

$$2m+3n=2\chi(W)+3\tau(W)-k-(4b_1(Z)+b_2(Z))\ell. \tag{27}$$

Consider the connected sum

$$X^{m,n}_{Z,M}=W\,\#\,M\,\#\,(k-1)\overline{\mathbb{C}P^2}\,\#\,\ell Z. \tag{28}$$

Notice that $X^{m,n}_{Z,M}$ is nonspin. On the other hand, by a direct computation, we obtain

$$\chi(M\,\#\,(k-1)\overline{\mathbb{C}P^2}\,\#\,\ell Z)=k+\chi(\ell Z)=k+2+b_2(Z)\ell-2b_1(Z)\ell.$$

Notice that $\chi(M)=3$ because we assume that $b_1(M)=0$, $b^+(M)=0$, and $b_2(M)=1$. Therefore we get

$$\chi(X^{m,n}_{Z,M})=\chi(W)+\chi(M\#(k-1)\overline{\mathbb{C}P^2})-2=\chi(W)+k+b_2(Z)\ell-2b_1(Z)\ell. \tag{29}$$

Similarly, we have

$$\begin{aligned}\tau(M\,\#\,(k-1)\overline{\mathbb{C}P^2}\,\#\,\ell Z)&=\tau(M)+\tau((k-1)\overline{\mathbb{C}P^2})+\tau(\ell Z)\\&=-1+(1-k)-b_2(Z)\ell\\&=-k-b_2(Z)\ell.\end{aligned}$$

Notice that $b^+(Z)=0$, so we have $\tau(Z)=-b_2(Z)$. Therefore

$$\tau(X^{m,n}_{Z,M})=\tau(W)+\tau(M\,\#\,(k-1)\overline{\mathbb{C}P^2}\,\#\,\ell Z)=\tau(W)-k-b_2(Z)\ell. \tag{30}$$

By (29) and (30), we get

$$\chi(X^{m,n}_{Z,M})+\tau(X^{m,n}_{Z,M})=\chi(W)+\tau(W)-2b_1(Z)\ell, \tag{31}$$

$$2\chi(X^{m,n}_{Z,M})+3\tau(X^{m,n}_{Z,M})=2\chi(W)+3\tau(W)-k-(4b_1(Z)+b_2(Z))\ell. \tag{32}$$

Then (26), (27), (31), and (32) immediately tell us that

$$\chi(X^{m,n}_{Z,M})=m,\quad \tau(X^{m,n}_{Z,M})=n. \tag{33}$$

On the other hand, by $b_1(W)=b_1(M)=b_1(\overline{\mathbb{C}P^2})=0$, we have

$$b_1(X^{m,n}_{Z,M})=b_1(Z)\ell. \tag{34}$$

Since there are infinitely many choices of ℓ, (33) and (34) tell us that, for (m,n) satisfying $m+n\equiv 0 \pmod 2$, there exist infinitely many nonhomeomorphic nonspin 4-manifolds $X^{m,n}_{Z,M}$ with $(\chi,\tau)=(m,n)$. Now set $N:=M\,\#\,(k-1)\overline{\mathbb{C}P^2}\,\#\,\ell Z$. We write $X^{m,n}_{Z,M}=W\,\#\,N$. Notice that N satisfies $b^+(N)=0$.

By considering an infinite sequence $\{K_{g_i}\}_{i\in\mathbb{N}}$ of fibered knots with $g_i\geq 1$ as the proof of Proposition 10 above, we obtain the sequence $\{W_{K_{g_i}}\,\#\,N\}_{i\in\mathbb{N}}$ which contains infinitely many diffeomorphism types by Proposition 10, and every 4-manifold $X_{g_i} := W_{K_{g_i}} \# N$ is homeomorphic to $X^{m,n}_{Z,M} = W \# N$. To prove that the 4-manifold $X^{m,n}_{Z,M}$ admits infinitely many distinct smooth structures for which nonzero monopole classes exist and Einstein metrics do not exist, it is enough to prove that the smooth 4-manifold X_{g_i} has nonzero special monopole classes and cannot admit any Einstein metric. It is clear that X_{g_i} has nonzero special monopole classes by the proof of Proposition 10, where the nontriviality of BF implies the existence of a special monopole class. On the other hand, since any symplectic 4-manifold with $b^+ > 1$ has nontrivial Seiberg–Witten invariants by a result of Taubes [1994] and N satisfies $b^+(N) = 0$, Theorem 7 tells us that if

$$k + (4b_1(Z) + b_2(Z))\ell > \tfrac{1}{3}(2\chi(W_{K_{g_i}}) + 3\tau(W_{K_{g_i}})), \tag{35}$$

the manifold $X_{g_i} := W_{K_{g_i}} \# N$ cannot admit any Einstein metric. By Park's construction of W and Theorem 9, $W_{K_{g_i}}$ is homeomorphic to W and we therefore have

$$2\chi(W_{K_{g_i}}) + 3\tau(W_{K_{g_i}}) = 2\chi(W) + 3\tau(W).$$

So the bound (35) is equivalent to

$$k + (4b_1(Z) + b_2(Z))\ell > \tfrac{1}{3}(2\chi(W) + 3\tau(W)). \tag{36}$$

However, the bound (36) automatically holds because we have (24) and (25). Therefore, $X_{g_i} := W_{K_{g_i}} \# N$ cannot admit any Einstein metric as desired. □

Theorem A immediately follows from Theorem 11. Indeed, it is enough to find a smooth closed 4-manifold satisfying (22) and a closed oriented smooth 4-manifold M with $b_1(M) = 0$, $b^+(M) = 0$, and $b_2(M) = 1$. For example, set $M := \overline{\mathbb{C}P^2}$ and $Z := 11K \# 4\overline{\mathbb{C}P^2}$, where K is a secondary Kodaira surface; cf. [Barth et al. 1984]. We have $b_1(K) = 1$, $b^+(K) = 0$, and $b_2(K) = 0$. It is clear that M satisfies $b_1(M) = 0$, $b^+(M) = 0$, and $b_2(M) = 1$. We also have $b_1(Z) = 11$, $b_2^+(Z) = 0$, and $b_2(Z) = 4$. Therefore, we get

$$\frac{b_2(Z)}{b_1(Z)} = \frac{4}{11} < \frac{19}{50}.$$

Hence $Z := 11K \# 4\overline{\mathbb{C}P^2}$ is a 4-manifold satisfying (22). Hence we have proved Theorem A by considering the connected sum

$$X^{m,n}_{11K\#4\overline{\mathbb{C}P^2},\,\overline{\mathbb{C}P^2}} = W \# k\overline{\mathbb{C}P^2} \# \ell(11K \# 4\overline{\mathbb{C}P^2});$$

see (28).

Remark 12. We are able to use another negative definite 4-manifold satisfying (22) to prove Theorem A. For example, we are able to use another connected sum

$aT\#b\overline{Y}$, or $aT\#b\overline{\mathbb{C}P^2}$ as Z above by taking a suitable pair of positive integers (a, b) for which the condition (22) is satisfied. Here $\overline{Y}$ is a Mumford fake projective plane [Mumford 1979] with the reversed orientation, and T is $S^1 \times S^3$, or a secondary Kodaira surface. If we take Z as $a(S^1 \times S^3)\#b\overline{\mathbb{C}P^2}$, then the resulting 4-manifold $X^{m,n}_{Z,\overline{\mathbb{C}P^2}}$ has a free fundamental group. See also [Del Rio Guerra 2002].

5. Proof of Theorem B

A method similar to that used in the proof of Theorem 11 enables us to prove Theorem B:

Theorem 13. *For any pair of integers* (m, n) *satisfying* $m + n \equiv 0 \pmod 2$ *and* $n \equiv 0 \pmod{16}$, *there exist infinitely many nonhomeomorphic topological spin* 4*-manifolds with* $(\chi, \tau) = (m, n)$ *and all such topological spin* 4*-manifolds admit infinitely many distinct smooth structures for which nonzero special monopole classes exist and Einstein metrics do not exist.*

Proof.

For any pair of integers (m, n) satisfying $m+n \equiv 0 \pmod 2$ and $n \equiv 0 \pmod{16}$, we are able to see that there exist infinitely many, sufficiently large positive integers ℓ satisfying

$$(x, y) := \Big(\frac{m+n+2\ell}{4},\ 2m+3n+4\ell\Big) \in \mathfrak{D},$$

$$4\ell > \tfrac{1}{3}y, \tag{37}$$

where $\mathfrak{D}$ is the set of all pairs of integers satisfying the conditions of Theorem 8. In particular, notice that $y \equiv 8x \pmod{16}$ must be satisfied, that is,

$$2m+3n+4\ell \equiv 8\Big(\frac{m+n+2\ell}{4}\Big) \pmod{16}.$$

Specifically, we have

$$2m+3n+4\ell \equiv 2m+2n+4\ell \pmod{16}.$$

This is nothing but $n \equiv 0 \pmod{16}$. By Theorem 8, for each (x, y) above, there is a simply connected spin noncomplex symplectic 4-manifold W with

$$\chi_h(W) := \frac{\chi(W)+\tau(W)}{4} = x = \frac{m+n+2\ell}{4},$$

$$c_1^2(W) := 2\chi(W)+3\tau(W) = y = 2m+3n+4\ell. \tag{38}$$

We obtain

$$m+n = \chi(W)+\tau(W)-2\ell, \tag{39}$$

$$2m+3n = 2\chi(W)+3\tau(W)-4\ell. \tag{40}$$

Let us consider the connected sum

$$X^{m,n} = W \# \ell(S^1 \times S^3).$$

Notice that $X^{m,n}$ is spin. We also get

$$\chi(X^{m,n}) + \tau(X^{m,n}) = \chi(W) + \tau(W) - 2\ell, \tag{41}$$

$$2\chi(X^{m,n}) + 3\tau(X^{m,n}) = 2\chi(W) + 3\tau(W) - 4\ell. \tag{42}$$

By (39), (40), (41), and (42), we obtain

$$\chi(X^{m,n}) = m, \quad \tau(X^{m,n}) = n. \tag{43}$$

On the other hand, we have $b_1(W) = 0$ and $b_1(S^1 \times S^3) = 1$. Therefore, we get

$$b_1(X^{m,n}) = \ell. \tag{44}$$

Since there are infinitely many choices of ℓ, (43) and (44) implies that, for (m, n) satisfying $m + n \equiv 0 \pmod 2$ and $n \equiv 0 \pmod{16}$, there exist infinitely many nonhomeomorphic spin 4-manifolds $X^{m,n}$ with $(\chi, \tau) = (m, n)$.

We set $N := \ell(S^1 \times S^3)$ and write $X^{m,n} = W \# N$. Notice that N satisfies $b^+(N) = 0$. By considering an infinite sequence $\{K_{g_i}\}_{i\in\mathbb{N}}$ of fibered knots with $g_i \geq 1$ as proof of Theorem 11, we obtain the sequence $\{W_{K_{g_i}} \# N\}_{i\in\mathbb{N}}$ which contains infinitely many diffeomorphism types. Every 4-manifold $X_{g_i} := W_{K_{g_i}} \# N$ is homeomorphic to $X^{m,n} = W \# N$ and has nonzero special monopole classes. Moreover, X_{g_i} cannot admit any Einstein metric as follows. By Theorem 7, if

$$4\ell > \tfrac{1}{3}(2\chi(W_{K_{g_i}}) + 3\tau(W_{K_{g_i}})), \tag{45}$$

the manifold $X_{g_i} := W_{K_{g_i}} \# N$ cannot admit any Einstein metric. Since

$$2\chi(W_{K_{g_i}}) + 3\tau(W_{K_{g_i}}) = 2\chi(W) + 3\tau(W),$$

the bound (45) is equal to

$$4\ell > \tfrac{1}{3}(2\chi(W) + 3\tau(W)). \tag{46}$$

However, the bound (46) automatically holds because of (37) and (38). Therefore, we conclude that $X_{g_i} := W_{K_{g_i}} \# N$ cannot admit any Einstein metric as desired. □

6. Remarks on the simply connected case

NonEinstein 4-manifolds constructed in Theorem A, Theorem B, and Theorem 1 are not simply connected. It is natural to ask whether one can prove simply connected versions of these theorems. This is an open problem, and is closely related to the following.

Question 14 [LeBrun 2001, Question 3.5]. For every $q \in (-1, 1) \cap \mathbb{Q}$, are there smooth, compact simply connected 4-manifolds with $\tau/\chi = q$ which do not admit Einstein metrics?

LeBrun [2001] gives a partial affirmative answer to this question under $0.351 \leq |q| < 1$ [LeBrun 2001, Corollary 3.6]. However, a complete solution to this question is still unknown.

On the other hand, in the present article, we have seen that 4-manifolds often admit infinitely many distinct smooth structures for which no Einstein metric exists. In light of this phenomenon, we would like to consider a generalization of Question 14:

Question 15. For every $q \in (-1, 1) \cap \mathbb{Q}$, are there compact simply connected topological 4-manifolds with $\tau/\chi = q$ which admit infinitely many distinct smooth structures for which no Einstein metrics exist?

To prove a result in this direction, we need to recall the following:

Theorem 16 [Park 2003, Theorem 1.1]. *There is an increasing sequence $\{m_i\}$ with $m_i \to 9$ such that every simply connected closed, nonspin, irreducible smooth 4-manifold X satisfying $0 \leq c(X) \leq m_i \chi_h(X)$ and $b^+(X) \geq C_i$, where C_i is an odd constant depending on m_i, admits infinitely many, both symplectic and nonsymplectic, exotic smooth structures.*

Applying the idea of the proof of [LeBrun 2001, Corollary 3.6] and the preceding result, we obtain:

Corollary 17. *For any rational number $q \in \mathbb{Q}$ satisfying*

$$\tfrac{1}{3} < |q| < 1,$$

there exist compact simply connected topological 4-manifolds with $\tau/\chi = q$ admitting infinitely many distinct smooth structures for which no Einstein metric exists.

Proof. We will actually prove that there is an increasing sequence $\{n_i\}$ such that $n_i \to -1/3$ satisfying the following property: For any rational number q with

$$-1 < q \leq n_i,$$

there exist compact simply connected topological 4-manifolds with $\tau/\chi = q$ admitting infinitely many distinct smooth structures for which no Einstein metrics exists.

Let α and β be any positive integers satisfying

$$\frac{s}{t} \in (0,\ m_i],$$

where m_i is an increasing sequence with $m_i \to 9$ in Theorem 16. Then, Theorem 16 specifically tells us that, for any sufficiently large integers ℓ, there is a simply connected nonspin 4-manifold X admitting infinitely many distinct symplectic structures and with

$$\chi_h(X) = \frac{\chi(X)+\tau(X)}{4} = t\ell, \tag{47}$$

$$c_1^2(X) = 2\chi(X) + 3\tau(X) = s\ell. \tag{48}$$

We denote the infinite family of symplectic 4-manifolds which are homeomorphic to X by $\{Y_n\}$. For each symplectic 4-manifold Y_n which is homeomorphic to X, consider the k-times blow-up $M_n^k := Y_n \# k\overline{\mathbb{C}P^2}$ of Y_n where k satisfies

$$k \geq \tfrac{1}{3}(2\chi(Y_n) + 3\tau(Y_n)) = \tfrac{1}{3}(2\chi(X)+3\tau(X)).$$

By (48), this is equivalent to

$$\frac{k}{s\ell} \geq \tfrac{1}{3}.$$

Theorem 7 tells us that M_n^k cannot admit any Einstein metric. Moreover, for each k, the infinite family $\{M_n^k\}$ of symplectic 4-manifolds also contains infinitely many diffeomorphism types because the difference of smooth structures survives under blow-ups. This means that, for each k, the underlying topological 4-manifold of $X_k := X \# k\overline{\mathbb{C}P^2}$ admits infinitely many smooth structures $\{M_n^k\}$ without Einstein metrics.

On the other hand, we have $\chi_h(M_n^k) = \chi_h(X)$ and $c_1^2(M_n^k) = c_1^2(X) - k$ because Y_n is homeomorphic to X. Using this fact, (47), and (48), we have

$$\frac{c_1^2(M_n^k)}{\chi_h(M_n^k)} = \frac{c_1^2(X)-k}{\chi_h(X)} = \frac{s\ell - k}{t\ell} = \frac{s}{t}\Big(1 - \frac{k}{s\ell}\Big).$$

Since $k/s\ell \in [1/3, \infty) \cap \mathbb{Q}$, we get

$$1 - \frac{k}{s\ell} \in (-\infty, \tfrac{2}{3}] \cap \mathbb{Q}.$$

Since we also have $s/t \in (0, m_i]$, the following holds.

$$\frac{c_1^2(M_n^k)}{\chi_h(M_n^k)} \in (-\infty, \tfrac{2}{3}m_i] \cap \mathbb{Q}. \tag{49}$$

On the other hand, as was already mentioned in the proof of [LeBrun 2001, Corollary 3.6], we get

$$\frac{\tau(M_n^k)}{\chi(M_n^k)} = \left(3 - \frac{1}{4}\frac{c_1^2(M_n^k)}{\chi_h(M_n^k)}\right)^{-1} - 1.$$

By (49), we have

$$\left(3 - \frac{1}{4}\frac{c_1^2(M_n^k)}{\chi_h(M_n^k)}\right)^{-1} \in \Big(0, \frac{6}{18-m_i}\Big] \cap \mathbb{Q}.$$

This tells us that

$$\frac{\tau(M^k_n)}{\chi(M^k_n)} \in \Big(-1, -1 + \frac{6}{18-m_i}\Big] \cap \mathbb{Q}.$$

Since the sequence $m_i \to 9$ is increasing, we have an increasing sequence $\{n_i\}$ such that

$$n_i \to -\tfrac{1}{3}$$

by setting

$$n_i := -1 + \frac{6}{18-m_i}.$$

Since we have $\tau(M^k_n) = \tau(X_k)$ and $\chi(M^k_n) = \tau(X_k)$, the above tells us that, for any rational number q with

$$-1 < q \leq n_i,$$

there exist compact simply connected topological 4-manifolds X_k with $\tau/\chi = q$ admitting infinitely many distinct smooth structures $\{M^k_n\}$ for which no Einstein metrics exists. The case where q is positive then follows by reversing the orientation of the above examples. □

Acknowledgement

The author thanks the referee for valuable comments and suggestions.

References

[Barth et al. 1984] W. Barth, C. Peters, and A. Van de Ven, *Compact complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **4**, Springer, Berlin, 1984. MR 86c:32026 Zbl 0718.14023

[Bauer 2004a] S. Bauer, "Refined Seiberg–Witten invariants", pp. 1–46 in *Different faces of geometry*, edited by S. Donaldson et al., Int. Math. Ser. (N. Y.) **3**, Kluwer/Plenum, New York, 2004. MR 2005h:57045 Zbl 1083.57039

[Bauer 2004b] S. Bauer, "A stable cohomotopy refinement of Seiberg–Witten invariants. II", *Invent. Math.* **155**:1 (2004), 21–40. MR 2005c:57041 Zbl 1051.57038

[Bauer and Furuta 2004] S. Bauer and M. Furuta, "A stable cohomotopy refinement of Seiberg–Witten invariants. I", *Invent. Math.* **155**:1 (2004), 1–19. MR 2005c:57040 Zbl 1050.57024

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Besson et al. 1995] G. Besson, G. Courtois, and S. Gallot, "Entropies et rigidités des espaces localement symétriques de courbure strictement négative", *Geom. Funct. Anal.* **5**:5 (1995), 731–799. MR 96i:58136 Zbl 1194.01076

[Brunnbauer et al. 2009] M. Brunnbauer, M. Ishida, and P. Suárez-Serrato, "An essential relation between Einstein metrics, volume entropy, and exotic smooth structures", *Math. Res. Lett.* **16**:3 (2009), 503–514. MR 2011g:53088 Zbl 1183.53038

[Del Rio Guerra 2002] H. Del Rio Guerra, "Seiberg–Witten invariants of nonsimple type and Einstein metrics", *Ann. Global Anal. Geom.* **21**:4 (2002), 319–339. MR 2003g:53062 Zbl 1011.53038

[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

[Hitchin 1974] N. Hitchin, "Compact four-dimensional Einstein manifolds", *J. Differential Geometry* **9** (1974), 435–441. MR 50 #3149 Zbl 0281.53039

[Ishida and LeBrun 2002] M. Ishida and C. LeBrun, "Spin manifolds, Einstein metrics, and differential topology", *Math. Res. Lett.* **9**:2-3 (2002), 229–240. MR 2003f:53079 Zbl 1017.53045

[Ishida and LeBrun 2003] M. Ishida and C. LeBrun, "Curvature, connected sums, and Seiberg–Witten theory", *Comm. Anal. Geom.* **11**:5 (2003), 809–836. MR 2005c:53042 Zbl 1084.58501

[Kotschick 1997] D. Kotschick, "The Seiberg–Witten invariants of symplectic four-manifolds (after C. H. Taubes)", pp. 195–220, exposé No. 812, 4, in *Séminaire Bourbaki, Vol. 1995/96*, Astérisque **241**, Soc. Math. de France, Paris, 1997. MR 98h:57057 Zbl 0882.57026

[Kotschick 1998] D. Kotschick, "Einstein metrics and smooth structures", *Geom. Topol.* **2** (1998), 1–10. MR 99h:57060 Zbl 0885.57014

[Kotschick 2004] D. Kotschick, "Monopole classes and Einstein metrics", *Int. Math. Res. Not.* **2004**:12 (2004), 593–609. MR 2004k:53062 Zbl 1051.57038

[Kronheimer 1999] P. B. Kronheimer, "Minimal genus in $S^1 \times M^3$", *Invent. Math.* **135**:1 (1999), 45–61. MR 2000c:57071 Zbl 0917.57017

[LeBrun 1995a] C. LeBrun, "Einstein metrics and Mostow rigidity", *Math. Res. Lett.* **2**:1 (1995), 1–8. MR 95m:53067 Zbl 0974.53035

[LeBrun 1995b] C. LeBrun, "Polarized 4–manifolds, extremal Kähler metrics, and Seiberg–Witten theory", *Math. Res. Lett.* **2**:5 (1995), 653–662. MR 96h:58038 Zbl 0874.53051

[LeBrun 1996] C. LeBrun, "Four-manifolds without Einstein metrics", *Math. Res. Lett.* **3**:2 (1996), 133–147. MR 97a:53072 Zbl 0856.53035

[LeBrun 2001] C. LeBrun, "Ricci curvature, minimal volumes, and Seiberg–Witten theory", *Invent. Math.* **145**:2 (2001), 279–316. MR 2002h:53061 Zbl 0999.53027

[LeBrun 2003] C. LeBrun, "Einstein metrics, four-manifolds, and differential topology", pp. 235–255 in *Surveys in differential geometry* (Boston, MA, 2002), edited by S.-T. Yau, Surv. Differ. Geom. **8**, Int. Press, Somerville, MA, 2003. MR 2005g:53078 Zbl 1051.53038

[LeBrun 2009] C. LeBrun, "Four-manifolds, curvature bounds, and convex geometry", pp. 119–152 in *Riemannian topology and geometric structures on manifolds*, edited by K. Galicki and S. R. Simanca, Progr. Math. **271**, Birkhäuser, Boston, MA, 2009. MR 2010h:53045 Zbl 1191.57023

[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

[Mumford 1979] D. Mumford, "An algebraic surface with K ample, $(K^2) = 9$, $p_g = q = 0$", *Amer. J. Math.* **101**:1 (1979), 233–244. MR 80j:14032 Zbl 0433.14021

[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 2003] J. Park, "Exotic smooth structures on 4-manifolds. II", *Topology Appl.* **132**:2 (2003), 195–202. MR 2004d:57033 Zbl 1028.57032

[Sambusetti 1998] A. Sambusetti, "An obstruction to the existence of Einstein metrics on 4-manifolds", *Math. Ann.* **311**:3 (1998), 533–547. MR 99k:53101 Zbl 0920.53026

[Taubes 1994] C. H. Taubes, "The Seiberg–Witten invariants and symplectic forms", *Math. Res. Lett.* **1**:6 (1994), 809–822. MR 95j:57039 Zbl 0853.57019

[Thorpe 1969] J. A. Thorpe, "Some remarks on the Gauss–Bonnet integral", *J. Math. Mech.* **18** (1969), 779–786. MR 41 #963 Zbl 0183.50503

[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

[Witten 1994] E. Witten, "Monopoles and four-manifolds", *Math. Res. Lett.* **1**:6 (1994), 769–796. MR 96d:57035 Zbl 0856.53035

Received October 31, 2010. Revised May 16, 2012.

MASASHI ISHIDA
DEPARTMENT OF MATHEMATICS
SOPHIA UNIVERSITY
7-1 KIOI-CHO
CHIYODA-KU, TOKYO 102-8554
JAPAN
ishida-m@hoffman.cc.sophia.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

NOETHER'S PROBLEM FOR $\widehat{S}_4$ AND $\widehat{S}_5$

MING-CHANG KANG AND JIAN ZHOU

Let k be a field, let G be a finite group and let $k(x_g : g \in G)$ be the rational function field over k, on which G acts by the k-automorphisms defined by $h \cdot x_g = x_{hg}$ for any $g, h \in G$. Noether's problem asks whether the fixed subfield $k(G) := k(x_g : g \in G)^G$ is k-rational, that is, purely transcendental over k. If $\widehat{S}_n$ is the double cover of the symmetric group S_n, in which the liftings of transpositions and products of disjoint transpositions are of order 4, Serre shows that $\mathbb{Q}(\widehat{S}_4)$ and $\mathbb{Q}(\widehat{S}_5)$ are not $\mathbb{Q}$-rational. We will prove that if k is a field such that char $k \neq 2, 3$, and $k(\zeta_8)$ is a cyclic extension of k, then $k(\widehat{S}_4)$ is k-rational. If it is assumed furthermore that char $k = 0$, then $k(\widehat{S}_5)$ is also k-rational.

1. Introduction

Let k be a field, and L be a finitely generated field extension of k. L is called k-rational (or rational over k) if L is purely transcendental over k; that is, L is isomorphic to some rational function field over k. L is called stably k-rational if $L(y_1, \dots, y_m)$ is k-rational for some $y_1, \dots, y_m$ that are algebraically independent over L. L is called k-unirational if L is k-isomorphic to a subfield of some k-rational field extension of k. It is easy to see that

$$k\text{-rational} \Rightarrow \text{stably } k\text{-rational} \Rightarrow k\text{-unirational}.$$

A notion of retract rationality was introduced in [Saltman 1984] (see also [Kang 2012]). It is known that if k is an infinite field, then

$$\text{stably } k\text{-rational} \Rightarrow \text{retract } k\text{-rational} \Rightarrow k\text{-unirational}.$$

Let k be a field and G a finite group. Let G act on the rational function field $k(x_g : g \in G)$ by k-automorphisms defined by $h \cdot x_g = x_{hg}$ for any $g, h \in G$.

Both authors were partially supported by the National Center for Theoretic Sciences (Taipei Office). The work of this paper was finished when the second-named author visited National Taiwan University under the support of the National Center for Theoretic Sciences (Taipei Office).
MSC2010: primary 14E08, 14M20; secondary 12F12, 13A50.
Keywords: Noether's problem, rationality problem, binary octahedral groups.

Denote by $k(G)$ the fixed subfield, that is, $k(G) = k(x_g : g \in G)^G$. Noether's problem asks under what conditions is the field $k(G)$ k-rational.

Noether's problem is related to the inverse Galois problem and the existence of generic G-Galois extensions over k. For the details, see Swan's survey paper [Swan 1983]. The purpose of this paper is to study Noether's problem for some double covers of the symmetric group S_n.

It is known that there are four different double covers of S_n when $n \geq 4$, that is, groups G that fit into a short exact sequence $1 \to C_2 \to G \to S_n \to 1$; see, for example, [Serre 1984, p. 653].

Definition 1.1 [Garibaldi et al. 2003, pp. 58, 90; Hoffman and Humphreys 1992, p. 18; Karpilovsky 1985, pp. 177–181]. Let $C_2 = \{\pm 1\}$ be the cyclic group of order 2. When $n \geq 4$, the group $\widehat{S}_n$ is the unique central extension of S_n by C_2, that is,

$$1 \to C_2 \to \widehat{S}_n \to S_n \to 1,$$

satisfying the condition that the transpositions and the product of two disjoint transpositions in S_n lift to elements of order 4 in $\widehat{S}_n$. On the other hand, the group $\widetilde{S}_n$ is the central extension

$$1 \to C_2 \to \widetilde{S}_n \to S_n \to 1,$$

such that a transposition in S_n lifts to an element of order 2 of $\widetilde{S}_n$, but a product of two disjoint transpositions in S_n lifts to an element of order 4.

Note that we follow the notation of $\widehat{S}_n$ and $\widetilde{S}_n$ adopted by Serre.

Theorem 1.2 (Serre [Garibaldi et al. 2003, p. 90]). *Both $\mathbb{Q}(\widehat{S}_4)$ and $\mathbb{Q}(\widehat{S}_5)$ are not retract $\mathbb{Q}$-rational. In particular, they are not $\mathbb{Q}$-rational.*

Serre proves this using cohomological invariants and trace forms over $\mathbb{Q}$ — the e-invariant method, in short. In pp. 89–90 of the same book, he proves that $\mathrm{Rat}(G/\mathbb{Q})$ is false for $G = \widehat{S}_4$ and $\widehat{S}_5$. Actually he proves a bit more. From Serre's proof it is easy to find that $\mathbb{Q}(\widehat{S}_4)$ and $\mathbb{Q}(\widehat{S}_5)$ are not retract $\mathbb{Q}$-rational (see [Kang 2012, Section 1] for the relationship of the property $\mathrm{Rat}(G/k)$ and the retract k-rationality of $k(G)$). This is the reason why we formulate Serre's theorem in the version above. In fact, Theorem 1.2 can be perceived also from Serre's own remark in [Garibaldi et al. 2003, p. 13, Remark 5.8].

We don't know whether Theorem 1.2 is valid for fields k other than the field $\mathbb{Q}$; for example, the field k satisfying the condition that $k(\zeta_8)$ is not cyclic over k. In fact, in a private communication, Serre told us that the e-invariant method remains valid (under the assumption that $k(\zeta_8)$ is not cyclic over k) if k is an algebraic number field of odd degree over $\mathbb{Q}$, or if $k = \mathbb{Q}(\sqrt{n})$, where $n \equiv 1 \pmod 8$. However, if $k = \mathbb{Q}(x, y)$ with $x^2 + y^2 = -1$, the assumption that $k(\zeta_8)$

is not cyclic over k is valid while the e-invariant method doesn't work any more [Serre 2011].

On the other hand, we have:

Theorem 1.3 [Plans 2007; 2009]. (1) *For any field* k, $k(\widetilde{S}_4)$ *is* k*-rational. Thus, if* k *is a field with* $\operatorname{char} k = 0$, $k(\widetilde{S}_5)$ *is also* k*-rational.*

(2) *For any infinite field* k *with* $\operatorname{char} k \neq 2$ *such that* $\sqrt{-1} \in k$, *both* $k(\widehat{S}_4)$ *and* $k(\widehat{S}_5)$ *are* k*-rational.*

The main result of this article is the following rationality criterion for $k(\widehat{S}_4)$ and $k(\widehat{S}_5)$.

Theorem 1.4. *Let* k *be a field with* $\operatorname{char} k \neq 2$ *or* 3, *and* ζ_8 *be a primitive eighth root of unity in some extension field of* k. *If* $k(\zeta_8)$ *is a cyclic extension of* k, *then* $k(\widehat{S}_4)$ *is* k*-rational*; *if it is assumed furthermore that* $\operatorname{char} k = 0$, *then* $k(\widehat{S}_5)$ *is also* k*-rational.*

When k is a field with $\operatorname{char} k = p > 0$ and $p \neq 2$, the assumption that $k(\zeta_8)$ is a cyclic extension of k is satisfied automatically. Thus $k(\widehat{S}_4)$ is k-rational provided that k is any field with $\operatorname{char} k \neq 2$ or 3.

Besides the groups $\widehat{S}_4$ and $\widehat{S}_5$, Serre shows that $\mathbb{Q}(G)$ is not retract $\mathbb{Q}$-rational if G is any one of the groups $\mathrm{SL}_2(\mathbb{F}_7)$, $\mathrm{SL}_2(\mathbb{F}_9)$ and the generalized quaternion group of order 16; see [Garibaldi et al. 2003, p. 90, Example 33.27]. In case G is the generalized quaternion group of order 16 and $k(\zeta_8)$ is cyclic over k, it is known that $k(G)$ is k-rational [Kang 2005]. We don't know whether analogous results as Theorem 1.4 are valid when the groups are $\mathrm{SL}_2(\mathbb{F}_7)$ and $\mathrm{SL}_2(\mathbb{F}_9)$.

The main idea of the proof of Theorem 1.4 is to use the method of Galois descent, namely we first enlarge the field k to $k(\zeta_8)$, solve the rationality of $k(\zeta_8)(\widehat{S}_4)$, and then descend the ground field to k.

The proof that $k(\zeta_8)(\widehat{S}_4)$ is $k(\zeta_8)$-rational requires at least two techniques. In order to decrease the number of variables (by applying Theorem 2.2), we will construct a 4-dimensional faithful representation V of $\widehat{S}_4$ defined over the field k. It seems the representation and the idea to find it are not well-known. Once we have this representation, we adjoin ζ_8 to the field k and write $\pi = \mathrm{Gal}(k(\zeta_8)/k)$. We will prove that $k(\zeta_8)(V)^{\langle \widehat{S}_4, \pi\rangle}$ is k-rational.

The rationality problem of $k(\zeta_8)(V)^{\langle \widehat{S}_4, \pi\rangle}$ is not straightforward. In several steps of computations we use computers to facilitate the process of symbolic computation. However, we emphasize that computers play only a minor role; we don't use particular codes of data bases such as GAP.

On the other hand, we point out that the first several steps in proving that $k(\zeta_8)(V)^{\langle \widehat{S}_4, \pi\rangle}$ is k-rational are rather similar to those in [Kang and Zhou 2012, Section 5]. This seems unsurprising because the group $\widetilde{S}_4$ considered in [Kang and Zhou 2012, Section 5] and the group $\widehat{S}_4$ here have a common subgroup $\widetilde{A}_4$.

For the rationality problem of $k(\widehat{S}_5)$, we apply Theorem 2.5 of Plans, which asserts that $k(\widehat{S}_5)$ is a rational extension of $k(\widehat{S}_4)$, whence the result.

We organize this paper as follows. We recall some preliminaries in Section 2 that will be used in the proof of Theorem 1.4. In Section 3, several low-dimensional faithful representations of $\widehat{S}_4$ over a field k with $\operatorname{char} k \neq 2$ will be constructed (the reader may find another explicit construction in [Karpilovsky 1985, p. 177–179]). Theorem 1.4 will be proved in Section 4. In Section 5 we will consider the rationality problem of $k(G_n)$ (see Definition 5.1 for the group G_n).

Throughout this article, whenever we write $k(x_1, x_2, x_3, x_4)$ or $k(x, y)$ without explanation, it is understood that it is a rational function field over k. We will denote by ζ_8 (or simply by ζ) a primitive eighth root of unity.

2. Preliminaries

We recall several results that will be used in tackling the rationality problem.

Theorem 2.1 [Ahmad et al. 2000, Theorem 3.1]. *Let L be any field, $L(x)$ the rational function field of one variable over L and G a finite group acting on $L(x)$. Suppose that for any $\sigma \in G$, $\sigma(L) \subset L$ and $\sigma(x) = a_\sigma \cdot x + b_\sigma$, where $a_\sigma, b_\sigma \in L$ and $a_\sigma \neq 0$. Then $L(x)^G = L^G(f)$ for some polynomial $f \in L[x]$. In fact, if $m = \min\{\deg g(x) : g(x) \in L[x]^G, \deg g(x) \geq 1\}$, any polynomial $f \in L[x]^G$ with $\deg f = m$ satisfies the property that $L(x)^G = L^G(f)$.*

Theorem 2.2 [Hajja and Kang 1995, Theorem 1]. *Let G be a finite group acting on the rational function field $L(x_1, \dots, x_n)$ of n variables over a field L. Suppose that:*

(i) *For any $\sigma \in G$, $\sigma(L) \subset L$.*

(ii) *The restriction of the action of G to L is faithful.*

(iii) *For any $\sigma \in G$,*

$$\begin{pmatrix} \sigma(x_1) \\ \sigma(x_2) \\ \vdots \\ \sigma(x_n) \end{pmatrix} = A(\sigma) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + B(\sigma),$$

where $A(\sigma) \in \mathrm{GL_n(L)}$ and $B(\sigma)$ is a $n \times 1$ matrix over L.

Then there exist elements $z_1, \dots, z_n \in L(x_1, \dots, x_n)$ that are algebraically independent over L and satisfy $L(x_1, \dots, x_n) = L(z_1, \dots, z_n)$ and $\sigma(z_i) = z_i$ for any $\sigma \in G$ and $1 \leq i \leq n$.

Theorem 2.3 [Yamasaki 2009]. *Let k be a field with* $\operatorname{char} k \neq 2$, *let $a \in k\backslash\{0\}$, and define a k-automorphism σ of the rational function field $k(x, y)$ by $\sigma(x) = a/x$ and $\sigma(y) = a/y$. Then $k(x, y)^{\langle\sigma\rangle} = k(u, v)$, where $u = (x - y)/(a - xy)$ and $v = (x + y)/(a + xy)$.*

Theorem 2.4 [Masuda 1955, Theorem 3; Hoshi and Kang 2010, Theorem 2.2]. *Let k be a field and let σ be the k-automorphism of the rational function field $k(x, y, z)$ defined by $\sigma : x \mapsto y \mapsto z \mapsto x$. Then $k(x, y, z)^{\langle\sigma\rangle} = k(s_1, u, v) = k(s_3, u, v)$, where s_1, s_2, s_3 are the elementary symmetric functions of degree one, two and three in x, y, z and u and v are defined by*

$$u = \frac{x^2y + y^2z + z^2x - 3xyz}{x^2 + y^2 + z^2 - xy - yz - zx} \quad \textit{and} \quad v = \frac{xy^2 + yz^2 + zx^2 - 3xyz}{x^2 + y^2 + z^2 - xy - yz - zx}.$$

Theorem 2.5 [Plans 2009, Theorem 11]. *Let $n \geq 5$ be an odd integer and let k be a field with* $\operatorname{char} k = 0$. *Then $k(\widehat{S}_n)$ is rational over $k(\widehat{S}_{n-1})$.*

Theorem 2.6 [Kang and Plans 2009, Theorem 1.9]. *Let k be a field and let G_1 and G_2 be two finite groups. If both $k(G_1)$ and $k(G_2)$ are k-rational, so is $k(G_1 \times G_2)$.*

3. Faithful representations of $\widehat{S}_4$

In this and the next section, the field k we consider is of $\operatorname{char} k \neq 2$ or 3. We will denote by $\zeta_8 = (1 + \sqrt{-1})/\sqrt{2}$ a primitive eighth root of unity.

In [Springer 1977, p. 92] a generating set of $\widehat{S}_4$ is given (where the group is called the binary octahedral group): $\widehat{S}_4 = \langle a', b, c\rangle$ with relations $a'^8 = b^4 = c^6 = 1$, $ba'b^{-1} = a'^{-1}$, $cbc^{-1} = a'^2$ and $(a'c)^2 = -a'^2 b$ (here -1 is the element that is equal to $a'^4 = b^2 = c^3$). Note that we have a short exact sequence of groups

$$1 \to \{\pm 1\} \to \widehat{S}_4 \xrightarrow{p} S_4 \to 1,$$

and that $p(a') = (1, 2, 3, 4)$, $p(b) = (1, 4)(2, 3)$ and $p(c) = (1, 2, 3)$. Note that $p(ba') = (1, 4)(2, 3)(1, 2, 3, 4) = (1, 3)$.

If $\zeta_8 \in k$, a faithful 2-dimensional representation $\Phi : \widehat{S}_4 \to \mathrm{GL}_2(\mathrm{k})$ is given in [Springer 1977, p. 92] as follows (we write $\zeta = \zeta_8$),

$$\text{(3-1)}\quad \Phi(a') = \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^7 \end{pmatrix}, \quad \Phi(b) = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \quad \Phi(c) = \frac{1}{\sqrt{2}} \begin{pmatrix} \zeta^7 & \zeta^7 \\ \zeta^5 & \zeta \end{pmatrix}.$$

Suppose that $\sqrt{2} \in k$ (but not necessarily that $\sqrt{-1} \in k$). We may obtain a 4-dimensional representation $\widehat{S}_4 \to \mathrm{GL}_4(\mathrm{k})$ by making in (3-1) the substitutions

$$\sqrt{-1} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \alpha \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix},$$

where k_0 is the prime field of k and $\alpha \in k_0(\sqrt{2})$. This process is an easy application of Weil's restriction [Weil 1956; Voskresenskii 1998, p. 38]. Thus we get

$$\text{(3-2)}\quad a' \mapsto \frac{1}{\sqrt{2}}\left(\begin{array}{cc|cc} 1 & -1 & & \\ 1 & 1 & & \\ \hline & & 1 & 1 \\ & & -1 & 1 \end{array}\right),\ b \mapsto \begin{pmatrix} & & & -1 \\ & & 1 & \\ & -1 & & \\ 1 & & & \end{pmatrix},\ c \mapsto \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}.$$

Similarly, when $\sqrt{-2}$ is in k (but possibly $\sqrt{-1}$ is not in k), write $\sqrt{-2} = \sqrt{-1} \cdot \sqrt{2}$. Thus represent $\sqrt{2}$ as $-\sqrt{-1} \cdot \sqrt{-2}$ and $\zeta = (1 + \sqrt{-1})/\sqrt{2}$ becomes $\sqrt{-2}(1 - \sqrt{-1})/2$. Make in (3-1) the substitutions

$$\sqrt{-1} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \alpha \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix},$$

where k_0 is the prime field of k and $\alpha \in k_0(\sqrt{-2})$. We get

$$\text{(3-3)}\quad a' \mapsto \frac{\sqrt{-2}}{2}\left(\begin{array}{cc|cc} 1 & 1 & & \\ -1 & 1 & & \\ \hline & & -1 & 1 \\ & & -1 & -1 \end{array}\right),\ b \mapsto \begin{pmatrix} & & & -1 \\ & & 1 & \\ & -1 & & \\ 1 & & & \end{pmatrix},\ c \mapsto \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}.$$

The same way, if $\sqrt{-1} \in k$ (but possibly $\sqrt{2} \notin k$), make in (3-1) the substitutions

$$\sqrt{2} \mapsto \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \alpha \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix},$$

where k_0 is the prime filed of k and $\alpha \in k_0(\sqrt{-1})$. We get

$$a' \mapsto \left(\begin{array}{cc|cc} 0 & 1+\sqrt{-1} & & \\ \frac{1+\sqrt{-1}}{2} & 0 & & \\ \hline & & 0 & 1-\sqrt{-1} \\ & & \frac{1-\sqrt{-1}}{2} & 0 \end{array}\right),$$

(3-4)

$$b \mapsto \left(\begin{array}{cc|cc} & & \sqrt{-1} & 0 \\ & & 0 & \sqrt{-1} \\ \hline \sqrt{-1} & 0 & & \\ 0 & \sqrt{-1} & & \end{array}\right),\quad c \mapsto \begin{pmatrix} \frac{1-\sqrt{-1}}{2} & 0 & \frac{1-\sqrt{-1}}{2} & 0 \\ 0 & \frac{1-\sqrt{-1}}{2} & 0 & \frac{1-\sqrt{-1}}{2} \\ \frac{-1-\sqrt{-1}}{2} & 0 & \frac{1+\sqrt{-1}}{2} & 0 \\ 0 & \frac{-1-\sqrt{-1}}{2} & 0 & \frac{1+\sqrt{-1}}{2} \end{pmatrix}.$$

Finally, from (3-2) we may get a faithful 8-dimensional representation of $\widehat{S}_4$ into $\mathrm{GL}_8(\mathrm{k}_0)$, where k_0 is the prime field of k. Explicitly, make in (3-2) the substitutions

$$\sqrt{2} \mapsto \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \alpha \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix},$$

where $\alpha \in k_0$. We get

$$a' \mapsto \frac{1}{2}\left(\begin{array}{cccc|cccc} 0 & 2 & 0 & -2 & & & & \\ 1 & 0 & -1 & 0 & & & & \\ 0 & 2 & 0 & 2 & & & & \\ 1 & 0 & 1 & 0 & & & & \\ \hline & & & & 0 & 2 & 0 & 2 \\ & & & & 1 & 0 & 1 & 0 \\ & & & & 0 & -2 & 0 & 2 \\ & & & & -1 & 0 & 1 & 0 \end{array}\right),$$

$$b \mapsto \left(\begin{array}{cccc|cccc} & & & & & & -1 & 0 \\ & & & & & & 0 & -1 \\ & & & & 1 & 0 & & \\ & & & & 0 & 1 & & \\ \hline & & -1 & 0 & & & & \\ & & 0 & -1 & & & & \\ 1 & 0 & & & & & & \\ 0 & 1 & & & & & & \end{array}\right), \tag{3-5}$$

$$c \mapsto \frac{1}{2}\left(\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ \hline -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \\ -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \end{array}\right).$$

4. Proof of Theorem 1.4

By Theorem 2.5, in case $\operatorname{char} k = 0$ and it is known that $k(\widehat{S}_4)$ is k-rational, it follows immediately that $k(\widehat{S}_5)$ is also k-rational. Hence, in proving Theorem 1.4, it suffices to prove the rationality of $k(\widehat{S}_4)$.

By assumption, $k(\zeta_8)$ is a cyclic extension of k. Hence at least one of $\sqrt{-1}$, $\sqrt{2}$ or $\sqrt{-2}$ belongs to k.

Case 1: $\zeta_8 \in k$. Since $\operatorname{char} k \neq 2$ or 3, the group algebra $k[\widehat{S}_4]$ is semisimple. Hence the 2-dimensional faithful representation provided by Equation (3-1) can be embedded into the regular representation whose dual space is $V_{\text{reg}} = \oplus_{g \in \widehat{S}_4} k \cdot x(g)$, where $\widehat{S}_4$ acts on V_{reg} by $h \cdot x(g) = x(hg)$ for any $g, h \in \widehat{S}_4$. By Theorem 2.2, we find that $k(\widehat{S}_4) = k(x(g) : g \in \widehat{S}_4)^{\widehat{S}_4}$ is rational over $k(x, y)^{\widehat{S}_4}$, where the actions given by Equation (3-1) are

$$\begin{aligned} a' &: x \mapsto \zeta x, & y &\mapsto \zeta^7 y, \\ b &: x \mapsto \sqrt{-1}y, & y &\mapsto \sqrt{-1}x, \\ c &: x \mapsto (\zeta^7 x + \zeta^5 y)/\sqrt{2}, & y &\mapsto (\zeta^7 x + \zeta y)/\sqrt{2}. \end{aligned}$$

Set $z = x/y$. Then $k(x, y) = k(z, x)$. By applying Theorem 2.1 we get that $k(z, x)^{\widehat{S}_4} = k(z)^{\widehat{S}_4}(t)$ for some element t fixed by $\widehat{S}_4$. The field $k(z)^{\widehat{S}_4}$ is k-rational by Lüroth's theorem. Hence $k(z, x)^{\widehat{S}_4}$ and $k(\widehat{S}_4)$ are k-rational.

Case 2: $\sqrt{2} \in k$ but $\sqrt{-1} \notin k$. We will use the 4-dimensional faithful representation of $\widehat{S}_4$ over k provided by Equation (3-2). This representation provides an action of $\widehat{S}_4$ on $k(x_1, x_2, x_3, x_4)$ given by

$$\begin{aligned} a' &: x_1 \mapsto (x_1 + x_2)/\sqrt{2}, & x_2 &\mapsto (-x_1 + x_2)/\sqrt{2}, \\ &\quad x_3 \mapsto (x_3 - x_4)/\sqrt{2}, & x_4 &\mapsto (x_3 + x_4)/\sqrt{2}, \\ b &: x_1 \mapsto x_4 \mapsto -x_1, & x_2 &\mapsto -x_3, \\ &\quad x_3 \mapsto x_2, \\ c &: x_1 \mapsto (x_1 - x_2 - x_3 - x_4)/2, & x_2 &\mapsto (x_1 + x_2 + x_3 - x_4)/2, \\ &\quad x_3 \mapsto (x_1 - x_2 + x_3 + x_4)/2, & x_4 &\mapsto (x_1 + x_2 - x_3 + x_4)/2. \end{aligned} \tag{4-1}$$

• Step 1. Apply Theorem 2.2 and use the arguments of Case 1. We find that $k(\widehat{S}_4)$ is rational over $k(x_1, x_2, x_3, x_4)^{\widehat{S}_4}$. It remains to show that $k(x_1, x_2, x_3, x_4)^{\widehat{S}_4}$ is k-rational.

• Step 2. Write $\pi = \mathrm{Gal}(k(\sqrt{-1})/k) = \langle\rho\rangle$, where $\rho(\sqrt{-1}) = -\sqrt{-1}$. Extend the actions of π and $\widehat{S}_4$ on $k(\sqrt{-1})$ and $k(x_1, x_2, x_3, x_4)$ to $k(\sqrt{-1})(x_1, x_2, x_3, x_4)$ by requiring that $\rho(x_i) = x_i$ for $1 \le i \le 4$ and $g(\sqrt{-1}) = \sqrt{-1}$ for all $g \in \widehat{S}_4$. It follows that

$$\begin{aligned} k(x_1, x_2, x_3, x_4)^{\widehat{S}_4} &= \{k(\sqrt{-1})(x_1, x_2, x_3, x_4)^{\langle\rho\rangle}\}^{\langle a', b, c\rangle} \\ &= k(\sqrt{-1})(x_1, x_2, x_3, x_4)^{\langle a', b, c, \rho\rangle}. \end{aligned}$$

Define $y_1, y_2, y_3, y_4 \in k(\sqrt{-1})(x_1, x_2, x_3, x_4)$ by

$$\begin{aligned} y_1 &= \sqrt{-1}x_1 + \sqrt{-1}x_2 - x_3 + x_4, & y_2 &= -\sqrt{-1}x_1 + \sqrt{-1}x_2 + x_3 + x_4, \\ y_3 &= x_1 - x_2 - \sqrt{-1}x_3 - \sqrt{-1}x_4, & y_4 &= x_1 + x_2 - \sqrt{-1}x_3 + \sqrt{-1}x_4. \end{aligned}$$

Then

$$k(\sqrt{-1})(x_1, x_2, x_3, x_4) = k(\sqrt{-1})(y_1, y_2, y_3, y_4)$$

and the actions in (4-1) become

$$
\begin{aligned}
a' &: y_1 \mapsto (y_1+y_2)/\sqrt{2}, && y_2 \mapsto (-y_1+y_2)/\sqrt{2},\\
&\quad y_3 \mapsto (y_3+y_4)/\sqrt{2}, && y_4 \mapsto (-y_3+y_4)/\sqrt{2},\\
b &: y_1 \mapsto \sqrt{-1}y_1, && y_2 \mapsto -\sqrt{-1}y_2,\\
&\quad y_3 \mapsto \sqrt{-1}y_3, && y_4 \mapsto -\sqrt{-1}y_4,\\
c &: y_1 \mapsto \frac{y_1-\sqrt{-1}y_2}{1+\sqrt{-1}}, && y_2 \mapsto \frac{y_1+\sqrt{-1}y_2}{1+\sqrt{-1}},\\
&\quad y_3 \mapsto \frac{y_3-\sqrt{-1}y_4}{1+\sqrt{-1}}, && y_4 \mapsto \frac{y_3+\sqrt{-1}y_4}{1+\sqrt{-1}},\\
\rho &: y_1 \mapsto -\sqrt{-1}y_4, && y_2 \mapsto \sqrt{-1}y_3,\\
&\quad y_3 \mapsto \sqrt{-1}y_2, && y_4 \mapsto -\sqrt{-1}y_1.
\end{aligned}
\tag{4-2}
$$

Note that the action of a'^2 is given by

$$a'^2 : y_1 \mapsto y_2 \mapsto -y_1, \quad y_3 \mapsto y_4 \mapsto -y_3.$$

The reader might find interesting to compare the actions in (4-2) with those in [Kang and Zhou 2012, Section 4]. It turns out that the formulae for b, a'^2, c^2 are completely the same as those for $\lambda_1, \lambda_2, \sigma$ in [Kang and Zhou 2012, Formula (4.3)]. As mentioned before, both the subgroups $\langle b, a'^2, c^2\rangle$ and $\langle \lambda_1, \lambda_2, \sigma\rangle$ are isomorphic to $\widetilde{A}_4$ (where $\widetilde{A}_4 = p^{-1}(A_4)$ in the notation of Section 3) as abstract groups.

• Step 3. Define $z_1 = y_1/y_2$, $z_2 = y_3/y_4$, $z_3 = y_1/y_3$. By Theorem 2.1, we find that

$$
\begin{aligned}
k(\sqrt{-1})(y_1,y_2,y_3,y_4)^{\langle \widehat{S}_4,\pi\rangle} &= k(\sqrt{-1})(z_1,z_2,z_3)(y_4)^{\langle \widehat{S}_4,\pi\rangle}\\
&= k(\sqrt{-1})(z_1,z_2,z_3)^{\langle \widehat{S}_4,\pi\rangle}(z_0),
\end{aligned}
$$

where z_0 is fixed by the actions of $\widehat{S}_4$ and π. There remains to show the k-rationality of $k(\sqrt{-1})(z_1,z_2,z_3)^{\langle \widehat{S}_4,\pi\rangle}$ is

Before we find $k(\sqrt{-1})(z_1,z_2,z_3)^{\langle \widehat{S}_4,\pi\rangle}$, we will find $k(\sqrt{-1})(z_1,z_2,z_3)^{\langle b,a'^2\rangle}$. The method is the same as in Steps 3 and 4 in [Kang and Zhou 2012, Section 4]. We will write down the details for the convenience of the reader.

Define $u_1 = z_1/z_2$, $u_2 = z_1z_2$, $u_3 = z_3$. Then

$$k(\sqrt{-1})(z_1,z_2,z_3)^{\langle b\rangle} = k(\sqrt{-1})(u_1,u_2,u_3).$$

The action of a'^2 is given by

$$a'^2 : u_1 \mapsto 1/u_1, \quad u_2 \mapsto 1/u_2, \quad u_3 \mapsto u_3/u_1.$$

Define

$$v_1=\frac{u_1-u_2}{1-u_1u_2},\quad v_2=\frac{u_1+u_2}{1+u_1u_2},\quad v_3=u_3\Big(1+\frac{1}{u_1}\Big).$$

Then

$$k(\sqrt{-1})(u_1,u_2,u_3)^{\langle a'^2\rangle}=k(\sqrt{-1})(u_1,u_2,v_3)^{\langle a'^2\rangle}=k(\sqrt{-1})(v_1,v_2,v_3)$$

by Theorem 2.3 (note that $a'^2(v_3)=v_3$). In summary,

$$k(\sqrt{-1})(z_1,z_2,z_3)^{\langle b,a'^2\rangle}=k(\sqrt{-1})(v_1,v_2,v_3).$$

• Step 4. The action of c on v_1, v_2, v_3 is given by

$$c: v_1\mapsto 1/v_2,\quad v_2\mapsto v_1/v_2,\quad v_3\mapsto v_3(v_1+v_2)/[v_2(1+v_1)].$$

Define $X_3=v_3(1+v_1+v_2)/[(1+v_1)(1+v_2)]$. Then $c(X_3)=X_3$ and

$$k(\sqrt{-1})(v_1,v_2,v_3)=k(\sqrt{-1})(v_1,v_2,X_3).$$

Thus we may apply Theorem 2.4 (regarding v_1, $1/v_2$, v_2/v_1 as x, y, z in its statement). More precisely, define

$$X_1=(v_1^3v_2^3+v_1^3+v_2^3-3v_1^2v_2^2)/(v_1^4v_2^2+v_2^4+v_1^2-v_1^2v_2^3-v_1v_2^2-v_1^3v_2),$$
$$X_2=(v_1v_2^4+v_1v_2+v_1^4v_2-3v_1^2v_2^2)/(v_1^4v_2^2+v_2^4+v_1^2-v_1^2v_2^3-v_1v_2^2-v_1^3v_2).$$

By Theorem 2.4 we get $k(\sqrt{-1})(v_1,v_2,X_3)^{\langle c\rangle}=k(\sqrt{-1})(X_1,X_2,X_3)$.

• Step 5. With the aid of computers, we find that the actions of a' and ρ on X_1, X_2, X_3 are given by

$$a': X_1\mapsto\frac{X_1}{X_1^2-X_1X_2+X_2^2},\quad X_2\mapsto\frac{X_2}{X_1^2-X_1X_2+X_2^2},\quad X_3\mapsto X_3,$$
$$\rho\ : X_1\mapsto\frac{X_2}{X_1^2-X_1X_2+X_2^2},\quad X_2\mapsto\frac{X_1}{X_1^2-X_1X_2+X_2^2},\quad X_3\mapsto -2A/X_3,$$

where $A=g_1g_2g_3^{-1}$ and

$$g_1=(1+X_1)^2-X_2(1+X_1)+X_2^2,$$
$$g_2=(1+X_2)^2-X_1(1+X_2)+X_1^2,$$
$$g_3=1+X_1+X_2+X_1^3+X_2^3+X_1X_2(3X_1X_2-2X_1^2-2X_2^2+2)+X_1^4+X_2^4.$$

Note that $\rho(g_1)=g_2/(X_1^2-X_1X_2+X_2^2)$ and $a'(g_1)=g_1/(X_1^2-X_1X_2+X_2^2)$. Define $Y_1=X_1/X_2$, $Y_2=X_1$, $Y_3=X_1X_3/g_1$. We find that

$$a': Y_1\mapsto Y_1,\quad Y_2\mapsto Y_1^2/(Y_2(1-Y_1+Y_1^2)),\quad Y_3\mapsto Y_3.$$

Thus

$$k(\sqrt{-1})(X_1, X_2, X_3)^{\langle a' \rangle} = k(\sqrt{-1})(Y_1, Y_2, Y_3)^{\langle a' \rangle} = k(\sqrt{-1})(Z_1, Z_2, Z_3),$$

where $Z_1 = Y_1$, $Z_2 = Y_2 + a'(Y_2)$, $Z_3 = Y_3$.

• Step 6. Using computers, we find that the action of ρ is given by

$$\rho : Z_1 \mapsto 1/Z_1, \quad Z_2 \mapsto Z_2/Z_1, \quad Z_3 \mapsto -2Z_1^3/(A'Z_3),$$

where A' is defined to be

$$-2Z_1^2+Z_1Z_2+Z_2^2+4Z_1^3-2Z_1Z_2^2-2Z_1^4+3Z_1^2Z_2^2+Z_1^4Z_2-2Z_1^3Z_2^2+Z_1^4Z_2^2.$$

Define $U_1 = Z_2+\rho(Z_2)$, $U_2 = \sqrt{-1}(Z_2-\rho(Z_2))$, $U_3 = Z_3+\rho(Z_3)$ and $U_4 = \sqrt{-1}(Z_3 - \rho(Z_3))$. We see that $k(\sqrt{-1})(Z_1, Z_2, Z_3)^{\langle \rho \rangle} = k(U_1, U_2, U_3, U_4)$ with a relation

$$U_3^2 + U_4^2 + 32(U_1^2 + U_2^2)/B = 0,$$

where $B = (U_1^2 - 3U_2^2)^2 + 4U_1(U_1^2 - 3U_2^2) + 32U_2^2$.

Dividing this relation by $16(U_1^2 + U_2^2)^2/B^2$, we get

$$\big(BU_3/(4U_1^2 + 4U_2^2)\big)^2 + \big(BU_4/(4U_1^2 + 4U_2^2)\big)^2 + 2B/(U_1^2 + U_2^2) = 0.$$

Multiply this relation by $U_1^2 + U_2^2$ and use the identity

$$(\alpha^2 + \beta^2)(\gamma^2 + \delta^2) = (\alpha\delta + \beta\gamma)^2 + (\alpha\gamma - \beta\delta)^2$$

to obtain the simplification

$$V_3^2 + V_4^2 + 2B = 0, \tag{4-3}$$

where

$$V_3 = B\frac{U_1U_3 + U_2U_4}{4U_1^2 + 4U_2^2} \quad \text{and} \quad V_4 = B\frac{U_1U_4 - U_2U_3}{4U_1^2 + 4U_2^2}.$$

Note that $k(U_1, U_2, U_3, U_4) = K(U_1, U_2, V_3, V_4)$.

Define $w_1 = 8U_1/(U_1^2-3U_2^2)$, $w_2 = 8U_2/(U_1^2-3U_2^2)$, $w_3 = V_3/(U_1^2-3U_2^2)$, $w_4 = V_4/(U_1^2 - 3U_2^2)$. Then $k(U_1, U_2, V_3, V_4) = k(w_1, w_2, w_3, w_4)$ and the relation (4-3) becomes

$$w_3^2 + w_4^2 + 2 + w_1 + w_2^2 = 0.$$

Hence $w_1 \in k(w_2, w_3, w_4)$. Thus $k(\sqrt{-1})(Z_1, Z_2, Z_3)^{\langle \rho \rangle} = k(w_2, w_3, w_4)$ is k-rational.

Case 3: $\sqrt{-2} \in k$ *but* $\sqrt{-1} \notin k$. We use the 4-dimensional faithful representation of $\widehat{S}_4$ over k provided by (3-3). This representation provides an action of $\widehat{S}_4$ on $k(x_1, x_2, x_3, x_4)$ given by

$$\begin{aligned}
a' &: x_1 \mapsto \sqrt{-2}(x_1 - x_2)/2, & x_2 &\mapsto \sqrt{-2}(x_1 + x_2)/2,\\
&\quad x_3 \mapsto \sqrt{-2}(-x_3 - x_4)/2, & x_4 &\mapsto \sqrt{-2}(x_3 - x_4)/2,\\
b &: x_1 \mapsto x_4 \mapsto -x_1, & x_2 &\mapsto -x_3,\\
&\quad x_3 \mapsto x_2, & &\\
c &: x_1 \mapsto (x_1 - x_2 - x_3 - x_4)/2, & x_2 &\mapsto (x_1 + x_2 + x_3 - x_4)/2,\\
&\quad x_3 \mapsto (x_1 - x_2 + x_3 + x_4)/2, & x_4 &\mapsto (x_1 + x_2 - x_3 + x_4)/2.
\end{aligned}$$

The proof of this case is very similar to that of Case 2.

• Step 1. Apply Theorem 2.2. We see that $k(\widehat{S}_4)$ is rational over $k(x_1, x_2, x_3, x_4)^{\widehat{S}_4}$. Hence the proof is reduced to proving that $k(x_1, x_2, x_3, x_4)^{\widehat{S}_4}$ is k-rational.

• Step 2. Write $\pi = \mathrm{Gal}(k(\sqrt{-1})/k) = \langle \rho \rangle$, where $\rho(\sqrt{-1}) = -\sqrt{-1}$. Extend the actions of π and $\widehat{S}_4$ to $k(\sqrt{-1})(x_1, x_2, x_3, x_4)$ as in Step 2 of Case 2. We find that

$$k(x_1, x_2, x_3, x_4)^{\widehat{S}_4} = k(\sqrt{-1})(x_1, x_2, x_3, x_4)^{\langle a', b, c, \rho \rangle}.$$

Define $y_1, y_2, y_3, y_4 \in k(\sqrt{-1})(x_1, x_2, x_3, x_4)$ by

$$\begin{aligned}
y_1 &= -x_1 - \sqrt{-1}x_2 + x_3 + \sqrt{-1}x_4, & y_2 &= \sqrt{-1}x_1 - x_2 + \sqrt{-1}x_3 - x_4,\\
y_3 &= x_1 - \sqrt{-1}x_2 + x_3 - \sqrt{-1}x_4, & y_4 &= \sqrt{-1}x_1 + x_2 - \sqrt{-1}x_3 - x_4.
\end{aligned}$$

We get $k(\sqrt{-1})(x_1, x_2, x_3, x_4) = k(\sqrt{-1})(y_1, y_2, y_3, y_4)$ and the actions are

$$\begin{aligned}
a' &: y_1 \mapsto (-y_1 - y_2)/\sqrt{2}, & y_2 &\mapsto (y_1 - y_2)/\sqrt{2},\\
&\quad y_3 \mapsto (y_3 + y_4)/\sqrt{2}, & y_4 &\mapsto (-y_3 + y_4)/\sqrt{2},\\
b &: y_1 \mapsto \sqrt{-1}y_1, & y_2 &\mapsto -\sqrt{-1}y_2,\\
&\quad y_3 \mapsto \sqrt{-1}y_3, & y_4 &\mapsto -\sqrt{-1}y_4,\\
c &: y_1 \mapsto \frac{y_1 - \sqrt{-1}y_2}{1 + \sqrt{-1}}, & y_2 &\mapsto \frac{y_1 + \sqrt{-1}y_2}{1 + \sqrt{-1}},\\
&\quad y_3 \mapsto \frac{y_3 - \sqrt{-1}y_4}{1 + \sqrt{-1}}, & y_4 &\mapsto \frac{y_3 + \sqrt{-1}y_4}{1 + \sqrt{-1}},\\
\rho &: y_1 \mapsto \sqrt{-1}y_4, & y_2 &\mapsto -\sqrt{-1}y_3,\\
&\quad y_3 \mapsto -\sqrt{-1}y_2, & y_4 &\mapsto \sqrt{-1}y_1.
\end{aligned} \tag{4-4}$$

Note that the action of ${a'}^2$ is

$${a'}^2 : y_1 \mapsto y_2 \mapsto -y_1, \quad y_3 \mapsto y_4 \mapsto -y_3.$$

(Compare with (4-2) and (4-4).) The actions of a'^2, b, c in both cases are the same.

• Step 3. Define $z_1 = y_1/y_2$, $z_2 = y_3/y_4$, $z_3 = y_1/y_3$. As in Step 3 of Case 2, it suffices to prove that $k(\sqrt{-1})(z_1, z_2, z_3)^{\langle \widehat{S}_4, \pi\rangle}$ is k-rational.

Define $u_1, u_2, u_3, v_1, v_2, v_3, X_1, X_2, X_3$ by the same formulae as in Step 3 and Step 4 of Case 2. We find that $k(\sqrt{-1})(z_1, z_2, z_3)^{\langle b, a'^2, c\rangle} = k(\sqrt{-1})(X_1, X_2, X_3)$.

• Step 4. The actions of a', ρ on X_1, X_2, X_3 are slightly different from Step 5 of Case 2. In the present case, we have

$$\begin{aligned} a' &: X_1 \mapsto \frac{X_1}{X_1^2 - X_1X_2 + X_2^2}, \quad X_2 \mapsto \frac{X_2}{X_1^2 - X_1X_2 + X_2^2}, \quad X_3 \mapsto -X_3, \\ \rho &: X_1 \mapsto \frac{X_2}{X_1^2 - X_1X_2 + X_2^2}, \quad X_2 \mapsto \frac{X_1}{X_1^2 - X_1X_2 + X_2^2}, \quad X_3 \mapsto -2A/X_3, \end{aligned}$$

where $A = g_1 g_2 g_3^{-1}$ and

$$\begin{aligned} g_1 &= (1 + X_1)^2 - X_2(1 + X_1) + X_2^2, \\ g_2 &= (1 + X_2)^2 - X_1(1 + X_2) + X_1^2, \\ g_3 &= 1 + X_1 + X_2 + X_1^3 + X_2^3 + X_1X_2(3X_1X_2 - 2X_1^2 - 2X_2^2 + 2) + X_1^4 + X_2^4. \end{aligned}$$

Note that the action of ρ is the same as in Step 5 of Case 2.

Define $Y_1 = X_1/X_2$, $Y_2 = X_1$, $Y_3 = X_1X_3/g_1$. We get

$$a' : Y_1 \mapsto Y_1, \quad Y_2 \mapsto Y_1^2/\big(Y_2(1 - Y_1 + Y_1^2)\big), \quad Y_3 \mapsto -Y_3.$$

Thus $k(\sqrt{-1})(X_1, X_2, X_3)^{\langle a'\rangle} = k(\sqrt{-1})(Y_1, Y_2, Y_3)^{\langle a'\rangle} = k(\sqrt{-1})(Z_1, Z_2, Z_3)$, where $Z_1 = Y_1$, $Z_2 = Y_2 + a'(Y_2)$, $Z_3 = Y_3(Y_2 - a'(Y_2))$.

• Step 5. Using computers, we find that the action of ρ is given by

$$\rho : Z_1 \mapsto 1/Z_1, \quad Z_2 \mapsto Z_2/Z_1, \quad Z_3 \mapsto C/Z_3,$$

where C is defined to be

$$\frac{2Z_1^2(-4Z_1^2 + Z_2^2 - Z_1Z_2^2 + Z_1^2Z_2^2)/(1 - Z_1 + Z_1^2)}{-2Z_1^2 + Z_1Z_2 + Z_2^2 + 4Z_1^3 - 2Z_1Z_2^2 - 2Z_1^4 + 3Z_1^2Z_2^2 + Z_1^4Z_2 - 2Z_1^3Z_2^2 + Z_1^4Z_2^2}.$$

Define $U_1 = Z_2 + \rho(Z_2)$, $U_2 = \sqrt{-1}(Z_2 - \rho(Z_2))$, $U_3 = Z_3 + \rho(Z_3)$ and $U_4 = \sqrt{-1}(Z_3 - \rho(Z_3))$. We find that $k(\sqrt{-1})(Z_1, Z_2, Z_3)^{\langle\rho\rangle} = k(U_1, U_2, U_3, U_4)$ with a relation

$$U_3^2 + U_4^2 = 8(U_1^2 + U_2^2)^2(-16 + U_1^2 - 3U_2^2)/B(U_1^2 - 3U_2^2), \tag{4-5}$$

where $B = (U_1^2 - 3U_2^2)^2 + 4U_1(U_1^2 - 3U_2^2) + 32U_2^2$.

Note that the above formula of B is identically the same as that in Step 6 of Case 2. It remains to simplify the relation (4-5). Dividing both sides by $(U_1^2 + U_2^2)^2$, we get

$$\big(U_3/(U_1^2+U_2^2)\big)^2 + \big(U_4/(U_1^2+U_2^2)\big)^2 = 8(-16+U_1^2-3U_2^2)/B(U_1^2-3U_2^2).$$

Divide both sides of the above identity by $(2(U_1^2-3U_2^2)/B)^2$. We get a relation

$$V_3^2+V_4^2 = 2(1-V_1^2+3V_2^2)(1+V_1+2V_2^2), \tag{4-6}$$

where

$$V_1 = \frac{4U_1}{U_1^2-3U_2^2}, \quad V_3 = \frac{BU_3}{(U_1^2-3U_2^2)(2U_1^2+2U_2^2)},$$
$$V_2 = \frac{4U_2}{U_1^2-3U_2^2}, \quad V_4 = \frac{BU_4}{(U_1^2-3U_2^2)(2U_1^2+2U_2^2)}.$$

Note that $k(U_1,U_2,U_3,U_4) = k(V_1,V_2,V_3,V_4)$.

Define $w_1 = 1/(1+V_1)$, $w_2 = V_2/(1+V_1)$, $w_3 = V_3/(1+V_1)^2$, $w_4 = V_4/(1+V_1)^2$. We get $k(V_1,V_2,V_3,V_4) = k(w_1,w_2,w_3,w_3)$ and the relation (4-6) becomes

$$w_3^2+w_4^2 = 2(-1+2w_1+3w_2^2)(w_1+2w_2^2).$$

Divide the above identity by $(w_1+2w_2^2)^2$. We get

$$\big(w_3/(w_1+2w_2^2)\big)^2 + \big(w_4/(w_1+2w_2^2)\big)^2 = 2(-1+2w_1+3w_2^2)/(w_1+2w_2^2).$$

Since $2(-1+2w_1+3w_2^2)/(w_1+2w_2^2)$ is a "fractional linear transformation" of w_1 and it belongs to $k(w_2, w_3/(w_1+2w_2^2), w_4/(w_1+2w_2^2))$, we find that w_1 is in $k(w_2, w_3/(w_1+2w_2^2), w_4/(w_1+2w_2^2))$. Thus

$$k(w_1,w_2,w_3,w_4) = k(w_2, w_3/(w_1+2w_2^2), w_4/(w_1+2w_2^2)).$$

We find that $k(\sqrt{-1})(Z_1,Z_2,Z_3)^{\langle\rho\rangle}$ is k-rational.

Case 4: $\sqrt{-1} \in k$ *but* $\sqrt{2} \notin k$. This is similar to Cases 2 or 3, so the detailed proof is omitted. In the case $\operatorname{char} k = 0$, we may apply Plans' result, Theorem 1.3. □

5. Other double covers of S_n

In this section we consider the rationality problem of G_n, which is a double cover of the symmetric group and different from both $\widehat{S}_n$ and $\widetilde{S}_n$.

There are four double covers of the symmetric group S_n when $n \geq 4$. The trivial case is the split group $S_n \times C_2$. The rationality problem of the group $S_n \times C_2$ is easy because we may apply Theorem 2.6. It remains to consider the non split cases: they are $\widehat{S}_n$, $\widetilde{S}_n$, and the group G_n defined below.

Definition 5.1. For $n \geq 3$, consider the group G_n such that the short exact sequence $1 \to \{\pm 1\} \to G_n \overset{p}{\to} S_n \to 1$ is induced by the cup product $\varepsilon_n \cup \varepsilon_n \in H^2(S_n, \{\pm 1\})$, (see, for example, [Serre 1984, page 654]) where $\varepsilon_n : S_n \to \{\pm 1\}$ is the signed map, that is, $\varepsilon_n(\sigma) = -1$ if and only if $\sigma \in S_n$ is an odd permutation. Note that the group G_n is denoted by $\overline{S}_n$ in [Plans 2009].

The group G_n can be constructed explicitly as follows. Let

$$1 \to \{\pm 1\} \to C_4 = \{\pm\sqrt{-1}, \pm 1\} \overset{p_0}{\to} \{\pm 1\} \to 1$$

be the short exact sequence defined by $p_0(\sqrt{-1}) = -1$. The group G_n can be realized as the pullback of the diagram

$$\begin{array}{ccc} & & S_n \\ & & \downarrow \varepsilon_n \\ C_4 & \overset{p_0}{\longrightarrow} & \{\pm 1\}. \end{array}$$

Explicitly, as a subgroup of $S_n \times C_4$,

$$\begin{aligned} G_n &= \{(\sigma, (\sqrt{-1})^i) \in S_n \times C_4 : \varepsilon_n(\sigma) = p_0((\sqrt{-1})^i)\} \\ &= (A_n \times \{\pm 1\}) \cup \{(\sigma, \pm\sqrt{-1}) \in S_n \times C_4 : \sigma \notin A_n\}. \end{aligned}$$

If k is a field with $\operatorname{char} k \neq 2$, a faithful $2n$-dimensional representation can be defined as follows. Let $X = \left(\bigoplus_{1 \leq i \leq n} k \cdot x_i\right) \oplus \left(\bigoplus_{1 \leq i \leq n} k \cdot y_i\right)$ and let G_n act on X by, for $1 \leq i \leq n$,

$$\begin{array}{rcll} t & : & x_i \mapsto -x_i, & y_i \mapsto -y_i, \\ \tau & : & x_i \mapsto x_{\tau(i)}, & y_i \mapsto y_{\sigma^{-1}\tau\sigma(i)}, \\ \overline{\sigma} & : & x_i \mapsto y_i \mapsto -x_i, & \end{array} \tag{5-1}$$

where $t = (1, -1) \in G_n \subset S_n \times C_4$, $\tau \in A_n$ and τ is identified with $(\tau, 1) \in G_n$, $\sigma = (1, 2) \in S_n$ and $\overline{\sigma} = (\sigma, \sqrt{-1}) \in G_n$.

The next result was proved in [Plans 2009, Theorem 14(b)] under the assumptions that $\operatorname{char} k = 0$ and $\sqrt{-1} \in k$. Our proof is different from Plans' even in the situation when $\operatorname{char} k = 0$.

Theorem 5.2. *Assume that k is a field that satisfies*:

(i) *Either* $\operatorname{char} k = 0$ *or* $\operatorname{char} k = p > 0$ *with* $p \nmid 2n$.

(ii) $\sqrt{-1} \in k$.

Then $k(G_n)$ is k-rational for $n \geq 3$.

Proof. The reader will find that (i) the assumption $\operatorname{char} k \neq 2$ is used throughout the proof; (ii) the assumption $\operatorname{char} k \nmid n$ is used in Step 2; (iii) the assumption $\sqrt{-1} \in k$ is used in Step 3.

• Step 1. Apply Theorem 2.2. We find that $k(G_n)$ is rational over

$$k(x_i, y_i : 1 \le i \le n)^{G_n},$$

where G_n acts on the rational function field $k(x_i, y_i : 1 \le i \le n)$ by (5-1).

• Step 2. Define $u_0 = \sum_{1\le i\le n} x_i$, $v_0 = \sum_{1\le i\le n} y_i$ and $u_i = x_i/u_0$, $v_i = y_i/v_0$ for $1 \le i \le n$. Note that $k(x_i, y_i : 1 \le i \le n) = k(u_j, v_j : 0 \le j \le n)$ with the relations $\sum_{1\le i\le n} u_i = \sum_{1\le i\le n} v_i = 1$. The action of G_n is given by

$$\begin{aligned}
t &: u_0 \mapsto -u_0, \quad v_0 \mapsto -v_0, \quad u_i \mapsto u_i, \quad v_i \mapsto v_i,\\
\tau &: u_0 \mapsto u_0, \quad v_0 \mapsto v_0, \quad u_i \mapsto u_{\tau(i)}, \quad v_i \mapsto v_{\sigma^{-1}\tau\sigma(i)},\\
\overline{\sigma} &: u_0 \mapsto v_0 \mapsto -v_0, \quad u_i \mapsto v_i \mapsto u_i,
\end{aligned}$$

where $1 \le i \le n$ and t, τ, $\overline{\sigma}$ are defined in (5-1).

Define $w_1 = u_0 v_0$, $w_2 = u_0/v_0$. Then

$$k(u_j, v_j : 0 \le j \le n)^{\langle t\rangle} = k(u_i, v_i : 1 \le i \le n)(w_1, w_2).$$

Note that $\tau(w_i) = w_i$ for $1 \le i \le 2$, $\overline{\sigma}(w_1) = -w_1$, $\overline{\sigma}(w_2) = -1/w_2$. By Theorem 2.1,

$$k(u_i, v_i : 1 \le i \le n)(w_1, w_2)^{G_n/\langle t\rangle} = k(u_i, v_i : 1 \le i \le n)(w_2)^{G_n/\langle t\rangle}(w')$$

for some w' fixed by the action of $G_n/\langle t\rangle$. Moreover, we may identify $G_n/\langle t\rangle$ with S_n and identify $\overline{\sigma}$ (modulo $\langle t\rangle$) with σ.

Define $U_i = u_i - (1/n)$, $V_i = v_i - (1/n)$ for $1 \le i \le n$. We find that

$$\sum_{1\le i\le n} U_i = \sum_{1\le i\le n} V_i = 0$$

and the action of S_n on $k(U_i, V_i : 1 \le i \le n)$ becomes linear. We will consider $k(U_i, V_i : 1 \le i \le n)(w_2)^{S_n}$. The action of S_n is given by

$$\begin{aligned}
\tau &: U_i \mapsto U_{\tau(i)}, \quad V_i \mapsto V_{\sigma^{-1}\tau\sigma(i)}, \quad w_2 \mapsto w_2,\\
\sigma &: U_i \mapsto V_i \mapsto U_i, \quad w_2 \mapsto -1/w_2,
\end{aligned} \tag{5-2}$$

where $1 \le i \le n$, $\tau \in A_n$, $\sigma = (1, 2)$ and $\sum_{1\le i\le n} U_i = \sum_{1\le i\le n} V_i = 0$.

• Step 3. Since $\sqrt{-1} \in k$, define $w = (\sqrt{-1} - w_2)/(\sqrt{-1} + w_2)$. We find that $\tau(w) = w$ for $\tau \in A_n$ and $\sigma(w) = -w$. Apply Theorem 2.1. We find that

$$k(u_i, v_i : 1 \le i \le n)(w_2)^{S_n} = k(U_i, V_i : 1 \le i \le n)^{S_n}(w'')$$

for some w'' fixed by the action of S_n.

It remains to show that $k(U_i, V_i : 1 \le i \le n)^{S_n}$ is k-rational. The following proof of this fact is due to the referee.

Define $W_i^{\pm} = U_i \pm V_{\sigma(i)}$. It is easy to verify that for $\tau(W_i^{\pm}) = W_{\tau(i)}^{\pm}$ for $\tau \in A_n$; and that for $\sigma = (1, 2)$, $\sigma(W_i^+) = W_{\sigma(i)}^+$ and $\sigma(W_i^-) = -W_{\sigma(i)}^-$.

Define subspaces W and W' by $W = \sum_{1\le i\le n} k\cdot W_i^+$ and $W' = \sum_{1\le i\le n} k\cdot W_i^-$. Note that

$$\sum_{1\le i\le n} k\cdot U_i \oplus \sum_{1\le i\le n} k\cdot V_i = W \oplus W'.$$

Moreover, W is the standard representation of S_n, that is, $W \simeq \sum_{1\le i\le n} k\cdot s_i$ with $\sum_{1\le i\le n} s_i = 0$ and $\lambda(s_i) = s_{\lambda(i)}$ for all $\lambda \in S_n$, for all $1 \le i \le n$. On the other hand, W' is the representation space of the tensor product of the standard representation and the linear character $\varepsilon_n : S_n \to \{\pm 1\}$.

• Step 4. Apply Theorem 2.2 to $k(U_i, V_i : 1 \le i \le n)^{S_n}$. We find that

$$k(U_i, V_i : 1 \le i \le n)^{S_n} = k(W \oplus W')^{S_n} = k(W_i^+ : 1 \le i \le n-1)^{S_n}(t_1, \dots, t_{n-1}),$$

where each t_i is fixed by S_n. Obviously the field $k(W_i^+ : 1 \le i \le n-1)^{S_n}$ is k-rational, whence the result. □

In the following theorem the assumption $\sqrt{-1} \in k$ from Theorem 5.2 will be dropped. The first part of the following theorem was proved by Plans [2009, Theorem 14 (b)]; there he assumed that $\operatorname{char} k = 0$.

Theorem 5.3. (1) *If k is a field with* $\operatorname{char} k \ne 2$ *or* 3, *then $k(G_3)$ is k-rational.*

(2) *If k is a field with* $\operatorname{char} k \ne 2$, *then $k(G_4)$ is k-rational. Moreover, if* $\operatorname{char} k = 0$, *then $k(G_5)$ is also k-rational.*

Proof. Case 1: $n = 3$. By Step 2 in the proof of Theorem 5.2, it suffices to consider $k(U_i, V_i : 1 \le i \le 3)(w_2)^{S_3}$, where $\sum_{1\le i\le 3} U_i = \sum_{1\le i\le 3} V_i = 0$. Define $\tau = (1, 2, 3) \in S_3$. The actions are given by

$$\tau : U_1 \mapsto U_2 \mapsto -U_1 - U_2, \quad V_2 \mapsto V_1 \mapsto -V_1 - V_2,$$
$$\sigma : U_1 \leftrightarrow V_1, \quad U_2 \leftrightarrow V_2.$$

Define $w_3 = U_1/V_2$, $w_4 = U_2/V_1$, $w_5 = V_1/V_2$. It follows that

$$k(U_i, V_i : 1 \le i \le 3)(w_2)^{S_3} = k(w_j : 2 \le j \le 5)(V_1)^{S_3} = k(w_j : 2 \le j \le 5)^{S_3}(w_0)$$

for some w_0 by Theorem 2.1.

It remains to show that $k(w_j : 2 \le j \le 5)^{S_3}$ is k-rational. Note that

$$\tau : w_2 \mapsto w_2, \quad w_3 \mapsto w_4 \mapsto (w_3 + w_4 w_5)/(1 + w_5),$$
$$\sigma : w_2 \mapsto -1/w_2, \quad w_3 \mapsto 1/w_4, \quad w_4 \mapsto 1/w_3, \quad w_5 \mapsto w_3/(w_4 w_5).$$

Define $w_6 = (w_3 + w_4 w_5)/(1 + w_5)$. Note that $k(w_3, w_4, w_5) = k(w_3, w_4, w_6)$ and

$$\tau : w_3 \mapsto w_4 \mapsto w_6 \mapsto w_3 \quad \text{and} \quad \sigma : w_6 \mapsto 1/w_6.$$

Define $w_7 = (1-w_3)/(1+w_3)$, $w_8 = (1-w_4)/(1+w_4)$, $w_9 = (1-w_6)/(1+w_6)$. Then $k(w_3, w_4, w_6) = k(w_7, w_8, w_9)$ and

$$\tau : w_7 \mapsto w_8 \mapsto w_9 \mapsto w_7,$$
$$\sigma : w_7 \mapsto -w_8, \quad w_8 \mapsto -w_7, \quad w_9 \mapsto -w_9.$$

By Theorem 2.4 we find that $k(w_2, w_3, w_4, w_5)^{\langle\tau\rangle} = k(w_2, X_1, X_2, X_3)$, where $X_1 = w_7 + w_8 + w_9$ and

$$X_2 = \frac{w_7^2 w_8 + w_8^2 w_9 + w_9^2 w_7 - 3w_7 w_8 w_9}{w_7^2 + w_8^2 + w_9^2 - w_7 w_8 - w_7 w_9 - w_8 w_9},$$
$$X_3 = \frac{w_7 w_8^2 + w_8 w_9^2 + w_9 w_7^2 - 3w_7 w_8 w_9}{w_7^2 + w_8^2 + w_9^2 - w_7 w_8 - w_7 w_9 - w_8 w_9}.$$

Moreover, the action of σ is given by

$$\sigma : w_2 \mapsto -1/w_2, \quad X_1 \mapsto -X_1, \quad X_2 \mapsto -X_3, \quad X_3 \mapsto -X_2.$$

Apply Theorem 2.2. We find that $k(w_2, X_1, X_2, X_3)^{\langle\sigma\rangle} = k(w_2)^{\langle\sigma\rangle}(Y_1, Y_2, Y_3)$ for some Y_1, Y_2, Y_3 fixed by σ. Since $k(w_2)^{\langle\sigma\rangle}$ is k-rational, it follows that $k(w_2, X_1, X_2, X_3)^{\langle\sigma\rangle}$ is k-rational.

Case 2: $n = 4$. Once again we use Step 2 in the proof of Theorem 5.2. It suffices to consider $k(U_i, V_i : 1 \le i \le 4)(w_2)^{S_4}$, where $\sum_{1\le i\le 4} U_i = \sum_{1\le i\le 4} V_i = 0$. Set $\lambda_1 = (1,2)(3,4)$, $\lambda_2 = (1,3)(2,4)$, $\rho = (1,2,3)$ and $\sigma = (1,2)$ as before. Then S_4 is generated by λ_1, λ_2, ρ and σ.

Define $t_1 = U_1 + U_2$, $t_2 = V_1 + V_2$, $t_3 = U_1 + U_3$, $t_4 = V_2 + V_3$, $t_5 = U_2 + U_3$ and $t_6 = V_1 + V_3$. The action of S_4 is given by

$$\lambda_1 : t_1 \mapsto t_1, \quad t_2 \mapsto t_2, \quad t_3 \mapsto -t_3, \quad t_4 \mapsto -t_4, \quad t_5 \mapsto -t_5, \quad t_6 \mapsto -t_6,$$
$$\lambda_2 : t_1 \mapsto -t_1, \quad t_2 \mapsto -t_2, \quad t_3 \mapsto t_3, \quad t_4 \mapsto t_4, \quad t_5 \mapsto -t_5, \quad t_6 \mapsto -t_6,$$
$$\rho : t_1 \mapsto t_5 \mapsto t_3 \mapsto t_1, \quad t_2 \mapsto t_6 \mapsto t_4 \mapsto t_2,$$
$$\sigma : t_1 \leftrightarrow t_2, \quad t_3 \leftrightarrow t_6, \quad t_4 \leftrightarrow t_5.$$

It follows that $k(t_i : 1 \le i \le 6)(w_2)^{<\lambda_1,\lambda_2>} = k(T_i : 1 \le i \le 6)(w_2)$, where $T_1 = t_1/t_2$, $T_2 = t_3/t_4$, $T_3 = t_5/t_6$, $T_4 = t_2 t_6/t_4$, $T_5 = t_4 t_6/t_2$, $T_6 = t_2 t_4/t_6$.

Moreover, the actions of ρ and σ are given by

$$\begin{aligned}&\rho : T_1 \mapsto T_3 \mapsto T_2 \mapsto T_1, \quad T_4 \mapsto T_5 \mapsto T_6 \mapsto T_4,\\ &\sigma : T_1 \mapsto 1/T_1, \quad T_2 \mapsto 1/T_3, \quad T_3 \mapsto 1/T_2,\\ &\quad T_4 \mapsto (T_1T_2/T_3)T_6, \quad T_5 \mapsto (T_2T_3/T_1)T_5, \quad T_6 \mapsto (T_1T_3/T_2)T_4.\end{aligned}$$

By Theorem 2.2, it suffices to show that $k(T_i : 1 \le i \le 3)(w_2)^{<\rho,\sigma>}$ is k-rational. Define $w_3 = (1-T_1)/(1+T_1)$, $w_4 = (1-T_2)/(1+T_2)$, $w_5 = (1-T_3)/(1+T_3)$. Then we find

$$\begin{aligned}&\rho : w_2 \mapsto w_2, \quad w_3 \mapsto w_5 \mapsto w_4 \mapsto w_3,\\ &\sigma : w_2 \mapsto -1/w_2, \quad w_3 \mapsto -w_3, \quad w_4 \mapsto -w_5, \quad w_5 \mapsto -w_4.\end{aligned}$$

Use Theorem 2.4 to find that $k(T_i : 1 \le i \le 3)(w_2)^{<\rho>}$. The remaining part of the proof is very similar to the last part of Case 1. The details are omitted.

Case 3: $n = 5$. By [Plans 2009, Theorem 11], $k(G_5)$ is rational over $k(G_4)$. Since $k(G_4)$ is k-rational by Case 2, we are done. □

Acknowledgment

We thank the referee for his or her critical comments. In particular, the referee helped simplify the proof in Theorem 5.2 (Step 3 in the proof). Because of this simplification, one assumption of this theorem has been relaxed from "char $k = p > 0$ with $p \nmid n!$" to "char $k = p > 0$ with $p \nmid 2n$".

References

[Ahmad et al. 2000] H. Ahmad, M. Hajja, and M.-C. Kang, "Rationality of some projective linear actions", *J. Algebra* **228**:2 (2000), 643–658. MR 2001e:12003 Zbl 0993.12003

[Garibaldi et al. 2003] S. Garibaldi, A. Merkurjev, and J.-P. Serre, *Cohomological invariants in Galois cohomology*, University Lecture Series **28**, American Mathematical Society, Providence, RI, 2003. MR 2004f:11034 Zbl 1159.12311

[Hajja and Kang 1995] M. Hajja and M.-C. Kang, "Some actions of symmetric groups", *J. Algebra* **177**:2 (1995), 511–535. MR 96i:20013 Zbl 0837.20054

[Hoffman and Humphreys 1992] P. N. Hoffman and J. F. Humphreys, *Projective representations of the symmetric groups: Q-functions and shifted tableaux*, Clarendon, New York, 1992. MR 94f:20027 Zbl 0777.20005

[Hoshi and Kang 2010] A. Hoshi and M.-C. Kang, "Twisted symmetric group actions", *Pacific J. Math.* **248**:2 (2010), 285–304. MR 2011k:12006 Zbl 1203.14018

[Kang 2005] M.-C. Kang, "Noether's problem for dihedral 2-groups, II", *Pacific J. Math.* **222**:2 (2005), 301–316. MR 2007a:12003 Zbl 1100.12003

[Kang 2012] M.-C. Kang, "Retract rational fields", *J. Algebra* **349** (2012), 22–37. MR 2853623 Zbl 06037563

[Kang and Plans 2009] M.-C. Kang and B. Plans, "Reduction theorems for Noether's problem", *Proc. Amer. Math. Soc.* **137**:6 (2009), 1867–1874. MR 2010c:12004 Zbl 1177.12006

[Kang and Zhou 2012] M.-C. Kang and J. Zhou, "The rationality problem for finite subgroups of $GL_4(\mathbb{Q})$", *J. Algebra* **368**:1 (2012), 53–69.

[Karpilovsky 1985] G. Karpilovsky, *Projective representations of finite groups*, Pure and Applied Mathematics **94**, Dekker, New York, 1985. MR 86m:20014 Zbl 0568.20016

[Masuda 1955] K. Masuda, "On a problem of Chevalley", *Nagoya Math. J.* **8** (1955), 59–63. MR 16:993c Zbl 0065.02601

[Plans 2007] B. Plans, "Noether's problem for GL(2, 3)", *Manuscripta Math.* **124**:4 (2007), 481–487. MR 2008i:12009 Zbl 1146.12004

[Plans 2009] B. Plans, "On Noether's problem for central extensions of symmetric and alternating groups", *J. Algebra* **321**:12 (2009), 3704–3713. MR 2010e:12007 Zbl 1177.12007

[Saltman 1984] D. J. Saltman, "Retract rational fields and cyclic Galois extensions", *Israel J. Math.* **47**:2-3 (1984), 165–215. MR 85j:13008 Zbl 0546.14013

[Serre 1984] J.-P. Serre, "L'invariant de Witt de la forme $\mathrm{Tr}(x^2)$", *Comment. Math. Helv.* **59**:4 (1984), 651–676. MR 86k:11067 Zbl 0565.12014

[Serre 2011] J.-P. Serre, April 23 2011. Private email to M. Kang.

[Springer 1977] T. A. Springer, *Invariant theory*, Lecture Notes in Mathematics **585**, Springer, Berlin, 1977. MR 56 #5740 Zbl 0346.20020

[Swan 1983] R. G. Swan, "Noether's problem in Galois theory", pp. 21–40 in *Emmy Noether in Bryn Mawr* (Bryn Mawr, PA, 1982), edited by B. Srinivasan and J. D. Sally, Springer, New York, 1983. MR 84k:12013 Zbl 0538.12012

[Voskresenskii 1998] V. E. Voskresenskii, *Algebraic groups and their birational invariants*, Translations of Mathematical Monographs **179**, American Mathematical Society, Providence, RI, 1998. MR 99g:20090 Zbl 0974.14034

[Weil 1956] A. Weil, "The field of definition of a variety", *Amer. J. Math.* **78** (1956), 509–524. MR 18,601a Zbl 0072.16001

[Yamasaki 2009] A. Yamasaki, "Negative solutions to three-dimensional monomial Noether problem", preprint, 2009. arXiv 0909.0586v1

Received October 11, 2011. Revised December 26, 2011.

MING-CHANG KANG
DEPARTMENT OF MATHEMATICS AND
TAIDA INSTITUTE OF MATHEMATICAL SCIENCES
NATIONAL TAIWAN UNIVERSITY
TAIPEI 106
TAIWAN
kang@math.ntu.edu.tw

JIAN ZHOU
SCHOOL OF MATHEMATICAL SCIENCES
PEKING UNIVERSITY
BEIJING, 100871
CHINA
zhjn@math.pku.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

REMARKS ON THE BEHAVIOR OF NONPARAMETRIC CAPILLARY SURFACES AT CORNERS

KIRK E. LANCASTER

Consider a nonparametric capillary or prescribed mean curvature surface $z = f(x)$ defined in a cylinder $\Omega \times \mathbb{R}$ over a two-dimensional region Ω whose boundary has a corner at $\mathcal{O}$ with an opening angle of 2α. Suppose the contact angle approaches limiting values γ_1 and γ_2 in $(0, \pi)$ as $\mathcal{O}$ is approached along each side of the opening angle. We will prove the nonconvex Concus–Finn conjecture, determine the exact sizes of the radial limit fans of f at $\mathcal{O}$ when $(\gamma_1, \gamma_1) \in D_1^{\pm} \cup D_2^{\pm}$ and discuss the continuity of the Gauss map.

1. Introduction

Let $\Omega \subset \mathbb{R}^2$ be a connected, open set. Consider the classical capillary problem in a cylinder

$$Nf = \kappa f + \lambda \quad \text{in } \Omega, \tag{1}$$

$$Tf \cdot \nu = \cos\gamma \quad \text{(a.e.) on } \partial\Omega, \tag{2}$$

and, more generally, the prescribed mean curvature problem in a cylinder

$$Nf = H(\,\cdot\,, f(\,\cdot\,)) \quad \text{in } \Omega, \tag{3}$$

$$Tf \cdot \nu = \cos\gamma \quad \text{(a.e.) on } \partial\Omega, \tag{4}$$

where $Tf = \nabla f/\sqrt{1 + |\nabla f|^2}$, $Nf = \nabla \cdot Tf$, ν is the exterior unit normal on $\partial\Omega$, $H(x, t)$ is a weakly increasing function of t for each $x \in \Omega$ and $\gamma = \gamma(x)$ is in $[0, \pi]$. We will let $\mathcal{S}_f$ denote the closure in $\mathbb{R}^3$ of the graph of f over Ω. When $H(x, t) = \kappa t + \lambda$ (i.e., f satisfies (1)–(2)) with κ and λ constants such that $\kappa \geq 0$, then the surface $\mathcal{S}_f \cap (\Omega \times \mathbb{R})$ represents the stationary liquid-gas interface formed by an incompressible fluid in a vertical cylindrical tube with cross-section Ω in a microgravity environment or in a downward-oriented gravitational field, the subgraph $U = \{(x, t) \in \Omega \times \mathbb{R} : t < f(x)\}$ represents the fluid filled portion of the cylinder and $\gamma(x)$ is the angle (within the fluid) at which the liquid-gas interface meets the vertical cylinder at $(x, f(x))$; Paul Concus and Robert Finn have made

MSC2010: primary 76B45, 35J93; secondary 53A10, 35J62.
Keywords: capillary graph, Concus–Finn conjecture, Gauss map, sizes of fans.

fundamental contributions to the mathematical theory of capillary surfaces and have discovered that capillary surfaces can behave in unexpected ways (cf. [Concus and Finn 1996; Finn 1986; 1999; 2002b; 2002a]). For a function $f \in C^2(\Omega)$, we let

$$\vec{n}(X) = \vec{n}_f(X) = \frac{(\nabla f(x), -1)}{\sqrt{1+|\nabla f(x)|^2}}, \quad X = (x, t) \in \Omega,$$

denote the downward unit normal to the graph of f; when f is a solution of (1)–(2) and $\kappa \ge 0$, $\vec{n}$ represents the inward unit normal with respect to the fluid region. Of interest here is the behavior of capillary surfaces and prescribed mean curvature surfaces over domains $\Omega \subset \mathbb{R}^2$ whose boundaries contain corners (e.g., [Concus and Finn 1996; Finn 1996]).

Let us suppose $\mathbb{O} = (0, 0) \in \partial\Omega$ and Ω is a connected, simply connected open set in $\mathbb{R}^2$ such that $\partial\Omega \setminus \{\mathbb{O}\}$ is a piecewise C^1 curve, Ω has a corner of size 2α at $\mathbb{O}$ and the tangent cone to $\partial\Omega$ at $\mathbb{O}$ is $L^+ \cup L^-$, where polar coordinates relative to $\mathbb{O}$ are denoted by r and θ, $L^+ = \{\theta = \alpha\}$ and $L^- = \{\theta = -\alpha\}$. We will assume there exist $\delta^* > 0$, $\rho^* \in (0, 1]$ such that $\partial^+\Omega = \partial\Omega \cap \overline{B(\mathbb{O}, \delta^*)} \cap T^+$ and $\partial^-\Omega = \partial\Omega \cap \overline{B(\mathbb{O}, \delta^*)} \cap T^-$ are connected, C^{1,ρ^*} arcs, where $T^+ = \{x \in \mathbb{R}^2 : x_2 \ge 0\}$, $T^- = \{x \in \mathbb{R}^2 : x_2 \le 0\}$ and $B(\mathbb{O}, \epsilon) = \{x \in \mathbb{R}^2 : \ |x| < \epsilon\}$; hence the tangent rays to $\partial^+\Omega$ and $\partial^-\Omega$ at $\mathbb{O}$ are L^+ and L^- respectively. Set $\Omega_0 = \Omega_0(\delta^*) = \Omega \cap B(\mathbb{O}, \delta^*)$. Let $\gamma^+(s)$ and $\gamma^-(s)$ denote γ along the arcs $\partial^+\Omega_0$ and $\partial^-\Omega_0$, respectively, where $s = 0$ corresponds to the point $\mathbb{O}$; here we have parametrized $\partial^+\Omega_0$ and $\partial^-\Omega_0$ by, for example, arclength s from $\mathbb{O}$ and write these parametrizations as x^+ and x^- respectively. We will assume there exist $\gamma_1, \gamma_2 \in (0, \pi)$ such that

$$\lim_{\partial^+\Omega \ni x \to \mathbb{O}} \gamma(x) = \gamma_1 \quad \text{and} \quad \lim_{\partial^-\Omega \ni x \to \mathbb{O}} \gamma(x) = \gamma_2. \tag{5}$$

Suppose first that $2\alpha \le \pi$ (i.e., the corner is convex or $\partial\Omega$ is C^1 at $\mathbb{O}$); such an Ω is illustrated in Figure 1. Figure 2 can then be used to illustrate our knowledge

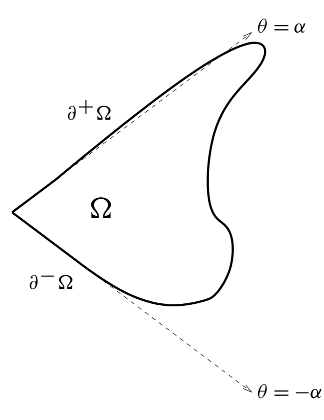


Figure 1. Ω with $2\alpha < \pi$.

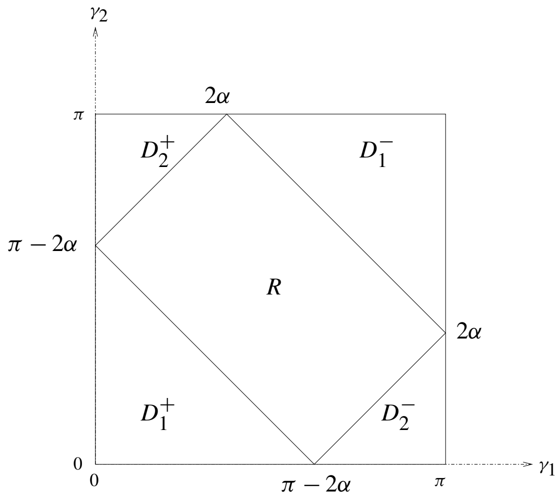


Figure 2. The Concus–Finn rectangle for convex corners.

of the behavior of a solution f of (3)–(4) at the corner $\mathcal{O}$; here let R, $D_1^\pm$, $D_2^\pm$ be the indicated open regions in the (open) square $(0, \pi) \times (0, \pi)$. If (γ_1, γ_2) is in $\overline{R} \cap (0, \pi) \times (0, \pi)$, then f is continuous at $\mathcal{O}$ [Concus and Finn 1996, Theorem 1; Lancaster and Siegel 1996b; 1996a, Corollary 4; Tam 1986]. If $(\gamma_1, \gamma_2) \in D_1^\pm$, then f is unbounded in any neighborhood of $\mathcal{O}$ and the capillary problem has no solution if $\kappa = 0$ [Concus and Finn 1996; Finn 1996]. If $(\gamma_1, \gamma_2) \in D_2^\pm$, then f is bounded [Lancaster and Siegel 1996a, Proposition 1] but its continuity at $\mathcal{O}$ was unknown until recently. Concus and Finn discovered bounded solutions of (1)–(2) in domains with corners whose unit normals (i.e., Gauss maps) cannot extend continuously as functions of x to a corner on the boundary of the domain [Finn 1988a, page 15; 1988b; 1996; Concus and Finn 1996, Example 2]. They formulated the conjecture that the solution f of (1)–(2) must be discontinuous at $\mathcal{O}$ when $(\gamma_1, \gamma_2) \in D_2^\pm$. Writing the conditions required for a pair of angles to be in $D_2^\pm$ yields the following formulation of their conjecture:

Concus–Finn conjecture. *Suppose* $0 < \alpha < \pi/2$, *that the limits* (5) *exist and* $0 < \gamma_1, \gamma_2 < \pi$. *If* $2\alpha + |\gamma_1 - \gamma_2| > \pi$, *then any solution of* (3)–(4), *with* $H(x, z) = \kappa z + \lambda$, $\kappa \geq 0$, *has a jump discontinuity at* O.

This conjecture was proven for solutions of (3)–(4) (i.e., without the restriction that $H(x, z) = \kappa z + \lambda$) in [Lancaster 2010].[1]

Thus, when $2\alpha \leq \pi$, $(\gamma_1, \gamma_2) \in D_2^\pm$, and f satisfies (3)–(4), f is discontinuous at $\mathcal{O}$ and there is a countable set $\mathcal{I} \subset (-\alpha, \alpha)$ such that the radial limit function of f at

[1] For convenience, we will abbreviate this reference as [L]. Similarly, [Lancaster and Siegel 1996a] and [Lancaster and Siegel 1996b] will be abbreviated [LS a] and [LS b], respectively.

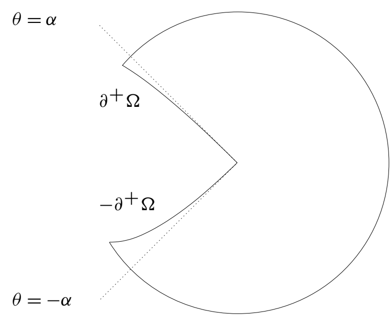


Figure 3. Ω with $2\alpha > \pi$.

$\mathcal{O}$, Rf, defined by $Rf(\alpha) = \lim_{\partial^+\Omega \ni x \to \mathcal{O}} f(x)$, $Rf(-\alpha) = \lim_{\partial^-\Omega \ni x \to \mathcal{O}} f(x)$ and

(6) $$Rf(\theta) = \lim_{r \downarrow 0} f(r\cos\theta, r\sin\theta), \quad -\alpha < \theta < \alpha,$$

is well-defined and continuous on $[-\alpha, \alpha] \setminus \mathcal{I}$ and behaves as in Proposition 1(i) of [LS b]; if $H(x, z)$ is strictly increasing in z [LS a, §5] or real-analytic [LS b] for x in a neighborhood of $\mathcal{O}$, then $\mathcal{I} = \varnothing$. (See [LS a], Step 3 of the proof of Theorem 1 and §5, and [LS b] regarding the sets $\mathcal{I}$ and cusp solutions.) We may assume for the moment that $(\gamma_1, \gamma_2) \in D_2^+$ since the other case follows by interchanging x_1 and x_2; then Theorems 1 and 2 of [LS a] and Proposition 1 and Theorem 1 of [LS b] imply there is a countable set $\mathcal{I} \subset [\alpha_1, \alpha_2]$ such that

$$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly increasing} & \text{on } [\alpha_1, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1], \end{cases}$$

with $\alpha_1 < \alpha_2$, $\alpha - \alpha_2 \geq \gamma_1$ and $\alpha_1 - (-\alpha) \geq \pi - \gamma_2$. In fact, determining the exact sizes of these radial limit fans when f is discontinuous at $\mathcal{O}$ follows easily from [L]. (Notice that $D_1^{\pm} = \varnothing$ if $2\alpha = \pi$.)

Proposition 1.1. *Let Ω be as above with $2\alpha < \pi$ and f be a bounded solution to* (3)–(4). *Suppose that $(\gamma_1, \gamma_2) \in D_2^{\pm}$ and that there exist constants $\underline{\gamma}^{\pm}$, $\overline{\gamma}^{\pm}$, $0 < \underline{\gamma}^{\pm} \leq \overline{\gamma}^{\pm} < \pi$, satisfying*

$$\underline{\gamma}^{+} + \underline{\gamma}^{-} > \pi - 2\alpha \quad \textit{and} \quad \overline{\gamma}^{+} + \overline{\gamma}^{-} < 2\alpha + \pi,$$

so that $\underline{\gamma}^{\pm} \leq \gamma^{\pm}(s) \leq \overline{\gamma}^{\pm}$ for all s, $0 < s < s_0$, for some s_0. Then $Rf(\theta)$ exists for $\theta \in [-\alpha, \alpha] \setminus \mathcal{I}$ and $Rf(\theta)$ is a continuous function of $\theta \in [-\alpha, \alpha] \setminus \mathcal{I}$, where $\mathcal{I}$ is a countable subset of $(-\alpha, \alpha)$.

Case (I). *If* $(\gamma_1,\gamma_2)\in D_2^+$ *(i.e.,* $\gamma_1-\gamma_2<2\alpha-\pi$*) then* $\alpha_1=-\alpha+\pi-\gamma_2$, $\alpha_2=\alpha-\gamma_1$ *and*

$$Rf \text{ is } \begin{cases} \textit{constant} & \textit{on } [\alpha-\gamma_1,\alpha], \\ \textit{strictly increasing} & \textit{on } [-\alpha+(\pi-\gamma_2),\alpha-\gamma_1]\setminus\mathcal{J}, \\ \textit{constant} & \textit{on } [-\alpha,-\alpha+(\pi-\gamma_2)], \end{cases}$$

where $\mathcal{J}$ *is a countable subset of* $[-\alpha+(\pi-\gamma_2),\alpha-\gamma_1]$.

Case (D). *If* $(\gamma_1,\gamma_2)\in D_2^-$ *(i.e.,* $\gamma_1-\gamma_2>\pi-2\alpha$*) then* $\alpha_1=-\alpha+\gamma_2$, $\alpha_2=\alpha-\pi+\gamma_1$ *and*

$$Rf \text{ is } \begin{cases} \text{constant} & \textit{on } [\alpha-(\pi-\gamma_1),\alpha], \\ \textit{strictly decreasing} & \textit{on } [-\alpha+\gamma_2,\alpha-(\pi-\gamma_1)]\setminus\mathcal{J}, \\ \textit{constant} & \textit{on } [-\alpha,-\alpha+\gamma_2], \end{cases}$$

where $\mathcal{J}$ *is a countable subset of* $[-\alpha+\gamma_2,\alpha-\pi+\gamma_1]$.

Proof. Using the information from [LS b] and [LS a] given above and assuming $(\gamma_1,\gamma_2)\in D_2^+$, we will argue by contradiction. Suppose that $\alpha_2<\alpha-\gamma_1$. Let

$$\Omega_0\subset\big\{(r\cos\theta,r\sin\theta)\in\Omega : r>0,\ \alpha_2<\theta<\alpha-\gamma_1/2\big\}.$$

be an open set whose boundary $\partial\Omega_0$ contains $\{\theta=\alpha-\gamma_1/2\}$ and is tangent to $\{\theta=\alpha_2\}$ at $\mathbb{O}$ so that the appropriate analogue of [L, (43)] tends to zero. Then f is continuous on $\overline{\Omega}$ and, from Theorem 2.1 of [L], we obtain

$$\lim_{r\downarrow 0}\vec{n}_f\big(r\cos(\alpha-\tfrac{1}{2}\gamma_1),r\sin(\alpha-\tfrac{1}{2}\gamma_1)\big)=\big(-\sin(\alpha-\gamma_1),\cos(\alpha-\gamma_1),0\big), \tag{7}$$

$$\lim_{\substack{x\to\mathbb{O}\\ x\in\partial\Omega_0\setminus\{\theta=\alpha-\frac{1}{2}\gamma_1\}}}\vec{n}_f(x)=(-\sin\alpha_2,\cos\alpha_2,0). \tag{8}$$

Notice that the limiting contact angles at $\mathbb{O}$ are $\frac{1}{2}\gamma_1$ (on $\theta=\alpha-\frac{1}{2}\gamma_1$) and π (on $\theta=\alpha_2$). Now, using Theorem 2.1 of [L], we see that the arguments in §3 of [L] yield a contradiction to the assumption that $\alpha_2<\alpha-\gamma_1$. (If $\gamma_2=\pi$ were allowed in Theorem 1.1 of [L], then a contradiction would follow immediately since $2\alpha=\alpha-\alpha_2-\frac{1}{2}\gamma_1$, $|\gamma_1-\gamma_2|=\pi-\frac{1}{2}\gamma_1$ and $2\alpha+|\gamma_1-\gamma_2|=\pi+\alpha-\gamma_1-\alpha_2>\pi$.) In the case that $\alpha_1>-\alpha+\pi-\gamma_2$ or $(\gamma_1,\gamma_2)\in D_2^-$, the proof follows in a similar manner. □

The focus of this note is to give a direct proof of the nonconvex Concus–Finn conjecture and, when $(\gamma_1,\gamma_1)\in D_1^\pm\cup D_2^\pm$, establish the exact sizes of radial limit fans at reentrant corners and discuss the continuity of the Gauss map. We note that Danzhu Shi assumes the (convex) Concus–Finn conjecture holds when $\gamma_1\in\{0,\pi\}$ or $\gamma_2\in\{0,\pi\}$ and then, in her extremely interesting paper [Shi 2006], gives an argument for the proof of the nonconvex Concus–Finn conjecture. Unfortunately, these cases (e.g., $\gamma_j \in \{0, \pi\}$, $j = 1, 2$) are not covered in [L]. Our interest in proving the nonconvex Concus–Finn conjecture arises from our need, when determining the exact sizes of fans at reentrant corners, for the information developed during its proof (e.g., analogs of Theorem 2.1 of [L]) and from a belief in the value of presenting a proof which directly uses the ideas and techniques in [L].

2. The nonconvex Concus–Finn conjecture

The following theorem implies that the nonconvex Concus–Finn conjecture (cf. [Shi 2006]) is true; the proof will be given in Section 2B.

Theorem 2.1. *Let* Ω *and* γ *be as above with* $\alpha \in [\frac{\pi}{2}, \pi]$. *Let*

$$f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathbb{O}, \delta) \cap \overline{\Omega} \setminus \{\mathbb{O}\})$$

be a bounded solution of (3)–(4) *with*

$$|H|_\infty = \sup_{\boldsymbol{x} \in \Omega} |H(\boldsymbol{x}, f(\boldsymbol{x}))| < \infty$$

for some $\delta > 0$ *and* $\rho \in (0, 1)$. *Suppose* (5) *holds and* $\gamma_1, \gamma_2 \in (0, \pi)$. *Then* f *is discontinuous at* $\mathbb{O}$ *whenever* $|\gamma_1 - \gamma_2| > 2\alpha - \pi$ *or* $|\gamma_1 + \gamma_2 - \pi| > 2\pi - 2\alpha$ (*i.e.*, $(\gamma_1, \gamma_2) \in D_1^\pm \cup D_2^\pm$).

Throughout this section, we will consider f to be a fixed solution of (3)–(4) that satisfies the hypotheses of this theorem. We may parametrize the graph of f as in [LS a], using the unit disk $E = \{(u, v) : u^2 + v^2 < 1\}$ as our parameter domain. From Step 1 of the proof of Theorem 1 of [LS a] and §3 of [L], we see that there is a parametric description $X : \overline{E} \to \mathbb{R}^3$ of the closure S of $S_0 = \{(x, f(x)) : x \in \Omega\}$,

$$X(u, v) = (x(u, v), y(u, v), z(u, v)), \quad (u, v) \in \overline{E},$$

such that:

(i) $X \in C^2(E : \mathbb{R}^3) \cap W^{1,2}(E : \mathbb{R}^3)$.

(ii) X is a homeomorphism of E onto S_0.

(iii) X maps ∂E onto $\{(x, f(x)) : x \in \partial\Omega\} \cup (\{\mathbb{O}\} \times [z_1, z_2])$, where

$$z_1 = \liminf_{\overline{\Omega} \ni x \to \mathbb{O}} f(x) \quad \text{and} \quad z_2 = \limsup_{\overline{\Omega} \ni x \to \mathbb{O}} f(x).$$

(iv) X is conformal on E: $X_u \cdot X_v = 0$, $|X_u| = |X_v|$ on E.

(v) Let $\tilde{H}(u, v) = H(X(u, v))$ denote the prescribed mean curvature of $\mathcal{S}_f$ at $X(u, v)$. Then $\triangle X := X_{uu} + X_{vv} = \tilde{H}\, X_u \times X_v$.

(vi) $X \in C^0(\overline{E})$.

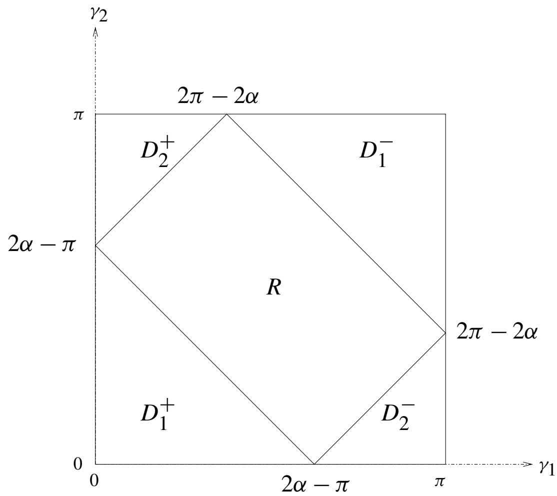


Figure 4. The Concus–Finn rectangle for nonconvex corners.

(vii) Writing $G(u, v) = (x(u, v), y(u, v))$, $G(\cos t, \sin t)$ moves clockwise about $\partial\Omega$ as t increases, $0 \le t \le 2\pi$, and G is an orientation reversing homeomorphism from E onto Ω; G maps $\overline{E}$ onto $\overline{\Omega}$ and, if f is continuous at $\mathbb{O}$, then G is a homeomorphism from $\overline{E}$ onto $\overline{\Omega}$.

(viii) Let $\pi_S : S^2 \to \mathbb{C}$ denote the stereographic projection from the North Pole and define $g(u + iv) = \pi_S(\vec{n}_f(G(u, v)))$, $(u, v) \in E$. Then

$$|g_{\bar{\zeta}}| = \tfrac{1}{2}|\tilde{H}|(1 + |g|^2)|X_u|, \tag{9}$$

where $\zeta = u + iv$, $\partial/\partial\zeta = \frac{1}{2}\big(\partial/\partial u - i\partial/\partial v\big)$ and $\partial/\partial\bar{\zeta} = \frac{1}{2}\big(\partial/\partial u + i\partial/\partial v\big)$. For convenience when working with complex variables, set $E_1 = \{\zeta \in \mathbb{C} : |\zeta| < 1\}$.

(ix) The parametric Gauss map $N : E \to S^2$ is $N = (X_u \times X_v)/|X_u \times X_v|$ and satisfies $N(u, v) = \vec{n}_f(G(u, v))$, $(u, v) \in E$; the domain of N is taken as the largest subset of $\overline{E}$ on which N extends continuously.

It is convenient to introduce some notation. Suppose $V \subset \mathbb{R}^2$ with $\mathbb{O} \in \partial V$. For $t > 0$, set $V_t = \{(x, y) \in V : x^2 + y^2 < t^2\}$. Let $s(V)$ denote the set of sequences in V that converge to $\mathbb{O}$. If $h \in C^1(V)$, we define $\Pi_h(V) = \bigcap_{t>0} \overline{\vec{n}_h(V_t)}$; then

$$\Pi_h(V) = \big\{Y \in S^2 : \text{ there exists } (x_j) \in s(V) \text{ such that } Y = \lim_{j\to\infty} \vec{n}_h(x_j)\big\}.$$

Without assuming that f is or is not continuous at $\mathbb{O}$, we have:

Lemma 2.2. *Let Λ be an open, connected, simply connected subset of Ω with $\mathbb{O}$ in $\partial\Lambda$ and suppose that there is a rotation M of $\mathbb{R}^2$ about $\mathbb{O}$ such that*

$$\{(M(y_1, y_2), y_3) : Y \in \Pi_f(\Lambda)\}$$

is contained in a compact subset of $\{Y \in S^2 : y_2 > 0,\ y_3 \le 0\}$. *Let* ϕ *be a conformal map from* E *to* $G^{-1}(\Lambda)$ *and define*

$$\tilde{g}(u+iv) = \pi_S(\vec{n}_f(G \circ \phi(u,v))), \quad (u,v) \in E. \tag{10}$$

Then there exists $p > 2$ *such that*

$$\tilde{g}(\zeta) = \psi(\zeta) + h(\zeta), \quad \zeta \in E_1, \tag{11}$$

where ψ *is a holomorphic function on* E_1 *and* $h \in L^\infty(E_1)$ *is a Hölder continuous function on* $\overline{E}_1$ *with Hölder exponent* $\mu = (p-2)/p$.

Proof. In §3 of [L], the fact that the limits at $\mathbb{O}$ of the Gauss map are contained in a compact subset of $\{Y \in S^2 : y_2 > 0,\ y_3 \le 0\}$ implies that $(u,v) \mapsto (z(u,v), x(u,v))$ is quasiconformal and has a quasiconformal extension to $\mathbb{R}^2$; Gehring's lemma and the isothermal parametrization imply $X \in W^{1,p}$ for some $p > 2$ and the classical literature implies $g = \psi + h$ with ψ and h as above. We can argue as in §3 of [L]; we find that $X \in W^{1,p}(E : \mathbb{R}^3)$ for some $p > 2$ and

$$\tilde{g}(\zeta) = \psi(\zeta) + h(\zeta), \tag{12}$$

where ψ is a holomorphic function and $h \in L^\infty(E_1)$ is an uniformly Hölder continuous function on E_1 with Hölder exponent μ. □

Remark 2.3. Notice that $\tilde{g} = g \circ \phi_1$, where ϕ_1 is a conformal map from E_1 onto $\{u+iv : (u,v) \in G^{-1}(\Lambda)\}$.

2A. ***Image of the Gauss map.*** The (nonparametric) Gauss map on $\mathcal{S}_f$ is the (downward) unit normal map to $\mathcal{S}_f$ when this is defined and equals $\vec{n}_f$ on $\mathcal{S}_f \cap (\Omega \times \mathbb{R})$; here we consider $\vec{n}_f : \Omega \times \mathbb{R} \to S^2_-$ by letting $(x,t) \mapsto \vec{n}_f(x)$. In this section, we characterize in Theorems 2.4 and 2.5 the behavior of the limits at points of $\{\mathbb{O}\} \times \mathbb{R}$ of the Gauss map for the graph of f when $(\gamma_1, \gamma_2) \notin \overline{R}$. Let $S^2_- = \{\omega \in \mathbb{R}^3 : |\omega| = 1, \omega_3 \le 0\}$ be the (closed) lower half of the unit sphere.

Theorem 2.4. *Let* $2\alpha > \pi$ *and* Ω *and* γ *be as in Section 1 and suppose* (5) *holds with* $\gamma_1, \gamma_2 \in (0, \pi)$. *Let* $\beta \in (-\alpha, \alpha)$ *and* $(x_j) \in s(\Omega)$ *such that*

$$\lim_{j\to\infty} \frac{x_j}{|x_j|} = (\cos\beta, \sin\beta). \tag{13}$$

Let us write $\omega(\theta) = (\cos\theta, \sin\theta, 0)$ *for* $\theta \in \mathbb{R}$.

(D_2^+) *If* $(\gamma_1, \gamma_2) \in D_2^+$ *(i.e.,* $\gamma_1 - \gamma_2 < \pi - 2\alpha$*) then*

$$\lim_{j\to\infty} \vec{n}_f(x_j) = \begin{cases} \omega(\alpha - \gamma_1 + \frac{\pi}{2}) & \text{if } \beta \in [\alpha - \gamma_1, \alpha), \\ \omega(\beta + \frac{\pi}{2}) & \text{if } \beta \in [-\alpha + (\pi - \gamma_2), \alpha - \gamma_1], \\ \omega(-\alpha - \gamma_2 + \frac{3\pi}{2}) & \text{if } \beta \in (-\alpha, -\alpha + (\pi - \gamma_2)]. \end{cases}$$

(D_2^-) *If* $(\gamma_1,\gamma_2)\in D_2^-$ (*i.e.*, $\gamma_1-\gamma_2>2\alpha-\pi$) *then*

$$\lim_{j\to\infty}\vec{n}_f(x_j)=\begin{cases}\omega(\alpha+\gamma_1-\frac{3\pi}{2}) & \text{if } \beta\in[\alpha+\gamma_1-\pi,\alpha),\\ \omega(\beta-\frac{\pi}{2}) & \text{if } \beta\in[-\alpha+\gamma_2,\alpha+\gamma_1-\pi],\\ \omega(-\alpha+\gamma_2-\frac{\pi}{2}) & \text{if } \beta\in(-\alpha,-\alpha+\gamma_2].\end{cases}$$

Proof. Let us assume $(\gamma_1,\gamma_2)\in D_2^{\pm}$. Let $\beta\in(-\alpha,\alpha)$ and (x_j) be an arbitrary sequence in Ω converging to $\mathbb{O}$ and satisfying (13). Since $(\vec{n}_f(x_j):j\in\mathbb{N})$ is a sequence in the compact set S^2_-, there is a subsequence of $(x_j:j\in\mathbb{N})$, still denoted $(x_j:j\in\mathbb{N})$, and $\theta\in(-\frac{\pi}{2},\frac{3\pi}{2}]$, $\tau\in[0,1]$ such that $(\vec{n}_f(x_{j_k}):k\in\mathbb{N})$ is convergent and

$$\lim_{k\to\infty}\vec{n}_f(x_{j_k})=(\tau\cos\theta,\tau\sin\theta,-\sqrt{1-\tau^2}).$$

Using [Jeffres and Lancaster 2008] and the techniques and arguments in §2 of [L], we see that $\tau=1$, $\lim_{k\to\infty}\vec{n}_f(x_{j_k})=\omega(\theta)$, and $\omega(\theta)$ is normal to $\partial\mathcal{P}$ and points into $\mathcal{P}$, where $\omega(\beta)\in\partial\mathcal{P}$ and $\mathcal{P}$ is given in Theorem 2.2 of [Jeffres and Lancaster 2008]. (In §2 of [L], the function $u(x)=f(x)-Rf(\beta)$ is blown up about $(0,0,0)$; that is, the graphs of a subsequence of the sequence (u_j) in $C^2(\Omega)$, where u_j is defined by $u_j(x)=(f(\epsilon_j x)-Rf(\beta))/\epsilon_j$ and $\epsilon_j=|x_j|$ for $j\in\mathbb{N}$, are shown to converge to the intersection of $\Omega\times\mathbb{R}$ with a vertical plane π_1. The (downward) unit normal to π_1 is shown to be normal to the vertical plane $\partial\mathcal{P}$ which contains $(\cos\beta,\sin\beta,0)$ and point into $\mathcal{P}$, where $\mathcal{P}$ satisfies Theorem 2.1 of [Jeffres and Lancaster 2007].)

If $(\gamma_1,\gamma_2)\in D_2^+$, then the conclusions of Theorem 2.4 follow from Corollary 2.4 of [Jeffres and Lancaster 2008]; Figure 5 illustrates the graph of the argument of $\vec{n}(\beta)=\lim_{r\downarrow 0}\vec{n}_f(r\cos\beta,r\sin\beta)$. If $(\gamma_1,\gamma_2)\in D_2^-$, then the conclusions of Theorem 2.4 follow from Corollary 2.5 of [ibid.]. □

Suppose that $\alpha\in(\frac{\pi}{2},\pi]$, $\gamma_1,\gamma_2\in(0,\pi)$ and $\gamma_1+\gamma_2<2\alpha-\pi$. Let us define $\mathcal{F}=\mathcal{F}(\alpha,\gamma_1,\gamma_2)$ as follows: Set

$$\mathcal{F}_1=[-\alpha,-\alpha-\gamma_2+\pi]\times\{-\alpha-\gamma_2-\pi/2\},$$
$$\mathcal{F}_2=[-\alpha,-\alpha+\gamma_2+\pi]\times\{-\alpha+\gamma_2-\pi/2\},$$
$$\mathcal{F}_3=[\alpha-\gamma_1-\pi,\alpha]\times\{\alpha-\gamma_1-3\pi/2\},$$
$$\mathcal{F}_4=[\alpha+\gamma_1-\pi,\alpha]\times\{\alpha+\gamma_1-3\pi/2\},$$
$$\mathcal{F}_5=\{(\beta,\beta-\pi/2):\beta\in[-\alpha+\gamma_2,\alpha+\gamma_1-\pi]\},$$
$$\mathcal{F}_6=\{(\beta,\beta-3\pi/2):\beta\in[-\alpha+\gamma_2+\pi,\alpha-\gamma_1]\},$$
$$\mathcal{F}_7=\left\{(\beta+t,-\alpha+\gamma_2-\pi/2+t):\begin{array}{c}\beta\in[-\alpha+\gamma_2,-\alpha+\gamma_2+\pi]\\ t\in[0,2\alpha-\pi-\gamma_1-\gamma_2]\end{array}\right\},$$

and define $\mathcal{F}=\bigcup_{j=1}^{7}\mathcal{F}_j=\mathcal{F}_1\cup\cdots\cup\mathcal{F}_7$ (see Figures 6, 7 and 8 for illustrations).

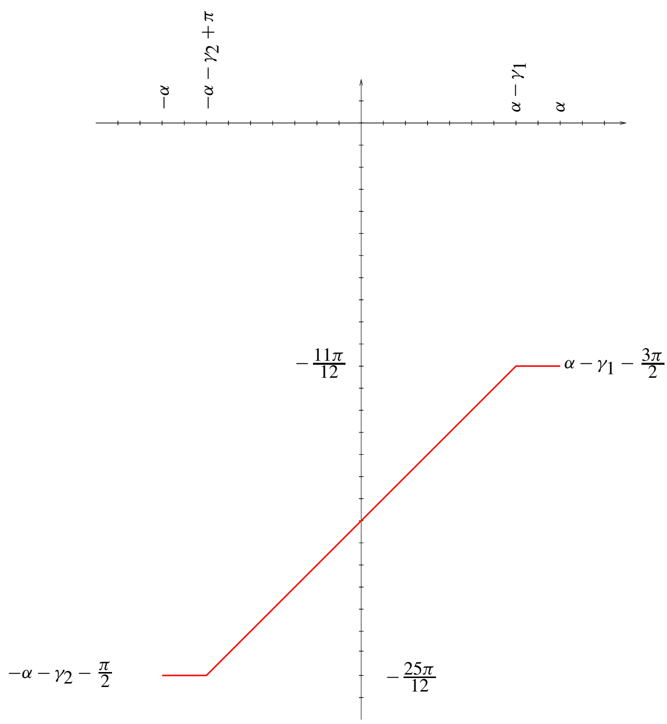


Figure 5. $\{(\beta, \arg(\vec{m}_f(\beta)))\}$ at a D_2^+ corner; $\alpha = \frac{3\pi}{4}$, $\gamma_1 = \frac{\pi}{6}$, $\gamma_2 = \frac{5\pi}{6}$.

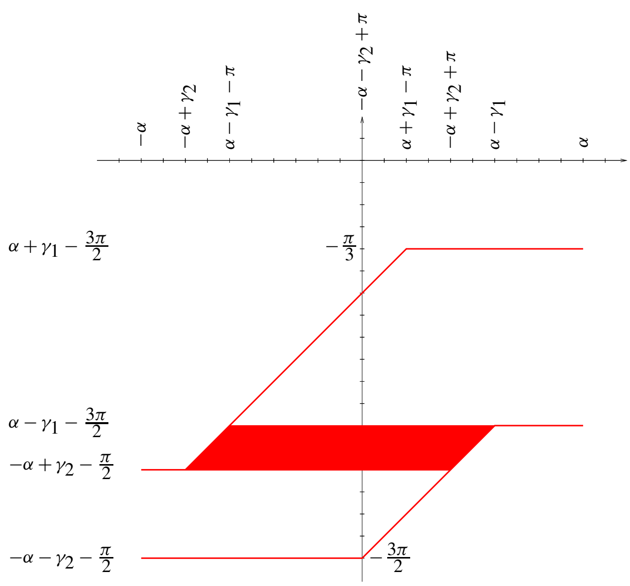


Figure 6. $\mathscr{F}$ at a D_1^+ corner; $\alpha = \frac{5\pi}{6}$, $\gamma_1 = \frac{\pi}{3}$, $\gamma_2 = \frac{\pi}{6}$.

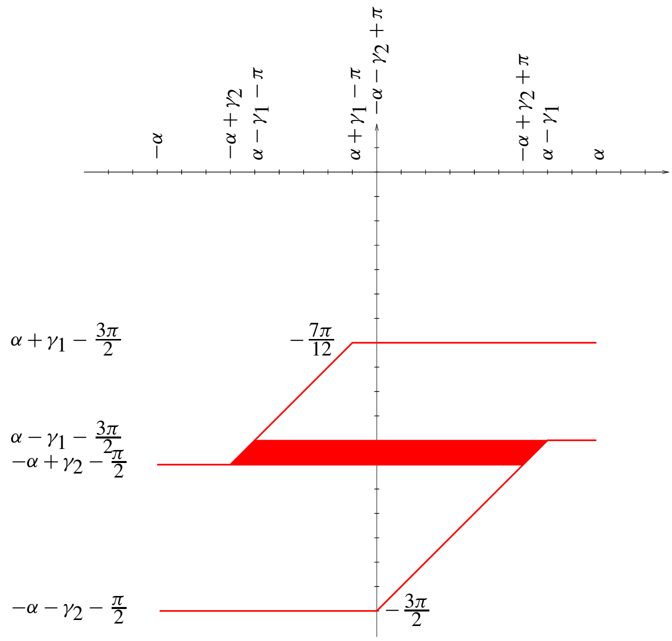


Figure 7. $\mathscr{F}$ at a D_1^+ corner; $\alpha=\frac{3\pi}{4}$, $\gamma_1=\frac{\pi}{6}$, $\gamma_2=\frac{\pi}{4}$.

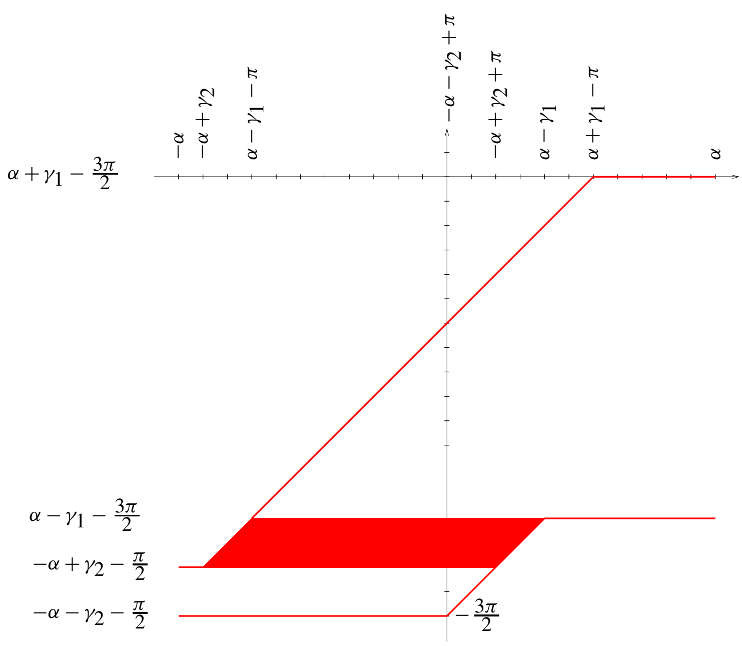


Figure 8. $\mathscr{F}$ at a D_1^+ corner; $\alpha=\frac{11\pi}{12}$, $\gamma_1=\frac{7\pi}{12}$, $\gamma_2=\frac{\pi}{12}$.

Theorem 2.5. *Let $2\alpha > \pi$ and Ω and γ be as in Section 1 and suppose* (5) *holds with $\gamma_1, \gamma_2 \in (0, \pi)$. Let $\beta \in (-\alpha, \alpha)$ and $(x_j) \in s(V)$ such that*

$$\lim_{j\to\infty} \frac{x_j}{|x_j|} = (\cos\beta, \sin\beta). \tag{14}$$

Continuing to write $\omega(\theta) = (\cos\theta, \sin\theta, 0)$ for $\theta \in \mathbb{R}$, we see:

(i) *Suppose $(\gamma_1, \gamma_2) \in D_1^+$ (i.e., $\gamma_1 + \gamma_2 < 2\alpha - \pi$), $\lim_{j\to\infty} \vec{n}_f(x_j)$ exists and*

$$\lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta) \quad \textit{for some } \theta \in \mathbb{R}.$$

Then $(\beta, \theta) \in \mathcal{F}$.

(ii) *Suppose $(\gamma_1, \gamma_2) \in D_1^-$ (i.e., $\gamma_1 + \gamma_2 > 2\alpha + \pi$), $\lim_{j\to\infty} \vec{n}_f(x_j)$ exists and*

$$\lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta) \quad \textit{for some } \theta \in \mathbb{R}.$$

Then $(-\beta, \theta) \in \mathcal{F}$.

(iii) *Connectedness at β*: *Suppose $(\gamma_1, \gamma_2) \in D_1^+$ and $(x_j), (y_j) \in s(\Omega)$ such that*

$$\lim_{j\to\infty} \frac{x_j}{|x_j|} = \lim_{j\to\infty} \frac{y_j}{|y_j|} = (\cos\beta, \sin\beta),$$
$$\lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta_1) \quad \textit{and} \quad \lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta_2),$$

for some $\theta_1 \le \theta_2$ such that $(\beta, \theta_1), (\beta, \theta_2) \in \mathcal{F}$. Then the set $\{\theta \in [\theta_1, \theta_2] : (\beta, \theta) \in \mathcal{F}\}$ must be connected.

(iv) *Connectedness*: *Suppose $(\gamma_1, \gamma_2) \in D_1^+$. Let $\beta_1, \beta_2 \in (-\alpha, \alpha)$ with $\beta_1 \le \beta_2$. Suppose $(x_j), (y_j) \in s(\Omega)$ such that*

$$\lim_{j\to\infty} \frac{x_j}{|x_j|} = (\cos\beta_1, \sin\beta_1), \qquad \lim_{j\to\infty} \frac{y_j}{|y_j|} = (\cos\beta_2, \sin\beta_2),$$
$$\lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta_1) \quad \textit{and} \quad \lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\theta_2),$$

for some θ_1, θ_2 such that $(\beta_1, \theta_1), (\beta_2, \theta_2) \in \mathcal{F}$. Set $L = [\min\{\theta_1, \theta_2\}, \max\{\theta_1, \theta_2\}]$. Then the set $\mathcal{F} \cap ([\beta_1, \beta_2] \times L)$ must be connected.

Proof. The proof of Theorem 2.5 (i) and (ii) is essentially the same as that of Theorem 2.4 with Corollaries 2.6 and 2.7 of [Jeffres and Lancaster 2008] replacing Corollaries 2.4 and 2.5 respectively. Conclusion (iii) follows from (i) by standard arguments (e.g., proof of Lemma 4.2). Conclusion (iv) follows from (i) by standard arguments which take into account the specific geometry of $\mathcal{F}$. □

2B. ***Proof of Theorem 2.1.*** Assume $(\gamma_1, \gamma_2) \in D_1^{\pm} \cup D_2^{\pm}$, f satisfies (3) in Ω and (4) on $B(\mathcal{O}, \delta^*) \cap \partial\Omega \setminus \{\mathcal{O}\}$ and f is continuous at $\mathcal{O}$; then f is bounded in a neighborhood of $\mathcal{O}$. Since f is continuous at $\mathcal{O}$, we have the following modifications of (i)–(viii) in Section 2A:

(iii)′ X maps ∂E strictly monotonically onto $\{(x, f(x)) : x \in \partial\Omega\}$.

(vi)′ $X \in C^0(\overline{E})$ and $X(1, 0) = (0, 0, z_0)$, where $z_0 = f(0, 0)$.

(vii)′ Continuing to write $G(u, v) = (x(u, v), y(u, v))$, $G(\cos t, \sin t)$ moves clockwise about $\partial\Omega$ as t increases, $0 \le t \le 2\pi$, and G is an orientation reversing homeomorphism from $\overline{E}$ onto $\overline{\Omega}$.

We will prove Theorem 2.1 in the cases $(\gamma_1, \gamma_2) \in D_2^+$ and $(\gamma_1, \gamma_2) \in D_1^+$; this will suffice to prove the lemma since the mapping

$$\mathbb{R}^3 \to \mathbb{R}^3, \quad (x_1, x_2, x_3) \mapsto (x_1, -x_2, -x_3),$$

converts a D_2^- corner into a D_2^+ corner and converts a D_1^- corner into a D_1^+ corner.

Suppose $(\gamma_1, \gamma_2) \in D_2^+$. Set $\theta_1 = (\pi - (\gamma_1 + \gamma_2))/2$ and let $\theta_2 \in (\alpha - \gamma_1, \alpha)$. By choosing $\delta_0 > 0$ small, we may assume

$$\Omega^* = \{(r\cos\theta, r\sin\theta) : 0 < r < \delta_0,\ \theta_1 < \theta < \theta_2\} \subset \Omega.$$

Notice that Theorem 2.4 (D_2^+) implies

$$\Pi_f(\Omega^*) = \{(\cos\theta, \sin\theta, 0) : \theta_1 + \tfrac{1}{2}\pi \le \theta \le \alpha - \gamma_1 + \tfrac{1}{2}\pi\}.$$

Since $\alpha - \gamma_1 - \theta_1 = \frac{1}{2}(2\alpha - \pi - \gamma_1 + \gamma_2) \in (2\alpha - \pi, \alpha) \subset (0, \pi)$, the hypotheses of Lemma 2.2 are satisfied (with M a rotation through an angle of $\pi/2 - \alpha$). If ϕ is a conformal map from E onto $G^{-1}(\Omega^*)$ which maps $(1, 0)$ to $(1, 0)$ and $\tilde{g}$ is defined by (10), then Lemma 2.2 implies there exists $p > 2$ such that

$$\tilde{g}(\zeta) = \psi(\zeta) + h(\zeta), \tag{15}$$

where ψ is a holomorphic function and $h \in L^\infty(E_1)$ is a Hölder continuous function on $\overline{E}_1$ with Hölder exponent $\mu = (p - 2)/p$. The assumption that f is continuous at $\mathcal{O}$ yields a contradiction as in §3 of [L] (i.e., the Phragmén–Lindelöf theorem is violated).

Now suppose $(\gamma_1, \gamma_2) \in D_1^+$. Let $\theta_1 \in (-\alpha, -\alpha + \gamma_2)$ and $\theta_2 \in (\alpha - \gamma_1, \alpha)$ and choose $\delta_0 > 0$ small enough that

$$\Omega^* = \{(r\cos\theta, r\sin\theta) : 0 < r < \delta_0,\ \theta_1 < \theta < \theta_2\} \subset \Omega.$$

Using Theorem 2.5, we see that

$$\Pi_f(\Omega^*) \subset \{(\cos\theta, \sin\theta, 0) : \beta \in (-\alpha, \alpha), (\beta, \theta) \in \mathcal{F}_L\},$$

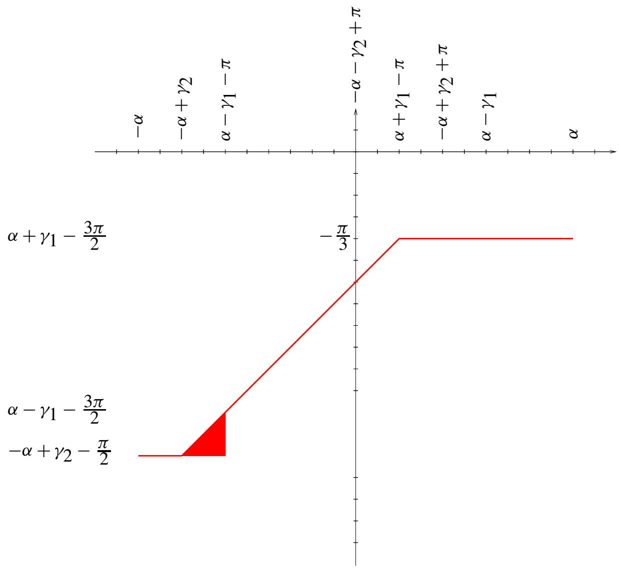


Figure 9. $\mathscr{F}_A$ at a D_1^+ corner; $\alpha=\frac{5\pi}{6}$, $\gamma_1=\frac{\pi}{3}$, $\gamma_2=\frac{\pi}{6}$.

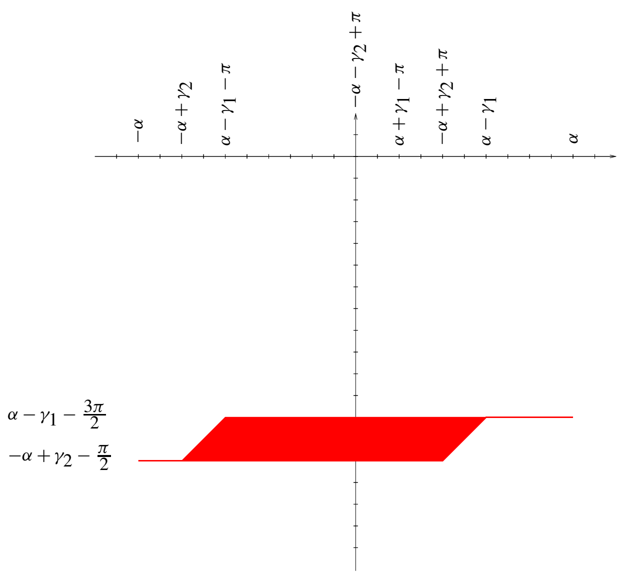


Figure 10. $\mathscr{F}_B$ at a D_1^+ corner; $\alpha=\frac{5\pi}{6}$, $\gamma_1=\frac{\pi}{3}$, $\gamma_2=\frac{\pi}{6}$.

where $\mathscr{F}_L$ is one of the sets $\mathscr{F}_A$, $\mathscr{F}_B$ or $\mathscr{F}_C$ illustrated in Figures 9, 10 and 11 respectively. When $\mathscr{F}_L$ is $\mathscr{F}_A$ or $\mathscr{F}_C$, the proof is essentially that same as that above for the case in which $(\gamma_1, \gamma_2) \in D_2^+$. When $\mathscr{F}_L$ is $\mathscr{F}_B$, the proof is essentially that same as that in §3 of [L]. $\square$

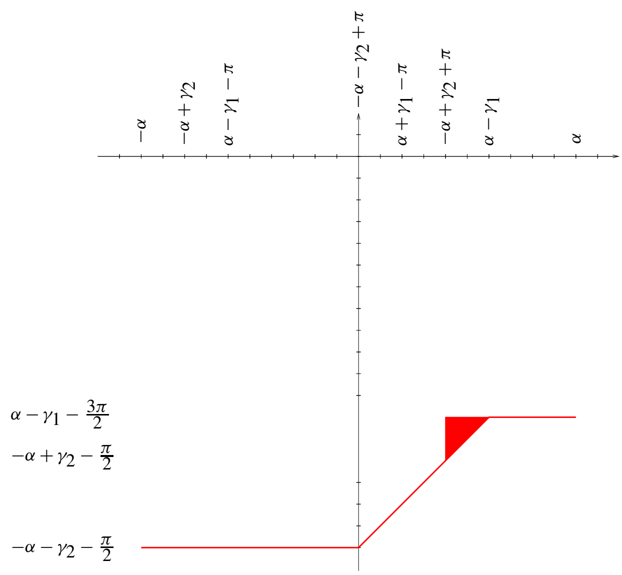


Figure 11. $\mathcal{F}_C$ at a D_1^+ corner; $\alpha = \frac{5\pi}{6}$, $\gamma_1 = \frac{\pi}{3}$, $\gamma_2 = \frac{\pi}{6}$.

3. The exact sizes of fans

We recall that a solution $f \in C^2(\Omega)\cap C^{1,\rho}(B(\mathbb{O},\delta)\cap\overline{\Omega}\setminus\{\mathbb{O}\})$ of (3)–(4) is unbounded if $\alpha < \pi/2$ and $(\gamma_1,\gamma_2) \in D_1^{\pm}$, for some $\delta > 0$ and $\rho \in (0,1)$. The following lemma justifies the definition of $\vec{m}_f : (-\alpha,\alpha) \to S^2_-$, given by

$$\vec{m}_f(\beta) = \lim_{j\to\infty} \vec{n}_f(x_j) \text{ whenever } (x_j) \in s(\Omega) \text{ with } \lim_{j\to\infty} \frac{x_j}{|x_j|} = (\cos\beta, \sin\beta),$$

when $(\gamma_1,\gamma_2) \in D_1^{\pm} \cup D_2^{\pm}$.

Lemma 3.1. *Let Ω and γ be as in Section 1, with $\alpha \in [0,\pi]$. For some $\rho \in (0,1)$ and $\delta > 0$, suppose $f \in C^2(\Omega)\cap C^{1,\rho}(B(\mathbb{O},\delta)\cap\overline{\Omega}\setminus\{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty = \sup_{x\in\Omega}|H(x, f(x))| < \infty$. Suppose* (5) *holds, $\gamma_1,\gamma_2 \in (0,\pi)$ and $(\gamma_1,\gamma_2) \in D_1^{\pm}\cup D_2^{\pm}$; that is, either $\alpha \in [0,\frac{\pi}{2})$ and $|\gamma_1-\gamma_2| > \pi - 2\alpha$ holds or $\alpha \in [\frac{\pi}{2},\pi]$ and one of $|\gamma_1-\gamma_2| > 2\alpha-\pi$ or $|\gamma_1+\gamma_2-\pi| > 2\pi-2\alpha$ holds. Then the Gauss map from $\mathcal{S}_f$ to S^2_- is continuous on $\mathcal{S}_f \cap (\overline{\Omega(\epsilon)}\times\mathbb{R})$ for each $\epsilon > 0$, where $\Omega(\epsilon) = \{(r\cos\theta, r\sin\theta) \in \Omega : r > 0, |\theta| < \alpha-\epsilon\}$. In particular, $\vec{m}_f(\beta)$ exists for all $\beta \in (-\alpha,\alpha)$ and $\vec{m}_f \in C^0((-\alpha,\alpha) : S^2_-)$.*

Proof. Using Theorem 2.1 of [LS a] when $\alpha < \pi/2$ and $(\gamma_1,\gamma_2) \in D_2^{\pm}$ and Theorems 2.4 and 2.5 and the proof of Theorem 2.1 when $\alpha \geq \pi/2$ and (γ_1,γ_2) is in $D_1^{\pm}\cup D_2^{\pm}$, we see that the hypotheses of Lemma 2.2 are satisfied when $\epsilon > 0$ and $\Lambda = \Omega(\epsilon)$. Therefore the restriction of the Gauss map to $\mathcal{S}_f \cap (\overline{\Omega(\epsilon)}\times\mathbb{R})$ is continuous. $\square$

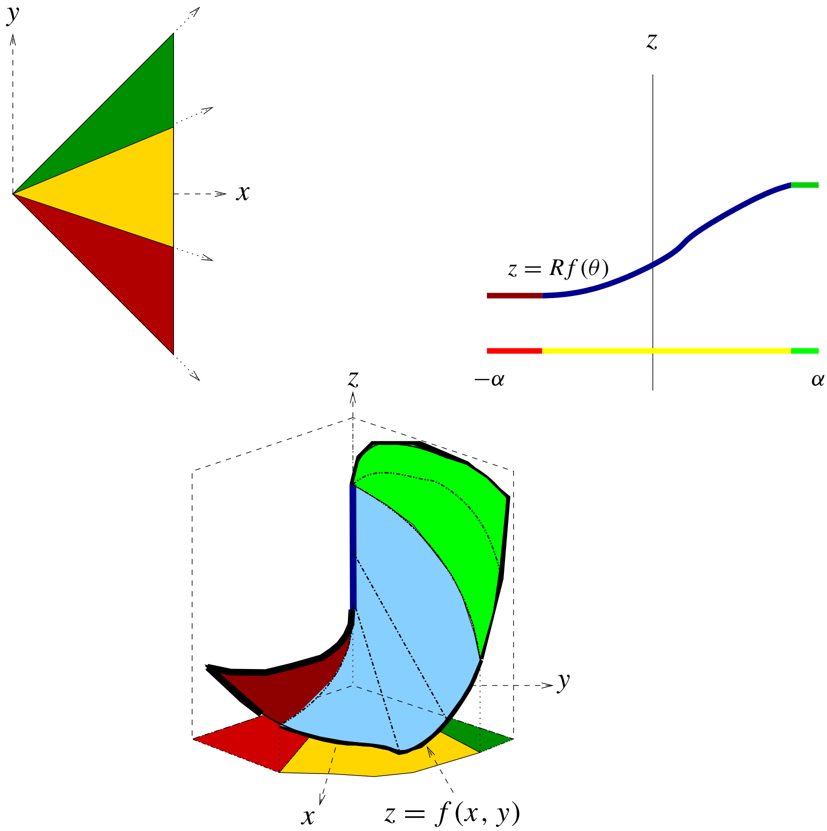


Figure 12. Radial limits: Side fans at a convex corner.

Now we wish to determine the exact sizes of the side fans (illustrated in Figure 12) when $(\gamma_1, \gamma_2) \in D_1^{\pm} \cup D_2^{\pm}$. From [LS a], Theorems 1 and 2, we know that if f is discontinuous at $\mathcal{O}$, then Rf and the limits at $\mathcal{O}$ of the Gauss map behave in the following ways; here $\mathcal{I}$ denotes a countable subset of the appropriate interval(s) and $\omega(\theta) = (\cos\theta, \sin\theta, 0)$ for $\theta \in \mathbb{R}$.

Case (I) $$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly increasing} & \text{on } [\alpha_1, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1]. \end{cases}$$

This case can only occur when $(\gamma_1, \gamma_2) \in R \cup D_2^{+} \cup D_1^{\pm}$. Theorem 2 of [LS a] implies $\alpha_2 \le \alpha - \gamma_1$ and $\alpha_1 \ge -\alpha + \pi - \gamma_2$. If $\beta \in (\alpha_1, \alpha_2)$ then $\vec{m}_f(\beta) = \omega(\beta + \frac{\pi}{2})$.

Case (D) $$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly decreasing} & \text{on } [\alpha_1, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1]. \end{cases}$$

This case can only occur when $(\gamma_1, \gamma_2) \in R \cup D_2^- \cup D_1^\pm$. Theorem 2 of [LS a] implies $\alpha_2 \le \alpha - \pi + \gamma_1$ and $\alpha_1 \ge -\alpha + \gamma_2$. If $\beta \in (\alpha_1, \alpha_2)$, then $\vec{m}_f(\beta) = \omega(\beta - \frac{\pi}{2})$.

Case (DI) There exists $\theta_0 \in (-\alpha + \gamma_2, \alpha - \gamma_1 - \pi)$ such that

$$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly increasing} & \text{on } [\theta_0 + \pi, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [\theta_0, \theta_0 + \pi], \\ \text{strictly decreasing} & \text{on } [\alpha_1, \theta_0] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_2]. \end{cases}$$

This case can only occur when $(\gamma_1, \gamma_2) \in D_1^+$. Theorem 2 of [LS a] implies $\alpha_2 \le \alpha - \gamma_1$ and $\alpha_1 \ge -\alpha + \gamma_2$. If $\beta \in (-\alpha, \alpha)$, then

$$\vec{m}_f(\beta) = \begin{cases} \omega(\beta + \frac{\pi}{2}) & \text{if } \beta \in (\theta_0 + \pi, \alpha_2), \\ \omega(\theta_0 - \frac{\pi}{2}) & \text{if } \beta \in [\theta_0, \theta_0 + \pi], \\ \omega(\beta - \frac{\pi}{2}) & \text{if } \beta \in (\alpha_1, \theta_0). \end{cases}$$

Case (ID) There exists $\theta_0 \in (-\alpha + \pi - \gamma_2, \alpha + \gamma_1 - 2\pi)$ such that

$$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly decreasing} & \text{on } [\theta_0 + \pi, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [\theta_0, \theta_0 + \pi], \\ \text{strictly increasing} & \text{on } [\alpha_1, \theta_0] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1]. \end{cases}$$

This case can only occur when $(\gamma_1, \gamma_2) \in D_1^-$. Theorem 2 of [LS a] implies $\alpha_2 \le \alpha - \pi + \gamma_1$ and $\alpha_1 \ge -\alpha + \pi - \gamma_2$. If $\beta \in (-\alpha, \alpha)$, then

$$\vec{m}_f(\beta) = \begin{cases} \omega(\beta - \frac{\pi}{2}) & \text{if } \beta \in (\theta_0 + \pi, \alpha_2), \\ \omega(\theta_0 + \frac{\pi}{2}) & \text{if } \beta \in [\theta_0, \theta_0 + \pi], \\ \omega(\beta + \frac{\pi}{2}) & \text{if } \beta \in (\alpha_1, \theta_0). \end{cases}$$

Theorem 3.2. *Let Ω and γ be as in Section 1, with $\alpha \in [\frac{\pi}{2}, \pi]$. For some $\rho \in (0, 1)$ and $\delta > 0$, suppose $f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathcal{O}, \delta) \cap \overline{\Omega} \setminus \{\mathcal{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty < \infty$. Suppose* (5) *holds, $\gamma_1, \gamma_2 \in (0, \pi)$ and $(\gamma_1, \gamma_2) \in D_1^\pm \cup D_2^\pm$. Then*:

(i) *In* Case (I), $\alpha_1 = -\alpha + \pi - \gamma_2$ *and* $\alpha_2 = \alpha - \gamma_1$.

(ii) *In* Case (D), $\alpha_1 = -\alpha + \gamma_2$ *and* $\alpha_2 = \alpha - \pi + \gamma_1$.

(iii) *In* Case (DI), $\alpha_1 = -\alpha + \gamma_2$ *and* $\alpha_2 = \alpha - \gamma_1$.

(iv) *In* Case (ID), $\alpha_1 = -\alpha + \pi - \gamma_2$ *and* $\alpha_2 = \alpha - \pi + \gamma_1$.

Proof. Suppose $(\gamma_1, \gamma_2) \in D_2^{\pm}$; the argument is the same when $\alpha < \pi/2$ and when $\alpha \geq \pi/2$. Let us assume $(\gamma_1, \gamma_2) \in D_2^{+}$; hence Case (I) holds. Then Figure 5 illustrates the conclusions of Theorem 2.1 of [L] and Theorem 2.4. Suppose there exists $\alpha_2 < \alpha - \gamma_1$ (and $\alpha_1 \geq -\alpha + \pi - \gamma_2$) such that

$$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly increasing} & \text{on } [\alpha_1, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1]. \end{cases}$$

If we define

$$\Omega' = \{(r\cos\beta, r\sin\beta) \in \Omega : 0 < r < \delta, \alpha_2 < \beta < \pi\}$$

for $\delta > 0$ sufficiently small, then $f \in C^0(\overline{\Omega'})$ and we may apply the arguments in the proof of Theorem 2.1, using Ω' as our domain, to obtain a contradiction. If $\alpha_1 > -\alpha + \pi - \gamma_2$, a similar argument yields a contradiction.

Now suppose $(\gamma_1, \gamma_2) \in D_1^{+}$, Case (I) holds and there exist $\alpha_2 < \alpha - \gamma_1$ and $\alpha_1 \geq -\alpha + \pi - \gamma_2$ such that

$$Rf \text{ is } \begin{cases} \text{constant} & \text{on } [\alpha_2, \alpha], \\ \text{strictly increasing} & \text{on } [\alpha_1, \alpha_2] \setminus \mathcal{I}, \\ \text{constant} & \text{on } [-\alpha, \alpha_1]. \end{cases}$$

Let $\theta_1 \in (-\alpha, -\alpha_1 + \pi - \gamma_2)$ and $\theta_2 \in (\alpha - \gamma_1, \alpha)$. By choosing $\delta_0 > 0$ small, we may assume $\Omega^* = \{(r\cos\theta, r\sin\theta) : 0 < r < \delta_0,\ \theta_1 < \theta < \theta_2\}$ is a subset of Ω. Set $\Omega' = \{(r\cos\beta, r\sin\beta) : 0 < r < \delta_0, \alpha_2 < \theta < \theta_2\}$ and notice that $f \in C^0(\overline{\Omega'})$. Now Theorem 2.5, Lemma 3.1 and the fact that

$$\lim_{j\to\infty} \vec{n}_f(x_j) = \omega(\beta + \tfrac{\pi}{2})$$

when $\beta \in (\alpha_1, \alpha_2)$ and $(x_j) \in s(\Omega)$ such that $\lim_{j\to\infty} x_j/|x_j| = (\cos\beta, \sin\beta)$ implies that

$$\Pi_f(\Omega^*) \subset \{\omega(\theta) : \beta \in [\theta_1, \theta_2], (\beta, \theta) \in \mathcal{F}_C\}$$

and $\vec{m}_f(\cdot) \in C^0((-\alpha, \alpha))$. If $\vec{m}_f(\alpha_2) \neq \omega(\alpha - \gamma_1 + \frac{\pi}{2})$, then we may apply the arguments in the proof of Theorem 2.1, using Ω' as our domain, to get a contradiction. If $\vec{m}_f(\alpha_2) = \omega(\alpha - \gamma_1 + \frac{\pi}{2})$, then $\vec{m}_f$ is discontinuous at α_2, which is a contradiction. Therefore $\alpha_2 = \alpha - \gamma_1$. The argument that $\alpha_1 = -\alpha + \pi - \gamma_2$ is similar.

The proof of the theorem when $(\gamma_1, \gamma_2) \in D_1^{+}$ and one of Cases (D), (DI) or (ID) occurs follows in a similar manner. The situation where $(\gamma_1, \gamma_2) \in D_1^{-}$ follows from this. □

4. Continuity of the Gauss map

Notice that Lemma 3.1 and the proof of Theorem 3.2 imply that the (nonparametric) Gauss map is continuous on $\mathcal{S}_f \cap (\overline{\Omega_\epsilon} \times \mathbb{R})$ for each $\epsilon > 0$ and, in each case, we have:

(I): $\lim_{\beta\to\alpha} \vec{m}_f(\beta) = \omega(\alpha_2 + \pi/2)$ and $\lim_{\beta\to-\alpha} \vec{m}_f(\beta) = \omega(\alpha_1 + \pi/2)$.

(D): $\lim_{\beta\to\alpha} \vec{m}_f(\beta) = \omega(\alpha_2 - \pi/2)$ and $\lim_{\beta\to-\alpha} \vec{m}_f(\beta) = \omega(\alpha_1 - \pi/2)$.

(DI): $\lim_{\beta\to\alpha} \vec{m}_f(\beta) = \omega(\alpha_2 + \pi/2)$ and $\lim_{\beta\to-\alpha} \vec{m}_f(\beta) = \omega(\alpha_1 - \pi/2)$.

(ID): $\lim_{\beta\to\alpha} \vec{m}_f(\beta) = \omega(\alpha_2 - \pi/2)$ and $\lim_{\beta\to-\alpha} \vec{m}_f(\beta) = \omega(\alpha_1 + \pi/2)$.

In order to conclude that the Gauss map is in $C^0(\mathcal{S}_f \cap (B(\mathbb{O}, \delta)\times\mathbb{R}) : S^2_-)$, it would be sufficient to blow up the graph of $u(x) = f(x) - Rf(\alpha)$ (or $u(x) = f(x) - Rf(-\alpha)$) about (0, 0, 0) tangent to $\partial^+\Omega$ (or $\partial^-\Omega$ respectively) and know that a subsequence converges to an appropriate cone. If one is willing to accept this hypothesis, then the claim that the Gauss map is in $C^0(\mathcal{S}_f \cap (B(\mathbb{O}, \delta) \times \mathbb{R}) : S^2_-)$ can be proven.

Hypothesis (B±). For all $(x_j) \in s(\Omega)$ with $\lim_{j\to\infty} x_j/|x_j| = (\cos(\pm\alpha), \sin(\pm\alpha))$, there is a subsequence (x_{j_k}) and a function $u_\infty : \Omega_\infty \to [-\infty, \infty]$ such that the subgraph $U_\infty = \{(x, t) \in \Omega_\infty \times \mathbb{R} : t < u_\infty(x)\}$ of u_∞ is a cone with respect to (0, 0, 0), there exists $\vec{\xi} \in S^2_-$ such that $\lim_{k\to\infty} \vec{n}(x_{j_k}) = \vec{\xi} = (\xi_1, \xi_2, \xi_3)$,

$$\lim_{k\to\infty} \operatorname{dist}\big(\{(x, u_{j_k}(x)) \in \Omega_{j_k}(\delta, b)\}, \partial U_\infty \cap \Omega_{j_k}(\delta, b)\big) = 0$$

for each $\delta > 0$ and $b > 0$, where $\epsilon_j = |x_j|$, $u_j(x) = (f(\epsilon_j x) - Rf(\pm\alpha))/\epsilon_j$ and $\Omega_j(\delta, b) = \{(x, t) \in \mathbb{R}^3 : x \in B(\mathbb{O}, \delta), \epsilon_j x \in \Omega, t \in (-b, b)\}$ for $j \in \mathbb{N}$, and

(a) if $\xi_3 < 0$, then $\partial U_\infty = \pi_1 \cap (\overline{\Omega_\infty} \times \mathbb{R})$, π_1 is a nonvertical plane with downward unit normal $\vec{\xi} \in S^2_-$, $\vec{\xi}$ makes an angle of γ_1 with the exterior unit normal to $\partial^+\Omega_\infty \times \mathbb{R}$ and an angle of γ_2 with the exterior unit normal to $\partial^-\Omega_\infty \times \mathbb{R}$ and $\vec{n}_{u_{j_k}} \to \vec{\xi}$ uniformly on compacta in $\Omega \times \mathbb{R}$ as $k \to \infty$,

(b) if $\xi_3 = 0$, then $\partial U_\infty = \partial\mathcal{P} \cap (\overline{\Omega_\infty} \times \mathbb{R})$, $\mathcal{P} = \{(r\cos\theta, r\sin\theta) : r > 0, \theta \in \Theta\}$ and, for each $x \in \partial\mathcal{P} \cap \Omega_\infty$, $\vec{n}_{u_{j_k}}(x) \to \vec{\xi}(x) \times \{0\}$, where $\mathcal{P} = \{x \in \Omega_\infty : u_\infty(x) = \infty\}$, $\vec{\xi}(x)$ is the interior (with respect to $\mathcal{P}$) unit normal vector to $\partial\mathcal{P}$ at x and Θ is one of the following sets: $(\alpha - \gamma_1, \alpha)$, $(-\alpha, -\alpha + \gamma_2)$, $(-\alpha + \pi - \gamma_2, \alpha - \pi + \gamma_1)$ (provided $(\alpha - \pi + \gamma_1) - (-\alpha + \pi - \gamma_2) \geq \pi$) or $(-\alpha, -\alpha + \gamma_2) \cup (\alpha - \gamma_1, \alpha)$ (provided $(\alpha - \gamma_1) - (-\alpha + \gamma_2) \geq \pi$).

Theorem 4.1. *Let Ω and γ be as in Section 1. For some $\rho \in (0, 1)$ and $\delta > 0$, suppose $f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathbb{O}, \delta) \cap \overline{\Omega} \setminus \{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty < \infty$. Suppose* (5) *holds, $\gamma_1, \gamma_2 \in (0, \pi)$ and either $\alpha < \pi/2$ and (γ_1, γ_2) is in $D^\pm_2$ or $\alpha \geq \pi/2$ and (γ_1, γ_2) is in $D^\pm_1 \cup D^\pm_2$. Suppose that* Hypotheses (B±) *are true. Then $\vec{n}_f : \mathcal{S}_f \times (\Omega \times \mathbb{R})$ extends to be continuous on $\mathcal{S}_f \cap (B(\mathbb{O}, \delta) \times \mathbb{R}))$ and $N \in C^0(E \cup \{(u, v) \in \partial E : G(u, v) \in \partial\Omega \cap B(\mathbb{O}, \delta)\})$.*

The proof of this theorem will follow from the information at the beginning of this section about the behavior of $\vec{m}_f$ and from Lemmas 4.2–4.5. Set

$$\begin{gathered}
C(x)=\{X\in S^2_-: X\cdot\nu(x)=\cos\gamma(x)\},\\
\Gamma_1=\{X\in S^2_-: X\cdot(\cos(\alpha+\tfrac{\pi}{2}),\sin(\alpha+\tfrac{\pi}{2}),0)=\cos\gamma_1\},\\
\Gamma_2=\{X\in S^2_-: X\cdot(\cos(-\alpha-\tfrac{\pi}{2}),\sin(-\alpha-\tfrac{\pi}{2}),0)=\cos\gamma_2\},\\
\vec{\xi}_A=\omega(\alpha-\gamma_1+\tfrac{\pi}{2})\in\Gamma_1,\quad \vec{\xi}_B=\omega(\alpha+\gamma_1-\tfrac{3\pi}{2})\in\Gamma_1,\\
\vec{\xi}_C=\omega(-\alpha-\gamma_2+\tfrac{3\pi}{2})\in\Gamma_2,\quad \vec{\xi}_D=\omega(-\alpha+\gamma_2-\tfrac{\pi}{2})\in\Gamma_2,\\
\Omega_\infty=\{(r\cos(\theta),\ r\sin(\theta)): r>0,\ -\alpha<\theta<\alpha\},\\
\Sigma^1_\infty=\{(r\cos(\alpha),\ r\sin(\alpha)): r>0\},\quad \Sigma^2_\infty=\{(r\cos(\alpha),\ -r\sin(\alpha)): r>0\},\\
\nu^+_\infty=(\cos(\alpha+\tfrac{\pi}{2}),\sin(\alpha+\tfrac{\pi}{2}),0)\quad\text{and}\quad \nu^-_\infty=(\cos(-\alpha-\tfrac{\pi}{2}),\sin(-\alpha-\tfrac{\pi}{2}),0).
\end{gathered}$$

Lemma 4.2. *Let Ω and γ be as in Section 1. For some $\rho\in(0,1)$ and $\delta>0$, suppose $f\in C^2(\Omega)\cap C^{1,\rho}(B(\mathbb{O},\delta)\cap\overline{\Omega}\setminus\{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty<\infty$. Suppose* (5) *holds, $\gamma_1,\gamma_2\in(0,\pi)$ and either $\alpha<\pi/2$ and (γ_1,γ_2) is in D^+_2, so that Rf behaves as in* Case (I), *or $\alpha\ge\pi/2$ and (γ_1,γ_2) is in $D^\pm_1\cup D^+_2$ and Rf behaves as in* Case (I) *or* Case (DI). *Assume* Hypothesis (B+) *is true. Then*

$$\lim_{j\to\infty}\vec{n}(x_j)=(\cos(\alpha-\gamma_1+\pi/2),\sin(\alpha-\gamma_1+\pi/2),0)=\vec{\xi}_A \tag{16}$$

for every $(x_j)\in s(\Omega)$ with $\lim_{j\to\infty}x_j/|x_j|=(\cos\alpha,\sin\alpha)$.

Proof. Since $\gamma(x)\to\gamma_1$ and $\nu(x)\to\nu^+_\infty$ as $x\in\partial^+\Omega$ converges to $\mathbb{O}$, we see that $\operatorname{dist}(C(x),\Gamma_1)\to 0$ as $x\in\partial^+\Omega$ converges to $\mathbb{O}$. Similarly, $\gamma(x)\to\gamma_2$, $\nu(x)\to\nu^-_\infty$ and $\operatorname{dist}(C(x),\Gamma_2)\to 0$ as $x\in\partial^-\Omega$ converges to $\mathbb{O}$. Thus

$$\operatorname{dist}(\vec{n}(x^+(s)),\Gamma_1)\to 0\quad\text{and}\quad\operatorname{dist}(\vec{n}(x^-(s)),\Gamma_2)\to 0 \tag{17}$$

as $s\to 0+$.

Suppose that $(x_j)\in s(\Omega)$ with $\lim_{j\to\infty}x_j/|x_j|=(\cos\alpha,\sin\alpha)$; then there is a subsequence, still denoted (x_j), and $\vec{\xi}\in S^2_-$ such that $\lim_{j\to\infty}\vec{n}(x_j,y_j)\to\vec{\xi}$. Notice that $\vec{\xi}\in\Gamma_1$ since $f\in C^{1,\rho}(B(\mathbb{O},\delta)\cap\overline{\Omega}\setminus\{\mathbb{O}\})$ and (17) holds.

Assume first that $\vec{\xi}=(\xi_1,\xi_2,\xi_3)$ with $\xi_3<0$. For each $j\in\mathbb{N}$, define $\epsilon_j=|x_j|$, $\Omega_j=\{x\in\mathbb{R}^2:\epsilon_j x\in\Omega\}$ and $u_j\in C^\infty(\Omega_j)\cap C^1(\overline{\Omega_j}\setminus\{\mathbb{O}\})$ by

$$u_j(x)=\frac{1}{\epsilon_j}(f(\epsilon_j x)-Rf(\alpha)). \tag{18}$$

Let γ_j be defined on $\partial\Omega_j\backslash\{\mathbb{O}\}$ by $\gamma_j(x)=\gamma(\epsilon_j x)$ and let ν_j denote the outward unit normal to $\partial\Omega_j$. Then u_j satisfies the prescribed mean curvature problem

$$Nu_j(x)=\epsilon_j H(\epsilon_j x,\ f(\epsilon_j x)),\quad x\in\Omega_j, \tag{19}$$

$$Tf_j\cdot\nu_j=\cos\gamma_j \quad \text{on } \partial\Omega_j\backslash\{\mathbb{O}\}. \tag{20}$$

Hypothesis (B+) implies that there is a nonvertical plane π_1 with downward unit normal $\vec{\xi}$ which meets $\partial^+\Omega_\infty$ in an angle of γ_1 and $\partial^-\Omega_\infty$ in an angle of γ_2 in the sense described in (a); however this is impossible since $(\gamma_1,\gamma_2)\in D_1^{\pm}\cup D_2^{\pm}$. Thus $\xi_3=0$ and so either $\vec{\xi}=\vec{\xi}_A$ or $\vec{\xi}=\vec{\xi}_B$. The intermediate value theorem implies that

$$\text{if } \vec{\xi}=\vec{\xi}_A,\quad \text{then } \vec{n}(x)\to\vec{\xi}_A \text{ as } x\in\partial^+\Omega \text{ converges to } \mathbb{O}, \tag{21}$$

$$\text{if } \vec{\xi}=\vec{\xi}_B,\quad \text{then } \vec{n}(x)\to\vec{\xi}_B \text{ as } x\in\partial^+\Omega \text{ converges to } \mathbb{O}. \tag{22}$$

Suppose (22) holds. Notice then that

$$\lim_{s\to 0+}\frac{d}{ds}f^+(s)=-\infty, \tag{23}$$

since $(\cos\alpha,\sin\alpha,0)\cdot\vec{\xi}_B=-\sin(\gamma_1)<0$, and so $f^+(s)=f(x^+(s))$ is a strictly decreasing function of s for $0\le s\le s_0$, where $s_0>0$ is sufficiently small. Since $\vec{m}_f(\beta)=\omega(\alpha-\gamma_1+\frac{\pi}{2})$ when $\beta\in[\alpha-\gamma_1,\alpha)$, we have

$$\lim_{r\to 0+}\nabla f(r\cos\beta,r\sin\beta)\cdot(\cos\beta,\sin\beta)=+\infty\quad \text{for } \beta\in(\alpha-\gamma_1,\alpha). \tag{24}$$

Since Rf behaves as in Case (I) or (DI), we have

$$Rf(\beta)<Rf(\alpha)=f^+(0)\quad \text{if } \beta\in[\alpha-\pi,\alpha-\gamma_1). \tag{25}$$

Let Ω_H be the connected component of

$$\{(r\cos\beta,r\sin\beta)\in\Omega : r>0,\alpha-\pi<\beta<5\pi/4\}$$

that contains $\{(r\cos\beta,r\sin\beta):0<r<\delta,\alpha-\gamma_1<\beta<\alpha-\epsilon\}$ for sufficiently small $\epsilon,\delta>0$. Consider the $k=f^+(0)$ $(=Rf(\alpha))$ level set of f in Ω_H. From (23), we see that there is a component C of $\{x\in\Omega_H : f(x)<k\}$ whose boundary contains $\partial^+\Omega\cap B(\mathbb{O},\tau)$ for $\tau>0$ sufficiently small; let c_α be the component of $\Omega_H\cap\partial C$ whose closure contains $\mathbb{O}$. Then (23) and (24) imply that for every $\beta_1<\alpha$ and $\beta_2>\alpha$, there exists $\epsilon>0$ such that

$$c_\alpha\cap B(\mathbb{O},\epsilon)\subset\{(r\cos\theta,r\sin\theta):0<r<\epsilon,\beta_1<\theta<\beta_2\};$$

in this sense, c_α is tangent to $\theta=\alpha$ at $\mathbb{O}$. Similarly, using (24) and (25), we see that there is a k-level curve of f, denoted $c_{\alpha-\gamma_1}$, which is tangent to $\theta=\alpha-\gamma_1$ at $\mathbb{O}$ in the sense that for every $\beta_1<\alpha-\gamma_1$ and $\beta_2>\alpha-\gamma_1$, there exists $\epsilon>0$ such that

$$c_{\alpha-\gamma_1}\cap B(\mathbb{O},\epsilon)\subset\{(r\cos\theta,r\sin\theta):0<r<\epsilon,\beta_1<\theta<\beta_2\}.$$

Now pick $\tau > 0$ small enough that the region bounded by c_α, $c_{\alpha-\gamma_1}$ and $\{r = \tau\}$ is well-defined, connected and simply connected; let us rotate this region about $\mathbb{O}$ through an angle $(\pi + \gamma_1)/2 - \alpha$ and denote this open set as Ω^τ, so that $\partial\Omega^\tau$ is tangent to $\theta = (\pi \pm \gamma_1)/2$ at $\mathbb{O}$. Notice that $\tilde{f} = f \circ R^{-1} \in C^0(\overline{\Omega^\tau})$ if R denotes the rotation above. We will let a particular portion of a suitable nodoid be the graph of a comparison function h over a domain $U^\tau \subset \Omega^\tau$ with $B(\mathbb{O}, \epsilon) \cap U^\tau = B(\mathbb{O}, \epsilon) \cap \Omega^\tau$ for some $\epsilon > 0$. Now ∂U^τ will be consist of two disjoint, connected curves, $\partial_1 U^\tau \subset \partial\Omega^\tau \setminus \{r = \tau\}$ and $\partial_2 U^\tau$, with $\mathbb{O} \in \partial_1 U^\tau$ and $\mathbb{O} \notin \overline{\partial_2 U^\tau}$. The comparison function $h \in C^0(\overline{U^\tau}) \cap C^1(U^\tau \cup \partial_1 U^\tau)$ will have the properties $h(\mathbb{O}) = k$, $\frac{\partial h}{\partial x_2}(\mathbb{O}) < \infty$, $h \geq k = \tilde{f}$ on $\partial_1 U^\tau$, $Nh \leq \inf_{x \in \Omega} Nf(x)$ on U^τ and $\frac{\partial h}{\partial \eta} = \nabla f \cdot \eta = +\infty$ on $\partial_2 U^\tau$, where η is the exterior unit normal to $\partial_2 U^\tau$. The comparison principle then implies $\tilde{f} \leq h$ on $\overline{U^\tau}$. This yields a contradiction of (22) since (24) implies

$$\lim_{x_2 \downarrow 0} \frac{\partial \tilde{f}}{\partial x_2}(0, x_2) = +\infty,$$

and the facts that $\tilde{f}(\mathbb{O}) = h(\mathbb{O})$, $\dfrac{\partial h}{\partial x_2}(\mathbb{O}) < \infty$ and $\tilde{f} \leq h$ imply

$$\liminf_{x_2 \downarrow 0} \frac{\partial \tilde{f}}{\partial x_2}(0, x_2) < \infty.$$

This implies (21) holds and completes the proof of Lemma 4.2 except for the construction of the comparison function h.

Let $\mathscr{C}$ be a nodary in $\{x \in \mathbb{R}^2 : x_2 > 0\}$ which, when rotated about the x_1-axis, generates a nodoid in $\mathbb{R}^3$ with constant mean curvature $H_D = |H|_\infty$, which we assume is positive; if not, set $H_D = 1$. (See, for example, [Eells 1987; Mladenov 2002; Rossman 2005] for discussions of Delaunay surfaces and nodoids.) Let the minimal and maximal radii of the nodary be r_0 and R_0 respectively, so that $r_0 \leq x_2 \leq R_0$ whenever $(x_1, x_2) \in \mathscr{C}$; we will assume $(0, r_0) \in \mathscr{C}$. Now let $\mathscr{D} \subset \mathscr{C}$ be the particular open inner loop of the nodary which contains $(0, r_0)$ (i.e., $(0, r_0) \in \mathscr{D}$ and $\mathscr{D}$ does not contain endpoints); notice that the unit normal to the nodary at the endpoints of $\overline{\mathscr{D}}$ are parallel to the axis of rotation of the nodoid and the surface

$$S_{\mathscr{D}} = \{(x_1, x_2 \cos\theta, x_2 \sin\theta) : (x_1, x_2) \in \mathscr{D}, -\pi \leq \theta \leq 0\}$$

obtained by partially rotating $\mathscr{D}$ about the x_1-axis has constant mean curvature $-H_D$ with respect to its upward unit normal.

Now fix t, $0 < t < r_0$, large enough that $\Phi_1 = \{(x_1, x_2 + t) : x \in \partial\Omega^\tau \cap R(c_\alpha)\}$ and $\Phi_2 = \{(x_1, x_2 + t) : x \in \partial\Omega^\tau \cap R(c_{\alpha-\gamma_1})\}$ both intersect $\mathscr{D}$. Let Σ denote the component of $\overline{\Phi}_1 \cup \overline{\Phi}_2 \setminus \mathscr{D}$ that contains $(0, t)$ and let W be the region bounded by Σ and $\mathscr{D}$. Set $W^\tau = \{(x_1, x_2) : (x_1, x_2 + t) \in W\}$, $\partial_1 W^\tau = \{(x_1, x_2) : (x_1, x_2 + t) \in \Sigma\}$ and $\partial_2 W^\tau = \partial W^\tau \setminus \partial_1 W^\tau$. Notice that $\partial_2 W^\tau \subset \{(x_1, x_2) : (x_1, x_2 + t) \in \mathscr{D}\}$. Now define $h \in C^0(\overline{W^\tau}) \cap C^1(W^\tau \cup \partial_1 W^\tau)$ by

$$h(x_1, x_2) = w(x_1, x_2 + t) - w(0, t) + k$$

for $x \in \overline{W^\tau}$, where $w : \mathfrak{D} \to \mathbb{R}$ such that $S_{\mathfrak{D}}$ is the graph of w. It follows that h has the properties mentioned previously. □

In a similar manner, we can prove each of the following lemmas.

Lemma 4.3. *Let Ω and γ be as in Section 1. For some $\rho \in (0, 1)$ and $\delta > 0$, suppose $f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathbb{O}, \delta) \cap \overline{\Omega} \setminus \{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty < \infty$. Suppose* (5) *holds, $\gamma_1, \gamma_2 \in (0, \pi)$ and either $\alpha < \pi/2$ and (γ_1, γ_2) is in D_2^-, so that Rf behaves as in* Case (D), *or $\alpha \geq \pi/2$ and (γ_1, γ_2) is in $D_1^\pm \cup D_2^-$ and Rf behaves as in* Case (D) *or* Case (ID). *Assume* Hypothesis (B+) *is true. Then*

$$\lim_{j\to\infty} \vec{n}(x_j) = (\cos(\alpha + \gamma_1 - 3\pi/2), \sin(\alpha + \gamma_1 - 3\pi/2), 0) \tag{26}$$

for every $(x_j) \in s(\Omega)$ with $\lim_{j\to\infty} x_j/|x_j| = (\cos\alpha, \sin\alpha)$.

Lemma 4.4. *Let Ω and γ be as in Section 1. For some $\rho \in (0, 1)$ and $\delta > 0$, suppose $f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathbb{O}, \delta) \cap \overline{\Omega} \setminus \{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty < \infty$. Suppose* (5) *holds, $\gamma_1, \gamma_2 \in (0, \pi)$ and either $\alpha < \pi/2$ and (γ_1, γ_2) is in D_2^+, so that Rf behaves as in* Case (I), *or $\alpha \geq \pi/2$ and (γ_1, γ_2) is in $D_1^\pm \cup D_2^+$ and Rf behaves as in* Case (I) *or* Case (ID). *Assume* Hypothesis (B-) *is true. Then*

$$\lim_{j\to\infty} \vec{n}(x_j) = (\cos(-\alpha - \gamma_1 + 3\pi/2), \sin(-\alpha - \gamma_1 + 3\pi/2), 0) \tag{27}$$

for every $(x_j) \in s(\Omega)$ with $\lim_{j\to\infty} x_j/|x_j| = (\cos(-\alpha), \sin(-\alpha))$.

Lemma 4.5. *Let Ω and γ be as in Section 1. For some $\rho \in (0, 1)$ and $\delta > 0$, suppose $f \in C^2(\Omega) \cap C^{1,\rho}(B(\mathbb{O}, \delta) \cap \overline{\Omega} \setminus \{\mathbb{O}\})$ is a bounded solution of* (3)–(4) *with $|H|_\infty < \infty$. Suppose* (5) *holds, $\gamma_1, \gamma_2 \in (0, \pi)$ and either $\alpha < \pi/2$ and (γ_1, γ_2) is in D_2^-, so that Rf behaves as in* Case (D), *or $\alpha \geq \pi/2$ and (γ_1, γ_2) is in $D_1^\pm \cup D_2^-$ and Rf behaves as in* Case (D) *or* Case (DI). *Assume* Hypothesis (B-) *is true. Then*

$$\lim_{j\to\infty} \vec{n}(x_j) = (\cos(-\alpha + \gamma_1 - \pi/2), \sin(-\alpha + \gamma_1 - \pi/2), 0) \tag{28}$$

for every $(x_j) \in s(\Omega)$ with $\lim_{j\to\infty} x_j/|x_j| = (\cos(-\alpha), \sin(-\alpha))$.

References

[Concus and Finn 1996] P. Concus and R. Finn, "Capillary wedges revisited", *SIAM J. Math. Anal.* **27**:1 (1996), 56–69. MR 96m:76006 Zbl 0843.76012

[Eells 1987] J. Eells, "The surfaces of Delaunay", *Math. Intelligencer* **9**:1 (1987), 53–57. MR 88h:53011 Zbl 0605.53002

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

[Finn 1988a] R. Finn, "Comparison principles in capillarity", pp. 156–197 in *Partial differential equations and calculus of variations*, edited by S. Hildebrandt and R. Leis, Lecture Notes in Math. **1357**, Springer, Berlin, 1988. MR 90d:53010 Zbl 0692.35006

[Finn 1988b] R. Finn, "Moon surfaces, and boundary behaviour of capillary surfaces for perfect wetting and nonwetting", *Proc. London Math. Soc.* (3) **57**:3 (1988), 542–576. MR 89m:49076 Zbl 0668.76019

[Finn 1996] R. Finn, "Local and global existence criteria for capillary surfaces in wedges", *Calc. Var. Partial Differential Equations* **4**:4 (1996), 305–322. MR 97f:53005 Zbl 0872.76017

[Finn 1999] R. Finn, "Capillary surface interfaces", *Notices Amer. Math. Soc.* **46**:7 (1999), 770–781. MR 2000g:76033 Zbl 1194.76049

[Finn 2002a] R. Finn, "Eight remarkable properties of capillary surfaces", *Math. Intelligencer* **24**:3 (2002), 21–33. MR 2003f:76041

[Finn 2002b] R. Finn, "Some properties of capillary surfaces", *Milan J. Math.* **70** (2002), 1–23. MR 2003m:53012 Zbl 1053.76009

[Jeffres and Lancaster 2007] T. Jeffres and K. Lancaster, "Vertical blow ups of capillary surfaces in $\mathbb{R}^3$. I. Convex corners", *Electron. J. Differential Equations* **152** (2007), 24 pp. MR 2009b:49103 Zbl 1139.49041

[Jeffres and Lancaster 2008] T. Jeffres and K. Lancaster, "Vertical blow ups of capillary surfaces in $\mathbb{R}^3$. II. Nonconvex corners", *Electron. J. Differential Equations* **160** (2008), 25 pp. MR 2009j:49098 Zbl 1163.49044

[Lancaster 2010] = [L] K. E. Lancaster, "A proof of the Concus–Finn conjecture", *Pacific J. Math.* **247**:1 (2010), 75–108. MR 2011m:53011 Zbl 05809032

[Lancaster and Siegel 1996a] = [LS a] K. E. Lancaster and D. Siegel, "Existence and behavior of the radial limits of a bounded capillary surface at a corner", *Pacific J. Math.* **176**:1 (1996), 165–194. MR 98g:58030a Zbl 0866.76018

[Lancaster and Siegel 1996b] = [LS b] K. E. Lancaster and D. Siegel, "Behavior of a bounded non-parametric H-surface near a reentrant corner", *Z. Anal. Anwendungen* **15**:4 (1996), 819–850. MR 97m:53011 Zbl 0866.35046

[Mladenov 2002] I. M. Mladenov, "Delaunay surfaces revisited", *C. R. Acad. Bulgare Sci.* **55**:5 (2002), 19–24. MR 2003g:53005 Zbl 1017.53010

[Rossman 2005] W. Rossman, "The first bifurcation point for Delaunay nodoids", *Experiment. Math.* **14**:3 (2005), 331–342. MR 2006j:53011 Zbl 1092.53010

[Shi 2006] D. Shi, "Capillary surfaces at a reentrant corner", *Pacific J. Math.* **224**:2 (2006), 321–353. MR 2008a:58013 Zbl 1125.76013

[Tam 1986] L.-F. Tam, "Regularity of capillary surfaces over domains with corners: borderline case", *Pacific J. Math.* **124**:2 (1986), 469–482. MR 87k:49049 Zbl 0604.49029

Received July 12, 2010. Revised July 8, 2012.

KIRK E. LANCASTER
DEPARTMENT OF MATHEMATICS AND STATISTICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67260-0033
UNITED STATES

lancaster@math.wichita.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.393

GENERALIZED NORMAL RULINGS AND INVARIANTS OF LEGENDRIAN SOLID TORUS LINKS

MIKHAIL LAVROV AND DAN RUTHERFORD

For Legendrian links in the 1-jet space of S^1 we show that the 1-graded ruling polynomial may be recovered from the Kauffman skein module. For such links a generalization of the notion of normal ruling is introduced. We show that the existence of such a generalized normal ruling is equivalent to sharpness of the Kauffman polynomial estimate for the Thurston–Bennequin number as well as to the existence of an ungraded augmentation of the Chekanov–Eliashberg DGA. Parallel results involving the HOMFLY-PT polynomial and 2-graded generalized normal rulings are established.

1. Introduction

In $\mathbb{R}^3$ interesting connections exist between the 2-variable knot polynomials and invariants of Legendrian knots. With respect to the standard contact structure on $\mathbb{R}^3$, Fuchs and Tabachnikov [1997] showed that an upper bound for the Thurston–Bennequin number arises from the Kauffman and HOMFLY-PT knot polynomials. Furthermore, when this estimate is sharp some nonclassical invariants exhibit nice properties. Specifically, combining results from [Fuchs 2003; Fuchs and Ishkhanov 2004; Sabloff 2005; Rutherford 2006] we have:

Theorem 1.1. *For a Legendrian link $L \subset \mathbb{R}^3$ the following three statements are all equivalent*:

(1) *The estimate $tb(L) \leq -\deg_a F_L$ (respectively $tb(L) \leq -\deg_a P_L$) is sharp, where $F_L, P_L \in \mathbb{Z}[a^{\pm 1}, z^{\pm 1}]$ denote the Kauffman and HOMFLY-PT polynomials.*

(2) *A front diagram for L has a* 1*-graded* (*respectively* 2*-graded*) *normal ruling.*

(3) *The Chekanov–Eliashberg DGA of L has a* 1*-graded* (*respectively* 2*-graded*) *augmentation.*

In this article, we establish analogous results for Legendrian knots in the 1-jet space of the circle, J^1S^1. The manifold J^1S^1 is topologically an open solid torus and carries a standard contact structure. Legendrian knots in J^1S^1 have attracted

Lavrov received support from NSF CAREER grant number DMS-0846346.
MSC2010: 57M27, 57R17.
Keywords: Legendrian knot, Kauffman polynomial, skein module, normal ruling.

a fair amount of attention in the literature; see [Ding and Geiges 2010; Ng and Traynor 2004; Traynor 1997]. The 1-jet space setting comes with convenient projections from which Legendrian knots may be presented via front or Lagrangian diagrams and Legendrian isotopy may be described in a combinatorial manner. 1-jet spaces also provide a natural setting for the use of generating families.

A convenient formal way to define a normal ruling, ρ, of L is as a family of fixed-point-free involutions of the strands of the front diagram of L subject to many restrictions. At least locally, this may be viewed as a decomposition of the front diagram into pairs of paths. Chekanov and Pushkar [2005] introduced normal rulings — albeit with different terminology — as well as related Legendrian isotopy invariants which have become known as ruling polynomials. In connection with augmentations, Fuchs independently defined normal rulings of knots in $\mathbb{R}^3$ and, in the case of the Kauffman polynomial, already conjectured the equivalence of (1) and (2) in [Fuchs 2003]. This conjecture was verified in [Rutherford 2006], where it was shown that in fact the 1-graded and 2-graded ruling polynomials appear as coefficients of the Kauffman and HOMFLY-PT polynomials respectively.

Relationships between the Kauffman/HOMFLY-PT invariants and Legendrians knots in J^1S^1 have already begun to be studied, and several factors make the situation more interesting. For instance, the HOMFLY-PT polynomial, P_L, of a solid torus link, L, belongs to a polynomial algebra over $R = \mathbb{Z}[a^{\pm 1}, z^{\pm 1}]$ with a countably infinite number of generators A_k, $k \in \mathbb{Z} \setminus \{0\}$; the Kauffman polynomial has a similar form. Chmutov and Goryunov [1997] proved Thurston–Bennequin number estimates analogous to those appearing in (1) of Theorem 1.1 using these many variable Kauffman and HOMFLY-PT polynomials. In the case of the HOMFLY-PT polynomial, it was shown in [Rutherford 2011] that the 2-graded ruling polynomial can be recovered from the HOMFLY-PT polynomial, but this requires first specializing via an R-module homomorphism $R[A_{\pm 1}, A_{\pm 2}, \dots] \to R$. In the present work we develop analogous results involving the 1-graded ruling polynomial and the Kauffman skein module. (See Theorems 3.4 and 3.6.)

The need to specialize the Kauffman and HOMFLY-PT invariants in order to recover the ruling polynomials has an interesting consequence. There are many solid torus links where the Kauffman or HOMFLY-PT polynomial estimate is sharp, yet the corresponding ruling polynomial vanishes. As a result, for Legendrians in J^1S^1 some adjustment is required to statement (2) of Theorem 1.1. For this purpose, we introduce a quite natural notion of *generalized normal ruling* where the fixed-point-free condition is relaxed. Our main result is the following analog of Theorem 1.1:

Theorem 1.2. *Let $L \subset J^1S^1$ be a Legendrian link.*

(1) *The estimate $tb(L) \le -\deg_a F_L$ (respectively $tb(L) \le -\deg_a P_L$) is sharp if and only if L has a* 1*-graded* (*respectively* 2*-graded*) *generalized normal ruling.*

(2) *Suppose L has been assigned a $\mathbb{Z}/p$-valued Maslov potential. The Chekanov–Eliashberg DGA of L has a p-graded augmentation if and only if a front diagram for L admits a p-graded generalized normal ruling.*

Remark 1.3. (i) Aside from allowing the more general p-graded condition in (2), it is natural to organize the three statements into these two equivalences. Even in $\mathbb{R}^3$, the authors do not know of any proof of an implication between the statements about the knot polynomial estimates and existence of augmentations which is able to avoid using normal rulings. There are settings, for instance certain contact lens spaces, where Legendrian contact homology [Licata 2011] and HOMFLY-PT polynomial estimates [Cornwell 2012a; 2012b] for *tb* have been established while an appropriate notion of normal ruling has yet to be formulated. For this reason, establishing a more direct link between Bennequin type inequalities and augmentations could prove interesting.

(ii) For Legendrians in $\mathbb{R}^3$, there is another interesting condition connected with the equivalent statements in Theorem 1.1. Specifically, the existence of a 0-graded normal ruling is equivalent to the existence of a linear at infinity generating family for L; see [Chekanov and Pushkar 2005; Fuchs and Rutherford 2011]. This statement remains true in J^1S^1. However, it is interesting to ask if links with 0-graded generalized normal rulings always admit reasonable (say, linear or quadratic at infinity) generating families. To allow for fixed point strands, it seems necessary to consider generating families $F : E \to \mathbb{R}$ defined on bundles $E \to S^1$ whose fiber has nontrivial homology. As an example, the basic front A_m defined in Section 2A may be generated by a function on an m-fold cover of S^1.

Organization. The article is arranged as follows: In Section 2, we provide the necessary background about normal rulings and the Kauffman and HOMFLY-PT invariants and also introduce generalized normal rulings.

Section 3 runs parallel to the results on the HOMFLY-PT skein module and 2-graded rulings from [Rutherford 2011]. We show how to recover the 1-graded ruling polynomial from an appropriate specialization of the Kauffman skein module. A natural basis for the Kauffman skein module is indexed by partitions, and for this basis we provide an explicit formula for the specialization.

In Section 4 we prove part (1) of Theorem 1.2 by combining the results of Section 3 (and of [Rutherford 2011] for the HOMFLY-PT case) with a linear independence argument.

Section 5 deals with part (2) of Theorem 1.2. For the forward implication we base all of our arguments on linear algebraic results from [Barannikov 1994], from which the reason behind the normality conditions, with or without fixed points, becomes clear.

2. Background on Legendrian solid torus links

We assume familiarity with basic concepts about Legendrian knots such as front projections, Legendrian Reidemeister moves, Thurston–Bennequin number, and rotation number, at least for knots in $\mathbb{R}^3$. See, for instance, [Geiges 2008], and also note that [Rutherford 2011] contains an alternate discussion of the case of Legendrian knots in J^1S^1.

We view the 1-jet space of the circle, J^1S^1, as $S^1 \times \mathbb{R}^2$ equipped with the contact structure $\xi = \ker(dz - y\,dx)$, where x is a circle-valued coordinate. We occasionally refer to a (Legendrian) link $L \subset J^1S^1$ as a (*Legendrian*) *solid torus link*. The front projection of a Legendrian solid torus link consists of some number of closed curves in the xz-annulus which we view as $[0, 1] \times \mathbb{R}$ with the identification $(0, z) \sim (1, z)$. Generically, front projections are immersed and embedded except at semicubical cusps and transverse double points, and two such projections represent Legendrian isotopic links if and only if they are related by a sequence of Legendrian Reidemeister moves.

We make the convention of extending the Thurston–Bennequin number to homologically nontrivial links by using the front projection formula

$$tb(L) = w(L) - c(L),$$

where $w(L)$ denotes the writhe of L (a signed sum of crossings) and $c(L)$ is half the number of cusps of L.

Similarly, for a Legendrian knot $L \subset J^1S^1$ we define the rotation number as

$$r(L) = \tfrac{1}{2}(d(L) - u(L)),$$

where $d(L)$ denotes the number of downward oriented cusps and $u(L)$ the number of upward oriented cusps.

2A. ***Products of basic fronts.*** Given two annular front diagrams, K and L, we define the product, $K \cdot L$, by stacking K above L. In contrast to the case of smooth knot diagrams, this product is noncommutative as the Legendrian isotopy types of $K \cdot L$ and $L \cdot K$ will not agree in general; see [Traynor 1997; Rutherford 2011].

In this article the *basic fronts*, A_m, will play an important role. Given $m \in \mathbb{Z}_{>0}$, A_m is the front diagram that winds m times around the annulus with $m-1$ crossings and no cusps; see Figure 1. When it is necessary to pay attention to orientations, for $m > 0$, we will use A_m (respectively A_{-m}) for the basic front oriented in the direction of the positive (respectively negative) x-axis.

If $\lambda = (\lambda_1, \dots, \lambda_\ell)$ is an ℓ-tuple of positive integers we write $A_\lambda = A_{\lambda_1} A_{\lambda_2} \cdots A_{\lambda_\ell}$ for the product of basic fronts and $A_{-\lambda}$ for the product with all orientations reversed.

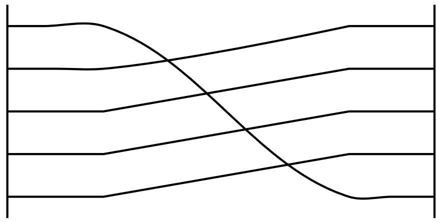
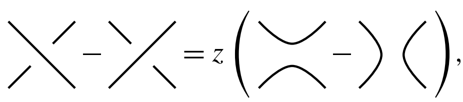
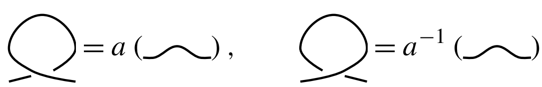


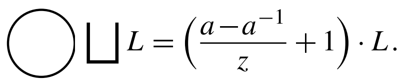
Figure 1. The basic front A_5.

2B. ***Kauffman polynomial in J^1S^1.*** We now describe a generalization from [Turaev 1988] of the Kauffman polynomial to smooth links (not necessarily Legendrian) in the solid torus. In practice, this invariant is computed by reducing a link diagram to products of basic fronts via skein relations. Whenever appropriate, we will view a front diagram of a Legendrian link as a smooth link diagram by placing the strand with lesser slope on top at crossings and smoothing cusps.

Let $\mathcal{D}$ denote the set of regular isotopy classes of unoriented link diagrams in the annulus. That is, we consider link diagrams up to the equivalence generated by Type II and Type III Reidemeister moves. Using the coefficient ring $R = \mathbb{Z}[a^{\pm 1}, z^{\pm 1}]$ we define the Kauffman skein module $\mathcal{F}$ as the quotient of the free R-module $R\mathcal{D}$ by the submodule generated by the Kauffman skein relations

(2-1) 

(2-2)  and

(2-3) 

The product of diagrams gives a well defined product on $\mathcal{F}$ which is commutative as we now consider diagrams of smooth links rather than front diagrams of Legendrian links. Turaev [1988] showed that $\mathcal{F}$ is a polynomial R-algebra in the basic fronts. Thus, to a link diagram L we may associate a polynomial $D_L(a, z; A_1, A_2, \ldots)$ according to

$$\mathcal{F} \cong R[A_1, A_2, \ldots], \quad [L] \leftrightarrow D_L.$$

The *Kauffman polynomial* of an oriented link $L \subset J^1S^1$ is then defined by the normalization $F_L = a^{-w(L)} D_L$, where $w(L)$ denotes the writhe of L.

Chmutov and Goryunov [1997] proved that for any Legendrian link $L \subset J^1S^1$,

$$tb(L) \leq -\deg_a F_L. \tag{2-4}$$

While Chmutov and Goryunov [1997] use a different projection annulus for computing F_L, a proof of (2-4) matching our conventions for F_L may be given precisely as in the case of the HOMFLY-PT polynomial addressed in Section 6.2 of [Rutherford 2011].

Remark 2.1. (i) Recall that a possibly empty sequence $\lambda = (\lambda_1, \dots, \lambda_\ell)$ of positive integers is called a *partition* if $\lambda_1 \geq \dots \geq \lambda_\ell$. The integers λ_i are called the *parts* of λ and we sometimes use the notation $\lambda = 1^{j_1} 2^{j_2} \dots n^{j_n}$ to indicate that λ is the partition with j_r parts equal to r, $r = 1, \dots, n$. We note that the collection of products A_λ with λ a partition forms an R-module basis for $\mathcal{F}$.

(ii) The HOMFLY-PT skein module is defined in a similar manner using oriented link diagrams and an appropriate modification of the skein relations (2-1)–(2-3) (see, for instance [Rutherford 2011]). The result is a polynomial algebra generated by the oriented basic fronts [Turaev 1988]. For a given oriented link $L \subset J^1S^1$ we denote the corresponding HOMFLY-PT polynomial as

$$P_L \in R[A_{\pm 1}, A_{\pm 2}, \dots].$$

2C. *Normal rulings in J^1S^1.* Let $L \subset S^1 \times \mathbb{R}$ be the front projection of a Legendrian link in the solid torus satisfying the additional assumption that all crossings and cusps have distinct x-coordinates none of which equals 0. A normal ruling can be viewed locally as a decomposition of L into pairs of paths. We make some notational preparation before giving the formal definition.

Denote by $\Sigma \subset S^1$ those x-coordinates which coincide with a crossing or cusp of L. We can write

$$S^1 \setminus \Sigma = \bigsqcup_{m=1}^{M} I_m$$

with each I_m an open interval (or all of S^1 if $\Sigma = \varnothing$). Making the convention that $I_0 = I_M$, we assume that the I_m are ordered so that I_{m-1} appears immediately to the left of I_m and I_M contains $x = 0$. On subsets of the form $I_m \times \mathbb{R}$ the front projection L consists of some number of nonintersecting components which project homeomorphically onto I_m. We refer to these components as the *strands* of L above I_m, and we number them from *top to bottom* as $1, \dots, N(m)$. Finally, for each $m = 1, \dots, M$ we choose a point $x_m \in I_m$.

Definition 2.2. A *normal ruling* of the front diagram L is a sequence of involutions $\rho = (\rho_1, \dots, \rho_M)$,

$$\rho_m : \{1, \dots, N(m)\} \to \{1, \dots, N(m)\}, \quad (\rho_m)^2 = id,$$

satisfying the following restrictions:

(1) Each ρ_m is fixed-point-free.

(2) If the strands above I_m labeled k and $k+1$ meet at a left cusp in the interval (x_{m-1}, x_m), then $\rho_m(k) = k+1$ and when $n \notin \{k, k+1\}$,

$$\rho_m(n) = \begin{cases} \rho_{m-1}(n) & \text{if } n < k, \\ \rho_{m-1}(n-2) & \text{if } n > k+1. \end{cases}$$

(3) A condition symmetric to (2) at right cusps.

(4) If strands above I_m labeled k and $k+1$ meet at a crossing on the interval (x_{m-1}, x_m), then $\rho_{m-1}(k) \neq k+1$ and either

(a) $\rho_m = (k\ k{+}1) \circ \rho_{m-1} \circ (k\ k{+}1)$, where $(k\ k{+}1)$ denotes the transposition, or

(b) $\rho_m = \rho_{m-1}$.

In the second case we refer to the crossing as a *switch* of ρ. Finally, we have a requirement at switches that is known as the *normality condition*.

(5) If there is a switch on the interval (x_{m-1}, x_m) then one of the following three orderings holds:

$$\rho_m(k+1) < \rho_m(k) < k < k+1,$$
$$\rho_m(k) < k < k+1 < \rho_m(k+1),$$
$$k < k+1 < \rho_m(k+1) < \rho_m(k).$$

Remark 2.3. This definition is a slight variation on those found elsewhere in the literature. Letting $\pi : S^1 \times \mathbb{R} \to S^1$ denote the projection, Chekanov and Pushkar defined a normal ruling as a continuous, fixed-point-free involution of $L \setminus \pi^{-1}(\Sigma)$ that preserves the x-coordinate and is subject to some requirements for continuous extension near crossings or cusps as well as a normality condition at switches. Such an involution is recovered from our definition by viewing the set $\{1, 2, \ldots, N(m)\}$ that ρ_m permutes as the set of strands above I_m.

From this perspective, the fixed-point-free condition causes the ρ_m to divide the strands above I_m into pairs, and in our figures we will present normal rulings by indicating this pairing. Beginning at $x = 0$ and working to the right, one may cover the front diagram with pairs of continuous paths with monotonically increasing x-coordinates, so that a given pair of paths corresponds to strands paired by the involutions. If a path proceeds all the way around the annulus, then it will not necessarily end up where it started. However, the division of the front diagram into pairs of points at $x = 0$ and $x = 1$ should match up.

Paired paths are only allowed to meet at common cusp endpoints. In particular, at any crossing the two paths of the ruling that meet should belong to different pairs and, for values of x near the crossing, each will have a "companion path" located somewhere above or below the crossing. The two paths can either follow the link diagram and cross each other (this corresponds to (4) (a) above) or they may switch

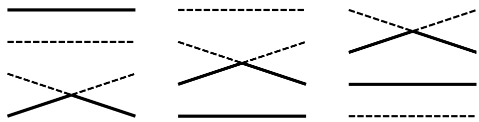


Figure 2. The normality condition.

strands by each turning a corner at the crossing. The normality condition provides a restriction on the location of the companion paths near a switch; out of six possible configurations for the switching strands and their companion strands only three are allowed. See Figure 2 for the normality condition and the right half of Figure 4 for an example of a normal ruling.

2D. ***Maslov potentials and graded normal rulings.*** Further grading restrictions may be placed on a normal ruling after the introduction of a Maslov potential for L. Let p be a divisor of $2r(L_i)$ for each component L_i of a Legendrian link L. A $\mathbb{Z}/p$-valued *Maslov potential* μ for L is a function from L to $\mathbb{Z}/p$ that is constant except at cusp points, where it increases by 1 when moving from the lower strand to the upper strand. Note that a chosen orientation provides L with a $\mathbb{Z}/2$-valued Maslov potential by following the convention that strands oriented to the right (respectively left) are assigned the value 0 (respectively 1) mod 2.

We say that a normal ruling ρ is *p-graded* with respect to a $\mathbb{Z}/p$-valued Maslov potential μ if whenever two strands S_1 and S_2 of L are paired by one of the ρ_m with S_1 above S_2 we have $\mu(S_1) = \mu(S_2) + 1$.

2E. ***Ruling polynomials.*** Suppose μ is a $\mathbb{Z}/p$-valued Maslov potential for a Legendrian link L. The p-graded ruling polynomial of L with respect to μ is

$$R^p_{(L,\mu)}(z) = \sum_\rho z^{j(\rho)},$$

where the sum is over all normal rulings of L that are p-graded with respect to μ and

$$j(\rho) = \#\,\text{switches} - \#\,\text{right cusps}.$$

The ruling polynomial does not depend on the choice of Maslov potential when $p = 1$; $p = 2$ and L is oriented; or L is connected. In any of these cases we denote the ruling polynomial simply as R^p_L. The ruling polynomials are Legendrian isotopy invariants [Chekanov and Pushkar 2005].

2F. ***Generalized normal rulings.*** In the following definition we relax the requirements from Definition 2.2 in a manner appropriate for Theorem 1.2 to hold.

Definition 2.4. A *generalized normal ruling* consists of a sequence of involutions $\rho = (\rho_1, \dots, \rho_M)$ as in Definition 2.2 subject to the following modifications:

(1) We remove the requirement that the ρ_m be fixed-point-free.

(2) If a crossing occurs in the interval (x_{m-1}, x_m) between the k and $k+1$ strands above I_{m-1} with exactly one of these two strands a fixed point of ρ_m, then we decide if the crossing is a switch precisely as in (4) of Definition 2.2. If the crossing is indeed a switch then we require the additional normality condition that either

$$\rho_m(k) = k < k+1 < \rho_m(k+1) \quad \text{or} \quad \rho_m(k) < k < k+1 = \rho_m(k+1).$$

(See Figure 3.)

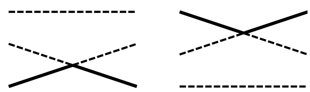


Figure 3. The normality condition for generalized rulings: The strand pictured in bold is a fixed point of ρ_m.

Remark 2.5. (i) If a crossing involving the k and $k+1$ strands occurs on (x_{m-1}, x_m) with both of the crossing strands fixed by the ruling, that is, $\rho_{m-1}(k) = k$ and $\rho_{m-1}(k+1) = k+1$, then $\rho_{m-1} = (k\,k+1) \circ \rho_{m-1} \circ (k\,k+1)$. Consequently, we will not consider such crossings to be switches.

(ii) In the presence of an appropriate Maslov potential, we can consider p-graded generalized normal rulings precisely as in Section 2D.

(iii) The number of generalized normal rulings of a Legendrian link is not invariant under Legendrian isotopy. However, in view of Lemma 2.6 below, the polynomials $R^p_{L\cdot A_\lambda}$ serve as some form of substitute for a "generalized ruling polynomial".

For establishing (1) of Theorem 1.2 we will use the following equivalent characterization of front diagrams that admit generalized rulings.

Lemma 2.6. *A front diagram L has a 1-graded (respectively 2-graded) generalized normal ruling if and only if there exists partitions λ and μ so that $R^1_{L\cdot A_\lambda}(z) \neq 0$ (respectively $R^2_{L\cdot A_\lambda A_{-\mu}}(z) \neq 0$).*

Proof. For simplicity, we treat the 1-graded case first. If $R^1_{L\cdot A_\lambda}(z) \neq 0$, then the diagram $L \cdot A_\lambda$ has a normal ruling, ρ. This produces a generalized normal ruling of L by restricting ρ to L and treating any strands of L which are paired with A_λ as fixed point strands. The normality condition from Definition 2.4 follows from that of Definition 2.2.

Now suppose that L has a generalized normal ruling. If one of the ρ_m has a fixed point strand, then we can continuously follow the fixed point strand around the diagram turning corners only at switches. The result is a portion of the front diagram, C_i, without cusps that we suppose winds λ_i times around the annulus.

There may be several fixed point components of this type. We may assume the λ_i are ordered so that they form a partition, λ. The product $L \cdot A_\lambda$ has a normal ruling where each C_i is paired with the component A_{λ_i} of λ. Such a ruling is completely determined once we specify the pairing between C_i and A_{λ_i} at a single point of C_i. Now, the normality condition of Definition 2.2 follows from that of Definition 2.4, and the ordering of the factors of A_λ is not important here since we do not have switches between any of the C_i; see Remark 2.5 and Figure 4.

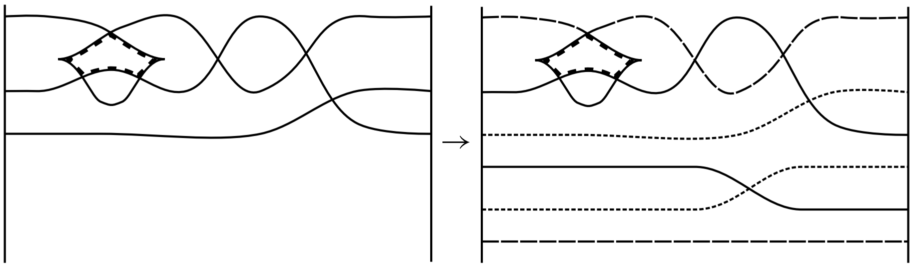


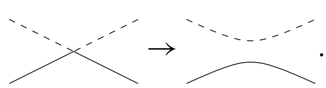
Figure 4. A generalized ruling with three fixed point strands producing a normal ruling of $L \cdot A_\lambda$ with $\lambda = (2, 1)$.

For the 2-graded case, observe that in a 2-graded ruling the orientation of strands meeting at a switch must agree. Therefore, the C_i each have a consistent orientation, and we choose an orientation on the component A_{λ_i} accordingly. □

3. Kauffman polynomial and computation of 1-graded ruling polynomials

An analysis of how to compute 2-graded ruling polynomials of Legendrian solid torus links from the HOMFLY-PT polynomial is done in [Rutherford 2011]. In this section, we will perform a similar analysis of the 1-graded case. We will derive formulas for the 1-graded ruling polynomial of A_λ, and then relate the general case to a coefficient of an appropriate specialization of the Kauffman polynomial.

3A. ***Normal rulings of the product A_λ.*** Given a front diagram L with normal ruling ρ we define the *decomposition* of L with respect to ρ as the Legendrian link L_ρ obtained by resolving the switches of L into parallel horizontal strands as



The involutions of the strands of L piece together to provide an involution, which we also denote as ρ, now defined on all of L_ρ. The involution ρ is continuous where we now view L_ρ as a subset of J^1S^1 rather than just a front diagram, and its only fixed points correspond to the cusps of the front projection of L_ρ. (Compare with Remark 2.3.) The normal ruling of L induces a normal ruling of L_ρ where none of the crossings are switches.

We record some observations about normal rulings of the products A_λ.

Lemma 3.1. *Suppose ρ is a normal ruling of $L = A_\lambda$.*

(1) *The decomposition, L_ρ, is also a product of basic fronts.*

(2) *The involution ρ must take a component of L_ρ isotopic to A_m to another component isotopic to A_m.*

(3) *If components C_1 and C_2 of L_ρ share a common switch of L, with C_1 above C_2 on the z-axis, then the vertical ordering of the four components C_1, C_2, $\rho(C_1)$, and $\rho(C_2)$ must be one of*

$$[\rho(C_2), \dots, \rho(C_1), \dots, C_1, C_2],$$
$$[\rho(C_1), \dots, C_1, C_2, \dots, \rho(C_2)],$$
$$[C_1, C_2, \dots, \rho(C_2), \dots, \rho(C_1)].$$

(4) *The restriction of ρ to a pair of components of L_ρ, C_1 and $C_2 = \rho(C_1)$, is completely determined by its value at a single point, $w \in C_1$. Moreover, if $C_1 \cong A_m$ then there are precisely m choices for $\rho(w) \in C_2$, and any one of them extends continuously to all of C_1.*

(5) *Two components of L_ρ of the form C_1 and $\rho(C_1)$ cannot correspond to subsets of the same component of L.*

Proof. Item (1) is clear; (2) follows from continuity of ρ; and (3) is a consequence of the normality condition. The first assertion of (4) follows from continuity of ρ. The second follows since $\rho(w)$ and w must have the same x-coordinate and C_2 also consists of m strands. That any such choice of $\rho(w)$ extends to all of C_1 is easily seen.

We prove (5) by contradiction. Suppose C_1 and $\rho(C_1)$ did come from the same component of L, and without loss of generality assume $\rho(C_1)$ is below C_1. They cannot meet at a switch as this would violate the normality condition. Thus, there is some other component C_2 on the other end of the switch below C_1. The only possible position of $\rho(C_2)$ is then between C_2 and $\rho(C_1)$. Then C_2 and $\rho(C_2)$ also came from the same component of L. They cannot meet at a switch, so there is some further component C_3 immediately below C_2, which is paired with a component $\rho(C_3)$ between C_3 and $\rho(C_2)$. We can continue this argument to produce arbitrarily many components of L_ρ between C_1 and $\rho(C_1)$. □

3B. *Computing $R^1_{A_m A_m}$*. The results in the previous section are sufficient to compute the ruling polynomial for the simplest possible product, $A_m A_m$ (the ruling polynomial of a single basic front A_m is 0 by (5) of Lemma 3.1). Although this agrees with $R^2_{A_m A_{-m}}$ which is computed in Lemma 4.1 of [Rutherford 2011], the form of the answer given here is simplified and the proof is quite different.

Lemma 3.2. *The ruling polynomial of* $L = A_m A_m$ *is*

$$\sum_{k=0}^{m-1} \binom{m+k}{2k+1} z^{2k}.$$

Proof. Normal rulings of $A_m A_m$ with $2k$ switches are in bijection with subdivisions of m ordered objects into $k+1$ consecutive parts, with a marked object chosen in each part.

The subdivision corresponds to choosing the location of k switches within the first A_m factor. Specifically, dividing m into parts $(\lambda_1, \dots, \lambda_{k+1})$ corresponds to choosing k switches so that in the decomposition, L_ρ, the first A_m factor becomes $A_{\lambda_1} \dots A_{\lambda_{k+1}}$. In L_ρ, the A_{λ_i} must be paired with $k+1$ components of the same size from the second A_m factor, by parts (2) and (5) of Lemma 3.1. Then, Lemma 3.1(3) determines the order of the components: they must be in the reverse order of the components from the first factor. The total number of switches is $2k$.

The choice of marked object within a part λ_i corresponds to choosing which strand within the A_{λ_i} component is paired with the top strand of $\rho(A_{\lambda_i})$ at $x = 0$. These choices may be arbitrary, and they uniquely determine a ruling by part (4) of Lemma 3.1. See Figure 5.

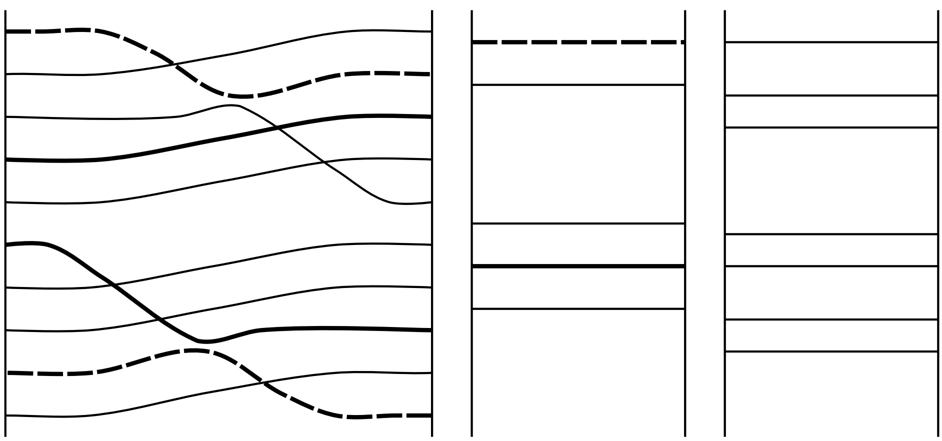


Figure 5. The bijection between rulings of $A_5 A_5$ with 2 switches, divisions of 5 objects into 2 parts with a marked object in each part, and compositions of 7 into 4 positive parts.

To complete the proof, observe that subdivisions of this type are in bijection with compositions of $m + (k+1)$ into $2(k+1)$ positive parts $(a_1, b_1, \dots, a_{k+1}, b_{k+1})$: two consecutive parts of size a_i and b_i correspond to a part $\lambda_i = a_i + b_i - 1$ with the a_i-th object marked in a subdivision of m. The number of ways to decompose $m+k+1$ objects into $2(k+1)$ parts of positive size is well-known to be

$$\binom{(m+k+1)-1}{2(k+1)-1} = \binom{m+k}{2k+1}.$$

This gives us the sum for the ruling polynomial. □

This formula will be used in the next section, so we will write $\langle m \rangle$ for the ruling polynomial $R_{A_m A_m}(z)$, following the convention in [Rutherford 2011].

3C. ***A formula for arbitrary products of basic fronts.*** We will use the formula for $\langle m \rangle$ to calculate the ruling polynomial of A_λ for an arbitrary λ.

Given a normal ruling ρ of $L = A_\lambda$, define the *block* B_{ij} to consist of those components of the decomposition L_ρ which originated in the i-th component of L, and are paired by ρ with components that originated in the j-th component of L. The size of the block, b_{ij}, is the number of points in B_{ij} with some fixed x-coordinate, away from crossings.

Lemma 3.3. *Given a normal ruling of*

$$L = A_\lambda = A_{\lambda_1} A_{\lambda_2} \dots A_{\lambda_n},$$

the blocks in the i-th component of L consist of vertically consecutive components of L_ρ, and are themselves vertically ordered as

$$B_{i,i-1} B_{i,i-2} \dots B_{i,1} B_{i,n} B_{i,n-1} \dots B_{i,i+1},$$

where some blocks may be empty.

Proof. Suppose that when we resolve A_{λ_i} at switches, we get the components $C_1, C_2, \dots, C_k$, in that vertical order. If, for some j, $\rho(C_j)$ is above C_j, then the normality condition demands that $\rho(C_{j-1})$ is between $\rho(C_j)$ and C_{j-1}. Similarly, if $\rho(C_j)$ is below C_j, then $\rho(C_{j+1})$ must be between C_{j+1} and $\rho(C_j)$.

As a result, if $\rho(C_{j_1})$ and $\rho(C_{j_2})$ come from the same component of L, then $\rho(C_j)$ for $j_1 \le j \le j_2$ are between $\rho(C_{j_1})$ and $\rho(C_{j_2})$. This implies each block is made up of some number of consecutive components. And due to the normality condition, the ordering of any two consecutive blocks must be either $B_{i,j+1} B_{i,j}$, with $j > i$, or $B_{i,j-1} B_{i,j}$, with $j < i$ (with the caveat that some of the blocks may be empty, if ρ does not pair two components of L at all). Putting this together yields the block ordering above. □

This means that once we pick the sizes of the blocks $b_{i,1}, \dots, b_{i,n}$, the locations of the blocks are determined. To complete the calculation of the ruling polynomial, observe that the choice of a normal ruling of the blocks B_{ij} and B_{ji}, with sizes $b_{ij} = b_{ji} = m$, is equivalent to the choice of a normal ruling of $A_m A_m$.

Theorem 3.4. *Let $\langle m\rangle$ denote the ruling polynomial of $A_m A_m$, with $\langle 0\rangle$ taken to be z^{-2}. Then the ruling polynomial of $A_\lambda = A_{\lambda_1} A_{\lambda_2} \dots A_{\lambda_n}$ is given by*

$$z^{n(n-2)} \sum_{(b_{ij})\in M_\lambda} \prod_{i<j} \langle b_{ij}\rangle,$$

where M_λ is the set of all symmetric matrices (b_{ij}) with nonnegative integer entries such that the row sums $\sum_{j=1}^{n} b_{ij} = \lambda_i$ and the trace $\operatorname{tr}(b_{ij}) = 0$.

Proof. The choice of a matrix in M_λ is equivalent to the choice of block sizes b_{ij}. By Lemma 3.3, this also fixes the locations of the blocks. A normal ruling of A_λ is then completely determined by its restriction to pairs of blocks B_{ij} and B_{ji}.

If the block size b_{ij} is nonzero, then $\langle b_{ij}\rangle$ describes the possible restrictions of the normal rulings to the union $B_{ij} \cup B_{ji}$. We take the product to combine these normal rulings, but we have to account for the switches between the blocks. If all block sizes are nonzero, then there will be $n-2$ switches in each of the n components of L, giving us a factor of $z^{n(n-2)}$. Any block B_{ij} of size 0 will reduce this number by 1 in component j, but the corresponding block B_{ji} will reduce the number of switches by 1 in component i; this gives a factor of z^{-2} which is accounted for by the convention of $\langle 0\rangle = z^{-2}$. □

Corollary 3.5. *The* 1*-graded ruling polynomial is commutative in front diagram products: that is, the ruling polynomials of*

$$A_{\lambda_1} A_{\lambda_2} \dots A_{\lambda_i} A_{\lambda_{i+1}} \dots A_{\lambda_n} \quad \text{and} \quad A_{\lambda_1} A_{\lambda_2} \dots A_{\lambda_{i+1}} A_{\lambda_i} \dots A_{\lambda_n}$$

are equal.

Proof. There is an easy bijection between the possibilities for the matrix M_λ and the new matrix $M_{\lambda'}$: we simply exchange the i-th and $(i+1)$-th columns and rows; the summands $\prod_{i<j} \langle b_{ij}\rangle$ do not change. □

3D. ***Calculating the ruling polynomial from the Kauffman polynomial.*** In $\mathbb{R}^3$, the 1-graded and 2-graded ruling polynomial of arbitrary Legendrian links may be easily recovered from the Kauffman and HOMFLY-PT polynomials. The second author shows in [Rutherford 2011] that the 1-graded (respectively 2-graded) ruling polynomial of a link L is the coefficient of $a^{-tb(L)}$ in the Kauffman polynomial (respectively HOMFLY-PT polynomial) of L. In the case of Legendrian solid torus links we first need to specialize the extra variables in a nonmultiplicative manner.

Using the notation of Section 2B, we consider the R-module homomorphism $\Psi : \mathcal{F} \cong R[A_1, A_2, \dots] \to R$ determined by $A_\lambda \mapsto R^1_{A_\lambda}(z)$ when λ is a partition. (Compare with Remark 2.1.) Given a link diagram L, we let $\widehat{D}_L(a, z) = \Psi(D_L)$, and $\widehat{F}_L(a, z) = a^{-w(L)} \widehat{D}_L(a, z)$.

Theorem 3.6. *Let $L \subset J^1S^1$ be any Legendrian solid torus link. Then the* 1*-graded ruling polynomial $R^1_L(z)$ is equal to the coefficient of $a^{-tb(L)}$ in $\widehat{F}_L(a, z)$.*

This result is analogous to Theorem 6.3 of [Rutherford 2011], where it is shown that we can recover the 2-graded ruling polynomial from such a specialization of the HOMFLY-PT polynomial. The proof, via induction on a certain measure of complexity of a front diagram, carries through in the 1-graded case as well. The base case consists of all products of basic fronts where the result follows from the crucial Corollary 3.5. Next, it is observed that the ruling polynomial and the coefficient of $a^{-tb(L)}$ in $\widehat{F}_L$ share common skein relations that are Legendrian analogs of equations (2-1)–(2-3) (see [Rutherford 2006; Rutherford 2011]). Then, just as in [Rutherford 2011], the inductive step is completed by an algorithm which uses these skein relations to evaluate the invariants in terms of front diagrams of lesser complexity.

Example. Consider the Legendrian knots L_1 and $L_2 = L_1 \cdot A_2A_1$ pictured in Figure 4, and suppose orientations are chosen so that all strands are oriented to the right when they pass through the vertical line $x = 0$. The Kauffman polynomials are given by

$$\begin{aligned} F_{L_1} = A_1 \times \big[a^{-1}(-z-z^3)+a^{-2}z^4+a^{-3}(z+2z^3)+a^{-4}z^2\big] \\ +A_3 \times \big[a^{-1}(z+z^3)+a^{-2}(-z^2-z^4)+a^{-3}(-z-z^3)\big] \\ +A_2A_1 \times \big[a^{-1}(1+z^2)-a^{-2}z^3-a^{-3}z^2\big], \end{aligned}$$

and $F_{L_2} = a^{-1}A_2A_1F_{L_1}$. We have $tb(L_1) = 1$ and $tb(L_2) = 2$, so in both cases the estimate (2-4) is sharp.

Theorem 3.4 gives $R^1_{A_{(2,1,1)}}(z) = z$, $R^1_{A_{(3,2,1)}}(z) = 2z + z^3$, and $R^1_{A_{(2,2,1,1)}}(z) = 2 + 3z^2$. This allows us to compute

$$\widehat{F}_{L_2} = a^{-2}(2+6z^2+5z^4+z^6)+a^{-3}(-4z^3-5z^5-z^7)+a^{-4}(-3z^2-4z^4-z^6)+a^{-5}z^3,$$

and Theorem 3.6 gives $R^1_{L_2}(z) = 2+6z^2+5z^4+z^6$, which can be verified directly.

4. Generalized normal rulings and the Thurston–Bennequin estimates

In this section we establish the equivalence (1) of Theorem 1.2 which follows from Lemma 2.6 together with the following:

Theorem 4.1. *Let L be a Legendrian link in the solid torus. Then the equality*

$$tb(L) = -\deg_a F_L$$

holds if and only if there exists a partition λ so that $L \cdot A_\lambda$ has a normal ruling.

Proof of Theorem 4.1. One direction is straightforward. Suppose that, for some λ, $L' = L \cdot A_\lambda$ has a normal ruling. Then the ruling polynomial of L' is nontrivial, so the coefficient of $a^{-tb(L)}$ is nonzero. Therefore $tb(L') \geq -\deg_a F_{L'}$ which, combined with the inequality (2-4), gives us an equality $tb(L') = -\deg_a F_{L'}$. However, $tb(L') = tb(L \cdot A_\lambda) = tb(L) + w(A_\lambda)$, since A_λ has no cusps. In addition, $D_{L'} = A_\lambda \cdot D_L$, so $F_{L'} = a^{-w(A_\lambda)} A_\lambda \cdot F_L$, and we compute

$$-\deg_a F_L = -w(A_\lambda) - \deg_a(F_{L'}) = -w(A_\lambda) + tb(L') = tb(L).$$

Now suppose $tb(L) = -\deg_a F_L$. We will find a λ such that $L \cdot A_\lambda$ has a normal ruling.

Let $\sum_\mu p_\mu(z) A_\mu$ be the coefficient of $a^{-tb(L)}$ in F_L, where the $p_\mu(z)$ are polynomials in z and z^{-1}. This coefficient is nonzero, or else the degree equality would not hold, so $p_\mu(z) \neq 0$ for at least one μ. Let k be the smallest integer such that at least one p_μ has a nonzero coefficient of z^k.

By Theorem 3.6, the ruling polynomial of $L \cdot A_\lambda$ is

$$\sum_\mu p_\mu(z) R_{A_\mu A_\lambda}(z).$$

We will prove that for some λ, this polynomial is nonzero (and therefore a normal ruling exists) by looking at the z^k coefficient of this polynomial. Since $R_{A_\mu A_\lambda}(z)$ is a polynomial in z with no terms of z^{-1} or lower degree, the only way to get a z^k coefficient is from the product of $p_\mu(z)[z^k]$ and $R_{A_\mu A_\lambda}(z)[z^0]$ for some μ (here, $f(z)[z^i]$ denotes the coefficient of z^i in $f(z)$). Denote $p_\mu(z)[z^k]$ by a_μ, and $R_{A_\mu}(z)[z^0]$ (which is the number of switchless rulings of A_μ) by $C(\mu)$.

The quantity $C(\mu)$ is easy to calculate. Without switches, each component of size k must simply be paired with another component of size k in one of k ways. In particular, this is only possible if there is an even number of each component size. Define the double factorial by

$$(2k-1)!! = (2k-1)(2k-3)(\cdots)(3)(1) = (2k)!/(2^k k!).$$

It counts the number of ways to divide $2k$ objects into pairs. It is clear that

$$C(\mu) = \begin{cases} \prod_{k=1}^n k^{a_k}(2a_k - 1)!! & \text{if } \mu = 1^{2a_1} 2^{2a_2} \dots n^{2a_n}, \\ 0 & \text{else.} \end{cases}$$

We wish to prove that for some λ, $\sum_\mu a_\mu C(\mu \cdot \lambda) \neq 0$. Here, if

$$\mu = 1^{a_1} 2^{a_2} \dots n^{a_n} \quad \text{and} \quad \lambda = 1^{b_1} 2^{b_2} \dots n^{b_n},$$

we will denote by $\mu \cdot \lambda$ the partition

$$1^{a_1+b_1} 2^{a_2+b_2} \dots n^{a_n+b_n}.$$

Let M be the collection of all partitions such that:

(1) The parts of the partition are all no larger than n, for some n.

(2) Parts of each size occur between 0 and $2m-1$ times, for some m.

We choose the parameters m and n such that we include all partitions μ with $a_\mu \neq 0$.

Let V be a n^{2m}-dimensional real vector space with basis vectors e_λ for $\lambda \in M$. For each $\mu \in M$, consider the vectors

$$v_\mu = \sum_{\lambda\in M} C(\mu\cdot\lambda)e_\lambda$$

in V. We will show that these vectors also form a basis of V, and are therefore linearly independent. From there, observe that

$$\sum_{\lambda\in M}\Big(\sum_{\mu\in M} a_\mu C(\mu\cdot\lambda)\Big)e_\lambda = \sum_{\mu\in M} a_\mu\Big(\sum_{\lambda\in M} C(\mu\cdot\lambda)e_\lambda\Big) = \sum_{\mu\in M} a_\mu v_\mu.$$

If the coefficients a_μ on the right are not all 0, then because the v_μ are linearly independent the resulting sum is a nonzero vector of V. Therefore the coefficients in terms of e_λ are not all 0 as well — that is, for some λ, $\sum_\mu a_\mu C(\mu\cdot\lambda) \neq 0$. So once we have the result of linear independence, we are done.

From the formula for $C(\mu)$, it's easy to calculate that $C(\mu\cdot\lambda)$ can be written as a product of $C(k^{a_k}\cdot k^{b_k})$, over all k, where a_k and b_k are the number of parts of size k in λ and μ respectively. Suppose we write V as the tensor product $\bigotimes_{i=1}^n \mathbb{R}^{2m}$, identifying the basis vector $e_{j_1}\otimes e_{j_2}\otimes\cdots\otimes e_{j_n}$ on the left with the basis vector e_λ on the right, where $\lambda = 1^{j_1}2^{j_2}\ldots n^{j_n}$. Here we use a slightly nonstandard basis of $\mathbb{R}^{2m}$: it is 0-indexed and consists of $\{e_0, e_1, \ldots, e_{2m-1}\}$, for ease of notation.

Then, if $\mu = 1^{a_1}2^{a_2}\ldots n^{a_n}$,

$$\begin{aligned} v_\mu = \sum_{\lambda\in M} C(\mu\cdot\lambda)\, e_\lambda &= \sum_{1^{b_1}\cdots n^{b_n}\in M}\Big(\prod_{i=1}^n C(i^{a_i}\cdot i^{b_i})\Big)\Big(\bigotimes_{i=1}^n e_{b_i}\Big) \\ &= \sum_{1^{b_1}\cdots n^{b_n}\in M}\Big(\bigotimes_{i=1}^n C(i^{a_i}\cdot i^{b_i})\, e_{b_i}\Big) = \bigotimes_{i=1}^n\Big(\sum_{j=0}^{2m-1} C(i^{a_i}\cdot i^j)\, e_j\Big). \end{aligned}$$

Therefore, rather than prove that the vectors v_μ are a basis of V, it suffices to prove that the vectors $u_k = \sum_{j=0}^{2m-1} C(i^k\cdot i^j)e_j$, as k goes from 0 to $2m-1$, are a basis of $\mathbb{R}^{2m}$. There are three simplifying observations to be made:

(1) $C(i^k\cdot i^j) = 0$ if $k \not\equiv j \pmod 2$. Therefore u_k is a linear combination only of the odd-indexed e_j if k is odd, and only of the even-indexed e_j if k is even. Furthermore, $C(i^k\cdot i^j) = C(i^{k-1}\cdot i^{j+1})$, so u_{2k} and u_{2k-1} have the same coefficients, just shifted over by one index. As a result, we will only show the independence of the vectors $u_0, u_2, \ldots, u_{2m-2}$ — the result for $u_1, u_3, \ldots, u_{2m-1}$ is similar.

(2) By the first observation, we have

$$u_{2k} = \sum_{j=0}^{m-1} C(i^{2k} \cdot i^{2j}) e_{2j} = i^k \sum_{j=0}^{m-1} C(1^{2k} \cdot 1^{2j})(i^j e_{2j}).$$

This corresponds to starting in the case $i = 1$, then scaling both the u_{2k} and the e_{2j} by powers of i – a scaling which doesn't change the question of linear independence one way or the other. Therefore it suffices to consider the case $i = 1$.

(3) Finally, we can scale each u_{2k} by $C(1^{2k})$ (which, too, doesn't affect linear independence). Now we want to look at

$$u'_{2k} = \sum_{j=0}^{m-1} C(1^{2k} \cdot 1^{2j})/C(1^{2k}) e_{2j} = \sum_{j=0}^{m-1} \Big(\prod_{\ell=1}^{j} (2k + 2\ell - 1) \Big) e_{2j}.$$

If we put the coefficients of u'_{2k} as columns of a matrix, (that is, j indexes the rows and k indexes the columns), we get

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 3 & \dots & 2m-1 \\ 1 \cdot 3 & 3 \cdot 5 & \dots & (2m-1)(2m+1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 \cdot 3 \cdots (2m-1) & 3 \cdot 5 \cdots (2m+1) & \dots & (2(m-1)+1)(\cdots)(4(m-1)-1) \end{pmatrix}$$

Here, the entries in the j-th row are given by $f_j(k) = \prod_{\ell=1}^{j} (2k + 2\ell - 1)$, which is a degree j polynomial function. In particular, $f_j(k)$ can be written as $(2k)^j$ plus lower-order terms; these lower-order terms are necessarily a linear combination of $f_1(k), \dots, f_{j-1}(k)$. Therefore, we can use row operations to eliminate the lower-order terms, so that the resulting matrix is

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & m \\ 1 & 4 & \dots & m^2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2^{m-1} & \dots & m^{m-1} \end{pmatrix}$$

This is a Vandermonde matrix whose determinant is $\prod_{j \neq k} (j - k) \neq 0$. Therefore the vectors u'_{2k} (and u_{2k}) form a basis of $\mathbb{R}^{2m}$, which completes the proof. □

4A. ***The 2-graded case and the HOMFLY-PT estimate.*** A similar approach applies in the case of the HOMFLY-PT polynomial, P_L. The proof of the reverse implication is identical. For the forward implication, we suppose $tb(L) = -\deg_a P_L$ and consider the coefficient of the lowest power z^k that appears in the $a^{-tb(L)}$ term of P_L,

$$\sum_{\alpha,\beta} b_{(\alpha,\beta)} A_\alpha A_{-\beta}.$$

Fix parameters m and n so that the set

$$M = \{(\mu, \nu) \mid \mu = 1^{a_1} \cdots n^{a_n},\ \nu = 1^{b_1} \cdots n^{b_n},\ 1 \le a_i, b_i \le m\}$$

contains all (α, β) such that $b_{(\alpha,\beta)} \neq 0$.

Using Theorem 6.3 in [Rutherford 2011], for any $(\mu, \nu) \in M$ the coefficient of z^k in the 2-graded ruling polynomial of $L \cdot A_\mu A_{-\nu}$ is given by

$$\sum_{\alpha,\beta} b_{(\alpha,\beta)} R^2_{A_{\alpha\cdot\mu} A_{-\beta\cdot\nu}}(0).$$

It suffices to show that the coefficient matrix

$$A = \big(R^2_{A_{\alpha\cdot\mu} A_{-\beta\cdot\nu}}(0)\big)_{(\alpha,\beta),(\mu,\nu)\in M}$$

is nonsingular. Writing $\alpha = 1^{a_1} \dots n^{a_n}$, $\beta = 1^{b_1} \dots n^{b_n}$, $\mu = 1^{c_1} \dots n^{c_n}$, and $\nu = 1^{d_1} \dots n^{d_n}$, one has

$$R^2_{A_{\alpha\cdot\mu} A_{-\beta\cdot\nu}}(0) = \prod_{k=1}^{n} \delta_{a_k+c_k, b_k+d_k} k^{a_k+c_k} (a_k + c_k)!.$$

Thus, A is a tensor product (Kronecker product) of matrices

$$A_k = \big(\delta_{a+c,b+d} k^{a+c} (a+c)!\big)_{(a,b),(c,d)}.$$

Due to the Kronecker delta, each A_k is a direct sum (block matrix) of matrices B_l, $l \in \mathbb{Z} \cap [-n, n]$ obtained from keeping rows and columns satisfying $a-b=d-c=l$.

The proof is completed by showing that each B_l is nonsingular. We treat the case $l \ge 0$ as $l < 0$ is similar. Then, $l \le a, d \le n$ and $B_l = \big(k^{a+d-l}(a+d-l)!\big)$. Dividing rows by k^{a-l} and columns by $k^d \cdot d!$ leaves

$$\frac{(a+d-l)!}{d!} = (f_a(d)),$$

where $f_a(x) = \prod_{j=1}^{a-l}(j + x)$ is a polynomial of degree $a - l$. Elementary row operations reduce this to a nonsingular Vandermonde matrix. □

5. Augmentations and generalized normal rulings

In this final section we complete the proof of Theorem 1.2 by establishing that:

For any Legendrian link $L \subset J^1S^1$ with $\mathbb{Z}/p$-graded Maslov potential, μ, the following are equivalent:

(A) The Chekanov–Eliashberg algebra $(\mathcal{A}(L), d)$ admits a p-graded augmentation.

(B) The front projection of L has a p-graded generalized normal ruling.

We begin by briefly recalling the aspects of the Chekanov–Eliashberg DGA that are important for the proof. The reader is referred to [Ng and Traynor 2004] for the original, more detailed treatment of this DGA in the J^1S^1 setting.

Given a Legendrian knot or link $L \subset J^1S^1$, the *Lagrangian projection* $\pi_{xy}(L)$ of L to the xy-annulus is an immersed curve. The Chekanov–Eliashberg DGA $(\mathcal{A}(L), d)$ is a graded algebra $\mathcal{A}(L)$ with a degree -1 differential d, defined via a generic Lagrangian projection of L.

After a small Legendrian isotopy, we may assume $\pi_{xy}(L)$ to have only finitely many transverse double points which we label as $q_1, \dots, q_n$. Then the algebra $\mathcal{A}(L)$ is the free associative $\mathbb{Z}/2$-algebra with unit generated by the double points $q_1, \dots, q_n$. The set of monic noncommutative monomials in the q_i forms a linear basis for $\mathcal{A}(L)$. If L is connected, then $\mathcal{A}(L)$ has a $\mathbb{Z}/2r(L)$ grading. In general, the grading depends on a choice of Maslov potential for L. The differential d is defined by counting certain immersed discs in the xy-annulus with boundary mapped to the Lagrangian projection of L.

Definition 5.1. An *augmentation* of $(\mathcal{A}(L), d)$ is an algebra homomorphism

$$\varepsilon : \mathcal{A}(L) \to \mathbb{Z}/2$$

satisfying

(i) $\varepsilon(1) = 1$, and

(ii) $\varepsilon \circ d = 0$.

In addition, ε is *p-graded* if $\varepsilon(q_i) \neq 0$ implies $|q_i| = 0 \bmod p$.

The existence of an augmentation of $(\mathcal{A}(L), d)$ is a property that is invariant under Legendrian isotopy. This follows from the fact that the "stable tame isomorphism type" (see [Chekanov 2002; Ng and Traynor 2004]) of $(\mathcal{A}(L), d)$ is unchanged by a Legendrian isotopy. Therefore, in establishing the equivalence of **(A)** and **(B)** we may work with the Chekanov–Eliashberg algebra of a Legendrian isotopic link L'. The links L' which we will consider have a standard form so that $(\mathcal{A}(L'), d)$ may be described in a formulaic manner from the front projection of L' (and this front projection is combinatorially the same as that of L). For this reason we do not present the differential or the grading of the Chekanov–Eliashberg DGA in full generality here.

5A. ***The DGA of a resolved front diagram with splashes.*** Given a Legendrian $L \subset J^1S^1$ we begin by modifying the front diagram of L via (a slight variation of) the resolution procedure of Ng and Traynor [2004]. Beginning near $x = 0$ and working from left to right, we alter the front projection of L by an isotopy in the xz-annulus as follows. We arrange so that, except for intervals near $x = 1$ or immediately prior to a crossing or right cusp, the slopes of the strands are constant and strictly decreasing as we move from the top to bottom. Further, we will assume that all strands usually have nonpositive slope. It is no problem to produce these conditions after a left cusp, but with crossings and right cusps the slopes of the two relevant strands will need to be interchanged prior to the crossing or cusp. As the y-coordinate is given by the slope dz/dx, this has the effect of producing double points on the Lagrangian projection corresponding to (but located to the left of) the crossings and right cusps of the front projection of L. Finally, when we near $x = 1$ the strands have become very spread out and moved below their original z values at $x = 0$. Beginning with the top strand and then proceeding successively to the lowest strands, we return each strand back to its initial position via a steep upward step. This creates several new crossings on the Lagrangian projection; see Figure 6.

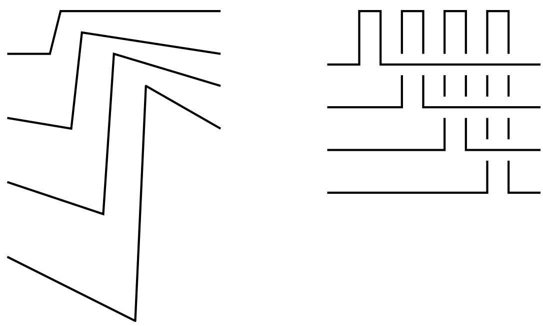


Figure 6. The front projection (left) and Lagrangian projection (right) of L' in an interval immediately to the left of $x = 1$.

Next, we add "splashes". This is a variant of a technique introduced in [Fuchs 2003]; see Remark 5.2. We view the S^1 factor of J^1S^1 as [0, 1] with 0 and 1 identified. In notation similar to that of Section 2, let $0 = x_0 < x_1 < \cdots < x_M = 1$ be a partitioning of the interval [0, 1] such that no x_m coincides with the x-coordinate of a crossing or cusp and each interval (x_{m-1}, x_m) contains exactly one crossing or cusp. For each $m = 1, \ldots, M-1$, we add a miniature version of the steps appearing in the part of the resolution procedure near $x = 1$ into a small interval centered at x_m. That is, beginning at the top strand and then working downward add a brief but steep (smooth) upward step into the diagram. This has a minimal effect on the front projection but alters the Lagrangian projection at each x_m by replacing what had been several parallel lines with a collection of crossings similar to those pictured in the right half of Figure 6. Denote the Legendrian link resulting from the combination of these two procedures as L'.

We now give a complete description of the Chekanov–Eliashberg DGA of L'. For each $1 \leq m \leq M$, let $N(m)$ denote the number of intersection points of L with the plane $x = x_m$. The generators of $\mathcal{A}(L')$ come from two sources. First, we have generators corresponding to the crossings and right cusps of the front projection of L via the resolution procedure. In addition, for each $1 \leq m \leq M$ we have two upper triangular matrices worth of generators, x^m_{ij} and y^m_{ij} with $1 \leq i < j \leq N(m)$. These correspond to the double points created by the splashes and the final step of the resolution procedure.

The grading. If L is equipped with a $\mathbb{Z}/p$-graded Maslov potential, μ, then $\mathcal{A}(L')$ is $\mathbb{Z}/p$-graded. We will describe the degree $|q_i| \in \mathbb{Z}/p$ assigned to the generators of $\mathcal{A}(L')$; degrees then extend additively as $|x \cdot y| = |x| + |y|$.

In the following, $\mu(m, i)$ denotes the value of the Maslov potential on the i-th strand at x_m. (As in Section 2, we label strands from top to bottom.) The generators of $\mathcal{A}(L')$ coming from splashes have degrees

$$|x^m_{ij}| = \mu(m, i) - \mu(m, j) \quad \text{and} \quad |y^m_{ij}| = \mu(m, i) - \mu(m, j) - 1. \tag{5-1}$$

In addition, a crossing b_m between the k and $k+1$ strands occurring in the interval (x_{m-1}, x_m) has $|b_m| = \mu(m, k + 1) - \mu(m, k)$, and all right cusps have degree 1.

The differential. Formulas for the differential d are most efficiently provided by placing the generators x^m_{ij} and y^m_{ij} into strictly upper triangular matrices

$$X_m = (x^m_{ij}) \quad \text{and} \quad Y_m = (y^m_{ij})$$

for each m. (Here, $x^m_{ij} = y^m_{ij} = 0$ if $i \geq j$.) As the x-coordinate is S^1-valued, it is important to make the convention that $X_0 = X_M$ and $Y_0 = Y_M$. Then, applying the differential to each entry, we have the formulas

$$\begin{aligned} dY_m &= (Y_m)^2 \quad \text{and} \\ dX_m &= Y_m(I + X_m) + (I + X_{m-1})\widetilde{Y}_{m-1} \end{aligned} \tag{5-2}$$

with I an identity matrix of the appropriate size. The precise form of $\widetilde{Y}_{m-1}$ depends on the tangle appearing on the interval (x_{m-1}, x_m) and is described presently.

Suppose that (x_{m-1}, x_m) contains a crossing, b_m, between the strands labeled k and $k + 1$. Then

$$db_m = y^{m-1}_{k,k+1} \quad \text{and} \quad \widetilde{Y}_{m-1} = B_{k,k+1}\widehat{Y}_{m-1}B^{-1}_{k,k+1},$$

where $B_{k,k+1}$ and $B^{-1}_{k,k+1}$ agree with the identity matrix except for a 2×2 block along the diagonal in rows k and $k+1$, having the form $\left[\begin{smallmatrix} 0 & 1 \\ 1 & b_m \end{smallmatrix}\right]$ for $B_{k,k+1}$ and $\left[\begin{smallmatrix} b_m & 1 \\ 1 & 0 \end{smallmatrix}\right]$ for $B^{-1}_{k,k+1}$, and $\widehat{Y}_{m-1}$ is the matrix Y_{m-1} with 0 replacing the entry $y^{m-1}_{k,k+1}$.

Next, we suppose (x_{m-1}, x_m) contains a single left cusp between the strands labeled k and $k + 1$ at x_m. Then,

$$\widetilde{Y}_{m-1} = J_k Y_{m-1} J_k^{\mathrm{T}} + E_{k,k+1},$$

where J_k is the $N(m) \times N(m)$ identity matrix with columns k and $k+1$ removed and $E_{k,k+1}$ is a matrix with a single nonzero entry in the $k, k+1$ position.

Finally, we suppose (x_{m-1}, x_m) contains a single right cusp, c_m, between the strands labeled k and $k+1$ at x_{m-1}. Then

$$dc_m = 1 + y^{m-1}_{k,k+1},$$

and the matrix $\widetilde{Y}_{m-1}$ is most easily described entry by entry. Let

$$\tau : \{1, \dots, N(m)\} \to \{1, \dots, N(m-1)\}, \qquad \tau(i) = \begin{cases} i & \text{if } i < k, \\ i+2 & \text{if } i \geq k. \end{cases}$$

The i, j entry of $\widetilde{Y}_{m-1}$ is given by $\tilde{y}^{m-1}_{ij} = y^{m-1}_{\tau(i),\tau(j)} + a_{ij}$, where

$$a_{ij} = y^{m-1}_{i,k+1}\, y^{m-1}_{k,\tau(j)} + y^{m-1}_{i,k}\, c_m\, y^{m-1}_{k,\tau(j)} + y^{m-1}_{i,k+1}\, c_m\, y^{m-1}_{k+1,\tau(j)} + y^{m-1}_{i,k+1} (c_m)^2 y^{m-1}_{k+1,\tau(j)}$$

when $i < k \leq j$ and $a_{ij} = 0$ otherwise.

Remark 5.2. The technique of adding some variation of splashes to simplify the differential first appeared in [Fuchs 2003] and has been used in several places in the literature. The version employed here is the same as that of [Fuchs and Rutherford 2011], to which we refer the reader for more details. For an alternate approach, we expect that a DGA of the same form would arise from iterating the "bordered Chekanov–Eliashberg algebra" construction introduced in [Sivek 2011].

5B. ***Proof of Theorem 1.2(2).*** We begin by introducing notation. Given an involution τ of $\{1, \dots, N\}$, $\tau^2 = id$, we let $B_\tau = (b_{ij})$ denote the $N \times N$ matrix with entries

$$b_{ij} = \begin{cases} 1 & \text{if } i < \tau(i) = j, \\ 0 & \text{else.} \end{cases}$$

(B) $\Rightarrow$ (A). Suppose that L the diagram admits a generalized normal ruling $\rho = (\rho_1, \dots, \rho_m)$. An augmentation ε of the algebra $\mathcal{A}(L')$ is defined as follows: on all right cusps c_m, $\varepsilon(c_m) = 0$; at crossings b_m, $\varepsilon(b_m)$ is 1 if b_m is a switch and 0 otherwise; for all m, $\varepsilon(Y_m) = B_{\rho_m}$; and $\varepsilon(x^m_{i,j}) = 0$ for all i, j except when a switch occurs between x_{m-1} and x_m. Assume the switch involves the k and $k+1$ strands. If one of the switching strands is also a fixed point strand, then of the generators x^m_{ij} augment only $x^m_{k,k+1}$. Else, note that due to the normality condition, near the switch the intervals connecting the switching strands and their companion strands (Remark 2.3) are either disjoint or nested. Assume that the switch occurs between the strands labeled k and $k+1$. If the switch is disjoint, augment only $x^m_{k,k+1}$. If the switch is nested, augment $x^m_{k,k+1}$ and also $x^m_{\tau(k),\tau(k+1)}$ or $x^m_{\tau(k+1),\tau(k)}$, depending on whether $\tau(k) < \tau(k+1)$ or $\tau(k+1) < \tau(k)$.

It is straightforward to verify from the formulas of the previous section that ε is an augmentation. If ρ is p-graded with respect to a Maslov potential μ, then ε is as well.

(A) $\Rightarrow$ (B). The proof of the reverse implication is based on some canonical form results from linear algebra due to Barannikov [1994].

Definition 5.3. An M-complex, $(V, \mathcal{B}, d)$ is a vector space V over a field $\mathbb{F}$ with a chosen ordered basis $\mathcal{B} = \{v_1, \dots, v_N\}$ together with a differential $d : V \to V$, $d^2 = 0$, of the form $dv_i = \sum_{i<j} c_{ij} v_j$.

Proposition 5.4. *If $(V, \mathcal{B}, d)$ is an M-complex, then there exists a triangular change of basis $\{\tilde{v}_1, \dots, \tilde{v}_N\}$, with $\tilde{v}_i = \sum_{i\le j} a_{ij} v_j$, and an involution*

$$\tau : \{1, \dots, N\} \to \{1, \dots, N\}$$

such that

$$d\tilde{v}_i = \begin{cases} \tilde{v}_j & \text{if } i < \tau(i) = j, \\ 0 & \text{else.} \end{cases}$$

Moreover, the involution τ is unique.

Remark 5.5. (i) Suppose in addition that the basis elements v_i are assigned degrees $|v_i| \in \mathbb{Z}/p$ so that V is $\mathbb{Z}/p$-graded and d has degree -1. Then, the change of basis may be assumed to preserve degree. Hence, if $i < \tau(i) = j$, then $|v_i| = |v_j| + 1$.

(ii) The classes $[\tilde{v}_i]$ such that $\tau(i) = i$ form a basis for the homology $H(V, d)$.

(iii) Proposition 5.4 has the following matrix interpretation: There is a unique function, $D \mapsto \tau(D)$ which assigns to every strictly upper triangular $N \times N$ matrix D with $D^2 = 0$ an involution $\tau = \tau(D)$ such that there exists an invertible upper triangular matrix P so that $PDP^{-1} = B_\tau$. Notice that the uniqueness assertion implies that $\tau(QDQ^{-1}) = \tau(D)$ if Q is nonsingular and upper triangular.

Proposition 5.6 [Barannikov 1994]. *Suppose that $(V, \mathcal{B}, d)$ is an M-complex, and $k \in \{1, \dots, N\}$ is such that $dv_k = \sum_{k+1<j} c_{kj} v_j$ so that the triple $(V, \mathcal{B}', d)$ with $\mathcal{B}' = \{v_1, \dots, v_{k+1}, v_k, \dots, v_N\}$ is also an M-complex. Then, the associated involutions τ and τ' are related as follows.*

1. *It is always possible to have $\tau' = (k\ k+1) \circ \tau \circ (k\ k+1)$, where $(k\ k+1)$ denotes the transposition.*
2. *In the following cases, it is also possible to have $\tau' = \tau$:*
   - (a) *If $\tau(k+1) < \tau(k) < k < k+1$, or*
     $\tau(k) < k < k+1 < \tau(k+1)$, or
     $k < k+1 < \tau(k+1) < \tau(k)$.
   - (b) *If $\tau(k) < k < k+1 = \tau(k+1)$ or $\tau(k) = k < k+1 < \tau(k+1)$.*
   - (c) *If $\tau(k) = k < k+1 = \tau(k+1)$.*

Remark 5.7. (i) From the matrix perspective, Proposition 5.6 puts restrictions on $\tau(P_{k,k+1} D P_{k,k+1})$ when $P_{k,k+1}$ is the permutation matrix of the transposition $(k\ k+1)$ and the $k, k+1$-entry of D is 0.

(ii) Propositions 5.4 and 5.6 are essentially the same as Lemma 2 and Lemma 4 of [Barannikov 1994]. Proposition 5.6 is proven quite directly by considering cases.

Proof of **(A)** $\Rightarrow$ **(B)**. Suppose now that ε is an augmentation of $\mathcal{A}(L')$.

For each m, the matrix $\varepsilon(Y_m)$ is strictly upper triangular and satisfies

$$[\varepsilon(Y_m)]^2 = \varepsilon \circ d(Y_m) = 0.$$

Letting $\tau_m = \tau(\varepsilon(Y_m))$ as in Remark 5.5 produces a sequence, $\tau_1, \dots, \tau_M$, with τ_m an involution of $\{1, \dots, N(m)\}$. We show that $\tau = (\tau_1, \dots, \tau_M)$ satisfies the requirements of a generalized normal ruling. This requires establishing that the restrictions provided by Definitions 2.2 and 2.4 on consecutive involutions τ_{m-1} and τ_m are satisfied.

Recall that each interval (x_{m-1}, x_m) contains a single crossing or cusp.

If (x_{m-1}, x_m) *contains a left cusp*, then (5-2) and the definition of augmentation allow us to compute

$$\varepsilon(Y_m) = (I + \varepsilon(X_m))\varepsilon(\widetilde{Y}_{m-1})(I + \varepsilon(X_m))^{-1}. \tag{5-3}$$

Using Remark 5.5 we conclude that

$$\tau_m = \tau(\varepsilon(Y_m)) = \tau(\varepsilon(\widetilde{Y}_{m-1})).$$

The M-complex associated with $\varepsilon(\widetilde{Y}_{m-1})$ is related to that of $\varepsilon(Y_{m-1})$ by adding two new generators v_k and v_{k+1} to $\mathcal{B}$. The complex is the split extension of that of $\varepsilon(Y_{m-1})$ by span$\{v_k, v_{k+1}\}$ with the differential $dv_k = v_{k+1}$. It can then be checked from the definition that the involutions τ_{m-1} and τ_m satisfy Definition 2.2(2).

If (x_{m-1}, x_m) *contains a right cusp*, let

$$\mathcal{C} = (V_{m-1}, \mathcal{B} = \{v_i \mid i = 1, \dots, N(m-1)\}, d)$$

denote the M-complex associated with the matrix $\varepsilon(Y_{m-1})$ by the formula

$$dv_i = \sum_{i<j} \varepsilon(y_{ij}^{m-1}) v_j. \tag{5-4}$$

Note that τ_{m-1} is precisely the involution associated to $\mathcal{C}$ by Proposition 5.4. From $0 = \varepsilon \circ d(c_m)$ we deduce that $1 = \varepsilon(y_{k,k+1}^{m-1})$, and it follows that $\tau_{m-1}(k) = k+1$.

Next, one observes that $\varepsilon(\widetilde{Y}_{m-1})$ is the matrix of the M-complex

$$\widetilde{\mathcal{C}} = \big(\widetilde{V}_{m-1}, \widetilde{\mathcal{B}} = \{[v_i] \mid i \neq k, k+1\}, \tilde{d}\big),$$

where $\widetilde{V}_{m-1}$ is the quotient of V_{m-1} by the subcomplex

$$\{v_k + \varepsilon(c_m)v_{k+1}, d(v_k + \varepsilon(c_m)v_{k+1})\}$$

and $\tilde{d}$ is the differential induced by d. If $\{\tilde{v}_i\}$ is a triangular change of basis for $\mathscr{C}$ satisfying the conditions of Proposition 5.4, then $\{[\tilde{v}_i] \mid i \neq k, k+1\}$ will be such a basis for $\widetilde{\mathscr{C}}$, so that the involution associated with $\varepsilon(\widetilde{Y}_{m-1})$ is related to τ_{m-1} as required in Definition 2.2(3). Finally, we get $\tau_m = \tau(\varepsilon(Y_m)) = \tau(\varepsilon(\widetilde{Y}_{m-1}))$ by using (5-3).

If (x_{m-1}, x_m) *contains a crossing* b_m, we have $0 = \varepsilon \circ d(b_m) = \varepsilon(y^{m-1}_{k,k+1})$. Thus, $\varepsilon(\widehat{Y}_{m-1}) = \varepsilon(Y_{m-1})$ with both matrices having 0 as their $(k, k+1)$ entry. Then, compute that

$$\varepsilon(B_{k,k+1})\varepsilon(\widehat{Y}_{m-1})\varepsilon(B^{-1}_{k,k+1}) = P_{k,k+1}[I + \varepsilon(b_m)E_{k,k+1}]\varepsilon(Y_{m-1})[I + \varepsilon(b_m)E_{k,k+1}]P_{k,k+1}.$$

Regardless of the value of $\varepsilon(b_m)$, the $(k, k+1)$-entry of

$$[I + \varepsilon(b_m)E_{k,k+1}]\varepsilon(Y_{m-1})[I + \varepsilon(b_m)E_{k,k+1}]$$

is 0, so the matrix $A = \varepsilon(B_{k,k+1})\varepsilon(\widehat{Y}_{m-1})\varepsilon(B^{-1}_{k,k+1})$ is strictly upper triangular and $\tau(A)$ is related to

$$\tau\left((I + \varepsilon(b_m)E_{k,k+1})\varepsilon(Y_{m-1})(I + \varepsilon(b_m)E_{k,k+1})\right) = \tau(Y_{m-1}) = \tau_{m-1}$$

as in Proposition 5.6. It follows that

$$\tau_m = \tau(\varepsilon(Y_m)) = \tau((I + \varepsilon(X_m))A(I + \varepsilon(X_m))^{-1}) = \tau(A)$$

and τ_{m-1} satisfy the requirements near crossings (including the normality conditions) of Definition 2.4.

The statement that τ is p-graded if ε is p-graded follows from (i) of Remark 5.5. As in (5-4), $\varepsilon(Y_m)$ is the matrix of an M-complex with basis $v_1, \dots, v_{N(m)}$ corresponding to the strands of L at x_m. If ε is p-graded with respect to μ, then we can assign a grading by $|v_i| = \mu(m, i)$ and the differential will have degree -1. □

Acknowledgements

This work was initiated through the PRUV program at Duke University. We thank David Kraines for supervising the program and encouraging our participation. Also, we thank Lenny Ng for his interest in the project. A portion of the writing was carried out while the second author was a visitor at the Max Planck Institute for Mathematics in Bonn, and it is a pleasure to acknowledge the MPIM for their hospitality.

References

[Barannikov 1994] S. A. Barannikov, "The framed Morse complex and its invariants", pp. 93–115 in *Singularities and bifurcations*, edited by V. I. Arnold, Adv. Soviet Math. **21**, Amer. Math. Soc., Providence, RI, 1994. MR 96a:57065 Zbl 0996.57514

[Chekanov 2002] Y. Chekanov, "Differential algebra of Legendrian links", *Invent. Math.* **150**:3 (2002), 441–483. MR 2003m:53153 Zbl 1029.57011

[Chekanov and Pushkar 2005] Y. V. Chekanov and P. E. Pushkar, "Combinatorics of fronts of Legendrian links, and Arnold's 4-conjectures", *Uspekhi Mat. Nauk* **60**:1(361) (2005), 99–154. In Russian; translated in *Russian Math. Surveys* **60**:1 (2005), 95–149. MR 2006h:58049 Zbl 1085.57008

[Chmutov and Goryunov 1997] S. Chmutov and V. Goryunov, "Polynomial invariants of Legendrian links and plane fronts", pp. 25–43 in *Topics in singularity theory*, edited by A. Khovanskiĭ et al., Amer. Math. Soc. Transl. Ser. 2 **180**, Amer. Math. Soc., Providence, RI, 1997. MR 2002c:57015 Zbl 0886.57005

[Cornwell 2012a] C. Cornwell, "Bennequin type inequalities in lens spaces", *Int. Math. Res. Not.* **2012**:8 (2012), 1890–1916. Zbl 06034372 arXiv 1002.1546

[Cornwell 2012b] C. Cornwell, "A polynomial invariant for links in lens spaces", *J. Knot Theory Ramifications* **21**:6 (2012), # 1250060. Zbl 06025162 arXiv 1002.1543v2

[Ding and Geiges 2010] F. Ding and H. Geiges, "Legendrian helix and cable links", *Commun. Contemp. Math.* **12**:3 (2010), 487–500. MR 2011j:53174 Zbl 1202.53082

[Fuchs 2003] D. Fuchs, "Chekanov–Eliashberg invariant of Legendrian knots: existence of augmentations", *J. Geom. Phys.* **47**:1 (2003), 43–65. MR 2004h:57007 Zbl 1028.57005

[Fuchs and Ishkhanov 2004] D. Fuchs and T. Ishkhanov, "Invariants of Legendrian knots and decompositions of front diagrams", *Mosc. Math. J.* **4** (2004), 707–717. MR 2005i:57033 Zbl 1073.53106

[Fuchs and Rutherford 2011] D. Fuchs and D. Rutherford, "Generating families and Legendrian contact homology in the standard contact space", *J. Topol.* **4**:1 (2011), 190–226. MR 2012e:57046 Zbl 1237.57026

[Fuchs and Tabachnikov 1997] D. Fuchs and S. Tabachnikov, "Invariants of Legendrian and transverse knots in the standard contact space", *Topology* **36**:5 (1997), 1025–1053. MR 99a:57006 Zbl 0904.57006

[Geiges 2008] H. Geiges, *An introduction to contact topology*, Cambridge Studies in Advanced Mathematics **109**, Cambridge University Press, 2008. MR 2008m:57064 Zbl 1153.53002

[Licata 2011] J. E. Licata, "Invariants for Legendrian knots in lens spaces", *Commun. Contemp. Math.* **13**:1 (2011), 91–121. MR 2772580 Zbl 05864098

[Ng and Traynor 2004] L. Ng and L. Traynor, "Legendrian solid-torus links", *J. Symplectic Geom.* **2**:3 (2004), 411–443. MR 2005k:57051 Zbl 1097.57014

[Rutherford 2006] D. Rutherford, "Thurston–Bennequin number, Kauffman polynomial, and ruling invariants of a Legendrian link: the Fuchs conjecture and beyond", *Int. Math. Res. Not.* **2006** (2006), Art. ID 78591. MR 2007a:57020 Zbl 1106.57012

[Rutherford 2011] D. Rutherford, "HOMFLY-PT polynomial and normal rulings of Legendrian solid torus links", *Quantum Topol.* **2**:2 (2011), 183–215. MR 2783131 Zbl 1230.57016

[Sabloff 2005] J. M. Sabloff, "Augmentations and rulings of Legendrian knots", *Int. Math. Res. Not.* **2005**:19 (2005), 1157–1180. MR 2006e:57010 Zbl 1082.57020

[Sivek 2011] S. Sivek, "A bordered Chekanov–Eliashberg algebra", *J. Topol.* **4**:1 (2011), 73–104. MR 2783378 Zbl 1219.57022

[Traynor 1997] L. Traynor, "Legendrian circular helix links", *Math. Proc. Cambridge Philos. Soc.* **122**:2 (1997), 301–314. MR 98f:58085 Zbl 0902.57006

[Turaev 1988] V. G. Turaev, "The Conway and Kauffman modules of a solid torus", *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **167**:Issled. Topol. 6 (1988), 79–89. In Russian; translated in *J. Soviet Math.* **52** (1990), 2799–2805. MR 90f:57012 Zbl 0673.57004

Received October 4, 2011. Revised December 8, 2011.

MIKHAIL LAVROV
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PA 15213
UNITED STATES
mlavrov@andrew.cmu.edu

DAN RUTHERFORD
UNIVERSITY OF ARKANSAS
FAYETTEVILLE, AR 72701
UNITED STATES
drruther@uark.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

CLASSIFICATION OF SINGULAR $\mathbb{Q}$-HOMOLOGY PLANES II: $\mathbb{C}^1$- AND $\mathbb{C}^*$-RULINGS

KAROL PALKA

A $\mathbb{Q}$-homology plane is a normal complex algebraic surface having trivial rational homology. We classify singular $\mathbb{Q}$-homology planes that are $\mathbb{C}^1$- or $\mathbb{C}^*$-ruled. We analyze their completions, the number of different rulings they have, and the number of affine lines on them; and we give constructions. Together with previously known results, this completes the classification of $\mathbb{Q}$-homology planes with smooth locus of nongeneral type. We show also that the dimension of a family of homeomorphic but nonisomorphic singular $\mathbb{Q}$-homology planes having the same weighted boundary, singularities and Kodaira dimension can be arbitrarily big.

We work with complex algebraic varieties.

1. Main results

A $\mathbb{Q}$-homology plane is a normal surface whose rational cohomology is the same as that of $\mathbb{C}^2$. This paper is the last piece of the classification of $\mathbb{Q}$-homology planes having smooth locus of nongeneral type. The classification is built on the work of many authors; for a summary of what is known about smooth and singular $\mathbb{Q}$-homology planes, see [Miyanishi 2001, §3.4] and [Palka 2011b]. In [Palka 2008], we classified singular $\mathbb{Q}$-homology planes with nonquotient singularities, showing in particular that they are quotients of affine cones over projective curves by actions of finite groups that respect the set of lines through the vertex. In [Palka 2011a], we classified singular $\mathbb{Q}$-homology planes whose smooth locus is of nongeneral type and admits no $\mathbb{C}^1$- or $\mathbb{C}^*$-ruling (*exceptional planes*). Here we classify singular $\mathbb{Q}$-homology planes that admit a $\mathbb{C}^1$- or a $\mathbb{C}^*$-ruling. We analyze completions and boundaries rather than the open surfaces themselves. To deal with nonuniqueness of these, we use the notions of a *balanced* and a *strongly balanced* weighted boundary and completion of an open surface (see Definitions 2.7 and 2.10).

We classify $\mathbb{C}^1$- and $\mathbb{C}^*$-ruled $\mathbb{Q}$-homology planes by giving necessary and sufficient conditions for a $\mathbb{C}^1$- or $\mathbb{C}^*$-ruled open surface to be a $\mathbb{Q}$-homology plane

The author was supported by Polish Grant NCN N N201 608640.
MSC2010: primary 14R05; secondary 14J17, 14J26.
Keywords: acyclic surface, homology plane, $\mathbb{Q}$-homology plane.

(see Lemmas 2.12, 3.2 and 4.4 and the remarks before the latter) and then giving general constructions (see Construction 3.3 and Section 4D). We compute the Kodaira dimension of a $\mathbb{C}^*$-ruled singular $\mathbb{Q}$-homology plane and of its smooth locus (Theorem 4.9) in terms of properties of singular fibers, and we list the planes with smooth locus of Kodaira dimension zero (Section 4C). As a corollary of the classification, we obtain the following result.

Theorem 1.1. *Let S' be a singular $\mathbb{Q}$-homology plane, and let S_0 be its smooth locus. Assume that S' is not affine-ruled and that $\bar{\kappa}(S_0) \neq 2$.*

(1) *Either S' has a unique balanced completion up to isomorphism, or it admits an untwisted $\mathbb{C}^*$-ruling with base $\mathbb{C}^1$ and more than one singular fiber. In the latter case, S' has exactly two strongly balanced completions.*

(2) *If S' has more than one singular point, then it has exactly two singular points, both of Dynkin type A_1, and there is a twisted $\mathbb{C}^*$-ruling of S' such that both singular points are contained in a unique fiber isomorphic to $\mathbb{C}^1$.*

(3) *If S' contains a quotient noncyclic singularity, then either $S' \cong \mathbb{C}^2/G$ for a small finite noncyclic subgroup of* GL(2, $\mathbb{C}$), *or S' has a twisted $\mathbb{C}^*$-ruling. In the latter case, the unique fiber isomorphic to $\mathbb{C}^1$ is of type* (*A*)(*iv*) (*see Theorem 4.9*) *and contains a singular point of Dynkin type D_k for some $k \geq 4$.*

We now comment on other corollaries of the classification. First, the case can occur when S' has exactly one singular point and it is a cyclic singularity. Second, we show that if S' is affine-ruled, then its strongly balanced weighted boundary is unique unless it is a chain, but that even if it is unique, there still may be infinitely many strongly balanced completions (see Example 3.6). Third, the singularities of affine-ruled S' are necessarily cyclic, but there may be arbitrarily many of them (see [Miyanishi and Sugie 1991] or Section 3). Regarding the remaining case $\bar{\kappa}(S_0) = 2$, which we do not analyze here, let us mention that it follows from the logarithmic Bogomolov–Miyaoka–Yau inequality (see [Palka 2008], for example) that S' has only one singular point and it is of quotient type.

It is known that smooth $\mathbb{Q}$-homology planes can have moduli [Flenner and Zaĭdenberg 1994]. The same is true for singular ones. We prove the following result.

Theorem 1.2. *There exist arbitrarily high-dimensional families of nonisomorphic singular $\mathbb{Q}$-homology planes having smooth locus of negative Kodaira dimension and having the same singularities, same homeomorphism type, and same weighted strongly balanced boundary.*

An important property of any $\mathbb{Q}$-homology plane with smooth locus of general type is that it does not contain topologically contractible curves. In fact, the number of contractible curves on a $\mathbb{Q}$-homology plane is known except in the case when the surface is singular and the smooth locus has Kodaira dimension zero (see Section 6).

In Theorem 6.1, we compute the number of different $\mathbb{C}^*$-rulings a $\mathbb{Q}$-homology plane can have. The computation of the number of contractible curves follows from it.

Theorem 1.3. *If a singular $\mathbb{Q}$-homology plane has smooth locus of Kodaira dimension zero, then it contains one or two irreducible topologically contractible curves if the smooth locus admits a $\mathbb{C}^*$-ruling, and no such curves otherwise.*

The notion of a balanced weighted boundary of an open surface (see Definition 2.10) is a more flexible version of the notion of a *standard graph* from [Flenner et al. 2007], which has its origin in [Daigle 2008]. It follows from above that every $\mathbb{Q}$-homology plane admits up to isomorphism one or two strongly balanced boundaries, but this is not so for the standard ones. The set of such boundaries is a useful invariant of the surface.

Integral homology groups and necessary conditions for singular fibers of $\mathbb{C}^1$- and $\mathbb{C}^*$-ruled $\mathbb{Q}$-homology planes have already been analyzed in [Miyanishi and Sugie 1991]. For $\mathbb{C}^*$-rulings, however, these conditions are not sufficient (see Examples 4.2 and 4.3), and a more detailed analysis is necessary. Also, some formulas for the Kodaira dimension in terms of singular fibers from [Miyanishi and Sugie 1991] require nontrivial corrections (see Section 4B).

2. Preliminaries

We follow the notational conventions and terminology of [Miyanishi 2001], [Fujita 1982] and [Palka 2008]. We recall some of them for the convenience of the reader.

2A. *Divisors and normal pairs.* Let $T = \sum t_i T_i$ be an snc-divisor on a smooth complete surface with distinct irreducible components T_i. Then $\underline{T} = \sum T_i$, where the sum runs over i with $t_i \neq 0$, is the reduced divisor with the same support as T, and $\beta_T(T_i) = \underline{T} \cdot (\underline{T} - T_i)$ is the *branching number* of T_i. A *tip* has $\beta_T(T_i) \leq 1$. By $Q(T)$ we denote the intersection matrix of T; we put $d(0) = 1$ and $d(T) = \det(-Q(T))$ for $T \neq 0$. The symbol "$\equiv$" denotes numerical equivalence of divisors.

If T is reduced and its dual graph is linear, it is called a *chain*, and in writing it as a sum of irreducible components $T = T_1 + \cdots + T_n$, we assume that $T_i \cdot T_{i+1} = 1$ for $1 \leq i \leq n-1$. We put $T^t = T_n + \cdots + T_1$. If T is a rational chain, then we write $T = [-T_1^2, \ldots, -T_n^2]$. A rational chain with all $T_i^2 \leq -2$ is called *admissible*. A *fork* is a rational tree for which the branching component is unique and has $\beta = 3$.

Let D be some reduced snc-divisor that is not an admissible chain. A rational chain with support contained in D, not containing branching components of D and containing one of its tips, is called a *twig* of D. For an admissible (ordered) chain, we put

$$e(T) = \frac{d(T - T_1)}{d(T)} \quad \text{and} \quad \tilde{e}(T) = e(T^t).$$

In general, $e(T)$ and $\tilde{e}(T)$ are defined as the sums of respective numbers computed for all maximal admissible twigs of T. Here we use the convention that the tip of the twig is the first component.

If X is a complete surface and D is a reduced snc-divisor contained in the smooth part of X, then we call (X, D) an *snc-pair* and we write $X - D$ for $X \setminus D$. The pair is *normal* (resp. *smooth*) if X is normal (resp. smooth). If X is a normal surface, then an embedding $\iota\colon X \to \overline{X}$, where $(\overline{X}, \overline{X} \setminus X)$ is a normal pair, is called a *normal completion* of X. If X is smooth, then $\overline{X}$ is smooth and $(\overline{X}, D, \iota)$ is called a *smooth completion* of X. A morphism of two completions $\iota_j : X \to \overline{X}_j$, with $j = 1, 2$, of a given surface X is a morphism $f\colon \overline{X}_1 \to \overline{X}_2$ such that $\iota_2 = f \circ \iota_1$.

Let $\pi : (X, D) \to (X', D')$ be a birational morphism of normal pairs. We put $\pi^{-1}D' = \underline{\pi^* D'}$; that is, $\pi^{-1}D'$ is the reduced total transform of D'. Assume $\pi^{-1}D' = D$. If π is a blow-up, then we call it *subdivisional* (resp. *sprouting*) for D' if its center belongs to two (resp. one) components of D'. In general, we say that π is *subdivisional* for D' (and for D) if for any component T of D' we have $\beta_{D'}(T) = \beta_D(\pi^{-1}T)$. The exceptional locus of a birational morphism between two surfaces $\eta : X \to X'$, denoted by $\mathrm{Exc}(\eta)$, is defined as the locus of points in X for which η is not a local isomorphism.

A b-curve is a smooth rational curve with self-intersection b. A divisor is snc-minimal if all of its (-1)-curves are branching. We write K_X for the canonical divisor on a complete surface X.

Definition 2.1. A birational morphism of surfaces $\pi : X \to X'$ is a *connected modification* if it is proper, $\pi(\mathrm{Exc}(\pi))$ is a smooth point on X', and $\mathrm{Exc}(\pi)$ contains a unique (-1)-curve. If π is a morphism of pairs $\pi : (X, D) \to (X', D')$ such that $\pi^{-1}(D') = D$ and $\pi(\mathrm{Exc}(\pi)) \in D'$, we call it a *connected modification over* D'.

A sequence of blow-downs (and its reversing sequence of blow-ups) whose composition is a connected modification is called a *connected sequence of blow-downs* (*blow-ups*).

2B. *Rational rulings.* A surjective morphism $p_0 : X_0 \to B_0$ of a normal surface onto a smooth curve is a *rational ruling* if general fibers are rational curves. By a *completion of* p_0, we mean a triple (X, D, p), where (X, D) is a normal completion of X_0 and $p\colon X \to B$ is an extension of p_0 to a $\mathbb{P}^1$-ruling, with B being a smooth completion of B_0. We say that p is a *minimal completion of* p_0 if p does not dominate any other completion of p_0. In this case we also say that D is *p-minimal*. It is easy to check that D is p-minimal if and only if all of its nonbranching (-1)-curves are horizontal. Let F be a fiber of p. An irreducible curve $G \subseteq X$ is an *n-section* of p if $G \cdot F = n$. A *section* is a 1-section. We call p_0 a $\mathbb{C}^{(n*)}$-ruling if $F \cdot D = n + 1$ for $n \geq 1$. In the case $n = 0$, we call p_0 a $\mathbb{C}^1$-ruling or an *affine ruling*; the arithmetic genus of F ($p_a(F) = \frac{1}{2} F \cdot (K_X + F) + 1$) vanishes and

$F^2 = 0$. Conversely, it is well-known that an effective divisor with these properties on a complete surface is a fiber of such a ruling [Barth et al. 2004, V.4.3]. If J is a component of F, then we denote by $\mu_F(J)$ the multiplicity of J; that is, $F = \mu_F(J)J + F'$, where F' is effective and $J \not\subseteq F'$. The structure of fibers of a $\mathbb{P}^1$-ruling is well known [Fujita 1982, §4].

Lemma 2.2. *Let F be a singular fiber of a $\mathbb{P}^1$-ruling of a smooth complete surface. Then F is a tree of rational curves and it contains a (-1)-curve. Each (-1)-curve of F meets at most two other components. If F contains a unique (-1)-curve C, then*:

(i) $\mu(C) > 1$. *There are exactly two components of F with multiplicity one*, *and they are tips of the fiber.*

(ii) *If $\mu(C) = 2$, then either $F = [2, 1, 2]$ or C is a tip of F*; *in the latter case either $\underline{F} - C = [2, 2, 2]$ or $\underline{F} - C$ is a (-2)-fork of type $(2, 2, n)$.*

(iii) *If $\underline{F}$ is not a chain*, *then the connected component of $\underline{F} - C$ not containing curves of multiplicity one is a chain* (*possibly empty*).

We define

$$\Sigma_{X-D} = \sum_{F \not\subseteq D} (\sigma(F) - 1),$$

where $\sigma(F)$ is the number of $(X - D)$-components of a fiber F [Fujita 1982, 4.16]. If p is a $\mathbb{P}^1$-ruling as above, then we call an irreducible curve G *vertical* (*for* p) if $p_*G = 0$; otherwise it is *horizontal*. A divisor is vertical (resp. horizontal) if all of its components are vertical (resp. horizontal). We decompose D as $D = D_h + (D - D_h)$, where D_h is horizontal and $D - D_h$ is vertical. The numbers h and ν are defined respectively as the number of irreducible components of D_h and as the number of fibers contained in D. We have [Fujita 1982, §4]

$$\Sigma_{X-D} = h + \nu + b_2(X) - b_2(D) - 2.$$

We call a connected component of $F \cap D$ a *D-rivet* (or *rivet* if this causes no confusion) if it meets D_h at more than one point or if it is a node of D_h.

Definition 2.3. Suppose (X, D, p) is a completion of a $\mathbb{C}^*$-ruling of a normal surface X. We say that the original ruling $p_0 = p_{|X-D}$ is *twisted* if D_h is a 2-section. If D_h consists of two sections, we say that p_0 is *untwisted*. Let F be a singular fiber of p that does not contain singular points of X. We say that F is *columnar* if $\underline{F}$ is a chain that can be written as

$$\underline{F} = A_n + \cdots + A_1 + C + B_1 + \cdots + B_m,$$

where C is a unique (-1)-curve and D_h meets F exactly in A_n and B_m. The chains $A = A_1 + \cdots + A_n$ and $B = B_1 + \cdots + B_m$ are called *adjoint chains*.

Remark. By expansion properties of determinants (see [Koras and Russell 2007, 2.1.1], for example) and the fact that $d(A)$ and $d(A-A_1)$ are coprime, we have $e(A)+e(B)=1$ and $d(A)=d(B)=\mu_F(C)$. In fact, we have also $\tilde{e}(B)+\tilde{e}(A)=1$ [Fujita 1982, 3.7].

2C. *Balanced completions.*

Definition 2.4. A pair (D, w) consisting of a complete curve D and a rationally valued function w defined on the set of irreducible components of D is called a *weighted curve*. If (X, D) is a normal pair, then (D, w) with w defined by $w(D_i)=D_i^2$ is a *weighted boundary* of $X-D$.

Definition 2.5. Let (X, D) be a normal pair.

(i) Let L be a 0-curve that is a nonbranching component of D, and let $c \in L$ be chosen so that if L intersects two other components of D, then c is one of the points of intersection. Make a blow-up of c and contract the proper transform of L. The resulting pair (X', D'), where D' is the reduced direct image of the total transform of D, is called an *elementary transform of* (X, D). The pair $\Phi=(\Phi^\circ, \Phi^\bullet)$ consisting of an assignment $\Phi^\circ\colon (X, D)\mapsto (X', D')$ together with the resulting rational mapping $\Phi^\bullet\colon X \dashrightarrow X'$ is called an *elementary transformation over* D. Φ is *inner* (*for* D) if $\beta_D(L)=2$, and *outer* (*for* D) if $\beta_D(L)=1$. The point $c\in L$ is the *center* of Φ.

(ii) For a sequence of (inner) elementary transformations

$$\Phi_i^\circ\colon (X_i, D_i)\mapsto (X_{i+1}, D_{i+1}),$$

with $i=1,\dots,n-1$, we put $\Phi^\circ=(\Phi_1^\circ,\dots,\Phi_{n-1}^\circ)$, $\Phi^\bullet=(\Phi_1^\bullet,\dots,\Phi_{n-1}^\bullet)$ and we call $\Phi=(\Phi^\circ, \Phi^\bullet)$ *an* (*inner*) *flow in* D_1. We denote it by $\Phi\colon (X_1, D_1)\rightsquigarrow (X_n, D_n)$.

$\Phi^\bullet=(\Phi_1^\bullet,\dots,\Phi_{n-1}^\bullet)$ induces a rational mapping $X_1 \dashrightarrow X_n$, which we also denote by $\Phi^\bullet$. There exists the largest open subset of X_1 on which $\Phi_1^\bullet$ is a morphism; the complement of this subset is called the *support of* Φ. Clearly, $\operatorname{Supp}\Phi_1\subseteq D_1$. If $\operatorname{Supp}\Phi=\varnothing$, then Φ is a *trivial flow*.

A weighted curve (D, w) determines the weighted dual graph of D. If (D, w) is a weighted boundary coming from a fixed normal pair (X, D), we omit the weight function w from the notation. For Φ as above, D_1 and D_n are isomorphic as curves. They have the same dual graphs, but usually different weights of components.

Example 2.6. Let $T=[0,0,a_1,\dots,a_n]$. Each chain of type $[0,b,a_1,\dots,a_n]$, $[a_1,\dots,a_{k-1},a_k-b,0,b,a_{k+1},\dots,a_n]$ or $[a_1,\dots,a_n,b,0]$, where $1\le k\le n$ and $b\in\mathbb{Z}$, can be obtained from T by a flow. This follows from the observation that an elementary transformation interchanges the chains $[w,x,0,y-1,z]$ and $[w,x-1,0,y,z]$. Looking at the dual graph, we see the weights can “flow” from one side of a 0-curve to another, and possibly vanish ($b = 0$ or $b = a_k$). If they do, then again the weights can flow through the new zero.

Definition 2.7. A rational chain $D = [a_1, \dots, a_n]$ is *balanced* if $a_1, \dots, a_n \in \{0, 2, 3, \dots\}$ or if $D = [1]$. A reduced snc-divisor whose dual graph contains no loops (snc-forest) is *balanced* if all rational chains contained in D that do not contain branching components of the divisor are balanced. A normal pair (X, D) is *balanced* if D is balanced.

Recall that if (X_i, D_i) for $i = 1, 2$ are normal pairs such that $X_1 - D_1 \cong X_2 - D_2$, then D_1 is a forest if and only if D_2 is a forest.

Proposition 2.8. *A normal surface that admits a normal completion with a forest as a boundary has a balanced completion. Two such completions differ by a flow.*

As we discovered after completing the proof, a more general version of this proposition was proved in a graph-theoretic context in [Flenner et al. 2007, Theorem 3.1 and Corollary 3.36]. We therefore leave our more direct arguments to be published elsewhere. In fact, some key observations were made earlier in [Daigle 2008, 4.23.1, 3.2, 5.2]. Let us restate some definitions from [Flenner et al. 2007] on the level of pairs.

Definition 2.9. Let (X, D) be a normal pair and assume D is an snc-forest.

(i) Connected components of the divisor that remains after subtracting all nonrational and all branching components of D are called the *segments of* D.

(ii) D is *standard* if for each of its connected components, either the component is equal to [1] or all of its segments are of types [0], [0, 0, 0] or $[0^{2k}, a_1, \dots, a_n]$, with $k \in \{0, 1\}$ and $a_1, \dots, a_n \geq 2$.

(iii) Let $D_0 = [0, 0, a_1, \dots, a_n]$, with $a_i \geq 2$ for $i = 1, \dots, n$, be a segment of D. A *reversion of* D_0 is a nontrivial flow $\Phi : (X, D) \rightsquigarrow (X', D')$ that is supported in D_0, is inner for D_0, and satisfies $D' - (\Phi^\bullet)_*(D - D_0) = [a_1, a_2, \dots, a_n, 0, 0]$.

The condition that Φ be nontrivial is introduced for the following reason: we want the reversion to transform the two zeros to the other end of the chain, and the condition in necessary to force this in case D is symmetric, that is, when $[a_1, \dots, a_n]^t = [a_1, \dots, a_n]$. Standard chains are called *canonical* in [Daigle 2008]. The Hodge index theorem implies that if (X, D) is a smooth pair and D is a forest, then it cannot have segments of type $[0^{2k+1}]$ or $[0^{2k}, a_1, \dots, a_n]$ for $k > 1$, and can have at most one such segment for $k = 1$.

Clearly, not every balanced forest is standard, but by a flow one can easily make it so. It follows from Proposition 2.8 that if D and D' are two standard boundaries of the same surface and D is a chain, then either D and D' are isomorphic as weighted curves or D' is the reversion of D. Unfortunately, the notion of a standard boundary is not as restrictive as one may imagine, and the difference between two standard boundaries can be more than just a reversion of some segments. An additional ambiguity is related to the existence of segments of type $[0^{2k+1}]$. Specifically, if $[0^{2k+1}]$ is a segment of D, then one can change by a flow the self-intersections of the components of D intersecting the segment. For example, consider a surface whose standard boundary is a rational fork with a dual graph

$$\begin{array}{ccccc} -2 & \text{——} & b & \text{——} & -2 \\ & & \vert & & \\ & & 0 & & \end{array}$$

for some $b \in \mathbb{Z}$. Then for any $b \in \mathbb{Z}$, there is a completion of this surface for which the boundary is standard and has the dual graph as above.[1] We therefore introduce the following more restrictive conditions.

Definition 2.10. A balanced snc-forest D is *strongly balanced* if it is standard and either D contains no segments of type $[0]$ or $[0, 0, 0]$, or for at least one such segment there is a component $B \subseteq D$ intersecting it such that $B^2 = 0$. A normal pair (X, D) for which D is a forest is *strongly balanced* if D is strongly balanced.

2D. ***Basic properties of $\mathbb{Q}$-homology planes.*** We assume that S' is a *singular $\mathbb{Q}$-homology plane*, that is, a normal nonsmooth complex algebraic surface with $H^*(S', \mathbb{Q}) \cong \mathbb{Q}$. Let $\epsilon\colon S \to S'$ be a resolution such that the inverse image of the singular locus is an snc-divisor, and let $(\bar{S}, D)$ be a smooth completion of S. Denote the singular points of S' by $p_1, \dots, p_q$ and the smooth locus by S_0. We put $\widehat{E}_i = \epsilon^{-1}(p_i)$ and assume that $\widehat{E} = \widehat{E}_1 + \widehat{E}_2 + \cdots + \widehat{E}_q$ is snc-minimal. Recall that S' is called *logarithmic* if and only if every singular point of S' is locally analytically isomorphic to $\mathbb{C}^2/G$ for some finite subgroup $G < \mathrm{GL}(2, \mathbb{C})$ (a *quotient singularity*). In [Palka 2008], we classified nonlogarithmic $\mathbb{Q}$-homology planes. In particular, it is known that they do not admit $\mathbb{C}^1$- or $\mathbb{C}^*$-rulings. Therefore, from now on we assume that S' is logarithmic. It follows that each $\widehat{E}_i$ is either an admissible chain or an admissible fork (that is, an snc-minimal fork with negative definite intersection matrix). By [Gurjar et al. 1997], S' is rational. By the argument in [Fujita 1982, 2.4], it is affine.

Proposition 2.11. *Let the notation be as above.*

(i) *D is a rational tree with $d(D) = -d(\widehat{E}) \cdot |H_1(S', \mathbb{Z})|^2$.*

(ii) *The embedding $D \cup \widehat{E} \to \bar{S}$ induces an isomorphism on $H_2(-, \mathbb{Q})$.*

[1] This observation was missed in [Flenner et al. 2007], whose Corollary 3.33 is false. See [Flenner et al. 2011] for corrections. In [Daigle 2008, Solution to problem 5] this ambiguity is implicitly taken into account without restricting to balanced divisors.

(iii) $\pi_1(S') \cong \pi_1(S)$ *and* $H_k(S', \mathbb{Z}) = 0$ *for* $k > 1$.

(iv) $b_i(S_0) = 0$ *for* $i = 1, 2, 4$ *and* $b_3(S_0) = q$.

(v) $\Sigma_{S_0} = h + \nu - 2$ *and* $\nu \leq 1$.

Proof. See [Palka 2008, 3.1, 3.2] and [Miyanishi and Sugie 1991, 2.2]. □

Lemma 2.12. *Let* $(\bar{S}, T)$ *be a smooth pair and let* $p \colon \bar{S} \to B$ *be a* $\mathbb{P}^1$*-ruling. Assume that*

(i) *there exists a unique connected component* D *of* T *that is not vertical*,

(ii) D *is a rational tree*,

(iii) $\Sigma_{\bar{S}-T} = h + \nu - 2$, *and*

(iv) $d(D) \neq 0$.

Then the surface S' *defined as the image of* $\bar{S} - D$ *after contraction of connected components of* $T - D$ *to points is a rational* $\mathbb{Q}$*-homology plane*, *and* p *induces a rational ruling of* S'. *Conversely*, *if* $p' \colon S' \to B$ *is a rational ruling of a rational* $\mathbb{Q}$*-homology plane* S', *then any completion* $(\bar{S}, T, p)$ *of the restriction of* p' *to the smooth locus of* S' *has the above properties.*

Proof. Since the base of p has some component of D as a branched cover, it is rational, and hence $\bar{S}$ is rational. We may assume that T is p-minimal. Put $\widehat{E} = T - D$. Since $\widehat{E}$ is vertical and since $\widehat{E} \cap D = \varnothing$, $Q(\widehat{E})$ is negative definite and $b_1(\widehat{E}) = 0$. Fujita's equation

$$\Sigma_{\bar{S}-T} = h + \nu - 2 + b_2(\bar{S}) - b_2(D + \widehat{E})$$

gives $b_2(\bar{S}) = b_2(T)$, so by (iv), the inclusion $T \to \bar{S}$ induces an isomorphism on $H_2(-, \mathbb{Q})$. By [Palka 2008, 2.6], S' is normal and affine, and in particular $b_4(S') = b_3(S') = 0$. Since $b_1(D) = 0$, the exact sequence of the pair $(\bar{S}, D)$ together with the Lefschetz duality give

$$b_2(S) = b_2(\bar{S}, D) = b_2(\bar{S}) - b_2(D) = b_2(\widehat{E}).$$

Since $b_1(\widehat{E}) = 0$, we get from the exact sequence of the pair $(S, \widehat{E})$ that $b_2(S') = b_2(S, \widehat{E}) = b_2(S) - b_2(\widehat{E}) = 0$. Now

$$\chi(S') = \chi(\bar{S}) - \chi(D \cup \widehat{E}) + b_0(\widehat{E}) = b_0(D) = 1,$$

so we obtain $b_1(S') = b_2(S') = 0$, and hence S' is $\mathbb{Q}$-acyclic.

Conversely, if p' is as above, then let $\widehat{E}$ be an exceptional divisor of a resolution of singularities of S', and let $D = T - \widehat{E}$. Since $\widehat{E}$ is vertical for the $\mathbb{P}^1$-ruling p, we have $b_1(\widehat{E}) = 0$. Then the necessity of the above conditions follows from [ibid., 3.1 and 3.2]. □

3. Smooth locus of negative Kodaira dimension

Here we assume that the smooth locus S_0 of the logarithmic $\mathbb{Q}$-homology plane S' has negative Kodaira dimension, implying that the Kodaira dimension of S' is also negative. This case was analyzed and a structure theorem given in [Miyanishi and Sugie 1991, 2.5–2.8]. We recover these results in Lemma 3.2 and Proposition 3.1, but we concentrate on analyzing possible completions and boundaries instead of S' itself. This gives more information, allowing us to give a construction and to answer the question of uniqueness of an affine ruling of S_0 (if it exists). The information about completions is also used in the analysis of an example where moduli occur.

Proposition 3.1. *If a singular $\mathbb{Q}$-homology plane has smooth locus of negative Kodaira dimension, then it is affine-ruled or isomorphic to $\mathbb{C}^2/G$ for some small finite, noncyclic subgroup $G < \mathrm{GL}(2,\mathbb{C})$. The surfaces $\mathbb{C}^2/G$ and $\mathbb{C}^2/G'$ are isomorphic if and only if G and G' are conjugate in* GL(2, $\mathbb{C}$). *The minimal normal completion of $\mathbb{C}^2/G$ is unique and the boundary is a nonadmissible rational fork with admissible twigs.*

Proof. For the first part of the statement, we follow the arguments of [Koras and Russell 2007, §3]. Assume that S' is not affine-ruled. Then S_0 is not affine-ruled. Since S' is affine, $D+\widehat{E}$ is not negative definite, so by [Miyanishi 2001, 2.5.1], S_0 contains a platonically $\mathbb{C}^*$-fibered open subset U, which is its almost minimal model. Also, $\chi(U) \le \chi(S_0)$ (see [Palka 2011a, 2.8]). The algorithm of construction of an almost minimal model [Miyanishi 2001, 2.3.8, 2.3.11] implies that $S_0 - U$ is a disjoint sum of s curves isomorphic to $\mathbb{C}$ and s' curves isomorphic to $\mathbb{C}^*$, for some $s, s' \in \mathbb{N}$. It follows that

$$0 = \chi(U) = \chi(S_0) - s = \chi(S') - q - s = 1 - q - s,$$

so $s = 0$, $q = 1$, and $s' \le 1$. If $s' \ne 0$, then the boundary divisor of U is connected, and hence U and S_0 are affine-ruled. Thus $s' = 0$ and $S_0 = U$, and by [Miyanishi and Tsunoda 1984], $S' \cong \mathbb{C}^2/G$, where G is a small finite noncyclic subgroup of GL(2, $\mathbb{C}$).

Suppose G and G' are two subgroups of GL(2, $\mathbb{C}$) such that $\mathbb{C}^2/G \cong \mathbb{C}^2/G'$. Then $\widehat{\mathcal{O}}_{\mathbb{C}^2/G,(0)} \cong \widehat{\mathcal{O}}_{\mathbb{C}^2/G',(0)}$, so if G and G' are small then they are conjugate, by [Prill 1967, Theorem 2]. The $\mathbb{C}^*$-ruling of S_0 does not extend to a ruling of S', so by [Palka 2008, 4.5], its boundary is a rational fork with admissible maximal twigs and its minimal normal completion is unique up to isomorphism. (For the description of the boundary, one could also use a more general result [Miyanishi 2001, 2.5.2.14].) □

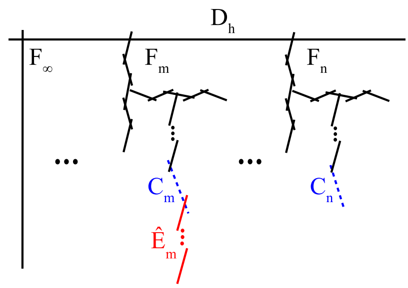


Figure 1. Affine-ruled S'.

3A. *Affine-ruled planes.* By Proposition 3.1, we may assume that S' is affine-ruled. This gives an affine ruling of S_0. We assume that $(\bar{S}, D+\widehat{E}, p)$ is a minimal completion of the latter. This weakens our initial snc-minimality assumption on D; that is, D is now p-minimal, but the unique section contained in D may be a nonbranching (-1)-curve. The base of p is rational because it is isomorphic to a section contained in $D+\widehat{E}$.

Lemma 3.2. *If S' is affine-ruled, then there exists exactly one fiber of p contained in D* (see Figure 1). *Each other singular fiber has a unique* (-1)*-curve, which is an S_0-component. The singularities of S' are cyclic.*

Proof. We have $\Sigma_{S_0} = \nu - 1$ and $\nu \le 1$ by Proposition 2.11, so $\Sigma_{S_0} = 0$ and there is exactly one fiber F_∞ contained in D. The fiber is smooth by the p-minimality of D. Each singular fiber F of p contains exactly one (-1)-curve. Indeed, if $D_0 \subseteq D$ is a vertical (-1)-curve, then by the p-minimality of D, it meets D_h and two D-components, so $\mu(D_0) > 1$. This is impossible because $D_h \cdot F = 1$. The (-1)-curve, say C, has $\mu(C) > 1$ and is the unique S_0-component of F. There are exactly two components of multiplicity one in F; they are tips of F and D_h intersects one of them. Thus the connected component of $\underline{F} - C$ not contained in D is a chain, so S' has only cyclic singularities. □

Remark. In Lemma 3.2, it was assumed (as in the whole paper — see Section 2D) that S' is logarithmic, but there is in fact no need for this. In any case $\widehat{E}$ is vertical, so it is a rational forest. Then D is a rational tree, and $\bar{S}$ and the base of p are rational by [Palka 2008, 3.4(i)]. The rest of the argument goes through.

Construction 3.3. Let $\mathbb{F}_1 = \mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1))$ be the first Hirzebruch surface with a (unique) projection $\tilde{p}\colon \mathbb{F}_1 \to \mathbb{P}^1$. Denote the section coming from the inclusion of the first summand by D'_h; then $D'^2_h = -1$. Choose $n+1$ distinct points $x_\infty, x_1, \dots, x_n \in D'_h$, and let F_∞ be the fiber containing x_∞. For each $i = 1, \dots, n$ starting from a blow-up of x_i, create a fiber F_i over $\tilde{p}(x_i)$ containing a unique (-1)-curve C_i. Let D_i be the connected component of $\underline{F_i} - C_i$ intersecting D_h, the proper transform of D'_h. By renumbering, we may assume there is $m \le n$ such that C_i is a tip of F_i if and only if $i > m$. Assume also that $m \ge 1$ (for $m = 0$ we would get a smooth surface). For $i \le m$, put $\widehat{E}_i = F_i - D_i - C_i$. Clearly, each $\widehat{E}_i$ is a chain. Let $\bar{S}$ be the resulting surface and let $p \colon \bar{S} \to \mathbb{P}^1$ be the induced $\mathbb{P}^1$-ruling. Put $D = F_\infty + D_h + \sum_{i=1}^{n} D_i$, $S = \bar{S} - D$ and $\widehat{E} = \sum_{i=1}^{m} \widehat{E}_i$. We define $\epsilon \colon S \to S'$ as the morphism contracting $\widehat{E}_i$'s.

Remark 3.4. Let $p \colon \bar{S} \to \mathbb{P}^1$ be as in 3.3, and for a fiber F denote the greatest common divisor of multiplicities of all S-components of F by $\mu_S(F)$. By Proposition 2.11, we have $H_1(S', \mathbb{Z}) = H_1(S, \mathbb{Z})$. By [Fujita 1982, 4.19, 5.9],

$$H_1(S, \mathbb{Z}) = \bigoplus_{i=1}^{n} \mathbb{Z}_{\mu_S(F_i)},$$

so $H_1(S', \mathbb{Z})$ can be any finite abelian group. It is easy to see that $\mu_S(F_i) = \mu(C_i)/d(\widehat{E}_i)$, where $d(\widehat{E}_i) = d(0) = 1$ if $i > m$. In particular, S' is a $\mathbb{Z}$-homology plane if and only if $m = n$ and each F_i is a chain. In fact in the latter case $\pi_1(S)$ vanishes and so S' is contractible.

Theorem 3.5. *The surface S' in Construction 3.3 is an affine-ruled singular $\mathbb{Q}$-homology plane. Conversely, each singular $\mathbb{Q}$-homology plane admitting an affine ruling can be obtained by Construction 3.3. Its strongly balanced boundary is unique if it is branched and is unique up to reversion if it is a chain. The affine ruling of S' is unique if and only if its strongly balanced boundary is not a chain.*

Proof. By definition, $\widehat{E}_i$'s are admissible chains, so S' is normal and has only cyclic singularities. We have $d(D) = -\prod_i d(D_i)$ [Koras and Russell 1999, 2.1.1], so $d(D) \neq 0$, and hence S' is a singular $\mathbb{Q}$-homology plane by Lemma 2.12. The last part of the statement almost follows from Lemma 3.2. It remains to note that by a flow (see Example 2.6), we can freely change the self-intersection of the horizontal boundary component without changing the rest of D, so we can assume that the construction starts with a negative section on $\mathbb{F}_1$. (We could, for instance, start with D'_h equal to the negative section on $\mathbb{F}_n$, so that the resulting boundary would be strongly balanced; see Definition 2.10). The uniqueness of a strongly balanced boundary follows from Proposition 2.8.

We now consider the uniqueness of an affine ruling. Let (V_i, D_i, p_i) be two minimal completions of two affine rulings of S' (see Section 2B). By Lemma 3.2, both D_i contain a 0-curve $F_{\infty,i}$ as a tip. By flows with supports in $F_{\infty,i}$, we may assume both D_i are standard (see Definition 2.9).

Assume that D_1 is not a chain. Then D_1 and D_2 are isomorphic as weighted curves (see Proposition 2.8). Let T_i be the unique maximal twig of D_i containing a 0-curve. Then either $T_i = F_{\infty,i} = [0]$, or we can write $T_i = [0, 0, a_1, \dots, a_n]$ with $[a_1, \dots, a_n]$ admissible. Then there is a flow $\Phi : (V_1, D_1) \rightsquigarrow (V_2, D_2)$ by

Proposition 2.8. Because D_1 is branched, $\operatorname{Supp}\Phi^\bullet \subseteq T_1$. Also, it follows from Proposition 2.8 and Example 2.6 that $\operatorname{Supp}\Phi^\bullet \subseteq F_{\infty,i}$. For $i = 1, 2$, let f_i be some fiber of p_i other than $F_{\infty,i}$. Since $\Phi^\bullet(f_1)$ is disjoint from $F_{\infty,2}$, we get $\Phi^\bullet(f_1)\cdot f_2 = 0$, so p_1 and p_2 agree on S'.

Let (V_1, D_1) be a standard completion of S' with $D_1 = [0, 0, a_1, \dots, a_n]$. We may assume that $[a_1, \dots, a_n]$ is admissible and nonempty; if it is empty, then $S' \cong \mathbb{C}^2$ is smooth, and if it is nonadmissible, then by the Hodge index theorem we necessarily have $D_1 = [0, 0, 0]$, which disagrees with Proposition 2.11(i). Let (V_2, D_2) be another completion of S', with D_2 being a reversion of D_1. The 0-tip T_i of each D_i induces an affine ruling on S'. Let (V, D) be a minimal normal pair dominating both (V_i, D_i), such that both affine rulings extend to $\mathbb{P}^1$-rulings of V. We argue that these affine rulings are different by proving that $\sigma_1^*T_1\cdot\sigma_2^*T_2 \neq 0$, where $\sigma_i : (V, D) \to (V_i, D_i)$ are the dominations. Suppose $\sigma_1^*T_1\cdot\sigma_2^*T_2 = 0$. Let H be an ample divisor on V and let $(\lambda_1, \lambda_2) \neq (0, 0)$ be such that $\widetilde{T}\cdot H = 0$ for $\widetilde{T} = \lambda_1\sigma_1^*T_1 + \lambda_2\sigma_2^*T_2$. We have $(\sigma_i^*T_i)^2 = T_i^2 = 0$, so

$$\widetilde{T}^2 = 2\lambda_1\lambda_2\sigma_1^*T_1\cdot\sigma_2^*T_2 = 0,$$

and hence $\widetilde{T} \equiv 0$ by the Hodge index theorem. But D has a nondegenerate intersection matrix, because $d(D) = d(D_1) \neq 0$, so $\widetilde{T}$ is a zero divisor. Then $\sigma_1^*T_1 = [0]$, for otherwise $\sigma_1^*T_1$ and $\sigma_2^*T_2$ would contain a common (-1)-curve, which contradicts the minimality of (V, D). It follows that σ_1 (and σ_2) are identities. This contradicts the fact that the reversion for nonempty $[a_1, \dots, a_n]$ is a nontrivial transformation of the completion (even if $[a_1, \dots, a_n]^t = [a_1, \dots, a_n]$). □

The following example shows that even if the strongly balanced boundary is unique, there might be infinitely many strongly balanced completions.

Example 3.6. Let (V, D, ι) be an snc-minimal completion (ι is the embedding; see Section 2A) of an affine-ruled singular $\mathbb{Q}$-homology plane S' as above. Assume that D_h is branched and that $D_h^2 = -1$. The only change of D that can be made by a flow is a change of the weight of D_h. If we now make an elementary transformation $(V, D) \mapsto (V_x, D_x)$ with a center $x \in F_\infty \setminus D_h$, then D becomes strongly balanced (see Definition 2.10). Denote the resulting completion by (V_x, D_x, ι_x) and let $F_{\infty,x}$ be the new fiber at infinity. The isomorphism type of the weighted boundary D_x does not depend on x, but for different x the completions (triples) are clearly different. In general, even the isomorphism type of the pair (V_x, D_x) depends on x. To see this, let $(V_x, D_x) \cong (V_y, D_y)$. Because the isomorphism maps $F_{\infty,x}$ to $F_{\infty,y}$, we get an automorphism of (V, D) mapping x to y. Taking a minimal resolution $\overline{S} \to V$, contracting all singular fibers to smooth fibers without touching D_h, and contracting D_h, we see that for $x \neq y$, this automorphism descends to a nontrivial automorphism of $\mathbb{P}^2$ fixing points that are images of contracted S_0-components and of D_h. In general such an automorphism does not exist.

3B. *Moduli.* Repeating Construction 3.3 in a special case, we obtain arbitrarily high-dimensional families of nonisomorphic singular $\mathbb{Q}$-homology planes with negative Kodaira dimension of the smooth locus and the same homeomorphism type. Example 3.7 gives a proof of Theorem 1.2. For smooth $\mathbb{Q}$-homology planes, a similar example was considered in [Flenner and Zaĭdenberg 1994, 4.16].

Example 3.7. Put $m = 2$ and $n = N + 2$ for some $N > 0$, and let $\bar{S}$, D, $\widehat{E}$, etc. be created as in the construction above, so that $D_1 = [3]$, $D_2 = [2]$ and $D_i = [2, 2, 2]$ for $3 \le i \le n$. Then $\widehat{E}_1 = [2, 2]$ and $\widehat{E}_2 = [2]$ (see Figure 2).

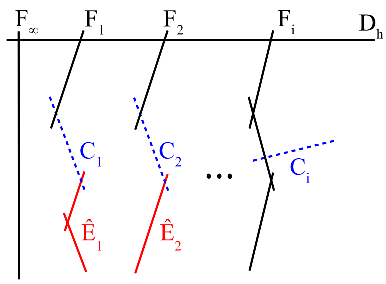


Figure 2. Singular fibers in Example 3.7.

Denoting the contraction of $\sum_{i=3}^{n} C_i$ by $\sigma : \bar{S} \to V$, we can factor the contraction $\bar{S} \to \mathbb{F}_1$ (which reverses the construction) as the composition $\bar{S} \xrightarrow{\sigma} V \xrightarrow{\sigma'} \mathbb{F}_1$. Put $y_i = \sigma(C_i)$ and $y = (y_3, \ldots, y_n)$. While σ'^{-1} is determined uniquely by the choice of $(x_1, \ldots, x_n)$, σ^{-1} and the resulting surface $\bar{S}$ (and hence S') can depend on the choice of y. Let us write $\bar{S}_y$ and S'_y to indicate this dependence. For $3 \le i \le n$, let D_i^0 be the open subset of the middle component of D_i remaining after subtracting two points belonging to other components of D_i. Put

$$U = D_4^0 \times \cdots \times D_n^0 \cong \mathbb{C}^{N-1}.$$

The family

$$\{S'_y\}_{y \in D_3^0 \times U} \to D_3^0 \times U$$

is N-dimensional. Since there is a compactly supported autodiffeomorphism of the pair $(\mathbb{C}^2, \mathbb{C}^* \times \{0\})$ mapping $(p, 0)$ to $(q, 0)$ for any $p, q \neq 0$, the choice of $y \in D_3^0 \times U$ is unique up to a diffeomorphism fixing irreducible components of $\sigma_*(D + \widehat{E} + C_1 + C_2)$. Thus all S'_y are homeomorphic.

Let $\pi : \mathfrak{X} \to U$ be the subfamily over $\{y_3^0\} \times U$. We show that the fibers of π are nonisomorphic. Suppose that $S'_y \cong S'_z$ for $y, z \in \{y_3^0\} \times U$. The isomorphism extends to snc-minimal resolutions. There is a flow $\Phi^\bullet : \bar{S}_y \dashrightarrow \bar{S}_z$ by Proposition 2.8, which is an isomorphism outside F_∞. Clearly, $\Phi^\bullet$ fixes $D_h \setminus \{x_\infty\}$, F_1 and F_2, and hence restricts to an identity on $D_h \setminus \{x_\infty\}$ and respects fibers. Since the C_i are unique (-1)-curves of the fibers, they are fixed by $\Phi^\bullet$. Therefore $\Phi^\bullet_{|\bar{S}-F_\infty-D_h}$ descends to an automorphism Φ_V of $V - F_\infty - D_h$ fixing the fibers, such that

$$\Phi_V(y_i) = z_i.$$

Also, Φ_V descends to an automorphism $\Phi_{\mathbb{F}_1}$ of $\mathbb{F}_1 - F_\infty - D'_h$ fixing fibers. If (x, y) are coordinates on $\mathbb{F}_1 - F_\infty - D'_h \cong \mathbb{C}^2$ such that x is a fiber coordinate, then

$$\Phi_{\mathbb{F}_1}(x, y) = (x, \lambda y + P(x))$$

for some $P \in \mathbb{C}[x]$ and $\lambda \in \mathbb{C}$. Introducing successive affine maps for the blow-ups, one can check that in some coordinates Φ_V acts on D_i^0 as $t \to \lambda^{\mu(C_i)} t$. Now the requirement $y_3 = y_3^0$ fixes $\lambda^2 = 1$, so because $\mu(C_i) = 2$ for each $3 \le i \le n$, we get that $y = z$.

Remark. By [Fujita 1982, 4.19 and 5.9], for S' as above, $\pi_1(S')$ is the N-fold free product of $\mathbb{Z}_2$. It follows from Remark 3.4 that given a weighted boundary, there exist only finitely many affine-ruled singular $\mathbb{Z}$-homology planes with this boundary. That is why in Example 3.7 we use branched fibers F_i for $3 \le i \le n$; so that the resulting surfaces are $\mathbb{Q}$-, but not $\mathbb{Z}$-homology planes.

4. $\mathbb{C}^*$-ruled $\mathbb{Q}$-homology planes

By [Palka 2008, 1.1(2) and 1.2] and Section 3A, to finish the classification of singular $\mathbb{Q}$-homology planes with smooth locus of nongeneral type, one needs to classify $\mathbb{Q}$-homology planes that are $\mathbb{C}^*$-ruled. Therefore, we assume here that S' is $\mathbb{C}^*$-ruled (and logarithmic; see Section 2D). The first homology group of S' and some necessary conditions for singular fibers of such rulings are analyzed in [Miyanishi and Sugie 1991, 2.9 and 2.10]. As before, we concentrate on completions rather than the affine part itself, because this gives more information and allows us to give a general method of construction. It also allows us to compute the number of different $\mathbb{C}^*$-rulings, and as a consequence the number of affine lines on S'.

4A. ***Properties of*** $\mathbb{C}^*$***-rulings.*** We can lift the $\mathbb{C}^*$-ruling of S' to a $\mathbb{C}^*$-ruling of the resolution and extend it to a $\mathbb{P}^1$-ruling $p\colon \bar{S} \to \mathbb{P}^1$ of a smooth completion. Assume that $D + \widehat{E}$ is p-minimal. By Proposition 2.11(v), $\Sigma_{S_0} = h + \nu - 2$ and $\nu \le 1$, so $(h, \nu, \Sigma_{S_0}) = (1, 1, 0)$, $(2, 1, 1)$ or $(2, 0, 0)$. The original $\mathbb{C}^*$-ruling of S' is twisted with base $\mathbb{C}^1$ in the first case, untwisted with base $\mathbb{C}^1$ in the second case, and untwisted with base $\mathbb{P}^1$ in the third case.

Lemma 4.1. *Denote by $F_1, \dots, F_n$ all the columnar fibers of $p : \bar{S} \to \mathbb{P}^1$ (see Definition 2.3). Let F_∞ be the fiber contained in D if $\nu = 1$. There is exactly one more singular fiber F_0; it contains $\widehat{E}$. Moreover*:

(i) *If* $(h, \nu) = (1, 1)$, *then* $F_\infty = [2, 1, 2]$, $\sigma(F_0) = 1$, *and* F_0 *and* F_∞ *contain branching points of* $p_{|D_h}$.

(ii) *If* $(h, \nu) = (2, 1)$, *then* F_∞ *is smooth and* $\sigma(F_0) = 2$.

(iii) *If* $(h, \nu) = (2, 0)$, *then* $\sigma(F_0) = 1$ *and* F_0 *contains a* D*-rivet.*

(iv) *If* $h = 2$, *then the components of* D_h *are disjoint.*

Proof. Let $(h, \nu) = (1, 1)$. Then $\Sigma_{S_0} = 0$, so by [Fujita 1982, 7.6], every singular fiber other than F_∞ is either columnar or contains a branching point of $p_{|D_h}$. Now D_h is rational and $p_{|D_h}$ has two branching points, one of them contained in F_∞, because D is a tree. Thus F_0 is unique. The p-minimality of D implies that $F_\infty = [2, 1, 2]$. Now let $h = 2$. We have $\Sigma_{S_0} = \nu \in \{0, 1\}$, and the p-minimality of D gives (ii), (iii) and the uniqueness of F_0. Suppose the components of D_h have a common point. D is a tree, so in this case $\nu = 0$, which gives $\sigma(F_0) = 1$. Because D is a simple normal crossing divisor, the common point belongs to the unique S_0-component of F_0, which therefore has multiplicity one. The connectedness of D implies that F_0 contains no D-components. But then F_0 has a unique (-1)-curve, which is impossible by Lemma 2.2. □

Lemma 4.1 is essentially [Miyanishi and Sugie 1991, 2.10]. While the conditions stated above are necessary, they are not sufficient. In the following examples the $\mathbb{C}^*$-ruling satisfies them, but the $\mathbb{C}^*$-ruled surface one obtains is not a $\mathbb{Q}$-homology plane.

Example 4.2. For $n \geq 0$, let $\mathbb{F}_n$ be the n-th Hirzebruch surface, and let D_0, D_∞ be sections with $D_0^2 = n$ and $D_\infty^2 = -n$. Let F_∞ be a fiber and put $D = D_0 + D_\infty + F_\infty$. Pick a point not belonging to D and make a connected sequence of blow-ups over it. Let C_0 be the unique (-1)-curve in the inverse image of the point, and let F_0 and C_1 be the reduced total and the proper transform of the fiber. Denote the resulting surface by $\overline{S}$, put $S = \overline{S} - D$ and $\widehat{E} = F_0 - C_0 - C_1$, and let $S \to S'$ be the morphism contracting $\widehat{E}$. In particular, $\widehat{E}$ might be any admissible chain, in which case S' has a unique cyclic singular point. S' is not a $\mathbb{Q}$-homology plane because $d(D) = 0$; see Lemma 2.12(iv).

Example 4.3. Take the pair $(\mathbb{F}_1, D_0 + D_\infty)$, where $\mathbb{F}_1$ is the first Hirzebruch surface and D_0 and D_∞ are sections with $D_0^2 = 1$ and $D_\infty^2 = -1$. Pick two points on D_0 and blow up over it to create two singular fibers $F_1 = [2, 1, 2]$, $F_2 = [2, 1, 2]$. Denote their (-1)-curves by C_1, C_2. These (-1)-curves separate two chains $T_0 = [2, 1, 2]$ and $T_\infty = [2, 1, 2]$, where the middle (-1)-curves are D_0 and D_∞, respectively. We have $d(T_0) = d(T_\infty) = 0$. Pick a point on some C_i, say C_1, that does not belong to $T_0 + T_\infty$, and make a connected sequence of blow-ups over it. Let C_0 be the unique (-1)-curve in the inverse image of the point, and let F_0 be the total reduced transform of the fiber. Denote the resulting complete surface by $\overline{S}$. If C_0 is not a tip of F_0, then denote the connected component of $F_0 - C_0$ not meeting $D_0 + D_\infty$ by $\widehat{E}$. Let D be the reduced divisor with support $T_0 \cup T_\infty \cup (F_0 - C_0 - \widehat{E})$. Put $S = \bar{S} - D$ and $\widehat{E} = F_0 - C_0 - C_1$, and let $S \to S'$ be the morphism contracting $\widehat{E}$ (which is necessarily an admissible chain). Again, S' is not a $\mathbb{Q}$-homology plane because $d(D) = 0$.

Theoretically, if X is a normal surface and $p' : X \to B$ is a $\mathbb{C}^*$-ruling, then by taking a completion of X and an extension of p' to a $\mathbb{P}^1$-ruling, with Lemma 2.12 we can recognize when X is a $\mathbb{Q}$-homology plane (B has to be rational). However, to give constructions we need to reformulate the condition $d(D) \neq 0$ in a way that is easier to verify by looking at the geometry of singular fibers. Recall that for a family of subsets $(A_i)_{i \in I}$ of a topological space Y, a subset $X \subseteq Y$ *separates* the subsets $(A_i)_{i \in I}$ (inside Y) if and only if each A_i is contained in a closure of some connected component of $Y \setminus X$ and none of these closures contains more than one A_i. Recall also that by convention, a twig of a fixed divisor is ordered so that its tip is the first component.

Lemma 4.4. *Let* $(\bar{S}, T, p)$ *be a triple satisfying conditions* (i)–(iii) *of Lemma 2.12. Assume also that* T *is* p*-minimal and that* $f \cdot T = 2$ *for a general fiber* f *of* p. *When* $(h, \nu) = (2, 0)$, *let* D_0 *be some horizontal component of* D, *let* F_0 *be a unique fiber containing a* D*-rivet, let* B *be a unique component of* D *separating* D_0, $D_h - D_0$ *and* $\widehat{E}$ *inside* $D \cup F_0$, *and let* $\widetilde{D}_0$ *be a connected component of* $D - B$ *containing* D_0. *Then* $d(D) \neq 0$ *if and only if the following conditions hold*:

(i) *The base of the fibration is* $\mathbb{P}^1$ *or* $\mathbb{C}^1$ (*that is*, $\nu \leq 1$).

(ii) *If* $(h, \nu) = (2, 1)$, *both* $\bar{S} - T$*-components of the fiber with* $\sigma = 2$ *intersect* D.

(iii) *If* $(h, \nu) = (2, 0)$, *then* $d(\widetilde{D}_0) \neq 0$.

The advantage of condition (iii) over $d(D) \neq 0$ is that $\widetilde{D}_0$ is simpler than D, containing at most one branching component.

Proof. Clearly, if $d(D) \neq 0$, then S' is a $\mathbb{Q}$-homology plane by Lemma 2.12, which implies (i) and (ii) (D meets each curve not contained in $D + \widehat{E}$ because S' is affine). Suppose now that (i) and (ii) are satisfied. We show that $d(D) \neq 0$ is equivalent to (iii) (which is an empty condition if $(h, \nu) \neq (2, 0)$). Note that $d(D) \neq 0$ is equivalent to $d(T) \neq 0$, because $T - D$ is negative definite.

Consider the case $h = 1$. We have $\Sigma_{\bar{S}-T} = \nu - 1$, and hence $\nu = 1$ and $\Sigma = 0$. The horizontal component D_h meets the unique fiber F_∞ contained in T in one point, because T is a forest. Let T_∞ be the component meeting D_h. We have $d(F_\infty) = 0$, so by [Koras and Russell 1999, 2.1.1(i)],

$$d(D) = d(F_\infty)d(D - F_\infty) - d(F_\infty - T_\infty)d(D - F_\infty - D_h),$$

and we obtain

$$d(D) = -d(F_\infty - T_\infty)d(D - F_\infty - D_h).$$

Since $F_\infty - T_\infty$ and $D - F_\infty - D_h$ are vertical and do not contain whole fibers, they are negative definite, and hence $d(D) < 0$.

We may now assume $h = 2$. Then $\Sigma = \nu \in \{0, 1\}$. Put $\widehat{E} = T - D$. When $\nu = 1$, let F_∞ be the unique fiber contained in D, and let F_0 be the unique singular fiber with $\sigma(F_0) = 2$. When $\nu = 0$, let F_0 be the unique fiber containing a D-rivet. All other singular fibers are columnar by [Fujita 1982, 7.6], so they contain no components of $\widehat{E}$. We need to prepare some tools to proceed. Recall that the Neron–Severi group of $\bar{S} - T$ is defined as the quotient of $\mathrm{NS}(\bar{S})$ by the subgroup generated by components of T. We put $\rho(\bar{S} - T) = \dim \mathrm{NS}(\bar{S} - T) \otimes \mathbb{Q}$.

Let (X, R) be a smooth pair with X rational. Suppose $R = R_1 + R_2$, where R_1 and R_2 meet in unique components $C_1 \subseteq R_1$, $C_2 \subseteq R_2$ respectively. If at least one of R_i is negative definite for $i = 1, 2$, then we call $R - C_1$ a *swap* of $R - C_2$ and vice versa. Similarly, $(X, R - C_i)$ are by definition swaps of each other, and so are $X - (R - C_i)$, for $i = 1, 2$. The basic property of this operation that we need is that

$$\rho(X - (R - C_1)) = \rho(X - (R - C_2)).$$

To see this, it is enough to show that C_1, C_2 do not belong to the subspace V of $\mathrm{NS}(X) \otimes \mathbb{Q}$ generated by components of $R_1 - C_1 + R_2 - C_2$. By symmetry, we can assume that R_2 is negative definite. Suppose that $C_1 \in V$ and write

$$C_1 \equiv U_1 + U_2,$$

where U_i is in the subspace generated by components of $R_i - C_i$. Then $0 = C_1 \cdot U_2 = U_1 \cdot U_2 + U_2^2 = U_2^2$, and hence $U_2 \equiv 0$ by the negative definiteness of R_2. Then $0 < C_1 \cdot C_2 = U_1 \cdot C_2 = 0$, a contradiction. Suppose $C_2 \in V$ and write $C_2 \equiv U_1 + U_2$ as above. Then $(C_2 - U_2)^2 = (C_2 - U_2) \cdot U_1 = 0$, so $C_2 \equiv U_2$ by the negative definiteness of R_2. Then $0 < C_1 \cdot C_2 = C_1 \cdot U_2 = 0$, a contradiction. Thus, swapping preserves ρ. Though the definition is of general use, we use only a special kind of swapping, when C_2 is a (-1)-curve and it is absorbed into the boundary (keeping the assumption that R_2 is negative definite); that is, we do the swap one way, changing $(X, R - C_2)$ to $(X, R - C_1)$.

Take $(\bar{S}, T)$ and interchangeably perform contractions of (-1)-curves in F_0 (and its images) that are nonbranching components of the boundary and swaps absorbing vertical (-1)-curves in F_0 (and its images) into the boundary. Denote the resulting smooth pair by (X, T'). By the properties of swaps and blow-ups, the rank of the Neron–Severi group of the open part and the difference between b_2 of the complete surface and the number of components in the boundary remain constant. Also, T' is a rational forest. Crucially, $d(T) = 0$ if and only if $d(T') = 0$. To see this, we may assume that (X, T') is simply a swap of $(\bar{S}, T)$ as above. Since the number of components of T equals $b_2(\bar{S})$, we know $d(T) \neq 0$ if and only if $\rho(\bar{S} - T) = 0$, which is equivalent to $\rho(X - T') = 0$ and then to $d(T') \neq 0$.

Consider the case $\Sigma = \nu = 0$. At some point, the process of swapping and contracting makes B into a 0-curve or a (-1)-curve. It is easy to see that the divisor $\widetilde{D}_0 + \widetilde{D}_\infty$ is not affected by the process, so we have $d(D) \neq 0$ if and only if $d(\widetilde{D}_0) \cdot d(\widetilde{D}_\infty) \neq 0$. All singular fibers of the induced $\mathbb{P}^1$-ruling at this stage are columnar, so they can be written as $R_{i,0} + C_i + R_{i,\infty}$, where $i = 1, \dots, n'$ enumerates these fibers, C_i^2 equals -1, and $R_{i,0}$ and $R_{i,\infty}$ are chains whose last components meet D_0 and D_∞, respectively. For $j = 0, \infty$, put $\tilde{e}_j = \tilde{e}(\widetilde{D}_j)$ (see Section 2A). Then $\tilde{e}_j = \sum_i \tilde{e}(R_{i,j})$. We have $d(\widetilde{D}_j) = (-D_j^2 - \tilde{e}_j) \cdot \prod_i d(R_{i,j})$. By the properties of columnar fibers,

$$d(\widetilde{D}_0) + d(\widetilde{D}_\infty) = -(D_0^2 + D_\infty^2 + n') \cdot \prod_i d(R_{i,0}).$$

When contracting singular fibers to smooth ones, $D_0 + D_\infty$ is touched n' times and its image consists of two disjoint sections on a Hirzebruch surface. It follows that $D_0^2 + D_\infty^2 + n' = 0$, and hence $d(\widetilde{D}_\infty) + d(\widetilde{D}_0) = 0$. Thus $d(D) \neq 0$ if and only if $d(\widetilde{D}_0) \neq 0$.

Consider the case $\Sigma = \nu = 1$. We show that T' has at most one horizontal component. Suppose that it has two. Then $\sigma(\widetilde{F}_0) = \sigma(F_0) = 2$, so $\widetilde{F}_0$ contains a (-1)-curve, say C_1. Because T' is p-minimal, $C_1 \not\subseteq T$. Because we assumed that every $\bar{S} - T$-component meets D, by the properties of swaps, every $X - T'$-component meets T'. By the definition of X, absorbing the (-1)-curve by a swap into the boundary is impossible. In particular, if $\widetilde{F}_0$ has no more (-1)-curves, then C_1 is not a tip of $\widetilde{F}_0$, so $\widetilde{F}_0$ is a chain. However, since $\sigma(\widetilde{F}_0) = 2$, a swap absorbing C_1 into the boundary is possible, which is a contradiction. Thus, $\widetilde{F}_0$ has two (-1)-curves, C_1 and C_2. One of them meets some horizontal component of T'; otherwise, either C_1 or C_2 is a tip or $\widetilde{F}_0 \cap T'$ has three connected components, and in either case a swap absorbing one of the C_i's into the boundary would be possible. A similar argument shows that the second (-1)-curve also meets a horizontal component of T'. Thus, $\widetilde{F}'_0$ is a chain with C_1 and C_2 as tips, and again a swap is possible, a contradiction. So T' has at most one horizontal component. But after the first swap where σ of the image of F_0 drops, the fiber has only one (-1)-curve, which therefore has multiplicity greater than one, so no more swaps of this kind are possible. Thus, T' has a unique horizontal component T'_h. Then

$$d(T') = d(F_\infty) d(T' - F_\infty) - d(T' - F_\infty - D_\infty) = -d(T' - F_\infty - D_\infty).$$

Now $T' - F_\infty - D_\infty$ is vertical and does not contain whole fibers, so it is negative definite and we obtain $d(T') = d(T' - F_\infty - D_\infty) \neq 0$. □

Remark. By Proposition 2.11, for any $\mathbb{Q}$-homology plane, we have $H_i(S', \mathbb{Z}) = 0$ for $i > 1$ and

$$|H_1(S', \mathbb{Z})|^2 = \frac{d(D)}{d(\widehat{E})},$$

and hence S' is a $\mathbb{Z}$-homology plane if and only if $d(D) = d(\widehat{E})$. For a $\mathbb{C}^*$-ruled S', more explicit computations are done in [Miyanishi and Sugie 1991], which we do not repeat here. For example, by [ibid., 2.17], if a $\mathbb{Z}$-homology plane with $\bar{\kappa}(S_0) \neq -\infty$ is $\mathbb{C}^*$-ruled, then $\bar{\kappa}(S_0) = 1$ and the ruling is untwisted with base $\mathbb{P}^1$. The conditions for S' having such a ruling to be contractible are given in [ibid., 2.11] (in particular $n = 2$).

4B. ***The Kodaira dimension.*** In [Miyanishi and Sugie 1991, 2.9–2.17] one can find formulas for the Kodaira dimension of the smooth locus $\bar{\kappa}(S_0)$ in terms of properties of singular fibers of the $\mathbb{C}^*$-ruling (there, $\bar{\kappa}(S')$ is by definition equal to $\bar{\kappa}(S_0)$). Unfortunately, their formulas 2.14(4), 2.15(2), and 2.16(2) are incorrect. The corrections require splitting into cases depending on additional properties of singular fibers. We also compute the Kodaira dimension of S'. We keep the notation for singular fibers as in Lemma 4.1. When $\nu = 0$, put $F_\infty = 0$. Let J be the reduced divisor with support equal to $D \cup F_0$. For $i = 1, \dots, n$, denote the (-1)-curve of the columnar fiber F_i by C_i and the multiplicity of C_i by μ_i. Put $J^+ = J + C_1 + \cdots + C_n$.

Lemma 4.5. *The divisor J^+ has simple normal crossings. Contract vertical (-1)-curves in J^+ and its images as long as the image is an snc-divisor. Let*

$$\zeta : (\overline{S}, J^+) \to (W, \zeta_* J^+)$$

be the composition of these contractions. Then the $\zeta_* F_i$ are smooth for $i = 1, \dots, n$; moreover:

(i) *If $h = 1$, then $\zeta_* F_0 = [2, 1, 2]$, $(\zeta_* D_h)^2 = 0$, and one can further contract $\zeta_* F_0$ and F_∞ to smooth fibers so that W maps to $\mathbb{F}_1$ and $\zeta_* D_h$ maps to a smooth 2-section of the $\mathbb{P}^1$-ruling of $\mathbb{F}_1$ disjoint from the negative section.*

(ii) *If $h = 2$, then $\zeta_* F_0$ is smooth, W is a Hirzebruch surface, and the components of $\zeta_* D_h$ are disjoint. Also, at least one of the components of D_h has negative self-intersection, and by changing ζ if necessary, one can assume that it is not affected by ζ.*

Proof. Suppose the crossings of J^+ at x are not simple normal. By Lemma 4.1, this only happens if $h = 2$. Also, x belongs to $D_h \cap F_0$ and is a branching point of $p_{|D_h}$, and two components of F_0 of multiplicity one meet at x. Because D has normal crossings, one of them is the unique S_0-component of F_0. By the p-minimality of D, it has to be a unique (-1)-curve of F_0 too, which is impossible by Lemma 2.2(i). Thus, J^+ is an snc-divisor. Because F_i for $i = 1, \dots, n$ are columnar, $\zeta_* F_i$ are smooth.

Suppose $h = 2$. Write $D_h = H + H'$. By Lemma 4.1, H and H' are disjoint. Since H and H' meet F_0 only in the components of multiplicity one, it follows from the definition of ζ that the images of H' and H intersect the same component of $\zeta_* F_0$. But this is possible only if $\zeta_* F_0$ is smooth. Since $\zeta_* J^+$ is snc, these images are disjoint. Say $H'^2 \le H^2$. Choosing the contracted (-1)-curves correctly, we may assume that H' is not affected by ζ. Since $\zeta_* D_h$ consists of two disjoint sections on a Hirzebruch surface, we have $(\zeta_* D_h)^2 = 0$, so $D_h^2 \le 0$. Suppose $H^2 = H'^2 = 0$. Then ζ does not affect D_h, so $n = 0$ and H and H' intersect the same component B of F_∞. If $\nu = 1$, then B is an S_0-component and the second S_0-component of F_0 does not intersect D, a contradiction with the affineness of S'. Thus $\nu = 0$ and Lemma 4.4 is not satisfied (in other words, $d(D) = 0$), a contradiction.

Suppose $h = 1$. By the definition of ζ, the image of D_h intersects the unique (-1)-curve of $\zeta_* F_0$. It follows that $\zeta_* F_0 = [2, 1, 2]$. Now after the contraction of F_0 and F_∞ to smooth fibers, the image of W is a Hirzebruch surface $\mathbb{F}_N$, where $N \ge 0$, and the image D'_h of D_h is a smooth 2-section. Write $D'_h \equiv \alpha f + 2H$, where H is a section with $H^2 = -N$ and f is a fiber of the induced $\mathbb{P}^1$-ruling of $\mathbb{F}_N$. We compute

$$p_a(\alpha f + 2H) = \alpha - N - 1,$$

so because D'_h is smooth, its arithmetic genus vanishes and $\alpha = N + 1$. Also, $D'_h \cdot H = \alpha - 2N$, and hence $D'_h \cdot H + N = 1$. Now if $N = 0$, then $\mathbb{F}_N = \mathbb{P}^1 \times \mathbb{P}^1$, and an elementary transformation with center equal to the point of tangency of D'_h and the image of F_∞ (which corresponds to a different choice of components to be contracted in F_∞) leads to $N = 1$ and $D'_h \cdot H = 0$. □

Remark 4.6. Let (X, D) be a smooth pair, and let L be the exceptional divisor of a blow-up $\sigma : X' \to X$ of a point in D. Then

$$K_{X'} + \sigma^{-1} D = \sigma^*(K_X + D)$$

if σ is subdivisional for D, and

$$K_{X'} + \sigma^{-1} D = \sigma^*(K_X + D) + L$$

if σ is sprouting for D.

Decompose ζ into a sequence of blow-downs $\zeta = \sigma_k \circ \cdots \circ \sigma_1$, and let $m \le k$ be the minimal number such that for $j > m$, the blow-up σ_j is subdivisional for $(\sigma_j \circ \cdots \circ \sigma_1)_* J^+$. Define $\eta\colon \bar{S} \to \widetilde{S}$ and $\theta\colon \widetilde{S} \to W$ as

$$\eta = \sigma_m \circ \cdots \circ \sigma_1 \quad \text{and} \quad \theta = \sigma_k \circ \cdots \circ \sigma_{m+1}.$$

Clearly, η is the identity outside F_0. We denote a general fiber of a $\mathbb{P}^1$-ruling by f.

Lemma 4.7. *Let $\eta : \bar{S} \to \widetilde{S}$ and $\theta : \widetilde{S} \to W$ be as above. Then*

$$K_{\widetilde{S}} + \eta_* J \equiv \Big(n + \nu - 1 - \sum_{i=1}^{n} \frac{1}{\mu_i}\Big) f + G + \theta^* \tfrac{1}{2}(U + U'),$$

where G is a negative definite effective divisor with support contained in the support of $F_\infty + \sum_{i=1}^{n} F_i$ and U, U' are the (-2)-tips of $\zeta_* F_0$ if p is twisted and are zero otherwise.

Proof. Let $V \subseteq W$ be defined as the sum of (four) (-2)-tips of $\underline{F_\infty} + \zeta_* \underline{F_0}$ if p is twisted and as zero otherwise. We check easily that

$$K_W + D_h + \underline{F_\infty} + \zeta_* \underline{F_0} \equiv (\nu - 1) f + \tfrac{1}{2} V.$$

Indeed, if p is untwisted, this is just $K_W + D_h + 2f \equiv 0$ on a Hirzebruch surface, and if p is twisted, then it follows from the numerical equivalences $K_W + D_h + f \equiv 0$ and $\underline{F_\infty} + \zeta_* \underline{F_0} - \frac{1}{2} V \equiv f$. By Remark 4.6,

$$K_{\widetilde{S}} + \eta_* J^+ \equiv (n + \nu - 1) f + \theta^* \tfrac{1}{2} V.$$

For every $i = 1, \ldots, n$, the divisor $G_i = (1/\mu_i) F_i - C_i$ is effective and negative definite because C_i is not contained in its support. We get

$$K_{\widetilde{S}} + \eta_* J \equiv (n + \nu - 1) f + \sum_{i=1}^{n} \Big(G_i - \frac{1}{\mu_i} F_i\Big) + \theta^* \tfrac{1}{2} V,$$

so

$$K_{\widetilde{S}} + \eta_* J \equiv \Big(n + \nu - 1 - \frac{1}{\mu_i}\Big) f + \sum_{i=1}^{n} G_i + \theta^* \tfrac{1}{2} V. \qquad \square$$

Remark 4.8. Because $K_{\bar{S}} + D + \widehat{E}$ and $K_{\bar{S}} + D$ intersect trivially with a general fiber, we can write $K_{\bar{S}} + D + \widehat{E} \equiv \kappa_0 f + G_0$ and $K_{\bar{S}} + D + \widehat{E} \equiv \kappa f + G$, where G_0 and G are some vertical effective and negative definite divisors and $\kappa_0, \kappa \in \mathbb{Q}$. It follows that $\bar{\kappa}(S_0)$ and $\bar{\kappa}(S)$ are determined by the signs of κ_0 and κ. More explicitly, $\bar{\kappa}(S_0)$ equals $-\infty$, 0, or 1 depending on whether $\kappa_0 < 0$, $\kappa_0 = 0$, or $\kappa_0 > 0$, respectively. An analogous statement holds for $\bar{\kappa}(S)$ and κ.

It turns out that κ and κ_0 depend in a quite involved way on the structure of F_0. This dependence can be stated in terms of the properties of $\eta : \bar{S} \to \widetilde{S}$ defined above. Denote the S_0-components of F_0 by C, $\widetilde{C}$ (or just C if there is only one) and their multiplicities by μ, $\tilde{\mu}$ respectively. Note that $\mu \geq 2$ if $\sigma(F_0) = 1$, but if $\sigma(F_0) = 2$, then it can happen that $\mu = 1$ or $\tilde{\mu} = 1$.

Theorem 4.9. *Let $\lambda = n+\nu-1-\sum_{i=1}^{n}(1/\mu_i)$. The numbers κ and κ_0 determining the Kodaira dimension of a $\mathbb{C}^*$-ruled singular $\mathbb{Q}$-homology plane S' and of its smooth locus S_0 defined in Remark 4.8 are as follows*:

(A) *Case* $(h, \nu) = (1, 1)$. *Denote the component of F_0 intersecting the* 2*-section contained in D by B.*

(i) *If* $\eta = \mathrm{id}$ *and* $F_0 = [2, 1, 2]$, *then* $\kappa = \kappa_0 = \lambda - \frac{1}{2}$.

(ii) *If* $\eta = \mathrm{id}$, *B is not a tip of F_0, and $C\cdot B > 0$, then* $(\kappa, \kappa_0) = (\lambda-\frac{1}{2}, \lambda-1/2\mu)$.

(iii) *If* $\eta = \mathrm{id}$, $C \cdot B = 0$, *and F_0 is a chain, then* $(\kappa, \kappa_0) = (\lambda - \frac{1}{2}, \lambda)$.

(iv) *If* $\eta = \mathrm{id}$ *and B is a tip of F_0, then* $(\kappa, \kappa_0) = (\lambda - \frac{1}{2}, \lambda - 1/\mu)$.

(v) *If* $\eta \neq \mathrm{id}$, *then* $\kappa = \kappa_0 = \lambda$.

(B) *Case* $(h, \nu) = (2, 1)$.

(i) *If* $\eta = \mathrm{id}$ *and* $C^2 = \widetilde{C}^2 = -1$, *then* $(\kappa, \kappa_0) = (\lambda - 1, \lambda - 1/\min(\mu, \tilde{\mu}))$.

(ii) *If* $\eta = \mathrm{id}$ *and* $C^2 \neq -1$ *or* $\widetilde{C}^2 \neq -1$, *then* $\kappa = \kappa_0 = \lambda - 1/\min(\mu, \tilde{\mu})$.

(iii) *If* $\eta \neq \mathrm{id}$, *then assuming that C is the S_0-component disjoint from $\widehat{E}$, we have* $\kappa = \kappa_0 = \lambda - 1/\mu$.

(C) *Case* $(h, \nu) = (2, 0)$. *Then* $\kappa = \kappa_0 = \lambda$.

Proof. (A) The unique S_0-component C of F_0 is a (-1)-curve. Otherwise, the p-minimality of D implies that B is the only (-1)-curve in F_0 and that it intersects two other D-components of F_0, giving $F_0 = [2, 1, 2] \subseteq D$, with no place for C. It is now easy to check that the list of cases in (A) is complete. Because $C^2 = -1$, $\underline{F_0} - C$ has at most two connected components. The only case when $\widehat{E}$ is not connected is when F_0 contains no D-components, which is only possible if $C = B$ and $F_0 = [2, 1, 2]$. Because C is the unique (-1)-curve in F_0, we know that $\zeta = \theta \circ \eta$ has at most one center on $\zeta_* F_0$, so by symmetry we can and do assume that it does not belong to U' (see Lemma 4.7). Suppose $\eta \neq \mathrm{id}$. The center of η belongs to a unique component of $\eta_* J$. Because D_h does not intersect components contracted by η, this component is a proper transform of a D-component, so $\eta_*(C + \widehat{E}) = 0$ by the connectedness of $\widehat{E}$. If we now factor η as $\eta = \sigma \circ \eta'$, where σ is a sprouting blow-up for $\eta_* J$, then by Lemma 4.7 and Remark 4.6,

$$K + \sigma^{-1}\eta_* J \equiv \lambda f + G + \sigma^*\theta^*\tfrac{1}{2}(U + U') + \mathrm{Exc}(\sigma),$$

where $\mathrm{Exc}(\sigma)$ is the exceptional (-1)-curve contracted by σ and K is a canonical divisor on a respective surface. Because $\eta_*(C + \widehat{E}) = 0$, each component of $C + \widehat{E}$ appears with positive integer coefficient in $\eta'^* \mathrm{Exc}(\sigma)$, which gives $K_{\bar{S}} + \eta^{-1}\eta_* J \equiv \lambda f + G + G_0$, where G_0 is a vertical effective and negative definite divisor for which $G_0 - \widehat{E} - C$ is still effective. Because $\eta^{-1}\eta_* J = J = D + \widehat{E} + C$, we get $\kappa = \kappa_0 = \lambda$. We can now assume that $\eta = \mathrm{id}$, so

$$K_{\bar{S}} + D + \widehat{E} + C \equiv \lambda f + G + \tfrac{1}{2}(U' + \theta^* U).$$

This can be written as

$$K_{\overline{S}}+D\equiv(\lambda-\tfrac{1}{2})f+G+\tfrac{1}{2}(U'+F_0+\theta^*U-2C-2\widehat{E}).$$

All components of F_0 appear in $U'+F_0+\theta^*U$ with coefficients bigger than 1, so $U'+F_0+\theta^*U-2C-2\widehat{E}$ is effective and negative definite, because its support does not contain the $\widehat{E}$-component that is a proper transform of U. This gives $\kappa=\lambda-\frac{1}{2}$. We now compute κ_0. If $F_0=[2,1,2]$, then $\theta^*U=U$ and $\widehat{E}=U+U'$, so $K_{\overline{S}}+D\equiv(\lambda-\frac{1}{2})f+G$ and we get $\kappa_0=\lambda-\frac{1}{2}$. Suppose B is a tip of F_0. Because $\mu(B)=2$, we know that F_0 is a fork with two (-2)-tips as maximal twigs (see Lemma 2.2(ii)) and that $\theta^*U=U$ (U and U' are components of $\widehat{E}$). The divisor $G_0=\frac{1}{2}(U+U')+(1/\mu)F_0-C$ is vertical effective and its support does not contain C. Writing

$$K_{\overline{S}}+D+\widehat{E}\equiv\Big(\lambda-\frac{1}{\mu}\Big)f+G+G_0,$$

we infer that $\kappa_0=\lambda-1/\mu$, and we obtain (iv). Consider the case (ii). Because B is not a tip of F_0, we know F_0 is a chain. The assumption $B\cdot C>0$ implies that $B^2\neq-1$ and $\theta^*U=C+\widehat{E}$. We obtain

$$K_{\overline{S}}+D+\widehat{E}\equiv\Big(\lambda-\frac{1}{2\mu}\Big)f+G+\tfrac{1}{2}\Big(U'+\widehat{E}+\frac{1}{\mu}F_0-C\Big),$$

and $U'+\widehat{E}+(1/\mu)F_0-C$ is effective with support not containing C. This gives $\kappa_0=\lambda-(1/2\mu)$. We are left with the case (iii). As in (ii), F_0 is a chain, and we have now

$$K_{\overline{S}}+D+\widehat{E}\equiv\lambda f+G+\tfrac{1}{2}(U'+\theta^*U-2C).$$

$U'+\theta^*U-2C$ is effective and does not contain B, because $B\cdot C=0$, so $\kappa_0=\lambda$.

(B) Suppose $\eta\neq\mathrm{id}$. Note that η_*F_0 contains a proper transform of one of C, $\widetilde{C}$, for otherwise, F_0 would contain a D-rivet. It follows that η is a connected modification and that its center lies on a birational transform of a D-component (the S_0-component contracted by η has to intersect D). Thus, η_*F_0 is a chain intersected by D_h in two different tips and containing C. Since $D\cap\widehat{E}=\varnothing$, we get $\eta_*(\widetilde{C}+\widehat{E})=0$. Writing $\eta=\sigma\circ\eta'$, where σ is a sprouting blow-down, we see that $\eta'^*\operatorname{Exc}(\sigma)$ is an effective negative definite divisor that does not contain C in its support and for which $\eta'^*\operatorname{Exc}(\sigma)-\widetilde{C}-\widehat{E}$ is effective. By Lemma 4.7, we have

$$K+\sigma^{-1}\eta_*D+C\equiv\lambda f+G+\operatorname{Exc}(\sigma),$$

where K is a canonical divisor on a respective surface. It follows from Remark 4.6 and from arguments analogous to those in part (A) that $\kappa=\kappa_0=\lambda-(1/\mu)$. We can now assume that $\eta=\mathrm{id}$. By Lemma 4.7,

$$K_{\overline{S}}+D+C+\widehat{E}+\widetilde{C}\equiv\lambda f+G,$$

which implies $\kappa_0 = \lambda - (1/\min(\mu, \tilde{\mu}))$. Writing

$$K_{\overline{S}} + D \equiv \Big(\lambda - \frac{1}{\alpha}\Big) f + G + \frac{1}{\alpha}\big(F_0 - \alpha(C + \widehat{E} + \widetilde{C})\big),$$

we see that $\kappa = \lambda - (1/\alpha)$, where α is the lowest multiplicity of a component of $C + \widehat{E} + \widetilde{C}$ in F_0. Note that $C + \widehat{E} + \widetilde{C}$ is a chain. Now if $C^2 \neq -1$, for instance, then F_0 is columnar, and factoring θ into blow-downs, we see that $\widehat{E}$ is contracted before C, and hence $\alpha = \mu \leq \tilde{\mu}$. Suppose $C^2 = \widetilde{C}^2 = -1$, and let θ' be the composition of successive contractions of (-1)-curves in F_0 different than C. Now either $\theta'_* F_0 = \theta'_* C = [0]$ or $\theta'_* F_0$ is columnar. Both possibilities imply that $C + \widehat{E}$ contains a component of multiplicity one, and hence $\alpha = 1$.

(C) C is a (-1)-curve. Indeed, $D \cap F_0$ contains at most one (-1)-curve, and if it does, then by the p-minimality of D, it meets both components of D_h and has multiplicity one, so there is another (-1)-curve in F_0. We infer that $\underline{F}_0 - C$ has two connected components, one being $\widehat{E}$ and the second containing a rivet. The existence of a rivet in F_0 implies that $\eta \neq \mathrm{id}$, so $\eta_*(C + \widehat{E}) = 0$. Factoring out a sprouting blow-down from η as above, we get

$$K + \sigma^{-1}\eta_* D \equiv \lambda f + G + \mathrm{Exc}(\sigma).$$

The divisor $\eta'^* \operatorname{Exc}(\sigma) - C - \widehat{E}$ is effective and does not contain all components of F_0, so by Remark 4.6, $\kappa = \kappa_0 = \lambda$. □

Remark. In case (B)(iii), it is not true in general that $\mu = \min(\mu, \tilde{\mu})$.

4C. ***Smooth locus of Kodaira dimension zero.*** As a corollary, we obtain the following information in case $\bar{\kappa}(S_0) = 0$.

Corollary 4.10. *Let S' be a $\mathbb{C}^*$-ruled singular $\mathbb{Q}$-homology plane, and let D be a p-minimal boundary for an extension p of this ruling to a normal completion, as above. Let D be the p-minimal boundary, and let n be the number of columnar fibers. Then $\bar{\kappa}(S_0) = 0$ exactly in the following cases*:

(i) *p is twisted, $n = 0$, and F_0 is of type* (A)(iii) *or* (A)(v).

(ii) *p is twisted, $n = 1$, $\mu = \mu_1 = 2$, and F_0 is of type* (A)(i) *or* (A)(iv) *with no D-components.*

(iii) *p is untwisted with base $\mathbb{C}^1$, $n = 1$, $\mu_1 = 2$, $\min(\mu, \tilde{\mu}) = 2$, and some connected component of $F_0 \cap D$ is a (-2)-curve.*

(iv) *p is untwisted with base $\mathbb{C}^1$, $n = 2$, $\mu_1 = \mu_2 = 2$, and some S_0-component of F_0 meets D_h.*

(v) *p is untwisted with base $\mathbb{P}^1$, $n = 2$, and $\mu_1 = \mu_2 = 2$.*

Proof. Note that $n - \sum_{i=1}^{n}(1/\mu_i) \geq n/2$ because $\mu_i \geq 2$ for each i. Suppose p is twisted. Then $\mu \geq 2$, and so by Theorem 4.9,

$$\lambda \geq \kappa_0 \geq \lambda - \tfrac{1}{2} \geq \frac{n-1}{2}.$$

If $n = 0$, then $\lambda = 0$, which gives $\kappa_0 = 0$ exactly in cases (A)(iii) and (A)(v). If $n = 1$, then $\kappa_0 = \lambda - \frac{1}{2} = 0$, which is possible in case (A)(i) if $\mu_1 = 2$ and in case (A)(iv) if $\mu = \mu_1 = 2$. In both cases, D_h meets the S_0-component, so F_0 contains no D-components. If p is untwisted with base $\mathbb{P}^1$, then

$$n - 1 \geq \lambda = \kappa_0 \geq \frac{n}{2} - 1,$$

so $n = 2$ (because $\lambda = -1/\mu_1 < 0$ for $n = 1$) and $\kappa_0 = 1 - 1/\mu_1 - 1/\mu_2$, which vanishes only if $\mu_1 = \mu_2 = 2$. Assume now that p is untwisted with base $\mathbb{C}^1$. Then

$$n > \kappa_0 \geq \lambda - 1 \geq \frac{n}{2} - 1,$$

so $n \in \{1, 2\}$. There are no (-1)-curves in $D \cap F_0$ by the p-minimality of D, so at least one S_0-component, say C, is a (-1)-curve. We can also assume that C is contracted by η in case $\eta \neq \mathrm{id}$ and that $\mu \geq \tilde{\mu}$ in case $\eta = \mathrm{id}$. Then $\kappa_0 = \lambda - 1/\tilde{\mu}$. The composition ξ of successive contractions of all (-1)-curves in $\underline{F}_0 - \widetilde{C}$ and its images is a connected modification. Suppose $n = 2$. The inequalities above give $\lambda = 1$, so $\mu_1 = \mu_2 = 2$ and $\tilde{\mu} = 1$. Then $\xi_* F_0 = [0]$, and because ξ is a connected modification, $\widetilde{C}$ is a tip of F_0. So one of C, $\widetilde{C}$ intersects D_h, because otherwise $F_0 - \widetilde{C} - C - \widehat{E}$ would be connected and would intersect both sections from D_h, and hence F_0 would contain a rivet. This gives (iv). Suppose that $n = 1$. Then $\mu_1 = \tilde{\mu} = 2$. By the choice of C, further contractions of F_0 to a smooth fiber are subdivisional for $\xi_* D \cup \xi_* F_0$, so we have $\xi_* F_0 = [2, 1, 2]$ with the birational transform of $\widetilde{C}$ in the middle, and the image of D_h intersects both (-2)-tips of $\xi_* F_0$. Since ξ is a connected modification, it does not touch one of these tips, so one of the connected components of $D \cap F_0$ is a (-2)-curve. If $\mu = 1$, then $\mu < \tilde{\mu}$, so by our assumption $\eta \neq \mathrm{id}$. But then $\mu > 1$, because $C^2 = -1$ and C intersects $\widehat{E}$ and D. This contradiction ends the proof of (iii). □

4D. ***Constructions.*** Lemmas 4.5 and 2.12 give a practical method of reconstructing all $\mathbb{C}^*$-ruled $\mathbb{Q}$-homology planes. We summarize it in here. We denote irreducible curves and their proper transforms by the same letters.

Construction 4.11. **Case 1** (twisted ruling). Let D_h be a smooth conic on $\mathbb{P}^2$, let L_0, L_∞ be tangents to D_h at distinct points x_0, x_∞, and let L_i, for $i = 1, \ldots, n$ and $n \geq 0$, be distinct lines through $L_0 \cap L_\infty$, other than L_0, L_∞. Blow up once at $L_0 \cap L_\infty$; let $p\colon \mathbb{F}_1 \to \mathbb{P}^1$ be the $\mathbb{P}^1$-ruling of the resulting Hirzebruch surface. Over each of $p(L_0)$, $p(L_\infty)$, blow up on D_h twice, creating singular fibers $\widetilde{F}_0 = [2, 1, 2]$

and $F_\infty = [2, 1, 2]$. For each $i = 1, \dots, n$, by a connected sequence of blow-ups subdivisional for $L_i + D_h$, create a column fiber F_i over $p(L_i)$ and denote its unique (-1)-curve by C_i. By some connected sequence of blow-ups with a center on $\widetilde{F}_0$, create a singular fiber F_0, and denote the newly created (-1)-curve by C (if the sequence is empty, define C as the (-1)-curve of $\widetilde{F}_0$). Denote the resulting surface by $\bar{S}$, put

$$T = D_h + \underline{F_\infty} + (\underline{F_1} - C_1) + \cdots + (\underline{F_n} - C_n) + \underline{F_0} - C,$$

and construct S' as in Lemma 2.12. S' is a $\mathbb{Q}$-homology plane (singular as long as T is not connected) because conditions 2.12(i)–(iii) are satisfied by construction and (iv) by Lemma 4.4. To see that each S' admitting a twisted $\mathbb{C}^*$-ruling can be obtained in this way, note that by the p-minimality of D, even if F_0 contains two (-1)-curves C and $B \subseteq D$, then B is not a tip of F_0 and ζ does not touch it, so in each case the modification $F_0 \to \zeta_* F_0$ induced by ζ is connected, and we are done by Lemma 4.5.

Case 2 (untwisted ruling with base $\mathbb{C}^1$). Let $x_0, x_1 \dots x_n, x_\infty, y \in \mathbb{P}^2$, for $n \geq 0$, be distinct points, such that all but y lie on a common line D_1. Let L_i be a line through x_i and y. Blow up y once and let D_2 be the negative section of the $\mathbb{P}^1$-ruling of the resulting Hirzebruch surface $p\colon \mathbb{F}_1 \to \mathbb{P}^1$. For each $i = 0, 1, \dots, n$, by a connected sequence of blow-ups (which can be empty if $i = 0$), with first center x_i and subdivisional for $D_1 + L_i$, create a column fiber F_i ($\widetilde{F}_0$ if $i = 0$) over $p(x_i)$ and denote its unique (-1)-curve by C_i if $i \neq 0$ and by $\widetilde{C}$ if $i = 0$ (put $\widetilde{C} = L_0$ if the sequence over $p(x_0)$ is empty). Choose a point $z \in F_0$ that lies on $D_1 + \underline{\widetilde{F}_0} - \widetilde{C}$, and by a nonempty connected sequence of blow-ups with first center z, create some singular fiber F_0 over $p(x_0)$. Let C be the new (-1)-curve. Denote the resulting surface by $\bar{S}$, put

$$T = D_1 + D_2 + L_\infty + (\underline{F_1} - C_1) + \cdots + (\underline{F_n} - C_n) + \underline{F_0} - C - \widetilde{C},$$

and construct S' as in Lemma 2.12. The surface S' is a $\mathbb{Q}$-homology plane by Lemma 4.4, because Lemma 4.4(ii) is satisfied by the choice of z. To see that all S' admitting an untwisted $\mathbb{C}^*$-ruling with base $\mathbb{C}^1$ can be obtained in this way, note that by changing the completion of S' by a flow if necessary, we can assume that one of the components of D_h is a (-1)-curve. $D \cap F_0$ contains no (-1)-curves, and as was shown in the proof of Theorem 4.9, η contracts at most one of C, $\widetilde{C}$. Then by Lemma 4.5, we are done.

Case 3 (untwisted ruling with base $\mathbb{P}^1$). Let D_2 be the negative section of the $\mathbb{P}^1$-ruling of a Hirzebruch surface $p\colon \mathbb{F}_N \to \mathbb{P}^1$, with $N > 0$. Let $x_0, x_1, \dots, x_n$, with $n \geq 0$ be points on a section D_1 of p disjoint from D_2. For each $i = 0, 1, \dots, n$, by a connected sequence of blow-ups (which can be empty if $i = 0$), with first center

x_i and subdivisional for $D_1 + p^{-1}(p(x_i))$, create a column fiber F_i ($\widetilde{F}_0$ if $i = 0$) over $p(x_i)$ and denote its unique (-1)-curve by C_i if $i \neq 0$ and by B if $i = 0$ (put $B = p^{-1}(p(x_0))$ if the sequence over $p(x_0)$ is empty). Assume that the intersection matrix of at least one of two connected components of

$$D_1 + D_2 + (\underline{F_1} - C_1) + \cdots + (\underline{F_n} - C_n) + (\underline{\widetilde{F}_0} - B)$$

is nondegenerate. By a connected sequence of blow-ups starting from a sprouting blow-up for $D_1 + \widetilde{F}_0$ with center on B, create some singular fiber F_0 over $p(x_0)$ and let C be the new (-1)-curve. Denote the resulting surface by $\overline{S}$, put

$$T = D_1 + D_2 + (\underline{F_1} - C_1) + \cdots + (\underline{F_n} - C_n) + (\underline{F_0} - C),$$

and construct S' as in Lemma 2.12. D is connected because the modification $F_0 + D_1 \to \widetilde{F}_0 + D_1$ is not subdivisional, so S' is a $\mathbb{Q}$-homology plane by Lemma 4.4. By Lemmas 4.5 and 4.4, each S' with an untwisted $\mathbb{C}^*$-ruling having a base $\mathbb{P}^1$ can be obtained in this way.

5. Corollaries

5A. ***Completions and singularities.*** Recall that $\mathbb{Q}$-homology planes with nonquotient singularities have unique snc-minimal completions (and hence also balanced ones) and unique singular points [Palka 2008, 1.2]. The completions and singularities in case $\bar{\kappa}(S_0) = -\infty$ are described in Section 3. In case $\bar{\kappa}(S_0) = 2$, the singular point is unique and of quotient type [ibid.]. Also, the snc-minimal boundary cannot contain nonbranching b-curves with $b \geq 0$, because these induce $\mathbb{C}^1$- or $\mathbb{C}^*$-rulings of S_0, and hence the snc-minimal completion is unique. Theorem 1.1 summarizes the remaining cases.

Proof of Theorem 1.1. (1) Suppose S' has at least two different balanced completions. These differ by a flow, which implies that the boundary contains a nonbranching rational component F_∞ with zero self-intersection. Then F_∞ is a fiber of a $\mathbb{P}^1$-ruling p of a balanced completion (V, D). We may assume that F_∞ is not contained in any maximal twig of D. Indeed, after moving the 0-curve by a flow to a tip of a new boundary, it gives an affine ruling of S', which is possible only if $\bar{\kappa}(S_0) = -\infty$. Because F_∞ is nonbranching, the induced ruling restricts to an untwisted $\mathbb{C}^*$-ruling of S'. It follows from the connectedness of the modification η (see the proof of Theorem 4.9) that $n > 0$, so this restriction has more than one singular fiber. Both components of D_h are branching in D. Since F_∞ is the only nonbranching 0-curve in D, centers of elementary transformations lie on the intersection of the fiber at infinity with D_h. If D is strongly balanced, then one of the components of D_h is a 0-curve, and hence there are at most two strongly balanced completions. Conversely, suppose that S' has an untwisted $\mathbb{C}^*$-ruling with base $\mathbb{C}^1$ and that $n > 0$, and let (V, D, p) be a completion of this ruling. Because S' is not affine-ruled, the horizontal components H, H' of D are branching, so (V, D) is balanced and we can assume $H'^2 = 0$. Because H, H' are proper transforms of two disjoint sections on a Hirzebruch surface, we have $H^2 + H'^2 + n \leq 0$, so $H^2 \neq 0$ and we can obtain a different strongly balanced completion of S' by a flow that makes H into a 0-curve.

(2), (3) By [Palka 2008, 4.5] and [Palka 2011a], we may assume that S' is $\mathbb{C}^*$-ruled. If this ruling is untwisted, it follows from the proof of Theorem 4.9 that S' has a unique singular point, and it is a cyclic singularity. In the twisted case, because $\widehat{E} \subseteq F_0$, if $\widehat{E}$ is not connected then F_0 is of type (A)(i), and if $\widehat{E}$ is not a chain then F_0 is of type (A)(iv). □

Remark. The set of isomorphism classes of strongly balanced boundaries that a given surface admits is an invariant of the surface, which allows us to easily distinguish between many $\mathbb{Q}$-acyclic surfaces.

5B. ***Singular planes of negative Kodaira dimension.*** As another corollary of Theorem 4.9 we give a detailed description of singular $\mathbb{Q}$-homology planes of negative Kodaira dimension. We assume that $\bar{\kappa}(S_0) \neq 2$, but as we show in [Palka and Koras 2010], this assumption is redundant.

Theorem 5.1. *Suppose that S' is a singular $\mathbb{Q}$-homology plane of negative Kodaira dimension and that S_0 is its smooth locus. If $\bar{\kappa}(S_0) \neq 2$, then exactly one of the following holds*:

(i) $\bar{\kappa}(S_0) = -\infty$; *$S'$ is affine-ruled or isomorphic to $\mathbb{C}^2/G$ for a small finite noncyclic subgroup $G < \mathrm{GL}(2, \mathbb{C})$.*

(ii) $\bar{\kappa}(S_0) \in \{0, 1\}$; *$S'$ is nonlogarithmic and is isomorphic to a quotient of an affine cone over a smooth projective curve by an action of a finite group acting freely off the vertex of the cone and preserving the set of lines through the vertex.*

(iii) $\bar{\kappa}(S_0) \in \{0, 1\}$; *$S'$ has an untwisted $\mathbb{C}^*$-ruling with base $\mathbb{C}^1$ and two singular fibers. One of them consists of two $\mathbb{C}^1$'s meeting in a cyclic singular point*; *after taking a resolution and completion*, *the respective completed singular fiber is of type* (B)(i) *with* $\mu, \tilde{\mu} \geq 2$ (see Figure 3 and Theorem 4.9).

Proof. By [Palka 2011a; Palka 2008, 4.5] and Section 3, we may assume that S' is logarithmic and $\mathbb{C}^*$-ruled and that $\bar{\kappa}(S_0) \geq 0$. We need to show (iii). Let (V, D, p) be a minimal completion of the $\mathbb{C}^*$-ruling. By Theorem 4.9, if p is twisted, then

$$0 > \kappa_0 \geq \lambda - \tfrac{1}{2} \geq \frac{n-1}{2},$$

so $n = \lambda = 0$. The inequalities $\kappa < 0$ and $\kappa_0 \geq 0$ can be satisfied only in case (A)(iii), and then $D_h^2 = 0$ by Lemma 4.5, so D_h induces an untwisted $\mathbb{C}^*$-ruling of

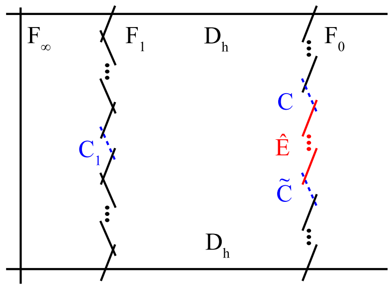


Figure 3. Untwisted $\mathbb{C}^*$-ruling, $\bar{\kappa}(S') = -\infty$.

S'. Suppose p is untwisted. Because $\kappa \neq \kappa_0$, p has base $\mathbb{C}^1$ and is of type (B)(i). Because

$$0 > \kappa = \lambda - 1 \geq \frac{n}{2} - 1,$$

we get $n \leq 1$, but for $n = 0$ we get $\kappa_0 < \lambda < 0$, so in fact $n = 1$. Then $0 \leq \kappa_0 = 1 - 1/\mu_1 - 1/\min(\mu, \tilde{\mu})$, and hence $\min(\mu, \tilde{\mu}) \geq 2$. □

By Proposition 2.11, $H_i(S', \mathbb{Z})$ vanishes for $i > 1$. If S' is of type $\mathbb{C}^2/G$ or of type (ii), then it is contractible. $H_1(S', \mathbb{Z})$ for affine-ruled S' was computed in Remark 3.4. For completeness, we now compute the fundamental group of S' of type (iii), which by Proposition 2.11 is the same as $\pi_1(S)$. Let E_0 be a component of $\widehat{E}$ intersecting C. Contract $\widetilde{C}$ and successive vertical (-1)-curves until C is the only (-1)-curve in the fiber (C cannot became a 0-curve, because it does not intersect D_h), and denote this contraction by θ. Let θ' be the contraction of $\theta_* F_0$ and F_1 to smooth fibers. Put $U = S_0 \setminus (C_1 \cup C \cup \widetilde{C})$ and let $\gamma_1, \gamma, t \in \pi_1(U)$ be the vanishing loops of the images of F_1, F_0 under $\theta' \circ \theta$ and of some component of D_h (see [Fujita 1982, 4.17]). We need to compute the kernel of the epimorphism $\pi_1(U) \to \pi_1(S)$. Because θ does not touch C, $\theta_* F_0$ is columnar and $\theta_* E_0 \neq 0$. Using [ibid., 7.17], one can show by induction on the number of components of a columnar fiber that because $E_0 \cdot C \neq 0$, the vanishing loops of E_0 and C, which are of type $\gamma^a t^b$ and $\gamma^c t^d$, satisfy $ad - bc = \pm 1$. Thus γ and t are in the kernel, and hence

$$\pi_1(S) = \langle \gamma_1 : \gamma^{\mu_1} \rangle \cong \mathbb{Z}_{\mu_1}.$$

In particular, S' is not a $\mathbb{Z}$-homology plane.

6. Uniqueness of $\mathbb{C}^*$-rulings

6A. ***The number of*** $\mathbb{C}^*$***-rulings.*** We consider the question of uniqueness of $\mathbb{C}^*$-rulings of S_0 and S'. Recall that a $\mathbb{C}^*$-ruling of S_0 is *extendable* if it extends to a ruling (morphism) of S'. Two rational rulings of a given surface are considered the same if they differ by an automorphism of the base. When a $\mathbb{C}^*$-ruling of S_0

exists, using the information on snc-minimal boundaries, we are able to compute the number of different $\mathbb{C}^*$-rulings.

Theorem 6.1. *Let S' be a singular $\mathbb{Q}$-homology plane that is not affine-ruled. Let $p_1, \dots, p_r$ for $r \in \mathbb{N} \cup \{\infty\}$ be all different $\mathbb{C}^*$-rulings of the smooth locus S_0 of S'. Let D be an snc-minimal boundary of S'.*

(1) *If $\bar\kappa(S_0) = 2$ or if S' is exceptional (so that $\bar\kappa(S_0) = 0$), then $r = 0$.*

(2) *If $\bar\kappa(S_0) = 1$ or if S' is nonlogarithmic, then $r = 1$.*

(3) *If $\bar\kappa(S_0) = -\infty$, then $r \geq 1$ and p_1 is nonextendable. Also, $r \neq 1$ only if the fork that is an exceptional divisor of the snc-minimal resolution of S' is of type $(2, 2, k)$. In this case we have:*

  (i) *If $k \neq 2$, then $r = 2$, p_2 is twisted, and it has a unique singular fiber, which is of type* (A)(iv).

  (ii) *If $k = 2$, then $r = 4$, p_2, p_3, p_4 are twisted, and they have unique singular fibers, which are of type* (A)(iv).

(4) *Assume that $\bar\kappa(S_0) = 0$ and that S' is logarithmic and not exceptional. Then all p_i extend to $\mathbb{C}^*$-rulings of S' and the following hold:*

  (i) *If the dual graph of D is*

$$\begin{array}{ccccccc} -2 & \text{———} & -1 & \text{———} & k & \text{———} & -2 \\ & & | & & | & & \\ & & -2 & & -2 & & \end{array}$$

  with $k \leq -2$, then $r = 1$ and p_1 is twisted.

  (ii) *If the dual graph of D is*

$$\begin{array}{ccccccc} -2 & \text{———} & -1 & \text{———} & -1 & \text{———} & -2 \\ & & | & & | & & \\ & & -2 & & -2 & & \end{array}$$

  then $r = 2$ and p_1, p_2 are twisted.

  (iii) *If the dual graph of D is*

$$\begin{array}{ccccccccc} -2 & \text{———} & k & \text{———} & 0 & \text{———} & m & \text{———} & -2 \\ & & | & & & & | & & \\ & & -2 & & & & -2 & & \end{array}$$

  then $r = 3$, p_1, p_2 are twisted and p_3 is untwisted with base $\mathbb{C}^1$.

  (iv) *In all other cases, $r = 2$, p_1 is twisted and p_2 is untwisted.*

Proof. (1) By definition, exceptional $\mathbb{Q}$-homology planes are not $\mathbb{C}^*$-ruled. If S_0 is of general type, then by Iitaka's easy addition formula [Iitaka 1982, 10.4], S_0 is not $\mathbb{C}^*$-ruled.

(2) If S' is nonlogarithmic, then by [Palka 2008, 4.1], the $\mathbb{C}^*$-ruling of S' is unique. Assume that $\bar{\kappa}(S_0)=1$. Let $(\bar{S}, D)$ be some normal completion of the snc-minimal resolution $S \to S'$. Denote the exceptional divisor of the resolution by $\widehat{E}$. By [Fujita 1982, 6.11], for some $n>0$, the base locus of $|n(K_{\bar{S}}+D+\widehat{E})^+|$ is empty and the linear system gives a $\mathbb{P}^1$-ruling of $\bar{S}$ that restricts to a $\mathbb{C}^*$-ruling of S_0; see also [Miyanishi 2001, 2.6.1]. Consider another $\mathbb{C}^*$-ruling of S_0. Modifying $\bar{S}$ if necessary, we can assume that it extends to a $\mathbb{P}^1$-ruling of $\bar{S}$. Let f' be a general fiber of this extension. Then

$$f'\cdot(K_{\bar{S}}+D+\widehat{E}) = f'\cdot K_{\bar{S}}+2=0,$$

and hence

$$f'\cdot(K_{\bar{S}}+D+\widehat{E})^+ + f'\cdot(K_{\bar{S}}+D+\widehat{E})^- = 0.$$

However, $(K_{\bar{S}}+D+\widehat{E})^-$ is effective and $(K_{\bar{S}}+D+\widehat{E})^+$ is numerically effective, so

$$f'\cdot(K_{\bar{S}}+D+\widehat{E})^+ = f'\cdot(K_{\bar{S}}+D+\widehat{E})^- = 0,$$

and we see that the rulings are the same.

(3), (4) We need to understand how to find all twisted $\mathbb{C}^*$-rulings of a given S'. Consider a twisted $\mathbb{C}^*$-ruling of S' and let $(\widetilde{V}, \widetilde{D}, \tilde{p})$ be a minimal completion of this ruling. By the $\tilde{p}$-minimality of $\widetilde{D}$, the only component of $\widetilde{D}$ that can be a non-branching (-1)-curve is $\widetilde{D}_h$, so there is a connected modification $(\widetilde{V}, \widetilde{D}) \to (V, D)$ with snc-minimal D. Let $\widetilde{D}_0 \subseteq \widetilde{D}$ be the (-1)-curve of the fiber at infinity (see Lemma 4.1). D is not a chain; otherwise S' would be affine-ruled. Let $D_0 \subseteq D$ be the image of $\widetilde{D}_0$, and let T be the connected component of $D-D_0$ containing the image of the horizontal component (which is a point if the modification is nontrivial). In this way, a twisted $\mathbb{C}^*$-ruling of S' determines a pair (D_0, T) (with D_0+T contained in a boundary of some snc-minimal completion), such that $\beta_D(D_0)=3$, $D_0^2 \geq -1$, T is a connected component of $D-D_0$ containing the image of the horizontal section, and both connected components of $D-D_0-T$ are (-2)-curves. Conversely, if we have an snc-minimal normal completion (V, D) and a pair as above, we make a connected modification $(\widetilde{V}, \widetilde{D}) \to (V, D)$ over D by blowing successively on the intersection of the total transform of T with the proper transform of D_0 until D_0 becomes a (-1)-curve. The (-1)-curve together with the transform of $D-T-D_0$ induce a $\mathbb{P}^1$-ruling of V' and constitute the fiber at infinity for this ruling. The restriction to S' is a twisted $\mathbb{C}^*$-ruling.

Suppose $\bar{\kappa}(S_0) = -\infty$. Since S_0 is not affine-ruled, $S' \cong \mathbb{C}^2/G$ for a finite noncyclic small subgroup $G < \mathrm{GL}(2, \mathbb{C})$ (see Section 3). Let (V, D) be an snc-minimal normal completion of S' and let $\bar{S} \to V$ be a minimal resolution with exceptional divisor $\widehat{E}$. We saw in the proof of Proposition 3.1 that S_0 admits a platonic $\mathbb{C}^*$-ruling, which extends to a $\mathbb{P}^1$-ruling of $\bar{S}$. Also, D and $\widehat{E}$ are forks for which D_h and $\widehat{E}_h$ are the unique branching components of D and E respectively. In particular, the $\mathbb{C}^*$-ruling does not extend to a ruling of S', and because nonbranching components of D have negative self-intersections, $(\bar{S}, D + \widehat{E})$ is a unique snc-minimal smooth completion of S_0 (and hence (V, D) is a unique snc-minimal normal completion of S'). It follows from the proof of [Palka 2008, 4.1] that the nonextendable $\mathbb{C}^*$-ruling of S_0 is unique. Suppose there is a $\mathbb{C}^*$-ruling of S_0 that does extend to S'. Since $\widehat{E}$ is not a chain, it follows from the proof of Theorem 4.9 that this ruling is twisted. Since maximal twigs of $\widehat{E}$ and D are adjoint chains of columnar fibers, we see that a maximal twig of $D - D_h$ is a (-2)-curve if and only if the respective maximal twig of $\widehat{E} - \widehat{E}_h$ is a (-2)-curve. Also, $0 < d(\widehat{E})$, so $\widehat{E}_h^2 \le -2$, and because $\widehat{E}_h^2 + D_h^2 = -3$, we have $D_h^2 \ge -1$. Therefore, S' admits a twisted $\mathbb{C}^*$-ruling if and only if $\widehat{E}$ is a fork of type $(2, 2, k)$ for some $k \ge 2$. If $k \ne 2$, then the choice of (D_0, T) as above is unique, and if $k = 2$, then there are three such choices. If (V', D', p) is a minimal completion of such a ruling, then D' is a fork, so because $\kappa_0 < 0$, we have $n = 0$ and F_0 is of type (A)(iv) (see the proof of Theorem 4.9). This gives (3).

We can now assume that $\bar{\kappa}(S_0) = 0$ and that S' is logarithmic and not exceptional. Then S_0 is $\mathbb{C}^*$-ruled and by [Palka 2008, 4.7(iii)], each $\mathbb{C}^*$-ruling of S_0 extends to a $\mathbb{C}^*$-ruling of S'. Let $r \in \{1, 2, \dots\} \cup \{\infty\}$ be the number of different (up to automorphism of the base) $\mathbb{C}^*$-rulings of S' and let (V_i, D_i, p_i), for $i \le r$, be their minimal completions. Minimality implies that nonbranching (-1)-curves in D_i are p_i-horizontal. We add consequently an upper index (i) to objects defined previously for any $\mathbb{C}^*$-ruling when we refer to the ruling p_i. If p_i is untwisted, we denote the horizontal components of $D_h^{(i)}$ by $H^{(i)}$, $H'^{(i)}$.

Suppose p_1 is untwisted with base $\mathbb{P}^1$. Then $F_0^{(1)}$ contains a rivet and by Corollary 4.10, $n^{(1)} = 2$, so D_1 does not contain nonbranching b-curves with $b \ge -1$. Then (V_1, D_1) is balanced and S' does not admit an untwisted $\mathbb{C}^*$-ruling with base $\mathbb{C}^1$, because it does not contain nonbranching 0-curves (see Lemma 4.1). By Corollary 4.10, each component of $D_h^{(1)}$ has $\beta_{D_1} = 3$ and intersects two (-2)-tips of D_1. Note that $\zeta^{(1)}$ (see Lemma 4.5) touches $D_h^{(1)}$ two times if both components of $D_h^{(1)}$ intersect the same horizontal component of $F_0^{(1)}$ and three times if not. By Lemma 4.5 and the properties of Hirzebruch surfaces, we get $-3 \le (D_h^{(1)})^2 \le -2$. In particular, one of the components of $D_h^{(1)}$, say $H^{(1)}$, has $(H^{(1)})^2 \ge -1$, so by the discussion about twisted $\mathbb{C}^*$-rulings above, $H^{(1)}$ together with two (-2)-tips of D_1 gives rise to a twisted $\mathbb{C}^*$-ruling p_2 of S'. Because $H'^{(1)}$ together with two

(-2)-tips of D_1 intersecting it are contained in a fiber of p_2, $(H'^{(1)})^2 \leq -2$. Thus p_2 is the only twisted ruling of S', because $H^{(1)}$ is the only possible choice for a middle component of the fiber at infinity of a twisted ruling. Suppose $r \geq 3$. Then p_3 is untwisted with base $\mathbb{P}^1$. Because D_1 does not contain nonbranching 0-curves, any flow in D_1 is trivial, so $V_3 = V_1$. Because p_3 and p_1 are different after restriction to S', the S_0-components $C^{(1)}$, $C^{(3)}$ contained respectively in $F_0^{(1)}$, $F_0^{(3)}$ are different. Because they both intersect $\widehat{E}$, they are contained in the same fiber of p_2, which contradicts $\Sigma_{S_0}^{(2)} = 0$. Because D contains no nonbranching 0-curves, D is not of type (4)(iii). Since $n^{(1)} = 2$, D contains at least seven components, so D is not of type (4)(i) or (4)(ii).

We can now assume that each untwisted $\mathbb{C}^*$-ruling of S' has base $\mathbb{C}^1$. Suppose p_1 is such a ruling. By Corollary 4.10, both horizontal components of D_1 have $\beta_{D_1} = 3$, and one of them, say $H'^{(1)}$, intersects two (-2)-tips T and T' of D_1. In particular, D_1 is snc-minimal. Because $F_\infty^{(1)} = [0]$, changing V_1 by a flow if necessary, we may assume that $H'^{(1)}$ is a (-1)-curve. Then

$$F_\infty^{(2)} = T + 2H'^{(1)} + T'$$

induces a $\mathbb{P}^1$-ruling $p_2 : V_1 \to \mathbb{P}^1$, which is a twisted $\mathbb{C}^*$-ruling after restricting it to S'. Suppose $r \geq 3$. If p_3 is untwisted, then its base is $\mathbb{C}^1$, and changing V_3 by a flow if necessary, we can assume that $V_3 = V_1$. But then $F_\infty^{(1)} = F_\infty^{(3)}$, because D_1 contains only one nonbranching 0-curve, so p_1 and p_3 have a common fiber and hence cannot be different after restriction to S', which is a contradiction. Thus p_3 is twisted. By the discussion above, p_3 can be recovered from a pair (D_0, T) on some snc-minimal completion of S'. All such completions of S' differ from (V_1, D_1) by a flow, which is an identity on $V_1 - F_\infty^{(1)}$, and hence the birational transform of D_0 on V_1 is either $H^{(1)}$ or $H'^{(1)}$. Because the restrictions of p_1 and p_2 to S' are different, it is $H^{(1)}$. It follows that $r = 3$ and that $D_1 - H'^{(1)}$ has two (-2)-tips as connected components, and hence the dual graph of D_1 is as in (iii). Conversely, if S' has a boundary as in (iii), then besides the untwisted $\mathbb{C}^*$-ruling induced by the 0-curve, it has also two twisted rulings, each with one of the branching components as the middle component of the fiber at infinity.

We can finally assume that all $\mathbb{C}^*$-rulings of S' are twisted. Let (V, D) be a balanced completion of S'. Because S' does not admit untwisted $\mathbb{C}^*$-rulings, D does not contain nonbranching 0-curves, so (V, D) is a unique snc-minimal completion of S'. Thus, to find all twisted $\mathbb{C}^*$-rulings of S', we need to determine all pairs (D_0, T) such that $D_0 + T \subseteq D$, $D_0^2 \geq -1$, $\beta_D(D_0) = 3$, and $D - T - D_0$ consists of two (-2)-tips. Let (D_0, T) and (D_0', T') be two such pairs. Suppose $D_0 \neq D_0'$ and, say, $D_0'^2 \geq D_0^2$. We have $D_0 \cdot D_0' \neq 0$, for otherwise the chain $D - T'$, which is not negative definite, would be contained in (and not equal to, because $\nu \leq 1$) a fiber of the twisted ruling associated with (D_0, T), which is impossible. Then D

has six components and we check that

$$d(D) = 16((D_0^2+1)(D_0'^2+1)-1),$$

so $(D_0^2+1)(D_0'^2+1) \leq 0$, because $d(D) < 0$. Then $D_0^2 = -1$ and D_0' is a 2-section of the twisted ruling associated with (D_0, T). Because $\beta_D(D_0') = 3$, by Corollary 4.10 and Lemma 4.5 for this ruling $n = 1$, D_0' is a (-1)-curve and D has dual graph as in (ii). Conversely, it is easy to see that S' with such a boundary has two twisted $\mathbb{C}^*$-rulings. Therefore, we can assume that the choice of D_0 for a pair (D_0, T) as above is unique. Let p_1 be a twisted $\mathbb{C}^*$-ruling associated with some pair (D_0, T). Suppose $n^{(1)} = 0$. By Lemma 4.5, $\zeta_* D_h^{(1)}$ is a 0-curve, so

$$F = \zeta^* \zeta_* D_h^{(1)}$$

induces a $\mathbb{P}^1$-ruling p of V. If ζ touches $D_h^{(1)}$, then F contains the S_0-component of $F_0^{(1)}$, so $F \nsubseteq D$ and p restricts to an untwisted $\mathbb{C}^*$-ruling of S' with base $\mathbb{P}^1$. If ζ does not touch $D_h^{(1)}$, then p restricts to a $\mathbb{C}^*$-ruling of S' with base $\mathbb{C}^1$. This contradicts the assumption. By Corollary 4.10 we get that $n^{(1)} = 1$, $F_0^{(1)}$ contains no D_1-components, and $\mu_1 = 2$. In particular, $D_1 = D$. By Lemma 4.5, $(D_h^{(1)})^2 \leq -1$ because $n^{(1)} = 1$, so D has a dual graph as in (i) or (ii). Conversely, if D is of type (i) or (ii), then $r = 2$ if $k = -1$ and $r = 1$ if $k \leq -2$. □

6B. ***The number of affine lines.*** Theorem 6.1 has interesting consequences. It is known [Zaĭdenberg 1987; Gurjar and Miyanishi 1992] that $\mathbb{Q}$-homology planes with smooth locus of general type (in particular the smooth ones) do not contain topologically contractible curves. In fact, the number $\ell \in \mathbb{N} \cup \{\infty\}$ of contractible curves on a $\mathbb{Q}$-homology plane S' is known except two cases: when S' is non-logarithmic and when S' is singular and $\bar{\kappa}(S_0) = 0$ (see [Palka 2011b, 10.1] and references there). Clearly, in the first case $\ell = \infty$ by the main result of [Palka 2008]. The case when S' is smooth and of Kodaira dimension zero has been considered in [Gurjar and Parameswaran 1995]. Theorem 1.3 is the missing piece of information, and the method can be easily applied to recover the result of Gurjar and Parameswaran.

Proof of Theorem 1.3. We can assume that S' is logarithmic. Suppose S' contains a topologically contractible curve L. We show that L is vertical for some $\mathbb{C}^*$-ruling of S'. The proper transform of L on $\bar{S}$ meets each connected component of $\widehat{E}$ in at most one point. We use the logarithmic Bogomolov–Miyaoka–Yau inequality as in [Koras and Russell 2007, 2.12] to show that $\bar{\kappa}(S_0 - L) \leq 1$. In case $\bar{\kappa}(S_0 - L) = 1$, the surface $S_0 - L$ is $\mathbb{C}^*$-ruled [Fujita 1982, 6.11], so we may assume that $\bar{\kappa}(S_0 - L) = 0$. Let $\mathbb{Z}[D + \widehat{E}]$ be a free abelian group generated by the components of $D + \widehat{E}$. Because

$$\operatorname{Pic} S_0 = \operatorname{Coker}(\mathbb{Z}[D + \widehat{E}] \to \operatorname{Pic} \bar{S})$$

is torsion, the class of L in Pic S_0 is torsion. So there is a surjection $f: S_0 - L \to \mathbb{C}^*$, and taking its Stein factorization, we get a $\mathbb{C}^*$-ruling of $S_0 - L$, which (because $\bar{\kappa}(S_0) \neq -\infty$) extends to a $\mathbb{C}^*$-ruling of S_0. Since S_0 is logarithmic, each $\mathbb{C}^*$-ruling of S_0 extends in turn to a $\mathbb{C}^*$-ruling of S'. Therefore L is vertical for some $\mathbb{C}^*$-ruling of S' and we are done. In particular, exceptional $\mathbb{Q}$-homology planes do not contain contractible curves. It follows from Corollary 4.10 that if the ruling is twisted or untwisted with base $\mathbb{P}^1$, then the vertical contractible curve is unique and is contained in the unique singular noncolumnar fiber. For an untwisted ruling with base $\mathbb{C}^1$, there are at most two such curves. In particular, in cases (4)(i) and (4)(ii) of Theorem 6.1, L needs to intersect the horizontal component of the boundary, so we get respectively $\ell = 1$ and $\ell = 2$. In case (4)(iii), the unique vertical contractible curves for the twisted rulings p_1 and p_3 are distinct and do not intersect the horizontal components of respective rulings, and hence are both vertical for the untwisted ruling p_3, so $\ell = 2$. In the remaining case (4)(iv), $r = 2$, p_1 is twisted and p_2 is untwisted. We can assume that the base of p_2 is $\mathbb{C}^1$ and the unique noncolumnar singular fiber contains two contractible curves, L_1 and L_2, for otherwise $\ell \leq 2$ by the above remarks and we are done. Since the twisted ruling is unique, there is exactly one horizontal component H of $D_h^{(2)}$ that meets two (-2)-tips of $D_h^{(1)}$ (together with these tips it induces the twisted ruling). Clearly, only one L_i can intersect H, so the second one is vertical for p_1 and we get $\ell \leq 2$ is this case too. □

References

[Barth et al. 2004] W. P. Barth, K. Hulek, C. A. M. Peters, and A. Van de Ven, *Compact complex surfaces*, 2nd ed., Ergebnisse der Math. (3) **4**, Springer, Berlin, 2004. MR 2004m:14070 Zbl 1036.14016

[Daigle 2008] D. Daigle, "Classification of linear weighted graphs up to blowing-up and blowing-down", *Canad. J. Math.* **60**:1 (2008), 64–87. MR 2008j:05294 Zbl 1138.14021 arXiv math/0305029

[Flenner and Zaĭdenberg 1994] H. Flenner and M. G. Zaĭdenberg, "$\mathbb{Q}$-acyclic surfaces and their deformations", pp. 143–208 in *Classification of algebraic varieties* (L'Aquila, 1992), edited by C. Ciliberto et al., Contemp. Math. **162**, Amer. Math. Soc., Providence, RI, 1994. MR 95b:14024 Zbl 0838.14027

[Flenner et al. 2007] H. Flenner, S. Kaliman, and M. G. Zaĭdenberg, "Birational transformations of weighted graphs", pp. 107–147 in *Affine algebraic geometry*, edited by T. Hibi, Osaka Univ. Press, Osaka, 2007. MR 2008g:14117 Zbl 1129.14056 arXiv math/0511063

[Flenner et al. 2011] H. Flenner, S. Kaliman, and M. G. Zaĭdenberg, "Corrigendum to our paper: Birational transformations of weighted graphs", pp. 35–38 in *Affine algebraic geometry*: *the Russell Festschrift* (Montréal, 2009), edited by D. Daigle et al., CRM Proc. Lecture Notes **54**, Amer. Math. Soc., Providence, RI, 2011. MR 2768632 Zbl 05994616 arXiv 0910.1939

[Fujita 1982] T. Fujita, "On the topology of noncomplete algebraic surfaces", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **29**:3 (1982), 503–566. MR 84g:14035 Zbl 0513.14018

[Gurjar and Miyanishi 1992] R. V. Gurjar and M. Miyanishi, "Affine lines on logarithmic $\mathbb{Q}$-homology planes", *Math. Ann.* **294**:3 (1992), 463–482. MR 94g:14018 Zbl 0757.14022

[Gurjar and Parameswaran 1995] R. V. Gurjar and A. J. Parameswaran, "Affine lines on $\mathbb{Q}$-homology planes", *J. Math. Kyoto Univ.* **35**:1 (1995), 63–77. MR 96b:14046 Zbl 0861.14031

[Gurjar et al. 1997] R. V. Gurjar, C. R. Pradeep, and A. R. Shastri, "On rationality of logarithmic $\mathbb{Q}$-homology planes, II", *Osaka J. Math.* **34**:3 (1997), 725–743. MR 99i:14038 Zbl 0890.14022

[Iitaka 1982] S. Iitaka, *Algebraic geometry*: *an introduction to birational geometry of algebraic varieties*, Graduate Texts in Math. **76**, Springer, New York, 1982. MR 84j:14001 Zbl 0491.14006

[Koras and Russell 1999] M. Koras and P. Russell, "$\mathbb{C}^*$-actions on $\mathbb{C}^3$: the smooth locus of the quotient is not of hyperbolic type", *J. Algebraic Geom.* **8**:4 (1999), 603–694. MR 2000g:14067

[Koras and Russell 2007] M. Koras and P. Russell, "Contractible affine surfaces with quotient singularities", *Transform. Groups* **12**:2 (2007), 293–340. MR 2008m:14008 Zbl 1124.14050

[Miyanishi 2001] M. Miyanishi, *Open algebraic surfaces*, CRM Monograph Series **12**, Amer. Math. Soc., Providence, RI, 2001. MR 2002e:14101 Zbl 0964.14030

[Miyanishi and Sugie 1991] M. Miyanishi and T. Sugie, "Homology planes with quotient singularities", *J. Math. Kyoto Univ.* **31**:3 (1991), 755–788. MR 92g:14034 Zbl 0790.14034

[Miyanishi and Tsunoda 1984] M. Miyanishi and S. Tsunoda, "Logarithmic del Pezzo surfaces of rank one with noncontractible boundaries", *Japan. J. Math.* (*N.S.*) **10**:2 (1984), 271–319. MR 88b:14030 Zbl 0596.14024

[Palka 2008] K. Palka, "Classification of singular $\mathbb{Q}$-homology planes, I: Structure and singularities", preprint, 2008. To appear in Israel J. Math. arXiv 0806.3110

[Palka 2011a] K. Palka, "Exceptional singular $\mathbb{Q}$-homology planes", *Ann. Inst. Fourier* (*Grenoble*) **61**:2 (2011), 745–774. MR 2895072 Zbl 05973460 arXiv 0909.0772

[Palka 2011b] K. Palka, "Recent progress in the geometry of $\mathbb{Q}$-acyclic surfaces", pp. 271–287 in *Affine algebraic geometry*: *the Russell Festschrift* (Montréal, 2009), edited by D. Daigle et al., CRM Proc. Lecture Notes **54**, Amer. Math. Soc., Providence, RI, 2011. MR 2768645 Zbl 05994629 arXiv 1003.2395

[Palka and Koras 2010] K. Palka and M. Koras, "Singular $\mathbb{Q}$-homology planes of negative Kodaira dimension have smooth locus of non-general type", preprint, 2010. To appear in *Osaka J. Math.* arXiv 1001.2256

[Prill 1967] D. Prill, "Local classification of quotients of complex manifolds by discontinuous groups", *Duke Math. J.* **34**:2 (1967), 375–386. MR 35 #1829 Zbl 0179.12301

[Zaĭdenberg 1987] M. G. Zaĭdenberg, "Изотривиальные семейства кривых на аффинных поверхностях и характеризация аффинной плоскости", *Izv. Akad. Nauk SSSR Ser. Mat.* **51**:3 (1987), 534–567. Translated as "Isotrivial families of curves on affine surfaces and characterization of the affine plane" in *Math. USSR Izv.* **30**:3 (1988), 503–532. MR 88k:14021 Zbl 0666.14018

Received May 11, 2011. Revised September 14, 2011.

KAROL PALKA

INSTITUTE OF MATHEMATICS
UNIVERSITY OF WARSAW
UL. BANACHA 2
02-097 WARSAW
POLAND

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
UL. ŚNIADECKICH 8
00-956 WARSAW
POLAND

palka@impan.pl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.459

A DYNAMICAL INTERPRETATION OF THE PROFILE CURVE OF CMC TWIZZLER SURFACES

OSCAR M. PERDOMO

Delaunay showed in 1841 that any surface of revolution of constant mean curvature in $\mathbb{R}^3$ has as its profile curve a roulette — specifically, the curve described by the focus of a quadric rolling on a line. Here we introduce a notion similar to the roulette that we call the *treadmill sled*, and we use it to provide a dynamical interpretation for the profile curves of *twizzlers* — helicoidal surfaces of nonzero constant mean curvature.

The treadmill sled is connected with a change of variables that allows us to solve the ordinary differential equation that produces twizzlers in a fairly easy way. This allows us to prove that all twizzlers are isometric to Delaunay surfaces; this is similar to work done by do Carmo and Dajczer.

We also provide a moduli space for twizzlers and Delaunay surfaces that shows the connection of each surface with its dynamical interpretation, and we explicitly show the foliation of our moduli space by curves of locally isometric CMC "associated surfaces" analogous to the well-known helicoid-to-catenoid deformation. Our dynamical interpretation for twizzlers also allows us to naturally define the notion of a *fundamental piece* of the profile curve of a twizzler, which yields the fact that, whenever a twizzler is not properly immersed, it is dense in the region bounded by two concentric cylinders if the twizzler does not contain the axis of symmetry, or dense in the region bounded by a cylinder otherwise.

Using the change of coordinates induced by the notion of the treadmill sled, we also provide a dynamical interpretation for helicoidal surfaces with constant Gauss curvature, and we find an easy way to describe Delaunay surfaces by a relatively simple first integral.

1. Introduction

Delaunay [1841] showed that if one rolls a conic section on a line in a plane and then rotates about that line the trace of a focus, one obtains a surface of revolution of constant mean curvature (CMC). When the conic is a parabola we obtain a catenoid; when the conic is an ellipse, the surface is embedded and it is called an

MSC2010: 53A10, 53C42.
Keywords: twizzler, constant mean curvature, helicoidal surfaces, Delaunay surfaces.

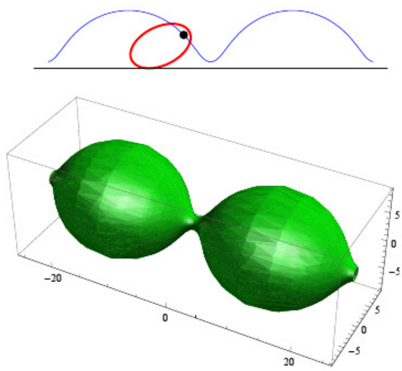


Figure 1. Dynamic interpretation of the profile curve of an unduloid.

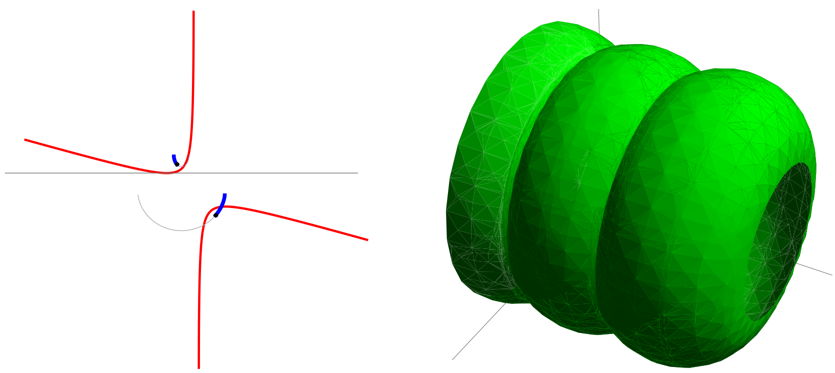


Figure 2. A nodoid and the construction of its profile curve.

unduloid; and when the conic is a hyperbola the surface is not embedded and it is called a *nodoid*. Unduloids and nodoids are called *Delaunay surfaces*. Figure 1 illustrates the relation between the ellipse and the trace of its focus. Notice that only one focus is used to get the curve that needs to get rotated in order to generate an unduloid. Figure 2 illustrates the relation between the hyperbola and the trace of its foci. Notice that both foci are used to get the curve that needs to get rotated in order to generate a nodoid.

Using the integrability of the Gauss equation and the Mainardi–Codazzi equation, Lawson [1970] showed that for any immersion $f_0 : U \to \mathbb{R}^3$ with constant mean curvature H defined in a simple connected surface U, there exists a 2π-periodic 1-parametric family of immersions $\{f_\theta : U \to \mathbb{R}^3 : \theta \in [0, 2\pi]\}$ with constant mean curvature H and with the same induced metric. This family is called the 2π-periodic isometric family *associated to* f_0.

Remark 1.1. The map $\theta \to f_\theta$ is continuous with respect to the parameter θ.

We can see this family of associated surfaces in the well-known deformation from a helicoid to a catenoid. See Figure 3.

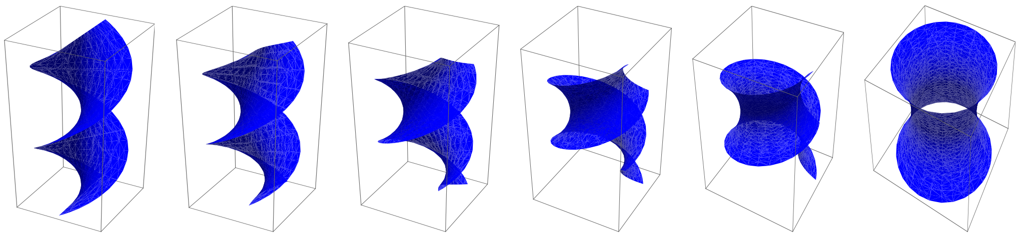


Figure 3. All these surfaces are isometric.

In this particular helicoid-to-catenoid deformation, the helicoid corresponds to $\theta = 0$ and the catenoid corresponds to $\theta = \pi/2$. The images in Figure 3 were taken by substituting $\theta = 0, \pi/10, \pi/5, 3\pi/10, 2\pi/5, \pi/2$ in the parametrization

$$\phi_\theta(u, v) = \big(\cos\theta \sinh v \sin u + \cos u \sin\theta \cosh v,\\ \cosh v \sin u \sin\theta - \cos u \cos\theta \sinh v,\ u\cos\theta + v\sin\theta\big).$$

A direct verification shows that the first and second fundamental form of ϕ_θ are given by

$$E = G = \cosh^2 v, \quad F = 0, \quad e = -\sin\theta, \quad f = \cos\theta, \quad g = \sin\theta,$$

from which we can infer that indeed all the elements in this family of surfaces are isometric. It is not difficult to show that the surfaces from $\theta = \pi/2$ to π are, up to a rigid motion, in the Euclidean space, the same as the surfaces from $\theta = 0$ to $\pi/2$. In this way, up to a rigid motion, all the surfaces in the 2π-periodic Lawson family of isometric surfaces to a helicoid are contained in those surfaces from $\theta = 0$ to $\theta = \pi/2$. We see in this paper that something similar happens for the isometric associated family to a Delaunay surface.

A surface is called helicoidal with pitch $h \in (-\infty, \infty)$ if it is invariant under the group $g_t : \mathbb{R}^3 \to \mathbb{R}^3$ of rigid motions

$$g_t(x, y, z) = \big(x\cos t + y\sin t, -x\sin t + y\cos t, z + ht\big).$$

When $h = 0$, the group g_t becomes a group of rotations and the helicoidal surfaces become surfaces of revolution. A *twizzler* is an immersion of the form

$$\phi(s, t) = \big(x(s)\cos(wt) + z(s)\sin(wt), t, -x(s)\sin(wt) + z(s)\cos(wt)\big), \tag{1-1}$$

with constant mean curvature. Assume that the curve $(x(s), z(s))$ is parametrized by arc length and call it the *profile curve* of the twizzler. Notice that twizzlers correspond to those helicoidal CMC surfaces with nonzero pitch. Geometrically, up to a rotation about the origin, the profile curve of a twizzler is the intersection of the surface with a plane perpendicular to the axis of symmetry. Here we give an interpretation of the profile curve of twizzlers similar to the interpretation for the profile curves of Delaunay surfaces.

To do this, we introduce an operator taking curves into curves (like the roulette operator), which we call the *treadmill sled*. Given a curve α, we imagine a movable plane supporting α rigidly. The trace of the origin of this plane on a stationary plane will be the new curve β, the treadmill sled of α. We now describe the motion of α (and its supporting plane).

First, we choose a point of α and place it at the origin of the fixed plane, so that α has a horizontal tangent there — the x-axis of the fixed plane. Then we move the supporting plane of α in such a way that α always remains tangent to the x-axis of the fixed plane at the origin. (Another way of thinking of this motion is to imagine a treadmill placed under, and aligned with, the x-axis of the fixed plane. The curve α rolls on the treadmill, always keeping one of its points at the origin.)

As already explained, β is described by the positions of the origin of α's supporting plane in this process. Obviously, the choice of the moving plane's origin plays an important role in this definition. For example, if α is a circle of radius R, its treadmill sled is just a point if the center of α is the origin and it is a circle of radius r with center at $(0, R)$ if the center of the circle is at a distance r from the origin.

Figure 4 shows the treadmill sled of an ellipse with center at the origin. The dot represents the center of the ellipse. (The Electronic Supplement to this article shows this example in motion.)

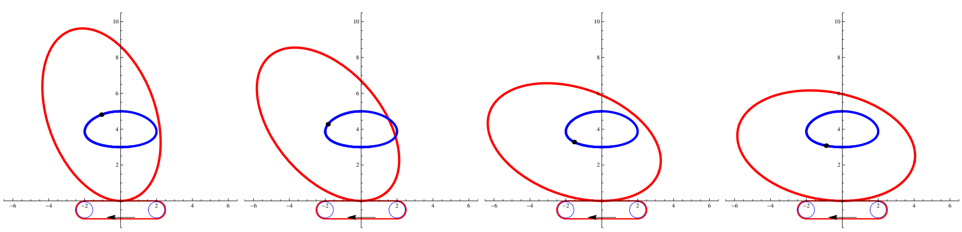


Figure 4. Treadmill sled of an ellipse centered at the origin.

We prove that a parametrization of the treadmill sled of an arc-length parametrized curve $(x(s), z(s))$ is given by $(\xi(s), \xi_2(s))$, where $(\xi(s), \xi_2(s))$ are the coordinates of the vector $(x(s), z(s))$ with respect to the orthonormal basis

$$\{(x'(s), z'(s)), (-z'(s), x'(s))\}.$$

It turns out that this treadmill sled notion is linked with the change of variables $x(s)$, $z(s)$ to the variables $\xi_1(s)$, $\xi_2(s)$, which ends up being very convenient for

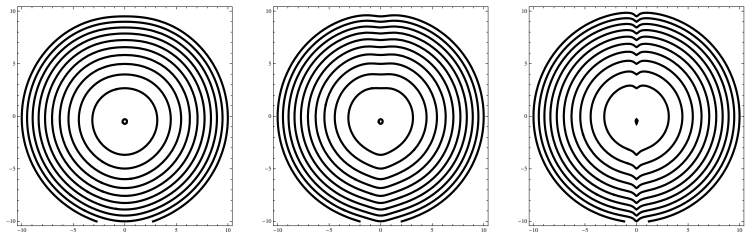


Figure 5. Contours, for different values of w, of the integral function $h_w(x, y) = x^2 + y^2 + y/\sqrt{1 + w^2x^2}$.

the study of helicoidal surfaces. This paper shows some of the applications of this change of variables. We can relatively easily solve the ODE that generates twizzlers, and as a bonus, we find a dynamical interpretation for their profile curve similar to the dynamical interpretation of the profile curve of Delaunay surfaces using conics. For twizzlers we do not use conics, but rather the level sets of the function

$$h_w(x, y) = x^2 + y^2 + \frac{y}{\sqrt{1 + w^2x^2}},$$

where w is a constant. It is not difficult to check that the range of the function h_w is the interval $[-\frac{1}{4}, \infty)$, that $h_w^{-1}(-\frac{1}{4}) = \{(0, -\frac{1}{2})\}$, that every $M > -\frac{1}{4}$ is a regular value of h_w, and that $h_w^{-1}(M)$ is a closed simple curve. We refer to these level sets as *heart-shaped* curves. Figure 5 shows some of them.

Let us denote the origin of the profile curve of a twizzler by O; that is, O is the intersection of the plane that contains the profile curve with the axis symmetry of the twizzler. We prove that the level sets of the function h_w are first integrals of the ODE for twizzlers with CMC 1 written in the coordinates ξ_1 and ξ_2, and therefore geometrically we can say that if we place the profile curve of a twizzler on a treadmill located at the origin and oriented in the positive direction of the x-axis, then the trace of the point O is a heart-shaped curve. In other words, the treadmill sled of the profile curve is a heart-shaped curve. It can be shown that the inverse of the treadmill sled of a curve is unique up to a rotation about the origin. Therefore we have a one-to-one correspondence between twizzlers with CMC 1 and the level sets of the function h_w. In this way, we can use the two parameters w and M that define the heart-shaped curves to describe twizzlers with CMC 1. Once we have all the twizzlers with CMC 1 described in terms of the treadmill sled of their profile curves, we explicitly describe which twizzlers are in the same associated family of isometric surfaces. Surprisingly for the author, the proof only uses the basic fact that, since Gauss curvatures are invariant under isometries, the quotient between the maximum and minimum of the Gauss curvature is the same for two isometric surfaces. An interesting fact that showed up is that in each one of these families of isometric associated surfaces, there is a twizzler that contains the axis of symmetry. Since such twizzlers are unique in each family and there is an easy formula that relates them with the isometric nodoid and unduloid, we call these twizzlers *special twizzlers*.

Lawson [1970] showed examples of helicoidal surfaces with nonzero pitch and constant mean curvature by proving that the family of CMC surfaces associated to a Delaunay surface is made out of helicoidal surfaces. It was known for a long time [Graustein 1935] that all the isometric surfaces in the associated family of a catenoid are helicoidal surfaces, and also that every helicoidal minimal surface belongs to the associated family of isometric surfaces of a catenoid. This result was generalized by do Carmo and Dajczer [1982] (see also [Haak 1998]), who showed that every helicoidal surface with CMC is in the associated family of a Delaunay surface. Do Carmo and Dajczer provided explicit parametrizations for almost all helicoidal surfaces with CMC. As we pointed out before, we show here that there are as many of these surfaces as unduloids, or as many as nodoids, by proving that there is one in each associated family of a Delaunay surface. The unduloids admit in their isometry group, besides the rotational symmetries, a discrete group of translations. This translational group shows up because the profile curve is periodic, and this periodicity happens because the profile curve is generated by an ellipse, which is a closed curve. The new dynamical interpretation for twizzlers allows us to easily visualize that, besides the helicoidal symmetry, twizzlers are invariant under a group of rotations about the axis of symmetry because the treadmill sled of their profile curve is a closed curve — a heart-shaped curve. If we define the fundamental piece of a twizzler as a connected part of the profile curve with the property that, when placed on a treadmill, the point O traces a heart-shaped curve exactly once, then we have that the whole profile curve is a union of fundamental pieces. Two fundamental pieces differ by a rotation about the origin, and when the angle made by the rays that connect the initial and final point of a fundamental piece is a rational multiple of 2π, then the whole profile curve is the union of only finitely many fundamental pieces, and therefore the twizzler is properly immersed. Otherwise, the twizzler is dense in either the region bounded by two cylinders or the region inside a cylinder. For twizzlers that do not contain the axes of symmetry, this property was shown in [Hitt and Roussos 1991].

2. The treadmill sled of a curve

According to the description given in the introduction, we define the treadmill sled of an arc-length parametrized curve $\alpha : [a, b] \to \mathbb{R}^2$ as

$$\mathrm{TS}(\alpha) = \big\{ T_s(0,0) : T_s \text{ is an oriented isometry of } \mathbb{R}^2, \\ T_s(\alpha(s)) = (0,0),\ \text{and } dT_s\alpha'(s)) = (1,0) \big\}.$$

As the following theorem shows, finding a parametrization for the treadmill sled of a curve is not difficult.

Theorem 2.1. *If* $\alpha(s) = (x(s), z(s))$ *is a curve parametrized by arc length and*

$$\xi_1(s) = x(s)\,x'(s) + y(s)\,y'(s) \quad \textit{and} \quad \xi_2(s) = -x(s)\,y'(s) + y(s)\,x'(s),$$

then the treadmill sled of the curve α *is* $-(\xi_1(s), \xi_2(s))$.

Proof. Let $\theta(s)$ be such that $\alpha'(s) = \big(\cos\theta(s), \sin\theta(s)\big)$. A direct computation shows that the transformation

$$\begin{aligned} T_s(X, Y) = \big(&\cos\theta(s)\,X + \sin\theta(s)\,Y, -\sin\theta(s)\,X + \cos\theta(s)\,Y\big) \\ &- \big(\cos\theta(s)\,x(s) + \sin\theta(s)\,y(s), -\sin\theta(s)x(s) + \cos\theta(s)y(s)\big) \end{aligned}$$

is the only oriented isometry of $\mathbb{R}$ that takes the point $\alpha(s)$ to the origin and for which $dT_{\alpha(s)}(\alpha'(s)) = (1, 0)$. From the definition of $TS(\alpha)$, it follows that

$$T_s(0, 0) = -\big(\cos\theta(s)\,x(s) + \sin\theta(s)\,y(s), -\sin\theta(s)\,x(s) + \cos\theta(s)\,y(s)\big)$$

must be a point in the treadmill sled of α. When we allow s to move through the domain of α we obtain the desired parametrization of $TS(\alpha)$. □

Remark 2.2. It easily follows, either from the geometric definition of treadmill sleds or from Theorem 2.1, that the maximum distance from the origin to a curve α equals the maximum distance from the origin to its treadmill sled. Likewise, the minimum distance from the origin to a curve α equals the minimum distance from the origin to its treadmill sled.

3. Treadmill sled coordinates on twizzlers: solution of the ODE

The following two results provide a solution for the ODE coming from the problem of finding all twizzlers with CMC 1. As mentioned before, the ODE is greatly simplified when we use treadmill sled coordinates.

Proposition 3.1. *The immersions given by* (1-1) *have mean curvature* 1 *if and only if the functions* ξ_1 *and* ξ_2 *defined in Theorem 2.1 satisfy the ordinary differential equations* $\xi_1'(s) = f_1(\xi_1(s), \xi_2(s))$ *and* $\xi_2'(s) = f_2(\xi_1(s), \xi_2(s))$, *where*

$$\begin{aligned} f_1(x_1, x_2) &= \frac{-w^2x_2 + 2(1 + w^2x_1^2)^{3/2}}{1 + w^2(x_1^2 + x_2^2)}\,x_2 + 1, \\ f_2(x_1, x_2) &= \frac{w^2x_2 - 2(1 + w^2x_1^2)^{3/2}}{1 + w^2(x_1^2 + x_2^2)}\,x_1. \end{aligned} \tag{3-1}$$

Moreover, the function h_w *is constant along all solutions* $(\xi_1(s), \xi_2(s))$.

Proof. Since the curve $(x(s), z(s))$ is parametrized by arc length, we can consider a function $\theta(s)$ such that

$$x'(s)=\cos\theta(s) \quad \text{and} \quad z'(s)=\sin\theta(s).$$

Let us define the functions $\xi_1(s)$ and $\xi_2(s)$ by

$$\xi_1 = x\cos\theta + z\sin\theta \quad \text{and} \quad \xi_2=-x\sin\theta+z\cos\theta.$$

A direct verification shows that

$$x=\xi_1\cos\theta-\xi_2\sin\theta,\quad z=\xi_1\sin\theta+\xi_2\cos\theta,\quad \theta'=x'z''-z'x''. \tag{3-2}$$

Moreover, it is not difficult to check that

$$\xi_1'=\theta'\xi_2+1,\quad \xi_2'=-\theta'\xi_1,\quad \xi_1^2+\xi_2^2=x^2+z^2.$$

A direct verification shows that the first fundamental form of ϕ is given by

$$E=\langle\phi_s,\phi_s\rangle=1,\quad F=\langle\phi_s,\phi_t\rangle=w(zx'-xz')=w\xi_2,$$
$$G=\langle\phi_t,\phi_t\rangle=1+w^2(x^2+z^2)=1+w^2(\xi_1^2+\xi_2^2),$$

and therefore,

$$EG-F^2=1+w^2(\xi_1^2+\xi_2^2)-w^2\xi_2^2=1+w^2\xi_1^2.$$

The Gauss map of the immersion ϕ is given by $\nu=\frac{1}{\sqrt{EG-F^2}}\,\phi_s\times\phi_t$. A direct verification shows that

$$\nu(s,t)=\frac{1}{\sqrt{1+w^2\xi_1^2(s)}}\big(\sin(wt-\theta(s)),\,w\xi_1,\,\cos(wt-\theta(s))\big).$$

A direct verification shows that the second fundamental form of ϕ is given by

$$e=\langle\phi_{ss},\nu\rangle=\frac{\theta'}{\sqrt{1+w^2\xi_1^2}},\qquad f=\langle\phi_{st},\nu\rangle=\frac{-w}{\sqrt{1+w^2\xi_1^2}},$$
$$g=\langle\phi_{tt},\nu\rangle=\frac{-w^2\xi_2}{\sqrt{1+w^2\xi_1^2}}.$$

Therefore, if we assume that the mean curvature $\frac{eG-2fF+gE}{2(EG-F^2)}$ equals 1, we obtain the ODE

$$\theta'=\frac{-w^2\xi_2+2(1+w^2\xi_1^2)^{3/2}}{1+w^2(\xi_1^2+\xi_2^2)}. \tag{3-3}$$

Using this expression for θ' in the equations $\xi_1'=\theta'\xi_2+1$ and $\xi_2'=-\theta'\xi_1$, we obtain that ξ_1 and ξ_2 satisfy the ODE

$$\xi_1'=f_1(\xi_1,\xi_2),\quad \xi_2'=f_2(x_1,x_2), \tag{3-4}$$

where

$$f_1(x_1, x_2) = \frac{-w^2 x_2 + 2(1 + w^2 x_1^2)^{3/2}}{1 + w^2(x_1^2 + x_2^2)} x_2 + 1,$$

$$f_2(x_1, x_2) = \frac{w^2 x_2 - 2(1 + w^2 x_1^2)^{3/2}}{1 + w^2(x_1^2 + x_2^2)} x_1.$$

A direct verification shows that if we define $h_w : \mathbb{R}^2 \to \mathbb{R}$ as

$$h_w(x_1, x_2) = \frac{x_2}{\sqrt{1 + w^2 x_1^2}} + x_1^2 + x_2^2,$$

then h_w is a first integral of the ODE for ξ_1 and ξ_2; that is, for any solution $\xi_1(s)$ and $\xi_2(s)$ of this system, we have that $h_w(\xi_1(s), \xi_2(s)) = M$, where M is a constant. This completes the proof of the proposition. □

As a consequence of the previous proposition, we have:

Theorem 3.2. *The treadmill sled of the profile curve of a twizzler with constant mean curvature* 1 *is a heart-shaped curve* $-h_w^{-1}(M)$ *for some* $M \geq -\frac{1}{4}$. *The value* $M = -\frac{1}{4}$ *is achieved by a cylinder of radius* $\frac{1}{2}$.

Figure 6 shows the profile curve of a twizzler and the heart-shaped curve associated with it. Figure 7 illustrates that, for this twizzler, the treadmill sled of the profile curve is indeed the negative of the heart-shaped curve.

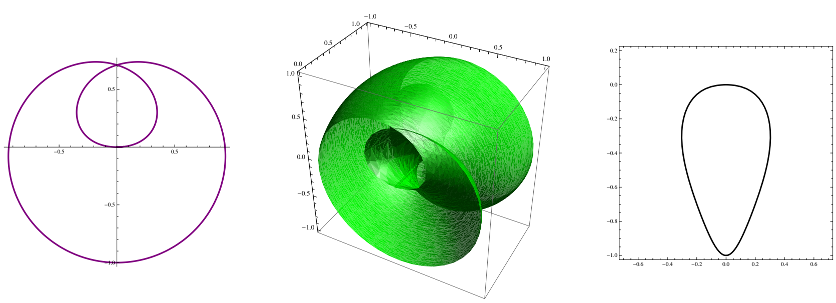


Figure 6. Profile curve, surface, and heart-shaped curve associated to a twizzler.

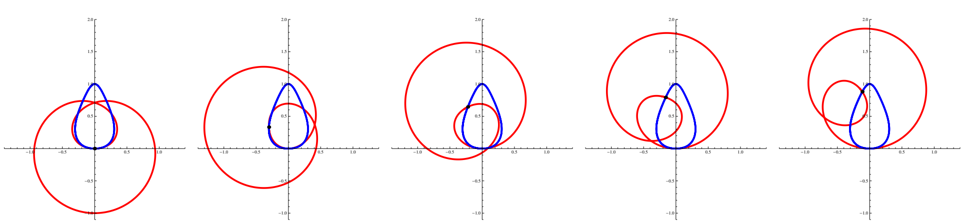


Figure 7. Treadmill sled of the profile curve of a twizzler.

4. Treadmill sled coordinates on flat surfaces

The following theorem gives us another application of the treadmill sled.

Theorem 4.1. *A surface of the form* (1-1) *is flat if and only if either the treadmill sled of the profile curve is a point in the y-axis other than the origin* (*in this case the surface is a cylinder*) *or the treadmill sled of the profile curve is contained in a vertical half-line that starts at a point in the x-axis other than the origin. The functions x and z can be explicitly computed*:

$$x(s) = \tfrac{1}{2}\cos\frac{2\sqrt{as+b}}{a} + \sqrt{as+b}\,\sin\frac{2\sqrt{as+b}}{a},$$
$$z(s) = \sqrt{as+b}\,\cos\frac{2\sqrt{as+b}}{a} - \tfrac{1}{2}\sin\frac{2\sqrt{as+b}}{a}.$$

Proof. If we define the functions θ, ξ_1, and ξ_2 as in the previous theorem, then the equation for Gauss curvature equal to zero, $eg - f^2 = 0$, reduces to $\theta' = -1/\xi_2$. Substituting this equation in the equations $\xi_1' = \theta'\xi_2 + 1$ and $\xi_2' = -\theta'\xi_1$, we obtain that ξ_1 and ξ_2 satisfy the ODE

$$\xi_1' = 0, \qquad \xi_2' = \frac{\xi_1}{\xi_2}. \tag{4-1}$$

It follows that $\xi_1(s) = a/2$ for some real number a. If $a = 0$, then ξ_2 is also a constant other than zero, and the surface ϕ is a cylinder. In the case that a is not zero, then $\xi_2 = \pm\sqrt{as+b}$ and $\theta(s) = \mp 2\sqrt{as+b}/a$. This completes the proof. □

Figure 8 illustrates that the treadmill sled of the profile curve of a flat helicoidal surface is a vertical half-line.

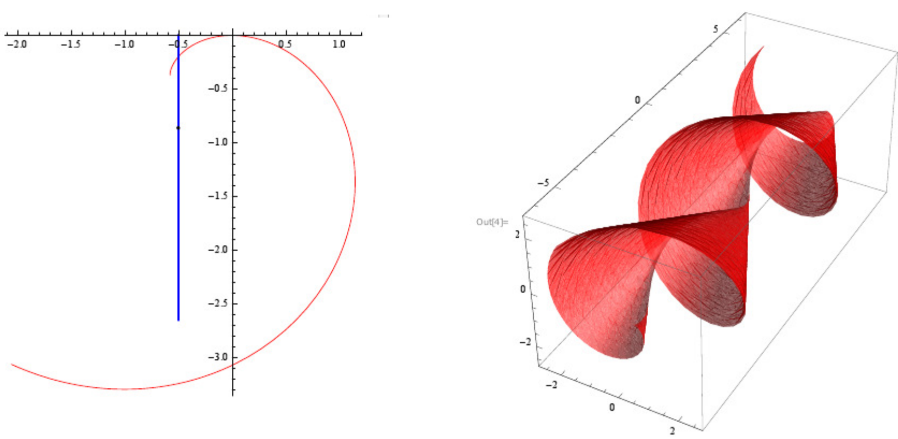


Figure 8. A surface with helicoidal symmetry is flat when its treadmill sled is a vertical half-line.

5. Treadmill sled coordinates on Delaunay surfaces

Extending the parallel between twizzlers and Delaunay surfaces, we now describe all Delaunay surfaces with CMC 1 using treadmill sled coordinates, and we provide an expression for the quotient of the maximum and minimum values of the Gauss curvature. We use this ratio to find out which unduloid-nodoid pairs are isometric.

Theorem 5.1. *For every nonzero real number $M \in (-\frac{1}{4}, \infty)$, the Delaunay surface $\mathbb{D}(M)$ generated by the conic $\{(x, y) : 4x^2 - y^2/M = 1\}$ has constant mean curvature* 1. *The quotient between the maximum value of the Gauss curvature and the minimum value of the Gauss curvature of $\mathbb{D}(M)$ is given by*

$$rs(M) = -\left(\frac{1-\sqrt{1+4M}}{1+\sqrt{1+4M}}\right)^2.$$

Proof. Let us assume that $\mathbb{D}(M)$ is parametrized as

$$\phi(s, t) = \big(x(s), z(s)\sin t, z(s)\cos t\big),$$

where the profile curve $(x(s), z(s))$ is parametrized by arc length. A direct verification shows that if $\theta(s)$ is a continuous function such that $x'(s) = \cos\theta(s)$ and $z'(s) = \sin\theta(s)$, then the mean curvature of $\mathbb{D}(M)$ is

$$\tfrac{1}{2}\left(\theta' - \frac{\cos\theta(s)}{z(s)}\right).$$

Since the mean curvature of $\mathbb{D}(M)$ is 1, the functions $\theta(s)$ and $z(s)$ satisfy

$$\theta' = 2 + \frac{\cos\theta}{z} \quad \text{and} \quad z' = \sin\theta.$$

This ODE has as a first integral the function $h(z, \theta) = z(\cos\theta + z)$. Recall that the function $z(s)$ is always positive. Since the minimum of the function h is $-\frac{1}{4}$, it follows that there exists a nonzero constant $k > -\frac{1}{4}$ such that $h(z(s), \theta(s)) = k$. When $k < 0$, the level sets of $h(z, \theta)$ are bounded, and therefore $\mathbb{D}(M)$ represents an unduloid. When $k > 0$, the level sets are not bounded, and $\mathbb{D}(M)$ represents a nodoid. In any case, the z-values of the level sets of $h(z, \theta)$ are bounded. A direct computation shows that the maximum and minimum of the z-values of the level set $h(z, \theta) = k$ are

$$\frac{1+\sqrt{1+4k}}{2} \quad \text{and} \quad \left|\frac{1-\sqrt{1+4k}}{2}\right|.$$

We can prove that k must be equal to M by comparing these critical values of $z(s)$ with the maximum and the minimum values of the profile curve viewed as the trace of the focus of a conic when it is rolled on a line. A direct computation shows that the Gauss curvature is $-(\theta' \cos\theta)/z$, and since $\theta' = 2 + (\cos\theta)/z$, the Gauss curvature reduces to

$$-\frac{\cos\theta\,(2z+\cos\theta)}{z^2}.$$

Using the Lagrange multiplier method, we see that the maximum and the minimum of the Gauss curvature subject to the constraint $h(z,\theta) = k$ are

$$\frac{4\sqrt{1+4k}}{(1+\sqrt{1+4k})^2} \quad \text{and} \quad -\frac{4\sqrt{1+4k}}{(-1+\sqrt{1+4k})^2},$$

respectively. It follows that the quotient between the maximum of the Gauss curvature and the minimum of the Gauss curvature is

$$-\left(\frac{-1+\sqrt{1+4k}}{1+\sqrt{1+4k}}\right)^2.$$

Since $k = M$, the theorem follows. □

The function rs defines a bijection between the intervals $(-\frac{1}{4}, 0)$ and $(0, 1)$, and it also defines a bijection between the intervals $(0, \infty)$ and $(0, 1)$. On the other hand, each unduloid is isometric to a nodoid [do Carmo and Dajczer 1982]. As a consequence of Theorem 5.1, we have:

Corollary 5.2. *Two Delaunay surfaces with CMC* 1 *are isometric if and only if the quotients of the maximum and minimum values of the Gauss curvatures are the same. In particular, for any* $u \in (0, 1)$, *the unduloid* $\mathbb{D}\big(-\sqrt{u}/(1+\sqrt{u})^2\big)$ *is isometric to the nodoid* $\mathbb{D}\big(\sqrt{u}/(1-\sqrt{u})^2\big)$.

6. Moduli space for twizzlers

If we exclude the cylinder and the value $M = -\frac{1}{4}$, Theorem 3.2 establishes a 1:1 correspondence between pairs (M, w) with $M > -\frac{1}{4}$ and $w > 0$ and twizzlers with mean curvature 1. Therefore, so far we have that the moduli space of all twizzlers with CMC 1 other than the cylinder is the set $\{(M, w) : M > -\frac{1}{4},\ w > 0\}$. In order to visualize better the boundary of the moduli space of twizzlers, we replace the parameter w with the bounded parameter $v = 1/(1+w^2)$. Therefore, the parameter v moves from 0 to 1 when w moves from ∞ to 0. Figure 9 shows pictures from an animation that produces a piece of the twizzler associated with values of M and v. We refer to this twizzler as $\mathfrak{T}(M, v)$ when the dependence of M and v is needed.

In [Perdomo 2011], a formula for the inverse of the treadmill sled of a curve is provided. Therefore we can get a parametrization for all twizzlers if we have a parametrization for all heart-shaped curves.

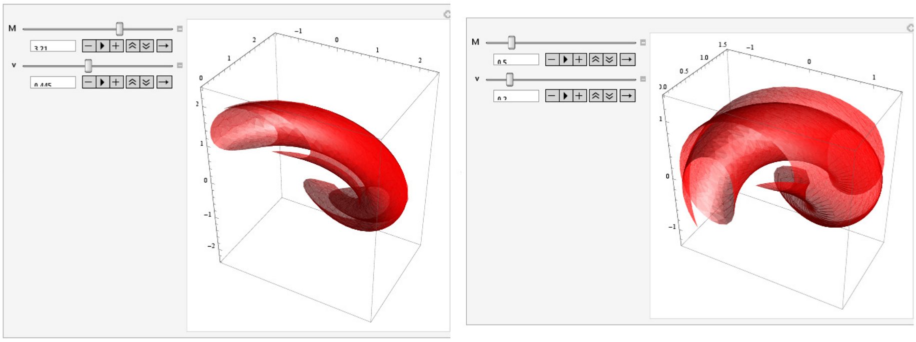


Figure 9. Moduli space of twizzler with mean curvature one.

Lemma 6.1. *For any* $M > -\frac{1}{4}$ *and* $w > 0$, *the curve* $\alpha(t) = (\rho_1(t), \ \rho_2(t))$ *defined on the interval* $[0, 2\pi]$ *and given by*

$$\rho_1(u) = A\cos u \quad \text{and} \quad \rho_2(u) = \frac{-1+\sqrt{1+4M+B\cos^2 u}\,\sin u}{2\sqrt{1+w^2A^2\cos^2 u}},$$

where

$$A = \frac{\sqrt{-1+Mw^2+\sqrt{1+(1+2M)w^2+M^2w^4}}}{\sqrt{2}\,w},$$

$$B = \frac{2+2M^2w^4+w^2+2(Mw^2-1)\sqrt{1+(1+2M)w^2+M^2w^4}}{w^2},$$

is a closed simple regular curve that parametrizes the heart-shaped curve $h_w^{-1}(M)$.

Proof. It is a direct verification. □

Since the maximum and minimum distances from a curve α to the origin agree with the maximum and minimum distances from its treadmill sled to the origin [Perdomo 2011], we have the following proposition.

Proposition 6.2. *The maximum distance from a special twizzler with CMC* 1 *to its axis of symmetry is* 1. *More generally, the maximum and minimum distances from the twizzler* $\mathfrak{T}(M, v)$ *to its axis of symmetry are given by*

$$r_1(M) = \left|\frac{\sqrt{1+4M}-1}{2}\right| \quad \text{and} \quad r_2(M) = \frac{\sqrt{1+4M}+1}{2}.$$

Proof. Since the maximum and minimum distances from a twizzler to its axis of symmetry are the same as the maximum and minimum distances from its profile curve to the origin, we only need to show that for any $M > -\frac{1}{4}$, the minimum and maximum distances from the origin to the heart-shaped curve $h_w^{-1}(M)$ are $r_1(M) = \left|(\sqrt{1+4M}-1)/2\right|$ and $r_2(M) = (\sqrt{1+4M}+1)/2$, respectively. We

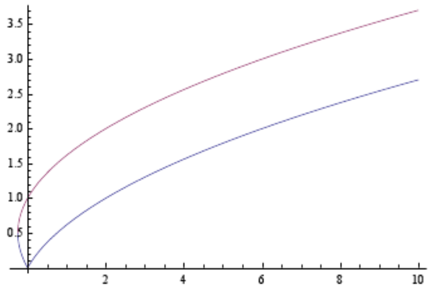


Figure 10. The maximum and minimum distances from the origin to the profile curve of $\mathfrak{T}(M, v)$ change with respect to M.

prove this by using the method of Lagrange multipliers to find the maximum and minimum values of the function $R(x_1, x_2) = x_1^2 + x_2^2$, subject to the restriction $h_w = M$. A direct verification shows that if (x_1, x_2) and λ_1 satisfy the Lagrange multiplier equations

$$\frac{\partial R}{\partial x_1} = \lambda_1 \frac{\partial h_w}{\partial x_1} \quad \text{and} \quad \frac{\partial R}{\partial x_2} = \lambda_1 \frac{\partial h_w}{\partial x_2},$$

then $x_1 = 0$. Once we know that x_1 must be zero, we obtain from the equation $h_w = M$ that x_2 is either $-(\sqrt{1+4M}+1)/2$ or $(\sqrt{1+4M}-1)/2$. Now the result easily follows. □

Remark 6.3. From this proposition we can understand the twizzlers in the moduli space that are near the boundary line $M = -\frac{1}{4}$. Since the limit when M goes to $-\frac{1}{4}$ of the functions $r_1(M)$ and $r_2(M)$ is $\frac{1}{2}$ (see Figure 10), then we have that when M is near $-\frac{1}{4}$, the twizzlers $\mathfrak{T}(M, v)$ are near the cylinder of radius $\frac{1}{2}$.

7. Fundamental piece of the profile curve of a twizzler and the immersed versus dense property

The fact that the treadmill sled of the profile curve of a twizzler is a closed curve allows us to define a *fundamental piece of the profile curve* as a connected piece of profile curve with the property that the treadmill sled motion of this piece goes exactly once over the heart-shaped curve. It is not difficult to see that the whole profile curve is the union of fundamental pieces. Figure 11 shows the fundamental piece of the profile curve of a properly immersed twizzler, along with the whole profile curve made up of four pieces in this case and the graph of the twizzler.

For the sake of comparison, for an unduloid we could define a fundamental piece as the trace of the focus of the ellipse when this ellipse rolls once. It is clear that the whole profile curve is the union of fundamental pieces, and therefore $\mathbb{Z}$ acts on the group of isometries of the unduloid in the form of translations. Theorem 7.2 shows that the group $\mathbb{Z}$ also acts on the set of isometries of twizzlers.

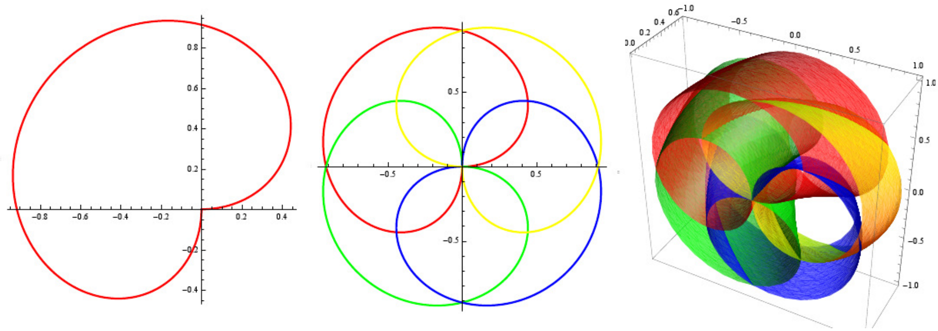


Figure 11. Fundamental piece of the profile curve.

Using the parametrization for the heart-shaped curve in Lemma 6.1, we get the following formula for the length of a fundamental piece of a twizzler. (This formula was used in the production of Figure 11.)

Lemma 7.1. *The length of the fundamental piece of the twizzler* $\mathfrak{T}(M, v)$ *is*

$$\int_0^{2\pi} \sqrt{\lambda/\mu}\, du,$$

where $\lambda(u) = \left(\frac{d\rho_1}{du}\right)^2 + \left(\frac{d\rho_2}{du}\right)^2$ *and* $\mu(u) = f_1^2(\rho_1(u), \rho_2(u)) + f_2^2(\rho_1(u), \rho_2(u))$.

The functions ρ_1, ρ_2, f_1, and f_2 are defined in Lemma 6.1 and Proposition 3.1. Recall that w and v are related by the equation $v = 1/(1 + w^2)$.

Proof. The proof is straightforward, and is actually included in the proof of the next result, Theorem 7.2. □

In [Perdomo 2011] we showed that two curves with the same treadmill sled differ only by a rotation about the origin. With this in mind, we have that two consecutive fundamental pieces of the same twizzler differ by a rotation about the origin, and therefore the whole profile curve is either a closed curve made out of a finite union of fundamental pieces or the union of infinitely many disjoint fundamental pieces. When the latter happens, it is not difficult to see that the profile curve is either dense in a circle or dense in an annulus depending on whether or not the profile curve passes through the origin. In order to better understand this property, given a twizzler, without loss of generality, let us consider a fundamental piece starting at a point p_1 other than the origin and ending in a point p_2. We have that $|p_1| = |p_2|$, so in polar coordinates $p_1 = re^{\theta_1}$ and $p_2 = re^{\theta_2}$. We prove that if $\theta_2 - \theta_1$ is a rational multiple of π, then the profile curve is a closed curve and the twizzler is properly immersed, for otherwise the twizzler is dense in either the region bounded by two concentric cylinders or dense in the region bounded by a cylinder.

The next theorem, along with Theorem 8.2 and Theorem 8.4, gives a precise picture of the moduli space for twizzlers.

Theorem 7.2. *The angle between the final and initial points of a fundamental piece of the twizzler $\mathfrak{T}(M, v)$ is given by $\theta_0 = \int_0^{2\pi} \psi\, du$, where*

$$\psi(u) = \frac{-w^2\rho_2(u) + 2(1+w^2\rho_1^2(u))^{3/2}}{1+w^2(\rho_1^2(u)+\rho_2^2(u))}\sqrt{\frac{\lambda(u)}{\mu(u)}}.$$

$\mathfrak{T}(M, v)$ is invariant under a group of rotations of the form $\{R(n\theta_0) : n \in \mathbb{Z}\}$. If $R(m\theta_0) = R(\theta_0)$ for some integer m, then the twizzler is properly immersed; otherwise it is dense in the interior of a cylinder of radius 1 when $M = 0$, or dense in the region bounded by two concentric cylinders of radii

$$r_1(M) = \left|\frac{\sqrt{1+4M}-1}{2}\right| \quad \textit{and} \quad r_2(M) = \frac{\sqrt{1+4M}+1}{2}$$

when $M \neq 0$. More precisely, we have that $\mathfrak{T}(M, v)$ is a properly immersed surface with a profile curve consisting of b fundamental pieces if and only if $\theta_0 = 2\pi(a/b)$, with a and b positive relatively prime integers. We also have another type of density: the set of points (M, v) associated with properly immersed twizzlers is uncountable and dense.

Proof. That the twizzler is bounded by a cylinder follows from Proposition 6.2. Let us assume that $(x(s), y(s))$ are such that the surface (1-1) has constant mean curvature 1. Since the curve (ρ_1, ρ_2) defined in Lemma 6.1 is regular, we see that

$$\lambda(u) = \left(\frac{d\rho_1}{du}\right)^2 + \left(\frac{d\rho_2}{du}\right)^2$$

is a periodic positive function. Likewise, since $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ only vanish simultaneously at $(x_1, x_2) = (0, -\frac{1}{2})$, we see that

$$\mu(u) = f_1^2(\rho_1(u), \rho_2(u)) + f_2^2(\rho_1(u), \rho_2(u))$$

is a positive periodic function. Notice that $\xi_1(s) = 0$ and $\xi_2(s) = -\frac{1}{2}$ is the only constant solution of the system (3-4). For any other solution, since h_w is a first integral of the system, there exist $M > -\frac{1}{4}$ and a function $\sigma(s)$ such that

$$\xi_1(s) = \rho_1(\sigma(s)) \quad \text{and} \quad \xi_2(s) = \rho_2(\sigma(s))$$

is a solution of the system (3-4). From the equations above, we have

$$\xi_1'(s)^2 + \xi_2'(s)^2 = \lambda(\sigma(s))\,\sigma'(s)^2. \tag{7-1}$$

On the other hand,

$$\xi_1'(s) = f_1(\xi_1(s), \xi_2(s)) = f_1(\rho_1(\sigma(s)), \rho_2(s)),$$
$$\xi_2'(s) = f_2(\xi_1(s), \xi_2(s)) = f_2(\rho_1(\sigma(s)), \rho_2(s)).$$

It follows that σ is either strictly increasing or strictly decreasing; without loss of generality, we can assume that σ is strictly increasing. Therefore we get

$$\sigma'(s) = \sqrt{\frac{\mu(\sigma(s))}{\lambda(\sigma(s))}}.$$

If $\kappa(u)$ is the inverse of the function $\sigma(s)$, we have that

$$\kappa'(u) = \frac{1}{\sigma'(\kappa(u))} = \sqrt{\frac{\lambda(u)}{\mu(u)}}. \tag{7-2}$$

If we change from the variable s to the variable u, that is, if we consider the functions

$$\tilde{\theta}(u) = \theta(\kappa(u)), \qquad \tilde{\xi}_1(u) = \xi_1(\kappa(u)), \qquad \tilde{\xi}_2(u) = \xi_2(\kappa(u)),$$
$$\tilde{x}(u) = x(\kappa(u)), \qquad \tilde{z}(u) = z(\kappa(u)),$$

it follows from (7-2) and (3-3) that $\tilde{\theta}'(u) = \psi(u)$, where

$$\psi(u) = \frac{-w^2\rho_2(u) + 2(1 + w^2\rho_1^2(u))^{3/2}}{1 + w^2(\rho_1^2(u) + \rho_2^2(u))}\sqrt{\frac{\lambda(u)}{\mu(u)}}.$$

Since the right side of this equation is a periodic function with period 2π, it follows by the existence and uniqueness theorem of ODEs that if $\tilde{\theta}(2\pi) = \theta_0$, then for any integer j,

$$\tilde{\theta}(u + 2j\pi) = j\theta_0 + \tilde{\theta}(u). \tag{7-3}$$

Since $\big|(x(s), z(s))\big| = \big|(\xi_1(s), \xi_2(s))\big|$, the piece of profile curve

$$C_{\text{fp}} = C_{\text{fundamental piece}} = \{(\tilde{x}(u), \tilde{z}(u)) : u \in [0, 2\pi]\}$$

also satisfies $r_1(M) = \min\{|q| : q \in C_{\text{fp}}\}$ and $r_2(M) = \min\{|q| : q \in C_{\text{fp}}\}$. Using (3-2) and (7-3), we get

$$\begin{pmatrix} \tilde{x}(u + 2j\pi) \\ \tilde{z}(u + 2j\pi) \end{pmatrix} = R_{\theta_0}^j \begin{pmatrix} \tilde{x}(u) \\ \tilde{z}(u) \end{pmatrix}, \quad \text{where } R_{\theta_0} = \begin{pmatrix} \cos\theta_0 & -\sin\theta_0 \\ \sin\theta_0 & \cos\theta_0 \end{pmatrix}. \tag{7-4}$$

This equality implies that the image of the profile curve can be viewed as the orbit of the group $\{R_{\theta_0}^j\}_{j\in\mathbb{Z}}$ acting on C_{fp}, that is,

$$C = \{(x(t), z(t)) : t \in \mathbf{R}\} = \{R_{\theta_0}^j p : j \in \mathbb{Z} \text{ and } p \in C_{\text{fundamental piece}}\}. \tag{7-5}$$

It follows from this equation that if $\theta_0/2\pi$ is a rational number, then C is a properly immersed curve, and if $\theta_0/2\pi$ is irrational, then C is dense in the annulus

$$\Big\{(x_1, x_2) : r_1(M) \le \sqrt{x_1^2 + x_2^2} \le r_2(M)\Big\}$$

when $M \neq 0$, or dense in the circle of radius 1 when $M = 0$. Therefore, twizzlers with constant mean curvature 1 have the following property: they are properly immersed, or they are dense in the region contained between two concentric cylinders, or they are dense in the interior of a cylinder of radius 1. We can prove that a surface corresponding to an irrational value $\theta_0/2\pi$ is dense by showing that the profile curve is dense, and we can prove that the profile curve is dense by showing that the intersection of this curve with a circle centered at the origin is either the empty set or dense in the circle. The problem of proving this last statement reduces to that of showing that for any irrational number ι, the set $\{\iota - [n\iota] : n \in \mathbb{Z}\}$ is dense in the interval [0, 1], which is a known fact. To finish, we notice that since the function $(x(s), z(s))$ is parametrized by arc length, the length of the fundamental piece is

$$\kappa(2\pi) = \int_0^{2\pi} \sqrt{\lambda(u)/\mu(u)}\, du.$$

Also, since $\tilde{\theta}'(u) = \psi(u)$, we have that

$$\theta_0 = \int_0^{2\pi} \psi(u) du. \qquad \square$$

For twizzlers that do not contain the axis of symmetry, the “properly immersed versus dense” property established in Theorem 7.2 was proved in [Hitt and Roussos 1991]. By numerically solving the equation $\int_0^{2\pi} \psi\, du = 2\pi(a/b)$ in that theorem, we can graph profile curves of twizzlers with any desired property.

In Figure 12, we solve the numerical equation $\int_0^{2\pi} \psi\, du = 2\pi(a/b)$, fixing $M = 0$ and taking several integer values for a and b. Since $M = 0$, these profile curves

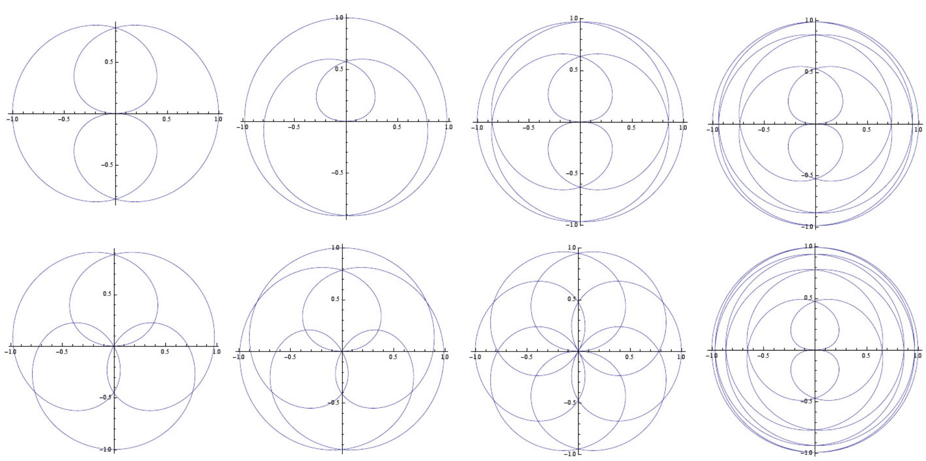


Figure 12. Profile curves of properly immersed twizzlers that contain the axis.

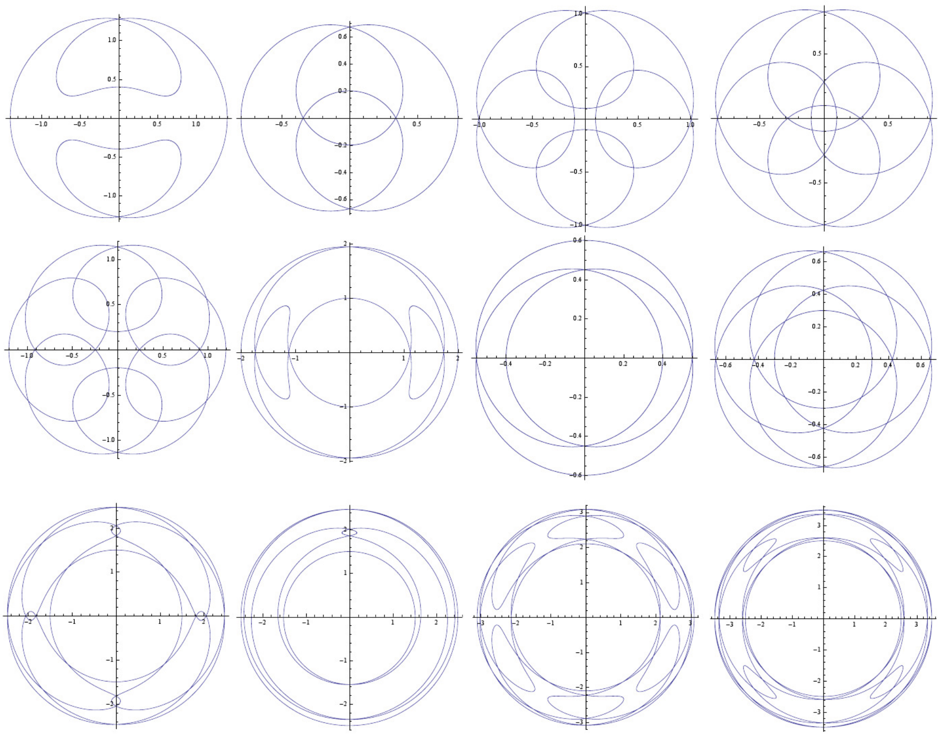


Figure 13. Profile curves of properly immersed twizzlers that do not contain their axis.

represent twizzlers that contain the axis of symmetry. In Figure 13, we take several values for $M \neq 0$ and a and b integers to produce properly immersed twizzlers that do not contain the axis of symmetry. In Figure 14, we take a and b such that a/b is not rational, so that the twizzler is not properly immersed. In Figure 15, we take $a = 5$, $b = 4$ and 4 values of M in order to produce properly immersed examples; we also show the points (M, w) associated with these twizzlers.

8. Isometric associate family of surfaces

As pointed out before, each nodoid is isometric to an unduloid, and therefore we can replace the word Delaunay by either the word unduloid or nodoid in the result proved in [do Carmo and Dajczer 1982]; that is, we can say that each twizzler is isometric to either a nodoid or an unduloid. Another family of surfaces that holds the same property is the set of twizzlers that contain the axis of symmetry, that is, the set of twizzlers corresponding to $M = 0$ in the moduli space. We call these surfaces *special twizzlers* and we denote them by $\mathbb{S}\mathfrak{T}(v)$; that is, $\mathbb{S}\mathfrak{T}(v) = \mathfrak{T}(0, v)$. Due to a singularity problem on the coordinates used so far to study helicoidal

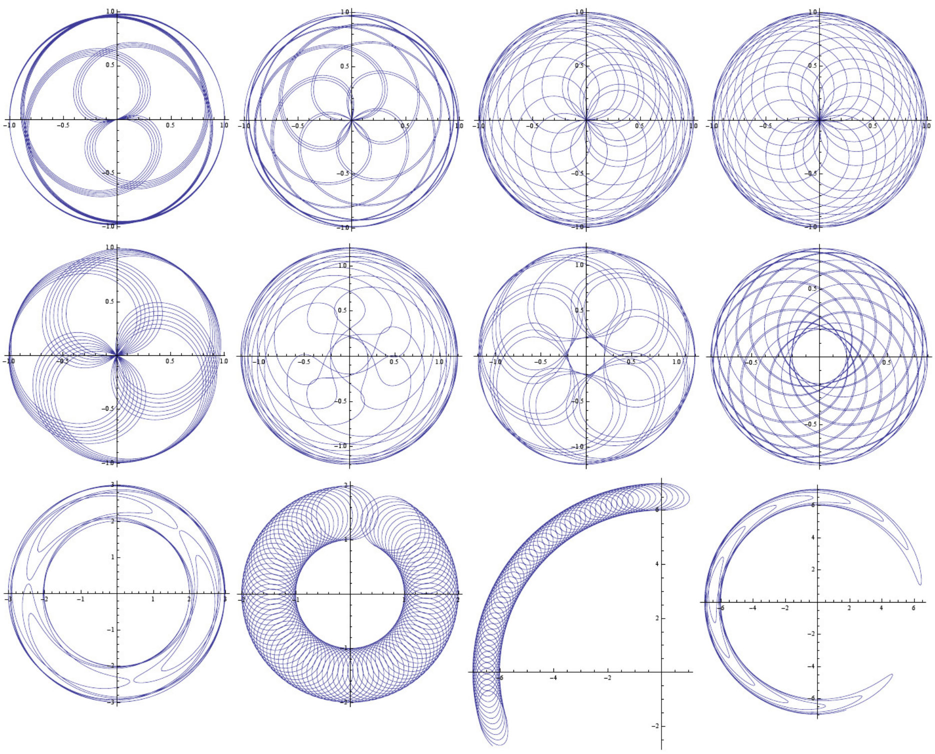


Figure 14. Profile curves of nonproperly immersed twizzlers.

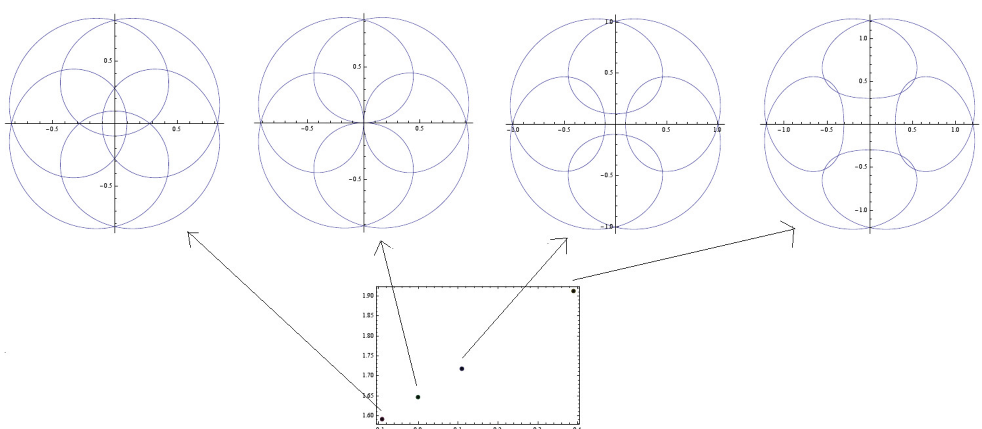


Figure 15. Profile curves of twizzlers consisting of four fundamental pieces and their corresponding values M and w.

surfaces, twizzlers that contain the axis of symmetry have been overlooked until now. The following theorem gives us the quotient of the maximum value and the minimum value of the Gauss curvature for special twizzlers. Figure 16 shows two sets of isometric nodoid-unduloid-special twizzler surfaces.

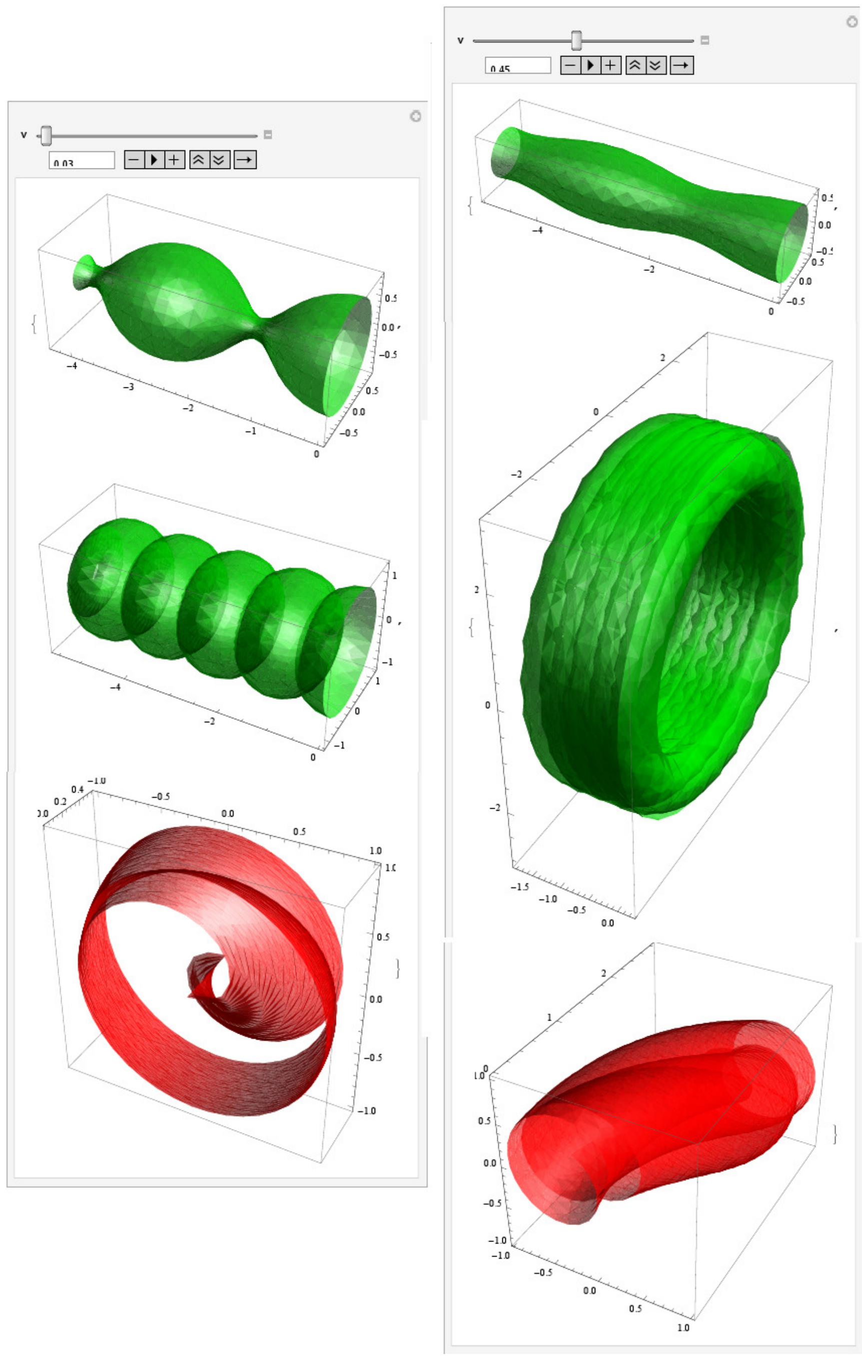


Figure 16. Isometric nodoid, unduloid and special twizzler.

Theorem 8.1. *For every nonzero real number* $v \in (0, 1)$, *the quotient between the maximum value of the Gauss curvature and the minimum value of the Gauss curvature of the special twizzler surface* $\mathbb{S}\mathfrak{T}(v)$ *is* $-v$. *Moreover*, $\mathbb{S}\mathfrak{T}(v)$ *is isometric to the unduloid* $\mathbb{D}\big(-\sqrt{v}/(1+\sqrt{v})^2\big)$ *and the nodoid* $\mathbb{D}\big(\sqrt{v}/(1-\sqrt{v})^2\big)$.

Proof. The proof is contained in the proof of Theorem 8.2. □

We can generalize Theorem 8.1 as follows:

Theorem 8.2. *If* $v = 1/(1+w^2)$, *then the quotient between the maximum and minimum values of the Gauss curvature of the twizzler surface* $\mathfrak{T}(M, v)$ *is*

$$-\frac{2+(1+2M-\sqrt{1+4M})\,w^2}{2+(1+2M+\sqrt{1+4M})\,w^2}.$$

Moreover, fixing $c \in (0, 1)$, *all the twizzlers in the set*

$$\left\{\mathfrak{T}\left(M, \frac{\sqrt{1+4M}-1-2M+c(\sqrt{1+4M}+1+2M)}{\sqrt{1+4M}+1-2M+c(\sqrt{1+4M}-1+2M)}\right) : M \in \left(-\frac{\sqrt{c}}{(1+\sqrt{c})^2}, \frac{\sqrt{c}}{(1-\sqrt{c})^2}\right)\right\}$$

are isometric.

Proof. Using the same notation as in the proof of Proposition 3.1, we see that the Gauss curvature K satisfies

$$K = \frac{eg-f^2}{EG-F^2} = -\frac{w^2(1+\theta'\xi_2)}{(1+w^2\xi_1^2)^2} = -\frac{w^2\big(1+2\,\xi_2\sqrt{1+w^2\xi_1^2}\big)}{(1+w^2\xi_1^2)(\xi_1^2+\xi_2^2)}.$$

Taking $\rho_1(u)$ and $\rho_2(u)$ as in Lemma 6.1, we get the following expression for the Gauss curvature in terms of the parameter u:

$$\frac{-4w^2\sqrt{1+4M+B\cos^2 u}\,\sin u}{\big(4+w^2+4A^4w^4\cos^4 u-2w^2\sqrt{1+4M+B\cos^2 u}\,\sin u + (1+4M)w^2\sin^2 u + w^2(8A^2+B\sin^2 u)\cos^2 u\big)}$$

A direct computation shows that the derivative of the function $K = K(u)$ is of the form $\cos u\ \mathrm{po}(u)$, where $\mathrm{po}(u)$ is a positive function, and therefore the maximum of the Gauss curvature occurs when $u = 3\pi/2$ and is equal to

$$\frac{2w^2\sqrt{1+4M}}{2+(1+2M+\sqrt{1+4M})w^2},$$

and the minimum of the Gauss curvature occurs when $u = \pi/2$ and is equal to

$$-\frac{2w^2\sqrt{1+4M}}{2+(1+2M-\sqrt{1+4M})w^2}.$$

We conclude that the quotient of the maximum value of the Gauss curvature and the minimum value of the Gauss curvature is

$$-\frac{2+(1+2M-\sqrt{1+4M})\,w^2}{2+(1+2M+\sqrt{1+4M})\,w^2}.$$

This expression in terms of v transforms into

$$-\frac{1+2M-\sqrt{1+4M}+v-2Mv+v\sqrt{1+4M}}{1+2M+\sqrt{1+4M}+v-2Mv-v\sqrt{1+4M}}.$$

A direct verification shows that this expression reduces to $-c$ when we replace v by

$$\frac{\sqrt{1+4M}-1-2M+c(\sqrt{1+4M}+1+2M)}{\sqrt{1+4M}+1-2M+c(\sqrt{1+4M}-1+2M)},$$

and therefore, for any $c \in (0, 1)$, all the twizzlers

$$\left\{\mathfrak{T}\left(M, \frac{\sqrt{1+4M}-1-2M+c(\sqrt{1+4M}+1+2M)}{\sqrt{1+4M}+1-2M+c(\sqrt{1+4M}-1+2M)}\right) : M \in \left(-\frac{\sqrt{c}}{(1+\sqrt{c})^2}, \frac{\sqrt{c}}{(1-\sqrt{c})^2}\right)\right\}$$

must be isometric. This follows because every twizzler with CMC 1 must be in the isometric associated family of a Delaunay surface [Lawson 1970], and it can be shown that the family of curves

$$\Omega_c = \left\{\left(M, \frac{\sqrt{1+4M}-1-2M+c(\sqrt{1+4M}+1+2M)}{\sqrt{1+4M}+1-2M+c(\sqrt{1+4M}-1+2M)}\right) : M \in \left(-\frac{\sqrt{c}}{(1+\sqrt{c})^2}, \frac{\sqrt{c}}{(1-\sqrt{c})^2}\right)\right\}$$

for $c \in (0, 1)$ defines a partition of the set $(-\frac{1}{4}, \infty) \times (0, 1)$. Figure 17 shows these curves Ω_c for different values of c. We know that two twizzlers corresponding to two points in different curves Ω_c cannot be isometric because their ratios of maximum to minimum Gauss curvatures are different. Using the continuity of the curve Ω_c and the fact that the 2π-periodic isometric family is continuous (see Remark 1.1), we see that all the isometric surfaces of the 2π-periodic associated family must be contained in a single Ω_c curve, and therefore all twizzlers there are

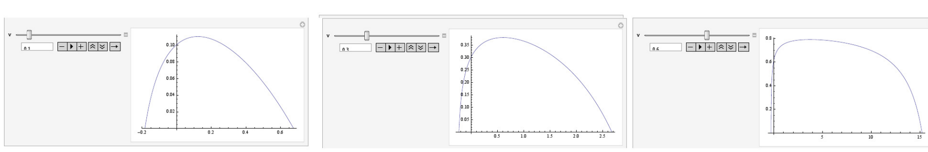


Figure 17. Points in the moduli space that represent isometric twizzlers.

isometric. As pointed out in the proof of Theorem 5.1, there are only two isometric Delaunay surfaces whose quotient between maximum and minimum values of the Gauss curvature is $-c$; they are the unduloid $\mathbb{D}(-\sqrt{c}/(1+\sqrt{c})^2)$ and the nodoid $\mathbb{D}(\sqrt{c}/(1-\sqrt{c})^2)$, and they correspond to the limit surfaces of the twizzlers that are in Ω_c. □

Remark 8.3. Helicoidal surfaces in the deformation helicoid-catenoid shown in Figure 3 are only a quarter of the whole 2π-periodic isometric family. All other elements in the 2π-periodic family are, up to a rigid motion, contained in the deformation shown in Figure 3. The same situation happens with twizzlers; the family of twizzlers given by points in Ω_c are only a quarter of the whole 2π-periodic isometric family. All other elements in the 2π-periodic family are, up to a rigid motion, contained in the twizzlers given in Ω_c.

Since the curve

$$\alpha_c(M) = \left(M, \frac{\sqrt{1+4M}-1-2M+c(\sqrt{1+4M}+1+2M)}{\sqrt{1+4M}+1-2M+c(\sqrt{1+4M}-1+2M)}\right)$$

satisfies that $\alpha_c\big(-\sqrt{c}/(1+\sqrt{c})^2\big) = \big(-\sqrt{c}/(1+\sqrt{c})^2, 0\big)$, $\alpha_c(0) = (0, c)$, and $\alpha_c\big(\sqrt{c}/(1-\sqrt{c})^2\big) = \big(\sqrt{c}/(1-\sqrt{c})^2, 0\big)$, as a corollary of Theorems 8.1 and 8.2 and Remark 1.1, we have:

Theorem 8.4. *Let* $\Omega = \{M+\mathrm{i}v \in \mathbb{C} : M \geq -\frac{1}{4},\ M \neq 0 \text{ and } 0 \leq v < 1\}$. *The function* ρ *from* Ω *to the set of immersions in* $\mathbb{R}^3$ *given by*

$$\begin{aligned} \rho(M+\mathrm{i}v) &= \mathfrak{T}(M, v) \text{ for any } v > 0 \text{ and } M \neq -\tfrac{1}{4}, \\ \rho(M) &= \mathbb{D}(M) \text{ for any } M \neq 0 \text{ and } M \neq -\tfrac{1}{4}, \\ \rho(-\tfrac{1}{4}+\mathrm{i}v) &= \{(x, y, z) \in \mathbb{R}^3 : x^2+z^2 = \tfrac{1}{4}\} \end{aligned}$$

is continuous in the sense that for every point p *in* Ω, *there exist a neighborhood* U *of* p *in* Ω *and a continuous function* $f : U \times \mathbb{R}^2 \to \mathbb{R}^3$ *such that for any* $M+\mathrm{i}v \in U$, *the map* $(s, t) \to f(M+\mathrm{i}v, s, t)$ *defines a parametrization of the surface* $\rho(M+\mathrm{i}v)$. *Moreover, the function* ρ *is one-to-one in the interior of* Ω.

The continuity at the points of form $-\frac{1}{4}+\mathrm{i}v$ follows from Theorem 7.2, because each twizzler $\mathfrak{T}(M, v)$ is contained in the region bounded by the two concentric cylinders of radii $r_1(M) = |(\sqrt{1+4M}-1)/2|$ and $r_2(M) = (\sqrt{1+4M}+1)/2$. Figure 17 shows the trace of the curve α_c for several values of c.

Summary. We collect some important facts on helicoidal surfaces with constant mean curvature one. Figure 18 shows a picture of the moduli space.

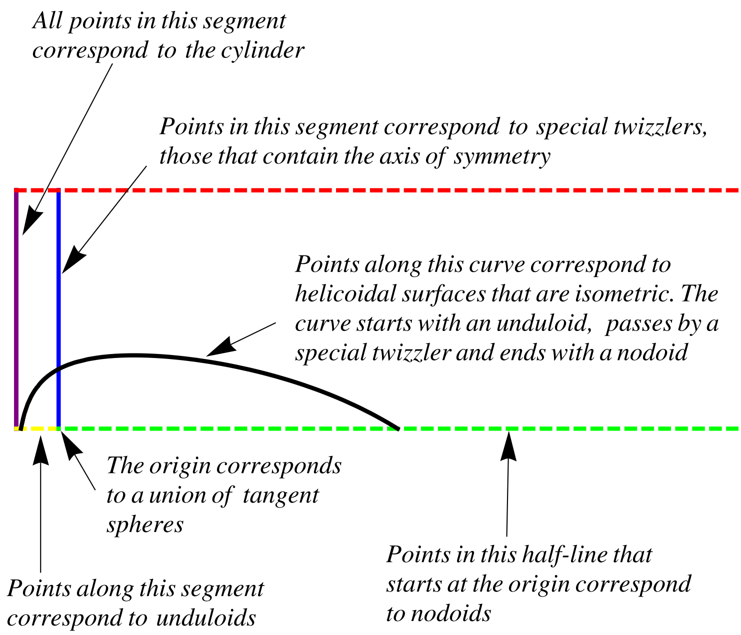


Figure 18. Moduli space of twizzler with CMC 1 and its boundary.

Dynamical interpretation of Delaunay surfaces. The trace of the focus of each conic $4x^2 - y^2/M = 1$ with $M \in (-\frac{1}{4}, 0) \cup (0, \infty)$, when it is rolled on a line, produces the profile curve of a Delaunay surface with constant mean curvature one. Moreover, every Delaunay surface with CMC 1 but the cylinder corresponds with one of these conics.

Dynamical interpretation of twizzlers. The treadmill sled of the profile curve of a twizzler with CMC 1 other than a cylinder is the closed curve

$$x^2 + y^2 - \frac{y}{\sqrt{1 + w^2 x^2}} = M$$

for some $M > -\frac{1}{4}$ and $w > 0$.

Moduli space of twizzlers. Denote by $\rho(M, v)$ the twizzler whose treadmill sled of its profile curve lies on the heart-shaped curve

$$x^2 + y^2 - \frac{y}{\sqrt{1 + w^2 x^2}} = M,$$

where $v = 1/(1 + w^2)$. Then ρ generates a one-to-one correspondence between the half-strip

$$\Omega = \{(M, v) : M > -\tfrac{1}{4} \text{ and } 0 < v < 1\}$$

and all twizzlers with CMC 1 but the cylinder.

Boundary of the moduli space of twizzlers. When M goes to $-\frac{1}{4}$, the surfaces $\rho(M, v)$ converge to a cylinder. When v goes to zero, the surfaces $\rho(M, v)$ converge to the Delaunay surface whose profile curve is traced by the focus of the conic $4x^2 - y^2/M = 1$. When M goes to zero, the Delaunay surface whose profile curve is traced by the focus of the conic $4x^2 - y^2/M = 1$ converges to a union of infinitely many tangent spheres [Kapouleas 1990, Appendix A].

Fundamental piece of the profile curve. For every twizzler other than a cylinder, we can define the fundamental piece of the profile curve as a connected part of the profile curve whose treadmill sled goes exactly once over the closed curve $x^2+y^2-y/\sqrt{1+w^2x^2}=M$. The function $\theta_0(M, v)$ given in Theorem 7.2 provides a formula for the angle between the initial and final positions of the fundamental piece of the profile curve. The function θ_0 defined on Ω is given in terms of an integral of an expression involving only sine and cosine functions.

Properties of twizzlers. If M is nonzero, then the twizzler $\rho(M, v)$ lies in the region $C_{r_1 r_2}$ bounded by two concentric cylinders of radii $r_1(M)=|(\sqrt{1+4M}-1)/2|$ and $r_2(M) = (\sqrt{1+4M}+1)/2$; also, $\rho(M, v)$ is properly immersed if and only if $\theta_0(M, v)/2\pi$ is a rational number, and otherwise it is dense in $C_{r_1 r_2}$. If $M = 0$, then, the twizzler $\rho(M, v)$ contains the axis of symmetry and lies inside a cylinder of radius 1; moreover, $\rho(M, v)$ is properly immersed if and only if $\theta_0(M, v)/2\pi$ is a rational number, and otherwise it is dense in the interior of this cylinder.

Isometric surfaces. Theorem 8.2 provides an explicit formula for a foliation of the moduli space Ω by curves with the property that all the twizzlers in each curve are isometric. Each one of these curves starts with an unduloid, passes through a special twizzler (a twizzler that contains the axis of symmetry), and ends with a nodoid. In particular, every twizzler different other than cylinder is isometric to a twizzler that contains the axis of symmetry.

Acknowledgements

The author would like to express his gratitude to Professors Robert Kusner, Ivan Sterling, Bruce Solomon, Wayne Rossman, Martin Kilian, and Ioannis Roussos for solving several of his doubts about Delaunay surfaces and twizzlers and for providing him with references.

References

[do Carmo and Dajczer 1982] M. P. do Carmo and M. Dajczer, "Helicoidal surfaces with constant mean curvature", *Tôhoku Math. J.* (2) **34**:3 (1982), 425–435. MR 84f:53003 Zbl 0501.53003

[Delaunay 1841] C. Delaunay, "Sur la surface de revolution dont la courbure mayenne est constante", *J. Math. Pures Appl.* **6** (1841), 309–320.

[Graustein 1935] W. C. Graustein, *Differential geometry*, Macmillan, New York, 1935. MR 33 #6510 Zbl 0139.14701

[Haak 1998] G. Haak, “On a theorem by do Carmo and Dajczer”, *Proc. Amer. Math. Soc.* **126**:5 (1998), 1547–1548. MR 98j:53011 Zbl 0902.53008

[Hitt and Roussos 1991] L. R. Hitt and I. M. Roussos, “Computer graphics of helicoidal surfaces with constant mean curvature”, *An. Acad. Bras. Ciênc.* **63**:3 (1991), 211–228. Zbl 0803.53010

[Kapouleas 1990] N. Kapouleas, “Complete constant mean curvature surfaces in Euclidean three-space”, *Ann. of Math.* (2) **131**:2 (1990), 239–330. MR 93a:53007a Zbl 0699.53007

[Lawson 1970] H. B. Lawson, Jr., “Complete minimal surfaces in S^3”, *Ann. of Math.* (2) **92** (1970), 335–374. MR 42 #5170 Zbl 0205.52001

[Perdomo 2011] O. Perdomo, “The treadmillsled of a curve”, preprint, 2011. arXiv 1105.3460v1

Received September 2, 2011. Revised April 5, 2012.

OSCAR M. PERDOMO
DEPARTMENT OF MATHEMATICS
CENTRAL CONNECTICUT STATE UNIVERSITY
NEW BRITAIN, CT 06050
UNITED STATES

perdomoosm@ccsu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.258.487

CLASSIFICATION OF ISING VECTORS IN THE VERTEX OPERATOR ALGEBRA V_L^+

HIROKI SHIMAKURA

Let L be an even lattice without roots. In this article, we classify all Ising vectors in the vertex operator algebra V_L^+ associated with L.

Introduction

In vertex operator algebra (VOA) theory, the simple Virasoro VOA $L(\frac{1}{2},0)$ of central charge $\frac{1}{2}$ plays important roles. In fact, for each embedding, an automorphism, called a τ-involution, is defined using the representation theory of $L(\frac{1}{2},0)$ [Miyamoto 1996]. This is useful for the study of the automorphism group of a VOA. For example, this construction gives a one-to-one correspondence between the set of subVOAs of the moonshine VOA isomorphic to $L(\frac{1}{2},0)$ and that of elements in certain conjugacy class of the Monster [Miyamoto 1996; Höhn 2010].

Many properties of τ-involutions are studied using Ising vectors, which are elements of weight 2 generating $L(\frac{1}{2},0)$. For example, the 6-transposition property of τ-involutions was proved in [Sakuma 2007] by classifying the subalgebra generated by two Ising vectors. Hence it is natural to classify Ising vectors in a VOA. For example, this was done in [Lam 1999; Lam et al. 2007] for code VOAs. However, in general, it is hard to even find an Ising vector.

Let L be an even lattice and V_L the lattice VOA associated with L. Then the subspace V_L^+ fixed by a lift of the -1-isometry of L is a subVOA of V_L. There are two constructions of Ising vectors in V_L^+ related to sublattices of L isomorphic to $\sqrt{2}A_1$ [Dong et al. 1994] and $\sqrt{2}E_8$ [Dong et al. 1998; Griess 1998].

The main theorem of this article is this:

Theorem 2.3. *Let L be an even lattice without roots and e an Ising vector in V_L^+. There is a sublattice U of L isomorphic to $\sqrt{2}A_1$ or $\sqrt{2}E_8$ and such that $e \in V_U^+$.*

This theorem was conjectured in [Lam et al. 2007], and proved there and in [Lam and Shimakura 2007] in the case that $L/\sqrt{2}$ is even and L is the Leech lattice. We

---

H. Shimakura was partially supported by Grants-in-Aid for Scientific Research (grant number 23540013), JSPS.
MSC2010: 17B69.
Keywords: vertex operator algebra, lattice vertex operator algebra, Ising vector.

note that if L has roots then the automorphism group of V_L^+ is infinite, and V_L^+ may have infinitely many Ising vectors.

In this article, we prove Theorem 2.3, and hence we classify all Ising vectors in V_L^+. Our result shows that the study of τ-involutions of V_L^+ is essentially equivalent to that of sublattices of L isomorphic to $\sqrt{2}E_8$ (see [Griess and Lam 2011; 2012]).

The key is to describe the action of the τ-involution on the Griess algebra B of V_L^+. Let e be an Ising vector in V_L^+ and $L(4; e)$ the norm 4 vectors in L which appear in the description of e with respect to the standard basis of $(V_L^+)_2$ (see Section 2 for the definition of $L(4; e)$). By [Lam and Shimakura 2007], the τ-involution τ_e associated to e is a lift of an automorphism g of L. We show in Lemma 2.1 that g is trivial on $\{\{\pm v\} \mid v \in L(4; e)\}$. This lemma follows from the decomposition of B with respect to the adjoint action of e [Höhn et al. 2012], the action of τ_e on it [Miyamoto 1996] and the explicit calculations on the Griess algebra [Frenkel et al. 1988]. By this lemma, we can obtain a VOA V containing e on which τ_e acts trivially. By [Lam et al. 2007] e is fixed by the group A generated by τ-involutions associated to elements in $L(4; e)$. Hence e belongs to the subVOA V^A of V fixed by A. Using the explicit action of A, we can find a lattice N satisfying $e \in V_N^+$ and $N/\sqrt{2}$ is even. This case was done in [Lam et al. 2007].

1. Preliminaries

VOAs associated with even lattices. In this subsection, we review the VOAs V_L and V_L^+ associated with even lattice L of rank n and their automorphisms. Our notation for lattice VOAs here is standard (see [Frenkel et al. 1988]).

Let L be a (positive-definite) even lattice with inner product $\langle \cdot , \cdot \rangle$. Let also $H = \mathbb{C} \otimes_{\mathbb{Z}} L$ be an abelian Lie algebra and $\hat{H} = H \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}c$ be its affine Lie algebra. Let $\hat{H}^- = H \otimes t^{-1}\mathbb{C}[t^{-1}]$ and let $S(\hat{H}^-)$ be the symmetric algebra of $\hat{H}^-$. Then $M_H(1) = S(\hat{H}^-) \cong \mathbb{C}[h(m) \mid h \in H, m < 0] \cdot \mathbf{1}$ is the unique irreducible $\hat{H}$-module such that $h(m) \cdot \mathbf{1} = 0$ for $h \in H$, $m \geq 0$ and $c = 1$, where $h(m) = h \otimes t^m$. Note that $M_H(1)$ has a VOA structure.

The twisted group algebra $\mathbb{C}\{L\}$ can be described as follows. Let $\langle \kappa \rangle$ be a cyclic group of order 2 and $1 \to \langle \kappa \rangle \to \hat{L} \to L \to 1$ a central extension of L by $\langle \kappa \rangle$ satisfying the commutator relation $[e^\alpha, e^\beta] = \kappa^{\langle \alpha, \beta \rangle}$ for $\alpha, \beta \in L$. Let $L \to \hat{L}, \alpha \mapsto e^\alpha$ be a section and $\varepsilon(,) : L \times L \to \langle \kappa \rangle$ the associated 2-cocycle, that is, $e^\alpha e^\beta = \varepsilon(\alpha, \beta) e^{\alpha+\beta}$. We may assume that $\varepsilon(\alpha, \alpha) = \kappa^{\langle \alpha, \alpha \rangle / 2}$ and $\varepsilon(,)$ is bilinear by [Frenkel et al. 1988, Proposition 5.3.1]. The twisted group algebra is defined by

$$\mathbb{C}\{L\} = \mathbb{C}[\hat{L}]/(\kappa + 1) \cong \mathrm{Span}_{\mathbb{C}}\{e^\alpha \mid \alpha \in L\},$$

where $\mathbb{C}[\hat{L}]$ is the usual group algebra of the group $\hat{L}$. The lattice VOA V_L associated with L is defined as $M_H(1) \otimes \mathbb{C}\{L\}$ [Borcherds 1986; Frenkel et al. 1988].

For any sublattice E of L, let $\mathbb{C}\{E\} = \mathrm{Span}_{\mathbb{C}}\{e^\alpha \mid \alpha \in E\}$ be a subalgebra of $\mathbb{C}\{L\}$ and let $H_E = \mathbb{C} \otimes_{\mathbb{Z}} E$ be a subspace of $H = \mathbb{C} \otimes_{\mathbb{Z}} L$. Then the subspace $S(\hat{H}_E^-) \otimes \mathbb{C}\{E\}$ forms a subVOA of V_L and it is isomorphic to the lattice VOA V_E.

Let $O(\hat{L})$ be the subgroup of $\mathrm{Aut}\,\hat{L}$ induced by $\mathrm{Aut}\,L$. By [Frenkel et al. 1988, Proposition 5.4.1] there is an exact sequence of groups

$$1 \longrightarrow \mathrm{Hom}(L, \mathbb{Z}/2\mathbb{Z}) \longrightarrow O(\hat{L}) \xrightarrow{\;-\;} \mathrm{Aut}\,L \longrightarrow 1.$$

Note that for $f \in O(\hat{L})$,

$$f(e^\alpha) \in \{\pm e^{\bar{f}(\alpha)}\}. \tag{1-1}$$

By [Frenkel et al. 1988, Corollary 10.4.8], $f \in O(\hat{L})$ acts on V_L as an automorphism by

$$\begin{aligned} f(h_{i_1}(n_1)h_{i_2}(n_2)\dots h_{i_k}(n_k) \otimes e^\alpha) \\ = \bar{f}(h_{i_1})(n_1)\bar{f}(h_{i_2})(n_2)\dots \bar{f}(h_{i_k})(n_k) \otimes f(e^\alpha), \end{aligned} \tag{1-2}$$

where $n_i \in \mathbb{Z}_{<0}$ and $\alpha \in L$. Hence $O(\hat{L})$ is a subgroup of $\mathrm{Aut}\,V_L$.

Let θ be the automorphism of $\hat{L}$ defined by $\theta(e^\alpha) = e^{-\alpha}$ for $\alpha \in L$. Then $\bar{\theta} = -1 \in \mathrm{Aut}\,L$. Using (1-2) we view θ as an automorphism of V_L. Let V_L^+ be the subspace $\{v \in V_L \mid \theta(v) = v\}$ of V_L fixed by θ. Then V_L^+ is a subVOA of V_L. Since θ is a central element of $O(\hat{L})$, the quotient group $O(\hat{L})/\langle\theta\rangle$ is a subgroup of $\mathrm{Aut}\,V_L^+$. Note that V_L^+ is a simple VOA of CFT type.

Later, we will consider the subVOA of V_L^+ generated by the weight 2 subspace.

Lemma 1.1 [Frenkel et al. 1988, Proposition 12.2.6]. *Let L be an even lattice without roots. Let N be the sublattice of L generated by $L(4)$. Then the subVOA of V_L^+ generated by $(V_L^+)_2$ is $(V_N \otimes M_{H'}(1))^+$, where $H' = (\langle N\rangle_{\mathbb{C}})^\perp$ in $\langle L\rangle_{\mathbb{C}}$.*

Ising vectors and τ-involutions. In this subsection, we review Ising vectors and corresponding τ-involutions.

Definition 1.2. A weight 2 element e of a VOA is called an *Ising vector* if the vertex subalgebra generated by e is isomorphic to the simple Virasoro VOA of central charge $\frac{1}{2}$ and e is its conformal vector.

For an Ising vector e, the automorphism τ_e, called the *τ-involution* or *Miyamoto involution*, was defined in [Miyamoto 1996, Theorem 4.2] based on the representation theory of the simple Virasoro VOA of central charge $\frac{1}{2}$ [Dong et al. 1994].

Let V be a VOA of CFT type with $V_1 = 0$. The first product $(a, b) \mapsto a \cdot b = a_{(1)}b$ provides a (nonassociative) commutative algebra structure on V_2. This algebra V_2 is called the *Griess algebra* of V, and τ_e acts on it as follows:

Lemma 1.3 [Höhn et al. 2012, Lemma 2.6]. *Let V be a simple VOA of CFT type with $V_1 = 0$ and e an Ising vector in V. Then $B = V_2$ has the decomposition*

$$B = \mathbb{C}e \oplus B^e(0) \oplus B^e(\tfrac{1}{2}) \oplus B^e(\tfrac{1}{16})$$

with respect to the adjoint action of e, where $B^e(k) = \{v \in B \mid e \cdot v = kv\}$. The automorphism τ_e acts on B as

$$1 \text{ on } \mathbb{C}e \oplus B^e(0) \oplus B^e(\tfrac{1}{2}) \quad \text{and} \quad -1 \text{ on } B^e(\tfrac{1}{16}).$$

In the proof of our main theorem, we need:

Lemma 1.4 [Lam et al. 2007, Lemma 3.7]. *Let V be a VOA of CFT type with $V_1 = 0$. Suppose that V has two Ising vectors e, f and that $\tau_e = \mathrm{id}$ on V. Then e is fixed by τ_f, namely $e \in V^{\tau_f}$.*

Let L be an even lattice of rank n without roots, that is,

$$L(2) = \{v \in L \mid \langle v, v\rangle = 2\} = \varnothing.$$

Then $(V_L^+)_1 = 0$, and we can consider the Griess algebra $B = (V_L^+)_2$ of V_L^+. Let $\{h_i \mid 1 \le i \le n\}$ be an orthonormal basis of the vector space $H = \mathbb{C}\otimes_{\mathbb{Z}} L = \langle L\rangle_{\mathbb{C}}$. Set $L(4) = \{v \in L \mid \langle v, v\rangle = 4\}$. For $1 \le i \le j \le n$ and $\alpha \in L(4)$, set $h_{ij} = h_i(-1)h_j(-1)\mathbf{1}$ and $x_\alpha = e^\alpha + e^{-\alpha} = e^\alpha + \theta(e^\alpha)$. Note that $x_\alpha = x_{-\alpha}$.

Lemma 1.5 [Frenkel et al. 1988, Section 8.9]. (1) *The set*

$$\{h_{ij}, x_\alpha \mid 1 \le i \le j \le n,\ \{\pm\alpha\} \subset L(4)\}$$

is a basis of B.

(2) *The products of the basis vectors of B given in* (1) *are*

$$\begin{aligned} h_{ij} \cdot h_{kl} &= \delta_{ik}h_{jl} + \delta_{il}h_{jk} + \delta_{jk}h_{il} + \delta_{jl}h_{ik},\\ h_{ij} \cdot x_\alpha &= \langle h_i, \alpha\rangle\langle h_j, \alpha\rangle x_\alpha,\\ x_\alpha \cdot x_\beta &= \begin{cases} \varepsilon(\alpha, \beta)x_{\alpha\pm\beta} & \text{if } \langle\alpha, \beta\rangle = \mp 2,\\ \alpha(-1)^2\mathbf{1} & \text{if } \alpha = \pm\beta,\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Let $\alpha \in L(4)$. Then the elements $\omega^+(\alpha)$ and $\omega^-(\alpha)$ of V_L^+ defined by

$$\omega^\pm(\alpha) = \tfrac{1}{16}\alpha(-1)^2 \cdot \mathbf{1} \pm \tfrac{1}{4}x_\alpha \tag{1-3}$$

are Ising vectors [Dong et al. 1994, Theorem 6.3]. The following lemma is easy:

Lemma 1.6. *The automorphisms $\tau_{\omega^\pm(\alpha)}$ of V_L^+ act by*

$$u \otimes x_\beta \mapsto (-1)^{\langle\alpha,\beta\rangle}u \otimes x_\beta \quad \text{for } u \in M_H(1) \text{ and } \beta \in L.$$

More generally:

Proposition 1.7 [Lam and Shimakura 2007, Lemma 5.5]. *Let L be an even lattice without roots and e an Ising vector in V_L^+. Then $\tau_e \in O(\hat{L})/\langle\theta\rangle$.*

When $L/\sqrt{2}$ is even, our main theorem reduces to something proved earlier:

Proposition 1.8 [Lam et al. 2007, Theorem 4.6]. *Let L be an even lattice and e an Ising vector in V_L^+. Assume that the lattice $L/\sqrt{2}$ is even. There is a sublattice U of L isomorphic to $\sqrt{2}A_1$ or $\sqrt{2}E_8$ and such that $e \in V_U^+$.*

2. Classification of Ising vectors in V_L^+

Let L be an even lattice of rank n without roots and e an Ising vector in V_L^+. Then by Lemma 1.5(1),

$$e = \sum_{i \le j} c_{ij}^e h_{ij} + \sum_{\{\pm\alpha\}\subset L(4)} d_{\{\pm\alpha\}}^e x_\alpha, \tag{2-1}$$

where $c_{ij}^e, d_{\{\pm\alpha\}}^e \in \mathbb{C}$. Set $L(4;e) = \{\alpha \in L(4) \mid d_{\{\pm\alpha\}}^e \neq 0\}$, $H_1 = \langle L(4;e)\rangle_{\mathbb{C}}$ and $H_2 = H_1^\perp$ in H. Note that if $\alpha \in L(4;e)$ then $-\alpha \in L(4;e)$. Without loss of generality, we may assume that $h_i \in H_1$ if $1 \le i \le \dim H_1$. Then we have $H_2 = \mathrm{Span}_{\mathbb{C}}\{h_j \mid \dim H_1 + 1 \le j \le n\}$.

By Proposition 1.7, $\tau_e \in O(\hat{L})/\langle\theta\rangle$. Since $e \in V_L$, we regard τ_e as an automorphism of V_L. Then $\tau_e \in O(\hat{L})$, and set $g = \bar{\tau_e} \in \mathrm{Aut}\, L$. Since τ_e is of order 1 or 2, so is g. We now state the key lemma in this article:

Lemma 2.1. *Let $\beta \in L(4;e)$. Then $g(\beta) \in \{\pm\beta\}$.*

Proof. By (1-1) and (1-2),

$$\tau_e(x_\beta) \in \{\pm x_{g(\beta)}\}. \tag{2-2}$$

On the other hand, $\tau_e(e) = e$, (1-2) and (2-1) show that

$$\tau_e(d_{\{\pm\beta\}}^e x_\beta) = d_{\{\pm g(\beta)\}}^e x_{g(\beta)}. \tag{2-3}$$

By (2-2) and (2-3),

$$d_{\{\pm g(\beta)\}}^e / d_{\{\pm\beta\}}^e \in \{\pm 1\}. \tag{2-4}$$

Suppose $g(\beta) \notin \{\pm\beta\}$. Then $x_\beta - \tau_e(x_\beta)$ is nonzero, and it is an eigenvector of τ_e with eigenvalue -1. By Lemma 1.3, we have

$$e \cdot (x_\beta - \tau_e(x_\beta)) = \tfrac{1}{16}(x_\beta - \tau_e(x_\beta)). \tag{2-5}$$

We calculate the image of both sides of (2-5) under the canonical projection $\mu : (V_L^+)_2 \to \mathrm{Span}_{\mathbb{C}}\{h_{ij} \mid 1 \le i \le j \le n\}$ with respect to the basis given in

Lemma 1.5(1). By (2-2) the image of the right side of (2-5) under μ is

$$\mu\left(\tfrac{1}{16}(x_\beta - \tau_e(x_\beta))\right) = 0. \tag{2-6}$$

Let us discuss the left side of (2-5). By Lemma 1.5(2) and (2-4), we have

$$\begin{aligned} e\cdot(x_\beta - \tau_e(x_\beta)) &= \Bigg(\sum_{i\le j} c^e_{ij} h_{ij} + \sum_{\{\pm\alpha\}\subset L(4)} d^e_{\{\pm\alpha\}} x_\alpha\Bigg)\cdot\big(x_\beta - \tau_e(x_\beta)\big) \\ &\in d^e_{\{\pm\beta\}}\big(\beta(-1)^2\mathbf{1} - g(\beta)(-1)^2\mathbf{1}\big) + \mathrm{Span}_{\mathbb{C}}\big\{x_\gamma \;\big|\; \{\pm\gamma\}\subset L(4)\big\}. \end{aligned}$$

Thus

$$\begin{aligned} \mu(e\cdot(x_\beta - \tau_e(x_\beta))) &= d^e_{\{\pm\beta\}}\big(\beta(-1)^2\mathbf{1} - g(\beta)(-1)^2\mathbf{1}\big) \\ &= d^e_{\{\pm\beta\}}\,(\beta - g(\beta))\,(-1)(\beta + g(\beta))(-1)\mathbf{1}. \end{aligned}$$

This is not zero since $g(\beta)\notin\{\pm\beta\}$, which contradicts (2-5) and (2-6). Therefore $g(\beta)\in\{\pm\beta\}$. □

For $\varepsilon\in\{\pm\}$, set

$$L(4;e,\varepsilon) = \{v\in L(4;e)\mid g(v) = \varepsilon v\},\quad L^{e,\varepsilon} = \langle L(4;e,\varepsilon)\rangle_{\mathbb{Z}},\quad H_1^\varepsilon = \langle L^{e,\varepsilon}\rangle_{\mathbb{C}}.$$

Since g preserves the inner product, $H_1 = H_1^+ \perp H_1^-$ and g acts on $H_2 = H_1^\perp$. Let $H_2^\pm$ be ±1-eigenspaces of g in H_2. For $\varepsilon\in\{\pm\}$, let W^ε be a lattice of full rank in H_2^ε isomorphic to an orthogonal direct sum of copies of $2A_1$. Then

$$M_{H_2^\varepsilon}(1)\subset V_{W^\varepsilon}. \tag{2-7}$$

Lemma 2.2. *The Ising vector e belongs to the VOA*

$$V^+_{L^{e,+}\oplus W^+}\otimes V^+_{L^{e,-}\oplus W^-},$$

and $\tau_e = \mathrm{id}$ on this VOA.

Proof. By Lemma 2.1, $L(4;e) = L(4;e,+)\cup L(4;e,-)$. Hence, by (2-1) and (2-7),

$$e\in (V_{L^{e,+}}\otimes M_{H_2^+}(1)\otimes V_{L^{e,-}}\otimes M_{H_2^-}(1))^+ \subset V^+_{L^{e,+}\oplus W^+\oplus L^{e,-}\oplus W^-}. \tag{2-8}$$

Since g acts by ±1 on $L^{e,\pm}\oplus W^\pm$, the subspace of (2-8) fixed by τ_e is

$$V^+_{L^{e,+}\oplus W^+}\otimes V^+_{L^{e,-}\oplus W^-}.$$

Since e is fixed by τ_e, we have the desired result. □

We now prove the main theorem.

Theorem 2.3. *Let L be an even lattice without roots and e an Ising vector in V_L^+. There is a sublattice U of L isomorphic to $\sqrt{2}A_1$ or $\sqrt{2}E_8$ and such that $e\in V_U^+$.*

Proof. Set $V = V^+_{L^{e,+}\oplus W^+} \otimes V^+_{L^{e,-}\oplus W^-}$. By Lemma 2.2, e belongs to V and $\tau_e = \mathrm{id}$ on V. Let $A = \langle \tau_{\omega^\pm(\beta)} \mid \beta \in L(4; e)\rangle$. By Lemma 1.4, e belongs to the subVOA V^A of V fixed by A. Since e is a weight 2 element, it is contained in the subVOA generated by $(V^A)_2$. By Lemmas 1.1 and 1.6 and (2-7) (see (2-8)),

$$e \in V^+_{N^+\oplus K^+} \otimes V^+_{N^-\oplus K^-} \subset V^+_N,$$

where for $\varepsilon \in \{\pm\}$, $N^\varepsilon = \mathrm{Span}_{\mathbb{Z}}\{v \in L(4; e, \varepsilon) \mid \langle v, L(4; e)\rangle \in 2\mathbb{Z}\}$, K^ε is a lattice of full rank in $(\langle N^\varepsilon\rangle_{\mathbb{C}})^\perp \cap (H_1^\varepsilon \oplus H_2^\varepsilon)$ isomorphic to an orthogonal direct sum of copies of $2A_1$, and $N = N^+ \oplus K^+ \oplus N^- \oplus K^-$. Since N is generated by norm 4 and 8 vectors, and the inner products of the generator belong to $2\mathbb{Z}$, the lattice $N/\sqrt{2}$ is even. By Proposition 1.8, there is a sublattice U of N isomorphic to $\sqrt{2}A_1$ or $\sqrt{2}E_8$ such that $e \in V_U^+$. It follows from $K^+(4) = K^-(4) = \varnothing$ that $N(4) = N^+(4) \cup N^-(4) \subset L$. Since $\sqrt{2}A_1$ and $\sqrt{2}E_8$ are spanned by norm 4 vectors as lattices, we have $U \subset L$. Hence V_U^+ is a subVOA of V_L^+. □

As an application of the main theorem, we count the total number of Ising vectors in V_L^+ for even lattice L without roots.

Let us describe Ising vectors in V_L^+. The Ising vector $\omega^\pm(\alpha)$ associated to α in $L(4)$ was described in (1-3) as

$$\omega^\pm(\alpha) = \tfrac{1}{16}\alpha(-1)^2 \cdot \mathbf{1} \pm \tfrac{1}{4}x_\alpha.$$

Let E be an even lattice isomorphic to $\sqrt{2}E_8$ and $\{u_i \mid 1 \le i \le 8\}$ an orthonormal basis of $\mathbb{C} \otimes_{\mathbb{Z}} E$. We consider the trivial 2-cocycle of $\mathbb{C}\{E\}$ for V_E. Then for $\varphi \in \mathrm{Hom}(E, \mathbb{Z}/2\mathbb{Z}) (\cong (\mathbb{Z}/2\mathbb{Z})^8)$,

$$\omega(E,\varphi) = \frac{1}{32}\sum_{i=1}^{8} u_i(-1)^2 \cdot \mathbf{1} + \frac{1}{32}\sum_{\{\pm\alpha\}\subset E(4)} (-1)^{\varphi(\alpha)} x_\alpha$$

is an Ising vector in V_E^+ [Dong et al. 1998; Griess 1998]. Since $E(4)$ spans E as a lattice, $\omega(E,\varphi) = \omega(E,\varphi')$ if and only if $\varphi = \varphi'$. Hence V_E^+ has 256 Ising vectors of form $\omega(E,\varphi)$. Thus $V^+_{\sqrt{2}A_1}$ and $V^+_{\sqrt{2}E_8}$ have exactly 2 and 496 Ising vectors, respectively [Lam et al. 2007, Propositions 4.2 and 4.3].

Corollary 2.4. *Let L be an even lattice without roots. Then the number of Ising vectors in V_L^+ is*

$$|L(4)| + 256 \times \left|\{U \subset L \mid U \cong \sqrt{2}E_8\}\right|.$$

Proof. Set $m = |L(4)| + 256 \times |\{E \subset L \mid E \cong \sqrt{2}E_8\}|$. Theorem 2.3 shows that the number of Ising vectors in V_L^+ is less than or equal to m. Let us show that there are exactly m Ising vectors in V_L^+, that is, the Ising vectors $\omega^\pm(\alpha)$ and $\omega(E,\varphi)$ are distinct. By Lemma 1.5(1), $\omega^\varepsilon(\alpha) = \omega^\delta(\beta)$ if and only if $\alpha = \beta$ and $\varepsilon = \delta$. Also $\omega^\varepsilon(\alpha) \ne \omega(E,\varphi)$ for all $\alpha \in L(4)$, $L \supset E \cong \sqrt{2}E_8$ and $\varphi \in \mathrm{Hom}(E, \mathbb{Z}/2\mathbb{Z})$.

Let E_1, E_2 be sublattices of L such that $E_1 \cong E_2 \cong \sqrt{2}E_8$. Let φ_i, $i = 1, 2$, be two elements of $\mathrm{Hom}(E_i, \mathbb{Z}/2\mathbb{Z})$. Then it follows from Lemma 1.5(1) and $\langle E_i(4)\rangle_{\mathbb{Z}} = E_i$ that $\omega(E_1, \varphi_1) = \omega(E_2, \varphi_2)$ if and only if $E_1 = E_2$ and $\varphi_1 = \varphi_2$. Therefore, there are exactly m Ising vectors in V_L^+. □

Acknowledgement

The author thanks the referee for valuable advice.

References

[Borcherds 1986] R. E. Borcherds, "Vertex algebras, Kac–Moody algebras, and the monster", *Proc. Nat. Acad. Sci. USA* **83**:10 (1986), 3068–3071. MR 87m:17033 Zbl 0613.17012

[Dong et al. 1994] C. Dong, G. Mason, and Y. Zhu, "Discrete series of the Virasoro algebra and the moonshine module", pp. 295–316 in *Algebraic groups and their generalizations: quantum and infinite-dimensional methods* (University Park, PA, 1991), edited by W. J. Haboush and B. J. Parshall, Proc. Sympos. Pure Math. **56**, Amer. Math. Soc., Providence, RI, 1994. MR 95c:17043 Zbl 0813.17019

[Dong et al. 1998] C. Dong, H. Li, G. Mason, and S. P. Norton, "Associative subalgebras of the Griess algebra and related topics", pp. 27–42 in *The monster and Lie algebras* (Columbus, OH, 1996), edited by J. Ferrar and K. Harada, Ohio State Univ. Math. Res. Inst. Publ. **7**, de Gruyter, Berlin, 1998. MR 99k:17048 Zbl 0946.17011

[Frenkel et al. 1988] I. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the monster*, Pure and Applied Mathematics **134**, Academic Press, Boston, 1988. MR 90h:17026 Zbl 0674.17001

[Griess 1998] R. L. Griess, Jr., "A vertex operator algebra related to E_8 with automorphism group $\mathrm{O}^+(10, 2)$", pp. 43–58 in *The monster and Lie algebras* (Columbus, OH, 1996), edited by J. Ferrar and K. Harada, Ohio State Univ. Math. Res. Inst. Publ. **7**, de Gruyter, Berlin, 1998. MR 99i:17030 Zbl 0946.17012

[Griess and Lam 2011] R. L. Griess, Jr. and C. H. Lam, "EE_8-lattices and dihedral groups", *Pure Appl. Math. Q.* **7**:3 (2011), 621–743. MR 2848589 Zbl 05942227

[Griess and Lam 2012] R. L. Griess, Jr. and C. H. Lam, "Diagonal lattices and rootless EE_8 pairs", *J. Pure Appl. Algebra* **216**:1 (2012), 154–169. MR 2826430 Zbl 05999997

[Höhn 2010] G. Höhn, "The group of symmetries of the shorter moonshine module", *Abh. Math. Semin. Univ. Hambg.* **80**:2 (2010), 275–283. MR 2012a:17052 Zbl 1210.17033

[Höhn et al. 2012] G. Höhn, C. H. Lam, and H. Yamauchi, "McKay's E_7 observation on the baby monster", *Int. Math. Res. Not.* **2012**:1 (2012), 166–212. MR 2874931 Zbl 06017887

[Lam 1999] C. H. Lam, "Code vertex operator algebras under coordinates change", *Comm. Algebra* **27**:9 (1999), 4587–4605. MR 2000i:17045 Zbl 0977.17028

[Lam and Shimakura 2007] C. H. Lam and H. Shimakura, "Ising vectors in the vertex operator algebra V_Λ^+ associated with the Leech lattice Λ", *Int. Math. Res. Not.* **2007**:24 (2007), Art. ID rnm 132. MR 2009a:17033 Zbl 1173.17022

[Lam et al. 2007] C. H. Lam, S. Sakuma, and H. Yamauchi, "Ising vectors and automorphism groups of commutant subalgebras related to root systems", *Math. Zeit.* **255**:3 (2007), 597–626. MR 2007h:17032 Zbl 1139.17010

[Miyamoto 1996] M. Miyamoto, "Griess algebras and conformal vectors in vertex operator algebras", *J. Algebra* **179**:2 (1996), 523–548. MR 96m:17052 Zbl 0964.17021

[Sakuma 2007] S. Sakuma, "6-transposition property of τ-involutions of vertex operator algebras", *Int. Math. Res. Not.* **2007**:9 (2007), Art. ID rnm 030. MR 2008h:17033 Zbl 1138.17013

Received October 2, 2011. Revised February 23, 2012.

HIROKI SHIMAKURA
GRADUATE SCHOOL OF INFORMATION SCIENCES
TOHOKU UNIVERSITY
ARAMAKI AZA AOBA 6-3-09, AOBA-KU
SENDAI-CITY 980-8579
JAPAN
shimakura@m.tohoku.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 258, No. 2, 2012

HIGHEST-WEIGHT VECTORS FOR THE ADJOINT ACTION OF GL_n ON POLYNOMIALS

RUDOLF TANGE

Let $G = \mathrm{GL}_n$ be the general linear group over an algebraically closed field k, and let $\mathfrak{g} = \mathfrak{gl}_n$ be its Lie algebra. Let U be the subgroup of G that consists of the upper unitriangular matrices. Let $k[\mathfrak{g}]$ be the algebra of polynomial functions on $\mathfrak{g}$, and let $k[\mathfrak{g}]^G$ be the algebra of invariants under the conjugation action of G. For certain special weights, we give explicit bases for the $k[\mathfrak{g}]^G$-module $k[\mathfrak{g}]^U_\lambda$ of highest-weight vectors of weight λ. For five of these special weights, we show that this basis is algebraically independent over $k[\mathfrak{g}]^G$ and generates the $k[\mathfrak{g}]^G$-algebra $\bigoplus_{r\geq 0} k[\mathfrak{g}]^U_{r\lambda}$. Finally, we formulate the question whether in characteristic zero, $k[\mathfrak{g}]^G$-module generators of $k[\mathfrak{g}]^U_\lambda$ can be obtained by applying one explicit highest-weight vector of weight λ in the tensor algebra $T(\mathfrak{g})$ to varying tuples of fundamental invariants.

Introduction

Let GL_n be the general linear group over an algebraically closed field k, and let $\mathfrak{gl}_n$ be its Lie algebra. We are interested in explicit formulas for highest-weight vectors in the ring $k[\mathfrak{gl}_n]$ of polynomial functions on $\mathfrak{gl}_n$ under the conjugation action. It is natural to take into account the fact that the highest-weight vectors of a given weight form a module over the invariant algebra $k[\mathfrak{gl}_n]^{\mathrm{GL}_n}$. A crude method would be to map the highest-weight vectors in the tensor algebra $T(\mathfrak{gl}_n)$ (see, for example, [Benkart et al. 1994]) into the symmetric algebra $S(\mathfrak{gl}_n)$, which is GL_n-equivariantly isomorphic to $k[\mathfrak{gl}_n]$. Mostly one will be projecting to zero. For example, in [Premet and Tange 2005, Section 5, Corollary 2], it was shown that the lowest degree in $k[\mathfrak{gl}_n]$ where the irreducible of highest weight $n\varpi_1$ occurs is $n(n-1)/2$. But the lowest degree in $T(\mathfrak{gl}_n)$ where this irreducible occurs is $n-1$. Our method involves differentiation of the fundamental invariants and applies to any relevant weight, although we can only prove that it provides a $k[\mathfrak{gl}_n]^{\mathrm{GL}_n}$-module basis for a special family of weights.

This research was funded by a research grant from The Leverhulme Trust.
MSC2010: primary 13A50; secondary 16W22, 20G05.
Keywords: highest-weight vectors, semi-invariants, adjoint action.

Kostant [1963] showed that, for any reductive group G over $\mathbb{C}$, the coordinate rings of the fibres of the adjoint quotient are all isomorphic as G-modules to the space H of harmonic functions, and determined the multiplicities of the irreducibles in H. Hesselink [1980] obtained a completely general formula for the graded character of H (or the coordinate ring of the nilpotent cone). For more results on multiplicities in the tensor, symmetric and exterior algebra of the Lie algebra we refer the reader to [Hanlon 1985; Stembridge 1987; Gupta 1987; Brylinski 1989; Reeder 1997; Bazlov 2001] and the references in there.

The paper is organised as follows. In Section 1, we introduce some basic notation and recall some results from the literature. Section 2 contains the main results: Theorem 1 gives explicit $k[\mathfrak{gl}_n]^{\mathrm{GL}_n}$-module bases for the space of highest-weight vectors for a family of $2(n-1)-1$ weights, and Theorem 2 extends this to all the multiples of 5 of these weights. Theorems 1 and 2 generalise the results in [Premet and Tange 2005, Section 5] for the weight $n\varpi_1$. See also [Dixmier 1976, lemme 3.4] for the case of the universal enveloping algebra of $\mathfrak{sl}_n$. In Section 3, we briefly consider the example GL_3. Here one can actually determine $k[\mathfrak{gl}_n]^{\mathrm{GL}_n}$-module bases for the space of highest-weight vectors for all relevant weights; that is, one can completely determine the algebra $k[\mathfrak{gl}_n]^{U_n}$, where U_n consists of the upper unitriangular matrices. In Section 4, we formulate the question whether in characteristic zero, $k[\mathfrak{gl}_n]^{\mathrm{GL}_n}$-module generators of $k[\mathfrak{gl}_n]^U_\lambda$ can be obtained by applying one explicit highest-weight vector of weight λ in the tensor algebra $T(\mathfrak{gl}_n)$ to varying tuples of fundamental invariants.

1. Preliminaries

Throughout this paper k is an algebraically closed field and $G=\mathrm{GL}_n$, $n\geq 2$, is the general linear group of invertible $n\times n$ matrices. Its natural module is $V=k^n$ and its Lie algebra is $\mathfrak{g}=\mathfrak{gl}_n\cong V\otimes V^*$. The standard basis elements of V are denoted by $e_1,\dots,e_n$ and the dual basis elements are denoted by $e_1^*,\dots,e_n^*$. We identify $\mathfrak{g}=\mathfrak{gl}_n$ with $\mathrm{End}(V)$, the endomorphisms of the vector space V. We denote by E_{ij} the matrix which is 1 on position (i,j) and 0 elsewhere. Under the isomorphism $\mathfrak{g}\cong V\otimes V^*$, E_{ij} corresponds to $e_i\otimes e_j^*$. The elements of the dual basis of E_{ij} are denoted by ξ_{ij}. So the algebra $k[\mathfrak{g}]$ of polynomial functions on $\mathfrak{g}$ is a polynomial algebra in the ξ_{ij}. The group G acts on $\mathfrak{g}$ via the adjoint action (conjugation) and therefore also on $k[\mathfrak{g}]$. For any group H and any kH-module W we denote the space of H-fixed vectors in W by W^H.

The Borel subgroup of G which consists of the invertible upper triangular matrices is denoted by B and its unipotent radical, which consists of the upper unitriangular matrices, by U. We denote by T the maximal torus of G which consist of the invertible diagonal matrices. The character group of T is denoted by X and its standard basis elements are denoted by $\varepsilon_1, \dots, \varepsilon_n$. Recall that the positive roots relative to B are the roots $\varepsilon_i - \varepsilon_j$ for $i < j$, and that $\lambda \in X$ is dominant if and only if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Furthermore, $\lambda \in X$ occurs in the root lattice if and only if its coordinate sum is 0. The all-zero and the all-one vector in X are denoted by $\underline{0}$ and $\underline{1}$ respectively. For $i \in \{1, \dots, n-1\}$ the i-th fundamental weight $\varpi_i \in \mathbb{Q} \otimes_{\mathbb{Z}} X$ is defined by

$$\varpi_i = \sum_{j=1}^{i} \varepsilon_j - \frac{i}{n}\underline{1} = \frac{1}{n}\Bigl((n-i)\sum_{j=1}^{i} \varepsilon_j - i \sum_{j=i+1}^{n} \varepsilon_j\Bigr).$$

The $\mathbb{Z}$-span of the fundamental weights contains the root lattice. For $\lambda \in X$ and W a T-module the weight space W_λ is defined by

$$W_\lambda = \{x \in W \mid t \cdot x = \lambda(t)x \text{ for all } t \in T\}.$$

We denote the irreducible $\mathrm{GL}_n(\mathbb{C})$-module of highest weight λ by $L_{\mathbb{C}}(\lambda)$. The Weyl group of G relative to T is the symmetric group Sym_n which permutes the coordinates. We denote the longest Weyl group element by w_0. We have $w_0(\varepsilon_i) = \varepsilon_{n-i+1}$, put differently, $w_0(\lambda)$ is the reversed tuple of λ.

For $i \in \{1, \dots, n\}$ we define $s_i \in k[\mathfrak{g}]$ by

$$s_i(x) = \operatorname{tr} \textstyle\bigwedge^i(x),$$

where $\bigwedge^i(x)$ denotes the i-th exterior power of x. Then the s_i are up to sign the coefficients of the characteristic polynomial. Note that $s_1 = \operatorname{tr}$ and $s_n = \det$. Furthermore, the s_i are algebraically independent generators of $k[\mathfrak{g}]^G$. See, e.g., [Jantzen 2004, Section 7].

The reader who only wants to understand the precise statements of the main results can now continue to Section 2, read definitions (1) and (2) and then Theorems 1 and 2.

We now state some auxiliary results that will be needed for the proofs of the main results. The result below was mentioned to me by S. Donkin.

Lemma 1. $$\dim k[\mathfrak{g}]^U = \dim B = n(n+1)/2.$$

Proof. For $m \in \{1, \dots, n\}$ put

$$\Delta_m = \det\bigl((\xi_{ij})_{n-m+1 \leq i \leq n,\, 1 \leq j \leq m}\bigr).$$

Then $\Delta_m \in k[\mathfrak{g}]^U$ for all $m \in \{1, \dots, n\}$ and $k[\mathfrak{g}][\Delta_1^{-1}, \dots, \Delta_n^{-1}] = k[Bw_0B]$. It follows that

$$k[\mathfrak{g}]^U[\Delta_1^{-1}, \dots, \Delta_n^{-1}] = k[Bw_0B]^U$$

and $\dim k[\mathfrak{g}]^U = \dim k[Bw_0B]^U$. Now $k[Bw_0B]^U \cong k[B]$ via the isomorphism that sends $f \in k[B]$ to the function $uw_0b \mapsto f(bu)$. □

We recall the graded Nakayama lemma. For its proof we refer to [Passman 1991, Chapter 13], Lemma 4, Exercise 3, Lemma 3.

Lemma 2 [Passman 1991, Chapter 13]. *Let $S = \bigoplus_{i\geq 0} S^i$ be a positively graded ring with S^0 a field, let M be a graded S-module and let $(x_i)_{i\in I}$ be a family of homogeneous elements of M. Put $S^+ = \bigoplus_{i>0} S^i$.*

(i) *If the images of the x_i in M/S^+M span the vector space M/S^+M over S^0, then the x_i generate M.*

(ii) *If M is projective and the images of the x_i in M/S^+M form an S^0-basis of M/S^+M, then $(x_i)_{i\in I}$ is an S-basis of M.*

The closed subvariety of $\mathfrak{g}$ which consists of the nilpotent matrices is denoted by $\mathcal{N}$. Since $\mathcal{N}$ is G-stable, G acts on the algebra $k[\mathcal{N}]$ of regular functions on $\mathcal{N}$. The two results below are actually valid, under some mild assumptions, for arbitrary reductive groups, but we will not need this generality.

Proposition 1 [Kostant 1963, Theorem 11; Jantzen 2004, Section 7; Donkin 1988, Theorem 2.2; Donkin 1990, Proposition 1.3b(i)].

(i) *The vanishing ideal of $\mathcal{N}$ in $k[\mathfrak{g}]$ is generated by $s_1, \dots, s_n$ and for each λ the restriction $k[\mathfrak{g}]^U_\lambda \to k[\mathcal{N}]^U_\lambda$ is surjective and has kernel $(k[\mathfrak{g}]^G)^+ k[\mathfrak{g}]^U_\lambda$.*

(ii) *We have $k[\mathfrak{g}]^U_\lambda \neq 0$ if and only if λ is dominant and lies in the root lattice.*

(iii) *If λ is dominant and lies in the root lattice, then $\dim k[\mathcal{N}]^U_\lambda = \dim L_{\mathbb{C}}(\lambda)_{\underline{0}}$ and $k[\mathfrak{g}]^U_\lambda$ is a free $k[\mathfrak{g}]^G$-module of rank $\dim L_{\mathbb{C}}(\lambda)_{\underline{0}}$.*

Note that $\dim L_{\mathbb{C}}(\lambda)_{\underline{0}} = \dim L_{\mathbb{C}}(-w_0(\lambda))_{\underline{0}}$, since the nondegenerate pairing between $L_{\mathbb{C}}(\lambda)$ and $L_{\mathbb{C}}(-w_0(\lambda)) = L_{\mathbb{C}}(\lambda)^*$ restricts to one between $L_{\mathbb{C}}(\lambda)_{\underline{0}}$ and $L_{\mathbb{C}}(-w_0(\lambda))_{\underline{0}}$.

We will call a weight $\lambda \in X$ *primitive* if it is nonzero, dominant, occurs in the root lattice and cannot be written as the sum of two such weights. Note that $k[\mathfrak{g}]$ is a unique factorisation domain, since it is isomorphic to a polynomial ring.

Lemma 3. *Let $u \in k[\mathfrak{g}]$ be nonzero. Assume that its top degree term does not vanish on $\mathcal{N}$ and is a B-semi-invariant of primitive weight λ. Then u is irreducible.*

Proof. If the top degree term of u is irreducible, then so is u. So we may assume that u is homogeneous. We now finish with the arguments from part 3 of the proof of [Premet and Tange 2005, Proposition 3]. Let $u = u_1^{m_1} \cdots u_r^{m_r}$ be the factorisation of u into irreducibles. Then the u_i are homogeneous. By a standard argument using the uniqueness of the prime factorisation and the connectedness of B, we get that the u_i are B-semi-invariants. Let $\lambda_1, \dots, \lambda_r$ be their weights. Then these are dominant by [Jantzen 2003, Proposition II.2.6] and we have $\lambda = \sum_{i=1}^r m_i\lambda_i$. So, by the primitivity of λ, we get that for precisely one i, $\lambda_i \neq 0$ and for this i we have $m_i = 1$. We may assume $i = 1$. Then $\lambda_1 = \lambda$ and $\lambda_2 = \cdots = \lambda_r = 0$. So

$u_2, \ldots, u_r$ are B-invariants and therefore G-invariants. Since u is nonzero on $\mathcal{N}$, we have by Proposition 1(i) that $r = 1$. □

2. The basic semi-invariants

For $t \in \{1, \ldots, n-1\}$ we define the weights

$$\lambda^t = \sum_{i=n-t+1}^{n} (\varepsilon_1 - \varepsilon_i) = (t, 0, \ldots, 0, -1, \ldots, -1),$$
$$\mu^t = \sum_{i=1}^{t} (\varepsilon_i - \varepsilon_n) = (1, \ldots, 1, 0, \ldots, 0, -t)\,. \tag{1}$$

Note that λ^t and μ^t are dominant and in the root lattice. We have $\lambda^1 = \mu^1 = \varepsilon_1 - \varepsilon_n$ and $\mu^t = -w_0(\lambda^t)$. Furthermore, we have $\lambda^t = t\varpi_1 + \varpi_{n-t}$ and $\mu^t = \varpi_t + t\varpi_{n-1}$. A weight $\sum_{i=1}^{n-1} m_i \varpi_i$ occurs in the root lattice if and only if $n \mid \sum_{i=1}^{n-1} i m_i$. From this we easily deduce that λ^t and μ^t are primitive.

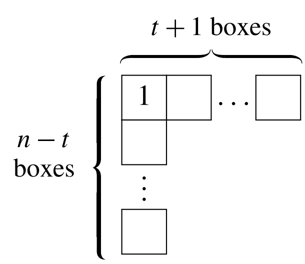
All (Young) tableaux that we consider will have entries in $\{1, \ldots, n\}$. Recall that a tableaux is called *standard* if the entries in the rows are increasing (i.e., non-decreasing) from left to right and if the entries in the columns are strictly increasing from top to bottom.

Lemma 4. *Let* $t \in \{1, \ldots, n-1\}$.

(i) *We have* $\dim k[\mathcal{N}]^U_{\lambda^t} = \dim k[\mathcal{N}]^U_{\mu^t} = \binom{n-1}{t}$.

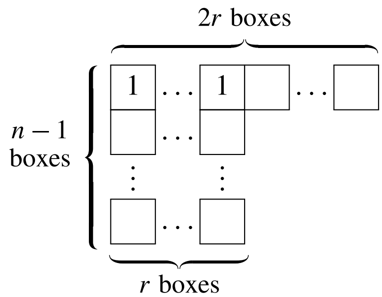
(ii) *Assume* $t = 1$ *or* $n \geq 3$ *and* $t \in \{1, n-2, n-1\}$, *let* $r \geq 0$ *be an integer and put* $s = \binom{n-1}{t}$. *Then* $\dim k[\mathcal{N}]^U_{r\lambda^t} = \dim k[\mathcal{N}]^U_{r\mu^t} = \binom{r+s-1}{r}$.

Proof. (i) We only have to consider the case of λ^t. The given dimension is by Proposition 1 equal to $\dim L_{\mathbb{C}}(\lambda^t)_{\underline{0}}$. Put $\nu := \lambda^t + \underline{1} = (t+1, 1, \ldots, 1, 0, \ldots, 0)$, where the number of zeros is t. Then $L_{\mathbb{C}}(\nu) = \det \otimes L_{\mathbb{C}}(\lambda^t)$. So it suffices to show that $\dim L_{\mathbb{C}}(\nu)_{\underline{1}} = \binom{n-1}{t}$. This dimension is well-known to be equal to the number of standard tableaux of shape ν and weight $\underline{1}$, that is, each integer in $\{1, \ldots, n\}$ must occur precisely once. The shape ν is a hook diagram as shown:



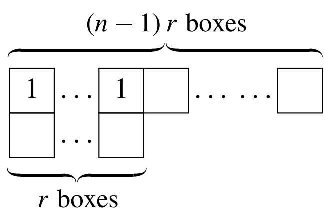
Clearly the box in the top left corner must contain 1 and the tableaux is completely determined by the choices for the other boxes in the first column. So our standard tableaux are in one-one correspondence with the $n-t-1$-subsets of $\{2,\dots,n\}$.

(ii) We only have to consider the case of λ^t. By the same arguments as in (i), it suffices to show that the number of standard tableaux of shape ν and weight $r\underline{1}$ is $\binom{r+s-1}{r}$, where $\nu := r\lambda^t + r\underline{1}$. So each integer in $\{1,\dots,n\}$ must occur precisely r times. First assume $t=1$. Then $s=n-1$ and the shape ν is a diagram as shown:



Clearly the first r boxes in the top row must contain 1. If we ignore the first row, then each column is a strictly increasing subsequence of $\{2,\dots,n\}$ of length $n-2$. So it is determined by an integer from $\{2,\dots,n\}$ (the one that does not occur). If we write these in the order of the columns, then the standardness implies that we get an increasing sequence. This sequence is what goes in the final r boxes in the first row and it determines the tableaux completely. The number of such sequences is the same as the number of monomials of degree r in $n-1$ variables, so it equals $\binom{n+r-2}{r}$.

Now assume that $t=n-2$. Then $s=n-1$ and the shape ν is a diagram as shown below.



Again the first r boxes in the top row must contain 1. Now the diagram is completely determined by the second row, which is an increasing subsequence of $\{2,\dots,n\}$. So again we get $\binom{n+r-2}{r}$ standard tableaux. The case $t=n-1$ is trivial, since the shape ν is then a single row of length nr. $\square$

We now define some basic B-semi-invariants in $k[\mathfrak{g}]$. For $t\in\{1,\dots,n-1\}$ and $I\subseteq\{2,\dots,n\}$ with $|I|=t$ we define

$$
\begin{aligned}
u_{t,I} &:= \det\big((\partial_{1i}s_j)_{n-t+1\le i\le n,\, j\in I}\big),\\
v_{t,I} &:= \det\big((\partial_{in}s_j)_{1\le i\le t,\, j\in I}\big).
\end{aligned}
\tag{2}
$$

Here the indices from I are taken in their natural order and ∂_{ij} is the partial derivative $\partial/\partial\xi_{ij}$. Note that $u_{t,I}$ and $v_{t,I}$ are homogeneous of degree $\left(\sum_{j\in I} j\right)-t$.

Define the involution φ of the vector space $\mathfrak{g}$ by $\varphi(A)=PA^TP$, where P is the permutation matrix corresponding to w_0 and A^T denotes the transpose of A. Then $\varphi(g\cdot A)=P(g^{-1})^TP\cdot\varphi(A)$, where the dot denotes conjugation action. If we denote the corresponding automorphism of $k[\mathfrak{g}]$ also by φ, then this formula also holds with A replaced by $f\in k[\mathfrak{g}]$. So $\varphi(k[\mathfrak{g}]^U_\lambda)=k[\mathfrak{g}]^U_{-w_0(\lambda)}$. In accordance with this we have $\varphi(u_{t,I})=\pm v_{t,I}$.

We set up some notation which will give another, more general, way to construct the elements $u_{t,I}$ and $v_{t,I}$. This will make clear why they are B-semi-invariants (see the proof of Theorem 1(ii) below). If λ is a partition, then we denote its length by $l(\lambda)$. For $\lambda^+,\lambda^-\in X$ we put $[\lambda^+,\lambda^-]:=\lambda^+-w_0(\lambda^-)$. It is easy to see that for any $\lambda\in X$ dominant there exist unique partitions λ^+ and λ^- with $l(\lambda^+)+l(\lambda^-)\le n$ and $\lambda=[\lambda^+,\lambda^-]$. In the sequel, when λ^+ and λ^- are introduced after λ, they are supposed to have these properties. Let λ be a partition of t. We define the tableau T_λ of shape λ by $T_\lambda(i,j)=\left(\sum_{l=1}^{i-1}\lambda_l\right)+j$. Furthermore we define the subgroup C_λ of the symmetric group Sym_t as the column stabiliser of T_λ. Define the element A_λ in the group algebra $k\langle\mathrm{Sym}_t\rangle$ by $A_\lambda=\sum_{\pi\in C_\lambda}\mathrm{sgn}(\pi)\pi$. Finally, define $e_\lambda\in V^{\otimes t}$ and $e^*_\lambda\in V^{*\otimes t}$ by

$$e_\lambda=\bigotimes_{i=1}^{l(\lambda)}e_i^{\otimes\lambda_i}\quad\text{and}\quad e^*_\lambda=\bigotimes_{i=1}^{l(\lambda)}{e^*_{n-i+1}}^{\otimes\lambda_i}.$$

Then, as is well-known (see, e.g., [Benkart et al. 1994]), $A_\lambda\cdot e_\lambda$ and $A_\lambda\cdot e^*_\lambda$ are highest-weight vectors of weight λ and $-w_0(\lambda)$ respectively.

Now let $\lambda=[\lambda^+,\lambda^-]$ be dominant and in the root lattice. Then λ^+ and λ^- are partitions of the same number, t say and we define $E_\lambda\in\mathfrak{g}^{\otimes t}$ as the element corresponding to $A_{\lambda^+}\cdot e_{\lambda^+}\otimes A_{\lambda^-}\cdot e^*_{\lambda^-}\in V^{\otimes t}\otimes V^{*\otimes t}$ under the isomorphism $\mathfrak{g}^{\otimes t}\cong V^{\otimes t}\otimes V^{*\otimes t}$. By the above, E_λ is a highest-weight vector of weight λ.

For each $x\in\mathfrak{g}$ we can extend the evaluation at x, considered as a linear map $\mathfrak{g}^*\to k\subseteq k[\mathfrak{g}]$, to a derivation of degree -1 of the algebra $k[\mathfrak{g}]$. Then the evaluation at E_{ij} extends to the derivation ∂_{ij}. So we obtain a G-equivariant linear map $\mathfrak{g}\to\mathrm{End}(k[\mathfrak{g}])$ and therefore also a G-equivariant linear map

$$\psi_t:\mathfrak{g}^{\otimes t}\to\mathrm{End}(k[\mathfrak{g}]^{\otimes t})\,.$$

We denote the G-equivariant multiplication map $k[\mathfrak{g}]^{\otimes t}\to k[\mathfrak{g}]$ by ϑ.

Theorem 1. *Let $t\in\{1,\dots,n-1\}$ and let $\lambda^t,\mu^t,u_{t,I},v_{t,I}$ be given by* (1) *and* (2).

(i) *The $u_{t,I}$, $I\subseteq\{2,\dots,n\}$ with $|I|=t$, form a basis of the $k[\mathfrak{g}]^G$-module $k[\mathfrak{g}]^U_{\lambda^t}$. The same holds for the $v_{t,I}$ and μ^t.*

(ii) *Any nontrivial k-linear combination of the* $u_{t,I}$, $I \subseteq \{2, \dots, n\}$ *with* $|I| = t$, *is an irreducible B-semi-invariant of weight* λ^t. *The same holds for the* $v_{t,I}$ *and* μ^t.

Proof. (i) Using the involution φ we see that we only have to prove the assertion for μ^t and the $v_{t,I}$. By Proposition 1 and Lemmas 2 and 4(i) it suffices to show that the restrictions of the $v_{t,I}$ to $\mathcal{N}$ are linearly independent. For $\Lambda_1, \Lambda_2 \subseteq \{1, \dots, n\}$ and $A = (a_{ij})_{1 \le i,j \le n} \in \mathfrak{g}$ set $A_{\Lambda_1, \Lambda_2} = (a_{ij})_{(i,j) \in \Lambda_1 \times \Lambda_2}$, where the indices are taken in their natural order. Furthermore, put $\mathscr{X} = (\xi_{ij})_{1 \le i,j \le n}$. If $|\Lambda_1| = |\Lambda_2|$, then we have, as in [Premet and Tange 2005], the following basic fact which follows from the Laplace expansion formulae for the determinant:

$$\partial_{ij}\big(\det(\mathscr{X}_{\Lambda_1, \Lambda_2})\big) = \begin{cases} \pm \det(\mathscr{X}_{\Lambda_1 \setminus \{i\}, \Lambda_2 \setminus \{j\}}) & \text{when } (i, j) \in \Lambda_1 \times \Lambda_2, \\ 0 & \text{when } (i, j) \notin \Lambda_1 \times \Lambda_2. \end{cases} \tag{3}$$

For $l \le n$ we have $s_l = \sum_\Lambda \det(\mathscr{X}_{\Lambda, \Lambda})$ where the sum ranges over all l-subsets Λ of $\{1, \dots, n\}$.

For a sequence $\sigma = (\sigma_1, \dots, \sigma_s)$ of distinct integers in $\{1, \dots, n\}$ we define $A_\sigma \in \mathrm{End}(V)$ by $A_\sigma(e_{\sigma_i}) = e_{\sigma_{i-1}}$ for $i \in \{2, \dots, s\}$ and $A_\sigma(e_i) = 0$ for $i \notin \{\sigma_2, \dots, \sigma_s\}$. Then A_σ is nilpotent and its restriction to the span of the e_{σ_i}, $1 \le i \le s$, is regular.

If $\Lambda_1, \Lambda_2 \subseteq \{1, \dots, n\}$ with $|\Lambda_1| = |\Lambda_2| > 0$ and $\det(\mathscr{X}_{\Lambda_1, \Lambda_2})(A_\sigma) \ne 0$, then

- $\Lambda_1 \subseteq \{\sigma_1, \dots, \sigma_{s-1}\}$ and $\Lambda_2 \subseteq \{\sigma_2, \dots, \sigma_s\}$,

(4)
- $\sigma_j \in \Lambda_1 \Rightarrow \sigma_{j+1} \in \Lambda_2$ for all $j \in \{1, \dots, s-1\}$,
- $\sigma_j \in \Lambda_2 \Rightarrow \sigma_{j-1} \in \Lambda_1$ for all $j \in \{2, \dots, s\}$.

Let σ be as above with $\sigma_1 = n$. Let $i \in \{1, \dots, n\}$, let $\Lambda \subseteq \{1, \dots, n\}$ with $|\Lambda| = l$ and assume that $\big(\partial_{in} \det(\mathscr{X}_{\Lambda, \Lambda})\big)(A_\sigma) \ne 0$. Then it follows from (3) and (4) that $i = \sigma_l$, that $\Lambda = \{\sigma_1, \dots, \sigma_l\}$ and that $\big(\partial_{in} \det(\mathscr{X}_{\Lambda, \Lambda})\big)(A_\sigma) = \pm 1$. So for such a σ we have $(\partial_{in} s_l)(A_\sigma) \ne 0 \Rightarrow l \le s$, $i = \sigma_l$ and $(\partial_{in} s_l)(A_\sigma) = \pm 1$.

So for $\sigma = (\sigma_1, \dots, \sigma_s)$ and $\tau = (\tau_1, \dots, \tau_t)$ sequences of distinct integers in $\{1, \dots, n\}$ and $\pi \in \mathrm{Sym}_t$ with $\sigma_1 = n$ and $(\partial_{\pi_1 n} s_{\tau_1}) \cdots (\partial_{\pi_t n} s_{\tau_t})(A_\sigma) \ne 0$ we have

(a) $\tau_i \le s$ for all $i \in \{1, \dots, t\}$,

(b) $\sigma \circ \tau = \pi$,

(c) $(\partial_{\pi_1 n} s_{\tau_1}) \cdots (\partial_{\pi_t n} s_{\tau_t})(A_\sigma) = \pm 1$.

Note that (a) implies that $\sigma(\{\tau_1, \dots, \tau_t\}) = \{1, \dots, t\}$, so the set $\{\tau_1, \dots, \tau_t\}$ is determined by σ.

Now we choose for each subset $I = \{i_1 > \dots > i_t\}$ of $\{2, \dots, n\}$ a sequence $\sigma(I)$ of $i_1 \ge t+1$ distinct integers in $\{1, \dots, n\}$ with $\sigma(I)_1 = n$ and $\sigma(I)_{i_j} = j$ for all $j \in \{1, \dots, t\}$. Then we get for $I, J \subseteq \{2, \dots, n\}$ with $|I| = |J| = t$ that

$$v_{t,I}(A_{\sigma(J)}) = \begin{cases} \pm 1 & \text{if } I = J, \\ 0 & \text{otherwise.} \end{cases}$$

So the linear map $f \mapsto f(A_{\sigma(J)})_J : k[\mathcal{N}] \to k^{\binom{n-1}{t}}$ sends the family $(v_{t,I}|_{\mathcal{N}})_I$ to a basis and therefore the restrictions of the $v_{t,I}$ to $\mathcal{N}$ are linearly independent.

(ii) Let $I \subseteq \{2, \dots, n\}$ with $|I| = t$ and write $I = \{i_1 < \cdots < i_t\}$. Then it follows immediately from the definitions that $u_{t,I} = \vartheta(\psi_t(F) \cdot s_{i_1} \otimes \cdots \otimes s_{i_t})$, where

$$F = \sum_{\pi} \operatorname{sgn}(\pi)\, E_{1,\pi_{n-t+1}} \otimes \cdots \otimes E_{1,\pi_n},$$

the sum over all permutations $\pi \in \operatorname{Sym}(\{n-t+1, \dots, n\})$. Now $\lambda_t^+ = t\varepsilon_1$ and $\lambda_t^- = \varepsilon_1 + \cdots + \varepsilon_t$. So $A_{\lambda_t^+} = \mathrm{id}$ and $A_{\lambda_t^-} = \sum_{\pi \in \mathrm{Sym}_t} \operatorname{sgn}(\pi)\pi$ for $e_{\lambda_t^+} = e_1^{\otimes t}$ and $e_{\lambda_t^-} = e_{n-t+1}^* \otimes \cdots \otimes e_n^*$. It follows that, under the isomorphism $\mathfrak{g}^{\otimes t} \cong V^{\otimes t} \otimes V^{*\otimes t}$, F corresponds to $A_{\lambda_t^+} \cdot e_{\lambda_t^+} \otimes A_{\lambda_t^-} \cdot e_{\lambda_t^-}^*$. So $F = E_{\lambda^t}$. Similarly, we get $v_{t,I} = \vartheta(\psi_t(E_{\mu^t}) \cdot s_{i_1} \otimes \cdots \otimes s_{i_t})$. Since the s_i are invariants, this shows that $u_{t,I}$ and $v_{t,I}$ are B-semi-invariants of the given weights. Since λ_t and μ_t are primitive, the assertion follows from Lemma 3 and the linear independence proved in (i). □

Remarks 1. 1. Kostant [1963, Remark 26] gave an explicit basis for the isotypic component of the space of harmonics H corresponding to the highest root. So the statement of Theorem 1 in the case of λ^1 extends to all complex reductive groups.

2. Assume $k = \mathbb{C}$, let $t \le s$ and let $\lambda = [\lambda^+, \lambda^-]$ be dominant and in the root lattice with λ^+ and λ^- partitions of t. Then

$$(\mathfrak{g}^{\otimes s})^U_\lambda \cong (V^{\otimes s} \otimes V^{*\otimes s})^U_\lambda$$

is a simple module for the *walled Brauer algebra* $\mathfrak{B}_{s,s}(n)$, see [Benkart et al. 1994]. Note that in the definition of the vectors $t_{\tau,\underline{m},\underline{n}}$ in Definition 2.4 of that reference, the symmetrisation can be omitted. Above we only considered the case $s = t$, the lowest tensor power of $\mathfrak{g}$ which contains $L_{\mathbb{C}}(\lambda)$. Then $(\mathfrak{g}^{\otimes s})^U_\lambda$ is an irreducible $\mathrm{Sym}_t \times \mathrm{Sym}_t$-module and the ideal of $\mathfrak{B}_{t,t}(n)$ spanned by the diagrams with at least one horizontal edge acts as 0.

3. Another natural definition of e_λ and e_λ^* is

$$e_\lambda = \bigotimes_{i=1}^{l(\lambda')} \bigotimes_{j=1}^{\lambda_i'} e_j \quad \text{and} \quad e_\lambda^* = \bigotimes_{i=1}^{l(\lambda')} \bigotimes_{j=1}^{\lambda_i'} e_{n-j+1}^*,$$

where λ' denotes the partition of t whose shape is the transpose of that of λ. In the definition of A_λ one then has to replace T_λ by its transpose (or C_λ by the row stabiliser R_λ). Then $A_\lambda \cdot e_\lambda$ and $A_\lambda \cdot e_\lambda^*$ are again highest-weight vectors of weight

λ and $-w_0(\lambda)$ and one can define E_λ as before. Note that this E_λ is $\mathrm{Sym}_t\times\mathrm{Sym}_t$-conjugate to the original one.

4. Assume $k=\mathbb{C}$. Theorem 1 answers the so-called *first occurrence* question for $k[\mathfrak{g}]$ and the weights λ^t and μ^t: The lowest degree where $L_{\mathbb{C}}(\lambda^t)$ (or $L_{\mathbb{C}}(\mu^t)$) occurs in $k[\mathfrak{g}]$ is $\left(\sum_{i=2}^{t+1} i\right)-t=t(t+1)/2$.

Theorem 2. *Assume $t=1$ or $n\geq 3$ and $t\in\{1,n-2,n-1\}$. Then the $u_{t,I}$, $I\subseteq\{2,\dots,n\}$ with $|I|=t$, are algebraically independent over $k[\mathfrak{g}]^G$ and generate the $k[\mathfrak{g}]^G$-algebra $\bigoplus_{r\geq 0}k[\mathfrak{g}]^U_{r\lambda^t}$. Furthermore, the same holds for the $v_{t,I}$ and μ^t.*

Proof. Using the involution φ we see that we only have to prove the assertion for μ^t and the $v_{t,I}$. By Proposition 1 and Lemmas 2 and 4(ii) it suffices to show that the restrictions of the $v_{t,I}$ to $\mathcal{N}$ are algebraically independent. If $t=n-1$, then this follows from the fact that $v_{n-1,\{2,\dots,n\}}|_{\mathcal{N}}$ is nonzero by Theorem 1(i) and of degree >0. Now we observe the following. If $f_1,\dots,f_l\in k[\mathcal{N}]$, then the morphism $(f_1,\dots,f_l):\mathcal{N}\to k^l$ is dominant if and only if the f_i are algebraically independent and its differential at a point $x\in\mathcal{N}$ is surjective if and only if the differentials at x of the f_i are linearly independent. So, by [Borel 1991, AG 17.3], it suffices to show that the differentials of the $v_{t,I}|_{\mathcal{N}}$ are linearly independent at some smooth point $x\in\mathcal{N}$. For $x\in\mathcal{N}$ we have that $T_x(\mathcal{N})$ is the intersection of the kernels of the differentials d_xs_i and x is a smooth point of $\mathcal{N}$ if and only if the d_xs_i are linearly independent. So it suffices to show that the differentials of the s_i and the $v_{t,I}$ at some nilpotent element x are together linearly independent. We will take $x=A=A_\sigma$, where $\sigma=(n,n-1,\dots,1)$ and the notation is as in the proof of Theorem 1(i). Put

$$\alpha=\big((1,1),\dots,(1,n),(n,1),\dots,(n,n-2),(2,1)\big).$$

Let M be the Jacobian matrix of $s_1,\dots,s_n$ and the $v_{t,I}$ and let M_α be the $(2n-1)$-square submatrix of M consisting of the columns with indices from α. We will show that $\det(M_\alpha)(A)=\pm 1$. This will prove the required linear independence.

From (3) and (4) we deduce easily that $(\partial_{ni}s_j)(A)=0$ and $(\partial_{21}s_j)(A)=0$ for all $i\in\{1,\dots,n-2\}$ and $j\in\{1,\dots,n\}$ and that $(\partial_{1i}s_j)(A)=\pm\delta_{ij}$ for all $i,j\in\{1,\dots,n\}$. So it suffices to show that the matrix $(\partial_{\alpha_i}v_{t,J})(A)_{n+1\leq i\leq 2n-1,J}$ is diagonal with the diagonal entries equal to ± 1, when the subsets J are suitably ordered.

Assume $t=n-2$. For $j\in\{2,\dots,n\}$ put $w_j=v_{t,\{2,\dots,n\}\setminus\{j\}}$. Put

$$\tau(j)=(2,\dots,j-1,j+1,\dots,n).$$

Then we have

$$\partial_{\alpha_m}w_j=\partial_{\alpha_m}\sum\pm(\partial_{\pi_1,n}s_{\tau(j)_1})\cdots(\partial_{\pi_{n-2},n}s_{\tau(j)_{n-2}}), \tag{5}$$

where the sum is over all $\pi \in \mathrm{Sym}(\{1, \dots, n-2\})$. We can expand this further by applying the product rule for differentiation. Then each term in (5) produces $n-2$ terms, the differentiation ∂_{α_m} being applied to each factor in turn. As in the proof of Theorem 1 we have

(6) $$(\partial_{in}s_l)(A) \neq 0 \quad \Rightarrow \quad (\partial_{in}s_l)(A) = \pm 1 \text{ and } i = \sigma_l = n-l+1.$$

Now assume $j \geq 3$, i.e., $\sigma_j \leq n-2$. Then $\sigma_{\tau(j)_1} = \sigma_2 = n-1$. Since π never takes the value $n-1$, the only term in the expanded form of

(7) $$\partial_{\alpha_m}\big((\partial_{\pi_1,n}s_{\tau(j)_1})\cdots(\partial_{\pi_{n-2},n}s_{\tau(j)_{n-2}})\big)$$

that can be nonzero at A is $\big(\partial_{\alpha_m}(\partial_{\pi_1,n}s_2)\big)(\partial_{\pi_2,n}s_{\tau(j)_2})\cdots(\partial_{\pi_{n-2},n}s_{\tau(j)_{n-2}})$. By (6) we must then have $\pi_i = \sigma_{\tau(j)_i}$ for all $i \in \{2, \dots, n-2\}$ and $\pi_1 = \sigma_j$. Finally (3) and (4) give us then that $\alpha_m = (n, \sigma_j)$ and that the value of the term is ± 1.

Now assume that $j = 2$. Then $\tau(2) = (3, \dots, n)$. So for a term in the expanded form of (7) to be nonzero at A we must, by (6), have $\pi_i = \sigma_{\tau(2)_i}$ for all but one and therefore for all $i \in \{1, \dots, n-2\}$. So $\pi = (n-2, \dots, 1)$. Now we check that $\big(\partial_{nl}(\partial_{\pi_i,n}s_{\tau(2)_i})\big)(A) = 0$ for all $l, i \in \{1, \dots, n-2\}$ by considering a term $\det(A_{\Lambda\setminus\{\pi_i,n\},\,\Lambda\setminus\{l,n\}})$ for $\Lambda \subseteq \{1, \dots, n\}$ with $|\Lambda| = \tau(2)_i = i+2$. Assume first $1 \in \Lambda$. Then $\pi_i = 1$, since otherwise the first row of $A_{\Lambda\setminus\{\pi_i,n\},\,\Lambda\setminus\{l,n\}}$ would be zero. So $i = n-2$ and $\Lambda = \{1, \dots, n\}$. But then the column of index $n-1$ in $A_{\Lambda\setminus\{\pi_i,n\},\,\Lambda\setminus\{l,n\}}$ is zero. So $1 \notin \Lambda$. The cases $l < \pi_i$ and $l = \pi_i$ are now easily dealt with using (3) and (4). So assume $\pi_i < l$. Then we get, using (3) and (4), $\Lambda = \{\pi_i, \dots, l, n\}$. Then $i+2 = |\Lambda| = l-n+i+3$, so $l = n-1$, which is impossible. Finally we check that $\big(\partial_{2,1}(\partial_{\pi_i,n}s_{i+2})\big)(A) = \pm\delta_{i,n-2}$, by considering a term $\det(A_{\Lambda\setminus\{2,\pi_i\},\,\Lambda\setminus\{1,n\}})$ for $\Lambda \subseteq \{1, \dots, n\}$ with $|\Lambda| = i+2$. Since $1 \in \Lambda$ we must have $\pi_i = 1$, so $i = n-2$ and $\Lambda = \{1, \dots, n\}$. The value of this term is then ± 1.

In conclusion we have shown that, for $m \in \{n+1, \dots, 2n-1\}$ and $j \in \{2, \dots, n\}$, $(\partial_{\alpha_m}w_j)(A) = \pm\delta_{m-n,w_0(j)}$.

Now assume $t = 1$. Then we put $w_j = v_{i,\{j\}} = \partial_{1,n}s_j$ and we show that, for $m \in \{n+1, \dots, 2n-1\}$ and $j \in \{2, \dots, n\}$, $(\partial_{\alpha_m}w_j)(A) = \pm\delta_{m-n,j-1}$. Since this case is much easier we leave it to the reader. □

Remarks 2. 1. Assume $k = \mathbb{C}$, let $t \in \{1, n-2, n-1\}$ and let $r \geq 0$. Then, by Theorem 2, the lowest degree where $L_{\mathbb{C}}(r\lambda^t)$ (or $L_{\mathbb{C}}(r\mu^t)$) occurs in $k[\mathfrak{g}]$ is $r\big((\sum_{i=2}^{t+1} i) - t\big) = rt(t+1)/2$.

2. Computer calculations suggest that, for $t \notin \{1, n-2, n-1\}$ and $r \geq 2$, $\dim k[\mathcal{N}]^U_{r\lambda^t} < \binom{r+s-1}{r}$, where $s = \dim k[\mathcal{N}]^U_{\lambda^t}$. So for such t one cannot expect the $u_{t,I}$ to be algebraically independent over $k[\mathfrak{g}]^G$, but one could still conjecture

that they generate the $k[\mathfrak{g}]^G$-algebra $\bigoplus_{r\geq 0} k[\mathfrak{g}]^U_{r\lambda^t}$. Similar remarks apply to μ^t and the $v_{t,I}$.

3. With a bit more effort one can show that the matrix $M_\alpha(A)$ from the proof of Theorem 2 is diagonal with the diagonal entries equal to ± 1.

3. The case of GL_3

In this section we describe the algebra $k[\mathfrak{g}]^U$ in the case of GL_3. So throughout this section $n=3$, $G=\mathrm{GL}_3$ and $\mathfrak{g}=\mathfrak{gl}_3$. We have $\lambda^1=\mu^1=\varpi_1+\varpi_2$, $\lambda^2=3\varpi_1=(2,-1,-1)$ and $\mu^2=3\varpi_2=(1,1,-2)$. Note that a weight $l_1\varpi_1+l_2\varpi_2$ is in the root lattice if and only if $3\mid(l_1-l_2)$. Put $\mathscr{X}=(\xi_{ij})_{1\leq i,j\leq 3}$. For $i, j\in\{1,2,3\}$ we denote by $\mathscr{X}^{(i,j)}$ the matrix $\mathscr{X}$ with the i-th row and j-th column omitted and we denote its determinant by $|\mathscr{X}^{(i,j)}|$. We put

$$d_1=\xi_{21}|\mathscr{X}^{(1,3)}|+\xi_{31}|\mathscr{X}^{(1,2)}|=-u_{2,\{2,3\}}\,,$$
$$d_2=\xi_{31}|\mathscr{X}^{(2,3)}|+\xi_{32}|\mathscr{X}^{(1,3)}|=v_{2,\{2,3\}}\,.$$

Lemma 5. *Let $\lambda=l_1\varpi_1+l_2\varpi_2$ be dominant and in the root lattice. Put $a=\min(l_1,l_2)$. Then* $\dim L_{\mathbb{C}}(\lambda)_{\underline{0}}=a+1$.

Proof. Put $b=(l_1+2l_2)/3$ and $\nu=\lambda+b\underline{1}=(l_1+l_2,l_2,0)$. Then $L_{\mathbb{C}}(\nu)=\det^b\otimes L_{\mathbb{C}}(\lambda)$. So it suffices to show that there are $a+1$ standard tableaux of shape ν and weight $b\underline{1}$. This we leave as an exercise for the reader. One has to distinguish the cases $l_1\geq l_2$ and $l_2\geq l_1$. □

Proposition 2. (i) *Let $\lambda=l_1\varpi_1+l_2\varpi_2$ be dominant and in the root lattice and put $a=\min(l_1,l_2)$. Put $d=d_1^{(l_1-l_2)/3}$ if $l_1\geq l_2$ and put $d=d_2^{(l_2-l_1)/3}$ otherwise. Then the elements $d\,\xi_{31}^i|\mathscr{X}^{(1,3)}|^{a-i}$, $0\leq i\leq a$, form a basis of the $k[\mathfrak{g}]^G$-module $k[\mathfrak{g}]^U_\lambda$.*

(ii) *The k-algebra $k[\mathfrak{g}]^U$ is generated by s_1, s_2, s_3, ξ_{31}, $|\mathscr{X}^{(1,3)}|$, d_1 and d_2. A defining relation is given by*

$$d_1d_2-|\mathscr{X}^{(1,3)}|^3-\xi_{31}|\mathscr{X}^{(1,3)}|^2s_1-\xi_{31}^2|\mathscr{X}^{(1,3)}|s_2-\xi_{31}^3s_3=0\,.$$

Proof. (i) By Proposition 1 and Lemmas 2 and 5 it suffices to show that the given elements are independent on $\mathcal{N}$. Since they all have different degrees, it suffices to show they are nonzero on $\mathcal{N}$. One easily checks that they are all nonzero on

$$\begin{bmatrix}0&0&0\\1&0&0\\1&1&0\end{bmatrix}.$$

(ii) By (i) the 7 given elements generate $k[\mathfrak{g}]^U$ and by Lemma 1 $\dim k[\mathfrak{g}]^U=6$. A straightforward computation shows that the given equation holds and it is clearly irreducible (by Gauss's lemma, for instance). □

Remark 3. Proposition 2 also shows that the k-algebra $k[\mathcal{N}]^U$ is generated by ξ_{31}, $|\mathscr{X}^{(1,3)}|$, d_1 and d_2 with defining relation $d_1d_2 - |\mathscr{X}^{(1,3)}|^3 = 0$.

4. The method in general

As the reader may have noticed after reading the proof of Theorem 1(ii) our method for producing highest-weight vectors applies to any dominant weight in the root lattice. So one may wonder whether we always get $k[\mathfrak{g}]^G$-module generators. We formulate this as a question. We assume that $k = \mathbb{C}$ and use the notation of Section 2 before Theorem 1.

Question. Let $\lambda = [\lambda^+, \lambda^-]$ be dominant and in the root lattice with λ^+ and λ^- partitions of t. Do the elements $\vartheta\big(\psi_t(E_\lambda) \cdot s_{i_1} \otimes \cdots \otimes s_{i_t}\big)$ for $2 \le i_1, \dots, i_t \le n$ generate the $k[\mathfrak{g}]^G$-module $k[\mathfrak{g}]^U_\lambda$? Equivalently, do their restrictions to $\mathcal{N}$ span $k[\mathcal{N}]^U_\lambda$?

Note that the only thing that varies here is the tuple $(i_1, \dots, i_t)$. Note also that we allow repetitions in the arguments s_{i_j}. As an example we consider the case $n = 4$ and $\lambda = 2\varpi_2 = (1, 1, -1, -1)$, a primitive weight. Then the Hesselink–Peterson formula [Hesselink 1980] shows that $k[\mathcal{N}]^U_\lambda$ has dimension 2 with a generator in degree 2 and one in degree 4. We have

$$\vartheta\big(\psi_t(E_\lambda) \cdot s_{i_1} \otimes s_{i_2}\big) = \pm \sum \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau)\,\partial_{\sigma_1\tau_3} s_{i_1}\,\partial_{\sigma_2\tau_4} s_{i_2},$$

where the sum is over all $\sigma \in \mathrm{Sym}(\{1,2\})$ and all $\tau \in \mathrm{Sym}(\{3,4\})$. It follows that $\vartheta\big(\psi_t(E_\lambda)\cdot s_2 \otimes s_2\big) = \pm 2\det(\mathscr{X}_{\{3,4\},\{1,2\}})$, where $\mathscr{X}_{\{3,4\},\{1,2\}}$ is defined as in the proof of Theorem 1. Clearly this is nonzero on the nilpotent cone. Note that the choice (s_2, s_2) is the only choice that gives the degree 2 generator. One can check that (s_3, s_3) and (s_2, s_4) both produce semi-invariants of degree 4 that are nonzero on $\mathcal{N}$. In the case (s_2, s_4) it is nonzero on $\mathcal{N}$ in any characteristic.

By Theorem 1 the answer to our question is affirmative for the weights λ_t and μ_t. The basis elements of the spaces $k[\mathfrak{g}]^U_{r\lambda^t}$ and $k[\mathfrak{g}]^U_{r\mu^t}$, $r > 1$ and $t \in \{1, n-2, n-1\}$, from Theorem 2 are not formed in accordance with our question.

One can probably formulate a more complicated question for k of arbitrary characteristic, where one divides the expression $\vartheta\big(\psi_t(E_\lambda)\cdot s_{i_1} \otimes \cdots \otimes s_{i_t}\big)$ by a suitable integer in case of repeated arguments.

References

[Bazlov 2001] Y. Bazlov, "Graded multiplicities in the exterior algebra", *Adv. Math.* **158**:2 (2001), 129–153. MR 2002c:17004 Zbl 0979.17003

[Benkart et al. 1994] G. Benkart, M. Chakrabarti, T. Halverson, R. Leduc, C. Lee, and J. Stroomer, "Tensor product representations of general linear groups and their connections with Brauer algebras", *J. Algebra* **166**:3 (1994), 529–567. MR 95d:20071 Zbl 0815.20028

[Borel 1991] A. Borel, *Linear algebraic groups*, 2nd ed., Graduate Texts in Mathematics **126**, Springer, New York, 1991. MR 92d:20001 Zbl 0726.20030

[Brylinski 1989] R. K. Brylinski, "Limits of weight spaces, Lusztig's q-analogs, and fiberings of adjoint orbits", *J. Amer. Math. Soc.* **2**:3 (1989), 517–533. MR 90g:17011 Zbl 0729.17005

[Dixmier 1976] J. Dixmier, "Sur les algèbres enveloppantes de $\mathfrak{sl}(n, C)$ et $\mathfrak{af}(n, C)$", *Bull. Sci. Math.* (2) **100**:1 (1976), 57–95. MR 55 #8131 Zbl 0328.17003

[Donkin 1988] S. Donkin, "On conjugating representations and adjoint representations of semisimple groups", *Invent. Math.* **91**:1 (1988), 137–145. MR 89a:20047 Zbl 0639.20021

[Donkin 1990] S. Donkin, "The normality of closures of conjugacy classes of matrices", *Invent. Math.* **101**:3 (1990), 717–736. MR 91j:14040 Zbl 0822.20045

[Gupta 1987] R. K. Gupta, "Characters and the q-analog of weight multiplicity", *J. London Math. Soc.* (2) **36**:1 (1987), 68–76. MR 88i:17010 Zbl 0649.17009

[Hanlon 1985] P. Hanlon, "On the decomposition of the tensor algebra of the classical Lie algebras", *Adv. Math.* **56**:3 (1985), 238–282. MR 88a:17011 Zbl 0577.17003

[Hesselink 1980] W. H. Hesselink, "Characters of the nullcone", *Math. Ann.* **252**:3 (1980), 179–182. MR 82c:17004 Zbl 0447.17006

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, American Mathematical Society, Providence, RI, 2003. MR 2004h:20061 Zbl 1034.20041

[Jantzen 2004] J. C. Jantzen, "Nilpotent orbits in representation theory", pp. 1–211 in *Lie theory*, edited by J.-P. Anker and B. Orsted, Progr. Math. **228**, Birkhäuser, Boston, 2004. MR 2005c:14055 Zbl 1169.14319

[Kostant 1963] B. Kostant, "Lie group representations on polynomial rings", *Amer. J. Math.* **85** (1963), 327–404. MR 28 #1252 Zbl 0124.26802

[Passman 1991] D. S. Passman, *A course in ring theory*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1991. MR 91m:16001 Zbl 0783.16001

[Premet and Tange 2005] A. Premet and R. Tange, "Zassenhaus varieties of general linear Lie algebras", *J. Algebra* **294**:1 (2005), 177–195. MR 2006i:17016 Zbl 1097.17010

[Reeder 1997] M. Reeder, "Exterior powers of the adjoint representation", *Canad. J. Math.* **49**:1 (1997), 133–159. MR 98m:22021 Zbl 0878.20028

[Stembridge 1987] J. R. Stembridge, "First layer formulas for characters of $\mathrm{SL}(n, \mathbb{C})$", *Trans. Amer. Math. Soc.* **299**:1 (1987), 319–350. MR 88g:20088 Zbl 0613.20027

Received August 23, 2011. Revised February 22, 2012.

RUDOLF TANGE
SCHOOL OF MATHEMATICS
TRINITY COLLEGE DUBLIN
COLLEGE GREEN
DUBLIN
IRELAND
tanger@tcd.ie

CONTENTS

Volume 258, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 258 No. 2 August 2012

0030-8730(201208)258:2;1-2