Pacific Journal of Mathematics

Volume 261 No. 2 February 2013

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

See inside back cover or msp.org/pjm for submission instructions.

The subscription price for 2013 is US $400/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published 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

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.257

GEOGRAPHY OF SIMPLY CONNECTED NONSPIN SYMPLECTIC 4-MANIFOLDS WITH POSITIVE SIGNATURE

ANAR AKHMEDOV, MARK C. HUGHES AND B. DOUG PARK

We construct new families of closed simply connected nonspin irreducible symplectic 4-manifolds with positive signature that are interesting with respect to the geography problem.

1. Introduction

Given a closed smooth 4-manifold M, let $e(M)$ and $\sigma(M)$ denote the Euler characteristic and the signature of M, respectively. We define

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

Note that $e(M)$ and $\sigma(M)$ are in turn completely determined by $\chi_h(M)$ and $c_1^2(M)$, that is,

$$e(M) = 12\chi_h(M) - c_1^2(M) \quad \text{and} \quad \sigma(M) = c_1^2(M) - 8\chi_h(M).$$

When M is a complex surface, $\chi_h(M)$ is the holomorphic Euler characteristic of M while $c_1^2(M)$ is the square of the first Chern class of M. The classical "geography problem" in algebraic geometry, originally posed by Persson [1981], asks which ordered pairs of positive integers can be realized as the pair $(\chi_h(M), c_1^2(M))$ for some minimal complex surface M of general type. The related "botany problem", which is a lot more difficult, asks for the classification of all minimal complex surfaces with a given pair of invariants (χ_h, c_1^2).

The symplectic geography problem, first posed in [McCarthy and Wolfson 1994], asks which ordered pairs of integers can be realized as $(\chi_h(M), c_1^2(M))$ for some minimal symplectic 4-manifold M. There has been steady progress on the symplectic geography problem in recent years and the problem has been completely solved for simply connected minimal symplectic 4-manifolds with negative signature

Akhmedov was partially supported by NSF grant DMS-1005741. Hughes was partially supported by an NSERC Canada graduate scholarship. Park was partially supported by an NSERC discovery grant.
MSC2010: primary 57R17, 57R55; secondary 57M60, 57M12.
Keywords: symplectic 4-manifold, geography, branched covering, semifree action, generalized fiber sum.

(cf. [Akhmedov et al. 2010a; Akhmedov and Park 2010a; Park and Szabó 2000]). The symplectic botany problem, that is, the classification problem for minimal symplectic 4-manifolds with a given pair of invariants (χ_h, c_1^2), seems to be an intractable problem at the moment. However, we now know that most ordered pairs are realized by infinitely many pairwise nondiffeomorphic simply connected minimal symplectic 4-manifolds; see [Gompf and Stipsicz 1999].

In this paper, we will focus our attention on the symplectic geography problem for simply connected minimal symplectic 4-manifolds with *nonnegative* signature. Unlike the negative signature case, the existing literature [Akhmedov and Park 2008; 2010b; Akhmedov et al. 2010b; Li and Stipsicz 2002; Niepel 2005; Park 2002; 2003; Stipsicz 1998; 1999] is far from capturing all possible (χ_h, c_1^2) coordinates, even if we allow nontrivial fundamental groups. The main goal of this paper is to summarize the current state of our knowledge when the simply connected symplectic 4-manifolds are required to be nonspin, or equivalently, are required to have odd intersection form. By Freedman's classification theorem [1982] for simply connected topological 4-manifolds, our problem is then equivalent to finding a minimal symplectic 4-manifold M with signature σ that is homeomorphic to $k\mathbb{CP}^2 \# (k-\sigma)\overline{\mathbb{CP}}^2$, where k is any odd positive integer and σ is any integer satisfying $0 \le \sigma \le k$. Here, $\mathbb{CP}^2$ is the complex projective plane, $\overline{\mathbb{CP}}^2$ is the underlying smooth 4-manifold $\mathbb{CP}^2$ equipped with the opposite orientation, and $k\mathbb{CP}^2 \# (k-\sigma)\overline{\mathbb{CP}}^2$ is the connected sum of k copies of $\mathbb{CP}^2$ and $k-\sigma$ copies of $\overline{\mathbb{CP}}^2$. Note that a simply connected symplectic 4-manifold M has odd $b_2^+(M)$, and hence our integer k must be odd.

A closed 4-manifold with signature σ corresponds to a point (χ_h, c_1^2) on the line $c_1^2 = 8\chi_h + \sigma$. For technical reasons, it will be convenient to fix the signature and deal with each of these lines separately. It is now well-known (see [Akhmedov and Park 2008; Park 2003]) that for each signature $\sigma \ge 0$, there exists a constant $\lambda(\sigma)$ depending only on σ such that any point (χ_h, c_1^2) on the line $c_1^2 = 8\chi_h + \sigma$ satisfying $\chi_h \ge \lambda(\sigma)$ is realized by at least one simply connected nonspin minimal symplectic 4-manifold and infinitely many simply connected nonspin irreducible nonsymplectic 4-manifolds (Definition 13 in Section 6). In other words, $k\mathbb{CP}^2 \# (k-\sigma)\overline{\mathbb{CP}}^2$ is homeomorphic to at least one minimal symplectic 4-manifold and infinitely many pairwise nondiffeomorphic irreducible nonsymplectic 4-manifolds, provided that k is odd and $k \ge 2\lambda(\sigma) - 1$ for some constant $\lambda(\sigma)$ that depends only on the signature σ.

The main result of this paper is the explicit formulation of the smallest values of $\lambda(\sigma)$ that are currently known to the authors. In [Akhmedov and Park 2008], small $\lambda(\sigma)$ values are given when $0 \le \sigma \le 4$, and these values are listed in Table 1. In this paper, we will concentrate on the cases when $\sigma \ge 5$ (see Table 2 in Section 6). For example, when $0 \le \sigma \le 100$, we realize more than 20,000 new (χ_h, c_1^2) points that were not covered by the results in [Akhmedov and Park 2008; Park 2003].

σ	0	1	2	3	4
$\lambda(\sigma) \le$	25	25	24	27	26

Table 1. Results from [Akhmedov and Park 2008].

If a 4-manifold M is simply connected, then $2\chi_h(M) - 1 = b_2^+(M) \ge \sigma(M)$. Thus we obtain an *a priori* lower bound $\chi_h \ge \lceil (\sigma+1)/2 \rceil$, where

$$\lceil x \rceil = \min\{k \in \mathbb{Z} \mid k \ge x\}$$

is the ceiling function. It is tempting to conjecture that our *a posteriori* lower bound for χ_h can eventually be improved down to $\lambda(\sigma) = \lceil (\sigma+1)/2 \rceil$, which will result in the complete solution of the geography problem for simply connected nonspin minimal symplectic 4-manifolds.

Our paper is organized as follows. In Section 2, we present a branched covering construction of Lefschetz fibrations with positive signature, which is a generalization of Stipsicz's constructions [1998; 1999]. In Section 3, we show how to glue together semifree cyclic group actions on closed 2-manifolds, and then we use these actions to construct new examples of Lefschetz fibrations with positive signature. In Section 4, we show how to obtain simply connected 4-manifolds from nonsimply connected Lefschetz fibrations by performing generalized fiber sums with certain 4-manifolds that were constructed in [Akhmedov and Park 2010a]. In Section 5, we implement the strategies from previous sections to construct new families of simply connected irreducible 4-manifolds with positive signature. In Section 6, we compute the lower bounds $\lambda(\sigma)$ for many small values of σ.

2. Branched covering construction

Let Σ_g be a closed 2-dimensional manifold of genus $g > 0$. Let $\zeta : \Sigma_g \to \Sigma_g$ be an orientation-preserving self-diffeomorphism of Σ_g with q fixed points $\{y_1, \dots, y_q\}$. Assume that

$$\zeta^p = \underbrace{\zeta \circ \cdots \circ \zeta}_{p} = \mathrm{id},$$

for some positive integer $p \ge 2$, and that ζ generates a semifree $\mathbb{Z}/p$ action on Σ_g. If $\zeta_* : H_1(\Sigma_g; \mathbb{Z}) \to H_1(\Sigma_g; \mathbb{Z})$ is the induced homomorphism on the first homology group, then we also assume that

$$\zeta_*^{p-1} + \zeta_*^{p-2} + \cdots + \zeta_* + \mathrm{id} = 0 \tag{1}$$

on $H_1(\Sigma_g; \mathbb{Z})$, which is equivalent to 1 not being an eigenvalue of ζ_*. See Examples 3 and 5 below for some concrete examples of ζ.

We will consider $\Sigma_g \times \Sigma_g$ as a symplectic 4-manifold equipped with a product symplectic form $\tilde{\omega} = \mathrm{pr}_1^*\omega + \mathrm{pr}_2^*\omega$, where ω is a symplectic volume form on Σ_g and $\mathrm{pr}_j : \Sigma_g \times \Sigma_g \to \Sigma_g$ $(j = 1, 2)$ is the projection map onto the j-th factor. For each $i = 1, \ldots, p$, let

$$\Gamma_i = \mathrm{graph}(\zeta^i) = \{(x, \zeta^i(x)) \mid x \in \Sigma_g\} \subset \Sigma_g \times \Sigma_g.$$

Note that Γ_p is equal to the diagonal $\{(x, x) \mid x \in \Sigma_g\}$. The graphs $\Gamma_1, \ldots, \Gamma_p$ are symplectic submanifolds of $\Sigma_g \times \Sigma_g$ with respect to $\tilde{\omega}$ (see Lemma 2.1 in [Akhmedov and Park 2008]), and the graphs intersect at q points

$$\{(y_j, y_j) \mid j = 1, \ldots, q\}.$$

If we symplectically blow up $\Sigma_g \times \Sigma_g$ at these q intersection points, then the proper transform B of the union $\Gamma_1 \cup \cdots \cup \Gamma_p$ consists of p disjoint genus g symplectic submanifolds of $(\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}^2}$.

Let $\{\gamma_k \mid k = 1, \ldots, 2g\}$ be a basis for $H_1(\Sigma_g; \mathbb{Z})$ and let $\{\gamma^\ell \mid \ell = 1, \ldots, 2g\}$ be the dual basis under the intersection product so that $\gamma_k \cdot \gamma^\ell = \delta_k^\ell$. If we introduce the notation

$$[\Delta] = [\Sigma_g \times \{\mathrm{pt}\}] + [\{\mathrm{pt}'\} \times \Sigma_g],$$

then the homology class of Γ_i is given by

$$[\Gamma_i] = [\Delta] - \sum_{k=1}^{2g} \gamma^k \times \zeta^i_*(\gamma_k).$$

Using (1), we can express the homology class of B as

$$[B] = p\left([\Delta] - \sum_{j=1}^{q} [E_j]\right),$$

where $E_1, \ldots, E_q$ are the exceptional spheres of the blowups. We also note that

$$c_1((\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}^2}) = \mathrm{PD}\left((2 - 2g)[\Delta] - \sum_{j=1}^{q} [E_j]\right),$$

where PD denotes the Poincaré duality isomorphism.

Since $[B]$ is divisible by p, we may take the cyclic p-fold branched cover of $(\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}^2}$ that is branched along B. We will denote this branched covering by $\beta : X^\zeta_{g,p,q} \to (\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}^2}$. The total space $X^\zeta_{g,p,q}$ inherits a symplectic structure from $(\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}}^2$, and we have

$$\begin{aligned} c_1(X^\zeta_{g,p,q}) &= \beta^*\Big(c_1((\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}}^2) - \frac{p-1}{p}\mathrm{PD}[B]\Big) \\ &= \beta^*\mathrm{PD}\Big((3-2g-p)[\Delta] + (p-2)\sum_{j=1}^{q}[E_j]\Big). \end{aligned}$$

The characteristic numbers of $X^\zeta_{g,p,q}$ can be computed as follows.

$$\begin{aligned} e(X^\zeta_{g,p,q}) &= pe((\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}}^2) - p(p-1)e(\Sigma_g) \\ &= p((2-2g)^2 + q) - p(p-1)(2-2g) \\ &= p(4g^2 + 2gp - 10g - 2p + q + 6), \\ c_1^2(X^\zeta_{g,p,q}) &= p\Big((3-2g-p)[\Delta] + (p-2)\sum_{j=1}^{q}[E_j]\Big)^2 \\ &= p(2(3-2g-p)^2 - q(p-2)^2) \\ &= p(-p^2q + 8g^2 + 2p^2 + 8gp + 4pq - 24g - 12p - 4q + 18), \\ \sigma(X^\zeta_{g,p,q}) &= \tfrac{1}{3}\big(c_1^2(X^\zeta_{g,p,q}) - 2e(X^\zeta_{g,p,q})\big) \\ &= \tfrac{1}{3}p(-p^2q + 2p^2 + 4gp + 4pq - 4g - 8p - 6q + 6), \\ \chi_h(X^\zeta_{g,p,q}) &= \tfrac{1}{4}\big(e(X^\zeta_{g,p,q}) + \sigma(X^\zeta_{g,p,q})\big) \\ &= \tfrac{1}{12}p(-p^2q + 12g^2 + 2p^2 + 10gp + 4pq - 34g - 14p - 3q + 24). \end{aligned}$$

Let $\epsilon : (\Sigma_g \times \Sigma_g)\# q\overline{\mathbb{CP}}^2 \to \Sigma_g \times \Sigma_g$ be the blowdown map. Then the composition of maps

$$X^\zeta_{g,p,q} \xrightarrow{\beta} (\Sigma_g \times \Sigma_g) \# q\overline{\mathbb{CP}}^2 \xrightarrow{\epsilon} \Sigma_g \times \Sigma_g \xrightarrow{\mathrm{pr}_1} \Sigma_g \tag{2}$$

gives a fibration of $X^\zeta_{g,p,q}$ over Σ_g. A regular fiber of this fibration is a cyclic p-fold branched cover of Σ_g that is branched over p points. Thus a regular fiber is a closed surface of genus equal to

$$\tfrac{1}{2}(p^2 + 2gp - 3p + 2). \tag{3}$$

The proper transform of each graph Γ_i $(i = 1, \ldots, p)$ gives rise to a section of (2) whose image is a genus g surface S_i in $X^\zeta_{g,p,q}$ with self-intersection equal to

$$\begin{aligned} [S_i]^2 &= \big\langle c_1(X^\zeta_{g,p,q}), [S_i]\big\rangle - e(\Sigma_g) \\ &= 2g - 2 + \frac{1}{p}\Big((3-2g-p)[\Delta] + (p-2)\sum_{j=1}^{q}[E_j]\Big)\cdot[B] \\ &= pq - 2g - 2p - 2q + 4. \end{aligned}$$

Lemma 1. *Let $f : X^{\zeta}_{g,p,q} \to \Sigma_g$ denote the composition of maps in* (2). *Then f is a relatively minimal Lefschetz fibration with pq critical points. Moreover, each critical point of f corresponds to a nonseparating vanishing cycle.*

Proof. Clearly the only singular fibers of f are $\{f^{-1}(y_j) \mid j = 1, \dots, q\}$. We will prove that each $f^{-1}(y_j)$ contains exactly p Lefschetz critical points. To describe each $f^{-1}(y_j)$ explicitly, we will view $X^{\zeta}_{g,p,q}$ as the minimal desingularization of another branched cover that we will define below.

Let $\Gamma = \Gamma_1 \cup \cdots \cup \Gamma_p$. Since $[\Gamma] = p[\Delta] \in H_2(\Sigma_g \times \Sigma_g; \mathbb{Z})$ is divisible by p, we may take the cyclic p-fold branched cover of $\Sigma_g \times \Sigma_g$ that is branched along Γ. We will denote this branched covering by $\widehat{\beta} : \widehat{X}^{\zeta}_{g,p,q} \to \Sigma_g \times \Sigma_g$. The total space $\widehat{X}^{\zeta}_{g,p,q}$ has q singular points, $\{\widehat{\beta}^{-1}(y_j, y_j) \mid j = 1, \dots, q\}$, each of which can be locally modeled by

$$\{(x, y, z) \in \mathbb{C}^3 \mid z^p = x^p + y^p\}. \tag{4}$$

In these local coordinates, the singular point $\widehat{\beta}^{-1}(y_j, y_j)$ corresponds to $(0, 0, 0)$, and a neighborhood of the singular point corresponds to the cyclic p-fold cover of the (x, y)-plane that is branched over p complex lines that intersect transversely at $(0, 0)$.

Next let $\widehat{f} : \widehat{X}^{\zeta}_{g,p,q} \to \Sigma_g$ denote the singular fibration given by the composition

$$\widehat{X}^{\zeta}_{g,p,q} \xrightarrow{\widehat{\beta}} \Sigma_g \times \Sigma_g \xrightarrow{\mathrm{pr}_1} \Sigma_g.$$

A regular fiber of $\widehat{f}$ is again a closed surface of genus equal to (3). There are exactly q singular fibers $\{\widehat{f}^{-1}(y_j) \mid j = 1, \dots, q\}$. For each $j = 1, \dots, q$, note that $\widehat{f}^{-1}(y_j) \setminus \{\widehat{\beta}^{-1}(y_j, y_j)\}$ is a smooth and connected surface since it is the unbranched cyclic p-fold cover of the once punctured surface $(\{y_j\} \times \Sigma_g) \setminus \{(y_j, y_j)\}$ coming from a surjective homomorphism

$$\pi_1((\{y_j\} \times \Sigma_g) \setminus \{(y_j, y_j)\}) \cong F_{2g} \longrightarrow \mathbb{Z}/p \subset S_p, \tag{5}$$

where F_{2g} is the free group with $2g$ generators and S_p is the symmetric group on p symbols. Since $\mathbb{Z}/p$ is abelian, (5) can be factored as the composition

$$\pi_1((\{y_j\} \times \Sigma_g) \setminus \{(y_j, y_j)\}) \longrightarrow \pi_1(\Sigma_g) \longrightarrow \mathbb{Z}/p.$$

Thus the cover $\widehat{f}^{-1}(y_j) \setminus \{\widehat{\beta}^{-1}(y_j, y_j)\} \to (\{y_j\} \times \Sigma_g) \setminus \{(y_j, y_j)\}$ can be viewed as a restriction of the unbranched cyclic p-fold cover of the closed surface Σ_g. In other words, $\widehat{f}^{-1}(y_j) \setminus \{\widehat{\beta}^{-1}(y_j, y_j)\}$ can be embedded into the unbranched cyclic p-fold cover of Σ_g. This implies that $\widehat{f}^{-1}(y_j) \setminus \{\widehat{\beta}^{-1}(y_j, y_j)\}$ is diffeomorphic to a surface of genus $gp - p + 1$ having p punctures, and $\widehat{f}^{-1}(y_j)$ is a connected surface that is smooth away from the point $\widehat{\beta}^{-1}(y_j, y_j)$, which is a multiple point of order p.

Now recall from [Gompf and Stipsicz 1999; Némethi 1999] that $X^{\zeta}_{g,p,q}$ is the minimal desingularization of $\widehat{X}^{\zeta}_{g,p,q}$. The standard algorithm for resolution of singularities (see [Némethi 1999, Example 1.20(h)]) replaces each singular point $\widehat{\beta}^{-1}(y_j, y_j)$ of $\widehat{X}^{\zeta}_{g,p,q}$ having local model (4) with a closed surface of genus $\frac{1}{2}(p^2-3p+2)$ and self-intersection $-p$. This surface is just $\beta^{-1}(E_j)$, which is a cyclic p-fold branched cover of the exceptional sphere E_j branched over p points. It follows that each singular fiber $f^{-1}(y_j)$ is the union of two closed surfaces that intersect each other transversely at p distinct points. One of the surfaces is $\beta^{-1}(E_j)$, and the other is a genus $gp-p+1$ surface of self-intersection $-p$, which is the smooth completion of $\widehat{f}^{-1}(y_j)\setminus\{\widehat{\beta}^{-1}(y_j, y_j)\}$. The p transverse intersection points between the two surfaces are exactly the p Lefschetz critical points of f that get mapped to y_j. Finally, comparing the sum of genera with (3), we observe that each union of the two surfaces is obtained by replacing the annular neighborhoods of p nonseparating circles in a regular fiber with p pairs of transversely intersecting disks. This implies that all the vanishing cycles are nonseparating. □

Remark 2. We can verify the number of critical points of f by computing the difference

$$e(X^{\zeta}_{g,p,q}) - e(\text{regular fiber})\cdot e(\text{base}) = pq.$$

We can split the singular fibers of f so that each new singular fiber contains only one critical point (cf. [Harris and Morrison 1998; Takamura 2004]) but we do not need to do so for our applications below.

Given a positive integer u, let $\eta_u : \Sigma_k \to \Sigma_g$ be a u-fold unbranched covering of Σ_g, where $k = u(g-1)+1$. We pull back the branched covering

$$X^{\zeta}_{g,p,q} \xrightarrow{\beta} (\Sigma_g \times \Sigma_g)\,\#\, q\overline{\mathbb{CP}}^2 \xrightarrow{\epsilon} \Sigma_g \times \Sigma_g$$

by the product map $\eta_{u_1}\times\eta_{u_2} : \Sigma_{k_1}\times\Sigma_{k_2} \to \Sigma_g\times\Sigma_g$, where u_i is a positive integer and $k_i = u_i(g-1)+1$ for each $i = 1, 2$. The total space of this pullback is a new symplectic 4-manifold $X^{\zeta}_{g,p,q}(u_1, u_2)$, which is a p-fold branched cover of $\Sigma_{k_1}\times\Sigma_{k_2}$ and a u_1u_2-fold unbranched cover of $X^{\zeta}_{g,p,q}$. The composition

$$f_{u_1,u_2} : X^{\zeta}_{g,p,q}(u_1, u_2) \longrightarrow \Sigma_{k_1}\times\Sigma_{k_2} \xrightarrow{\mathrm{pr}_1} \Sigma_{k_1}$$

gives a new relatively minimal Lefschetz fibration, where $X^{\zeta}_{g,p,q}(1,1) = X^{\zeta}_{g,p,q}$ and $f_{1,1} = f$. A regular fiber of f_{u_1,u_2} is a u_2-fold unbranched cover of the fiber of f (or equivalently a p-fold branched cover of Σ_{k_2} branched along u_2p points) and hence has genus equal to

$$1+\frac{u_2}{2}(p^2+2gp-3p).$$

A section of f gives rise to a section of f_{u_1,u_2} whose image is a genus k_1 surface of self-intersection equal to

$$u_1(pq-2g-2p-2q+4).$$

Since $X^{\zeta}_{g,p,q}(u_1,u_2)$ is a u_1u_2-fold unbranched cover of $X^{\zeta}_{g,p,q}$, we have

$$e(X^{\zeta}_{g,p,q}(u_1,u_2)) = u_1u_2\cdot e(X^{\zeta}_{g,p,q}), \qquad \sigma(X^{\zeta}_{g,p,q}(u_1,u_2)) = u_1u_2\cdot \sigma(X^{\zeta}_{g,p,q}),$$
$$\chi_h(X^{\zeta}_{g,p,q}(u_1,u_2)) = u_1u_2\cdot \chi_h(X^{\zeta}_{g,p,q}), \quad c_1^2(X^{\zeta}_{g,p,q}(u_1,u_2)) = u_1u_2\cdot c_1^2(X^{\zeta}_{g,p,q}).$$

Example 3. Recall from Section 2 of [Akhmedov and Park 2008] that there exists a semifree $\mathbb{Z}/(g+1)$ action on Σ_g with 4 fixed points satisfying (1). Applying the above machinery, we obtain a family of symplectic 4-manifolds $X^g_{u_1,u_2} = X^{\zeta}_{g,g+1,4}(u_1,u_2)$, where g, u_1 and u_2 are positive integers, satisfying

$$\begin{aligned}
e(X^g_{u_1,u_2}) &= 2u_1u_2(g+1)(3g^2-5g+4),\\
\sigma(X^g_{u_1,u_2}) &= \tfrac{2}{3}u_1u_2(g+1)(g^2+2g-6),\\
\chi_h(X^g_{u_1,u_2}) &= \tfrac{1}{6}u_1u_2(g+1)(10g^2-13g+6),\\
c_1^2(X^g_{u_1,u_2}) &= 2u_1u_2(g+1)(7g^2-8g+2).
\end{aligned}$$

For each triple of positive integers g, u_1, u_2, there exists a relatively minimal Lefschetz fibration $f_{u_1,u_2} : X^g_{u_1,u_2} \to \Sigma_{k_1}$ such that the genus of a regular fiber is equal to $1+\frac{1}{2}u_2(g+1)(3g-2)$ and there is a section whose image is a surface of genus $k_1 = u_1(g-1)+1$ and self-intersection $-2u_1$.

Remark 4. The 4-manifolds X_g, $X_g(n)$ and $\tilde{X}_g(n^2)$ in [Akhmedov and Park 2008] are equal to $X^g_{1,1}$, $X^g_{n,1}$ and $X^g_{n,n}$, respectively.

3. Gluing self-diffeomorphisms of surfaces

In light of the machinery in Section 2, it will be desirable to find lots of semifree $\mathbb{Z}/p$ actions on closed surfaces. One way to produce such actions is to glue together semifree $\mathbb{Z}/p$ actions on surfaces of low genera as we explain below.

Let $v \geq 2$ be an integer. For each $i = 1, \ldots, v$, let $\alpha_i : \Sigma_{g_i} \to \Sigma_{g_i}$ be an orientation-preserving self-diffeomorphism of a closed surface of genus g_i with q_i fixed points $\{y_{i,1}, \ldots, y_{i,q_i}\}$. Assume that each α_i generates a semifree $\mathbb{Z}/p$ action on Σ_{g_i}. For each $j = 1, \ldots, q_i$, let $\rho_{i,j}$ be the rotational number of α_i at the fixed point $y_{i,j}$ so that α_i induces rotation by angle $2\pi\rho_{i,j}/p$ in the tangent space at $y_{i,j}$. The rotational numbers are well-defined mod p and are relatively prime to p. They satisfy (see [Nielsen 1937])

$$\sum_{j=1}^{q_i} \frac{1}{\rho_{i,j}} \equiv 0 \pmod{p},$$

where $1/\rho_{i,j}$ denotes the multiplicative inverse of $\rho_{i,j}$ in $(\mathbb{Z}/p)^\times$. We can reverse the signs of $\rho_{i,1}, \dots, \rho_{i,q_i}$ simultaneously by reversing the orientation of Σ_{g_i}.

Now choose a single fixed point of α_i for $i = 1, v$, and choose two fixed points of α_i for $i = 2, \dots, v-1$. Without loss of generality, we may choose $y_{1,2}$, $y_{v,1}$ and $y_{i,1}$, $y_{i,2}$ for $i = 2, \dots, v-1$. We remove small $\mathbb{Z}/p$-equivariant neighborhoods of these chosen fixed points and then glue the boundary circle at $y_{i,2}$ to the boundary circle at $y_{i+1,1}$ for $i = 1, \dots, v-1$. Such gluing of one-holed and two-holed surfaces results in a closed surface of genus $g = \sum_{i=1}^{v} g_i$. If $\rho_{i,2} = -\rho_{i+1,1}$ for all $i = 1, \dots, v-1$, that is, the rotational numbers are negatives of each other at the gluing points, then the restrictions of α_i's to the punctured surfaces can also be glued together to form an orientation-preserving self-diffeomorphism $\zeta : \Sigma_g \to \Sigma_g$ with q fixed points, where

$$q = -2(v-1) + \sum_{i=1}^{v} q_i.$$

We will say that ζ is an *equivariant sum* of $\alpha_1, \dots, \alpha_v$, and write $\zeta = \alpha_1 \# \cdots \# \alpha_v$. In case when $\alpha_1 = \cdots = \alpha_v$, we will write $\zeta = v\alpha_1$ for short.

Example 5. For each odd integer $p \geq 3$, there exists a semifree $\mathbb{Z}/p$ action on $\Sigma_{(p-1)/2}$ as follows. Consider $\Sigma_{(p-1)/2}$ as the quotient of a regular $2p$-gon by identifying the opposite sides. The rotation of the $2p$-gon by angle $2\pi/p$ gives an orientation-preserving self-diffeomorphism $\tau_p : \Sigma_{(p-1)/2} \to \Sigma_{(p-1)/2}$ with 3 fixed points. The fixed points of τ_p are the center of the $2p$-gon and the 2 points coming from the vertices. The center of the $2p$-gon has rotational number 1, and the other 2 fixed points both have rotational number -2.

We can find a basis of $H_1(\Sigma_{(p-1)/2}; \mathbb{Z})$ such that the induced homomorphism $(\tau_p)_* : H_1(\Sigma_{(p-1)/2}; \mathbb{Z}) \to H_1(\Sigma_{(p-1)/2}; \mathbb{Z})$ is represented by the $(p-1) \times (p-1)$ matrix

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & -1 \\ \hline & & & -1 \\ & I_{p-2} & & \vdots \\ & & & -1 \end{array}\right], \tag{6}$$

where I_{p-2} is the identity $(p-2) \times (p-2)$ matrix. It is easy to check that this matrix satisfies (1).

For each positive integer v, let $\zeta = v\tau_p$ be the equivariant sum of v copies of τ_p. (We glue along fixed points with rotational number -2, and we alternate the orientations of the punctured $\Sigma_{(p-1)/2}$'s so that the rotational numbers are $+2$ and -2 at each gluing.) Then $\zeta : \Sigma_{v(p-1)/2} \to \Sigma_{v(p-1)/2}$ generates a semifree $\mathbb{Z}/p$ action on $\Sigma_{v(p-1)/2}$ with $v+2$ fixed points. The induced homomorphism

$\zeta_* : H_1(\Sigma_{v(p-1)/2};\mathbb{Z}) \to H_1(\Sigma_{v(p-1)/2};\mathbb{Z})$ satisfies (1) since it can be represented by a block diagonal matrix each of whose blocks is conjugate to (6).

From the branched covering construction in Section 2, we obtain a family of symplectic 4-manifolds $W^{p,v}_{u_1,u_2} = X^{v\tau_p}_{v(p-1)/2,p,v+2}(u_1,u_2)$, where $p \geq 3$ is an odd integer and v, u_1, u_2 are positive integers, satisfying

$$\begin{aligned}
e(W^{p,v}_{u_1,u_2}) &= pu_1u_2[(v^2+v)p^2 - 2(v^2+3v+1)p + v^2+6v+8],\\
\sigma(W^{p,v}_{u_1,u_2}) &= \tfrac{1}{3}pu_1u_2(vp^2-4v-6),\\
\chi_h(W^{p,v}_{u_1,u_2}) &= \tfrac{1}{12}pu_1u_2[(3v^2+4v)p^2 - 6(v^2+3v+1)p + 3v^2+14v+18],\\
c_1^2(W^{p,v}_{u_1,u_2}) &= pu_1u_2[(2v^2+3v)p^2 - 4(v^2+3v+1)p + 2v^2+8v+10].
\end{aligned}$$

Moreover, for each quadruple of positive integers p, v, u_1, u_2 with odd $p \geq 3$, we have a relatively minimal Lefschetz fibration $f_{u_1,u_2} : W^{p,v}_{u_1,u_2} \to \Sigma_{k_1}$ such that the genus of a regular fiber is equal to $1+\frac{1}{2}pu_2[(v+1)p-v-3]$ and there is a section whose image is a surface of genus $k_1 = 1+u_1[-1+v(p-1)/2]$ and self-intersection $-u_1v$.

Note that $c_1^2(W^{p,v}_{u_1,u_2}) \leq 9\chi_h(W^{p,v}_{u_1,u_2})$, with equality if and only if $p=5$ and $v=1$. If we view the quotient $c_1^2(W^{p,v}_{u_1,u_2})/\chi_h(W^{p,v}_{u_1,u_2})$ as a function of p and v, then its gradient vector is

$$\begin{bmatrix}
-\dfrac{24\big((v^3+3v^2+v)p^2-(5v^3+16v^2+14v)p+4v^3+18v^2+22v+6\big)}{\big((3v^2+4v)p^2-6(v^2+3v+1)p+3v^2+14v+18\big)^2}\\[2ex]
-\dfrac{12\big((p^2-4)(p-1)^2v^2-12(p-1)^2v+2p^3-14p^2+28p-4\big)}{\big((3v^2+4v)p^2-6(v^2+3v+1)p+3v^2+14v+18\big)^2}
\end{bmatrix}$$

When $p \geq 7$ and $v \geq 1$, both components of this gradient vector are negative and hence $c_1^2(W^{p,v}_{u_1,u_2})/\chi_h(W^{p,v}_{u_1,u_2})$ is decreasing as p and v increase. We observe that $\lim_{v\to\infty} c_1^2(W^{p,v}_{u_1,u_2})/\chi_h(W^{p,v}_{u_1,u_2}) = 8$, and

$$\lim_{p\to\infty} \frac{c_1^2(W^{p,v}_{u_1,u_2})}{\chi_h(W^{p,v}_{u_1,u_2})} = \frac{12(2v+3)}{3v+4} \leq \frac{60}{7},$$

where the rational function $12(2v+3)/(3v+4)$ is decreasing for $v \geq 1$. Therefore most $W^{p,v}_{u_1,u_2}$'s lie well below the Bogomolov–Miyaoka–Yau (BMY) line, $c_1^2 = 9\chi_h$.

Remark 6. According to Section 4.5 of [Luo 2000], there is a unique $\mathbb{Z}/3$ action on Σ_g with $g+2$ fixed points. It follows that $W^{3,2}_{u_1,u_2}$ is exactly equal to $X^2_{u_1,u_2}$ in Example 3. More generally, for each odd integer $p \geq 5$, we conjecture that $W^{p,2}_{u_1,u_2}$ is diffeomorphic to $X^{p-1}_{u_1,u_2}$ in Example 3. We also conjecture that the 4-manifolds Z_g, $Z_g(n)$ and $\tilde{Z}_g(n^2)$ in Section 3 of [Akhmedov and Park 2008] are diffeomorphic to $W^{2g+1,1}_{1,1}$, $W^{2g+1,1}_{n,1}$ and $W^{2g+1,1}_{n,n}$, respectively. In particular, we conjecture that

$W^{5,1}_{1,1}$, $W^{5,1}_{n,1}$ and $W^{5,1}_{n,n}$, lying on the BMY line $c_1^2 = 9\chi_h$, are diffeomorphic to complex surfaces $H = H(1)$, $H(n)$ and $H(n^2)$ in [Chen 1991; Stipsicz 1998; 1999], respectively.

4. Generalized fiber sums

Let Σ_b denote a closed Riemann surface of genus $b > 0$. Suppose $f : X \to \Sigma_b$ is a Lefschetz fibration with generic fiber F diffeomorphic to a closed Riemann surface Σ_a with genus $a > 0$. Assume that f is a relatively minimal Lefschetz fibration (i.e., no fiber contains a sphere of self-intersection -1) so that X is a minimal symplectic 4-manifold (Theorem 1.4 of [Stipsicz 2000]). Also assume that f has a section whose image S in X has self-intersection d. From Theorem 10.2.18 in [Gompf and Stipsicz 1999], X can be equipped with a symplectic structure such that both F and S are symplectic submanifolds. From Proposition 8.1.9 in [Gompf and Stipsicz 1999], we have an exact sequence

$$\pi_1(F) \longrightarrow \pi_1(X) \xrightarrow{f_*} \pi_1(\Sigma_b) \longrightarrow 1. \tag{7}$$

Let $t > 0$ be an integer. By symplectically resolving the intersection points, we can find a symplectic genus $ta + b$ surface $\Sigma \subset X$ representing the homology class $t[F] + [S] \in H_2(X; \mathbb{Z})$ with self-intersection $2t + d$. By taking t large enough, we can assume that $2t + d \geq 0$. Let $\widetilde{X} = X \# (2t + d)\overline{\mathbb{CP}}^2$, where each of the $2t + d$ symplectic blowups take place at points on $\Sigma \subset X$. The proper transform $\widetilde{\Sigma} \subset \widetilde{X}$ is a symplectic submanifold with genus $ta + b$ and self-intersection 0. Note that we have

$$e(\widetilde{X}) = e(X) + 2t + d,$$
$$\sigma(\widetilde{X}) = \sigma(X) - 2t - d.$$

Lemma 7. *Let $\tilde{i} : \widetilde{\Sigma}^{\parallel} \hookrightarrow \widetilde{X} \setminus \nu\widetilde{\Sigma}$ be the inclusion map of a parallel copy of $\widetilde{\Sigma}$ into the complement of a tubular neighborhood $\nu\widetilde{\Sigma}$ in $\widetilde{X} = X \# (2t + d)\overline{\mathbb{CP}}^2$. Then we have*

$$\frac{\pi_1(\widetilde{X} \setminus \nu\widetilde{\Sigma})}{\langle \tilde{i}_*(\pi_1(\widetilde{\Sigma}^{\parallel}))\rangle} = 1, \tag{8}$$

where $\langle \tilde{i}_*(\pi_1(\widetilde{\Sigma}^{\parallel}))\rangle$ is the normal subgroup of $\pi_1(\widetilde{X} \setminus \nu\widetilde{\Sigma})$ generated by the image $\tilde{i}_*(\pi_1(\widetilde{\Sigma}^{\parallel}))$.

Proof. Let $i : \Sigma^{\parallel} \hookrightarrow X \setminus \nu\Sigma$ be the inclusion map of a parallel copy of Σ. From exact sequence (7), we deduce that $\pi_1(X)/\langle i_*(\pi_1(\Sigma^{\parallel}))\rangle = 1$. Since the blowups do not effect the fundamental groups, we conclude that $\pi_1(\widetilde{X})/\langle \tilde{i}_*(\pi_1(\widetilde{\Sigma}^{\parallel}))\rangle = 1$. If

$2t+d>0$, then any meridian $\mu(\widetilde{\Sigma})$ of $\widetilde{\Sigma}$ in $\pi_1(\widetilde{X}\setminus\nu\widetilde{\Sigma})$ bounds a disk that comes from a punctured exceptional sphere. Hence $\pi_1(\widetilde{X}\setminus\nu\widetilde{\Sigma})=\pi_1(\widetilde{X})$ and (8) follows from our last conclusion.

If $2t+d=0$, then $\widetilde{X}=X$, $\widetilde{\Sigma}=\Sigma$, $\widetilde{\Sigma}^{\parallel}=\Sigma^{\parallel}$, and $\tilde{i}=i$. Any meridian $\mu(\Sigma)$ in $\pi_1(X\setminus\nu\Sigma)$ is conjugate to a meridian of S. Since $[F]\cdot[S]=1$, $\mu(\Sigma)$ is in the normal subgroup generated by the generators of $\pi_1(F)$, which in turn lies in $\langle i_*(\pi_1(\Sigma^{\parallel}))\rangle$. This implies that $\pi_1(X\setminus\nu\Sigma)/\langle i_*(\pi_1(\Sigma^{\parallel}))\rangle=\pi_1(X)/\langle i_*(\pi_1(\Sigma^{\parallel}))\rangle=1$. □

For each pair of integers $m\geq 1$ and $n\geq 2$, let $Y_n(m)$ denote the irreducible 4-manifold constructed in Section 2 of [Akhmedov and Park 2010a] that has the same cohomology ring as the connected sum $(2n-3)(S^2\times S^2)$. Recall that $Y_n(m)$ is obtained by performing $2n+4$ surgeries along Lagrangian tori in the product 4-manifold $\Sigma_2\times\Sigma_n$. Thus $Y_n(m)$ contains a pair of submanifolds $\Sigma_2=\Sigma_2\times\{\mathrm{pt}\}$ and $\Sigma_n=\{\mathrm{pt}'\}\times\Sigma_n$, both of self-intersection 0. When $m=1$, $Y_n(1)$ is a minimal symplectic 4-manifold. Moreover, Σ_2 and Σ_n are symplectic submanifolds of $Y_n(1)$. When $n\geq 3$, there exist $2n-4$ pairs of geometrically dual Lagrangian tori which, together with Σ_2 and Σ_n, form a basis for $H_2(Y_n(1);\mathbb{Z})\cong\mathbb{Z}^{4n-6}$.

Theorem 8. *Let $f:X\to\Sigma_b$ be a relatively minimal Lefschetz fibration as above having at least one nonseparating vanishing cycle. Suppose that $n=ta+b\geq 2$. For a suitable choice of the gluing diffeomorphism $\varphi:\partial(\nu\widetilde{\Sigma})\to\partial(\nu\Sigma_n)$, the generalized fiber sum*

$$P_n^m(X)=\widetilde{X}\#_{\varphi}Y_n(m)=(\widetilde{X}\setminus\nu\widetilde{\Sigma})\cup_{\varphi}(Y_n(m)\setminus\nu\Sigma_n) \tag{9}$$

along $\widetilde{\Sigma}$ and Σ_n is simply connected, and satisfies

$$\begin{aligned}
e(P_n^m(X))&=e(X)+d+(8a+2)t+8b-8,\\
\sigma(P_n^m(X))&=\sigma(X)-2t-d,\\
\chi_h(P_n^m(X))&=\chi_h(X)+2at+2b-2,\\
c_1^2(P_n^m(X))&=c_1^2(X)-d+(16a-2)t+16b-16,\\
b_2^+(P_n^m(X))&=b_2^+(X)-b_1(X)+4at+4b-4\geq 3,\\
b_2^-(P_n^m(X))&=b_2^-(X)-b_1(X)+d+(4a+2)t+4b-4.
\end{aligned}$$

If $\sigma(P_n^m(X))$ is not divisible by 16 or if $2t+d>0$, then $P_n^m(X)$ is nonspin and the set $\{P_n^m(X)\mid m\geq 1\}$ contains infinitely many homeomorphic but pairwise nondiffeomorphic irreducible 4-manifolds. When $m=1$, $P_n^1(X)$ is symplectic and irreducible. If $n=ta+b\geq 3$, then $P_n^1(X)$ contains disjoint symplectic tori T_1 and T_2 of self-intersection 0 satisfying $\pi_1(P_n^1(X)\setminus(T_1\cup T_2))=1$.

Proof. Recall from [Akhmedov and Park 2010a] that $e(Y_n(m)) = 4n - 4$ and $\sigma(Y_n(m)) = 0$ since torus surgeries change neither e nor σ. Hence we have

$$\begin{aligned} e(P_n^m(X)) &= e(\widetilde{X}) + e(Y_n(m)) - 2e(\Sigma_n) \\ &= e(X) + 2t + d + 4n - 4 - 2(2 - 2n) \\ &= e(X) + 2t + d + 8n - 8 \\ &= e(X) + 2t + d + 8ta + 8b - 8, \end{aligned}$$

$$\sigma(P_n^m(X)) = \sigma(\widetilde{X}) + \sigma(Y_n(m)) = \sigma(X) - 2t - d.$$

The other characteristic numbers can be computed from the formulas $\chi_h = \frac{1}{4}(e+\sigma)$, $c_1^2 = 2e + 3\sigma$, $b_2^+ = b_1 - 1 + \frac{1}{2}(e+\sigma)$, and $b_2^- = b_1 - 1 + \frac{1}{2}(e-\sigma)$.

To compute $\pi_1(P_n^m(X))$, we first choose a standard presentation

$$\pi_1(\Sigma_n) = \Big\langle c_1, d_1, \dots, c_n, d_n \;\Big|\; \prod_{j=1}^{n} [c_j, d_j] = 1 \Big\rangle.$$

From the presentation of $\pi_1(Y_n(m))$ in [Akhmedov and Park 2010a], we know that $\pi_1(Y_n(m))/\langle z \rangle = 1$, where $\langle z \rangle$ is the normal subgroup generated by the image z of any one of the four generators c_1, d_1, c_2, d_2 of $\pi_1(\Sigma_n)$ under the inclusion induced homomorphism $\pi_1(\Sigma_n) \to \pi_1(Y_n(m))$. We also know that any meridian of Σ_n is conjugate to the image of $[a_1, b_1][a_2, b_2]$ in $\pi_1(Y_n(m) \setminus \nu\Sigma_n)$, where a_i, b_i $(i = 1, 2)$ are the images of standard generators of $\pi_1(\Sigma_2 \times \{\mathrm{pt}\})$. All relations of $\pi_1(Y_n(m))$ listed in [Akhmedov and Park 2010a], except $[a_1, b_1][a_2, b_2] = 1$, continue to hold in $\pi_1(Y_n(m) \setminus \nu\Sigma_n)$ since these relations come from torus surgeries that occur away from $\nu\Sigma_n$. Since $z = 1$ still implies $a_i = b_i = 1$ $(i = 1, 2)$ in $\pi_1(Y_n(m) \setminus \nu\Sigma_n)$, we deduce that $\pi_1(Y_n(m) \setminus \nu\Sigma_n)/\langle z \rangle = 1$.

When forming the generalized fiber sum $P_n^m(X)$, we choose the gluing diffeomorphism φ such that the induced homomorphism φ_* maps the element of $\pi_1(\widetilde{\Sigma}^{\parallel})$ represented by a nonseparating vanishing cycle of the Lefschetz fibration X to z, viewed as an element of $\pi_1(\Sigma_n^{\parallel})$. Thus $z = 1$ in $\pi_1(P_n^m(X))$, which then implies that the inclusion induced homomorphism

$$\pi_1(Y_n(m) \setminus \nu\Sigma_n) \longrightarrow \pi_1(P_n^m(X)) \tag{10}$$

is trivial. Note that the inclusion induced homomorphism $\pi_1(\widetilde{\Sigma}^{\parallel}) \to \pi_1(P_n^m(X))$ is also trivial since it can be factored through homomorphism (10) after $\widetilde{\Sigma}^{\parallel}$ is identified with $\Sigma_n^{\parallel}$ via φ. It follows from Lemma 7 that the inclusion induced homomorphism $\pi_1(\widetilde{X} \setminus \nu\widetilde{\Sigma}) \to \pi_1(P_n^m(X))$ is trivial as well. By the Seifert–van Kampen theorem, we conclude that $\pi_1(P_n^m(X)) = 1$.

If $2t + d > 0$, then $P_n^m(X)$ contains a genus 2 surface of self-intersection -1 that is the internal sum of a punctured exceptional sphere in $\widetilde{X} \setminus \nu\widetilde{\Sigma}$ and a punctured

Σ_2 in $Y_n(m) \setminus \nu\Sigma_n$. In this case, the intersection form of $P_n^m(X)$ is odd and $P_n^m(X)$ is nonspin. Also recall that the signature of a spin 4-manifold is divisible by 16 according to Rohlin's theorem [1952].

Note that $e(P_n^m(X))$ and $\sigma(P_n^m(X))$ are independent of m. If $\sigma(P_n^m(X))$ is not divisible by 16 or if $2t+d>0$, then for fixed n, the set $\{P_n^m(X) \mid m \geq 1\}$ consists of homeomorphic simply connected nonspin 4-manifolds by Freedman's classification theorem (cf. [Freedman 1982]).

Since $Y_n(1)$ is symplectic, the corresponding fiber sum $P_n^1(X)$ is symplectic as well (cf. [Gompf 1995; McCarthy and Wolfson 1994]). Since $(\widetilde{X}, \widetilde{\Sigma})$ is a relatively minimal pair (i.e., every sphere of self-intersection -1 intersects $\widetilde{\Sigma}$) by Corollary 3 in [Li 1999], $P_n^1(X)$ is minimal by Usher's theorem [2006]. Recall from [Hamilton and Kotschick 2006; Kotschick 1997] that a simply connected minimal symplectic 4-manifold is irreducible, and thus $P_n^1(X)$ is irreducible.

Any Lefschetz fibration X with fiber genus a and base genus b satisfies $b_1(X) \leq 2a+2b$. Since X has at least one nonseparating vanishing cycle, we have $b_1(X) < 2a+2b \leq 2at+2b$. Thus we deduce that $b_2^+(P_n^m(X)) > b_2^+(X) \geq 1$. Since $P_n^1(X)$ is symplectic and simply connected, $b_2^+(P_n^1(X)) = b_2^+(P_n^m(X))$ is odd. It follows that $b_2^+(P_n^m(X)) \geq 3$ and the Seiberg–Witten invariant of $P_n^m(X)$ is well defined.

Let Y_0 denote the symplectic 4-manifold that is obtained by performing the same torus surgeries on $\Sigma_2 \times \Sigma_n$ as for $Y_n(m)$, except $(a_1'' \times d_2', d_2', +m)$ surgery (cf. [Akhmedov and Park 2010a]). Let $P_0 = \widetilde{X} \#_\varphi Y_0$ be the generalized fiber sum of $\widetilde{X}$ and Y_0 along $\widetilde{\Sigma}$ and Σ_n using the same gluing diffeomorphism φ that was used in the construction of $P_n^m(X)$. Note that P_0 is symplectic and minimal for the same reasons as $P_n^1(X)$. We have $b_2(P_0) = b_2(P_n^m(X)) + 2$, and there is an orthogonal decomposition $H^2(P_0; \mathbb{Z}) = H \oplus H^\perp$, where H is the 2-dimensional hyperbolic summand generated by the Poincaré duals of $[a_1 \times d_2]$ and $[b_1 \times c_2]$. Using the adjunction inequality, we can easily see that every Seiberg–Witten basic class of P_0 lies in $H^\perp$.

Since $P_n^m(X)$ can be obtained from $P_n^1(X)$ by performing a $1/(m-1)$ surgery on a null-homologous torus, we can apply the product formula in [Morgan et al. 1997] as in [Akhmedov et al. 2008; Fintushel et al. 2007; Szabó 1998] and deduce that there exist surjective homomorphisms

$$\xi_m : H^\perp \longrightarrow H^2(P_n^m(X); \mathbb{Z})$$

that preserve the cup product pairing and satisfy

$$\mathrm{SW}_{P_n^m(X)}(\xi_m(L_0)) = \mathrm{SW}_{P_n^1(X)}(\xi_1(L_0)) + (m-1)\,\mathrm{SW}_{P_0}(L_0), \tag{11}$$

for every characteristic element $L_0 \in H^\perp \subset H^2(P_0; \mathbb{Z})$. We note that the right side of (11) contains only one SW_{P_0} term for the reasons given in the proof of Corollary 2 in [Fintushel et al. 2007]. By a theorem of Taubes [1994], we have

$\mathrm{SW}_{P_0}(c_1(P_0)) = \pm 1$. By setting $L_0 = c_1(P_0)$ in (11) and observing that there are infinitely many values for the Seiberg–Witten invariants of $P_n^m(X)$, we conclude that $\{P_n^m(X) \mid m \geq 1\}$ contains infinitely many pairwise nondiffeomorphic 4-manifolds.

Next we prove that $P_n^m(X)$ is irreducible for all m large enough, or more specifically when $\mathrm{SW}_{P_n^m(X)}(\xi_m(c_1(P_0))) \neq 0$. We will argue the same way as in the proof of Theorem 5.4 in [Kotschick 1997]. Suppose $P_n^m(X) = M \# N$ is a connected sum of two smooth 4-manifolds M and N. Both M and N are simply connected since $P_n^m(X)$ is. If $b_2^+(M)$ and $b_2^+(N)$ are both positive, then the Seiberg–Witten invariant of $P_n^m(X)$ is trivial (cf. [Witten 1994]), a contradiction. Without loss of generality, assume $b_2^+(N) = 0$. If $b_2(N) = 0$, then the simply connected 4-manifold N must be homeomorphic to S^4 by Freedman's theorem in [Freedman 1982]. Thus it remains to rule out the case when $b_2(N) = b_2^-(N) > 0$. In this case, the intersection form of N is a nontrivial negative definite form, so by Donaldson's theorem in [Donaldson 1983], it is equivalent to the standard diagonal form. Let $e_1, \dots, e_{b_2(N)}$ be a basis for $H^2(N; \mathbb{Z})$ such that $e_i^2 = -1$ for each $i = 1, \dots, b_2(N)$, and $e_i \cdot e_j = 0$ when $i \neq j$. Using the neck pinching argument as in [Donaldson 1996; Kotschick 1997], we deduce that M has nontrivial Seiberg–Witten invariant. Moreover, if L is any Seiberg–Witten basic class of M, then the cohomology classes

$$L + \sum_{i=1}^{b_2(N)} a_i e_i, \tag{12}$$

where $a_i = \pm 1$ for each $i = 1, \dots, b_2(N)$, are all Seiberg–Witten basic classes of $P_n^m(X) = M \# N$. Furthermore, every Seiberg–Witten basic class of $P_n^m(X)$ can be written as (12).

Let $L_m = \xi_m(c_1(P_0))$ be a Seiberg–Witten basic class of $P_n^m(X)$. By changing any basis element e_i to $-e_i$ if necessary, we can assume that $L_m = L - e_1 - \cdots - e_{b_2(N)}$ for some L. Thus $L_m + 2e_1 = L + e_1 - e_2 - \cdots - e_{b_2(N)}$ is also a Seiberg–Witten basic class of $P_n^m(X)$. By the adjunction inequality, we can assume that $\xi_1(c_1(P_0)) = c_1(P_n^1(X))$. It now follows from (11) that there exists $\bar{e}_1 \in \xi_m^{-1}(e_1) \subset H^\perp$ such that $c_1(P_n^1(X)) + 2\xi_1(\bar{e}_1)$ or $c_1(P_0) + 2\bar{e}_1$ is a Seiberg–Witten basic class of $P_n^1(X)$ or P_0, respectively. By a theorem of Taubes [1996], we can then deduce that the Poincaré dual of $\xi_1(\bar{e}_1)$ or $\bar{e}_1$ is represented by an embedded symplectic sphere of self-intersection -1 in $P_n^1(X)$ or P_0, respectively (cf. Remark 10.1.16(b) in [Gompf and Stipsicz 1999]). This implies that $P_n^1(X)$ or P_0 is not minimal, a contradiction.

Finally, if $n \geq 3$, then $Y_n(1)$ contains $2n-4$ pairs of geometrically dual Lagrangian tori that are all disjoint from Σ_n. The images of these $4n-8$ tori in the fiber sum $P_n^1(X)$ are again Lagrangian submanifolds (cf. [Gompf 1995]). Let T_1 and T_2 be two of these $4n-8$ Lagrangian tori in $P_n^1(X)$ that are not geometrically dual to each other. By perturbing the symplectic form on $P_n^1(X)$, we can turn both T_1 and T_2 into symplectic submanifolds of $P_n^1(X)$ (cf. [Gompf 1995, Lemma 1.6]).

To show $\pi_1(P_n^1(X) \setminus (T_1 \cup T_2)) = 1$, it will be convenient to fix T_1 and T_2, say $T_1 = a_1' \times c_3''$ and $T_2 = a_2' \times d_3''$. Here, a_1', a_2', c_3'' and d_3'' are parallel copies of a_1, a_2, c_3 and d_3 as defined in [Fintushel et al. 2007]. Then $\pi_1(P_n^1(X) \setminus (T_1 \cup T_2))$ is normally generated by meridians of T_1 and T_2, which are all conjugate to the commutators $[b_1^{-1}, d_3]$ or $[b_2^{-1}, c_3]$. Note that the generators b_1, b_2, c_3 and d_3 are still trivial in $\pi_1(P_n^1(X) \setminus (T_1 \cup T_2))$ since the Luttinger surgery relations in Section 2 of [Akhmedov and Park 2010a] still hold true in $\pi_1(P_n^1(X) \setminus (T_1 \cup T_2))$. It follows that meridians of T_1 and T_2 are all trivial and hence $\pi_1(P_n^1(X) \setminus (T_1 \cup T_2)) = \pi_1(P_n^1(X)) = 1$. □

Instead of using $Y_n(m)$ summand in generalized fiber sum (9), we may use $Y_{n-2}(m) \# 2\overline{\mathbb{CP}}^2$ when $n \geq 4$. Specifically, we resolve the intersection between Σ_2 and Σ_{n-2} in $Y_{n-2}(m)$ to obtain a genus n submanifold of $Y_{n-2}(m)$ with self-intersection 2. Next we blow up two points on this submanifold to obtain a genus n submanifold Σ_n' of self-intersection 0 in $Y_{n-2}(m) \# 2\overline{\mathbb{CP}}^2$. When $m = 1$, the resolution and the blowups can be performed symplectically, and hence $(Y_{n-2}(1) \# 2\overline{\mathbb{CP}}^2, \Sigma_n')$ is a relatively minimal pair of symplectic manifolds. The advantage of using $Y_{n-2}(m) \# 2\overline{\mathbb{CP}}^2$ summand is that the resulting generalized fiber sum has slightly smaller characteristic numbers than $P_n^m(X)$.

Theorem 9. *Let $f : X \to \Sigma_b$ be a relatively minimal Lefschetz fibration as above having at least one nonseparating vanishing cycle. Suppose that $n = ta + b \geq 4$. For a suitable choice of the gluing diffeomorphism $\psi : \partial(\nu\widetilde{\Sigma}) \to \partial(\nu\Sigma_n')$, the generalized fiber sum*

$$\begin{aligned} Q_n^m(X) &= \widetilde{X} \#_\psi (Y_{n-2}(m) \# 2\overline{\mathbb{CP}}^2) \\ &= (\widetilde{X} \setminus \nu\widetilde{\Sigma}) \cup_\psi \big((Y_{n-2}(m) \# 2\overline{\mathbb{CP}}^2) \setminus \nu\Sigma_n'\big) \end{aligned}$$

along $\widetilde{\Sigma}$ and Σ_n' is simply connected, nonspin, and satisfies

$$\begin{aligned} e(Q_n^m(X)) &= e(X) + d + (8a+2)t + 8b - 14, \\ \sigma(Q_n^m(X)) &= \sigma(X) - 2t - d - 2, \\ \chi_h(Q_n^m(X)) &= \chi_h(X) + 2at + 2b - 4, \\ c_1^2(Q_n^m(X)) &= c_1^2(X) - d + (16a-2)t + 16b - 34, \\ b_2^+(Q_n^m(X)) &= b_2^+(X) - b_1(X) + 4at + 4b - 8 \geq 3, \\ b_2^-(Q_n^m(X)) &= b_2^-(X) - b_1(X) + d + (4a+2)t + 4b - 6. \end{aligned}$$

The set $\{Q_n^m(X) \mid m \geq 1\}$ contains infinitely many homeomorphic but pairwise nondiffeomorphic irreducible 4-manifolds. When $m = 1$, $Q_n^1(X)$ is symplectic and irreducible. If $n = ta + b \geq 5$, then $Q_n^1(X)$ contains disjoint symplectic tori T_1' and T_2' of self-intersection 0 satisfying $\pi_1(Q_n^1(X) \setminus (T_1' \cup T_2')) = 1$.

Proof. We compute that

$$\begin{aligned} e(Q_n^m(X)) &= e(\widetilde{X}) + e(Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2) - 2e(\Sigma_n') \\ &= e(X) + 2t + d + 4(n-2) - 4 + 2 - 2(2-2n) \\ &= e(X) + 2t + d + 8n - 14 \\ &= e(X) + 2t + d + 8ta + 8b - 14, \end{aligned}$$

$$\sigma(Q_n^m(X)) = \sigma(\widetilde{X}) + \sigma(Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2) = \sigma(X) - 2t - d - 2.$$

The other characteristic numbers can be computed from these as before.

Since the exceptional sphere of a blowup intersects Σ_n' once transversely, any meridian of Σ_n' is null-homotopic in the complement of a tubular neighborhood $\nu\Sigma_n'$. Hence we conclude that

$$\pi_1\big((Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2) \setminus \nu\Sigma_n'\big) = \pi_1\big(Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2\big) = \pi_1(Y_{n-2}(m)).$$

From [Akhmedov and Park 2010a], we know that $\pi_1(Y_{n-2}(m))/\langle z\rangle = 1$, where z is the image of any one of the generators c_1, d_1, c_2, d_2 of $\pi_1(\Sigma_{n-2})$ under the inclusion induced homomorphism.

Let $\widetilde{\Sigma}^{\|}$ and $\Sigma_n'^{\|}$ denote parallel copies of $\widetilde{\Sigma}$ and Σ_n' in the boundaries $\partial(\nu\widetilde{\Sigma})$ and $\partial(\nu\Sigma_n')$, respectively. When forming the generalized fiber sum $Q_n^m(X)$, we choose the gluing diffeomorphism ψ such that ψ_* maps the element of $\pi_1(\widetilde{\Sigma}^{\|})$ represented by a nonseparating vanishing cycle of X to z, viewed as an element of $\pi_1(\Sigma_n'^{\|})$. Thus $z = 1$ in $\pi_1(Q_n^m(X))$, which then implies that the inclusion induced homomorphism

$$\pi_1\big((Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2) \setminus \nu\Sigma_n'\big) \longrightarrow \pi_1(Q_n^m(X)) \tag{13}$$

is trivial. Note that the inclusion induced homomorphism $\pi_1(\widetilde{\Sigma}^{\|}) \to \pi_1(Q_n^m(X))$ is also trivial since it can be factored through homomorphism (13) after $\widetilde{\Sigma}^{\|}$ is identified with $\Sigma_n'^{\|}$. It follows from Lemma 7 that the inclusion induced homomorphism $\pi_1(\widetilde{X} \setminus \nu\widetilde{\Sigma}) \to \pi_1(Q_n^m(X))$ is trivial as well. By Seifert–van Kampen theorem, we conclude that $\pi_1(Q_n^m(X)) = 1$.

$Q_n^m(X)$ is nonspin since it contains a surface of self-intersection -1 and genus $a > 0$, namely the internal sum of the image of a punctured fiber of X in $\widetilde{X} \setminus \nu\widetilde{\Sigma}$ and a punctured exceptional sphere in $(Y_{n-2}(m)\,\#\,2\overline{\mathbb{CP}}^2) \setminus \nu\Sigma_n'$. Since $Y_{n-2}(1)\,\#\,2\overline{\mathbb{CP}}^2$ is symplectic, the corresponding fiber sum $Q_n^1(X)$ is symplectic as well. The irreducibility of $Q_n^1(X)$ and the fact that $\{Q_n^m(X) \mid m \geq 1\}$ contains infinitely many homeomorphic but pairwise nondiffeomorphic irreducible 4-manifolds can be proved exactly the same way as in the proof of Theorem 8.

Finally, if $n \geq 5$, then $Y_{n-2}(1)$ contains $2n-8$ pairs of geometrically dual Lagrangian tori. The images of these $4n-16$ tori in the blowup $Y_{n-2}(1)\,\#\,2\overline{\mathbb{CP}}^2$ are disjoint from Σ'_n, and hence their images in $Q_n^1(X)$ are Lagrangian submanifolds of $Q_n^1(X)$. Let T'_1 and T'_2 denote two of these $4n-16$ Lagrangian tori, say $T'_1 = a'_1 \times c''_3$ and $T'_2 = a'_2 \times d''_3$. By perturbing the symplectic form on $Q_n^1(X)$, we can turn both T'_1 and T'_2 into symplectic submanifolds of $Q_n^1(X)$. We can deduce that $\pi_1(Q_n^1(X)\setminus(T'_1\cup T'_2)) = 1$ in exactly the same way as in the proof of Theorem 8. $\square$

For comparison, we note that

$$(14)\qquad \begin{aligned} e(Q_n^m(X)) &= e(P_n^m(X)) - 6, & \sigma(Q_n^m(X)) &= \sigma(P_n^m(X)) - 2,\\ \chi_h(Q_n^m(X)) &= \chi_h(P_n^m(X)) - 2, & c_1^2(Q_n^m(X)) &= c_1^2(P_n^m(X)) - 18,\\ b_2^+(Q_n^m(X)) &= b_2^+(P_n^m(X)) - 4, & b_2^-(Q_n^m(X)) &= b_2^-(P_n^m(X)) - 2. \end{aligned}$$

Remark 10. The irreducible symplectic 4-manifolds M and N (homeomorphic to $47\mathbb{CP}^2\,\#\,45\overline{\mathbb{CP}}^2$ and $51\mathbb{CP}^2\,\#\,47\overline{\mathbb{CP}}^2$, respectively) in Section 4 of [Akhmedov and Park 2008] are respectively equal to $Q_n^1(X)$ and $P_n^1(X)$ with $a = 7$, $b = 2$, $t = 1$, $d = -2$, $n = 9$, $e(X) = 36$, and $\sigma(X) = 4$.

5. Simply connected 4-manifolds with positive signature

We now apply Theorems 8 and 9 to Lefschetz fibrations in Sections 2 and 3 to obtain new families of simply connected irreducible 4-manifolds with positive signature.

Example 11. For each triple of positive integers g, u_1, u_2, recall from Example 3 that there is a Lefschetz fibration $f_{u_1,u_2} : X^g_{u_1,u_2} \to \Sigma_b$ such that the genus of a regular fiber is $a = 1 + \frac{1}{2}u_2(g+1)(3g-2)$ and there is a section whose image is a surface of genus $b = u_1(g-1)+1$ and self-intersection $d = -2u_1$. Since $2t + d \geq 0$, we require $t \geq u_1$. Let

$$n = t + \tfrac{1}{2}tu_2(g+1)(3g-2) + u_1(g-1) + 1.$$

Applying Theorem 8 to $f_{u_1,u_2} : X^g_{u_1,u_2} \to \Sigma_b$, we obtain a family of simply connected 4-manifolds $P_n^m(X^g_{u_1,u_2})$, with $m \geq 1$ and $n \geq 3$, satisfying

$$(15)\qquad \begin{aligned} e(P_n^m(X^g_{u_1,u_2})) &= 2u_1u_2(g+1)(3g^2-5g+4)\\ &\qquad + 4tu_2(g+1)(3g-2) + 8u_1g + 10t - 10u_1,\\ \sigma(P_n^m(X^g_{u_1,u_2})) &= \tfrac{2}{3}u_1u_2(g+1)(g^2+2g-6) - 2t + 2u_1,\\ \chi_h(P_n^m(X^g_{u_1,u_2})) &= \tfrac{1}{6}u_1u_2(g+1)(10g^2-13g+6)\\ &\qquad + tu_2(g+1)(3g-2) + 2t + 2u_1(g-1),\\ c_1^2(P_n^m(X^g_{u_1,u_2})) &= 2u_1u_2(g+1)(7g^2-8g+2)\\ &\qquad + 8tu_2(g+1)(3g-2) + 16u_1g + 14t - 14u_1, \end{aligned}$$

$$b_2^+(P_n^m(X_{u_1,u_2}^g)) = \tfrac{1}{3}u_1u_2(g+1)(10g^2-13g+6) \\ + 2tu_2(g+1)(3g-2) + 4t + 4u_1(g-1) - 1,$$

$$b_2^-(P_n^m(X_{u_1,u_2}^g)) = \tfrac{1}{3}u_1u_2(g+1)(8g^2-17g+18) \\ + 2tu_2(g+1)(3g-2) + 4u_1g + 6t - 6u_1 - 1. \tag{16}$$

From Theorem 9, we obtain another family of simply connected nonspin 4-manifolds $Q_n^m(X_{u_1,u_2}^g)$, with $m \geq 1$ and $n \geq 5$, whose characteristic numbers can be computed from (14) (15), and (16). Moreover, when $m = 1$, both $P_n^1(X_{u_1,u_2}^g)$ and $Q_n^1(X_{u_1,u_2}^g)$ are irreducible symplectic 4-manifolds and contain symplectic tori T_j and T_j' ($j = 1, 2$) of self-intersection 0 such that

$$\pi_1(P_n^1(X_{u_1,u_2}^g) \setminus (T_1 \cup T_2)) = 1 \quad \text{and} \quad \pi_1(Q_n^1(X_{u_1,u_2}^g) \setminus (T_1' \cup T_2')) = 1.$$

Example 12. For each quadruple of positive integers p, v, u_1, u_2 with odd $p \geq 3$, recall from Example 5 that there is a Lefschetz fibration $f_{u_1,u_2} : W_{u_1,u_2}^{p,v} \to \Sigma_b$ such that the genus of a regular fiber is $a = 1 + \frac{1}{2}pu_2[(v+1)p - v - 3]$ and there is a section whose image is a surface of genus $b = 1 + u_1[-1 + v(p-1)/2]$ and self-intersection $d = -u_1 v$. Since $2t + d \geq 0$, we require

$$t \geq \lceil u_1 v/2 \rceil,$$

where $\lceil x \rceil = \min\{k \in \mathbb{Z} \mid k \geq x\}$. From Theorems 8 and 9, we obtain two families of simply connected 4-manifolds $P_n^m(W_{u_1,u_2}^{p,v})$ and $Q_n^m(W_{u_1,u_2}^{p,v})$ with $m \geq 1$ and

$$n = t + \tfrac{1}{2}tpu_2[(v+1)p - v - 3] + u_1[-1 + v(p-1)/2] + 1 \geq 5.$$

We compute that

$$e(P_n^m(W_{u_1,u_2}^{p,v})) = pu_1u_2\big[(v^2+v)p^2 - 2(v^2+3v+1)p + v^2 + 6v + 8\big] \\ + 4tu_2(v+1)p^2 + 4\big[u_1v - tu_2(v+3)\big]p + 10t - 5u_1v - 8u_1,$$

$$\sigma(P_n^m(W_{u_1,u_2}^{p,v})) = \tfrac{1}{3}pu_1u_2(vp^2 - 4v - 6) - 2t + u_1v,$$

$$\chi_h(P_n^m(W_{u_1,u_2}^{p,v})) = \tfrac{1}{12}pu_1u_2\big[(3v^2+4v)p^2 - 6(v^2+3v+1)p + 3v^2 + 14v + 18\big] \\ + tu_2(v+1)p^2 + \big[u_1v - tu_2(v+3)\big]p + 2t - u_1v - 2u_1,$$

$$c_1^2(P_n^m(W_{u_1,u_2}^{p,v})) = pu_1u_2\big[(2v^2+3v)p^2 - 4(v^2+3v+1)p + 2v^2 + 8v + 10\big] \\ + 8tu_2(v+1)p^2 + 8\big[u_1v - tu_2(v+3)\big]p + 14t - 7u_1v - 16u_1,$$

$$b_2^+(P_n^m(W_{u_1,u_2}^{p,v})) = \tfrac{1}{6}pu_1u_2\big[(3v^2+4v)p^2 - 6(v^2+3v+1)p + 3v^2 + 14v + 18\big] \\ + 2tu_2(v+1)p^2 + 2\big[u_1v - tu_2(v+3)\big]p + 4t - 2u_1v - 4u_1 - 1,$$

$$b_2^-(P_n^m(W_{u_1,u_2}^{p,v})) = \tfrac{1}{6}pu_1u_2\big[(3v^2+2v)p^2 - 6(v^2+3v+1)p + 3v^2 + 22v + 30\big] \\ + 2tu_2(v+1)p^2 + 2\big[u_1v - tu_2(v+3)\big]p + 6t - 3u_1v - 4u_1 - 1.$$

The characteristic numbers of $Q_n^m(W_{u_1,u_2}^{p,v})$ can be computed from these values via (14). When $m = 1$, both $P_n^1(W_{u_1,u_2}^{p,v})$ and $Q_n^1(W_{u_1,u_2}^{p,v})$ are irreducible symplectic 4-manifolds and contain symplectic tori T_j and T_j' ($j = 1, 2$) of self-intersection 0 such that $\pi_1(P_n^1(W_{u_1,u_2}^{p,v}) \setminus (T_1 \cup T_2)) = 1$ and $\pi_1(Q_n^1(W_{u_1,u_2}^{p,v}) \setminus (T_1' \cup T_2')) = 1$.

6. Upper bounds for the lower bound

We start this section by giving a more rigorous definition of $\lambda(\sigma)$ from the introduction.

Definition 13. Given an integer $\sigma \geq 0$, let $\lambda(\sigma)$ be the smallest positive integer with the following properties.

(i) $\lambda(\sigma) \geq \lceil(\sigma+1)/2\rceil$.

(ii) Every point (χ_h, c_1^2) on the line $c_1^2 = 8\chi_h + \sigma$ satisfying $\chi_h \geq \lambda(\sigma)$ is realized as $(\chi_h(M_i), c_1^2(M_i))$, where $\{M_i \mid i \in \mathbb{Z}\}$ is an infinite family of homeomorphic but pairwise nondiffeomorphic closed simply connected nonspin irreducible 4-manifolds such that M_i is symplectic for each $i \geq 0$ and M_i is nonsymplectic for each $i < 0$.

As in the introduction, we make the following conjecture.

Conjecture 14. $\lambda(\sigma) = \lceil(\sigma+1)/2\rceil$ for every integer $\sigma \geq 0$.

Our goal in this section is to calculate explicit upper bounds on $\lambda(\sigma)$ for many small values of σ. First we restate a result from [Akhmedov and Park 2008] (see also [Akhmedov et al. 2010a, Theorem 23; Akhmedov and Park 2010a, Theorem 2]).

Theorem 15 [Akhmedov and Park 2008, Theorem 5.3]. *Let X be a closed symplectic 4-manifold that contains a symplectic torus T of self-intersection 0. Let νT be a tubular neighborhood of T and $\partial(\nu T)$ its boundary. Suppose that the homomorphism $\pi_1(\partial(\nu T)) \to \pi_1(X \setminus \nu T)$ induced by the inclusion is trivial. Then for any pair of integers (χ, c) satisfying*

$$\chi \geq 1 \quad \textit{and} \quad 0 \leq c \leq 8\chi, \tag{17}$$

there exists a symplectic 4-manifold Y with $\pi_1(Y) = \pi_1(X)$,

$$\chi_h(Y) = \chi_h(X) + \chi \quad \textit{and} \quad c_1^2(Y) = c_1^2(X) + c.$$

Moreover, if X is minimal then Y is minimal as well. If $c < 8\chi$, or if $c = 8\chi$ and X has an odd intersection form, then the corresponding Y has an odd indefinite intersection form. □

The next theorem gives us a means for constructing infinitely many distinct smooth structures on some topological 4-manifolds.

Theorem 16. *Let Y be a closed simply connected minimal symplectic 4-manifold with $b_2^+(Y) > 1$. Assume that Y contains a symplectic torus T of self-intersection 0 such that $\pi_1(Y \setminus T) = 1$. Then there exist an infinite family of pairwise nondiffeomorphic irreducible symplectic 4-manifolds and an infinite family of pairwise nondiffeomorphic irreducible nonsymplectic 4-manifolds, all of which are homeomorphic to Y.*

Proof. We can perform a knot surgery on Y along T using a knot $K \subset S^3$ (see [Fintushel and Stern 2009, Lecture 3]). Let Y_K denote the resulting 4-manifold. Since $\pi_1(Y \setminus T) = 1$, Y_K is homeomorphic to Y. By varying the knot K, we obtain infinitely many pairwise nondiffeomorphic 4-manifolds. If K is a fibered knot, then Y_K can be viewed as a symplectic fiber sum [Fintushel and Stern 1998], is minimal by Usher's theorem [2006], and hence is irreducible [Hamilton and Kotschick 2006; Kotschick 1997].

Given an integer $k \neq 0$, let $T(k)$ denote the k-twist knot on page 372 of [Fintushel and Stern 1998] with Alexander polynomial $kt - (2k+1) + kt^{-1}$. If $k = \pm 1$, then $T(\pm 1)$ is fibered, and thus $Y_{T(\pm 1)}$ is symplectic and irreducible. If $k \neq 0, \pm 1$, then $Y_{T(k)}$ is nonsymplectic. It only remains to prove that $Y_{T(k)}$ is irreducible when $k \neq 0, \pm 1$. We will argue the same way as in the proof of Theorem 8. The computation of the Seiberg–Witten invariant of $Y_{T(k)}$ in [Fintushel and Stern 2009] implies that there exists an isomorphism $\xi_{T(k)} : H^2(Y_{T(1)}; \mathbb{Z}) \longrightarrow H^2(Y_{T(k)}; \mathbb{Z})$ that preserves the cup product pairing and restricts to a one-to-one correspondence between the Seiberg–Witten basic classes of $Y_{T(1)}$ and $Y_{T(k)}$. Suppose that $Y_{T(k)}$ is not irreducible. Then there will be some $e_1 \in H^2(Y_{T(k)}; \mathbb{Z})$ such that $e_1^2 = -1$ and $\xi_{T(k)}(c_1(Y_{T(1)})) + 2e_1$ is a Seiberg–Witten basic class of $Y_{T(k)}$. This will imply that $c_1(Y_{T(1)}) + 2\xi_{T(k)}^{-1}(e_1)$ is a Seiberg–Witten basic class of $Y_{T(1)}$. By a result of Taubes [1996], we can then conclude that the Poincaré dual of $\xi_{T(k)}^{-1}(e_1)$ is represented by an embedded symplectic sphere of self-intersection -1 in $Y_{T(1)}$. Hence $Y_{T(1)}$ is not minimal, a contradiction. □

By combining Theorems 15 and 16, we may deduce the following.

Corollary 17. *Let X be a closed simply connected nonspin minimal symplectic 4-manifold with $b_2^+(X) > 1$ and $\sigma(X) \geq 0$. Assume that X contains disjoint symplectic tori T_1 and T_2 of self-intersection 0 such that $\pi_1(X \setminus (T_1 \cup T_2)) = 1$. Suppose σ is a fixed integer satisfying $0 \leq \sigma \leq \sigma(X)$. If $\lceil x \rceil = \min\{k \in \mathbb{Z} \mid k \geq x\}$ and if we define*

$$\ell(\sigma) = \left\lceil \frac{\sigma(X) - \sigma}{8} - 1 \right\rceil,$$

then

$$\lambda(\sigma) \leq \chi_h(X) + \ell(\sigma) + 1.$$

In other words, if k is any odd integer satisfying $k \geq b_2^+(X) + 2\ell(\sigma) + 2$, then there exist an infinite family of pairwise nondiffeomorphic irreducible symplectic 4-manifolds and an infinite family of pairwise nondiffeomorphic irreducible nonsymplectic 4-manifolds, all of which are homeomorphic to $k\mathbb{CP}^2 \# (k-\sigma)\overline{\mathbb{CP}}^2$.

Proof. We can write $\sigma(X) - \sigma = 8\ell(\sigma) + r(\sigma)$ for integers $\ell(\sigma)$ and $r(\sigma)$ satisfying $\ell(\sigma) \geq -1$ and $1 \leq r(\sigma) \leq 8$. Since $\pi_1(X \setminus \nu T_1) = 1$, we can apply Theorem 15 to the pair, X and T_1. Let (χ, c) and Y be as in the conclusion of Theorem 15. Since $\pi_1(Y) = \pi_1(X) = 1$, we have $b_2^+(Y) = b_2^+(X) + 2\chi$ and $b_2^-(Y) = b_2^-(X) + 10\chi - c$. By Freedman's classification theorem [1982], Y must be homeomorphic to

$$(b_2^+(X) + 2\chi)\mathbb{CP}^2 \# (b_2^-(X) + 10\chi - c)\overline{\mathbb{CP}}^2.$$

By setting $c = 8\chi + \sigma - \sigma(X)$ in (17), we obtain a minimal symplectic 4-manifold Y that is homeomorphic to $k\mathbb{CP}^2 \# (k-\sigma)\overline{\mathbb{CP}}^2$, where $k = b_2^+(X) + 2\chi$. Since c is nonnegative, we must have $8\chi + \sigma - \sigma(X) = 8(\chi - \ell(\sigma)) - r(\sigma) \geq 0$, which implies that $\chi \geq \ell(\sigma) + 1$. It follows that $\chi_h(Y) \geq \chi_h(X) + \ell(\sigma) + 1$ and $k \geq b_2^+(X) + 2\ell(\sigma) + 2$.

We recall from [Akhmedov et al. 2010a; Akhmedov and Park 2008; 2010a] that for each pair of integers (χ, c) satisfying (17), there exist a minimal symplectic 4-manifold Z with $\chi_h(Z) = \chi$, $c_1^2(Z) = c$, and a symplectic torus $T'' \subset Z$ of self-intersection 0 such that Y is the generalized fiber sum of X and Z along T_1 and T''. Note that $T_2 \subset (X \setminus \nu T_1) \subset Y$ is a symplectic torus of self-intersection 0 in Y (cf. [Gompf and Stipsicz 1999, Theorem 10.2.1]). Since $\pi_1(X \setminus (\nu T_1 \cup T_2)) = 1$, we have $\pi_1(Y \setminus T_2) = 1$. We can now apply Theorem 16 to the pair, Y and T_2, and conclude that there are infinitely many distinct smooth structures on Y. □

Next we show that $\lambda(\sigma)$ is subadditive in the following sense.

Corollary 18. *Let σ_1 and σ_2 be positive integers such that $\sigma_1 + \sigma_2$ is not divisible by 16. For each $j = 1, 2$, suppose that there exists a closed simply connected nonspin minimal symplectic 4-manifold N_j containing a symplectic torus $T_j \subset N_j$ of self-intersection 0 such that*

(i) $\pi_1(N_j \setminus T_j) = 1$,

(ii) $\chi_h(N_j) = \lambda(\sigma_j)$, *and* $\sigma(N_j) = \sigma_j$.

Then we have $\lambda(\sigma_1 + \sigma_2) \leq \lambda(\sigma_1) + \lambda(\sigma_2)$.

Proof. Let X be the generalized fiber sum of N_1 and N_2 along T_1 and T_2. It is easy to check that X is a closed simply connected minimal symplectic 4-manifold. Since

$$\sigma(X) = \sigma(N_1) + \sigma(N_2) = \sigma_1 + \sigma_2 \not\equiv 0 \pmod{16},$$

X is nonspin by Rohlin's theorem [1952]. Let T be a parallel copy of T_1 (and T_2) in X. From (i), there are topological disks bounding the meridians of T_1 and T_2, and these disks can be glued together to form a topological sphere that intersects T transversely once. It follows that $\pi_1(X \setminus T) = 1$ and thus we can apply Corollary 17 with $\sigma = \sigma(X)$ and conclude that

$$\lambda(\sigma_1 + \sigma_2) \le \chi_h(X) = \chi_h(N_1) + \chi_h(N_2) = \lambda(\sigma_1) + \lambda(\sigma_2). \qquad \square$$

We now proceed to list the smallest upper bounds on $\lambda(\sigma)$ currently known to the authors. We begin by first finding parameters g, p, v, u_1, u_2 and t in Examples 11 and 12 that yield 4-manifolds with small χ_h values. By Rohlin's theorem, these 4-manifolds are nonspin if their signatures are not divisible by 16. Unfortunately, given an integer $\sigma \ge 0$, there is no clear pattern as to which family or parameters

σ	$\lambda(\sigma) \le$	X	σ	$\lambda(\sigma) \le$	X
0–1	25	$Q^1_9(W^{3,2}_{1,1})$	50	86	$P^1_{19}(W^{5,1}_{2,1})$
2	24	$Q^1_9(W^{3,2}_{1,1})$	51	111	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_9(W^{3,2}_{1,1})$
3	27	$P^1_9(W^{3,2}_{1,1})$	52	110	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_9(W^{3,2}_{1,1})$
4	26	$P^1_9(W^{3,2}_{1,1})$	53	113	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi P^1_9(W^{3,2}_{1,1})$
5	47	$Q^1_{15}(W^{3,2}_{1,2})$	54	112	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi P^1_9(W^{3,2}_{1,1})$
6	46	$Q^1_{15}(W^{3,2}_{1,2})$	55	133	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_{15}(W^{3,2}_{1,2})$
7	49	$P^1_{15}(W^{3,2}_{1,2})$	56	132	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_{15}(W^{3,2}_{1,2})$
8	48	$P^1_{15}(W^{3,2}_{1,2})$	57	135	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi P^1_{15}(W^{3,2}_{1,2})$
9–13	59	$Q^1_{18}(W^{5,1}_{1,1})$	58	134	$P^1_{19}(W^{5,1}_{2,1})\,\#_\varphi P^1_{15}(W^{3,2}_{1,2})$
14–21	58	$Q^1_{18}(W^{5,1}_{1,1})$	59–61	143	$Q^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_{18}(W^{5,1}_{1,1})$
22	57	$Q^1_{18}(W^{5,1}_{1,1})$	62–69	142	$Q^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_{18}(W^{5,1}_{1,1})$
23	60	$P^1_{18}(W^{5,1}_{1,1})$	70	141	$Q^1_{19}(W^{5,1}_{2,1})\,\#_\varphi Q^1_{18}(W^{5,1}_{1,1})$
24	59	$P^1_{18}(W^{5,1}_{1,1})$	71	144	$Q^1_{36}(W^{5,1}_{3,1})$
25	84	$P^1_{18}(W^{5,1}_{1,1})\,\#_\varphi Q^1_9(W^{3,2}_{1,1})$	72	143	$Q^1_{36}(W^{5,1}_{3,1})$
26	83	$P^1_{18}(W^{5,1}_{1,1})\,\#_\varphi Q^1_9(W^{3,2}_{1,1})$	73	146	$P^1_{36}(W^{5,1}_{3,1})$
27	86	$P^1_{18}(W^{5,1}_{1,1})\,\#_\varphi P^1_9(W^{3,2}_{1,1})$	74	145	$P^1_{36}(W^{5,1}_{3,1})$
28	85	$P^1_{18}(W^{5,1}_{1,1})\,\#_\varphi P^1_9(W^{3,2}_{1,1})$	75–81	167	$Q^1_{34}(W^{5,1}_{2,2})$
29–31	87	$Q^1_{19}(W^{5,1}_{2,1})$	82–89	166	$Q^1_{34}(W^{5,1}_{2,2})$
32–39	86	$Q^1_{19}(W^{5,1}_{2,1})$	90–97	165	$Q^1_{34}(W^{5,1}_{2,2})$
40–47	85	$Q^1_{19}(W^{5,1}_{2,1})$	98	164	$Q^1_{34}(W^{5,1}_{2,2})$
48	84	$Q^1_{19}(W^{5,1}_{2,1})$	99	167	$P^1_{34}(W^{5,1}_{2,2})$
49	87	$P^1_{19}(W^{5,1}_{2,1})$	100	166	$P^1_{34}(W^{5,1}_{2,2})$

Table 2. Upper bounds on $\lambda(\sigma)$.

will yield a simply connected nonspin 4-manifold X with $\sigma(X) \geq \sigma$ having the smallest $\chi_h(X) + \ell(\sigma) + 1$. Hence we had to resort to a computer search.

Table 2 on the previous page lists some of the smallest upper bounds on $\lambda(\sigma)$ that we found. For example, when $\sigma = 10$, Table 2 says that $\lambda(10) \leq 59$, that is, for each odd integer $k \geq 2 \cdot 59 - 1 = 117$, there exist an infinite family of pairwise nondiffeomorphic irreducible symplectic 4-manifolds and an infinite family of pairwise nondiffeomorphic irreducible nonsymplectic 4-manifolds, all of which are homeomorphic to $k\mathbb{CP}^2 \# (k-10)\overline{\mathbb{CP}}^2$. The third column in Table 2 lists the simply connected 4-manifold X that was used to obtain the upper bound via Corollary 17. The $\#\varphi$ symbol denotes a generalized fiber sum along the tori T_j and/or T'_j. We have compiled upper bounds on $\lambda(\sigma)$ for σ up to about 1,000,000 but we will only list a small sample here.

Acknowledgements

The authors thank Vincent Chan, David McKinnon and Shigeru Takamura for very helpful comments. We also thank the referee for suggesting numerous ways to improve the paper. Some of the computations for Table 2 were done with the aid of Maple and Mathematica.

References

[Akhmedov and Park 2008] A. Akhmedov and B. D. Park, "New symplectic 4-manifolds with nonnegative signature", *J. Gökova Geom. Topol. GGT* **2** (2008), 1–13. MR 2009m:57044 Zbl 1184.57017

[Akhmedov and Park 2010a] A. Akhmedov and B. D. Park, "Exotic smooth structures on small 4-manifolds with odd signatures", *Invent. Math.* **181**:3 (2010), 577–603. MR 2012b:57048 Zbl 1206.57029

[Akhmedov and Park 2010b] A. Akhmedov and B. D. Park, "Geography of simply connected spin symplectic 4-manifolds", *Math. Res. Lett.* **17**:3 (2010), 483–492. MR 2011h:57042 Zbl 05937386

[Akhmedov et al. 2008] A. Akhmedov, R. İ. Baykur, and B. D. Park, "Constructing infinitely many smooth structures on small 4-manifolds", *J. Topol.* **1**:2 (2008), 409–428. MR 2010c:57040 Zbl 1146.57041

[Akhmedov et al. 2010a] A. Akhmedov, S. Baldridge, R. İ. Baykur, P. Kirk, and B. D. Park, "Simply connected minimal symplectic 4-manifolds with signature less than -1", *J. Eur. Math. Soc.* (*JEMS*) **12**:1 (2010), 133–161. MR 2011b:57032 Zbl 1185.57023

[Akhmedov et al. 2010b] A. Akhmedov, B. D. Park, and G. Urzúa, "Spin symplectic 4-manifolds near Bogomolov–Miyaoka–Yau line", *J. Gökova Geom. Topol. GGT* **4** (2010), 55–66. MR 2755993

[Chen 1991] Z. J. Chen, "The existence of algebraic surfaces with preassigned Chern numbers", *Math. Z.* **206**:2 (1991), 241–254. MR 92a:14036 Zbl 0695.14018

[Donaldson 1983] S. K. Donaldson, "An application of gauge theory to four-dimensional topology", *J. Differential Geom.* **18**:2 (1983), 279–315. MR 85c:57015 Zbl 0507.57010

[Donaldson 1996] S. K. Donaldson, "The Seiberg–Witten equations and 4-manifold topology", *Bull. Amer. Math. Soc.* (*N.S.*) **33**:1 (1996), 45–70. MR 96k:57033 Zbl 0872.57023

[Fintushel and Stern 1998] R. Fintushel and R. J. Stern, "Knots, links, and 4-manifolds", *Invent. Math.* **134**:2 (1998), 363–400. MR 99j:57033 Zbl 0914.57015

[Fintushel and Stern 2009] R. Fintushel and R. J. Stern, "Six lectures on four 4-manifolds", pp. 265–315 in *Low dimensional topology*, edited by T. S. Mrowka and P. S. Ozsváth, IAS/Park City Math. Ser. **15**, Amer. Math. Soc., Providence, RI, 2009. MR 2010g:57035 Zbl 1195.57001

[Fintushel et al. 2007] R. Fintushel, B. D. Park, and R. J. Stern, "Reverse engineering small 4-manifolds", *Algebr. Geom. Topol.* **7** (2007), 2103–2116. MR 2009h:57044 Zbl 1142.57018

[Freedman 1982] M. H. Freedman, "The topology of four-dimensional manifolds", *J. Differential Geom.* **17**:3 (1982), 357–453. MR 84b:57006 Zbl 0528.57011

[Gompf 1995] R. E. Gompf, "A new construction of symplectic manifolds", *Ann. of Math.* (2) **142**:3 (1995), 527–595. MR 96j:57025 Zbl 0849.53027

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, 4-*manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038 Zbl 0933.57020

[Hamilton and Kotschick 2006] M. J. D. Hamilton and D. Kotschick, "Minimality and irreducibility of symplectic four-manifolds", *Int. Math. Res. Not.* **2006**:2 (2006), Art. ID 35032, 13. MR 2007i:57023 Zbl 1101.53052

[Harris and Morrison 1998] J. Harris and I. Morrison, *Moduli of curves*, Graduate Texts in Mathematics **187**, Springer, New York, 1998. MR 99g:14031 Zbl 0913.14005

[Kotschick 1997] D. Kotschick, "The Seiberg–Witten invariants of symplectic four-manifolds (after C. H. Taubes)", pp. 195–220 in *Séminaire Bourbaki* 1995/1996 (Exposé 812), Astérisque **241**, Société Mathématique de France, Paris, 1997. MR 98h:57057 Zbl 0882.57026

[Li 1999] T.-J. Li, "Smoothly embedded spheres in symplectic 4-manifolds", *Proc. Amer. Math. Soc.* **127**:2 (1999), 609–613. MR 99c:57055 Zbl 0911.57018

[Li and Stipsicz 2002] T.-J. Li and A. I. Stipsicz, "Minimality of certain normal connected sums", *Turkish J. Math.* **26**:1 (2002), 75–80. MR 2003c:57028 Zbl 1002.57047

[Luo 2000] F. Luo, "Torsion elements in the mapping class group of a surface", preprint, 2000. arXiv 0004048

[McCarthy and Wolfson 1994] J. D. McCarthy and J. G. Wolfson, "Symplectic normal connect sum", *Topology* **33**:4 (1994), 729–764. MR 95h:57038 Zbl 0812.53033

[Morgan et al. 1997] J. W. Morgan, T. S. Mrowka, and Z. Szabó, "Product formulas along T^3 for Seiberg–Witten invariants", *Math. Res. Lett.* **4**:6 (1997), 915–929. MR 99f:57039 Zbl 0892.57021

[Némethi 1999] A. Némethi, "Five lectures on normal surface singularities", pp. 269–351 in *Low dimensional topology* (Eger, 1996/Budapest, 1998), edited by K. Böröczky, Jr. et al., Bolyai Soc. Math. Stud. **8**, János Bolyai Math. Soc., Budapest, 1999. MR 2001g:32066 Zbl 0958.32026

[Nielsen 1937] J. Nielsen, "Die Struktur periodischer Transformationen von Flächen", *Danske Vid. Selsk. Mat.-Fys. Medd.* **15**:1 (1937), 1–77. Zbl 0017.13302

[Niepel 2005] M. Niepel, "Examples of symplectic 4-manifolds with positive signature", pp. 235–242 in *Geometry and topology of manifolds*, edited by H. U. Boden et al., Fields Inst. Commun. **47**, Amer. Math. Soc., Providence, RI, 2005. MR 2006k:57073 Zbl 1095.57024

[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

[Park and Szabó 2000] B. D. Park and Z. Szabó, "The geography problem for irreducible spin four-manifolds", *Trans. Amer. Math. Soc.* **352**:8 (2000), 3639–3650. MR 2000m:57037 Zbl 0947.57023

[Persson 1981] U. Persson, "Chern invariants of surfaces of general type", *Compositio Math.* **43**:1 (1981), 3–58. MR 83b:14012 Zbl 0479.14018

[Rohlin 1952] V. A. Rohlin, "New results in the theory of four-dimensional manifolds", *Doklady Akad. Nauk SSSR* (*N.S.*) **84** (1952), 221–224. In Russian. MR 14,573b Zbl 0046.40702

[Stipsicz 1998] A. I. Stipsicz, "Simply connected 4-manifolds near the Bogomolov–Miyaoka–Yau line", *Math. Res. Lett.* **5**:6 (1998), 723–730. MR 2000h:57047 Zbl 0947.57032

[Stipsicz 1999] A. I. Stipsicz, "Simply connected symplectic 4-manifolds with positive signature", *Turkish J. Math.* **23**:1 (1999), 145–150. MR 2001e:57029 Zbl 0945.57010

[Stipsicz 2000] A. I. Stipsicz, "Chern numbers of certain Lefschetz fibrations", *Proc. Amer. Math. Soc.* **128**:6 (2000), 1845–1851. MR 2000j:57062 Zbl 0982.57012

[Szabó 1998] Z. Szabó, "Simply-connected irreducible 4-manifolds with no symplectic structures", *Invent. Math.* **132**:3 (1998), 457–466. MR 99f:57033 Zbl 0906.57014

[Takamura 2004] S. Takamura, "Towards the classification of atoms of degenerations, I: Splitting criteria via configurations of singular fibers", *J. Math. Soc. Japan* **56**:1 (2004), 115–145. MR 2005c:14009 Zbl 1054.14016

[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

[Taubes 1996] C. H. Taubes, "SW $\Rightarrow$ Gr: From the Seiberg–Witten equations to pseudo-holomorphic curves", *J. Amer. Math. Soc.* **9**:3 (1996), 845–918. MR 97a:57033 Zbl 0867.53025

[Usher 2006] M. Usher, "Minimality and symplectic sums", *Int. Math. Res. Not.* **2006**:16 (2006), Art. ID 49857, 17. MR 2007h:53139 Zbl 1110.57017

[Witten 1994] E. Witten, "Monopoles and four-manifolds", *Math. Res. Lett.* **1**:6 (1994), 769–796. MR 96d:57035 Zbl 0867.57029

Received March 6, 2011. Revised January 9, 2013.

ANAR AKHMEDOV
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES
akhmedov@math.umn.edu

MARK C. HUGHES
DEPARTMENT OF MATHEMATICS
STONY BROOK UNIVERSITY
STONY BROOK, NY 11794-3651
UNITED STATES
hughes@math.sunysb.edu

B. DOUG PARK
DEPARTMENT OF PURE MATHEMATICS
UNIVERSITY OF WATERLOO
WATERLOO, ON
N2L 3G1
CANADA
bdpark@uwaterloo.ca

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.283

SCHUR–HORN THEOREMS IN II_∞-FACTORS

MARTÍN ARGERAMI AND PEDRO MASSEY

We describe majorization between selfadjoint operators in a σ-finite II_∞ factor $(\mathcal{M}, \tau)$ in terms of simple spectral relations. For a diffuse abelian von Neumann subalgebra $\mathcal{A} \subset \mathcal{M}$ that admits a (necessarily unique) trace-preserving conditional expectation, denoted by $E_{\mathcal{A}}$, we characterize the closure in the measure topology of the image through $E_{\mathcal{A}}$ of the unitary orbit of a selfadjoint operator in $\mathcal{M}$ in terms of majorization (i.e., a Schur–Horn theorem). We also obtain similar results for the contractive orbit of positive operators in $\mathcal{M}$ and for the unitary and contractive orbits of τ-integrable operators in $\mathcal{M}$.

1. Introduction

Given two vectors $x, y \in \mathbb{R}^n$, we say that x is *majorized* by y ($x \prec y$) if

$$\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}, \quad k = 1, \dots, n-1; \quad \sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j,$$

where $x^{\downarrow} \in \mathbb{R}^n$ denotes the vector obtained from x by rearranging the entries in nonincreasing order. The first systematic study of the notion of majorization is attributed to Hardy, Littlewood, and Pólya [Hardy et al. 1929]. We refer the reader to [Bhatia 1997] and [Marshall et al. 2011] for further references and properties of majorization. It is well known that (vector) majorization is intimately related with the theory of doubly stochastic matrices. Indeed, $x \prec y$ if and only if $x = Dy$ for some doubly stochastic matrix D; then, as a consequence of Birkhoff's characterization [1946] of the extreme points of the set of doubly stochastic matrices, one can conclude that

$$\{x \in \mathbb{R}^n : x \prec y\} = \operatorname{conv}\{y_\sigma : \sigma \in \mathbb{S}_n\}, \tag{1-1}$$

where $\operatorname{conv}\{y_\sigma : \sigma \in \mathbb{S}_n\}$ denotes the convex hull of the set of vectors y_σ that are obtained from y by rearrangement of its components through permutations $\sigma \in \mathbb{S}_n$.

Argerami was supported in part by the NSERC Discovery Grant Program. Massey was partially supported by PIP 0435 - CONICET and UNLP-11X585.
MSC2010: primary 46L51; secondary 46L10, 52A05, 15A18.
Keywords: II_∞ factors, majorization, Schur–Horn theorem.

It turns out that majorization also characterizes the relation between the spectrum and the diagonal of a selfadjoint matrix. Let $M_n(\mathbb{C})$ denote the algebra of complex $n \times n$ matrices. For $A \in M_n(\mathbb{C})$, let $\operatorname{diag}(A) = (a_{11}, a_{22}, \ldots, a_{nn}) \in \mathbb{C}^n$, and let $\lambda(A) \in \mathbb{C}^n$ be the vector whose coordinates are the eigenvalues of A, counted with multiplicity. I. Schur [1923] proved that for $A \in M_n(\mathbb{C})$ selfadjoint, $\operatorname{diag}(A) \prec \lambda(A)$; while A. Horn [1954] proved the converse: given $x, y \in \mathbb{R}^n$ with $x \prec y$, there exists a selfadjoint matrix $A \in M_n(\mathbb{C})$, with $\operatorname{diag}(A) = x$, $\lambda(A) = y$. For $y \in \mathbb{C}^n$ let $M_y \in M_n(\mathbb{C})$ denote the diagonal matrix with main diagonal y and let $\mathcal{U}_n \subset M_n(\mathbb{C})$ denote the group of unitary matrices. The results from Schur and Horn can then be combined in the following assertion: given $y \in \mathbb{R}^n$,

$$\{x \in \mathbb{R}^n : x \prec y\} = \{\operatorname{diag}(U M_y U^*) : U \in \mathcal{U}_n\}, \tag{1-2}$$

usually known as the Schur–Horn Theorem. The fact that majorization relations imply a family of entropic-like inequalities makes the Schur–Horn theorem an important tool in matrix analysis theory [Bhatia 1997]. It has also been observed that the Schur–Horn theorem plays a crucial role in frame theory [Antezana et al. 2007; Dhillon et al. 2005; Massey and Ruiz 2010].

Majorization in the context of von Neumann algebras has been widely studied (see for instance [Argerami and Massey 2008b; Hiai 1987; 1992; Hiai and Nakamura 1987; Kamei 1983; 1984]). F. Hiai showed several characterizations of majorization in a semifinite von Neumann algebra, including a generalization of (1-1), i.e., a "Birkhoff" theorem. Nevertheless, the lack of the corresponding "Schur–Horn" theorems in the general context of von Neumann factors was only recently observed. Early work on this topic was developed by A. Neumann [1999; 2002] in relation with an extension to infinite dimensions of the linear Kostant convexity theorem in Lie theory.

W. Arveson and R.V. Kadison [2006] conjectured a Schur–Horn theorem in II_1 factors. Although this conjecture remains an open problem, there has been progress on related (but weaker) Schur–Horn theorems in this context [Argerami and Massey 2007; 2008a; 2009]. There has also been significant improvements of Neumann's work on majorization between sequences in $c_0(\mathbb{R}^+)$ due to V. Kaftal and G. Weiss [2008; 2010] because of the relations between infinite dimensional versions of the Schur–Horn theorem (via majorization of bounded structured real sequences) and arithmetic mean ideals (see also [Arveson and Kadison 2006] for improvements in the compact case in $B(H)$).

In this paper we prove versions of the Schur–Horn theorem (i.e., generalizations of (1-2)) in the case of a σ-finite II_∞-factor. These results extend those obtained in [Argerami and Massey 2007; 2008a; Neumann 1999]. Our results are in the vein of Neumann's work, and they are related with a weak version of Arveson and Kadison's scheme for Schur–Horn theorems, but modeled in II_∞ factors. These extensions are formally analogous to the Schur–Horn theorems in [Argerami and Massey 2007; 2008a], but the techniques are more involved in the infinite case. We show that our results are optimal, in the sense that they can not be strengthened for a general selfadjoint operator in a II_∞ factor.

The paper is organized as follows. In Section 2 we develop notation and some basic results on the measure topology and the τ-singular values in von Neumann algebras. Section 3 deals with majorization in $B(H)$, including some results complementing those in [Neumann 1999]. In Section 4 we consider a notion of majorization between selfadjoint operators in a II_∞ factor $(\mathcal{M}, \tau)$ — in line with Neumann's idea — together with several of its basic properties. Although majorization in II_∞ factors is not a new notion [Hiai 1987; 1992], our approach is quite different from the previous presentations. In Section 5 we state and prove the generalizations of the Schur–Horn theorem in II_∞ factors. Our strategy is to reduce the problem to a discrete version, where we can apply the Schur–Horn theorems developed in Section 3 for $B(H)$. We then proceed to show that Hiai's notion of majorization in terms of Choquet's theory of comparison of measures [Hiai 1992] coincides with ours. We finally consider similar results for the contractive orbit of a positive operator and for the unitary and contractive orbits of bounded τ-measurable operators.

2. Preliminaries

Let $(\mathcal{M}, \tau)$ be a σ-finite, semifinite, diffuse von Neumann algebra. The real subspace of selfadjoint elements in $\mathcal{M}$ is denoted by $\mathcal{M}^{sa}$; the group of unitary operators by $\mathcal{U}_{\mathcal{M}}$; and the set of selfadjoint projections by $\mathcal{P}(\mathcal{M})$. Given $p \in \mathcal{P}(\mathcal{M})$, we use the notation $p^\perp = I - p$. For any $a \in \mathcal{M}^{sa}$ and any Borel set $\Delta \subset \mathbb{R}$, $p^a(\Delta) \in \mathcal{P}(\mathcal{M})$ denotes the spectral projection of a corresponding to Δ.

T. Fack [1982] considered in $\mathcal{M}$ the ideals $\mathcal{F}(\mathcal{M}) = \{x \in \mathcal{M} : \tau(\operatorname{supp} x^*) < \infty\}$ — the *τ-finite rank operators* — and $\mathcal{K}(\mathcal{M}) = \overline{\mathcal{F}(\mathcal{M})}$, the ideal of *$\tau$-compact operators*. The quotient C*-algebra $\mathcal{M}/\mathcal{K}(\mathcal{M})$ is called the generalized Calkin algebra. The *essential spectrum* of x — denoted $\sigma_e(x)$ — is the spectrum of $x + \mathcal{K}(\mathcal{M})$ as an element of $\mathcal{M}/\mathcal{K}(\mathcal{M})$. The complement of $\sigma_e(x)$ within $\sigma(x)$ is the *discrete spectrum* $\sigma_d(x)$ of x. As shown in [Hiai 1992], for $x \in \mathcal{M}^{sa}$,

$$\sigma_e(x) = \{t \in \sigma(x) : \tau(p^x(t - \varepsilon, t + \varepsilon)) = \infty \text{ for all } \varepsilon > 0\}.$$

It follows from the previous definitions that $x \in \mathcal{M}^{sa}$ is τ-compact if and only if $\sigma_e(x) = \{0\}$.

We consider in $\mathcal{M}$ the *measure topology* $\mathcal{T}$, which is the linear topology given by the neighborhoods of $0 \in \mathcal{M}$,

$$V(\varepsilon, \delta) = \{r \in \mathcal{M} : \text{there exists } p \in \mathcal{P}(\mathcal{M}) \text{ such that } \|rp\| < \varepsilon,\ \tau(p^\perp) < \delta\},$$

where $\varepsilon, \delta > 0$. For a II_1 factor, $\mathcal{T}$ reduces to the σ-strong topology on bounded sets, while in a type I_∞ factor it reduces to the norm topology.

Definition 2.1. The *upper spectral scale* of $b \in \mathcal{M}^{\mathrm{sa}}$ is the nonincreasing right-continuous real function

$$\lambda_t(b) = \min\{s \in \mathbb{R} : \tau(p^b(s,\infty)) \le t\}, \quad t \in [0,\infty).$$

The *lower spectral scale* of b is the nondecreasing right-continuous function

$$\mu_t(b) = -\lambda_t(-b) = \max\{s \in \mathbb{R} : \tau(p^b(-\infty,s)) \le t\}, \quad t \in [0,\infty).$$

A direct consequence of these definitions is that $\lambda_t(b), \mu_t(b) \in \sigma(b)$ for every $t \in \mathbb{R}^+$. The function $t \mapsto \lambda_t(b)$ is the analogue of the rearrangement of the eigenvalues (in nonincreasing order and counting multiplicities) of a self-adjoint matrix.

For $x \in \mathcal{M}$ we can consider the τ*-singular values* of x given by $\nu_t(x) = \lambda_t(|x|)$, $t \in [0,\infty)$. The spectral scale and τ-singular values have been extensively studied [Fack 1982; Fack and Kosaki 1986; Hiai and Nakamura 1987; Kadison 2004; Petz 1985] in the broader context of τ-measurable operators affiliated to $(\mathcal{M},\tau)$.

The elements of $\mathcal{K}(\mathcal{M})$ can be described in terms of τ-singular values. Indeed, $x \in \mathcal{M}$ is τ-compact if and only if $\lim_{t\to\infty} \nu_t(x) = 0$ [Hiai 1987]. We will make frequent use of the fact that (since $\mathcal{M}$ is diffuse) a given τ-compact $x \in \mathcal{M}^+$ admits a complete flag, i.e., an increasing assignment $\mathbb{R}^+ \ni t \mapsto e(t) \in \mathcal{P}(\mathcal{M})$ such that $\tau(e(t)) = t$, and

$$x = \int_0^\infty \lambda_t(x)\, de(t). \tag{2-1}$$

Unlike the finite case [Argerami and Massey 2007], the equality in (2-1) does not hold for arbitrary τ-compact selfadjoint operators in $\mathcal{M}$. This is possibly one of the reasons why majorization has been considered mainly between positive operators in the semifinite algebras (see the remarks at the end of [Hiai 1987]). We shall overcome this issue by considering both the upper and lower spectral scale, as done in [Neumann 1999] in the case of separable I_∞ factors.

The following fact is used in [Hiai 1992] (in the context of possibly unbounded operators) but we do not know of an explicit proof in the literature. For $x \in \mathcal{M}$, we denote its usual one-norm or trace norm in $(\mathcal{M},\tau)$ by $\|x\|_1 = \tau(|x|) \in [0,\infty]$.

Proposition 2.2. *Let $(\mathcal{M},\tau)$ be a semifinite von Neumann algebra. For $s > 0$ let $\|\cdot\|_{(s)}$ be the norm given by*

$$\|x\|_{(s)} = \inf\{\|x_1\|_1 + s\|x_2\| : x = x_1 + x_2,\ x_1, x_2 \in \mathcal{M}\}, \quad x \in \mathcal{M}.$$

Then $\|x\|_{(s)} = \int_0^s \nu_t(x)\, dt$, and the topology induced by $\|\cdot\|_{(s)}$ agrees with the measure topology on bounded sets.

Proof. The equality $\|x\|_{(s)} = \int_0^s \nu_t(x)\,dt$ is proven in [Fack and Kosaki 1986] in the argument after Theorem 4.4. We now show that the topology induced by $\|\cdot\|_{(s)}$ and the measure topology agree on bounded sets. Indeed, if $0 < s \le r$ then there exists $k \in \mathbb{N}$ such that $r \le ks$ and therefore $\|x\|_{(s)} \le \|x\|_{(r)} \le k\|x\|_{(s)}$, since $t \mapsto \nu_t(x)$ is a nonincreasing function. This shows that the norms $\|\cdot\|_{(s)}$, for $s > 0$, are all equivalent and induce the same topology. Hence we can assume without loss of generality that $s = 1$.

If $\|x\|_{(1)} < d$, then $\int_0^1 \nu_t(x)\,dt < d$. Using that $\nu_t(x)$ is nonincreasing, there exists t_0 with $0 < t_0 < \sqrt{d}$ such that $\nu_{t_0}(x) < \sqrt{d}$. By [Fack and Kosaki 1986, Proposition 2.2],

$$\nu_{t_0}(x) = \inf\{\|xq\| : \tau(q^\perp) \le t_0\}, \tag{2-2}$$

so there is a projection $q \in \mathcal{P}(\mathcal{M})$ such that $\|xq\| < \sqrt{d}$ and $\tau(q^\perp) < \sqrt{d}$; that is, $x \in V(\sqrt{d}, \sqrt{d})$.

Conversely, if $x \in V(\varepsilon, \delta)$ and $\|x\| \le k$, there exists a projection $q \in \mathcal{P}(\mathcal{M})$ such that $\|xq\| < \varepsilon$, $\tau(q^\perp) < \delta$. Since $x = xq^\perp + xq$,

$$\|x\|_{(1)} \le \|xq^\perp\|_1 + \|xq\| \le k\delta + \varepsilon;$$

that is, $V(\varepsilon, \delta) \cap \{x \in \mathcal{M} : \|x\| \le k\} \subset \{x \in \mathcal{M} : \|x\|_{(1)} \le k\delta + \varepsilon\}$. □

Corollary 2.3. *Let $\mathcal{N}$ be a II_1-factor with trace $\tau_{\mathcal{N}}$, and let $\{x_j\}$ be a bounded net. Then $x_j \xrightarrow{\|\cdot\|_1} x$ if and only if $x_j \xrightarrow{\mathcal{T}} x$.*

Proof. For any $x \in \mathcal{N}^{\mathrm{sa}}$ we have $\|x\|_1 = \tau_{\mathcal{N}}(|x|) = \int_0^1 \nu_t(x)\,ds$. Then $\|\cdot\|_1 = \|\cdot\|_{(1)}$ and Proposition 2.2 yields the result. □

We will often and without mention make use of the following properties of the measure topology.

Corollary 2.4. *Let $\mathcal{A} \subset \mathcal{M}$ be a von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathcal{A}}$. Let $\{x_j\} \subset \mathcal{M}^{\mathrm{sa}}$ satisfy $x_j \xrightarrow{\mathcal{T}} x$, and let $\alpha, \beta \in \mathbb{R}$ with $\alpha I \le x_j \le \beta I$ for every j. Then:*

(i) *$x \in \mathcal{M}^{\mathrm{sa}}$ and $\alpha \le x \le \beta$.*

(ii) *$E_{\mathcal{A}}(x_j) \xrightarrow{\mathcal{T}} E_{\mathcal{A}}(x)$.*

Proof. In order to prove (i) first notice that if $x_j \xrightarrow{\mathcal{T}} x$ with $x_j \ge 0$ for every j then $x \in \mathcal{M}^{\mathrm{sa}}$; indeed, this follows from the facts that the operation of taking adjoint is continuous in the measure topology and that this topology is Hausdorff. If $x \notin \mathcal{M}^+$, there exists a nonzero projection $q \in \mathcal{M}$ and $k \in \mathbb{R}^+$ such that $qxq \le (-k)q$. By replacing q by a smaller projection if necessary, we may assume that $\tau(q) < \infty$. We have $qx_jq \xrightarrow{\mathcal{T}} qxq$, so for j big enough there exists a projection p such that

$\|(qxq - qx_jq)p\| < k/3$ and $\tau(p^\perp) < \tau(q)/2$. Then $pqp \neq 0$, since

$$\tau(pqp) = \tau(pq) = \tau(q) - \tau(p^\perp q) \geq \tau(q) - \tau(q)/2 = \tau(q)/2 > 0.$$

We also get from above that $\tau(q) \leq 2\tau(pqp)$. But then $\tau(pq(x_j - x)qp) = \tau(q[q(x_j - x)qp]) \leq \frac{1}{3}k\tau(q)$, so

$$\begin{aligned} 0 \leq \tau(pqx_jqp) &= \tau(pqxqp) + \tau(pq(x_j - x)qp) \leq (-k)\tau(pqp) + \tfrac{1}{3}k\tau(q) \\ &\leq (-k)\tau(pqp) + \tfrac{2}{3}k\tau(pqp) = -\tfrac{1}{3}k\tau(pqp) < 0, \end{aligned}$$

a contradiction. This shows that $x \geq 0$. By linearity we get that if $x_j \xrightarrow{\mathcal{T}} x$ and $\alpha \leq x_j \leq \beta$ then $\alpha \leq x \leq \beta$.

Item (ii) follows from the fact that $E_{\mathcal{A}}$ is contractive with respect to $\|\cdot\|_{(1)}$ together with Proposition 2.2. Indeed, it is well known that $\|E_{\mathcal{A}}(x)\| \leq \|x\|$ for $x \in \mathcal{M}$. Using that $\tau(E_{\mathcal{A}}(x)y) = \tau(xE_{\mathcal{A}}(y)) \leq \|E_{\mathcal{A}}(y)\|\tau(|x|)$ we get

$$\|E_{\mathcal{A}}(x)\|_1 = \sup\{|\tau(E_{\mathcal{A}}(x)y)| : y \in \mathcal{M}, \|y\| \leq 1\} \leq \|x\|_1.$$

For any decomposition $x = y + z$, since $E_{\mathcal{A}}(x) = E_{\mathcal{A}}(y) + E_{\mathcal{A}}(z)$,

$$\|E_{\mathcal{A}}(x)\|_{(1)} \leq \|E_{\mathcal{A}}(y)\|_1 + \|E_{\mathcal{A}}(z)\| \leq \|y\|_1 + \|z\|.$$

So, by Proposition 2.2, $\|E_{\mathcal{A}}(x)\|_{(1)} \leq \|x\|_{(1)}$ for all $x \in \mathcal{M}$, and so $E_{\mathcal{A}}$ is $\mathcal{T}$-continuous. □

3. Majorization in $\ell^\infty(\mathbb{N})$ and $B(H)$ revisited

Let H be a complex separable Hilbert space. In this section we revise and complement A. Neumann's [1999] theory on majorization between self-adjoint operators in $B(H)$. These results will play a key role in our proof of the Schur–Horn theorem in II_∞-factors (Theorem 5.5). For conceptual and notational convenience, we shall follow the exposition in [Antezana et al. 2007] (see also [Kadison 2004]).

In $B(H)$ we consider the canonical trace Tr. We write $\mathcal{U}(H)$ for the group of unitary operators in H, and $\mathcal{C}(H)$ for the semigroup of contractive operators in $B(H)$, i.e.,

$$\mathcal{C}(H) = \{v \in B(H) : v^*v \leq I\}.$$

For $k \in \mathbb{N}$, let $\mathcal{P}_k$ be the set of orthogonal projections $p \in B(H)$ such that $\mathrm{Tr}(p) = k$. For $b \in B(H)^{\mathrm{sa}}$, $k \in \mathbb{N}$, we consider

$$U_k(b) = \sup_{p \in \mathcal{P}_k} \mathrm{Tr}(bp), \quad \text{and} \quad L_k(b) = \inf_{p \in \mathcal{P}_k} \mathrm{Tr}(bp). \tag{3-1}$$

For each $k \in \mathbb{N}$, both $b \mapsto U_k(b)$ and $b \mapsto L_k(b)$ are norm-continuous in $B(H)$, with $L_k(b) = -U_k(-b)$. Moreover, $U_k(u^*bu) = U_k(b)$ for every $b \in B(H)^{\mathrm{sa}}$, $u \in \mathcal{U}(H)$.

Following [Neumann 1999] (but with a different notation) we define, for $f \in \ell^\infty(\mathbb{N})$ and $k \in \mathbb{N}$,

$$U_k(f) = \sup\Big\{\sum_{j\in K} f_j : |K| = k\Big\}, \quad L_k(f) = \inf\Big\{\sum_{j\in K} f_j : |K| = k\Big\}. \tag{3-2}$$

Again, for each $k \in \mathbb{N}$, $L_k(f) = -U_k(-f)$. The similarity of the notations in (3-1) and (3-2) is justified by the following fact: if $b \in B(\mathcal{H})$ is selfadjoint and there exists an orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ of H and $f = (f_i)_{i\in\mathbb{N}} \in \ell^\infty_{\mathbb{R}}(\mathbb{N})$ such that $be_i = f_i e_i$, $i \in \mathbb{N}$ (i.e., if b is diagonal), then by [Antezana et al. 2007, Proposition 3.3]

$$U_k(b) = U_k(f), \quad L_k(b) = L_k(f), \quad k \in \mathbb{N}. \tag{3-3}$$

Definition 3.1 (operator majorization in $B(H)$ [Antezana et al. 2007]). Let a, $b \in B(H)^{\mathrm{sa}}$.

(i) We say that a is *submajorized* by b, and write $a \prec_w b$, if $U_k(a) \le U_k(b)$ for every $k \in \mathbb{N}$.

(ii) We say that a is *majorized* by b, and write $a \prec b$, if $a \prec_w b$ and $L_k(a) \ge L_k(b)$ for every $k \in \mathbb{N}$.

We will also use the notion of vector majorization in $\ell^\infty_{\mathbb{R}}(\mathbb{N})$ (used implicitly in [Neumann 1999]) as follows:

Definition 3.2 (vector majorization in $\ell^\infty_{\mathbb{R}}(\mathbb{N})$). Let f, $g \in \ell^\infty_{\mathbb{R}}(\mathbb{N})$.

(i) We say that f is *submajorized* by g, and write $f \prec_w g$, if $U_k(f) \le U_k(g)$ for every $k \in \mathbb{N}$.

(ii) We say that f is *majorized* by g, and write $f \prec g$, if $f \prec_w g$ and $L_k(f) \ge L_k(g)$ for every $k \in \mathbb{N}$.

We fix an orthonormal basis $\mathcal{B} = \{e_i\}_{i\in\mathbb{N}}$ on H, with associated system of matrix units $\{e_{ij}\}_{i,j\in\mathbb{N}}$ in $B(H)$. For each $f \in \ell^\infty(\mathbb{N})$ we denote by $M_f \in B(H)$ the induced diagonal operator with respect to $\mathcal{B}$, i.e., $M_f = \sum_{i\in\mathbb{N}} f_i e_{ii}$. By (3-3), it is immediate that for all $f, g \in \ell^\infty_{\mathbb{R}}(\mathbb{N})$,

$$M_f \prec M_g \iff f \prec g, \quad M_f \prec_w M_g \iff f \prec_w g. \tag{3-4}$$

We denote by $P_D : B(H) \to B(H)$ the trace preserving conditional expectation onto the (discrete) diagonal masa with respect to the fixed orthonormal basis. Explicitly, for each $x \in B(H)$,

$$P_D(x) = \sum_i e_{ii} x e_{ii} = \sum_i f_i e_{ii} = M_f, \quad \text{where } f_i = \langle xe_i, e_i\rangle,\ i \in \mathbb{N}.$$

The next theorem is a combination of Theorems 2.18 and 3.13 of [Neumann 1999]. Although Neumann phrases the result in terms of vectors in $\ell^\infty_{\mathbb{R}}(\mathbb{N})$, we phrase it in terms of operators in $B(H)$, as in [Antezana et al. 2007, Theorem 3.10].

Theorem 3.3 (A Schur–Horn theorem for $B(H)$). *Let H be a separable complex Hilbert space and let P_D denote the unique trace preserving conditional expectation onto the discrete masa of diagonal operators with respect to the orthonormal basis $\mathcal{B}$ of H. Then, for $b \in B(H)^{\mathrm{sa}}$,*

$$\overline{\{P_D(ubu^*) : u \in \mathcal{U}(H)\}}^{\|\,\|} = \{M_f : f \in \ell^\infty_{\mathbb{R}}(\mathbb{N}),\, M_f \prec b\}.$$

As a consequence of Theorem 3.3 and (3-4) we recover Neumann's result for majorization in $\ell^\infty_{\mathbb{R}}(\mathbb{N})$ which states that, for $f, g \in \ell^\infty_{\mathbb{R}}(\mathbb{N})$,

$$M_f \in \overline{\{P_D(uM_gu^*) : u \in \mathcal{U}(H)\}}^{\|\,\|} \quad \text{if and only if} \quad f \prec g. \tag{3-5}$$

In the rest of this section we will develop a contractive version of Theorem 3.3 for positive operators of $B(H)$ (Theorem 3.7). We will need a few preliminary results.

A proof of the following elementary inequality can be found in [Kadison 2004, Lemma 24].

Lemma 3.4. *Let $y_1 \geq y_2 \geq \cdots$ be positive real numbers and $\alpha_1, \alpha_2, \ldots \in [0, 1]$ with $\sum_{j=1}^\infty \alpha_j \leq k$. Then*

$$\sum_{j=1}^\infty \alpha_j y_j \leq \sum_{j=1}^k y_j. \tag{3-6}$$

Lemma 3.5. *For any $g \in \ell^\infty(\mathbb{N})^+$, $k \in \mathbb{N}$ we have*

$$U_k(g) = \sup\{\mathrm{Tr}(M_g x) : x \in \mathcal{C}(H)^+,\, \mathrm{Tr}(x) \leq k\}.$$

Proof. The inequality "$\leq$" is clear by (3-1) and (3-3). To prove the reverse inequality, fix $k \in \mathbb{N}$, let $\varepsilon > 0$, and fix $x \in \mathcal{C}(H)^+$ with $\mathrm{Tr}(x) \leq k$. As x is a compact and positive contraction, $x = \sum_j \gamma_j h_j$, where $\{h_j\}_j$ is a pairwise-orthogonal family of rank-one projections, $0 \leq \gamma_j \leq 1$ for all j, and $\sum_j \gamma_j \leq k$. We also have that $M_g = \sum_i g_i e_{ii}$, where $\{e_{ii}\}_i$ is the pairwise-orthogonal family of rank-one projections associated with the canonical basis $\mathcal{B}$. Let $\beta = \limsup_n g_n = \max \sigma_{\mathrm{e}}(M_g)$ and define $g' \in \ell^\infty(\mathbb{N})$ by

$$g'_i = \begin{cases} g_i & \text{if } g_i \geq \beta + \varepsilon, \\ \beta & \text{otherwise.} \end{cases}$$

Using [Neumann 1999, Lemma 2.17] it is readily seen that $|U_k(g') - U_k(g)| < k\varepsilon$. Notice that the set $D = \{i : g'_i > \beta\}$ is finite. So there is a unitary $u \in \mathcal{U}(H)$ (induced by an appropriate permutation) such that g'' given by $M_{g''} = uM_{g'}u^*$ satisfies $g''_1 \geq g''_2 \geq \cdots \geq g''_m$, where $m = |D|$, and $g''_i = \beta$ if $i > m$. For each $j \in \mathbb{N}$, let $h'_j = u^* h_j u$; then $\{h'_j\}_j$ is another family of pairwise orthogonal rank-one projections with sum I. We have

$$\sum_i \Bigl(\sum_j \gamma_j \operatorname{Tr}(e_{ii} h'_j)\Bigr) = \sum_j \gamma_j \operatorname{Tr}(h'_j) = \sum_j \gamma_j \le k$$

and

$$0 \le \sum_j \gamma_j \operatorname{Tr}(e_{ii} h'_j) \le \sum_j \operatorname{Tr}(e_{ii} h'_j) = \operatorname{Tr}(e_{ii}) = 1.$$

Since $x \ge 0$ and $g \le g'$,

$$\operatorname{Tr}(M_g x) \le \operatorname{Tr}(M_{g'} x) = \operatorname{Tr}(M_{g''} u^* x u) = \sum_i g''_i \Bigl(\sum_j \gamma_j \operatorname{Tr}(e_{ii} h'_j)\Bigr). \tag{3-7}$$

Now, starting from (3-7) and applying the inequality (3-6) to the numbers $g''_1 \ge g''_2 \ge \cdots \ge 0$ and $\{\sum_j \gamma_j \operatorname{Tr}(e_{ii} h_j)\}_i$, we get

$$\begin{aligned}\operatorname{Tr}(M_g x) &\le \sum_i g''_i \Bigl(\sum_j \gamma_j \operatorname{Tr}(e_{ii} h'_j)\Bigr) \le \sum_{i=1}^k g''_i \\ &= U_k(g'') = U_k(g') < U_k(g) + \varepsilon k.\end{aligned}$$

As ε and x were arbitrary, we have proven the reverse inequality. □

Remark 3.6. Two operators $a, b \in B(H)$ are said to be *approximately unitarily equivalent* if there exists a sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}(H)$ such that

$$\lim_{n\to\infty} \|a - u_n b u_n^*\| = 0.$$

This equivalence is well-known to operator theorists and operator algebraists. As a consequence of the Weyl – von Neumann theorem, it follows from the proof of Theorem II.4.4 of [Davidson 1996] that $a, b \in B(H)^{\mathrm{sa}}$ are approximately unitarily equivalent if and only if their essential spectra (with respect to the classical Calkin algebra) coincide and $\dim\ker(a - \lambda I) = \dim\ker(b - \lambda I)$ for every λ that is not in the essential spectrum of these operators. From this it can be deduced, again as in the proof of the result just cited, that for every $b \in B(H)^+$ and every orthonormal basis $\mathcal{B}$ of H, there exists $M_g \in B(H)^+$ — diagonal with respect to $\mathcal{B}$ — that is approximately unitarily equivalent to b.

The following is the main result of this section.

Theorem 3.7 (A contractive Schur–Horn theorem for $B(H)$). *Let H be a separable complex Hilbert space and let P_D denote the unique trace preserving conditional expectation onto the discrete masa of diagonal operators with respect to the orthonormal basis $\mathcal{B}$ of H. Then, for $b \in B(H)^+$,*

$$\overline{\{P_D(vbv^*) : v \in \mathcal{C}(H)\}}^{\|\,\|} = \{M_f : f \in \ell^\infty(\mathbb{N})^+,\ M_f \prec_w b\}.$$

Proof. We first consider a reduction to the case where b is diagonalizable with respect to the orthonormal basis $\mathcal{B}$. Indeed, by Remark 3.6 there exists $g \in \ell^\infty(\mathbb{N})^+$ such that b and M_g are approximately unitarily equivalent. It is then straightforward to see that

$$\overline{\{vbv^* : v \in \mathscr{C}(H)\}}^{\|\,\|} = \overline{\{vM_gv^* : v \in \mathscr{C}(H)\}}^{\|\,\|},$$

and that

$$\overline{\{P_D(v^*bv) : v \in \mathscr{C}(H)\}}^{\|\,\|} = \overline{\{P_D(v^*M_gv) : v \in \mathscr{C}(H)\}}^{\|\,\|}. \tag{3-8}$$

By (3-3), $U_k(b) = U_k(M_g)$ and $L_k(b) = L_k(M_g)$ for all $k \in \mathbb{N}$. These identities, together with (3-8), imply that — without loss of generality — we can assume that $b = M_g$ for some $g \in \ell^\infty(\mathbb{N})^+$.

Let $v \in \mathscr{C}(H)$ and let $p \in B(H)$ be a projection with $\mathrm{Tr}(p) = k$. Since $vv^* \le I$ and $0 \le P_D(p) \le I$ we have $v^*P_D(p)v \in \mathscr{C}(H)^+$ and $\mathrm{Tr}(v^*P_D(p)v) = \mathrm{Tr}(P_D(p)^{1/2}vv^*P_D(p)^{1/2}) \le \mathrm{Tr}(P_D(p)) = k$. Put $M_f = P_D(vM_gv^*)$. Then

$$\begin{aligned} U_k(M_f) &= \sup\{\mathrm{Tr}(P_D(vM_gv^*)p) : \mathrm{Tr}(p) = k\} \\ &= \sup\{\mathrm{Tr}((vM_gv^*)P_D(p)) : \mathrm{Tr}(p) = k\} \\ &= \sup\{\mathrm{Tr}(M_g(v^*P_D(p)v)) : \mathrm{Tr}(p) = k\} \le U_k(M_g), \end{aligned}$$

where in the last inequality we are using Lemma 3.5 and the fact that $v^*P_D(p)v \in \mathscr{C}(H)^+$. Thus, $M_f \prec_w M_g$ and, as $U_k(\cdot)$ is norm-continuous for every $k \in \mathbb{N}$, we get the inclusion "$\subset$".

For the reverse inclusion, assume that $M_f \prec_w M_g$ (i.e., $f \prec_w g$) and let $\varepsilon > 0$. We follow the idea of the proof of [Bhatia 1997, Theorem II.2.8]. Consider $f', g' \in \ell^\infty(\mathbb{N}) \oplus \ell^\infty(\mathbb{N})$, given by

$$f' = (f + \varepsilon e) \oplus \varepsilon e, \quad g' = (g + \varepsilon e) \oplus 0.$$

where $e \in \ell^\infty(\mathbb{N})$ is the identity. Note that $\|f \oplus 0 - f'\|_\infty, \|g \oplus 0 - g'\|_\infty < \varepsilon$. Since $f, g \ge 0$, we have $U_k(f') = U_k(f) + k\varepsilon$, $U_k(g') = U_k(g) + k\varepsilon$, $L_k(f') = k\varepsilon$, $L_k(g') = 0$, for all $k \in \mathbb{N}$. Hence we have $f' \prec g'$. By Theorem 3.3, there exists a unitary operator $u \in B(H \oplus H)$ such that

$$\|M_{f'} - P_{D\oplus D}(uM_{g'}u^*)\| < \varepsilon. \tag{3-9}$$

We have

$$\|M_{g\oplus 0} - M_{g'}\| < \varepsilon, \quad \|M_{f\oplus 0} - M_{f'}\| < \varepsilon. \tag{3-10}$$

Now let $q = I \oplus 0 \in B(H \oplus H)$, and let $c = quq$ (clearly a contraction), seen as an operator in $B(H)$. Then, as $qP_{D\oplus D} = P_D \oplus 0$ and $qM_{f\oplus 0} = qM_{f\oplus 0}q = M_{f\oplus 0}$, we can use (3-9) and (3-10) to get

$$\begin{aligned}\|M_f - P_D(cM_gc^*)\| &= \|q(M_{f\oplus 0} - P_{D\oplus D}(uM_{g\oplus 0}u^*))q\| \\ &\le \|M_{f\oplus 0} - P_{D\oplus D}(uM_{g\oplus 0}u^*)\| \\ &< 2\varepsilon + \|M_{f'} - P_{D\oplus D}(uM_{g'}u^*)\| < 3\varepsilon.\end{aligned}$$

As ε was arbitrary, we conclude that $M_f \in \overline{\{P_D(v^*M_gv) : v \in \mathcal{C}(H)\}}^{\|\ \|}$. □

Remark 3.8. The positivity assumption in Theorem 3.7 is not just a technicality: even in dimension one we have $-1 \prec_w 0$, and $\{v0v^* : |v| \le 1\} = \{0\}$.

As a consequence of Theorem 3.7 we get that, for $f, g \in \ell^\infty(\mathbb{N})^+$,

$$M_f \in \overline{\{P_D(vM_gv^*) : v \in \mathcal{C}(H)\}}^{\|\ \|} \quad \text{if and only if} \quad f \prec_w g. \tag{3-11}$$

4. Majorization in II_∞-factors

Recall that $(\mathcal{M}, \tau)$ denotes a σ-finite and semifinite diffuse von Neumann algebra. Given $a \in \mathcal{M}^{\mathrm{sa}}$, we consider the functions

$$U_t(a) = \int_0^t \lambda_s(a)\, ds \quad \text{and} \quad L_t(a) = \int_0^t \mu_s(a)\, ds, \quad t \in \mathbb{R}^+,$$

where $t \mapsto \lambda_t(a)$ and $t \mapsto \mu_t(a)$ denote the upper and lower spectral scales (Definition 2.1).

Our next goal is to describe the maps $b \mapsto U_t(b)$ and $b \mapsto L_t(b)$ by means of [Fack and Kosaki 1986, Lemma 4.1]. We will make use of the following relation between spectral scales and singular values:

$$\lambda_t(a) = \nu_t(a + \gamma I) - \gamma, \quad \mu_t(a) = \rho - \nu_t(-a + \rho I), \quad a \in \mathcal{M}^{\mathrm{sa}}, \tag{4-1}$$

for any $\gamma, \rho \in \mathbb{R}$ such that $a + \gamma I,\ -a + \rho I \in \mathcal{M}^+$. We will denote by $\mathcal{P}_t(\mathcal{M})$ the set of all projections in $\mathcal{M}$ of trace t, i.e.,

$$\mathcal{P}_t(\mathcal{M}) = \{p \in \mathcal{P}(\mathcal{M}) : \tau(p) = t\}.$$

Since $(\mathcal{M}, \tau)$ is diffuse and semifinite, $\mathcal{P}_t(\mathcal{M}) \ne \varnothing$ for every $t \ge 0$.

Lemma 4.1. *For any* $a \in \mathcal{M}^{\mathrm{sa}}$,

$$U_t(a) = \sup\{\tau(ap) : p \in \mathcal{P}_t(\mathcal{M})\}, \quad L_t(a) = \inf\{\tau(ap) : p \in \mathcal{P}_t(\mathcal{M})\}, \quad t \in \mathbb{R}^+.$$

Proof. The equalities are an immediate consequence of the identities (4-1) together with [Fack and Kosaki 1986, Lemma 4.1] and the fact that, for every $t \in \mathbb{R}^+$,

$$\sup\{\tau(ap) : p \in \mathcal{P}_t(\mathcal{M})\} = \sup\{\tau((a + \gamma I)p) : p \in \mathcal{P}_t(\mathcal{M})\} - \gamma t. \qquad □$$

Remark 4.2. If $a \in \mathcal{K}(\mathcal{M})^+$, then $\mu_t(a^+) = 0$ for $t \in \mathbb{R}^+$. Let $\{e(t)\}_{t\in\mathbb{R}^+} \subset \mathcal{M}$ be a complete flag for a such that $a = \int_0^\infty \lambda_t(a)\, de(t)$ (which exists by the assumptions on $\mathcal{M}$). Then, using [Fack and Kosaki 1986, Proposition 2.7] and (4-1), we have

$$U_t(a) = \int_0^t \lambda_s(a)\, ds = \tau(ae(t)) \quad \text{and} \quad L_t(a) = 0, \quad t \in \mathbb{R}^+.$$

Thus, for a positive τ-compact operator a the supremum in Lemma 4.1 is attained explicitly by means of the projection $e(t)$ in $\mathcal{P}_t(\mathcal{M}) \cap \{a\}'$.

Lemma 4.3. *Let $b \in \mathcal{M}^{sa}$. Then, for each $t \in \mathbb{R}^+$, the functions $b \mapsto U_t(b)$, $b \mapsto L_t(b)$ are $\|\cdot\|_1$-continuous, and they are also $\mathcal{T}$-continuous on bounded sets of $\mathcal{M}^{sa}$.*

Proof. It is enough to prove the statement for $U_t(\cdot)$, since $L_t(b) = -U_t(-b)$. Given $\varepsilon > 0$, by Lemma 4.1 there exists $p \in \mathcal{P}_t(\mathcal{M})$ with $U_t(x) \le \tau(xp) + \varepsilon$. Then

$$U_t(x) - U_t(y) \le \tau(xp) + \varepsilon - \tau(yp) \le \|x - y\|_{(t)} + \varepsilon \le \|x - y\|_1 + \varepsilon,$$

where we used the inequality $\tau((x - y)p) \le \tau(|x - y|p) \le \|x - y\|_{(t)}$ that follows from Lemma 4.1. By letting $\varepsilon \to 0$ and reversing the roles of x and y we conclude the $\mathcal{T}$ and $\|\cdot\|_1$ continuity of $b \mapsto U_t(b)$ on bounded sets, by Proposition 2.2. □

From now on we will specialize $(\mathcal{M}, \tau)$ to be a σ-finite II_∞-factor with faithful normal semifinite tracial weight τ.

We begin by describing the notion of majorization between selfadjoint operators in the II_∞-factor $\mathcal{M}$. In the setting of nonfinite von Neumann algebras, this concept was developed for selfadjoint operators in [Hiai 1992]. Our presentation, inspired by Neumann's work [1999], is fairly different (see Remark 4.5 below).

Definition 4.4. Let $a, b \in \mathcal{M}^{sa}$.

(i) We say that a is *submajorized* by b, and write $a \prec_w b$, if

$$U_t(a) \le U_t(b) \quad \text{for every } t \in \mathbb{R}^+.$$

(ii) We say that a is *majorized* by b, and write $a \prec b$, if $a \prec_w b$ and

$$L_t(a) \ge L_t(b) \quad \text{for every } t \in \mathbb{R}^+.$$

Remark 4.5. If $b \in \mathcal{K}(\mathcal{M})^+$, then $\mu_t(b) = 0$ for all $t \in \mathbb{R}^+$ and therefore $L_t(b) = 0$ for all $t \in \mathbb{R}^+$. Thus, if $a \in \mathcal{M}^+$ and $a \prec_w b$, then $a \prec b$.

For a, $b \in \mathcal{M}^+$, our notion of majorization is strictly stronger than the one considered in [Hiai 1987]. As we have already mentioned, our notion of majorization does coincide with that of [Hiai 1992] for selfadjoint operators in a II_∞-factor (see Corollary 5.7). It is worth pointing out that in [Hiai 1992] majorization is described (for normal operators) in terms of Choquet's theory on comparison of measures, rather than in the simple terms used above: Lemma 4.1 shows that the notion of majorization in a II_∞-factor from Definition 4.4 is an analogue of the notion of operator majorization in $B(H)$ as described in Definition 3.1.

For a fixed $b \in \mathcal{M}^{\mathrm{sa}}$, we write $\Omega_{\mathcal{M}}(b)$ for the set of all elements in $\mathcal{M}^{\mathrm{sa}}$ that are majorized by b, i.e.,

$$\Omega_{\mathcal{M}}(b) = \{a \in \mathcal{M}^{\mathrm{sa}} : a \prec b\}.$$

Proposition 4.6. *Let $b \in \mathcal{M}^{\mathrm{sa}}$. Then $\Omega_{\mathcal{M}}(b)$ is a bounded $\mathcal{T}$-closed convex set that contains the unitary orbit $\mathcal{U}_{\mathcal{M}}(b)$.*

Proof. For any $x \in \mathcal{M}^{\mathrm{sa}}$, the definition of $U_t(x)$ and $L_t(x)$, together with the right-continuity of $\lambda_t(x)$ and $\mu_t(x)$, imply that

$$\lim_{t\to 0^+} \frac{U_t(x)}{t} = \lambda_t(0) = \max \sigma(x) \quad \text{and} \quad \lim_{t\to 0^+} \frac{L_t(x)}{t} = \mu_t(0) = \min \sigma(x).$$

Hence, $a \prec b$ implies $\sigma(a) \subset [\min \sigma(b), \max \sigma(b)]$; in particular $\|a\| \le \|b\|$, so $\Omega_{\mathcal{M}}(b)$ is a bounded set. Lemma 4.3 immediately implies that it is closed in the measure topology. Moreover, if $u \in \mathcal{U}_{\mathcal{M}}$, it is easy to see that $\lambda_t(ubu^*) = \lambda_t(b)$. So $U_t(ubu^*) = U_t(b)$ and, similarly, $L_t(ubu^*) = L_t(b)$. Thus $ubu^* \prec b$, and $\mathcal{U}_{\mathcal{M}}(b) \subset \Omega_{\mathcal{M}}(b)$.

Let $a_1, a_2 \in \mathcal{M}^{\mathrm{sa}}$, $\gamma \in [0, 1]$, with $a_1 \prec b$, $a_2 \prec b$. Using Lemma 4.1,

$$\begin{aligned} U_t(\gamma a_1 + (1-\gamma)a_2) &= \sup\{\tau(p(\gamma a_1 + (1-\gamma)a_2)) : \tau(p) = t\} \\ &= \sup\{\gamma\tau(pa_1) + (1-\gamma)\tau(pa_2) : \tau(p) = t\} \\ &\le \gamma U_t(a_1) + (1-\gamma)U_t(a_2) \le U_t(b). \end{aligned}$$

Similarly,

$$L_t(\gamma a_1 + (1-\gamma)a_2) \ge \gamma L_t(a_1) + (1-\gamma)L_t(a_2) \ge L_t(b),$$

so $\gamma a_1 + (1-\gamma)a_2 \prec b$, and $\Omega_{\mathcal{M}}(b)$ is convex. □

Remark 4.7. Let $b \in \mathcal{M}^{\mathrm{sa}}$. The function $t \mapsto \lambda_t(b)$ is nonincreasing and bounded; therefore the numbers $\lambda^{\mathrm{e}}_{\max}(b) = \lim_{t\to\infty} \lambda_t(b)$ and $\lambda^{\mathrm{e}}_{\min}(b) = \lim_{t\to\infty} \mu_t(b)$ exist. Indeed, we have

$$\lambda^{\mathrm{e}}_{\max}(b) = \max \sigma_{\mathrm{e}}(b) = \lim_{t\to\infty} \frac{U_t(b)}{t}, \quad \lambda^{\mathrm{e}}_{\min}(b) = \min \sigma_{\mathrm{e}}(b) = \lim_{t\to\infty} \frac{L_t(b)}{t}. \tag{4-2}$$

Consider the operators $\bar{b}, \underline{b} \in \mathcal{M}^+$ given by

$$\bar{b} = (b - \lambda^{\mathrm{e}}_{\max}(b)I)^+ \quad \text{and} \quad \underline{b} = (\lambda^{\mathrm{e}}_{\min}(b)I - b)^+. \tag{4-3}$$

Both $\bar{b}, \underline{b}$ are positive τ-compact operators with orthogonal support. It is easy to check that, for all $t \ge 0$, $U_t(b) = U_t(\bar{b}) + t\lambda^{\mathrm{e}}_{\max}(b)$, $L_t(b) = -U_t(\underline{b}) + t\lambda^{\mathrm{e}}_{\min}(b)$, and $L_t(\underline{b}) = L_t(\bar{b}) = 0$. If $a \prec b$ then, by (4-2),

$$\lambda^{\mathrm{e}}_{\min}(b) \le \lambda^{\mathrm{e}}_{\min}(a) \le \lambda^{\mathrm{e}}_{\max}(a) \le \lambda^{\mathrm{e}}_{\max}(b).$$

We finish the section with three lemmas on perturbations to be used later.

Lemma 4.8. *Let $x \in \mathcal{K}(\mathcal{M})^+$, $z \in \mathcal{P}(\mathcal{M})$ infinite with $zx = 0$ and $\varepsilon > 0$. Then there exists $x' \in \mathcal{K}(\mathcal{M})^+$ such that*

(i) *the support of x' contains z;*

(ii) $\|x' - x\| < \varepsilon$;

(iii) $\lambda_t(x') = \lambda_t(x) + \varepsilon/(6+t)$, $t \in [0, \infty)$.

Proof. Since x is τ-compact, there exists $s_0 > 0$ such that $\lambda_{s_0}(x) < \varepsilon/6$. Let $p_1 = p^x(\lambda_{s_0}(x), \infty)$. The τ-compactness of x guarantees that $\tau(p_1) < \infty$.

As x is τ-compact and positive, there exists a complete flag $e_x(t)$ with $x = \int_0^\infty \lambda_t(x)\, de_x(t)$. Note that $p_1 = e_x(s_0)$. Let $e_1(t)$ be a complete flag over z, and define

$$x' = \int_0^{s_0} \left(\lambda_t(x) + \frac{\varepsilon}{6+t}\right) de_x(t) + \int_0^\infty \left(\lambda_{t+s_0}(x) + \frac{\varepsilon}{6+t+s_0}\right) de_1(t).$$

The second term above equals $x' p_1^\perp = x'z$ and its norm is less than $\varepsilon/3$; so

$$\|x - x'\| \le \left\| \int_0^{s_0} \frac{\varepsilon}{6+t}\, de_x(t) \right\| + \|x p_1^\perp\| + \|x' p_1^\perp\| < \frac{\varepsilon}{6} + \frac{\varepsilon}{6} + \frac{\varepsilon}{3} < \varepsilon.$$

It is clear by construction (since $e_x(t)e_1(s) = 0$ for all t, s) that

$$\lambda_t(x') = \lambda_t(x) + \frac{\varepsilon}{6+t}, \qquad t \in [0, \infty),$$

and this implies $x' \in \mathcal{K}(\mathcal{M})$. □

Lemma 4.9. *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse von Neumann subalgebra. Let $a \in \mathcal{A}^{\mathrm{sa}}$, $b \in \mathcal{M}^{\mathrm{sa}}$ with $a \prec b$, and fix $\varepsilon > 0$. Then there exist $a' \in \mathcal{A}^{\mathrm{sa}}$, $b' \in \mathcal{M}^{\mathrm{sa}}$ such that*

(i) $\|a - a'\| < \varepsilon$, $\|b - b'\| < \varepsilon$;

(ii) $a' \prec b'$;

(iii) $\overline{a'}$, $\underline{a'}$, $\overline{b'}$, $\underline{b'}$ (as defined in Remark 4.7) *have infinite support.*

Proof. We first consider a partition of the identity

$$s_1 = p^b\Big[\lambda^{\mathrm{e}}_{\max}(b) + \frac{\varepsilon}{8}, \infty\Big), \quad s_2 = p^b\Big(\lambda^{\mathrm{e}}_{\min}(b) - \frac{\varepsilon}{8}, \lambda^{\mathrm{e}}_{\max}(b) + \frac{\varepsilon}{8}\Big),$$
$$s_3 = p^b\Big(-\infty, \lambda^{\mathrm{e}}_{\min}(b) - \frac{\varepsilon}{8}\Big].$$

The projection s_2 is infinite, while the others may or may not be infinite. We consider a decomposition $s_2 = z_1 + z_2 + z_3$ into three mutually orthogonal infinite projections, such that

$$z_1 \le p^b\Big(\lambda^{\mathrm{e}}_{\max}(b) - \frac{\varepsilon}{8}, \lambda^{\mathrm{e}}_{\max}(b) + \frac{\varepsilon}{8}\Big), \quad z_3 \le p^b\Big(\lambda^{\mathrm{e}}_{\min}(b) - \frac{\varepsilon}{8}, \lambda^{\mathrm{e}}_{\min}(b) + \frac{\varepsilon}{8}\Big).$$

Let $\underline{a}, \bar{a} \in \mathcal{K}(\mathcal{A})^+$ and $\underline{b}, \bar{b} \in \mathcal{K}(\mathcal{M})^+$ be as in (4-3). Apply Lemma 4.8 to $\bar{b}s_1$ with the projection z_1 and to $\underline{b}s_3$ with z_3, to obtain $(\bar{b})', (\underline{b})' \in \mathcal{K}(\mathcal{M})^+$, both with infinite support and such that $\|(\bar{b})' - \bar{b}s_1\| < \varepsilon/4$, $\|(\underline{b})' - \underline{b}s_3\| < \varepsilon/4$. Define

$$b' = \big((\bar{b})' + \lambda^{\mathrm{e}}_{\max}(b)(s_1 + z_1)\big) + (s_2 - z_1 - z_3)b - \big((\underline{b})' - \lambda^{\mathrm{e}}_{\min}(b)(s_3 + z_3)\big).$$

As $b = (\bar{b}s_1 + \lambda^{\mathrm{e}}_{\max}(b)s_1) + bs_2 - (\underline{b}s_3 - \lambda^{\mathrm{e}}_{\min}(b)s_3)$, we get

$$\begin{aligned}\|b' - b\| &\le \|(\bar{b})' - \bar{b}s_1\| + \|\lambda^{\mathrm{e}}_{\max}(b)z_1 - bz_1\| + \|\lambda^{\mathrm{e}}_{\min}(b)z_3 - bz_3\| + \|(\underline{b})' - \underline{b}s_3\| \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

Note that $\lambda^{\mathrm{e}}_{\max}(b') = \lambda^{\mathrm{e}}_{\max}(b)$; then $\overline{b'} = (\bar{b})'$, $\underline{b'} = (\underline{b})'$ have infinite support,

$$\begin{aligned}\lambda_t(b') &= \lambda_t(\overline{b'}) + \lambda^{\mathrm{e}}_{\max}(b') = \lambda_t((\bar{b})') + \lambda^{\mathrm{e}}_{\max}(b) \\ &= \lambda_t(\bar{b}) + \frac{\varepsilon}{6+t} + \lambda^{\mathrm{e}}_{\max}(b) = \lambda_t(b) + \frac{\varepsilon}{6+t}\end{aligned} \tag{4-4}$$

and similarly

$$\mu_t(b') = \mu_t(b) - \frac{\varepsilon}{6+t}.$$

Proceeding with a in the same way we did for b, we obtain $a' \in \mathcal{A}^{\mathrm{sa}}$ with $\|a - a'\| < \varepsilon$, with $\overline{a'}$ and $\underline{a'}$ having infinite support, and such that

$$\lambda_t(a') = \lambda_t(a) + \frac{\varepsilon}{6+t}, \quad \mu_t(a') = \mu_t(a) - \frac{\varepsilon}{6+t}, \quad t \in [0, \infty). \tag{4-5}$$

From (4-4), (4-5), and the fact that $a \prec b$, we deduce that $a' \prec b'$. □

Let $\mathcal{N}$ be a semifinite diffuse von Neumann algebra with fns (faithful, normal, semifinite) trace τ. We consider the set $L^1(\mathcal{N}) \cap \mathcal{N}$, which consists of those $x \in \mathcal{N}$ with $\|x\|_1 < \infty$. The elements in $L^1(\mathcal{N}) \cap \mathcal{N}$ are necessarily compact, since $\int_0^\infty \lambda_t(|x|)\, dt < \infty$ forces $\nu_t(x) = \lambda_t(|x|) \xrightarrow[t\to\infty]{} 0$.

Lemma 4.10. *Let $\mathcal{N}$ be a semifinite diffuse von Neumann algebra with fns trace τ, and let $x \in L^1(\mathcal{N})^{\mathrm{sa}}$, $\varepsilon > 0$. Then there exists $x' \in L^1(\mathcal{N})^{\mathrm{sa}}$ such that*

(i) $\|x' - x\|_1 < \varepsilon$;

(ii) $\lambda_t(x') = \lambda_t(x) + \varepsilon/(10 + 4t^2)$;

(iii) $\mu_t(x') = \mu_t(x) - \varepsilon/(10 + 4t^2)$;

(iv) $\tau(p^{x'}(0, \infty)) = \infty$, $\tau(p^{x'}(-\infty, 0)) = \infty$;

(v) $p^{x'}(-\infty, 0) + p^{x'}(0, \infty) = I$.

Proof. Since x is τ-compact, its essential spectrum contains zero. Then $\lambda_t(x) \geq 0$, $\mu_t(x) \leq 0$ for all t. With that in mind, the proof runs as the proof of Lemma 4.8, using the L^1 property instead of compactness to choose p_1 and considering the positive and negative parts of x separately. □

5. Schur–Horn theorems in II_∞-factors

In this section we prove versions of the Schur–Horn theorem in the σ-finite II_∞-factor $(\mathcal{M}, \tau)$ (Theorems 5.5 and 5.8), in the spirit of Neumann's work [1999]. We also consider versions of these results for τ-integrable operators (Theorems 5.10 and 5.12).

We begin with the following result, which comprises the main technical part of the proof of Theorem 5.5 (by allowing us to reduce the argument to a discrete case). Recall that $V(\varepsilon, \delta)$ denotes the canonical basis of neighborhoods of 0 in the measure topology, indexed by $\varepsilon, \delta > 0$.

Proposition 5.1. *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse von Neumann subalgebra. Let $a \in \mathcal{A}^{\mathrm{sa}}$, $b \in \mathcal{M}^{\mathrm{sa}}$ be such that $a \prec b$ and fix $m \in \mathbb{N}$. Then there exist $\{p_n\}_{n\geq 1} \subset \mathcal{P}(\mathcal{A})$, $\{q_n\}_{n\geq 1} \subset \mathcal{P}(\mathcal{M})$ such that*

(i) $p_i p_j = q_i q_j = 0$ *for* $i \neq j$;

(ii) $\tau(p_n) = \tau(q_n) = \tau(p_1)$ *for all* $n \in \mathbb{N}$;

(iii) $\tau(1 - \sum_{n\geq 1} p_n) = \tau(1 - \sum_{n\geq 1} q_n) < \frac{1}{m}$;

(iv) *there exist* $f, g \in \ell^\infty_{\mathbb{R}}(\mathbb{N})$ *such that*

   (a) $f \prec g$;

   (b) $\Big(a - \sum_{n\geq 1} f(n) p_n\Big), \Big(b - \sum_{n\geq 1} g(n) q_n\Big) \in V\Big(\frac{1}{m}, \frac{1}{m}\Big)$.

Proof. By Lemma 4.9 there exist $a' \in \mathcal{A}^{\mathrm{sa}}$, $b' \in \mathcal{M}^{\mathrm{sa}}$ with $\|a - a'\| < 1/2m$, $\|b - b'\| < 1/2m$, $a' \prec b'$, and such that $\bar{a}$, $\underline{a}$, $\bar{b}$, $\underline{b}$ (as defined in Remark 4.7) have infinite support. So, at the cost of replacing $1/m$ with $2/m$ in (b) above, we can assume without loss of generality that $\tau(r_1) = \tau(s_1) = \tau(r_3) = \tau(s_3) = \infty$, where $r_1, s_1, r_3, s_3 \in \mathcal{P}(\mathcal{M})$ are as in the proof of Lemma 4.9.

Since $\mathcal{A}$ is diffuse, there exist complete flags $\{e_{\bar{a}}(t)\}_{t\in[0,\infty)}$, $\{e_{\underline{a}}(t)\}_{t\in[0,\infty)}$ in $\mathcal{A}$ over r_1 and r_3 respectively such that $\tau(e_{\bar{a}}(t)) = \tau(e_{\underline{a}}(t)) = t$ for $t \geq 0$ and

$$\bar{a} = \int_0^\infty \lambda_s(\bar{a})\, de_{\bar{a}}(s), \quad \underline{a} = \int_0^\infty \lambda_s(\underline{a})\, de_{\underline{a}}(s).$$

Similarly, there exist complete flags $\{e_{\bar{b}}(t)\}_{t\in[0,\infty)}$, $\{e_{\underline{b}}(t)\}_{t\in[0,\infty)}$ over s_1 and s_3 respectively such that $\tau(e_{\bar{b}}(t)) = \tau(e_{\underline{b}}(t)) = t$ for $t \geq 0$ and

$$\bar{b} = \int_0^\infty \lambda_s(\bar{b})\, de_{\bar{b}}(s), \quad \underline{b} = \int_0^\infty \lambda_s(\underline{b})\, de_{\underline{b}}(s).$$

Let $q_t = I-(e_{\bar b}(t)+e_{\underline{b}}(t))$, $p_t = I-(e_{\bar a}(t)+e_{\underline{a}}(t))$. Then $\{q_t\}$, $\{p_t\}$ are decreasing nets of projections that converge strongly to s_2, r_2 respectively. For the rest of the proof, we will fix $t>0$ big enough so that the following three properties hold (all guaranteed by the fact that $\lambda_t(x)\to 0$ as $t\to\infty$ if $x\in\mathcal{K}(\mathcal{M})$):

$$\Big(\lambda^{\mathrm{e}}_{\min}(b)-\frac{1}{m}\Big)q_t \le bq_t \le \Big(\lambda^{\mathrm{e}}_{\max}(b)+\frac{1}{m}\Big)q_t, \tag{5-1}$$

$$\Big(\lambda^{\mathrm{e}}_{\min}(b)-\frac{1}{m}\Big)p_t \le ap_t \le \Big(\lambda^{\mathrm{e}}_{\max}(b)+\frac{1}{m}\Big)p_t, \tag{5-2}$$

$$\max\{\lambda_t(\bar a),\lambda_t(\bar b),\lambda_t(\underline{a}),\lambda_t(\underline{b})\}<\frac{1}{m}. \tag{5-3}$$

Now apply [Argerami and Massey 2007, Lemma 3.2] and Corollary 2.3 to $ae_{\bar a}(t)$ in the II_1 factor $e_{\bar a}(t)\mathcal{M}e_{\bar a}(t)$ and to $ae_{\underline{a}}(t)$ in the II_1-factor $e_{\underline{a}}(t)\mathcal{M}e_{\underline{a}}(t)$. This way we get $N\in\mathbb{N}$ with $N\ge t\cdot 3m\cdot(2\|b\|m+3)$, partitions $\{p_j\}_{j=1}^N$ and $\{p'_j\}_{j=1}^N$ of $e_{\bar a}(t)$ and $e_{\underline{a}}(t)$ respectively given by

$$p_j=e_{\bar a}\Big(\frac{jt}{N}\Big)-e_{\bar a}\Big(\frac{(j-1)t}{N}\Big),\quad p'_j=e_{\underline{a}}\Big(\frac{jt}{N}\Big)-e_{\underline{a}}\Big(\frac{(j-1)t}{N}\Big),\quad 1\le j\le N,$$

and coefficients $\alpha'_1\ge\alpha'_2\ge\cdots\ge\alpha'_N$, $\alpha''_1\ge\alpha''_2\ge\cdots\ge\alpha''_N$ given by

$$\alpha'_j=\frac{N}{t}\int_{(j-1)t/N}^{jt/N}\lambda_s(ae_{\bar a}(t))\,ds=\frac{N}{t}\tau(ap_j),\quad \alpha''_j=\frac{N}{t}\tau(ap'_j),$$

such that

$$\Big(ae_{\bar a}(t)-\sum_{j=1}^N\alpha'_jp_j\Big),\Big(ae_{\underline{a}}(t)-\sum_{j=1}^N\alpha''_jp'_j\Big)\in V\Big(\frac{1}{m},\frac{1}{2m}\Big) \tag{5-4}$$

(recall that $\|x\|_{(1)}\le\|x\|_1$ and that if $\|x\|_{(1)}<1/4m^2$, then $x\in V(1/2m,1/2m)$; see the proof of Proposition 2.2). Similarly, we obtain for b partitions $\{q_j\}_{j=1}^N$ and $\{q'_j\}_{j=1}^N$ of $e_{\bar b}(t)$ and $e_{\underline{b}}(t)$ respectively such that

$$q_j=e_{\bar b}\Big(\frac{jt}{N}\Big)-e_{\bar b}\Big(\frac{(j-1)t}{N}\Big),\quad q'_j=e_{\underline{b}}\Big(\frac{jt}{N}\Big)-e_{\underline{b}}\Big(\frac{(j-1)t}{N}\Big),\quad 1\le j\le N,$$

and coefficients $\beta'_1\ge\beta'_2\ge\cdots\ge\beta'_N$, $\beta''_1\ge\beta''_2\ge\cdots\ge\beta''_N$ given by

$$\beta'_j=\frac{N}{t}\tau(bq_j),\quad \beta''_j=\frac{N}{t}\tau(bq'_j)$$

with

$$\Big(be_{\bar b}(t)-\sum_{j=1}^N\beta'_jq_j\Big),\Big(be_{\underline{b}}(t)-\sum_{j=1}^N\beta''_jq'_j\Big)\in V\Big(\frac{1}{m},\frac{1}{2m}\Big). \tag{5-5}$$

Consider now a partition $\{I_j\}_{j=1}^L$ of $\left[\lambda^{\mathrm{e}}_{\min}(b)-\frac{1}{m},\lambda^{\mathrm{e}}_{\max}(b)+\frac{1}{m}\right]$ into L consecutive disjoint subintervals with $2\le L\le 2\|b\|m+3$, with $I_1=\left[\lambda^{\mathrm{e}}_{\min}(b)-\frac{1}{m},\lambda^{\mathrm{e}}_{\min}(b)\right)$, $I_L=\left(\lambda^{\mathrm{e}}_{\max}(b),\lambda^{\mathrm{e}}_{\max}(b)+\frac{1}{m}\right]$, and such that the length of each I_j is no greater than $\frac{1}{m}$. Define

$$a_e=p_ta,\quad b_e=q_tb.$$

Let $\gamma_1=\lambda^{\mathrm{e}}_{\min}(b)$, $\gamma_L=\lambda^{\mathrm{e}}_{\max}(b)$, and choose $\gamma_j\in I_j$ for $2\le j\le L-1$. The choice of the γ_j, together with (5-1) and (5-2), imply that

$$\text{(5-6)}\qquad \Big\|a_e-\sum_{j=1}^L\gamma_j\,p^{a_e}(I_j)\Big\|<\frac{1}{m},\quad \Big\|b_e-\sum_{j=1}^L\gamma_j\,p^{b_e}(I_j)\Big\|<\frac{1}{m}.$$

For $j\in\{1,\dots,L\}$ let

$$t_j^a=\begin{cases}\left\lfloor\dfrac{\tau(p^{a_e}(I_j))N}{t}\right\rfloor & \text{if } \tau(p^{a_e}(I_j))<\infty,\\ \infty & \text{if } \tau(p^{a_e}(I_j))=\infty,\end{cases}$$

where $\lfloor x\rfloor$ denotes the integer part of $x\in\mathbb{R}$. We construct $\{t_j^b\}_{j=1}^L$ in the same way. For each j, if $t_j^a=\infty$ we consider a partition

$$\{p_i^{(j)}\}_{i\in\mathbb{N}}\subset\mathcal{P}(\mathcal{A})$$

of $p^{a_e}(I_j)$ with $\tau(p_i^{(j)})=t/N$ for all $i\in\mathbb{N}$; otherwise, if $t_j^a<\infty$, we consider a partition

$$\{p_i^{(j)}\}_{i=1}^{t_j^a+1}\subset\mathcal{P}(\mathcal{A})$$

with $\tau(p_i^{(j)})=t/N$ for $1\le i\le t_j^a$, and $\tau(p^{(j)}_{t_j^a+1})<t/N$.

Analogously, we consider partitions $\{q_i^{(j)}\}_i\subset\mathcal{P}(\mathcal{M})$ of $p^{b_e}(I_j)$ for $1\le j\le L$. Since $\overline{b}$ and $\underline{b}$ have infinite support, we have

$$\text{(5-7)}\qquad t_1^b=t_L^b=\infty,\quad \lambda^{\mathrm{e}}_{\min}(b)\le\min_{1\le j\le L}\gamma_j\le\max_{1\le j\le L}\gamma_j\le\lambda^{\mathrm{e}}_{\max}(b)$$

and there exists $i_0\in\{1,\dots,L\}$ with $t^a_{i_0}=\infty$. And, since $L\le 2\|b\|m+3$ and $N\ge t\cdot 3m\cdot(2\|b\|m+3)$, we have

$$\text{(5-8)}\qquad \sum_{j:t_j^a<\infty}\tau(p^{(j)}_{t_j^a+1})\le\sum_{i=1}^L\frac{t}{N}\le\frac{1}{3m},\quad \sum_{j:t_j^b<\infty}\tau(q^{(j)}_{t_j^b+1})\le\frac{1}{3m}.$$

We can assume that the projections $\sum_{j:t_j^a<\infty}p^{(j)}_{t_j^a+1}$ and $\sum_{j:t_j^b<\infty}q^{(j)}_{t_j^b+1}$ have equal trace; indeed we can take the necessary mass (which will be certainly less than $1/2m$) from one of the projections $p^{a_e}(I_{i_0})$, $p^{b_e}(I_L)$ respectively (since each of them is an infinite projection) before considering the partitions of these projections (this, at the cost of replacing both occurrences of "$< 1/m$" in (5-6) by "$\in V(1/m, 1/2m)$"). From (5-6) and (5-8),

$$\Big(a_e - \sum_{j=1}^{L} \gamma_j \sum_{i=1}^{t_j^a} p_i^{(j)}\Big), \Big(b_e - \sum_{j=1}^{L} \gamma_j \sum_{i=1}^{t_j^b} q_i^{(j)}\Big) \in V\Big(\frac{1}{m}, \frac{1}{m}\Big). \tag{5-9}$$

Let $\{(\alpha_i, p_i)\}_{i\geq 1}$ be an enumeration of the countable set

$$\{(\alpha'_j, p_j) : 1 \leq j \leq N\} \cup \{(\alpha''_j, p'_j) : 1 \leq j \leq N\} \cup \{(\gamma_j, p_i^{(j)}) : 1 \leq j \leq L, 1 \leq i \leq t_j^a\}$$

and let $\{(\beta_i, q_i)\}_{i\geq 1}$ be an enumeration of the countable set

$$\{(\beta'_j, q_j) : 1 \leq j \leq N\} \cup \{(\beta''_j, q'_j) : 1 \leq j \leq N\} \cup \{(\gamma_j, q_i^{(j)}) : 1 \leq j \leq L, 1 \leq i \leq t_j^b\}.$$

By construction, $\{p_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$. It also follows that (i), (ii), and (iii) in the statement of the theorem hold. Moreover, from (5-4), (5-5) and (5-9) we get part (b) of (iv) (with $f = \{\alpha_n\}_{n\geq 1}$, $g = \{\beta_n\}_{n\geq 1}$). It remains to show that $f \prec g$ in the sense of Definition 3.1. We will only prove that $U_k(f) \leq U_k(g)$ for $k \geq 1$, since the L_k inequalities follow in a similar way. We have

$$U_k(g) = \begin{cases} \sum_{i=1}^{k} \beta'_j & \text{if } 1 \leq k \leq N, \\ \sum_{i=1}^{N} \beta'_j + (k-N)\lambda^{\mathrm{e}}_{\max}(b) & \text{if } N < k \end{cases}$$

(recall that $\gamma_L = \lambda^{\mathrm{e}}_{\max}(b)$ and that there is an infinity of γ_L in the list $\{\beta_n\}$). For $U_k(f)$ we get

$$U_k(f) = \begin{cases} \sum_{i=1}^{k} \alpha'_j & \text{if } 1 \leq k \leq N, \\ \sum_{i=1}^{N} \alpha'_j + \sum_{i=N+1}^{k} \gamma_{\sigma(i)} & \text{if } N < k, \end{cases}$$

for appropriate choices $\sigma(i) \in \{1, \dots, L\}$. If $1 \leq k \leq N$, then

$$\begin{aligned} U_k(g) &= \sum_{i=1}^{k} \beta'_i = \frac{N}{t} \int_0^{\frac{kt}{N}} \lambda_s(b)\, ds = \frac{N}{t} U_{kt/N}(b) \\ &\geq \frac{N}{t} U_{kt/N}(a) = \frac{N}{t} \int_0^{\frac{kt}{N}} \lambda_s(a)\, ds = \sum_{i=1}^{k} \alpha'_i = U_k(f). \end{aligned}$$

If $N < k$,

$$\begin{aligned} U_k(g) &= \frac{N}{t} \int_0^{t} \lambda_s(b)\, ds + (k-N)\lambda^{\mathrm{e}}_{\max}(b) \\ &\geq \frac{N}{t} \int_0^{t} \lambda_s(a)\, ds + \sum_{i=N+1}^{k} \gamma_{\sigma(i)} = U_k(f) \end{aligned}$$

since, by (5-7), $\gamma_{\sigma(i)} \leq \lambda^{\mathrm{e}}_{\max}(b)$ for all i. □

Remark 5.2. Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse von Neumann subalgebra. Fix $a \in \mathcal{A}^+$, $b \in \mathcal{M}^+$ such that $a \prec_w b$ and let $m \in \mathbb{N}$. Then a slightly modified version of the proof of Proposition 5.1 (with $r_3 = s_3 = 0$, $\lambda^{\mathrm{e}}_{\min}(b) = \lambda^{\mathrm{e}}_{\min}(a) = 0$) shows that there exist $\{p_n\}_{n\geq 1} \subset \mathcal{P}(\mathcal{A})$, $\{q_n\}_{n\geq 1} \subset \mathcal{P}(\mathcal{M})$ and $f, g \in \ell^\infty(\mathbb{N})^+$ such that conditions (i)–(iii) and (b) hold, and such that $f \prec_w g$. We will use these facts for the proof of the contractive Schur–Horn theorem (Theorem 5.8).

The following result is standard, so its proof is omitted.

Lemma 5.3. *Let $\mathcal{N} \subset \mathcal{M}$ be a von Neumann subalgebra that admits a (unique) trace-preserving conditional expectation, denoted by $E_{\mathcal{N}}$. Let $\{p_j\}_{j\in\mathbb{N}} \subset \mathfrak{X}(\mathcal{N})$ be a family of mutually orthogonal projections, pairwise equivalent in $\mathcal{M}$. Let $\{e_{ij}\}$ be a system of matrix units in $B(H)$. Then there exists a (possibly nonunital) normal *-monomorphism $\pi : B(H) \to \mathcal{M}$ such that*

$$\pi(e_{jj}) = p_j, \quad j \in \mathbb{N}, \tag{5-10}$$

and

$$E_{\mathcal{N}}(\pi(x)) = \pi(P_D(x)), \quad x \in B(H). \tag{5-11}$$

The characterization of U_t in Lemma 4.1 allows us to prove that conditional expectations are "contractive" from a majorization point of view:

Lemma 5.4. *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathcal{A}}$. Then, for every $b \in \mathcal{M}^{\mathrm{sa}}$, we have $E_{\mathcal{A}}(b) \prec b$.*

Proof. Fix $t > 0$ and let $\varepsilon > 0$. Then we can apply Lemma 4.1 in $\mathcal{A}$ to get a projection $q \in \mathcal{P}(\mathcal{A})$ with $\tau(q) = t$ and such that $U_t(E_{\mathcal{A}}(b)) \leq \tau(E_{\mathcal{A}}(b)q) + \varepsilon$. Since $\tau(E_{\mathcal{A}}(b)q) = \tau(E_{\mathcal{A}}(bq)) = \tau(bq) \leq U_t(b)$, we conclude that $U_t(E_{\mathcal{A}}(b)) \leq U_t(b) + \varepsilon$ for all $\varepsilon > 0$; so, $U_t(E_{\mathcal{A}}(b)) \leq U_t(b)$. Applying the same proof to $-b$, we get $L_t(E_{\mathcal{A}}(b)) = -U_t(E_{\mathcal{A}}(-b)) \geq -U_t((-b)) = L_t(b)$. As t was arbitrary, we get $E_{\mathcal{A}}(b) \prec b$. ☐

We are finally in position to state and prove our main theorem.

Theorem 5.5 (Schur–Horn theorem for II_∞-factors). *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathcal{A}}$. Then, for any $b \in \mathcal{M}^{\mathrm{sa}}$,*

$$\overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\mathcal{T}} = \{a \in \mathcal{A}^{\mathrm{sa}} : a \prec b\}.$$

Proof. By Proposition 4.6 and Lemma 5.4, $\overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\mathcal{T}} \subset \{a \in \mathcal{A} : a \prec b\}$. To show the reverse inclusion, fix $a \in \mathcal{A}^{\mathrm{sa}}$ with $a \prec b$ and fix $m \in \mathbb{N}$. Applying Proposition 5.1

to a, b we obtain sequences $f=\{\alpha_n\}$, $g=\{\beta_n\}\subset\ell^\infty_{\mathbb{R}}(\mathbb{N})$, $\{p_n\}\subset\mathcal{P}(\mathcal{A})$, $\{q_n\}\subset\mathcal{P}(\mathcal{M})$ with

$$p_ip_j=q_iq_j=0 \quad \text{if } i\neq j, \qquad \tau(p_1)=\tau(p_j)=\tau(q_j) \quad \text{for all } j, \tag{5-12}$$

$$\tau\Big(1-\sum_{n\geq 1}p_n\Big)=\tau\Big(1-\sum_{n\geq 1}q_n\Big)<\frac{1}{m}, \tag{5-13}$$

$$\Big(a-\sum_{n\geq 1}\alpha_np_n\Big),\Big(b-\sum_{n\geq 1}\beta_nq_n\Big)\in V\Big(\frac{1}{m},\frac{1}{m}\Big), \tag{5-14}$$

and $f\prec g$. By Theorem 3.3 there exists a unitary $v\in B(H)$ such that

$$\|M_f-P_D(vM_gv^*)\|<\frac{1}{m}.$$

The conditions on the projections in (5-12) and (5-13) guarantee that we can choose $w\in\mathcal{U}_{\mathcal{M}}$ with $wq_nw^*=p_n$ for all n. Let $p=\sum_n p_n$, $q=\sum_n q_n$; then by (5-13) there exists a partial isometry $z\in\mathcal{M}$ with $z^*z=p^\perp$, $zz^*=q^\perp$. Let u be the unitary $u=(\pi(v)+z)w$, where π is the *-monomorphism from Lemma 5.3 with respect to the projections $\{p_n\}_n$. From (5-14),

$$a-\pi(M_f)\in V\Big(\frac{1}{m},\frac{1}{m}\Big), \qquad wbw^*-\pi(M_g)\in V\Big(\frac{1}{m},\frac{1}{m}\Big).$$

Note that by (5-13) we have $\tau(p^\perp)<1/m$, $\tau(q^\perp)<1/m$, so $z,z^*\in V(\varepsilon,1/m)$ for any $\varepsilon>0$. From this we conclude that

$$(\pi(v)+z)\pi(M_g)(\pi(v)+z)^*-\pi(vM_gv^*)\in V\Big(\varepsilon,\frac{2}{m}\Big), \qquad \varepsilon>0.$$

It follows that

$$ubu^*-\pi(vM_gv^*)\in V\Big(\frac{2}{m},\frac{3}{m}\Big).$$

Letting m vary all along $\mathbb{N}$, we have constructed sequences of unitaries $\{u_m\}_m\subset\mathcal{M}$ and $\{v_m\}_m\subset\mathcal{U}(H)$, and sequences $\{f_m\}_m,\{g_m\}_m\subset\ell^\infty_{\mathbb{R}}(\mathbb{N})$ with

$$\pi(M_{f_m})-a\xrightarrow[m\to\infty]{\mathcal{T}}0, \qquad M_{f_m}-P_D(v_mM_{g_m}v_m^*)\xrightarrow[m\to\infty]{\|\ \|}0, \tag{5-15}$$
$$u_mbu_m^*-\pi(v_mM_{g_m}v_m^*)\xrightarrow[m\to\infty]{\mathcal{T}}0.$$

Using that π is a *-monomorphism, the $\mathcal{T}$-continuity of $E_{\mathcal{A}}$ (Corollary 2.4) and the fact that $E_{\mathcal{A}}\circ\pi=\pi\circ P_D$ (Lemma 5.3) we get from (5-15) that

$$\pi(M_{f_m})-\pi(P_D(v_mM_{g_m}v_m^*))\xrightarrow[m\to\infty]{\|\ \|}0 \tag{5-16}$$

and

$$E_{\mathcal{A}}(u_mbu_m^*)-\pi(P_D(v_mM_{g_m}v_m^*))\xrightarrow[m\to\infty]{\mathcal{T}}0. \tag{5-17}$$

From (5-15), (5-16), and (5-17), we get $E(u_m b u_m^*) - a \xrightarrow[m\to\infty]{\mathcal{T}} 0$. That is, a lies in $\overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\mathcal{T}}$. □

Remark 5.6. Consider the notations and hypothesis in the statement of Theorem 5.5. It is natural to ask whether one can remove the closure bar in the description of the set $\{a \in \mathcal{A}^{\mathrm{sa}} : a \prec b\}$ given in Theorem 5.5. Next we show an example in which

$$E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b)) \subset E_{\mathcal{A}}\big(\overline{\mathcal{U}_{\mathcal{M}}(b)}^{\mathcal{T}}\big) \subsetneq \overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\mathcal{T}}.$$

This implies that the characterization of $\{a \in \mathcal{A}^{\mathrm{sa}} : a \prec b\}$ given in Theorem 5.5 cannot be strengthened in the II_∞ case.

We consider $p \in \mathcal{P}(\mathcal{M})$ an infinite projection with $p^\perp$ also infinite. Then $U_t(p) = t$, $L_t(p) = 0$ for all t. Since $U_t(I) = t$, $L_t(I) = t$, we have $I \prec p$; then

$$I \in \overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(p))}^{\mathcal{T}} \quad \text{but} \quad I \notin E_{\mathcal{A}}\big(\overline{\mathcal{U}_{\mathcal{M}}(p)}^{\mathcal{T}}\big). \tag{5-18}$$

Indeed, Theorem 5.5 guarantees the claim to the left in (5-18). On the other hand, assume that there exists $x \in \overline{\mathcal{U}_{\mathcal{M}}(p)}^{\mathcal{T}}$ with $I = E_{\mathcal{A}}(x)$. By Corollary 2.4, $0 \le x \le I$ and then

$$0 = \tau(I - E_{\mathcal{A}}(x)) = \tau(E_{\mathcal{A}}(I - x)) = \tau(I - x).$$

This last fact implies that $I = x \in \overline{\mathcal{U}_{\mathcal{M}}(p)}^{\mathcal{T}}$ by the faithfulness of τ. But as $\|\cdot\|_{(1)}$ is a unitarily invariant norm, for any $u \in \mathcal{U}_{\mathcal{M}}$ we get

$$\|I - upu^*\|_{(1)} = \|u(I - p)u^*\|_{(1)} = \|I - p\|_{(1)} > 0$$

as $p \neq I$. Since $\|\cdot\|_{(1)}$ is $\mathcal{T}$-continuous (see Proposition 2.2), there is positive distance from I to the $\mathcal{T}$-closure of the unitary orbit of p, a contradiction.

It would be interesting to have a description of the set $E_{\mathcal{A}}(\overline{\mathcal{U}_{\mathcal{M}}(b)}^{\mathcal{T}})$ for an abelian diffuse von Neumann subalgebra $\mathcal{A}$ of a general σ-finite semifinite factor $(\mathcal{M}, \tau)$, that admits a trace preserving conditional expectation $E_{\mathcal{A}}$. But even in the I_∞ factor case this problem is known to be hard (see [Kadison 2002, Theorem 15; Arveson 2007; Arveson and Kadison 2006] for further discussion). In the II_1-factor case Arveson and Kadison [2006] conjectured that

$$E_{\mathcal{A}}\big(\overline{\mathcal{U}_{\mathcal{M}}(b)}^{\mathcal{T}}\big) = \{a \in \mathcal{A}^{\mathrm{sa}} : a \prec b\}, \tag{5-19}$$

which is still an open problem (see [Argerami and Massey 2007; 2008a; 2009] for a detailed discussion). □

The next result shows that the notion of majorization in $\mathcal{M}^{\mathrm{sa}}$ from Definition 4.4 coincides with the majorization introduced in [Hiai 1992]. Thus, several other characterizations of majorization can be obtained from Hiai's work. Following Hiai, we say that a map is *doubly stochastic* if it is unital, positive and preserves the trace.

Corollary 5.7. *Let $\mathscr{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathscr{A}}$. Given $a, b \in \mathcal{M}^{\mathrm{sa}}$, the following statements are equivalent:*

(i) $a \prec b$.

(ii) $a \in \overline{E_{\mathscr{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\mathcal{T}}$.

(iii) $a \in \overline{\operatorname{conv}\{\mathcal{U}_{\mathcal{M}}(b)\}}^{\mathcal{T}}$.

(iv) *There exists a doubly stochastic map F on $\mathcal{M}$ with $a = F(b)$.*

(v) *There exists a completely positive doubly stochastic map F on $\mathcal{M}$ with $a = F(b)$.*

(vi) $\tau(f(a)) \le \tau(f(b))$ *for every convex function $f : I \to [0, \infty)$ with $\sigma(a) \subset I$ and $\sigma(b) \subset I$.*

(vii) *a is spectrally majorized by b* (in the sense of [Hiai 1992]).

Proof. By Theorem 5.5, (i) and (ii) are equivalent. The statements (iii)–(vii) are mutually equivalent by [Hiai 1992, Theorem 2.2]. Also, (iii) implies (i) by Proposition 4.6. So it will be enough to show that (i) implies (iv).

Let $a \in \mathscr{A}$ with $a \prec b$. By Theorem 5.5, there exist unitaries $\{u_j\} \subset \mathcal{M}$ such that $a = \lim_{\mathcal{T}} E_{\mathscr{A}}(u_j b u_j^*)$. Consider the sequence of completely positive contractions $E_{\mathscr{A}}(u_j \cdot u_j^*) : \mathcal{M} \to \mathscr{A}$; by compactness in the BW topology [Paulsen 2002, Theorem 7.4], this sequence admits a convergent (pointwise ultraweakly) subnet $\{E_{\mathscr{A}}(u_{j_k} \cdot u_{j_k}^*)\}$. Let F be the limit of such subnet. Since $a = \lim_{\mathcal{T}} E_{\mathscr{A}}(u_j b u_j^*)$ and $F(b) = \lim_{\sigma-\mathrm{wot}} E_{\mathscr{A}}(u_{j_k} b u_{j_k}^*)$, we conclude (mimicking the argument in the proof of Lemma 3.3 in [Hiai 1992]) that $F(b) = a$. It is easy to check that F is unital and that it preserves the trace. □

We finish this section with contractive and L^1 analogs of Theorem 5.5.

Theorem 5.8. *Let $\mathscr{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathscr{A}}$. If $b \in \mathcal{M}^+$ then*

$$\overline{E_{\mathscr{A}}(\{cbc^* : \|c\| \le 1\})}^{\mathcal{T}} = \{a \in \mathscr{A}^+ : a \prec_w b\}. \tag{5-20}$$

Proof. If $c \in \mathcal{M}$ is a contraction, then $\lambda_t(cbc^*) \le \lambda_t(b)$ [Fack and Kosaki 1986, Lemma 2.5]. So $cbc^* \prec_w b$ and then Lemmas 5.4 and 4.3 give the inclusion "$\subset$" above.

For the reverse inclusion, the proof runs exactly as that of Theorem 5.5, but instead of using Proposition 5.1 and (3-5) to obtain a sequence of unitary operators in $\mathcal{M}$, we use (3-11) and Remark 5.2 to obtain a convenient sequence of contractions in $\mathcal{M}$. □

Remark 5.9. The positivity condition in Theorem 5.8 cannot be relaxed to selfadjointness. As a trivial example, take $b = 0$; then $-I \prec_w b$, but $cbc^* = 0$ for all c, so the set on the left in (5-20) is $\{0\}$.

Recall that $L^1(\mathcal{M}) \cap \mathcal{M}$ consists of those $x \in \mathcal{M}$ with $\tau(|x|) < \infty$, and that such elements are necessarily τ-compact.

Theorem 5.10. *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathcal{A}}$. If $b \in L^1(\mathcal{M}) \cap \mathcal{M}^{\mathrm{sa}}$ then*

$$\overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\|\cdot\|_1} = \{a \in L^1(\mathcal{M}) \cap \mathcal{A}^{\mathrm{sa}} : a \prec b,\, \tau(a) = \tau(b)\}.$$

Proof. Proposition 4.6 together with Lemma 5.4 show that $E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b)) \subset \{a \in \mathcal{A}^{\mathrm{sa}} : a \prec b,\, \tau(a) = \tau(b)\}$. Then Lemma 4.3 and the $\|\cdot\|_1$-continuity of the trace imply the inclusion of the corresponding closure.

Conversely, suppose that $a \prec b$ and $\tau(a) = \tau(b)$. First assume that $b \in \mathcal{M}^+$. Then $a \in \mathcal{A}^+$. By Theorem 5.5, there exists a sequence of unitaries $\{u_j\}$ such that

$$E_{\mathcal{A}}(u_j b u_j^*) \xrightarrow{\mathcal{T}} a.$$

Since b is positive, $\|E_{\mathcal{A}}(u_j b u_j^*)\|_1 = \tau(E_{\mathcal{A}}(u_j b u_j^*)) = \tau(b) = \tau(a) = \|a\|_1$. Then [Fack and Kosaki 1986, Theorem 3.7] guarantees that $\|E_{\mathcal{A}}(u_j b u_j^*) - a\|_1 \to 0$.

If b is not positive, we apply Lemma 4.10 to obtain $a' \in \mathcal{A}$, $b' \in \mathcal{M}$, with

(i) $a' \prec b'$;

(ii) $\|a' - a\|_1 < \varepsilon$, $\|b' - b\|_1 < \varepsilon$;

(iii) $\tau(p^{a'}(0,\infty)) = \tau(p^{b'}(0,\infty)) = \infty$;

(iv) $\tau(p^{a'}(-\infty,0)) = \tau(p^{b'}(-\infty,0)) = \infty$;

(v) $p^{a'}(-\infty,0) + p^{a'}(0,\infty) = p^{b'}(-\infty,0) + p^{b'}(0,\infty) = I$.

Let $r_1 = p^{a'_+}(0,\infty)$, $r_2 = p^{a'_-}(0,\infty)$. The last three conditions above guarantee that we can find a unitary $v \in \mathcal{U}_{\mathcal{M}}$ with

$$v(p^{b'_+}(0,\infty))v^* = r_1, \quad v(p^{b'_-}(0,\infty))v^* = r_2.$$

Let $b'' = v b' v^*$. Then $a' \prec b''$. Since both are τ-compact, we deduce that $a'_+ \prec b''_+$, $a'_- \prec b''_-$. Note that

$$a'_+, b''_+ \in r_1 \mathcal{M} r_1, \quad a'_-, b''_- \in r_2 \mathcal{M} r_2.$$

As both $r_1, r_2 \in \mathcal{A}$ are infinite projections, the factors $r_1\mathcal{M}r_1$ and $r_2\mathcal{M}r_2$ are II_∞. So we can apply the first part of the proof to obtain unitaries $\{u_j^{(1)}\} \subset \mathcal{U}(r_1\mathcal{M}r_1)$, $\{u_j^{(2)}\} \subset \mathcal{U}(r_2\mathcal{M}r_2)$, with

$$\|E_{\mathcal{A}}(u_j^{(1)} b''_+ (u_j^{(1)})^*) - a'_+\|_1 \to 0, \quad \|E_{\mathcal{A}}(u_j^{(2)} b''_- (u_j^{(2)})^*) - a'_-\|_1 \to 0.$$

Since $r_1 + r_2 = I$, $r_1 r_2 = 0$, the operators $u_j = (u_j^{(1)} + u_j^{(2)})v$ are unitaries in $\mathcal{M}$.

Then

$$\begin{aligned}
&\|E_{\mathcal{A}}(u_j b u_j^*) - a\|_1 \\
&\quad \le \|E_{\mathcal{A}}(u_j b u_j^*) - E_{\mathcal{A}}(u_j b' u_j^*)\|_1 + \|E_{\mathcal{A}}(u_j b' u_j^*) - a'\|_1 + \|a' - a\|_1 \\
&\quad \le \|b'-b\|_1 + \|a'-a\|_1 + \|E_{\mathcal{A}}(u_j^{(1)} b'' (u_j^{(1)})^*) - a'_+\|_1 + \|E_{\mathcal{A}}(u_j^{(2)} b'' (u_j^{(2)})^*) - a'_-\|_1 \\
&\quad \le 2\varepsilon + \|E_{\mathcal{A}}(u_j^{(1)} b''_+ (u_j^{(1)})^*) - a'_+\|_1 + \|E_{\mathcal{A}}(u_j^{(2)} b''_- (u_j^{(2)})^*) - a'_-\|_1.
\end{aligned}$$

So $\limsup_j \|E_{\mathcal{A}}(u_j b u_j^*) - a\|_1 < 2\varepsilon$, and as ε was arbitrary we conclude that $\lim_j \|E_{\mathcal{A}}(u_j b u_j^*) - a\|_1 = 0$, i.e., $a \in \overline{E_{\mathcal{A}}(\mathcal{U}_{\mathcal{M}}(b))}^{\|\cdot\|_1}$. □

Remark 5.11. The condition $\tau(a) = \tau(b)$ in Theorem 5.10 cannot be removed because of the $\|\cdot\|_1$-continuity of the trace τ. Actually, below we characterize the case where the trace restriction is removed but only in the case of positive operators.

Theorem 5.12. *Let $\mathcal{A} \subset \mathcal{M}$ be a diffuse abelian von Neumann subalgebra that admits a (unique) trace preserving conditional expectation, denoted by $E_{\mathcal{A}}$. If $b \in L^1(\mathcal{M}) \cap \mathcal{M}^+$ then*

$$\overline{E_{\mathcal{A}}(\{cbc^* : \|c\| \le 1\})}^{\|\cdot\|_1} = \{a \in \mathcal{A}^+ : a \prec_w b\} = \{a \in \mathcal{A}^+ : a \prec b\}.$$

Proof. If $b \in L^1(\mathcal{M}) \cap \mathcal{M}^+$ and $a \prec_w b$ then, since $\lambda_t(b) \in L^1(\mathbb{R}^+)$, we get $\lambda_t(a) \in L^1(\mathbb{R}^+)$. In particular, $a \in \mathcal{K}(\mathcal{M})^+$. Thus, the second equality is immediate from the fact that for positive τ-compact operators one has $L_t = 0$. So for the rest of the proof we focus on the first equality.

The inclusion "$\subset$" is obtained by combining the arguments at the beginning of the proofs of Theorems 5.8 and 5.10.

Conversely, let $a \prec_w b$ for some $a \in \mathcal{A}^+$ (so that $a \in \mathcal{K}(\mathcal{A})^+$). We write both a and b in terms of complete flags in $\mathcal{A}$ and $\mathcal{M}$ respectively, i.e.,

$$a = \int_0^\infty \lambda_t(a)\, de_a(t), \qquad b = \int_0^\infty \lambda_t(b)\, de_b(t),$$

with $e_a(t) \in \mathcal{A}$ for all t (this can be done since $\mathcal{A}$ is diffuse). Then $a \prec_w b$ means that, for any $s > 0$, $\int_0^s \lambda_t(a)\, dt \le \int_0^s \lambda_t(b)\, dt$. For each $s > 0$, let $p_s = e_a(s) \vee e_b(s)$, a finite projection. So we have $ae_a(s) \prec_w be_b(s)$ in the II_1-factor $p_s \mathcal{M} p_s$. By [Argerami and Massey 2008a, Theorem 3.4], there exists a contraction $c_s \in p_s \mathcal{M} p_s \subset \mathcal{M}$ with

$$k_s := \tau_s(|ae_a(s) - E_{\mathcal{A}e_a(s)}(c_s e_b(s) b e_b(s) c_s^*)|) < \frac{1}{\tau(p_s)^2}.$$

The trace τ_s is given by $\tau_s = \tau/\tau(p_s)$; using the fact that $e_a(s) \in \mathcal{A}$ and that $\mathcal{A}$ is abelian, we get that $E_{\mathcal{A}e_a(s)}(\,\cdot\,) = e_a(s) E_{\mathcal{A}}(\,\cdot\,)$. So

$$\tau(|ae_a(s) - E_{\mathcal{A}}(e_a(s) c_s e_b(s) b e_b(s) c_s^* e_a(s))|) = \tau(p_s) k_s < \frac{1}{\tau(p_s)} \le \frac{1}{s}$$

(note that $p_s \geq e_a(s)$, so $\tau(p_s) \geq s$). Let $\varepsilon > 0$; fix $s > 0$ such that $s > 2/\varepsilon$ and $\int_s^\infty \lambda_t(a)\,dt < \varepsilon/2$. Put $c = e_a(s)c_s e_b(s)$, a contraction in $\mathcal{M}$. Then

$$\begin{aligned}\|a - E_{\mathcal{A}}(cbc^*)\|_1 &\leq \|a - ae_a(s)\|_1 + \|ae_a(s) - E_{\mathcal{A}}(e_a(s)c_s e_b(s)be_b(s)c_s^* e_a(s))\|_1 \\ &= \int_s^\infty \lambda_a(t)\,dt + \tau\big(|ae_a(s) - E_{\mathcal{A}}(e_a(s)c_s e_b(s)be_b(s)c_s^* e_a(s))|\big) \\ &\leq \frac{\varepsilon}{2} + \frac{1}{s} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

As ε was arbitrary, this shows that $a \in \overline{E_{\mathcal{A}}(\{cbc^* : \|c\| \leq 1\})}^{\|\cdot\|_1}$. □

Remark 5.13. The proof of Theorem 5.12 uses a reduction to a II_1 case, under the hypothesis that the operators belong to $L^1(\mathcal{M})$. This last assumption seems to be essential for such a reduction, and there is no immediate hope of using the same idea to obtain results like Theorems 5.5 and 5.8. Conversely, one cannot expect to use those results to obtain Theorem 5.12, since convergence in measure does not imply $\|\cdot\|_1$-convergence.

References

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

[Argerami and Massey 2007] M. Argerami and P. Massey, "A Schur–Horn theorem in II_1 factors", *Indiana Univ. Math. J.* **56**:5 (2007), 2051–2059. MR 2008m:46120 Zbl 1136.46043

[Argerami and Massey 2008a] M. Argerami and P. Massey, "A contractive version of a Schur–Horn theorem in II_1 factors", *J. Math. Anal. Appl.* **337**:1 (2008), 231–238. MR 2008m:46121 Zbl 1130.46038

[Argerami and Massey 2008b] M. Argerami and P. Massey, "The local form of doubly stochastic maps and joint majorization in II_1 factors", *Integral Equations Operator Theory* **61**:1 (2008), 1–19. MR 2010m:46102 Zbl 1152.46053

[Argerami and Massey 2009] M. Argerami and P. Massey, "Towards the carpenter's theorem", *Proc. Amer. Math. Soc.* **137**:11 (2009), 3679–3687. MR 2011a:46087 Zbl 1183.46058

[Arveson 2007] W. Arveson, "Diagonals of normal operators with finite spectrum", *Proc. Natl. Acad. Sci. USA* **104**:4 (2007), 1152–1158. MR 2008f:47027 Zbl 1191.47027

[Arveson and Kadison 2006] W. Arveson and R. V. Kadison, "Diagonals of self-adjoint operators", pp. 247–263 in *Operator theory, operator algebras, and applications*, edited by D. Han et al., Contemp. Math. **414**, Amer. Math. Soc., Providence, RI, 2006. MR 2007k:46116 Zbl 1113.46064

[Bhatia 1997] R. Bhatia, *Matrix analysis*, Graduate Texts in Mathematics **169**, Springer, New York, 1997. MR 98i:15003 Zbl 0863.15001

[Birkhoff 1946] G. Birkhoff, "Three observations on linear algebra", *Univ. Nac. Tucumán. Revista A.* **5** (1946), 147–151. MR 8,561a

[Davidson 1996] K. R. Davidson, *C^*-algebras by example*, Fields Institute Monographs **6**, Amer. Math. Soc., Providence, RI, 1996. MR 97i:46095 Zbl 0958.46029

[Dhillon et al. 2005] I. S. Dhillon, R. W. Heath, Jr., M. A. Sustik, and J. A. Tropp, "Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum", *SIAM J. Matrix Anal. Appl.* **27**:1 (2005), 61–71. MR 2006i:15019 Zbl 1087.65038

[Fack 1982] T. Fack, "Sur la notion de valeur caractéristique", *J. Operator Theory* **7**:2 (1982), 307–333. MR 84m:47012 Zbl 0493.46052

[Fack and Kosaki 1986] T. Fack and H. Kosaki, "Generalized s-numbers of τ-measurable operators", *Pacific J. Math.* **123**:2 (1986), 269–300. MR 87h:46122 Zbl 0617.46063

[Hardy et al. 1929] G. H. Hardy, J. E. Littlewood, and G. Pólya, "Some simple inequalities satisfied by convex functions", *Messenger Math* **58** (1929), 145–152. JFM 55.0740.04

[Hiai 1987] F. Hiai, "Majorization and stochastic maps in von Neumann algebras", *J. Math. Anal. Appl.* **127**:1 (1987), 18–48. MR 88k:46076 Zbl 0634.46051

[Hiai 1992] F. Hiai, "Spectral majorization between normal operators in von Neumann algebras", pp. 78–115 in *Operator algebras and operator theory* (Craiova, 1989), edited by W. B. Arveson et al., Pitman Res. Notes Math. Ser. **271**, Longman Sci. Tech., Harlow, 1992. MR 94a:46094 Zbl 0790.46045

[Hiai and Nakamura 1987] F. Hiai and Y. Nakamura, "Majorizations for generalized s-numbers in semifinite von Neumann algebras", *Math. Z.* **195**:1 (1987), 17–27. MR 88g:46070 Zbl 0598.46039

[Horn 1954] A. Horn, "Doubly stochastic matrices and the diagonal of a rotation matrix", *Amer. J. Math.* **76** (1954), 620–630. MR 16,105c Zbl 0055.24601

[Kadison 2002] R. V. Kadison, "The Pythagorean theorem, II: The infinite discrete case", *Proc. Natl. Acad. Sci. USA* **99**:8 (2002), 5217–5222. MR 2003e:46108b

[Kadison 2004] R. V. Kadison, "Non-commutative conditional expectations and their applications", pp. 143–179 in *Operator algebras, quantization, and noncommutative geometry* (Baltimore, MD, 2003), edited by R. S. Doran and R. V. Kadison, Contemp. Math. **365**, Amer. Math. Soc., Providence, RI, 2004. MR 2005i:46072 Zbl 1080.46044

[Kaftal and Weiss 2008] V. Kaftal and G. Weiss, "A survey on the interplay between arithmetic mean ideals, traces, lattices of operator ideals, and an infinite Schur–Horn majorization theorem", pp. 101–135 in *Hot topics in operator theory* (Timişoara, 2006), edited by R. G. Douglas et al., Theta Ser. Adv. Math. **9**, Theta, Bucharest, 2008. MR 2010b:47050 Zbl 1199.47166 arXiv 0707.3271

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

[Kamei 1983] E. Kamei, "Majorization in finite factors", *Math. Japon.* **28**:4 (1983), 495–499. MR 84j:46086 Zbl 0527.47016

[Kamei 1984] E. Kamei, "Double stochasticity in finite factors", *Math. Japon.* **29**:6 (1984), 903–907. MR 88a:46067a Zbl 0557.46038

[Marshall et al. 2011] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, 2nd ed., Springer, New York, 2011. MR 2012g:26001 Zbl 1219.26003

[Massey and Ruiz 2010] P. Massey and M. Ruiz, "Minimization of convex functionals over frame operators", *Adv. Comput. Math.* **32**:2 (2010), 131–153. MR 2011b:42109 Zbl 1191.42017

[Neumann 1999] A. Neumann, "An infinite-dimensional version of the Schur–Horn convexity theorem", *J. Funct. Anal.* **161**:2 (1999), 418–451. MR 2000a:22030 Zbl 0926.52001

[Neumann 2002] A. Neumann, "An infinite dimensional version of the Kostant convexity theorem", *J. Funct. Anal.* **189**:1 (2002), 80–131. MR 2003d:47100 Zbl 1035.17032

[Paulsen 2002] V. Paulsen, *Completely bounded maps and operator algebras*, Cambridge Studies in Advanced Mathematics **78**, Cambridge University Press, Cambridge, 2002. MR 2004c:46118 Zbl 1029.47003

[Petz 1985] D. Petz, "Spectral scale of selfadjoint operators and trace inequalities", *J. Math. Anal. Appl.* **109**:1 (1985), 74–82. MR 87c:47055 Zbl 0655.47032

[Schur 1923] I. Schur, "Über eine Klasse von Mittelbildungen mit Anwendung auf die Determinantentheorie", *S.-Ber. Berliner Math. Ges.* **22** (1923), 9–20. JFM 49.0054.01

Received May 16, 2011.

MARTÍN ARGERAMI
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF REGINA
REGINA, SK S4S 0A2
CANADA
argerami@math.uregina.ca

PEDRO MASSEY
DEPARTAMENTO DE MATEMÁTICA - FCE
UNIVERSIDAD NACIONAL DE LA PLATA AND
INSTITUTO ARGENTINO DE MATEMÁTICA "ALBERTO P. CALDERÓN" – CONICET
1083 BUENOS AIRES
ARGENTINA
massey@mate.unlp.edu.ar

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.311

CLASSIFICATION OF POSITIVE SOLUTIONS FOR AN ELLIPTIC SYSTEM WITH A HIGHER-ORDER FRACTIONAL LAPLACIAN

JINGBO DOU AND CHANGZHENG QU

We discuss properties of solutions to the following elliptic PDE system in $\mathbb{R}^n$:

$$\begin{cases}(-\Delta)^{\alpha/2}u = \lambda_1 u^{p_1} + \mu_1 v^{p_2} + \beta_1 u^{p_3} v^{p_4},\\ (-\Delta)^{\alpha/2}v = \lambda_2 u^{q_1} + \mu_2 v^{q_2} + \beta_2 u^{q_3} v^{q_4},\end{cases}$$

where $0 < \alpha < n$, λ_j, μ_j, β_j $(j = 1, 2)$ are nonnegative constants and p_i and q_i $(i = 1, 2, 3, 4)$ satisfy some suitable assumptions. It is shown that this PDE system is equivalent to the integral system

$$\begin{cases}u(x) = \displaystyle\int_{\mathbb{R}^n} \frac{\lambda_1 u^{p_1}(y) + \mu_1 v^{p_2}(y) + \beta_1 u^{p_3}(y) v^{p_4}(y)}{|x-y|^{n-\alpha}}\,dy,\\ v(x) = \displaystyle\int_{\mathbb{R}^n} \frac{\lambda_2 u^{q_1}(y) + \mu_2 v^{q_2}(y) + \beta_2 u^{q_3}(y) v^{q_4}(y)}{|x-y|^{n-\alpha}}\,dy\end{cases}$$

in $\mathbb{R}^n$. The radial symmetry, monotonicity and regularity of positive solutions are proved via the method of moving plane in integral forms and a regularity lifting lemma. For the special case with

$$p_1 = p_2 = q_1 = q_2 = p_3 + p_4 = q_3 + q_4 = \frac{n+\alpha}{n-\alpha},$$

positive solutions of the integral system (or the PDE system) are classified. Furthermore, our symmetry results, together with some known results on nonexistence of positive solutions, imply that, under certain integrability conditions, the PDE system has no positive solution in the subcritical case.

1. Introduction

In this paper, we study positive solutions of the following higher-order elliptic system in $\mathbb{R}^n$:

$$\begin{cases}(-\Delta)^{\alpha/2}u = \lambda_1 u^{p_1} + \mu_1 v^{p_2} + \beta_1 u^{p_3} v^{p_4},\\ (-\Delta)^{\alpha/2}v = \lambda_2 u^{q_1} + \mu_2 v^{q_2} + \beta_2 u^{q_3} v^{q_4},\end{cases} \tag{1}$$

MSC2010: 35J99, 45K05, 35R09, 35J48, 35B65, 45G15.

Keywords: system of integral equations, regularity, moving plane method in integral form, classification of solutions.

where $(-\Delta)^{\alpha/2}$ is a higher-order fractional Laplacian, $0 < \alpha < n$, $\lambda_i, \mu_i, \beta_i \geq 0$ $(i = 1, 2)$ are constants, and p_i and q_i $(i = 1, 2, 3, 4)$ satisfy some suitable assumptions.

System (1) arises from N-coupled higher-order nonlinear Schrödinger systems

$$
(2) \qquad \begin{cases} i\dfrac{\partial \Phi_j}{\partial t} - (-\Delta)^m \Phi_j + \sum\limits_{i=1}^{N} \beta_{ij} |\Phi_i|^2 \Phi_j = 0, & y \in \mathbb{R}^n,\ t > 0, \\ \Phi_j(y, t) \to 0, & \text{as } y \to \infty,\ t > 0,\ j = 1, 2, \dots, N, \end{cases}
$$

for $m \in \mathbb{N}$, $\beta_{ij} = \beta_{ji}$. System (2) appears in some physical problems, especially in nonlinear optics. When $m = 1$, $n \leq 3$, it describes physical phenomena such as the propagation in birefringent optical fibers, Kerr-like photo refractive media in optics (see [Akhmediev and Ankiewicz 1999]) and Bose–Einstein condensates (see [Esry et al. 1997]). When the spatial dimension is one, i.e., $n = 1$, system (2) has applications in quantum mechanics (see [Liu et al. 2007; Fu et al. 2009]).

Letting $\Phi_j(y, t) = e^{-i\lambda_j t} u_j(y)$, system (2) is transformed into the elliptic system

$$
(3) \qquad \begin{cases} (-\Delta)^m u_j = \lambda_j u_j + \sum\limits_{i=1}^{N} \beta_{ij} |u_i|^2 u_j & \text{in } \mathbb{R}^n, \\ u_j(x) \to 0, & \text{as } |x| \to \infty,\ j = 1, 2, \dots, N. \end{cases}
$$

Clearly, in some sense, system (1) extends system (3) with $N = 2$.

For further discussion, we need an integral form of system (1). In Section 5, we will show that system (1) is equivalent to the following system:

$$
(4) \qquad \begin{cases} u(x) = \displaystyle\int_{\mathbb{R}^n} \frac{\lambda_1 u^{p_1}(y) + \mu_1 v^{p_2}(y) + \beta_1 u^{p_3}(y) v^{p_4}(y)}{|x - y|^{n-\alpha}}\, dy, \\ v(x) = \displaystyle\int_{\mathbb{R}^n} \frac{\lambda_2 u^{q_1}(y) + \mu_2 v^{q_2}(y) + \beta_2 u^{q_3}(y) v^{q_4}(y)}{|x - y|^{n-\alpha}}\, dy. \end{cases}
$$

In particular, when $\mu_1 = \lambda_2 = 1$ and $\lambda_1 = \mu_2 = \beta_1 = \beta_2 = 0$, system (4) reduces to

$$
(5) \qquad \begin{cases} u(x) = \displaystyle\int_{\mathbb{R}^n} \frac{v^{p_2}(y)}{|x - y|^{n-\alpha}}\, dy, \\ v(x) = \displaystyle\int_{\mathbb{R}^n} \frac{u^{q_1}(y)}{|x - y|^{n-\alpha}}\, dy, \end{cases}
$$

which is closely related to the maximizer of the best constant in a Hardy–Littlewood–Sobolev (HLS) inequality; see [Chen et al. 2005; Chen and Li 2005].

In recent years many works have been devoted to the study of the special cases of system (1) or system (4). In the case of $\alpha = 2$, under certain assumptions, the existence of bound state solutions and radially symmetric solutions of (3) was studied in [Bartsch et al. 2007; 2010; Busca and Sirakov 2000; Dancer et al. 2010; Liu and Wang 2008; Guo and Liu 2008; Hioe 1999; Lin and Wei 2005; Maia et al. 2006; Sirakov 2007; Wei and Weth 2007; 2008]. In particular, for $\alpha = 2$ ($n \geq 3$) and $\lambda_i = \mu_i = 1$, $\beta_i \geq 0$ ($i = 1, 2$), de Figueiredo and Sirakov [2005] proved the nonexistence of positive solutions for system (1) under some subcritical exponent conditions. When $m = n = 1$, system (2) is integrable, and there are many analytical and numerical results on solitary wave solutions of higher-order nonlinear Schrödinger equations (e.g., see [Liu et al. 2007; Fu et al. 2009]).

In the case of $\alpha = 2m$ ($m = 1, 2, \dots$) and $\mu_1 = \lambda_2 = 1$, $\lambda_1 = \mu_2 = \beta_1 = \beta_2 = 0$, system (1) becomes

$$\begin{cases} (-\Delta)^m u = v^{p_2}, \\ (-\Delta)^m v = u^{q_1}, \end{cases} \tag{6}$$

in $\mathbb{R}^n$. This system is equivalent to the integral system (5) with $\alpha = 2m$ (see [Chen and Li 2009b]). Guo, Liu and Zhang [Liu et al. 2006; Zhang 2007] proved that any positive solutions of (6) are radially symmetric for critical exponents $p_2 = q_1 = \frac{n+2m}{n-2m}$. Moreover, they also showed that there are no positive solutions of (6) if $p_2, q_1 \geq 1$, but are not both equal to 1, and satisfy the following subcritical exponent condition:

$$\frac{1}{p_2+1} + \frac{1}{q_1+1} > \frac{n-2m}{n}.$$

Assuming that p_2 and q_1 satisfy $\frac{\alpha}{n-\alpha} < p_2, q_1 < \infty$, under natural integrability conditions on u and v, Chen, Li and Ou [Chen et al. 2005; Chen and Li 2005] and Hang [2007] discussed the symmetry, monotonicity and regularity of positive solutions of system (5) with the critical exponent condition

$$\frac{1}{p_2+1} + \frac{1}{q_1+1} = \frac{n-\alpha}{n}.$$

Furthermore, Chen and Li [2009b] proved the nonexistence of positive solutions of system (5) satisfying some subcritical exponents assumptions.

In [Dou et al. 2011], we studied the symmetry, monotonicity and regularity of positive solutions of integral system (5) with weighted functions for $\max\{1, \frac{\alpha}{n-\alpha}\} < p_2, q_1 < \infty$ and

$$\frac{1}{p_2+1} + \frac{1}{q_1+1} \geq \frac{n-\alpha}{n}.$$

In addition, the nonexistence result for positive solutions of system (5) with $0 < p_2, q_1 < \frac{n+\alpha}{n-\alpha}$ was established.

In the case of $\lambda_i = 1$, $\mu_i = \beta_i = 0$ and $u(x) = v(x)$, system (1) reduces to the single elliptic equation

$$(-\Delta)^{\alpha/2} u = u^p, \quad \text{in } \mathbb{R}^n. \tag{7}$$

For $p = \frac{n+\alpha}{n-\alpha}$, Chen et al. [2006] and Li [2004] proved that any positive solutions u of Equation (7) are radially symmetric and monotonic about some point. Indeed all the positive solutions are given by

$$u(x)=\left(\frac{C_\alpha}{d+|x-\bar{x}|^2}\right)^{(n-\alpha)/2}, \tag{8}$$

where $d>0$ is a constant and $C_\alpha=\bigl(2^{-\alpha}\Gamma(\frac{n+\alpha}{2})\Gamma(\frac{n-\alpha}{2})^{-1}d\bigr)^{1/2}$. When $\alpha=2m$ is any even number, the above result was also proved by Wei and Xu [1999], and they showed that there exist no positive solutions of (7) with $0<\tau<\frac{n+2m}{n-2m}$. Moreover, for $\alpha=2$, the problem is the so-called Yamabe problem, and the radial symmetry of solutions was discussed by Gidas, Ni and Nirenberg [Gidas et al. 1981].

In this paper, we show that system (1) is equivalent to integral system (4). By the discussion of the symmetry, monotonicity and regularity of positive solutions of integral system (4), we are able to perform the classification of positive solutions to system (1).

Throughout the paper, we use the following notation:

$$\Pi_1=\bigl\{f(x)\mid x\in\mathbb{R}^n,\ f\in L^{s_{11}}(\mathbb{R}^n)\cap L^{s_{21}}(\mathbb{R}^n)\cap L^{k_0}(\mathbb{R}^n)\bigr\},$$
$$\Pi_2=\bigl\{f(x)\mid x\in\mathbb{R}^n,\ f\in L^{s_{12}}(\mathbb{R}^n)\cap L^{s_{22}}(\mathbb{R}^n)\cap L^{k_0}(\mathbb{R}^n)\bigr\},$$

where $s_{1i}=n(p_i-1)/\alpha$, $s_{2i}=n(q_i-1)/\alpha$, $i=1,2$, and $k_0=n(p_3+p_4-1)/\alpha=n(q_3+q_4-1)/\alpha$ with $n/(n-\alpha)<p_i, q_i, p_3+p_4, q_3+q_4<\infty$, and $p_3+p_4=q_3+q_4$.

We are now in a position to state our main results.

Theorem 1.1. *Assume that $\lambda_i, \mu_i, \beta_i\ge 0$ $(i=1,2)$, and they are not equal to zero simultaneously. Let (u,v) be a pair of solutions to system* (4) *with $u\in\Pi_1$, $v\in\Pi_2$. Then $u, v\in L^s(\mathbb{R}^n)\cap L^\infty(\mathbb{R}^n)$ for any $\frac{n}{n-\alpha}<s<\infty$. Furthermore, $u, v\in C^\infty$.*

Theorem 1.2. *Assume that $\lambda_i, \mu_i, \beta_i\ge 0$ $(i=1,2)$ and they are not equal to zero at the same time. Let $(u,v)\in\Pi_1\times\Pi_2$ be a pair of solutions to system* (4). *Then u and v are radially symmetric and decreasing about some point.*

For system (4) with critical exponents, i.e., $p_1=p_2=q_1=q_2=p_3+p_4=q_3+q_4=\frac{n+\alpha}{n-\alpha}$, we have:

Theorem 1.3. *Let $(u,v)\in L^{2n/(n-\alpha)}(\mathbb{R}^n)\times L^{2n/(n-\alpha)}(\mathbb{R}^n)$ be a pair of positive solutions to system* (4) *with $\lambda_i, \mu_i, \beta_i\ge 0$ $(i=1,2)$ but not equal to zero at the same time. If $p_1=p_2=q_1=q_2=p_3+p_4=q_3+q_4=\frac{n+\alpha}{n-\alpha}$, then $u, v\in L^s(\mathbb{R}^n)\cap L^\infty(\mathbb{R}^n)$ for any $\frac{n}{n-\alpha}<s<\infty$, and $u, v\in C^\infty$. Moreover, u and v are radially symmetric and decreasing about some point, and u, v must be of the following forms*:

$$u(x)=\left(\frac{c_1}{d+|x-\bar{x}|^2}\right)^{(n-\alpha)/2},\quad v(x)=\left(\frac{c_2}{d+|x-\bar{x}|^2}\right)^{(n-\alpha)/2}, \tag{9}$$

where $\bar{x}\in\mathbb{R}^n$, $c_1, c_2>0$, $d>0$ and satisfy the conditions

$$C_\alpha^2 c_1^{(n-\alpha)/2} = \lambda_1 c_1^{(n+\alpha)/2} + \mu_1 c_2^{(n+\alpha)/2} + \beta_1 \big(c_1^{p_3} c_2^{p_4}\big)^{(n-\alpha)/2},$$
$$C_\alpha^2 c_2^{(n-\alpha)/2} = \lambda_2 c_1^{(n+\alpha)/2} + \mu_2 c_2^{(n+\alpha)/2} + \beta_2 \big(c_1^{q_3} c_2^{q_4}\big)^{(n-\alpha)/2}.$$

Theorem 1.4. *System* (1) *is equivalent to integral system* (4).

Combining our symmetry and equivalence results with the known results on nonexistence of positive solutions in the subcritical case (see [Dancer et al. 2010; de Figueiredo and Sirakov 2005]), we can obtain results on the nonexistence of positive solutions (u, v) of system (1) with some suitable conditions.

Theorem 1.5. (i) *Suppose* $n \ge 3$, $\alpha = 2$, $\lambda_i = \mu_i = 1$ *and* $\beta_i = 0$ *for* $i = 1, 2$, *and*

$$\frac{n}{n-2} < p_1, q_2 < \frac{n+2}{n-2}, \quad p_2 = \frac{p_1(q_2-1)}{p_1-1}, \quad q_1 = \frac{q_2(p_1-1)}{q_2-1}.$$

Then system (1) *has no positive solutions* (u, v) *satisfying* $u \in L^{n(p_1-1)/2}(\mathbb{R}^n) \cap L^{n(q_1-1)/2}(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ *and* $v \in L^{n(p_2-1)/2}(\mathbb{R}^n) \cap L^{n(q_2-1)/2}(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$.

(ii) *Assume that* $n \ge 3$, $\alpha = 2$ *and* $\lambda_i, \mu_i > 0$, $\beta_i \ge 0$, $\beta_i \not\equiv 0$, *and* p_j, q_j *satisfy*

$$\frac{n}{n-2} < p_1, q_2 < \frac{n+2}{n-2} \quad \textit{with} \quad p_2 = \frac{p_1(q_2-1)}{p_1-1}, \quad q_1 = \frac{q_2(p_1-1)}{q_2-1},$$

and

$$\frac{p_3}{p_1} + \frac{p_4}{p_2} = \frac{q_3}{q_1} + \frac{q_4}{q_2} = 1$$

with $0 \le p_3 \le p_1$, $0 \le p_4 \le p_2$, $0 \le q_3 \le q_1$, $0 \le q_4 \le q_2$, $p_3 + p_4 = q_3 + q_4$. *Then system* (1) *has no positive solutions* (u, v) *satisfying* $u \in \Pi_1 \cap L^\infty(\mathbb{R}^n)$ *and* $v \in \Pi_2 \cap L^\infty(\mathbb{R}^n)$.

(iii) *Assume that* $n = 3$, $\alpha = 2$, $\lambda_2 = \mu_1 = 0$, $\beta_1 = \beta_2 > -\sqrt{\lambda_1 \mu_2}$. *Then system* (1) *has no positive solutions* (u, v) *satisfying* $u \in \Pi_1 \cap L^\infty(\mathbb{R}^n)$ *and* $v \in \Pi_2 \cap L^\infty(\mathbb{R}^n)$, *where* $p_1 = q_2 = 3$, $p_3 = q_4 = 1$, $p_4 = q_3 = 2$.

Remark 1.6. We can show that the results above hold for the more general system

$$\begin{cases} (-\Delta)^{\alpha/2} u = \lambda_1 u^{p_1} + \mu_1 v^{p_2} + \beta_1 u^{p_3} v^{p_4}, \\ (-\Delta)^{\kappa/2} v = \lambda_2 u^{q_1} + \mu_2 v^{q_2} + \beta_2 u^{q_3} v^{q_4} \end{cases} \tag{10}$$

in $\mathbb{R}^n$, where $0 < \alpha, \kappa < n$, $\lambda_i, \mu_i, \beta_i \ge 0$ $(i = 1, 2)$. That is, if

$$p_1, p_2, p_3+p_4 > \frac{n}{n-\alpha}, \quad q_1, q_2, q_3+q_4 > \frac{n}{n-\kappa}, \quad s_{1i} = \frac{n(p_i-1)}{\alpha}, \quad s_{2i} = \frac{n(q_i-1)}{\kappa}$$

for $i = 1, 2$, and $u, v \in L^{k_0}(\mathbb{R}^n) \cap L^{k_1}(\mathbb{R}^n)$, where

$$k_0 = \frac{n(p_3 + p_4 - 1)}{\alpha}, \quad k_1 = \frac{n(q_3 + q_4 - 1)}{\kappa},$$

then the results of Theorem 1.1, Theorem 1.2 and Theorem 1.4 are still valid.

We remark that a more general system of m equations has been discussed by Chen and Li [2009a]. That is,

$$\begin{cases} u_j(x) = \displaystyle\int_{\mathbb{R}^n} \frac{f_j(u(y))}{|x-y|^{n-\alpha}}\,dy, \quad j = 1, 2, \dots, m, \\ u(x) = (u_1(x), u_2(x), \dots, u_m(x)), \end{cases} \quad \text{in } \mathbb{R}^n \tag{11}$$

where $f_j(u) \geq 0$ are continuous real-valued functions and homogeneous of degree $\frac{n+\alpha}{n-\alpha}$, and satisfy $\partial f_i/\partial u_j \geq 0$ for $i = 1, 2, \dots, m$. System (11) includes only the critical exponent case of system (4). It was shown in [Chen and Li 2009a] any positive solutions of (11) are radially symmetric under the assumptions $u_j \in L^\infty_{\text{loc}}(\mathbb{R}^n)$. Furthermore, based on the Kelvin transformation and the results in [Chen et al. 2006], any positive solutions of (11) must be the form of (8). In our proof of Theorem 1.3, a key calculus lemma due to Li and Zhu [1995] and the Kelvin transformation are used to show that all positive solutions of (4) are given by (9).

The main difficulty in our proof is the lack of a maximum principle for the higher-order fractional Laplace operator. Theorem 1.4 says that system (1) is equivalent to the integral system (4), which is helpful for our discussion since we can use the method of moving planes in integral forms (see [Chen et al. 2006]) to discuss the radial symmetry and monotonicity of positive solution of the integral system (4). Furthermore, the regularity of solutions to system (4) is proved by the regularity lifting lemma introduced in [Chen and Li 2010; Ma et al. 2011].

The paper is organized as follows. In Section 2, we prove the regularity of solutions of system (4) (Theorem 1.1). The radially symmetric property and monotonicity of solutions are studied in Section 3 (Theorem 1.2). In Section 4, positive solutions of system (4) with critical exponents are classified. Namely, Theorem 1.3 is proved. In Section 5, we obtain some nonexistence results by proving Theorems 1.4 and 1.5.

Throughout the paper, we always assume that $\lambda_i, \mu_i, \beta_i \geq 0$ $(i = 1, 2)$ and they are not equal to zero simultaneously. Moreover, for convenience of presentation we shall use c, c_1, C, etc. for a suitable positive constants unless indicated otherwise.

2. Regularity

In this section, we prove the regularity of solutions to system (4). To this end, we need the following regularity lifting lemma (see [Chen and Li 2010; Ma et al. 2011]). An earlier version was introduced in [Chen and Li 2005].

Let V be a topological vector space. Suppose there are two extended norms (i.e., the norm of an element in V might be infinity) defined on V,

$$\|\cdot\|_X, \|\cdot\|_Y : V \to [0, \infty].$$

Let

$$X := \{f \in V : \|f\|_X < \infty\} \quad \text{and} \quad Y := \{f \in V : \|f\|_Y < \infty\}.$$

Lemma 2.1. *Let T be a contraction map from X into itself and from Y into itself. Assume that for any $f \in X$ there exists a function $g \in Z := X \cap Y$ such that $f = Tf + g$ in X. Then $f \in Z$.*

We also need an equivalent form of the HLS inequality (see [Chen and Li 2005; 2010]): let $C(n, \alpha, p)$ be a uniform positive constant and define

$$Tf(x) = \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-\alpha}}\, dy.$$

Assume that $f \in L^p(\mathbb{R}^n)$ for $\frac{n}{n-\alpha} < p < \infty$. Then

$$\|Tf\|_{L^p(\mathbb{R}^n)} \le C(n, \alpha, p)\|f\|_{L^{\frac{np}{n+\alpha p}}(\mathbb{R}^n)}. \tag{12}$$

Denote

$$u_R(x) = \begin{cases} u(x), & |u(x)| > R, \\ 0, & |u(x)| \le R. \end{cases}$$

Assume that $\phi \in L^r(\mathbb{R}^n)$, $\varphi \in L^s(\mathbb{R}^n)$ for $\frac{n}{n-\alpha} < r, s < \infty$. Define

$$T_1(\phi, \varphi) = \int_{\mathbb{R}^n} \frac{\lambda_1 u_R^{p_1-1}(y)}{|x-y|^{n-\alpha}} \phi(y)\, dy + \int_{\mathbb{R}^n} \frac{\mu_1 v_R^{p_2-1}(y) + \beta_1 u_R^{p_3}(y) v_R^{p_4-1}(y)}{|x-y|^{n-\alpha}} \varphi(y)\, dy,$$

$$T_2(\phi, \varphi) = \int_{\mathbb{R}^n} \frac{\mu_2 v_R^{q_2-1}(y)}{|x-y|^{n-\alpha}} \varphi(y)\, dy + \int_{\mathbb{R}^n} \frac{\lambda_2 u_R^{q_1-1}(y) + \beta_2 v_R^{q_4}(y) u_R^{q_3-1}(y)}{|x-y|^{n-\alpha}} \phi(y)\, dy.$$

Let $u_b(x) = u(x) - u_R(x)$, and

$$f_R(x) = \int_{\mathbb{R}^n} \frac{\mu_1 v_b^{p_2-1}(y) + \beta_1 u_b^{p_3}(y) v_b^{p_4-1}(y)}{|x-y|^{n-\alpha}} v(y)\, dy + \int_{\mathbb{R}^n} \frac{\lambda_1 u_b^{p_1-1}(y)}{|x-y|^{n-\alpha}} u(y)\, dy,$$

$$g_R(x) = \int_{\mathbb{R}^n} \frac{\lambda_2 u_b^{q_1-1}(y) + \beta_2 v_b^{q_4}(y) u_b^{q_3-1}(y)}{|x-y|^{n-\alpha}} u(y)\, dy + \int_{\mathbb{R}^n} \frac{\mu_2 v_b^{q_2-1}(y)}{|x-y|^{n-\alpha}} v(y)\, dy.$$

Denote the norm in the cross product space $L^r(\mathbb{R}^n) \times L^s(\mathbb{R}^n)$ by

$$\|(u, v)\|_{r\times s} = \|u\|_r + \|v\|_s,$$

and define the mapping $T : L^r(\mathbb{R}^n) \times L^s(\mathbb{R}^n) \to L^r(\mathbb{R}^n) \times L^s(\mathbb{R}^n)$ by

$$T(\phi, \varphi) = \big(T_1(\phi, \varphi), T_2(\phi, \varphi)\big).$$

Throughout the paper, we use the notation $\|u\|_s = \|u\|_{L^s(\mathbb{R}^n)}$.

Consider the equation

$$(\phi, \varphi) = T(\phi, \varphi) + (f_R, g_R). \tag{13}$$

Notice that there is no intersection between the supports of u_R, v_R and u_b, v_b, so (u, v) is a pair of solutions of (13).

Proof of Theorem 1.1. The proof is divided into three steps.

Step 1. Firstly, we show $u, v \in L^s(\mathbb{R}^n)$ for all $s > \frac{n}{n-\alpha}$. To this end, we show that

(i) T is a contracting map from $L^s(\mathbb{R}^n) \times L^s(\mathbb{R}^n)$ to itself for R large enough;

(ii) f_R and g_R belong to $L^s(\mathbb{R}^n)$.

We first show (i). For any $\phi, \varphi \in L^s(\mathbb{R}^n)$, using the HLS inequality (12) and the Minkowski inequality, we have

$$\|T_1(\phi,\varphi)\|_s \le C(n,\alpha,\gamma)\big(\lambda_1\big\|u_R^{p_1-1}\phi\big\|_\theta + \mu_1\big\|{v_R}^{p_2-1}\varphi\big\|_\theta + \beta_1\big\|{u_R}^{p_3}{v_R}^{p_4-1}\varphi\big\|_\theta\big), \tag{14}$$

where $\theta = \frac{ns}{n+\alpha s}$. By the Hölder inequality, we have

$$\big\|u_R^{p_1-1}\phi\big\|_\theta \le \|u_R\|_{s_{11}}^{p_1-1}\|\phi\|_s, \qquad \big\|{v_R}^{p_2-1}\varphi\big\|_\theta \le \|v_R\|_{s_{12}}^{p_2-1}\|\varphi\|_s, \tag{15}$$

where $s_{1j} = n(p_j-1)/\alpha$, $j = 1, 2$, and

$$\begin{aligned}\big\|{u_R}^{p_3}{v_R}^{p_4-1}\varphi\big\|_\theta &\le \left(\int_{\mathbb{R}^n} u_R^{p_3t_1\theta}(y)\,dy\right)^{\frac{1}{t_1\theta}}\left(\int_{\mathbb{R}^n} v_R^{(p_4-1)t_2\theta}(y)\,dy\right)^{\frac{1}{t_2\theta}}\left(\int_{\mathbb{R}^n}\varphi^{t_3\theta}(y)\,dy\right)^{\frac{1}{t_3\theta}}\\ &= \|u_R\|_{k_0}^{p_3}\|v_R\|_{k_0}^{p_4-1}\|\varphi\|_s.\end{aligned} \tag{16}$$

In the above inequality we have chosen $t_3 = (n+s\alpha)/n > 1$, so we take $1/t_1 + 1/t_2 = s\alpha/(n+s\alpha)$ with

$$k_0 = t_1p_3\theta = t_2(p_4-1)\theta,$$

and then

$$\frac{p_3}{k_0} + \frac{p_4-1}{k_0} = \frac{\alpha}{n}.$$

Substituting (15) and (16) into (14), we deduce that

$$\|T_1(\phi,\varphi)\|_s \le c\|u_R\|_{s_{11}}^{p_1-1}\|\phi\|_s + c\big(\|v_R\|_{s_{12}}^{p_2-1} + \|u_R\|_{k_0}^{p_3}\|v_R\|_{k_0}^{p_4-1}\big)\|\varphi\|_s. \tag{17}$$

Since $u \in L^{s_{11}}(\mathbb{R}^n) \cap L^{k_0}(\mathbb{R}^n)$, $v \in L^{s_{12}}(\mathbb{R}^n) \cap L^{k_0}(\mathbb{R}^n)$, we may choose R large enough such that

$$\|u_R\|_{s_{11}}^{p_2-1} \le \tfrac{1}{4}, \qquad \|v_R\|_{s_{12}}^{p_2-1} + \|u_R\|_{k_0}^{p_3}\|v_R\|_{k_0}^{p_4-1} \le \tfrac{1}{4}.$$

Hence, from (17) we obtain

$$\|T_1(\phi,\varphi)\|_s \le \tfrac{1}{4}\big(\|\phi\|_s + \|\varphi\|_s\big). \tag{18}$$

Similarly, we have

$$\|T_2(\phi,\varphi)\|_s \le \tfrac{1}{4}\big(\|\phi\|_s + \|\varphi\|_s\big). \tag{19}$$

Combining (18) and (19), one obtains

$$\|T(\phi,\varphi)\|_{s\times s} \le \tfrac{1}{2}\big(\|\phi\|_s + \|\varphi\|_s\big).$$

It turns out that T is the contracting map from $L^s(\mathbb{R}^n)\times L^s(\mathbb{R}^n)$ to itself.

(ii) Next we estimate f_R and g_R. We write

$$\begin{aligned} f_R(x) &= \int_{\mathbb{R}^n} \frac{\mu_1 v_b^{p_2-1}(y) + \beta_1 u_b^{p_3}(y) v_b^{p_4-1}(y)}{|x-y|^{n-\alpha}} v(y)\,dy + \int_{\mathbb{R}^n} \frac{\lambda_1 u_b^{p_1-1}(y)}{|x-y|^{n-\alpha}} u(y)\,dy \\ &=: J_1 + J_2. \end{aligned} \tag{20}$$

For any $s > \frac{n}{n-\alpha}$, we apply the HLS inequality, Minkowski inequality and Hölder inequality to get

$$\begin{aligned} \|J_1\|_s &\le c\big\|{v_b}^{p_2-1} v\big\|_\theta + c\big\|{u_b}^{p_3}{v_b}^{p_4-1} v\big\|_\theta \\ &\le c\|v_b\|_{k_1}^{p_2-1}\|v\|_{k_2} + c\|u_b\|_{k_3}^{p_3}\|v_b\|_{k_4}^{p_4-1}\|v\|_{k_5}, \end{aligned} \tag{21}$$

and

$$\|J_2\|_s \le c\big\|u_b^{p_1-1} u\big\|_\theta \le c\|u_b\|_{k_6}^{p_1-1}\|u\|_{k_7}, \tag{22}$$

where

$$\frac{p_2-1}{k_1} + \frac{1}{k_2} = \frac{p_3}{k_3} + \frac{p_4-1}{k_4} + \frac{1}{k_5} = \frac{p_1-1}{k_6} + \frac{1}{k_7} = \frac{n+\alpha s}{ns} = \frac{1}{s} + \frac{\alpha}{n}.$$

Since v_b, u_b are bounded, k_1, k_3, k_4, k_6 can be chosen arbitrarily. Notice that $\frac{n}{n-\alpha} < p_3 + p_4 = q_3 + q_4$, so it follows that $k_0 = n(p_3+p_4-1)/\alpha = n(q_3+q_4-1)/\alpha > n/(n-\alpha)$. In view of $u, v \in L^{k_0}(\mathbb{R}^n)$, we may choose $k_2 = k_5 = k_7 = k_0$ such that

$$\begin{aligned} \frac{1}{s} = \frac{p_2-1}{k_1} + \frac{1}{k_0} - \frac{\alpha}{n} &= \frac{p_2-1}{k_1} + \frac{n-\alpha k_0}{nk_0} \\ &= \frac{p_3}{k_3} + \frac{p_4-1}{k_4} + \frac{n-\alpha k_0}{nk_0} \\ &= \frac{p_1-1}{k_6} + \frac{n-\alpha k_0}{nk_0}. \end{aligned}$$

Now, letting $k_1, k_3, k_4, k_6 \to \infty$, the previous equation implies that

$$s \to \frac{nk_0}{n-\alpha k_0}.$$

We conclude that $f_R \in L^{nk_0/(n-\alpha k_0)-\epsilon}(\mathbb{R}^n)$ for any small $\epsilon > 0$. Obviously, $nk_0/(n-\alpha k_0) > k_0$. Similarly, we can show $g_R \in L^{nk_0/(n-\alpha k_0)-\epsilon}(\mathbb{R}^n)$.

By Lemma 2.1, if $n \le \alpha k_0$, we are done. If $n > \alpha k_0$, we repeat the above process, and after a few steps, we obtain

$$u, v \in L^s(\mathbb{R}^n), \qquad \frac{n}{n-\alpha} < s < \infty.$$

Step 2. We show $u, v \in L^\infty(\mathbb{R}^n)$. We split $u(x)$ into two parts, i.e., $u(x)$ can be written as

$$\begin{aligned} u(x) &= \int_{B_1(x)} \frac{\lambda_1 u^{p_1}(y)+\mu_1 v^{p_2}(y)+\beta_1 u^{p_3}(y)v^{p_4}(y)}{|x-y|^{n-\alpha}}\,dy \\ &\qquad + \int_{\mathbb{R}^n\setminus B_1(x)} \frac{\lambda_1 u^{p_1}(y)+\mu_1 v^{p_2}(y)+\beta_1 u^{p_3}(y)v^{p_4}(y)}{|x-y|^{n-\alpha}}\,dy \\ &=: J_3 + J_4. \end{aligned}$$

We estimate J_3 and J_4 separately. First consider J_4. Since $1/|x-y|^{n-\alpha} < 1$, and $u, v \in L^s(\mathbb{R}^n)$ for any $\frac{n}{n-\alpha} < s$, according to the assumptions that $\frac{n}{n-\alpha} < p_1, p_2, p_3 + p_4$, and using the Hölder inequality, we have

$$\begin{aligned} J_4 &\le c\int_{\mathbb{R}^n\setminus B_1(x)} u^{p_1}(y)\,dy + c\int_{\mathbb{R}^n\setminus B_1(x)} v^{p_2}(y)\,dy + c\int_{\mathbb{R}^n\setminus B_1(x)} u^{p_3}(y)v^{p_4}(y)\,dy \\ &< \infty. \end{aligned}$$

Next, we compute J_3:

$$\begin{aligned} J_3 &\le c\biggl(\int_{B_1(x)} \frac{1}{|x-y|^{(n-\alpha)p}}\,dy\biggr)^{1/p} \biggl(\int_{B_1(x)} |u(y)|^{p_1 p/(p-1)}\,dy\biggr)^{(p-1)/p} \\ &\quad + c\biggl(\int_{B_1(x)} \frac{1}{|x-y|^{(n-\alpha)p}}\,dy\biggr)^{1/p} \biggl(\int_{B_1(x)} |v(y)|^{p_2 p/(p-1)}\,dy\biggr)^{(p-1)/p} \\ &\quad + c\biggl(\int_{B_1(x)} \frac{1}{|x-y|^{(n-\alpha)p}}\,dy\biggr)^{1/p} \biggl(\int_{B_1(x)} \bigl(|u^{p_3}(y)v^{p_4}(y)|\bigr)^{p/(p-1)}\,dy\biggr)^{(p-1)/p}. \end{aligned}$$

Choose the constant p such that $(n-\alpha)p < n$, and then

$$\biggl(\int_{B_1(x)} \frac{1}{|x-y|^{(n-\alpha)p}}\,dy\biggr)^{1/p} < C.$$

Since $u, v \in L^s(\mathbb{R}^n)$ for any $\frac{n}{n-\alpha} < s < \infty$, we get

$$\biggl(\int_{B_1(x)} |u(y)|^{p_1 p/(p-1)}\,dy\biggr)^{(p-1)/p} < C, \qquad \biggl(\int_{B_1(x)} |v(y)|^{p_2 p/(p-1)}\,dy\biggr)^{(p-1)/p} < C,$$

and by the Hölder inequality, we obtain

$$\left(\int_{B_1(x)} |u^{p_3}(y)v^{p_4}(y)|^{\frac{p}{p-1}}\,dy\right)^{\frac{p-1}{p}} \le \left(\int_{B_1(x)} |u(y)|^{p_3 l_1 \frac{p}{p-1}}\,dy\right)^{\frac{p-1}{l_1 p}} \left(\int_{B_1(x)} |v(y)|^{p_4 l_2 \frac{p}{p-1}}\,dy\right)^{\frac{p-1}{l_2 p}} < C,$$

where $l_1, l_2 > 1$ and $1/l_1 + 1/l_2 = 1$. (We may choose $l_1 = (p_3 + p_4)/p_3$ and $l_2 = (p_3 + p_4)/p_4$.) So $u \in L^\infty(\mathbb{R}^n)$. Arguing as above, it also follows that $v \in L^\infty(\mathbb{R}^n)$.

Step 3. Using the usual bootstrap method, as in [Li 2004], we conclude $u, v \in C^\infty(\mathbb{R}^n)$. □

3. Radial symmetry and monotonicity

In this section, we use the method of moving plane in integral form to prove Theorem 1.2. The moving plane method in integral form used here was introduced by Chen, Li and Ou [2006] and exploits global properties of integral equations instead of using the amount of local properties of differential operators as the traditional moving plane method (e.g., see [Guo and Liu 2008; de Figueiredo and Sirakov 2005; Liu et al. 2006; Zhang 2007; Wei and Xu 1999; Gidas et al. 1981]).

We first deduce two representation formulas related to $u(x)$ and $v(x)$, respectively. Let λ be a real number. Define

$$\Sigma_\lambda = \{x = (x_1 \cdots x_n) \mid x_1 \ge \lambda\},$$

and set

$$x^\lambda = (2\lambda - x_1, x_2, \ldots, x_n), \quad u_\lambda(x) = u(x^\lambda) \quad \text{and} \quad v_\lambda(x) = v(x^\lambda).$$

For convenience, we set $Q_y(u, v) = \lambda_1 u^{p_1}(y) + \mu_1 v^{p_2}(y) + \beta_1 u^{p_3}(y) v^{p_4}(y)$ and $K_y(u, v) = \lambda_2 u^{q_1}(y) + \mu_2 v^{q_2}(y) + \beta_2 u^{q_3}(y) v^{q_4}(y)$. In view of (4), we have

$$\begin{aligned} u_\lambda(x) &= \int_{\mathbb{R}^n} \frac{Q_y(u, v)}{|x^\lambda - y|^{n-\alpha}}\,dy \\ &= \int_{\Sigma_\lambda} \frac{Q_y(u, v)}{|x^\lambda - y|^{n-\alpha}}\,dy + \int_{\mathbb{R}^n \setminus \Sigma_\lambda} \frac{Q_y(u, v)}{|x^\lambda - y|^{n-\alpha}}\,dy \\ &= \int_{\Sigma_\lambda} \frac{Q_y(u, v)}{|x^\lambda - y|^{n-\alpha}}\,dy + \int_{\Sigma_\lambda} \frac{Q_y(u_\lambda, v_\lambda)}{|x^\lambda - y^\lambda|^{n-\alpha}}\,dy. \end{aligned}$$

We also have

$$v_\lambda(x) = \int_{\Sigma_\lambda} \frac{K_y(u, v)}{|x^\lambda - y|^{n-\alpha}}\,dy + \int_{\Sigma_\lambda} \frac{K_y(u_\lambda, v_\lambda)}{|x^\lambda - y^\lambda|^{n-\alpha}}\,dy.$$

Noting that $|x^\lambda - y^\lambda| = |x-y|$, it is easy to see that

$$
\begin{aligned}
(23)\quad & u_\lambda(x)-u(x) \\
&= \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right)(Q_y(u_\lambda, v_\lambda)-Q_y(u,v))\,dy \\
&= \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right)\big[\lambda_1\big(u_\lambda^{p_1}(y)-u^{p_1}(y)\big)+\mu_1\big(v_\lambda^{p_2}(y)-v^{p_2}(y)\big) \\
&\qquad\qquad +\beta_1\big(u_\lambda^{p_3}(y)v_\lambda^{p_4}(y)-u^{p_3}(y)v^{p_4}(y)\big)\big]\,dy
\end{aligned}
$$

and

$$
\begin{aligned}
(24)\quad & v_\lambda(x)-v(x) \\
&= \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right)(K_y(u_\lambda, v_\lambda)-K_y(u,v))\,dy \\
&= \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right)\big[\lambda_2\big(u_\lambda^{q_1}(y)-u^{q_1}(y)\big)+\mu_2\big(v_\lambda^{q_2}(y)-v^{q_2}(y)\big) \\
&\qquad\qquad +\beta_2\big(u_\lambda^{q_3}(y)v_\lambda^{q_4}(y)-u^{q_3}(y)v^{q_4}(y)\big)\big]\,dy.
\end{aligned}
$$

The next lemma shows the plane can start moving from $x_1 = -\infty$ to the right.

Lemma 3.1. *Let* $(u, v) \in \Pi_1 \times \Pi_2$ *be a pair of positive solutions of* (4). *Then, for* λ *sufficiently negative*,

$$
(25)\qquad u(x) \ge u_\lambda(x) \quad \textit{and} \quad v(x) \ge v_\lambda(x) \quad \textit{for all } x \in \Sigma_\lambda.
$$

Proof. Define

$$
\Sigma_\lambda^u = \{x \in \Sigma_\lambda \mid u(x) < u_\lambda(x)\} \quad \text{and} \quad \Sigma_\lambda^v = \{x \in \Sigma_\lambda \mid v(x) < v_\lambda(x)\}.
$$

Let Σ_λ^c be the complement of Σ_λ. From (23) and the mean value theorem, we have

$$
\begin{aligned}
(26)\quad & u_\lambda(x)-u(x) \\
&= \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right) \\
&\qquad \times\big[\lambda_1(u_\lambda^{p_1}(y)-u^{p_1}(y))+\mu_1(v_\lambda^{p_2}(y)-v^{p_2}(y)) \\
&\qquad\qquad +\beta_1\big(u_\lambda^{p_3}(y)(v_\lambda^{p_4}(y)-v^{p_4}(y))+v^{p_4}(y)(u_\lambda^{p_3}(y)-u^{p_3}(y))\big)\big]\,dy \\
&\le \int_{\Sigma_\lambda}\left(\frac{1}{|x-y|^{n-\alpha}}-\frac{1}{|x^\lambda-y|^{n-\alpha}}\right) \\
&\qquad \times\big[p_1\lambda_1\phi_1^{p_1-1}(u)(u_\lambda(y)-u(y))+p_2\mu_1\phi_2^{p_2-1}(v)(v_\lambda(y)-v(y)) \\
&\qquad\qquad +p_4\beta_1u_\lambda^{p_3}(y)\phi_4^{p_4-1}(v)(v_\lambda(y)-v(y)) \\
&\qquad\qquad +p_3\beta_1v^{p_4}(y)\phi_3^{p_3-1}(u)(u_\lambda(y)-u(y))\big]\,dy,
\end{aligned}
$$

where $u(y) \le \phi_i(u) \le u_\lambda(y)$, $i = 1, 3$, on Σ_λ^u, and $v(y) \le \phi_j(v) \le v_\lambda(y)$, $j = 2, 4$ on Σ_λ^v. It follows that we can write

$$u_\lambda(x) - u(x) \le c(I_1 + I_2 + I_3 + I_4),$$

where

$$I_1 := \int_{\Sigma_\lambda^u} \frac{u_\lambda^{p_1-1}(y)}{|x-y|^{n-\alpha}}(u_\lambda(y)-u(y))\,dy, \quad I_2 := \int_{\Sigma_\lambda^v} \frac{v_\lambda^{p_2-1}(y)}{|x-y|^{n-\alpha}}(v_\lambda(y)-v(y))\,dy,$$

$$I_3 := \int_{\Sigma_\lambda^v} \frac{u_\lambda^{p_3}(y)v_\lambda^{p_4-1}(y)}{|x-y|^{n-\alpha}}(v_\lambda(y)-v(y))\,dy,$$

$$I_4 := \int_{\Sigma_\lambda^u} \frac{v^{p_4}(y)u_\lambda^{p_3-1}(y)}{|x-y|^{n-\alpha}}(u_\lambda(y)-u(y))\,dy.$$

Using the HLS inequality and the Hölder inequality, we get

$$\left(\int_{\Sigma_\lambda^u} |I_1|^\gamma\right)^{1/\gamma} \le C(n,\alpha,\gamma)\left\|{u_\lambda}^{p_1-1}(u_\lambda - u)\right\|_{L^\theta(\Sigma_\lambda^u)} \tag{27}$$

for any $\frac{n}{n-\alpha} < \gamma < \infty$ and $\theta = \frac{n\gamma}{n+\alpha\gamma}$. Let $m_1 = \frac{n+\alpha\gamma}{\alpha\gamma} > 1$ and $m_2 = \frac{n+\alpha\gamma}{n} > 1$. Thus, we invoke the Hölder inequality to obtain

$$\begin{aligned}
&\left\{\int_{\Sigma_\lambda^u}\left[{u_\lambda}^{p_1-1}(y)(u_\lambda(y)-u(y))\right]^\theta dy\right\}^{1/\theta} \\
&\le \left\{\left[\int_{\Sigma_\lambda^u}(u_\lambda(y))^{\theta(p_1-1)m_1}\,dy\right]^{1/m_1}\left[\int_{\Sigma_\lambda^u}(u_\lambda(y)-u(y))^{\theta m_2}\,dy\right]^{1/m_2}\right\}^{1/\theta} \\
&= \|u_\lambda\|_{L^{s_{11}}_{(\Sigma_\lambda^u)}}^{p_1-1}\|u_\lambda - u\|_{L^\gamma_{(\Sigma_\lambda^u)}},
\end{aligned} \tag{28}$$

where $s_{11} = n(p_1 - 1)/\alpha$. Substituting (28) into (27), we get

$$\left(\int_{\Sigma_\lambda^u} |I_1|^\gamma\right)^{1/\gamma} \le C(n,\alpha,\gamma)\|u_\lambda\|_{L^{s_{11}}_{(\Sigma_\lambda^u)}}^{p_1-1}\|u_\lambda - u\|_{L^\gamma_{(\Sigma_\lambda^u)}}. \tag{29}$$

Similarly, one has

$$\|I_2\|_{L^\gamma_{(\Sigma_\lambda^u)}} \le C(n,\alpha,\gamma)\|v_\lambda\|_{L^{s_{12}}_{(\Sigma_\lambda^v)}}^{p_2-1}\|v_\lambda - v\|_{L^\gamma_{(\Sigma_\lambda^v)}}, \tag{30}$$

where $s_{12} = n(p_2 - 1)/\alpha$.

Next, we estimate I_3 and I_4. By the HLS inequality we have

$$\|I_3\|_{L^\gamma_{(\Sigma_\lambda^u)}} \le C(n,\alpha,\gamma)\left\|{u_\lambda}^{p_3}{v_\lambda}^{p_4-1}(v_\lambda - v)\right\|_{L^\theta_{(\Sigma_\lambda^v)}}. \tag{31}$$

Letting $1/t_1 + 1/t_2 + 1/t_3 = 1$ for $t_1 > 1$, it follows that

$$
\begin{aligned}
(32)\quad & \left\| {u_\lambda}^{p_3} {v_\lambda}^{p_4-1}(v_\lambda - v) \right\|_{L^\theta_{(\Sigma^v_\lambda)}} \\
& \le \left[\int_{\Sigma^v_\lambda} u_\lambda^{p_3\theta t_1}(y)\,dy \right]^{\frac{1}{t_1\theta}} \left[\int_{\Sigma^v_\lambda} v_\lambda^{(p_4-1)\theta t_2}(y)\,dy \right]^{\frac{1}{t_2\theta}} \left[\int_{\Sigma^v_\lambda} (v_\lambda(y)-v(y))^{\theta t_3}\,dy \right]^{\frac{1}{t_3\theta}} \\
& \le \|u_\lambda\|^{p_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}.
\end{aligned}
$$

Arguing as Section 2, we choose $t_3 = (n+\alpha\gamma)/n > 1$, $t_1 = (n+\alpha\gamma)(p_3+p_4-1)/(p_3\alpha\gamma)$ and $t_2 = (n+\alpha\gamma)(p_3+p_4-1)/((p_4-1)\alpha\gamma)$, satisfying $1/t_1+1/t_2+1/t_3 = 1$. Then $k_0 = t_1 p_3\theta = t_2(p_4-1)\theta$. Substituting (32) into (31), we conclude

$$
(33)\qquad \|I_3\|_{L^\gamma_{(\Sigma^u_\lambda)}} \le C(n,\alpha,\gamma)\|u_\lambda\|^{p_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}.
$$

In the same way, one has

$$
(34)\qquad \|I_4\|_{L^\gamma_{(\Sigma^u_\lambda)}} \le C(n,\alpha,\gamma)\|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda\|^{p_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}}.
$$

Now, we compute the norm $L^\gamma(\Sigma^u_\lambda)$ of (26) for any $\frac{n}{n-\alpha} < \gamma < \infty$. Using the Minkowski inequality and combining (29), (30), (33) and (34), we arrive at

$$
\begin{aligned}
(35)\quad & \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
& \le c\|u_\lambda\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^u_\lambda)}} \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + c\|v_\lambda\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^v_\lambda)}} \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \\
& \quad + c\|u_\lambda\|^{p_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} + c\|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda\|^{p_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
& \le c\Big(\|u_\lambda\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^u_\lambda)}} + \|u_\lambda\|^{p_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \Big) \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
& \quad + c\Big(\|v_\lambda\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{p_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} \Big) \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}.
\end{aligned}
$$

Along the same line, noting that $p_3 + p_4 = q_3 + q_4$, we have

$$
\begin{aligned}
(36)\quad \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} & \le c\Big(\|u_\lambda\|^{q_1-1}_{L^{s_{21}}_{(\Sigma^u_\lambda)}} + \|u_\lambda\|^{q_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \Big) \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
& \quad + c\Big(\|v_\lambda\|^{q_2-1}_{L^{s_{22}}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{q_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{q_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} \Big) \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}},
\end{aligned}
$$

where $s_{21} = n(q_1-1)/\alpha$, $s_{22} = n(q_2-1)/\alpha$, $k_0 = n(q_3+q_4-1)/\alpha$.

By adding (35) and (36), we obtain

$$
\begin{aligned}
(37)\quad & \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \\
&\le c\Big(\|u_\lambda\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^u_\lambda)}} + \|u_\lambda\|^{q_1-1}_{L^{s_{21}}_{(\Sigma^u_\lambda)}} + \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda\|^{p_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} + \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u_\lambda\|^{q_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}}\Big) \\
&\qquad \times \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
&\quad + c\Big(\|v_\lambda\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^v_\lambda)}} + \|v_\lambda\|^{q_2-1}_{L^{s_{22}}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{p_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{q_3}_{L^{k_0}_{(\Sigma^v_\lambda)}} \|v_\lambda\|^{q_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}}\Big) \\
&\qquad \times \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \\
&\le c\Big(\|u\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^c_\lambda)}} + \|u\|^{q_1-1}_{L^{s_{21}}_{(\Sigma^c_\lambda)}} + \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u\|^{p_3-1}_{L^{k_0}_{(\Sigma^c_\lambda)}} + \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u\|^{q_3-1}_{L^{k_0}_{(\Sigma^c_\lambda)}}\Big) \\
&\qquad \times \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
&\quad + c\Big(\|v\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^c_\lambda)}} + \|v\|^{q_2-1}_{L^{s_{22}}_{(\Sigma^c_\lambda)}} + \|u\|^{p_3}_{L^{k_0}_{(\Sigma^c_\lambda)}} \|v\|^{p_4-1}_{L^{k_0}_{(\Sigma^c_\lambda)}} + \|u\|^{q_3}_{L^{k_0}_{(\Sigma^c_\lambda)}} \|v\|^{q_4-1}_{L^{k_0}_{(\Sigma^c_\lambda)}}\Big) \\
&\qquad \times \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}.
\end{aligned}
$$

Since $u \in \Pi_1$ and $v \in \Pi_2$, we can choose λ sufficiently negative such that

$$
c\Big(\|u\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^c_\lambda)}} + \|u\|^{q_1-1}_{L^{s_{21}}_{(\Sigma^c_\lambda)}} + \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u\|^{p_3-1}_{L^{k_0}_{(\Sigma^c_\lambda)}} + \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}} \|u\|^{q_3-1}_{L^{k_0}_{(\Sigma^c_\lambda)}}\Big) \le \tfrac{1}{2},
$$
$$
c\Big(\|v\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^c_\lambda)}} + \|v\|^{q_2-1}_{L^{s_{22}}_{(\Sigma^c_\lambda)}} + \|u\|^{p_3}_{L^{k_0}_{(\Sigma^c_\lambda)}} \|v\|^{p_4-1}_{L^{k_0}_{(\Sigma^c_\lambda)}} + \|u\|^{q_3}_{L^{k_0}_{(\Sigma^c_\lambda)}} \|v\|^{q_4-1}_{L^{k_0}_{(\Sigma^c_\lambda)}}\Big) \le \tfrac{1}{2}.
$$

Hence

$$
\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \le \tfrac{1}{2}\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \tfrac{1}{2}\|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}.
$$

This implies that $\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} = \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} = 0$, and therefore Σ^u_λ and Σ^v_λ must be empty. Thus, (25) is proved. □

Next we define

$$
\lambda_0 = \sup\{\lambda \in \mathbb{R} \mid u_\mu(x) \le u(x),\ v_\mu(x) \le v(x) \text{ for all } \mu \le \lambda \text{ and all } x \in \Sigma_\mu\}.
$$

By the regularity of positive solutions to system (4), we observe the fact that u and v are bounded as $|x| \to \infty$. Combining this and noting $u, v > 0$, we conclude $\lambda_0 < \infty$. Thus, we will move the plane to the limiting position to derive symmetry. That is, we have the following lemma.

Lemma 3.2. *Under the assumptions of Theorem 1.2, we have*

$$
(38)\qquad u_{\lambda_0}(x) \equiv u(x) \quad \textit{and} \quad v_{\lambda_0}(x) \equiv v(x) \quad \textit{for all } x \in \Sigma_{\lambda_0}.
$$

Proof. We use argument by contradiction. Assume that there exists a $\lambda_0 < 0$ such that $u(x) \geq u_{\lambda_0}(x)$, and $v_\lambda(x) \geq v_{\lambda_0}(x)$, but $u(x) \not\equiv u_{\lambda_0}(x)$ or $v_\lambda(x) \not\equiv v_{\lambda_0}(x)$ for any $x \in \Sigma_{\lambda_0}$.

We show that the plane can be moved further to the right. More precisely, there exists an ε depending on n, α, and the solution (u, v) itself such that

$$u(x) \geq u_\lambda(x) \quad \text{and} \quad v(x) \geq v_\lambda(x), \quad \text{on } \Sigma_\lambda$$

for $\lambda \in [\lambda_0, \lambda_0 + \varepsilon)$.

In the case of $v(x) \not\equiv v_{\lambda_0}(x)$ on Σ_{λ_0}, from (23) and (24) we obtain $u(x) \not\equiv u_{\lambda_0}(x)$, that is, $u(x) > u_{\lambda_0}(x)$ in the interior of Σ_{λ_0}. Let

$$\overline{\Sigma^u_{\lambda_0}} = \{x \in \Sigma_{\lambda_0} \mid u(x) \leq u_{\lambda_0}(x)\} \quad \text{and} \quad \overline{\Sigma^v_{\lambda_0}} = \{x \in \Sigma_{\lambda_0} \mid v(x) \leq v_{\lambda_0}(x)\}.$$

Obviously, $\overline{\Sigma^u_{\lambda_0}}$ has measure zero, and $\lim_{\lambda\to\lambda_0} \Sigma^u_\lambda \subset \overline{\Sigma^u_{\lambda_0}}$. The same fact holds for that of v. Let $(\Sigma^u_\lambda)^*$ be the reflection of set Σ^u_λ about the plane $x_1 = \lambda$. Similarly to (37), we have

$$\begin{aligned}
(39)\quad & \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \\
&\leq c\Big(\|u_\lambda\|^{p_1-1}_{L^{s_{11}}_{(\Sigma^u_\lambda)}} + \|u_\lambda\|^{q_1-1}_{L^{s_{21}}_{(\Sigma^u_\lambda)}} + \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}}\|u_\lambda\|^{p_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}} + \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}}\|u_\lambda\|^{q_3-1}_{L^{k_0}_{(\Sigma^u_\lambda)}}\Big) \\
&\qquad \times \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
&\quad + c\Big(\|v_\lambda\|^{p_2-1}_{L^{s_{12}}_{(\Sigma^v_\lambda)}} + \|v_\lambda\|^{q_2-1}_{L^{s_{22}}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{q_3}_{L^{k_0}_{(\Sigma^v_\lambda)}}\|v_\lambda\|^{p_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}} + \|u_\lambda\|^{q_3}_{L^{k_0}_{(\Sigma^v_\lambda)}}\|v_\lambda\|^{q_4-1}_{L^{k_0}_{(\Sigma^v_\lambda)}}\Big) \\
&\qquad \times \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \\
&\leq c\Big(\|u\|^{p_1-1}_{L^{s_{11}}_{((\Sigma^u_\lambda)^*)}} + \|u\|^{q_1-1}_{L^{s_{21}}_{((\Sigma^u_\lambda)^*)}} + \|v\|^{p_4}_{L^{k_0}_{(\Sigma^u_\lambda)}}\|u\|^{p_3-1}_{L^{k_0}_{((\Sigma^u_\lambda)*)}} + \|v\|^{q_4}_{L^{k_0}_{(\Sigma^u_\lambda)}}\|u\|^{q_3-1}_{L^{k_0}_{((\Sigma^u_\lambda)^*)}}\Big) \\
&\qquad \times \|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} \\
&\quad + c\Big(\|v\|^{p_2-1}_{L^{s_{12}}_{((\Sigma^v_\lambda)^*)}} + \|v\|^{q_2-1}_{L^{s_{22}}_{((\Sigma^v_\lambda)^*)}} + \|u\|^{q_3}_{L^{k_0}_{((\Sigma^v_\lambda)^*)}}\|v\|^{p_4-1}_{L^{k_0}_{((\Sigma^v_\lambda)^*)}} + \|u\|^{q_3}_{L^{k_0}_{((\Sigma^v_\lambda)^*)}}\|v\|^{q_4-1}_{L^{k_0}_{((\Sigma^v_\lambda)^*)}}\Big) \\
&\qquad \times \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}
\end{aligned}$$

for any $\frac{n}{n-\alpha} < \gamma < \infty$. Since $u \in \Pi_1$, $v \in \Pi_2$, we can choose ε small enough, such that for all $\lambda \in [\lambda_0, \lambda_0 + \varepsilon)$, we have

$$\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} \leq \tfrac{1}{2}\Big(\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} + \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}}\Big).$$

This implies that $\|u_\lambda - u\|_{L^\gamma_{(\Sigma^u_\lambda)}} = \|v_\lambda - v\|_{L^\gamma_{(\Sigma^v_\lambda)}} = 0$. So Σ^u_λ and Σ^v_λ must be empty. The proof of (38) is then completed. □

Proof of Theorem 1.2. From Lemma 3.1, it follows that $u(x) \geq u_\lambda(x)$ and $v(x) \geq v_\lambda(x)$ on Σ_λ for λ enough negative. This implies the possibility of moving the plane from near $x_1 = -\infty$, so we can invoke Step 2: move the plane to the limiting position to derive symmetry. Furthermore, it follows from Lemma 3.2 that if the plane stops at $x_1 = \lambda_0$ for some $\lambda_0 < 0$, then $u(x)$ and $v(x)$ must be symmetric and monotonic about the plane $x_1 = \lambda_0$. Otherwise, we can move the plane all the way to $x_1 = 0$. Since the direction of x_1 can be chosen arbitrarily, we deduce that $u(x)$ and $v(x)$ must be radially symmetric and monotonically decreasing about some point. This completes the proof of Theorem 1.2. □

4. Classification of positive solutions to system (4) with critical exponents

In this section, we prove Theorem 1.3. Since we have established the regularity and radial symmetry of solutions to system (4) in previous sections, we may employ a proposition in [Li and Zhu 1995; Li and Zhang 2003] to show the form of radially symmetric solutions of (4) with critical exponents. Throughout this section, we always assume that $p_1 = p_2 = q_1 = q_2 = p_3 + p_4 = q_3 + q_4 = \frac{n+\alpha}{n-\alpha}$ in system (4) and $(u, v) \in L^{2n/(n-\alpha)}(\mathbb{R}^n) \times L^{2n/(n-\alpha)}(\mathbb{R}^n)$. It is well known that system (4) (or (1)) is invariant with respect to scaling, translation, and inversion transformations with the above exponent conditions.

For $x \in \mathbb{R}^n$ and $\lambda > 0$, consider the Kelvin transformation of w:

$$w_{x,\lambda}(y) = \left(\frac{\lambda}{|y-x|}\right)^{n-\alpha} w\left(x + \frac{\lambda^2(y-x)}{|y-x|^2}\right).$$

To classify solutions, we need the following lemma.

Lemma 4.1. *Let* (u, v) *be a pair of solutions of system* (4) *with the assumptions of Theorem 1.3. Then there exist* $\lambda > 0$ *and* $x_0 \in \mathbb{R}^n$ *such that*

$$u_{x_0,\lambda}(y) = u(y), \tag{40}$$

$$v_{x_0,\lambda}(y) = v(y). \tag{41}$$

Proof. It suffices to prove (40). The proof of (41) is similar. Consider $x_0 = 0$, for otherwise we make a translation transform and a scaling transform on $u_{x_0,\lambda}(y)$. Let (u, v) be a pair of solutions of (4). By radial symmetry we assume without loss of generality that $u(x)$ and $v(x)$ are symmetric about the origin and $\lim_{|x|\to\infty} |x|^{n-\alpha} u(x) = u_\infty = u(0)$. Let $\lambda^{n-\alpha} = u_\infty / u(0)$ and e be any unit vector in $\mathbb{R}^n$. We define

$$w(y) = \frac{1}{|y|^{n-\alpha}} u\left(\frac{y}{|y|^2} - e\right).$$

Then

$$w(0) = u_\infty \quad \text{and} \quad w(e) = u(0).$$

Thus, w must be symmetric about $\frac{1}{2}e$.

Now, choosing $y = (\frac{1}{2} - h)e$ for any h, as in [Chen et al. 2006], it is easy to see

$$w\big((\tfrac{1}{2}-h)e\big) = \left(\frac{1}{\left|\frac{1}{2}-h\right|}\right)^{n-\alpha} u\left(\frac{\frac{1}{2}-h}{\left|\frac{1}{2}-h\right|^2}e - e\right) = \left(\frac{1}{\left|\frac{1}{2}-h\right|}\right)^{n-\alpha} u\left(\frac{\frac{1}{2}+h}{\frac{1}{2}-h}e\right),$$

where

$$\frac{\frac{1}{2}-h}{\left|\frac{1}{2}-h\right|^2}e - e = e\left(\frac{(\frac{1}{2}-h)-(\frac{1}{2}-h)^2}{\left|\frac{1}{2}-h\right|^2}\right) = e\frac{(\frac{1}{2}-h)(1-\frac{1}{2}+h)}{(\frac{1}{2}-h)^2} = e\frac{\frac{1}{2}+h}{\frac{1}{2}-h}.$$

Taking $y = (\frac{1}{2} - h)e$, we have

$$w\big((\tfrac{1}{2}+h)e\big) = \left(\frac{1}{\left|\frac{1}{2}+h\right|}\right)^{n-\alpha} u\left(\frac{\frac{1}{2}-h}{\frac{1}{2}+h}e\right).$$

Since w is symmetric about $\frac{1}{2}e$, by scaling we have

$$\frac{\lambda^{(n-\alpha)/2}}{\left|\frac{1}{2}-h\right|^{n-\alpha}} u\left(\lambda\frac{\frac{1}{2}+h}{\frac{1}{2}-h}e\right) = \frac{\lambda^{(n-\alpha)/2}}{\left|\frac{1}{2}+h\right|^{n-\alpha}} u\left(\lambda\frac{\frac{1}{2}-h}{\frac{1}{2}+h}e\right).$$

Letting $t = (\frac{1}{2} - h)/(\frac{1}{2} + h)$, it follows that

$$u\left(\frac{\lambda e}{t}\right) = t^{n-\alpha}u(\lambda t e).$$

Replacing t, e by $\lambda/|x-y|$, $y-x/|x-y|$, respectively, it follows that $u(y-x) = (\lambda/|y-x|)^{n-\alpha}u(\lambda^2(y-x)/|y-x|^2)$. Furthermore, we can take the translation transform to obtain (40). □

To prove our main result, we also need the following proposition from [Li and Zhang 2003]. Earlier versions with stronger assumptions were first proved by Li and Zhu [1995].

Proposition 4.2 [Li and Zhang 2003]. *Let $f \in C^1(\mathbb{R}^n)$, $\lambda > 0$ and $\mu > 0$. Suppose that for every $x \in \mathbb{R}^n$, there exists $\lambda(x) > 0$ such that*

$$\left(\frac{\lambda}{|y-x|}\right)^{\mu} f\left(x + \frac{\lambda^2(y-x)}{|y-x|^2}\right) = f(y), \quad y \in \mathbb{R}^n \setminus \{x\}.$$

Then for some $a \geq 0$, $d > 0$ and $\bar{x} \in \mathbb{R}^n$, we have

$$f(x) \equiv \pm a\left(\frac{1}{d+|x-\bar{x}|^2}\right)^{\mu/2}.$$

Proof of Theorem 1.3. From Lemma 4.1 and Proposition 4.2, we obtain directly that the solution of system (4) must be of the form (9). □

5. Equivalence of system (1) and system (4)

In this section, we show the equivalence of system (1) and the integral system (4). The proof is similar to that in [Chen and Li 2011] which is based on properties and the Fourier transform of the Riesz potential. For completeness and convenience of the reader, the details will be included. However, by choosing a suitable cut-off function, we provide a different approach for the case of even numbers $\alpha = 2m$.

First, we define a positive solution of (1) in the distribution sense, i.e, $u, v \in H^{\alpha/2}(\mathbb{R}^n)$, and they satisfy

$$\int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} u (-\Delta)^{\alpha/4} \phi \, dx = \int_{\mathbb{R}^n} \big(\lambda_1 u^{p_1} + \mu_1 v^{p_2} + \beta_1 u^{p_3} v^{p_4}\big) \phi \, dx, \tag{42}$$

$$\int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} v (-\Delta)^{\alpha/4} \phi \, dx = \int_{\mathbb{R}^n} \big(\lambda_2 u^{q_1} + \mu_2 v^{q_2} + \beta_2 u^{q_3} v^{q_4}\big) \phi \, dx \tag{43}$$

for any $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi(x) > 0$. As usual, by the Fourier transform we have

$$\int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} u (-\Delta)^{\alpha/4} \phi \, dx = c_n \int_{\mathbb{R}^n} |\xi|^\alpha \hat{u}(\xi) \hat{\phi}(\xi) \, d\xi, \tag{44}$$

where $\hat{u}$ and $\hat{\phi}$ are the Fourier transforms of u and ϕ, respectively.

For $\alpha = 2m$, where m is a positive integer, we prove that every positive solution of PDE system (1) satisfies integral system (4). Here we don't use the maximum principles for higher-order elliptic operators; the method be used here comes from [Lu and Zhu 2011].

Lemma 5.1. *Any positive solutions of system* (1) *with* $\alpha = 2m$ *satisfy the integral system* (4).

Proof. We define the cut-off function on $B_R(0)$:

$$\eta(x) = \begin{cases} 1, & x \in B_1(0), \\ 0, & x \notin B_2(0), \end{cases}$$

and $0 < \eta^{(i)} < 2$ on $B_2(0)$ for $i = 1, 2, \ldots, 2m$. Let $\eta_R(x-y) = \eta(\frac{|x-y|}{R})$ on $B_2(x)$ and choose $\phi(x-y) = \eta_R(x-y)/|x-y|^{n-2m}$. It is easy to check that $\phi \in C_0^\infty(\mathbb{R}^n)$. Hence, for any $u, v \in H^m(\mathbb{R}^n)$, by definition (42) and integration by parts, we have

$$\begin{aligned} \int_{\mathbb{R}^n} (-\Delta)^{m/2} u (-\Delta)^{m/2} \phi \, dy &= \int_{\mathbb{R}^n} u (-\Delta)^m \phi \, dy \\ &= \int_{\mathbb{R}^n} u (-\Delta)^m \left(\frac{\eta_R(x-y)}{|x-y|^{n-2m}} \right) dy \\ &= \int_{\mathbb{R}^n} Q_y(u, v) \frac{\eta_R(x-y)}{|x-y|^{n-2m}} \, dy, \end{aligned} \tag{45}$$

where $Q_y(u, v)$ is defined in Section 3. Since

$$(-\Delta)^m\left(\frac{\eta_R(x-y)}{|x-y|^{n-2m}}\right) = (-\Delta)^m\left(\frac{1}{|x-y|^{n-2m}}\right)\eta_R(x-y)+\sum_{i=1}^{2m} c_i|x-y|^{-n+i}\eta_R^{(i)}R^{-i},$$

one has

$$\int_{\mathbb{R}^n} u(-\Delta)^m\left(\frac{\eta_R(x-y)}{|x-y|^{n-2m}}\right)dy = \int_{\mathbb{R}^n} u(-\Delta)^m\left(\frac{1}{|x-y|^{n-2m}}\right)\eta_R(x-y)\,dy + \sum_{i=1}^{2m} c_i\int_{\mathbb{R}^n} R^{-i}u|x-y|^{-n+i}\eta_R^{(i)}\,dy. \tag{46}$$

As in [Lu and Zhu 2011], for $u \in L^{2n/(n-2m)}(\mathbb{R}^n)$, using the Hölder inequality we get

$$\begin{aligned}\int_{\mathbb{R}^n} u(x-y)^{-n-i}\eta_R^{(i)}R^{-i}\,dy &\le c_i R^{-i}\left(\int_{\mathbb{R}^n} u^{2n/(n-2m)}dy\right)^{\frac{n-2m}{2n}}\left(\int_{B_{2R}\setminus B_R}|x-y|^{2n(i-n)/(n+2m)}dy\right)^{\frac{n+2m}{2m}}\\ &\le \frac{c_i}{R^i}\int_R^{2R} r^{2n(i-n)/(n+2m)}r^{n-1}dr \to 0,\end{aligned} \tag{47}$$

as $R \to \infty$. We also note that

$$\int_{\mathbb{R}^n} u(-\Delta)^m\left(\frac{\eta_R(x-y)}{|x-y|^{n-2m}}\right)dy = \int_{\mathbb{R}^n}\delta(x-y)u(y)\,dy = u(x). \tag{48}$$

Therefore, combining (45), (46), (47) with (48), we have

$$u(x) = \int \frac{\lambda_1 u^{p_1}(y)+\mu_1 v^{p_2}(y)+\beta_1 u^{p_3}(y)v^{p_4}(y)}{|x-y|^{n-2m}}\,dy.$$

In the same way, we obtain

$$v(x) = \int_{\mathbb{R}^n} \frac{\lambda_2 u^{q_1}(y)+\mu_2 v^{q_2}(y)+\beta_2 u^{q_3}(y)v^{q_4}(y)}{|x-y|^{n-2m}}\,dy.$$

The proof of the lemma is completed. □

Now, we consider the case that α is not even, that is, system (1) is equivalent to the integral system (4) for any α.

Proof of Theorem 1.4. (i) For any $\phi \in C_0^\infty(\mathbb{R}^n)$, set

$$\psi(x) = \int \frac{\phi(x)}{|x-y|^{n-\alpha}}\,dy,$$

so that $(-\Delta)^{\alpha/2}\psi = \phi$, and then $\psi \in H^{\alpha}(\mathbb{R}^n) \subset H^{\alpha/2}(\mathbb{R}^n)$, and satisfies

$$\int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} u (-\Delta)^{\alpha/4} \psi \, dx = \int_{\mathbb{R}^n} Q_x(u, v)\psi(x)\, dx.$$

This implies

$$\int_{\mathbb{R}^n} u(-\Delta)^{\alpha/2}\psi \, dx = \int_{\mathbb{R}^n} u\phi \, dx = \int_{\mathbb{R}^n} \left\{ \int_{\mathbb{R}^n} \frac{Q_y(u, v)}{|x-y|^{n-\alpha}} dy \right\} \phi(x)\, dx$$

for any nonnegative $\phi \in C_0^\infty(\mathbb{R}^n)$. Thus, we get

$$u(x) = \int \frac{\lambda_1 u^{p_1}(y) + \mu_1 v^{p_2}(y) + \beta_1 u^{p_3}(y) v^{p_4}(y)}{|x-y|^{n-\alpha}}\, dy.$$

Similarly, we have

$$v(x) = \int \frac{\lambda_2 u^{q_1}(y) + \mu_2 v^{q_2}(y) + \beta_2 u^{q_3}(y) v^{q_4}(y)}{|x-y|^{n-\alpha}}\, dy.$$

(ii) Now we show that any positive solutions of the integral system (4) satisfy system (1). Assume that $u, v \in L^{2n/(n-2m)}(\mathbb{R}^n)$ are the solutions of the integral system (4). Invoking the Fourier transform on both sides of the first equation of (4), we have

$$\hat{u}(\xi) = c_n |\xi|^{-\alpha} \widehat{Q_\xi(u, v)}.$$

Then

$$|\xi|^\alpha \hat{u}(\xi) = c_n \widehat{Q_\xi(u, v)}(\xi).$$

Hence, for any $\phi \in C_0^\infty(\mathbb{R}^n)$, by (44) one has

$$\begin{aligned} \int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} u (-\Delta)^{\alpha/4} \phi \, dx &= c_n \int_{\mathbb{R}^n} |\xi|^\alpha \hat{u}(\xi) \hat{\phi}(\xi)\, d\xi \\ &= c_n \int_{\mathbb{R}^n} \widehat{Q_\xi(u, v)} \hat{\phi}(\xi)\, d\xi \\ &= c_n \int_{\mathbb{R}^n} Q_x(u, v)\phi(x)\, dx. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \int_{\mathbb{R}^n} (-\Delta)^{\alpha/4} v (-\Delta)^{\alpha/4} \phi \, dx &= c_n \int_{\mathbb{R}^n} \widehat{K_\xi(u, v)}(\xi) \hat{\phi}(\xi)\, d\xi \\ &= c_n \int_{\mathbb{R}^n} K_x(u, v)\phi(x)\, dx. \end{aligned}$$

This means that (u, v) is a pair of solutions of

$$\begin{cases} (-\Delta)^{\alpha/2} u = c_n(\lambda_1 u^{p_1} + \mu_1 v^{p_2} + \beta_1 u^{p_3} v^{p_4}), \\ (-\Delta)^{\alpha/2} v = c_n(\lambda_2 u^{q_1} + \mu_2 v^{q_2} + \beta_2 u^{q_3} v^{q_4}), \end{cases}$$

for $x \in \mathbb{R}^n$, in the sense of distributions. This completes the proof of the theorem. □

Now, we can combine Theorems 1.2 and 1.4 to show the nonexistence results.

Proof of Theorem 1.5. It suffices to verify the condition for exponents.

(i) and (ii) Under conditions (i) and (ii), respectively, the nonexistence results have been proved in [de Figueiredo and Sirakov 2005]. Combining this with our symmetry results, we find that there exist no nontrivial positive solutions (u, v) with $u \in \Pi_1 \cap L^\infty(\mathbb{R}^n)$, $v \in \Pi_2 \cap L^\infty(\mathbb{R}^n)$ satisfying conditions (i) and (ii), respectively.

(iii) Combining the nonexistence results of Dancer, Wei and T. Weth [2010] and our symmetry results, we conclude that there exist no nontrivial positive solutions (u, v) with $u \in \Pi_1 \cap L^\infty(\mathbb{R}^n)$ and $v \in \Pi_2 \cap L^\infty(\mathbb{R}^n)$ with $p_1 = q_2 = 3$, $p_3 = q_4 = 1$, $p_4 = q_3 = 2$. □

Acknowledgements

We would like to thank the anonymous referees for their valuable comments and suggestions. This work was supported by the Chinese National Science Foundation for Distinguished Young Scholars (grant 10925104), National Natural Science Foundation of China (grant 11101319), Foundation of Shaanxi Province Education Department (grant 2010JK549) and Zhejiang Provincial Natural Science Foundation of China (grant Y6110118).

References

[Akhmediev and Ankiewicz 1999] N. Akhmediev and A. Ankiewicz, "Partially coherent solitons on a finite background", *Phys. Rev. Lett.* **82**:13 (1999), 2661–2664.

[Bartsch et al. 2007] T. Bartsch, Z.-Q. Wang, and J. Wei, "Bound states for a coupled Schrödinger system", *J. Fixed Point Theory Appl.* **2**:2 (2007), 353–367. MR 2009i:35073 Zbl 1153.35390

[Bartsch et al. 2010] T. Bartsch, N. Dancer, and Z.-Q. Wang, "A Liouville theorem, a-priori bounds, and bifurcating branches of positive solutions for a nonlinear elliptic system", *Calc. Var. Partial Differential Equations* **37**:3-4 (2010), 345–361. MR 2011a:35110 Zbl 1189.35074

[Busca and Sirakov 2000] J. Busca and B. Sirakov, "Symmetry results for semilinear elliptic systems in the whole space", *J. Differential Equations* **163**:1 (2000), 41–56. MR 2001m:35100 Zbl 0952.35033

[Chen and Li 2005] W. Chen and C. Li, "Regularity of solutions for a system of integral equations", *Commun. Pure Appl. Anal.* **4**:1 (2005), 1–8. MR 2006g:45006 Zbl 1073.45004

[Chen and Li 2009a] W. Chen and C. Li, "Classification of positive solutions for nonlinear differential and integral systems with critical exponents", *Acta Math. Sci. Ser. B Engl. Ed.* **29**:4 (2009), 949–960. MR 2010i:35078 Zbl 1212.35103

[Chen and Li 2009b] W. Chen and C. Li, "An integral system and the Lane–Emden conjecture", *Discrete Contin. Dyn. Syst.* **24**:4 (2009), 1167–1184. MR 2010d:35068 Zbl 1176.35067

[Chen and Li 2010] W. Chen and C. Li, *Methods on nonlinear elliptic equations*, AIMS Series on Differential Equations & Dynamical Systems **4**, AIMS, Springfield, MO, 2010. MR 2012k:35002 Zbl 1214.35023

[Chen and Li 2011] W. Chen and C. Li, "Super polyharmonic property of solutions for PDE systems and its applications", preprint, 2011. arXiv 1110.2539v1

[Chen et al. 2005] W. Chen, C. Li, and B. Ou, "Classification of solutions for a system of integral equations", *Comm. Partial Differential Equations* **30**:1-3 (2005), 59–65. MR 2006a:45007 Zbl 1073.45005

[Chen et al. 2006] W. Chen, C. Li, and B. Ou, "Classification of solutions for an integral equation", *Comm. Pure Appl. Math.* **59**:3 (2006), 330–343. MR 2006m:45007a Zbl 1093.45001

[Dancer et al. 2010] E. N. Dancer, J. Wei, and T. Weth, "A priori bounds versus multiple existence of positive solutions for a nonlinear Schrödinger system", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **27**:3 (2010), 953–969. MR 2011d:35117 Zbl 1191.35121

[Dou et al. 2011] J. Dou, C. Qu, and Y. Han, "Symmetry and nonexistence of positive solutions to an integral system with weighted functions", *Sci. China Math.* **54**:4 (2011), 753–768. MR 2012c:45013 Zbl 1222.45003

[Esry et al. 1997] B. D. Esry, C. H. Greene, J. P. Burke, Jr., and J. L. Bohn, "Hartree–Fock theory for double condensates", *Phys. Rev. Lett.* **78**:19 (1997), 3594–3597.

[de Figueiredo and Sirakov 2005] D. G. de Figueiredo and B. Sirakov, "Liouville type theorems, monotonicity results and a priori bounds for positive solutions of elliptic systems", *Math. Ann.* **333**:2 (2005), 231–260. MR 2006i:35072 Zbl 1165.35360

[Fu et al. 2009] F. Fu, L. Kong, and L. Wang, "Symplectic Euler method for nonlinear high order Schrödinger equation with a trapped term", *Adv. Appl. Math. Mech.* **1**:5 (2009), 699–710. MR 2010j:65125

[Gidas et al. 1981] B. Gidas, W. M. Ni, and L. Nirenberg, "Symmetry of positive solutions of nonlinear elliptic equations in $\mathbf{R}^n$", pp. 369–402 in *Mathematical analysis and applications, Part A*, edited by L. Nachbin, Adv. in Math. Suppl. Stud. **7**, Academic, New York, 1981. MR 84a:35083 Zbl 0469.35052

[Guo and Liu 2008] Y. Guo and J. Liu, "Liouville type theorems for positive solutions of elliptic system in $\mathbb{R}^N$", *Comm. Partial Differential Equations* **33**:1-3 (2008), 263–284. MR 2009b:35075 Zbl 1139.35305

[Hang 2007] F. Hang, "On the integral systems related to Hardy–Littlewood–Sobolev inequality", *Math. Res. Lett.* **14**:3 (2007), 373–383. MR 2008j:26037 Zbl 1144.26031

[Hioe 1999] F. T. Hioe, "Solitary waves for N coupled nonlinear Schrödinger equations", *Phys. Rev. Lett.* **82**:6 (1999), 1152–1155.

[Li 2004] Y. Y. Li, "Remark on some conformally invariant integral equations: The method of moving spheres", *J. Eur. Math. Soc.* (*JEMS*) **6**:2 (2004), 153–180. MR 2005e:45007 Zbl 1075.45006

[Li and Zhang 2003] Y. Li and L. Zhang, "Liouville-type theorems and Harnack-type inequalities for semilinear elliptic equations", *J. Anal. Math.* **90** (2003), 27–87. MR 2004i:35118 Zbl 1173.35477

[Li and Zhu 1995] Y. Li and M. Zhu, "Uniqueness theorems through the method of moving spheres", *Duke Math. J.* **80**:2 (1995), 383–417. MR 96k:35061 Zbl 0846.35050

[Lin and Wei 2005] T.-C. Lin and J. Wei, "Ground state of N coupled nonlinear Schrödinger equations in $\mathbb{R}^n$, $n \leq 3$", *Comm. Math. Phys.* **255**:3 (2005), 629–653. MR 2006g:35044 Zbl 1119.35087

[Liu and Wang 2008] Z. Liu and Z.-Q. Wang, "Multiple bound states of nonlinear Schrödinger systems", *Comm. Math. Phys.* **282**:3 (2008), 721–731. MR 2009k:58022 Zbl 1156.35093

[Liu et al. 2006] J. Liu, Y. Guo, and Y. Zhang, "Liouville-type theorems for polyharmonic systems in $\mathbb{R}^N$", *J. Differential Equations* **225**:2 (2006), 685–709. MR 2007d:35092 Zbl 1147.35316

[Liu et al. 2007] X.-S. Liu, Y.-Y. Qi, J.-F. He, and P.-Z. Ding, "Recent progress in symplectic algorithms for use in quantum systems", *Commun. Comput. Phys.* **2**:1 (2007), 1–53. MR 2008a:81002

[Lu and Zhu 2011] G. Lu and J. Zhu, "Symmetry and regularity of extremals of an integral equation related to the Hardy–Sobolev inequality", *Calc. Var. Partial Differential Equations* **42**:3-4 (2011), 563–577. Zbl 1231.35290

[Ma et al. 2011] C. Ma, W. Chen, and C. Li, "Regularity of solutions for an integral system of Wolff type", *Adv. Math.* **226**:3 (2011), 2676–2699. MR 2011k:45009 Zbl 1209.45006

[Maia et al. 2006] L. A. Maia, E. Montefusco, and B. Pellacci, "Positive solutions for a weakly coupled nonlinear Schrödinger system", *J. Differential Equations* **229**:2 (2006), 743–767. MR 2007h:35070 Zbl 1104.35053

[Sirakov 2007] B. Sirakov, "Least energy solitary waves for a system of nonlinear Schrödinger equations in $\mathbb{R}^n$", *Comm. Math. Phys.* **271**:1 (2007), 199–221. MR 2007k:35477 Zbl 1147.35098

[Wei and Weth 2007] J. Wei and T. Weth, "Nonradial symmetric bound states for a system of coupled Schrödinger equations", *Rend. Lincei Mat. Appl.* **18**:3 (2007), 279–293. MR 2008g:35078 Zbl 1229.35019

[Wei and Weth 2008] J. Wei and T. Weth, "Radial solutions and phase separation in a system of two coupled Schrödinger equations", *Arch. Ration. Mech. Anal.* **190**:1 (2008), 83–106. MR 2009k:35059 Zbl 1161.35051

[Wei and Xu 1999] J. Wei and X. Xu, "Classification of solutions of higher order conformally invariant equations", *Math. Ann.* **313**:2 (1999), 207–228. MR 2000a:58093 Zbl 0940.35082

[Zhang 2007] Y. Zhang, "A Liouville type theorem for polyharmonic elliptic systems", *J. Math. Anal. Appl.* **326**:1 (2007), 677–690. MR 2007g:35070 Zbl 1142.35020

Received July 22, 2011. Revised September 7, 2012.

JINGBO DOU
SCHOOL OF STATISTICS
XI'AN UNIVERSITY OF FINANCE AND ECONOMICS
710100 XI'AN
CHINA
jbdou@xaufe.edu.cn

CHANGZHENG QU
CENTER FOR NONLINEAR STUDIES
NINGBO UNIVERSITY
315211 NINGBO
CHINA
quchangzheng@nbu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.335

BOUND STATES OF ASYMPTOTICALLY LINEAR SCHRÖDINGER EQUATIONS WITH COMPACTLY SUPPORTED POTENTIALS

MINGWEN FEI AND HUICHENG YIN

We study the existence and concentration of bound states to N-dimensional nonlinear Schrödinger equation $-\varepsilon^2\Delta u_\varepsilon + V(x)u_\varepsilon = K(x)f(u_\varepsilon)$, where $N \geq 3$, $\varepsilon > 0$ is sufficiently small, and the function $f(s)$ is nonnegative and asymptotically linear at infinity. More concretely, when $f(s) \sim O(s)$ as $s \to +\infty$, the potential function $V(x)$ lies in $C_0^1(\mathbb{R}^N)$ with $V(x) \geq 0$ and $V(x) \not\equiv 0$, and $K(x) \geq 0$ is permitted to be unbounded under some other necessary restrictions, we can show that a positive $H^1(\mathbb{R}^N)$-solution $u_\varepsilon(x)$ exists and concentrates around the local maximum point of the corresponding ground energy function.

1. Introduction and statements of main results

This paper deals with the problem on the existence and concentration of bound states for the nonlinear Schrödinger equation

$$\begin{cases} -\varepsilon^2\Delta u_\varepsilon + V(x)u_\varepsilon = K(x)f(u_\varepsilon), & x \in \mathbb{R}^N, \\ u_\varepsilon \in H^1(\mathbb{R}^N), \quad u_\varepsilon(x) > 0, \end{cases} \tag{1-1}$$

where $N \geq 3$, $\varepsilon > 0$ is small, $K(x) \geq 0$, $V(x) \geq 0$ with $V(x) \not\equiv 0$, $f(s) \geq 0$ and $f(s) \sim O(s)$ as $s \to +\infty$, which is asymptotically linear. Such a solution u_ε is called as a bound state for $u_\varepsilon \in H^1(\mathbb{R}^N)$ and $u_\varepsilon(x) > 0$.

Consider in particular the superlinear problem given by the equation

$$\begin{cases} -\varepsilon^2\Delta u_\varepsilon + V(x)u_\varepsilon = K(x)|u_\varepsilon|^{p-1}u_\varepsilon, & x \in \mathbb{R}^N, \\ u_\varepsilon \in H^1(\mathbb{R}^N), \quad u_\varepsilon > 0, \end{cases} \tag{1-2}$$

for $N \geq 3$ and $1 < p < \frac{N+2}{N-2}$. Under various assumptions on the potential function $V(x) \geq C_0 > 0$ for large $|x|$ or $\lim_{|x|\to\infty} V(x) = 0$ or even $V(x)$ is compactly

This research is supported by NSFC (No. 10931007, No. 11025105), and by the Priority Academic Program Development of Jiangsu Higher Education Institutions. Yin is the corresponding author.
MSC2010: 35J10, 35J60, 35Q55, 35J91.
Keywords: nonlinear Schrödinger equation, bound state, asymptotically linear, Harnack inequality, concentration-compactness.

supported with $V(x) \geq 0$ and $V(x) \not\equiv 0$, the existence of H^1-positive solutions has been established, and the concentration properties of u_ε can be obtained at a global or local minimum point of the ground energy function $G(\xi) \equiv V^{\theta}(\xi)K^{-2/(p-2)}(\xi)$ with $\theta = \frac{p}{p-2} - \frac{N}{2}$ (one can see [Ambrosetti et al. 2005; Ambrosetti and Malchiodi 2007; Ambrosetti and Wang 2005; Berestycki and Lions 1983; Bonheure and Van Schaftingen 2008; Byeon and Wang 2006; Dávila et al. 2007; del Pino and Felmer 1996; del Pino et al. 2007; Fei and Yin 2010; Gui 1996; Rabinowitz 1992; Wang and Zeng 1997; Yin and Zhang 2009]).

For the asymptotically linear problem (1-1) with $\varepsilon = 1$, there are many papers on the existence of solution in recent years. For examples, in the case of $V(x) \geq C_0 > 0$ for large $|x|$, one can see [Costa and Tehrani 2001; Jeanjean and Tanaka 2002; Liu et al. 2006; Liu and Wang 2004; Stuart and Zhou 1999]; in the special case that $V(x)$ vanishes at infinity like $a/(1+|x|^{\sigma}) \leq V(x) \leq A$ (the constants $\sigma \in (0, 2)$, $a > 0$ and $A > 0$) and some other restrictions, the authors in [Liu et al. 2008] established the existence of bound states.

We now consider the following interesting problems indicated in [Ambrosetti and Malchiodi 2007]: if the potential function $V(x)$ decays faster than $1/(1+|x|^{\sigma})$ with $\sigma \in (0, 2)$ at infinity or is compactly supported with $V(x) \geq 0$ and $V(x) \not\equiv 0$, does the bound state of (1-1) still exist? If it exists, what is the concentration profile of $u_\varepsilon(x)$ as $\varepsilon \to 0$? In this paper, we will treat these two problems. We only focus on the case that $V(x)$ is compactly supported, since the other cases of $V(x) = O(1/(1+|x|^{\sigma}))$ with $\sigma \in \mathbb{R}$ can be treated analogously and even more simply.

To proceed, we define the *ground energy function* $G(\xi)$. The constant coefficient asymptotically linear equation is as follows:

$$(1\text{-}3) \qquad \begin{cases} -\triangle u(x) + V(\xi)u(x) = K(\xi)f(u), & x \in \mathbb{R}^N, \\ u \in H^1(\mathbb{R}^N), \quad u(x) > 0, \end{cases}$$

where $V(\xi)$, $K(\xi) > 0$ with $\xi \in \bar{\Lambda}$, and the meaning of Λ is given in assumption (H_4) below.

The associated Euler functional is defined as

$$(1\text{-}4) \qquad I^{\xi}(u) = \frac{1}{2}\int_{\mathbb{R}^N} |\nabla u|^2\,dx + \frac{V(\xi)}{2}\int_{\mathbb{R}^N} |u|^2\,dx - K(\xi)\int_{\mathbb{R}^N} F(u)\,dx,$$

where $F(u) = \int_0^u f(x,\tau)\,d\tau$.

In the terminology in [Wang and Zeng 1997], the function $G(\xi) = \inf_{u \in \mathcal{M}^{\xi}} I^{\xi}(u)$ is the *ground energy function* of (1-3) and $\omega(x)$ is a *ground state* of the functional

I^ξ if $G(\xi) = I^\xi(\omega)$, where $\mathcal{M}^\xi$ is the Nehari manifold, defined as

$$\mathcal{M}^\xi = \left\{u \in H^1(\mathbb{R}^N)\backslash\{0\} : \int_{\mathbb{R}^N} |\nabla u|^2\,dx + V(\xi)\int_{\mathbb{R}^N} |u|^2\,dx = K(\xi)\int_{\mathbb{R}^N} f(u)u\,dx\right\}. \tag{1-5}$$

Under certain assumptions, we will solve the constant coefficient asymptotically linear problem (1-3) and prove that the ground state exists and $G(\xi)$ is a continuous function in $\bar{\Lambda}$ in Section 3 below. The assumptions are as follows:

(H_1) $V(x) \in C_0^1(\mathbb{R}^N)$, $V(x) \geq 0$; $K(x) \in C^1(\mathbb{R}^N)$, $K(x) \geq 0$.

(H_2) $f \in C(\mathbb{R}, \mathbb{R}^+) \cap C^{1,\gamma}_{\text{loc}}(\mathbb{R})$ with some constant γ satisfying $0 < \gamma \leq 1$; $f(s) = 0$ for $s \leq 0$; $f(s) = O(s^\alpha)$ with some $\alpha > 1$ near $s = 0$.

(H_3) $f(s)/s$ is a nondecreasing function for $s > 0$ and

$$\frac{f(s)}{s} \to l \in (0, +\infty) \quad \text{as } s \to +\infty. \tag{1-6}$$

(H_4) There exists a smooth bounded domain Λ of $\mathbb{R}^N$ such that $V(x) > 0$, $K(x) > 0$ on $\bar{\Lambda}$ and

$$\mu^* \equiv \max_{\xi \in \bar{\Lambda}} \frac{V(\xi)}{K(\xi)} < l, \tag{1-7}$$

$$0 < c_0 \equiv \inf_{\xi \in \Lambda} G(\xi) < \inf_{\xi \in \partial\Lambda} G(\xi). \tag{1-8}$$

(H_5) Let $N \geq 5$. There exist some constants $k > 0$ and $\beta < (\alpha - 1)(N - 2) - 2$ such that

$$0 \leq K(x) \leq k(1 + |x|)^\beta \quad \text{in } \mathbb{R}^N. \tag{1-9}$$

Our main results in this paper can be stated as follows:

Theorem 1.1 (existence and concentration). *Let assumptions* (H_1)–(H_5) *hold.*

(i) *Equation* (1-1) *has at least one bound state* u_ε *provided that* ε *is small.*

(ii) u_ε *has exactly one maximum point* $x_\varepsilon \in \Lambda$, *which satisfies*

$$C_1 \leq u_\varepsilon(x_\varepsilon) \leq C_2 \tag{1-10}$$

and

$$\operatorname{dist}(x_\varepsilon, M) \to 0 \quad \textit{as } \varepsilon \to 0, \tag{1-11}$$

where C_1, C_2 *are positive constants independent of* ε, *and the set* M *is defined by* $M = \{x \in \Lambda : G(x) = c_0\}$. *Moreover, if* M *only contains a single point* x_0, *then* u_ε *is a single peak solution; more precisely,*

$$u_\varepsilon(x) = v\left(\frac{x - x_\varepsilon}{\varepsilon}\right) + w_\varepsilon(x), \tag{1-12}$$

where $w_\varepsilon(x) \to 0$ *in* $C^2_{\text{loc}}(\mathbb{R}^N) \cap L^\infty(\mathbb{R}^N)$ *as* $\varepsilon \to 0$ *and* $v \in C^2(\mathbb{R}^N) \cap H^1(\mathbb{R}^N)$ *is the positive solution of the equation*

$$-\Delta v + V(x_0)v = K(x_0) f(v), \quad x \in \mathbb{R}^N. \tag{1-13}$$

Remark 1.1. In the assumption (H_5), $N \geq 5$ can not be removed to obtain $u_\varepsilon \in L^2(\mathbb{R}^N)$ in Theorem 1.1 since this is also necessary even for the N-dimensional linear Laplacian equation. For more details, one can see Remark 1.2 of [Yin and Zhang 2009]. On the other hand, if we do not require $u_\varepsilon \in L^2(\mathbb{R}^N)$ in Theorem 1.1, for example, only $u_\varepsilon \in L^q(\mathbb{R}^N)$ is permitted for some $q > 1$, then Theorem 1.1 still holds for all $N \geq 2$ by our proof procedure since $N \geq 5$ is only used in (4-52) of Section 4 to derive $u_\varepsilon \in L^2(\mathbb{R}^N)$ through the whole paper.

Remark 1.2. In the assumption (H_2), due to $f \in C^{1,\gamma}_{\text{loc}}(\mathbb{R})$, $f(s) = 0$ for $s \leq 0$ and $f(s) = O(s^\alpha)$ near $s = 0$ with $\alpha > 1$, then we actually have $0 < \gamma \leq \min\{1, \alpha - 1\}$.

Remark 1.3. With respect to the assumption (1-7) in (H_4), if $V(x) \sim l^*/(1+|x|^{\beta_1})$ with $\beta_1 > 0$ and $K(x) \sim 1/(1+|x|^{\beta_2})$ with $0 < \beta_2 < \beta_1$ or $V(x) \sim l^* e^{-|x|^{\beta_1}}$ with $\beta_1 > 0$ and $K(x) \sim e^{-|x|^{\beta_2}}$ with $0 < \beta_2 < \beta_1$, then for $0 < l^* < l$, we have $\mu^* \leq l^* < l$, namely, (1-7) holds true. However, assumption (1-7) does not satisfy the condition (K_1) in [Liu et al. 2008], to the effect that $\sup\bigl\{\frac{f(s)}{s} : s > 0\bigr\} < \inf\bigl\{\frac{V(x)}{K(x)} : |x| \geq R_0\bigr\}$ for some $R_0 > 0$, which seems to be crucial to the proof there. On the other hand, the main assumptions (K_1) and (1.8) in Theorem 1.1 of [Liu et al. 2008] are rather restricted. If we use instead of (K_1) the more natural assumption $\sup\bigl\{\frac{f(s)}{s} : s > 0\bigr\} < \inf\bigl\{\frac{V(x)}{K(x)} : x \in \mathbb{R}^n\bigr\}$, one can easily derive $l < \inf\bigl\{\frac{V(x)}{K(x)} : x \in \mathbb{R}^n\bigr\}$ and

$$\begin{aligned}\mu^* &= \inf f \int_{\mathbb{R}^N} \bigl(|\nabla u|^2 + V(x)u^2\bigr)\,dx \int_{\mathbb{R}^N} K(x)u^2\,dx \\ &\geq \inf \frac{\int_{\mathbb{R}^N} (|\nabla u|^2 + lK(x)u^2)\,dx}{\int_{\mathbb{R}^N} K(x)u^2\,dx} \geq l,\end{aligned}$$

which yields an obvious contradiction between the main assumption $l > \mu^*$ of (1.8) and (K_1) in Theorem 1.1 of [Liu et al. 2008].

Remark 1.4. The function $K(x)$ in (1-1) can be permitted to be unbounded if $\alpha > \frac{N}{N-2}$ in view of the assumption (1-9). Moreover, as in Remark 1.2 of [Yin and Zhang 2009], we can illustrate that the restriction on $\beta < (\alpha - 1)(N - 2) - 2$ in (1-9) is optimal in order to obtain the existence of H^1-positive solution to (1-1).

Remark 1.5. The assumption in (H_3) that $f(s)/s$ is a nondecreasing function for $s > 0$ can be removed by more careful analysis than that employed in this paper. This will be done in a forthcoming paper.

Next let's make some comments on the proof of Theorem 1.1. First, we modify the nonlinear term $K(x) f(u_\varepsilon)$ of (1-1) outside Λ to $g_\varepsilon(x, u_\varepsilon)$, as in [Yin and Zhang

2009], with the expression

$$g_\varepsilon(x,u)=\min\bigl\{K(x)f(u),\ \varepsilon^3/(1+|x|^{\theta_0})u^+,\ \varepsilon/(1+|x|^N)\bigr\}$$

for $x\in\mathbb{R}^N$ and $u\in\mathbb{R}$, for a positive constant θ_0 to be chosen suitably. Then we study the modified equation

$$-\varepsilon^2\Delta u_\varepsilon+V(x)u_\varepsilon=\chi_\Lambda(x)K(x)f(u_\varepsilon)+(1-\chi_\Lambda(x))g_\varepsilon(x,u_\varepsilon) \tag{1-14}$$

instead of $-\varepsilon^2\Delta u_\varepsilon+V(x)u_\varepsilon=K(x)f(u_\varepsilon)$ in (1-1). It can be shown that the corresponding Euler functional I_ε of the modified equation is well-defined and has a mountain pass geometry in the weighted Sobolev space

$$E_\varepsilon\equiv\left\{u\in\mathfrak{D}^{1,2}(\mathbb{R}^N):\int_{\mathbb{R}^N}(\varepsilon^2|\nabla u|^2+V(x)|u|^2)\,dx<\infty\right\},$$

with $\mathfrak{D}^{1,2}(\mathbb{R}^N)=\{u\in L^{2N/(N-2)}(\mathbb{R}^N):\nabla u\in L^2(\mathbb{R}^N)\}$. Motivated by techniques in Chapter IV of [Ekeland 1990] or [Jeanjean and Tanaka 2002], we can use a variant of the mountain pass theorem to find a so-called Cerami sequence, and further show by contradiction that such a Cerami sequence is bounded and prove the existence of a positive solution u_ε to the modified equation.

In order to show such a solution u_ε is just the solution of the original problem (1-1), we require to derive the decay property of solution u_ε and further show $g_\varepsilon(x,u_\varepsilon)=K(x)f(u_\varepsilon)$ outside the domain Λ. To this end, we establish a compactness estimate of integral type to prove that u_ε is small away from their extreme points (see Lemma 4.6 below). Based on such an integral estimate together with the Harnack inequality, we obtain the pointwise decay property of u_ε at infinity and then complete the proof of Theorem 1.1.

Here we point out that some phenomena arising from the asymptotically linear case are quite different from those in superlinear cases, since the exponent $p>1$ of $f(u)\sim u^p$ plays a crucial role in showing the concentration-compactness of u_ε and deriving the decay property of u_ε at infinity. (Especially important is the property $F(s)\equiv\int_0^s f(\tau)\,d\tau\le k_0f(s)s$, with a positive constant $k_0<\frac{1}{2}$ and $s>0$ in superlinear cases; one can see details in [Yin and Zhang 2009; Fei and Yin 2010] and the illustrations before Lemma 4.3 in this paper.) This means that some methods used in [Yin and Zhang 2009] cannot be employed directly here.

Our paper is organized as follows. In Section 2, we replace the nonlinearity $K(x)f(u_\varepsilon)$ outside Λ by a suitably truncated function $g_\varepsilon(x,u_\varepsilon)$ and give a detailed analysis of the modified equation (1-14), so that the existence of nontrivial positive solution u_ε can be established. In Section 3, we give some preliminary results regarding the properties of the nonlinear Schrödinger equation $-\Delta u+V(\xi)u=K(\xi)f(u)$. In Section 4, we derive an integral decay estimate and use the Harnack inequality to derive the pointwise decay estimate of u_ε at infinity, inspired by

Lemma 17 of [Ambrosetti et al. 2005] and Lemmas 4.3 and 4.4 of [Yin and Zhang 2009]. From these, together with some involved analysis, we can complete the proof of Theorem 1.1.

We will use the following notations:

B_r denotes the ball centered at the origin with the radius r.

For a set $A \subset \mathbb{R}^N$, we put $A^\varepsilon = \{\varepsilon^{-1}x : x \in A\}$.

2. Existence of critical points for a modified nonlinear equation

We define a class of weighted Sobolev spaces as follows:

$$E_\varepsilon : \left\{ u \in \mathfrak{D}^{1,2}(\mathbb{R}^N) : \int_{\mathbb{R}^N} (\varepsilon^2 |\nabla u|^2 + V(x)|u|^2)\,dx < \infty \right\}$$

with $\mathfrak{D}^{1,2}(\mathbb{R}^N) = \{u \in L^{2N/(N-2)}(\mathbb{R}^N) : \nabla u \in L^2(\mathbb{R}^N)\}$.

The norm of the space E_ε is denoted by

$$\|u\|_\varepsilon = \left(\int_{\mathbb{R}^N} (\varepsilon^2 |\nabla u|^2 + V(x)|u|^2)\,dx \right)^{1/2} \quad \text{for } u \in E_\varepsilon.$$

Towards proving Theorem 1.1, it is necessary to modify (1-1) and further discuss the existence of solution to the modified equation.

To this end, we define a function $g_\varepsilon(x,\xi)$ by

$$g_\varepsilon(x,\xi) = \min\left\{ K(x) f(\xi),\ \frac{\varepsilon^3}{1+|x|^{\theta_0}}\xi^+,\ \frac{\varepsilon}{1+|x|^N} \right\}, \quad x \in \mathbb{R}^N,\ \xi \in \mathbb{R},$$

where $\xi^+ = \max\{\xi, 0\}$, and $\theta_0 > 2$ will be suitably chosen in (4-51).

Set

$$h_\varepsilon(x,\xi) = \chi_\Lambda(x) K(x) f(\xi) + (1-\chi_\Lambda(x)) g_\varepsilon(x,\xi),$$

where $\chi_\Lambda(x)$ represents the characteristic function of the set Λ.

We now consider the modified nonlinear equation

$$-\varepsilon^2 \Delta u + V(x)u = h_\varepsilon(x,u), \quad x \in \mathbb{R}^N. \tag{2-1}$$

The functional corresponding to (2-1) is

$$I_\varepsilon(u) = \tfrac{1}{2}\|u\|_\varepsilon^2 - \int_\Lambda K(x)F(u)\,dx - \int_{\mathbb{R}^N\setminus\Lambda} G_\varepsilon(x,u)\,dx, \tag{2-2}$$

where $F(s) = \int_0^s f(\tau)\,d\tau$ and $G_\varepsilon(x,s) = \int_0^s g_\varepsilon(x,\tau)\,d\tau$.

By (H_2) and (H_3), for any $\delta > 0$, there exists $C_\delta > 0$ such that $f(s) \le \delta s + C_\delta |s|^{2^*-1}$

and further

$$\int_{\Lambda} K(x)F(u)\,dx \le C\delta\|u\|_\varepsilon^2 + C\varepsilon^{-2^*}\|u\|_\varepsilon^{2^*}. \tag{2-3}$$

On the other hand, a direct computation yields for $u \in E_\varepsilon$

$$\int_{\mathbb{R}^N\setminus\Lambda} G_\varepsilon(x,u)\,dx \le \int_{\mathbb{R}^N\setminus\Lambda} g_\varepsilon(x,u)u\,dx \le C\varepsilon\|u\|_\varepsilon^2. \tag{2-4}$$

It follows from (2-3) and (2-4) that $I_\varepsilon(u)$ is well-defined on E_ε. That I_ε lies in $C^1(E_\varepsilon,\mathbb{R})$ is obvious.

Next we show that I_ε has a mountain pass geometry. Given small $\varepsilon > 0$, by (2-3) and (2-4), there are two small numbers δ and $r > 0$ such that

$$I_\varepsilon(u) \ge \tfrac{1}{2}\|u\|_\varepsilon^2 - C\delta\|u\|_\varepsilon^2 - C\varepsilon^{-2^*}\|u\|_\varepsilon^{2^*} - C\varepsilon\|u\|_\varepsilon^2 \ge \tfrac{1}{4}\|u\|_\varepsilon^2 \quad \text{for } \|u\|_\varepsilon \le r. \tag{2-5}$$

We now claim that

$$\inf_{\psi\in H^1(\mathbb{R}^N)\setminus\{0\}} \frac{\int_{\mathbb{R}^N}|\nabla\psi|^2\,dx}{\int_{\mathbb{R}^N}\psi^2\,dx} = 0. \tag{2-6}$$

Indeed, if $\psi_0(x) \ne 0 \in H^1(\mathbb{R}^N)$, then for any fixed $\lambda \in \mathbb{R}$, one has $\psi_0(\lambda x) \in H^1(\mathbb{R}^N)$. A direct computation yields that

$$\int_{\mathbb{R}^N} |\nabla(\psi_0(\lambda x))|^2\,dx = \lambda^{2-N}\int_{\mathbb{R}^N}|\nabla\psi_0(x)|^2\,dx$$

and

$$\int_{\mathbb{R}^N} |\psi_0(\lambda x)|^2\,dx = \lambda^{-N}\int_{\mathbb{R}^N}|\psi_0(x)|^2\,dx.$$

Therefore, we arrive at

$$\frac{\int_{\mathbb{R}^N}|\nabla(\psi_0(\lambda x))|^2\,dx}{\int_{\mathbb{R}^N}|\psi_0(\lambda x)|^2\,dx} = \lambda^2\frac{\int_{\mathbb{R}^N}|\nabla\psi_0(x)|^2\,dx}{\int_{\mathbb{R}^N}|\psi_0(x)|^2\,dx} \to 0 \quad \text{as } \lambda \to 0, \tag{2-7}$$

proving (2-6).

From (2-6), we obtain for any fixed $\xi \in \Lambda$,

$$\inf_{\psi\in H^1(\mathbb{R}^N)\setminus\{0\}} \frac{\int_{\mathbb{R}^N}(|\nabla\psi|^2 + V(\xi)|\psi|^2)\,dx}{\int_{\mathbb{R}^N}K(\xi)\psi^2\,dx} = \frac{V(\xi)}{K(\xi)}. \tag{2-8}$$

This, together with (1-7), yields that for fixed $\xi \in \Lambda$ there exists a function $\varphi \in C_0^\infty(\mathbb{R}^N)$ such that

$$\frac{\int_{\mathbb{R}^N}(|\nabla\varphi|^2 + V(\xi)|\varphi|^2)\,dx}{\int_{\mathbb{R}^N}K(\xi)\varphi^2\,dx} < l. \tag{2-9}$$

Choose $R > 0$ such that $B_R(\xi) \subset \Lambda$. We define a smooth cut-off function

$\eta:\mathbb{R}^+\to\mathbb{R}^+$ satisfying $\eta(t)=1$ if $0\le t\le \frac{R}{4}$, $\eta(t)=0$ if $t\ge\frac{R}{2}$ and $|\eta'(t)|\le\frac{8}{R}$.

Set

$$\varphi_\varepsilon(x)=\eta(|x-\xi|)\varphi\Big(\frac{x-\xi}{\varepsilon}\Big)\in C_0^\infty(\Lambda).$$

Then

$$I_\varepsilon(t\varphi_\varepsilon) = \varepsilon^N\Big(\frac{t^2}{2}\int_{\mathbb{R}^N}\big(|\nabla\varphi|^2+V(\xi)|\varphi|^2\big)\,dx-\frac{1}{2}\int_{\mathbb{R}^N}K(\xi)F(t\varphi)\,dx+o_\varepsilon(1)\Big); \tag{2-10}$$

here and below the notation $o_\varepsilon(1)$ stands for a quantity which satisfies $o_\varepsilon(1)\to 0$ as $\varepsilon\to 0$.

Thus we have, for $\varepsilon\le 1$,

$$\liminf_{t\to+\infty}\frac{I_\varepsilon(t\varphi_\varepsilon)}{t^2}\le\frac{1}{2}\int_{\mathbb{R}^2}\big(|\nabla\varphi|^2+V(\xi)|\varphi|^2\big)\,dx-\frac{l}{2}\int_{\mathbb{R}^N}K(\xi)\varphi^2\,dx<0. \tag{2-11}$$

Consequently, there exists some $t_0>0$ such that $I_\varepsilon(t_0\varphi_\varepsilon)<0$. This, together with (2-5), means that I_ε has a mountain pass geometry. Let

$$c_\varepsilon=\inf_{\gamma\in\Gamma_\varepsilon}\max_{0\le t\le 1}I_\varepsilon(\gamma(t)),$$

where $\Gamma_\varepsilon=\{\gamma\in C([0,1],E_\varepsilon):\gamma(0)=0,\ I_\varepsilon(\gamma(1))<0\}$. By the mountain pass theorem in Chapter IV of [Ekeland 1990], as in [Liu et al. 2008], one has the following lemma.

Lemma 2.1. *Under the assumptions* (H_1)–(H_4), *for small* $\varepsilon>0$, *there exists a sequence* $\{u_n\}\subset E_\varepsilon$ *such that* $I_\varepsilon(u_n)\to c_\varepsilon$ *and* $\|I'_\varepsilon(u_n)\|_{E'_\varepsilon}(1+\|u_n\|_\varepsilon)\to 0$ *as* $n\to\infty$, *where* E'_ε *and* $\|I'_\varepsilon(u_n)\|_{E'_\varepsilon}$ *denote by the dual space of* E_ε *and the norm of* $I'_\varepsilon(u_n)$ *in* E'_ε.

Such a sequence is called a Cerami sequence. Next we will prove the sequence $\{u_n\}$ is bounded in E_ε. We reason by contradiction: we assume up to a subsequence that $\|u_n\|_\varepsilon\to+\infty$ as $n\to+\infty$, and derive a contradiction in Lemmas 2.2 and 2.3.

So assume $\|u_n\|_\varepsilon\to\infty$ and set $\omega_n=u_n/\|u_n\|_\varepsilon$. By the boundedness of $\{\omega_n\}$ in E_ε there exists $\omega\in E_\varepsilon$ satisfying, after passing to a subsequence if necessary,

$$\omega_n\rightharpoonup\omega\quad\text{weakly in } E_\varepsilon, \tag{2-12}$$

$$\omega_n\to\omega\quad\text{strongly in } L^t_{\mathrm{loc}}(\mathbb{R}^N)\text{ with } 2\le t<\frac{2N}{N-2}, \tag{2-13}$$

$$\omega_n\to\omega\quad\text{almost everywhere in } \mathbb{R}^N. \tag{2-14}$$

Lemma 2.2. *Under the assumptions* (H_1)–(H_3), *if* $\|u_n\|_\varepsilon\to+\infty$, *then* $\omega(x)\ge 0$ *with* $\omega(x)\not\equiv 0$ *and* ω *solves the following equation weakly in* E_ε:

$$-\varepsilon^2\triangle u+V(x)u=\chi_\Lambda(x)lK(x)u. \tag{2-15}$$

Proof. Since it follows from Lemma 2.1 that $I'_\varepsilon(u_n)u_n^- = o_n(1)$, then $\|u_n^-\|_\varepsilon = o_n(1)$ holds true. This means $\|\omega_n^-\|_\varepsilon = o_n(1)$; hence $\omega^- = 0$ and $\omega \geq 0$.

On the other hand, by Lemma 2.1 and (2-4), we have

$$o_n(1) = \frac{I'_\varepsilon(u_n)u_n}{\|u_n\|_\varepsilon^2} = 1 - \int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n^2\,dx - \int_{\mathbb{R}^N\setminus\Lambda}\frac{g_\varepsilon(x,u_n)u_n}{\|u_n\|_\varepsilon^2}dx$$
$$\geq 1 - \int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n^2\,dx - C\varepsilon;$$

here and below $o_n(1)$ denotes a quantity that vanishes as $n \to \infty$.

From this, for small ε and large n we obtain

$$C\int_\Lambda \omega_n^2\,dx \geq \int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n^2\,dx \geq 1 - o_n(1) - C\varepsilon \geq \tfrac{1}{2}. \tag{2-16}$$

Combining (2-13) with (2-16) yields $\int_\Lambda \omega^2\,dx \geq C$, which obviously leads to $\omega \not\equiv 0$.

Next we prove that ω satisfies (2-15).

In fact, for any $\phi \in C_0^\infty(\mathbb{R}^N)$, we have $\dfrac{I'_\varepsilon(u_n)\phi}{\|u_n\|_\varepsilon} = o_n(1)$, which is equivalent to

$$\int_{\mathbb{R}^N}(\varepsilon^2\nabla\omega_n\nabla\phi + V(x)\omega_n\phi)\,dx = \int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n\phi\,dx - \int_{\mathbb{R}^N\setminus\Lambda}\frac{g_\varepsilon(x,u_n)}{\|u_n\|_\varepsilon}\phi\,dx + o_n(1). \tag{2-17}$$

Due to (2-12) and (2-17), there holds

$$\int_{\mathbb{R}^N}(\varepsilon^2\nabla\omega\nabla\phi + V(x)\omega\phi)\,dx = \lim_{n\to\infty}\left(\int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n\phi\,dx - \int_{\mathbb{R}^N\setminus\Lambda}\frac{g_\varepsilon(x,u_n)}{\|u_n\|_\varepsilon}\phi\,dx\right). \tag{2-18}$$

Noting that

$$\int_\Lambda\left(K(x)\frac{f(u_n)}{u_n}\omega_n\right)^2dx \leq C\int_\Lambda V(x)\omega_n^2\,dx \leq C$$

and

$$K(x)\frac{f(u_n)}{u_n}\omega_n \to lK(x)\omega \quad \text{almost everywhere in } \Lambda,$$

we get

$$\lim_{n\to\infty}\int_\Lambda K(x)\frac{f(u_n)}{u_n}\omega_n\phi\,dx = \int_\Lambda lK(x)\omega\phi\,dx. \tag{2-19}$$

In addition, one has

$$\lim_{n\to\infty}\int_{\mathbb{R}^N\setminus\Lambda}\frac{g(x,u_n)}{\|u_n\|_\varepsilon}\phi\,dx=0. \tag{2-20}$$

Substituting (2-19) and (2-20) into (2-18) yields the conclusion of Lemma 2.2. □

Lemma 2.3. *Under the assumptions* (H_1)–(H_4), *Equation* (2-15) *has no nontrivial solution* $\omega(x)$ *with* $\omega(x)\ge 0$.

Proof. By (1-7), along the proof line of (2-9), there exists $v_\varepsilon\in C_0^\infty(\Lambda)$ such that

$$\frac{\int_\Lambda(\varepsilon^2|\nabla v_\varepsilon|^2+V(x)|v_\varepsilon|^2)\,dx}{\int_\Lambda K(x)v_\varepsilon^2\,dx}<l.$$

Let Λ_0 be a set satisfying $\operatorname{supp} v_\varepsilon\subsetneq\Lambda_0\subsetneq\Lambda$ and

$$\mu_0=\inf_{\varphi\in C_0^\infty(\Lambda_0)}\frac{\int_{\Lambda_0}(\varepsilon^2|\nabla\varphi|^2+V(x)|\varphi|^2)\,dx}{\int_{\Lambda_0}K(x)\varphi^2\,dx};$$

then $\mu_0<l$.

Due to the compactness of the embedding $H_0^1(\Lambda_0)\hookrightarrow L^2(\Lambda_0)$, a direct argument then shows there exists a nontrivial nonnegative function $v_0\in H_0^1(\Lambda_0)$ such that

$$-\varepsilon^2\triangle v_0+V(x)v_0=\mu_0K(x)v_0,\quad x\in\Lambda_0. \tag{2-21}$$

In addition, by the strong maximum principle [Gilbarg and Trudinger 1983, Lemma 3.4 and Theorem 3.5], one has

$$v_0>0,\quad x\in\Lambda_0,\qquad \frac{\partial v_0}{\partial\nu}<0,\quad x\in\partial\Lambda_0.$$

Moreover, we can assert that if $\omega\ge 0$ is a nontrivial solution of (2-15), then $\omega\not\equiv 0$ in Λ for small ε. Indeed, if $\omega\equiv 0$ in Λ, we get $\|\omega\|_\varepsilon^2=0$ by (2-15), which yields a contradiction since ω is nontrivial.

Hence, we can choose the domain Λ_0 so that $\int_{\Lambda_0}K(x)v_0\omega\,dx>0$. In this case, we have

$$\begin{aligned}\mu_0\int_{\Lambda_0}K(x)v_0\omega\,dx&=\int_{\Lambda_0}(-\varepsilon^2\triangle v_0+V(x)v_0)\omega\,dx\\&=l\int_{\Lambda_0}K(x)v_0\omega\,dx-\int_{\partial\Lambda_0}\varepsilon^2\frac{\partial v_0}{\partial\nu}\omega\,d\sigma\ge l\int_{\Lambda_0}K(x)v_0\omega\,dx.\end{aligned}$$

This means $\mu_0\ge l$, which contradicts with $\mu_0<l$. Hence we complete the proof of Lemma 2.3. □

Combining Lemma 2.2 with Lemma 2.3, we immediately obtain the announced result:

Lemma 2.4. *Under the assumptions* (H_1)–(H_4), *the sequence* $\{u_n\}$ *in Lemma 2.1 is bounded in* E_ε.

Next we state the main result in this section.

Lemma 2.5. *Under the assumptions* (H_1)–(H_4), *for small* $\varepsilon > 0$, *the modified functional* I_ε *of* (2-1) *has a nontrivial critical point* $u_\varepsilon \in E_\varepsilon$ *with the level* $I_\varepsilon(u_\varepsilon) = c_\varepsilon$.

Proof. The boundedness of $\{u_n\}$ in E_ε implies that there exists $u_\varepsilon \in E_\varepsilon$ satisfying, after passing to a subsequence if necessary,

$$u_n \rightharpoonup u_\varepsilon \quad \text{weakly in } E_\varepsilon, \tag{2-22}$$

$$u_n \to u_\varepsilon \quad \text{strongly in } L^t_{\text{loc}}(R^N) \text{ with } 2 \le t < \frac{2N}{N-2}. \tag{2-23}$$

Next we show $\|u_n\|_\varepsilon \to \|u_\varepsilon\|_\varepsilon$ as $n \to \infty$, which together with (2-22) leads to the strong convergence of $\{u_n\}$ in E_ε.

In fact, by $I'_\varepsilon(u_n)u_\varepsilon \to 0$ and (2-22), we arrive at

$$\begin{aligned} o_n(1) = \int_{\mathbb{R}^N} (\varepsilon^2 \nabla u_n \cdot \nabla u_\varepsilon + V(x) u_n u_\varepsilon)\, dx \\ - \int_\Lambda K(x) f(u_n) u_\varepsilon\, dx - \int_{\mathbb{R}^N \setminus \Lambda} g_\varepsilon(x, u_n) u_\varepsilon\, dx, \end{aligned} \tag{2-24}$$

which implies

$$\|u_\varepsilon\|_\varepsilon^2 - \int_\Lambda K(x) f(u_n) u_\varepsilon\, dx - \int_{\mathbb{R}^N \setminus \Lambda} g_\varepsilon(x, u_n) u_\varepsilon\, dx = o_n(1). \tag{2-25}$$

In addition, we have

$$\|u_n\|_\varepsilon^2 - \int_\Lambda K(x) f(u_n) u_n\, dx - \int_{\mathbb{R}^N \setminus \Lambda} g_\varepsilon(x, u_n) u_n\, dx = I'_\varepsilon(u_n) u_n = o_n(1). \tag{2-26}$$

On the other hand, by use of (2-23), we find

$$\lim_{n\to\infty} \int_\Lambda K(x) f(u_n) u_n\, dx = \lim_{n\to\infty} \int_\Lambda K(x) f(u_n) u_\varepsilon\, dx, \tag{2-27}$$

and for any fixed large $R > 0$ (without loss of generality, $\Lambda \subset B_R$ is assumed),

$$\lim_{n\to\infty} \int_{B_R \setminus \Lambda} g_\varepsilon(x, u_n) u_n\, dx = \lim_{n\to\infty} \int_{B_R \setminus \Lambda} g_\varepsilon(x, u_n) u_\varepsilon\, dx. \tag{2-28}$$

Thus, in order to obtain $\|u_n\|_\varepsilon \to \|u_0\|_\varepsilon$, it follows from (2-25)–(2-28) that we only need to prove the following statement:

For any given $\delta > 0$, there exists $R > 0$ such that for all n

$$\left| \int_{\mathbb{R}^N \setminus B_R} g_\varepsilon(x, u_n) u_\varepsilon\, dx \right| < \delta, \quad \left| \int_{\mathbb{R}^N \setminus B_R} g_\varepsilon(x, u_n) u_n\, dx \right| < \delta. \tag{2-29}$$

It is only enough to check the first inequality in (2-29) since the second one is similar. By direct computations, we have

$$\left|\int_{\mathbb{R}^N\setminus B_R} g_\varepsilon(x,u_n)u_\varepsilon\,dx\right| \le \frac{C\varepsilon}{R^{(\theta_0-2)/2}}\|u_n\|_\varepsilon\|u_\varepsilon\|_\varepsilon \to 0 \quad \text{as } R\to\infty.$$

The last estimate follows from the choice of $\theta_0 > 2$ and the boundedness of $\{u_n\}$. Thus we have shown that $u_n \to u_\varepsilon$ in E_ε, which completes the proof of Lemma 2.5. □

Remark 2.1. Since $h_\varepsilon(x,\xi)$ is Lipschitzian continuous in ξ for fixed x, it follows from second order elliptic regularity theory that u_ε is a classical solution of (2-1). Furthermore, $u_\varepsilon > 0$.

3. Solving a related constant coefficient problem

In this section, toward the proof of Theorem 1.1 in Section 4, we study the asymptotically linear problem (1-3) with constant coefficients. Some conclusions and techniques in this section are very similar to those in Section 2, but we give the argument anyway, for the reader's convenience.

We consider the functional $I^\xi(u)$ defined in (1-4) for $u\in E\equiv H_r^1(\mathbb{R}^N)=\{u\in H^1(\mathbb{R}^N): u(x)=u(|x|)\}$. Set

$$\|u\|_\xi = \left(\int_{\mathbb{R}^N}(|\nabla u|^2+V(\xi)|u|^2)\,dx\right)^{1/2},$$

which is a norm equivalent to the $H^1(\mathbb{R}^N)$ norm. We now verify that I^ξ has a mountain pass geometry. Similar to the proof of (2-5), there are two small numbers $\delta, r>0$ such that

$$I^\xi(u) \ge \tfrac{1}{2}\|u\|_\xi^2 - C\delta\|u\|_\xi^2 - C\|u\|_\xi^{2^*} \ge \tfrac{1}{4}\|u\|_\xi^2 \quad \text{for } \|u\|_\xi \le r. \tag{3-1}$$

In addition, by (2-9), there exists a function $\varphi\in H^1(\mathbb{R}^N)\setminus\{0\}$ such that

$$\frac{\int_{\mathbb{R}^N}(|\nabla\varphi|^2+V(\xi)|\varphi|^2)\,dx}{\int_{\mathbb{R}^N}K(\xi)\varphi^2\,dx} < l. \tag{3-2}$$

Let φ^* be the symmetrization of φ (see [Berestycki and Lions 1983, Appendix A.III]). Then $\varphi^*(x)=\varphi^*(|x|)$ is a nonnegative function. Moreover, for any continuous function $H(s)$ such that $H(\varphi(x))$ is integrable in $\mathbb{R}^N$ there holds

$$\int_{\mathbb{R}^N} H(\varphi^*)\,dx = \int_{\mathbb{R}^N} H(\varphi)\,dx \tag{3-3}$$

and

$$\int_{\mathbb{R}^N} |\nabla\varphi^*|^2 dx \le \int_{\mathbb{R}^N}|\nabla\varphi|^2 dx. \tag{3-4}$$

By (3-2)–(3-4), we have

$$\frac{\int_{\mathbb{R}^N}(|\nabla\varphi^*|^2+V(\xi)|\varphi^*|^2)\,dx}{\int_{\mathbb{R}^N}K(\xi)|\varphi^*|^2\,dx}<l; \tag{3-5}$$

by the same argument as in (2-11) we can derive

$$\liminf_{t\to+\infty}\frac{I^\xi(t\varphi^*)}{t^2}<0. \tag{3-6}$$

Thus there exists $t_0>0$ such that $I^\xi(t_0\varphi^*)<0$, showing that I^ξ has a mountain pass geometry. Define the mountain level

$$c_1=\inf_{\gamma\in\Gamma}\max_{0\le t\le 1}I^\xi(\gamma(t)), \tag{3-7}$$

where $\Gamma=\{\gamma\in C([0,1],E):\gamma(0)=0,\ I^\xi(\gamma(1))<0\}$.

The next two lemmas are established analogously to Lemma 2.1 and Lemma 2.4, respectively.

Lemma 3.1. *There exists a sequence* $\{u_n\}\subset E$ *such that* $I^\xi(u_n)\to c_1$ *and*

$$\|(I^\xi)'(u_n)\|_{H^{-1}}(1+\|u_n\|_\xi)\to 0 \quad \text{as } n\to\infty.$$

Lemma 3.2. *The sequence* $\{u_n\}$ *given in Lemma 3.1 is bounded in* E.

Based on Lemma 3.2, we have:

Lemma 3.3. *The functional* I^ξ *has a positive critical point* $\omega\in H_r^1(\mathbb{R}^N)$ *with the level* $I^\xi(\omega)=c_1$. *That is,* ω *is a radially symmetric solution to the problem* (1-3).

Proof. It follows from the boundedness of $\{u_n\}$ in Lemma 3.2 that there exists $\omega\in E$ satisfying, after passing to a subsequence if necessary,

$$u_n\rightharpoonup\omega \quad \text{weakly in } E, \tag{3-8}$$

$$u_n\to\omega \quad \text{strongly in } L^t_{\text{loc}}(R^N) \text{ with } 2\le t<\frac{2N}{N-2}. \tag{3-9}$$

As in Lemma 2.5, we only need to show $\|u_n\|_\xi\to\|\omega\|_\xi$ as $n\to\infty$, which together with (3-8) leads to the strong convergence of $\{u_n\}$ in E.

Since $(I^\xi)'(u_n)\omega\to 0$ and using (3-8), we arrive at

$$o_n(1)=\int_{\mathbb{R}^N}(\nabla u_n\cdot\nabla\omega+V(\xi)u_n\omega)\,dx-\int_{\mathbb{R}^N}K(\xi)f(u_n)\omega\,dx.$$

This implies

$$\|\omega\|_\xi^2-\int_{\mathbb{R}^N}K(\xi)f(u_n)\omega\,dx=o_n(1). \tag{3-10}$$

In addition, we have

$$\|u_n\|_\xi^2 - \int_{\mathbb{R}^N} K(\xi) f(u_n) u_n \, dx = o_n(1). \tag{3-11}$$

On the other hand, it follows from (3-9) and the Hölder inequality that

$$\begin{aligned} \left| \int_{\mathbb{R}^N} f(u_n)(u_n - \omega)\, dx \right| &\le C \int_{\mathbb{R}^N} |u_n|\, |u_n - \omega|\, dx \\ &\le C \|u_n\|_{L^2} \|u_n - \omega\|_{L^2} = o_n(1). \end{aligned} \tag{3-12}$$

Hence, collecting (3-10)–(3-12) yields $\|u_n\|_\xi \to \|\omega\|_\xi$ as $n \to \infty$ and $I^\xi(\omega) = c_1$. Moreover, ω is a nontrivial critical point of I^ξ to E. By the principle of symmetric criticality (see [Willem 1996, Theorem 1.28]), ω is also a nontrivial critical point of I^ξ to $H^1(\mathbb{R}^N)$. In addition, $\omega > 0$ can be shown as in Remark 2.1. Therefore, Lemma 3.3 is proved. □

Next we assert that the radial function $\omega(x) = \omega(V(\xi), K(\xi); x)$ found in Lemma 3.3 is a ground state of the functional I^ξ, that is,

$$G(\xi) = I^\xi(\omega). \tag{3-13}$$

Obviously, $G(\xi) \le I^\xi(\omega)$ since $\omega \in \mathcal{M}^\xi$, ω being defined in (1-5). What is left is to show $I^\xi(\omega) \le G(\xi)$ in order to get (3-13).

For any $u \in \mathcal{M}^\xi$, let u^* be the symmetrization of u. Then $u^* \in H^1(\mathbb{R}^N)$ and $u^* \ge 0$. Consider the function

$$J(t) = I^\xi(tu^*) = \frac{t^2}{2} \int_{\mathbb{R}^N} \big(|\nabla u^*|^2 + V(\xi)|u^*|^2\big)\, dx - K(\xi) \int_{\mathbb{R}^N} F(tu^*)\, dx. \tag{3-14}$$

A direct computation yields

$$\begin{aligned} \lim_{t\to\infty} \frac{J(t)}{t^2} &= \frac{1}{2} \int_{\mathbb{R}^N} \big(|\nabla u^*|^2 + V(\xi)|u^*|^2\big)\, dx - \frac{lK(\xi)}{2} \int_{\mathbb{R}^N} |u^*|^2\, dx \\ &\le \frac{K(\xi)}{2} \int_{\mathbb{R}^N} \Big(\frac{f(u^*)}{u^*} - l\Big)|u^*|^2\, dx. \end{aligned} \tag{3-15}$$

In addition, by the Strauss inequality [Willem 1996, Lemma 4.5], we have $u^*(x) \to 0$ as $|x| \to +\infty$. On the other hand, it follows from $\lim_{s\to 0^+} f(s)/s = 0$ that there exists $\Omega \subset \mathbb{R}^N$ with $|\Omega| > 0$ such that

$$\Big(\frac{f(u^*(x))}{u^*(x)} - l\Big)|u^*(x)|^2 < 0 \tag{3-16}$$

for $x \in \Omega$. If $x \in R^N \setminus \Omega$, the left-hand side of (3-16) is nonnegative, by (H_3). Thus, we have

$$\int_{\mathbb{R}^N} \Big(\frac{f(u^*)}{u^*} - l\Big)|u^*|^2\, dx < 0.$$

This, together with (3-15), yields that there exists $t_0 = t_0(u^*) > 0$ such that $I^\xi(t_0 u^*) < 0$. Define $\gamma(t) = t t_0 u^*$; then $\gamma(t) \in \Gamma$. By the definition of c_1, we see that

$$I^\xi(\omega) = c_1 \le \max_{0\le t\le 1} I^\xi(t t_0 u^*) \le \max_{0\le t\le 1} I^\xi(t t_0 u) \le \max_{t\ge 0} I^\xi(tu) = I^\xi(u).$$

Since u is arbitrary, we have $I^\xi(\omega) \le G(\xi)$ and (3-13) is shown.

Remark 3.1. By the Gidas–Ni–Nirenberg result [Fei and Yin 2010, Theorem 2 and following remark], 0 is the unique maximum point of $\omega(x)$ in $\mathbb{R}^N$. This motivates us to establish a similar result in Lemma 4.5 in Section 4 below.

Finally, we show that the ground energy function $G(\xi)$ is continuous for $\xi \in \bar{\Lambda}$. Here we point out that the continuity of $G(\xi)$ corresponding to the superlinear case of $f(u)$ in (1-3) has been proved in [Wang and Zeng 1997].

Lemma 3.4. *$G(\xi)$ is continuous with respect to $\xi \in \bar{\Lambda}$.*

Proof. Consider a sequence $\{\xi_j\} \subseteq \bar{\Lambda}$ such that $\xi_j \to \xi_0 \in \bar{\Lambda}$ as $j \to +\infty$. Then $V(\xi_j) \to V(\xi_0)$, $K(\xi_j) \to K(\xi_0)$ as $j \to \infty$. Set

$$I_j(u) = \frac{1}{2}\int_{\mathbb{R}^N} |\nabla u|^2\,dx + \frac{V(\xi_j)}{2}\int_{\mathbb{R}^N} |u|^2\,dx - K(\xi_j)\int_{\mathbb{R}^N} F(u)\,dx,$$

$$I_0(u) = \frac{1}{2}\int_{\mathbb{R}^N} |\nabla u|^2\,dx + \frac{V(\xi_0)}{2}\int_{\mathbb{R}^N} |u|^2\,dx - K(\xi_0)\int_{\mathbb{R}^N} F(u)\,dx,$$

and

$$\Gamma_j = \{\gamma \in C([0,1], E) : \gamma(0) = 0,\ I_j(\gamma(1)) < 0\},$$

$$\Gamma_0 = \{\gamma \in C([0,1], E) : \gamma(0) = 0,\ I_0(\gamma(1)) < 0\}.$$

From (3-7) and (3-13), we have

$$G(\xi_j) = \inf_{\gamma\in\Gamma_j}\max_{0\le t\le 1} I_j(\gamma(t)) \quad\text{and}\quad G(\xi_0) = \inf_{\gamma\in\Gamma_0}\max_{0\le t\le 1} I_0(\gamma(t)).$$

The proof of the continuity of $G(\xi)$ now proceeds in two steps.

Step 1: $\limsup_{j\to\infty} G(\xi_j) \le G(\xi_0)$.

For any fixed path $\gamma(t)$ satisfying $\gamma(0) = 0$ and $I_0(\gamma(1)) < 0$, we have $I_j(\gamma(1)) < 0$ for large j and

$$\limsup_{j\to\infty} G(\xi_j) \le \limsup_{j\to\infty} \max_{0\le t\le 1} I_j(\gamma(t)) = \max_{0\le t\le 1} I_0(\gamma(t)).$$

Since the path $\gamma(t)$ is arbitrary, this yields

$$\limsup_{j\to\infty} G(\xi_j) \le G(\xi_0). \tag{3-17}$$

Step 2: $\liminf_{j\to\infty} G(\xi_j) \geq G(\xi_0)$.

We split this step into four parts.

Let $\omega_j(x) \in H_r^1(\mathbb{R}^N)$ satisfy $G(\xi_j) = I_j(\omega_j(x))$ (the existence of $\omega_j(x)$ has been shown in Lemma 3.3).

Part 1. $\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx$ is uniformly bounded with respect to j.

According to Pohozaev identity [Willem 1996, Appendix], we have

$$\frac{N-2}{2N}\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx = -\frac{V(\xi_j)}{2}\int_{\mathbb{R}^N} |\omega_j|^2\,dx + K(\xi_j)\int_{\mathbb{R}^N} F(\omega_j)\,dx.$$

This implies

$$G(\xi_j) = I_j(\omega_j) = \frac{1}{N}\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx. \tag{3-18}$$

It follows from (3-17) and (3-18) that there is a positive constant C such that

$$\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx \leq C \quad \text{for any } j. \tag{3-19}$$

Part 2. $\int_{\mathbb{R}^N} \omega_j^2\,dx$ has a uniform upper bound independent of j.

Note that up to a subsequence, there exists a radial symmetric function $\omega(x)$ such that, as $j \to \infty$,

$$\omega_j \rightharpoonup \omega, \quad \text{weakly in } \mathcal{D}^{1,2}(\mathbb{R}^N), \tag{3-20}$$

$$\omega_j \to \omega, \quad \text{strongly in } L^t_{\text{loc}}(\mathbb{R}^N), \quad 1 \leq t < \frac{2N}{N-2}, \tag{3-21}$$

$$\omega_j \to \omega, \quad \text{almost everywhere in } \mathbb{R}^N. \tag{3-22}$$

By the Strauss inequality [Berestycki and Lions 1983, Lemma A.III, p. 340] for the radial function in $\mathcal{D}^{1,2}(\mathbb{R}^N)$, we have

$$|\omega_j(x)|^2 \leq C(N)|x|^{2-N}\int_{\mathbb{R}^N} |\nabla\omega_j(x)|^2\,dx, \quad \text{for all } |x| \geq 1, \tag{3-23}$$

where the positive constant $C(N)$ only depends on N.

Since $f(s)/s \to 0$ as $s \to 0$ by the assumption (H_2), we get from (3-23) and the fact that $N \geq 5$ that

$$\frac{f(\omega_j(x))}{\omega_j(x)} \to 0 \quad \text{as } |x| \to \infty \text{ uniformly with respect to } j.$$

This implies that there exists a large number $R > 0$ such that

$$\int_{|x|\geq R}\Bigl(V(\xi_j) - K(\xi_j)\frac{f(\omega_j)}{\omega_j}\Bigr)|\omega_j|^2\,dx \geq C\int_{|x|\geq R} |\omega_j|^2\,dx, \tag{3-24}$$

where $C > 0$ is independent of R and j.

It follows from (3-24) and the partial differential equation satisfied by ω_j that for large R,

$$C\int_{|x|\ge R}|\omega_j|^2\,dx \le \int_{|x|\ge R}\Big(V(\xi_j)-K(\xi_j)\frac{f(\omega_j)}{\omega_j}\Big)|\omega_j|^2\,dx \le C\int_{|x|\le R}|\omega_j|^2\,dx \to C\int_{|x|\le R}|\omega|^2\,dx \quad \text{as } j\to\infty. \tag{3-25}$$

Combining (3-24) with (3-25) yields that $\int_{\mathbb{R}^N}|\omega_j|^2\,dx$ has a uniform supper bound with respect to j. Thus $\omega\in L^2(\mathbb{R}^N)$ and further $\omega\in H^1(\mathbb{R}^N)$. Moreover, ω is a solution of the equation

$$-\triangle\omega(x)+V(\xi_0)\omega(x)=K(\xi_0)f(\omega),\quad x\in\mathbb{R}^N. \tag{3-26}$$

Part 3. $\int_{\mathbb{R}^N}|\omega_j|^2\,dx$ has a uniform positive lower bound with respect to j.

We now show that $\int_{\mathbb{R}^N}|\omega_j|^2\,dx$ has a uniform positive lower bound with respect to j. If so, this assertion together with (3-21) and (3-25) will yield

$$\omega\not\equiv 0. \tag{3-27}$$

Note that $V(\xi_0)/K(\xi_0)<l$ and $V(\xi_j)\to V(\xi_0)$, $K(\xi_j)\to K(\xi_0)$ as $j\to\infty$. Thus we can choose a fixed small number $\eta>0$ satisfying

$$\frac{V(\xi_0)-\eta}{K(\xi_0)+\eta}<l, \tag{3-28}$$

and, for large j,

$$V(\xi_j)>V(\xi_0)-\eta,\quad K(\xi_j)<K(\xi_0)+\eta. \tag{3-29}$$

Let m_0 be the ground energy of the functional

$$H^1(\mathbb{R}^N)\ni u\mapsto \frac12\int_{\mathbb{R}^N}|\nabla u|^2\,dx+\frac{V(\xi_0)-\eta}{2}\int_{\mathbb{R}^N}|u|^2\,dx-(K(\xi_0)+\eta)\int_{\mathbb{R}^N}F(u)\,dx$$

in the Nehari manifold $\mathcal{M}^\eta$, which is defined as

$$\mathcal{M}^\eta=\Big\{u\in H^1(\mathbb{R}^N)\setminus\{0\}:\int_{\mathbb{R}^N}|\nabla u|^2\,dx+(V(\xi_0)-\eta)\int_{\mathbb{R}^N}|u|^2\,dx=(K(\xi_0)+\eta)\int_{\mathbb{R}^N}f(u)u\,dx\Big\}.$$

By (3-28) and the similar proof on Lemma 3.3, one can show that m_0 is achieved and is positive (in the arguments of Lemma 3.3, we have used the condition $V(\xi)/K(\xi)<l$ parallel to (3-28)).

Consider the function

$$g_j(t) = \int_{\mathbb{R}^N} |\nabla(t\omega_j)|^2\,dx + (V(\xi_0)-\eta)\int_{\mathbb{R}^N} |t\omega_j|^2\,dx - (K(\xi_0)+\eta)\int_{\mathbb{R}^N} f(t\omega_j)t\omega_j\,dx.$$

Recalling that $\lim_{s\to 0} F(s)/s^2 = \lim_{s\to 0} f(s)/(2s) = 0$, we get $g_j(t) > 0$ for $0 < t \ll 1$. In addition, by (3-29) we get $g_j(1) < I'_j(\omega_j)\omega_j = 0$. Therefore there exists a $t_j \in (0, 1)$ such that $g_j(t_j\omega_j) = 0$, that is,

$$\frac{1}{2}\int_{\mathbb{R}^N} |\nabla(t_j\omega_j)|^2\,dx + \frac{V(\xi_0)-\eta}{2}\int_{\mathbb{R}^N} |t_j\omega_j|^2\,dx - (K(\xi_0)+\eta)\int_{\mathbb{R}^N} F(t_j\omega_j)\,dx \ge m_0. \tag{3-30}$$

Set

$$h_j(t) = \frac{1}{2}\int_{\mathbb{R}^N} |\nabla(t\omega_j)|^2\,dx + \frac{V(\xi_j)}{2}\int_{\mathbb{R}^N} |t\omega_j|^2\,dx - K(\xi_j)\int_{\mathbb{R}^N} F(t\omega_j)\,dx.$$

It follows from a direct computation and the assumption (H_3) that, for $t \in (0, 1]$,

$$h'_j(t) = t\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx + tV(\xi_j)\int_{\mathbb{R}^N} |\omega_j|^2\,dx - K(\xi_j)\int_{\mathbb{R}^N} f(t\omega_j)\omega_j\,dx \ge 0. \tag{3-31}$$

Combining (3-29), (3-30), and (3-31), we obtain, for large j,

$$I_j(\omega_j) \ge m_0.$$

Together with (3-18), this yields, for large j,

$$\frac{1}{N}\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx = I_j(\omega_j) \ge m_0. \tag{3-32}$$

In addition, since

$$\left(\frac{F(s)}{s^2}\right)' = \frac{f(s)s - 2F(s)}{s^3} \ge 0 \quad \text{and} \quad \lim_{s\to+\infty}\frac{F(s)}{s^2} = \lim_{s\to+\infty}\frac{f(s)}{2s} = \frac{l}{2},$$

we have

$$0 \le \frac{F(s)}{s^2} \le \frac{l}{2}, \quad s \ne 0. \tag{3-33}$$

Therefore, by (3-32), (3-33), and the Pohozaev identity we find that

$$0 < C \le \frac{N-2}{2N}\int_{\mathbb{R}^N} |\nabla\omega_j|^2\,dx \le C\int_{\mathbb{R}^N} \omega_j^2\,dx, \tag{3-34}$$

where C is a generic positive constant independent of j, that is, $\int_{\mathbb{R}^N} |\omega_j|^2\,dx$ have a uniform positive lower bound with respect to j.

Part 4. $\lim_{j\to\infty}\int_{\mathbb{R}^N} F(\omega_j)\,dx = \int_{\mathbb{R}^N} F(\omega)\,dx$.

In order to show

$$\lim_{j\to\infty}\int_{\mathbb{R}^N} F(\omega_j)\,dx = \int_{\mathbb{R}^N} F(\omega)\,dx, \tag{3-35}$$

then by (3-21) we only need to prove:

For any given $\delta > 0$, there exists $R > 0$ such that, for large j,

$$\left|\int_{\mathbb{R}^N\setminus B_R} F(\omega_j)\,dx\right| < \delta. \tag{3-36}$$

In fact, if we set η_R to be a smooth cut-off function such that $\eta_R = 0$ for $|x| \le \frac{R}{2}$, $\eta_R = 1$ for $|x| \ge R$ and $|\nabla\eta| \le \frac{4}{R}$, then multiplying by $\eta_R\omega_j$ the equation

$$-\Delta\omega_j + V(\xi_j)\omega_j = K(\xi_j)f(\omega_j), \quad x \in \mathbb{R}^N,$$

yields, for large R and j,

$$C\int_{|x|\ge R}\big(|\nabla\omega_j|^2 + |\omega_j|^2\big)\,dx \le \frac{C}{R} \to 0 \quad \text{as } R \to +\infty,$$

which means that (3-36) and further (3-35) hold.

Finally, we show $\liminf_{j\to\infty} G(\xi_j) \ge G(\xi_0)$. In view of (3-35), (3-26)–(3-27) and the fact that $G(\xi_0)$ is the ground energy of the functional I_0, we have

$$\begin{aligned}\liminf_{j\to\infty} G(\xi_j) &\\ = \liminf_{j\to\infty}&\left\{\frac{1}{2}\int_{\mathbb{R}^N}\big(|\nabla\omega_j|^2 + V(\xi_j)|\omega_j|^2\big)\,dx - K(\xi_j)\int_{\mathbb{R}^N} F(\omega_j)\,dx\right\} \ge G(\xi_0).\end{aligned} \tag{3-37}$$

Thus the continuity of $G(\xi)$ is derived from (3-17) and (3-37), that is, Lemma 3.4 is proved. □

4. The proof of Theorem 1.1

At first, we intend to obtain an upper bound estimate of the critical value c_ε corresponding to the functional $I_\varepsilon(u)$ defined in Section 2, which will play a crucial role in establishing the concentration and decay estimates of solution u_ε to Equation (2-1). From the decay estimates of u_ε we can show $g_\varepsilon(x, u_\varepsilon) \equiv K(x)f(u_\varepsilon)$ in $\mathbb{R}^N \setminus \Lambda$ and subsequently complete the proof of Theorem 1.1.

Lemma 4.1. *Under the hypotheses* (H_1)–(H_4), *and with* c_0 *as in* (H_4), *we have, for small* $\varepsilon > 0$,

$$c_\varepsilon \le (c_0 + o_\varepsilon(1))\varepsilon^N. \tag{4-1}$$

Proof. For $\xi \in \Lambda$, choose $R > 0$ such that $B_R(\xi) \subset \Lambda$. Define a smooth cut-off function $\eta : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\eta(t) = 1$ if $0 \le t \le \frac{R}{4}$, $\eta(t) = 0$ if $t \ge \frac{R}{2}$ and $|\eta'(t)| \le \frac{8}{R}$. Set

$$w_\varepsilon(x) = \eta(|x-\xi|)\omega\Big(\frac{x-\xi}{\varepsilon}\Big),$$

where $\omega(x) = \omega(V(\xi), K(\xi); x)$ is the solution of (1-3).

Noting that w_ε is compactly supported in Λ, one can get $G_\varepsilon(x, tw_\varepsilon) = 0$ for all $t \ge 0$ and $x \in \Lambda$, where $G_\varepsilon(x, u)$ is the function defined in (2-2). Then as in the argument in (2-11), there exists a sufficiently large $T > 0$ such that $I_\varepsilon(Tw_\varepsilon) < 0$. This implies that the path $\gamma_\varepsilon(t) = \{tTw_\varepsilon : t \in [0, 1]\}$ is an element of Γ_ε satisfying $c_\varepsilon \le \max_{0\le t\le 1} I_\varepsilon(\gamma_\varepsilon(t))$. Also, similar to the proof of (2-10), we infer that $I_\varepsilon(tTw_\varepsilon) = \varepsilon^N(I^\xi(tTw) + o_\varepsilon(1))$. Hence

$$\max_{0\le t\le 1} I_\varepsilon(\gamma_\varepsilon(t)) = \max_{0\le t\le 1} I_\varepsilon(tTw_\varepsilon) = \varepsilon^N(\max_{0\le t\le 1} I^\xi(tTw) + o_\varepsilon(1)) = \varepsilon^N(G(\xi) + o_\varepsilon(1)).$$

Since ξ is arbitrary and the smallness of ε is independent of the choice of ξ, then Lemma 4.1 is proved. □

The next result illustrates that the maximum of u_ε on $\overline{\Lambda}$ has a uniform positive lower bound.

Lemma 4.2. *Let x_ε be the maximum point of u_ε on $\overline{\Lambda}$, then there exists a positive constant C independent of ε such that*

$$u_\varepsilon(x_\varepsilon) \ge C. \tag{4-2}$$

Proof. By (H_2) and (H_3), for any $\delta > 0$, there exists $C_\delta > 0$ such that $f(s) \le \delta s + C_\delta |s|^2$. From $I'_\varepsilon(u_\varepsilon)u_\varepsilon = 0$, one has, for small δ and ε,

$$\begin{aligned}\|u_\varepsilon\|_\varepsilon^2 &= \int_\Lambda K(x) f(x, u_\varepsilon) u_\varepsilon\, dx + \int_{\mathbb{R}^N\setminus\Lambda} g_\varepsilon(x, u_\varepsilon) u_\varepsilon\, dx \\ &\le \frac{1}{2}\|u_\varepsilon\|_\varepsilon^2 + C\|u_\varepsilon\|_\varepsilon^2 \max_{\overline{\Lambda}} u_\varepsilon.\end{aligned}$$

Obviously this means that there exists a positive number C independent of ε such that $u_\varepsilon(x_\varepsilon) \ge C$ holds true due to $\|u_\varepsilon\|_\varepsilon \ne 0$, then the proof of Lemma 4.2 is completed. □

Note that since $f(s)$ is asymptotically linear, then in the general case, there is no number $\theta > 0$ such that $(2+\theta)F(s) \le f(s)s$ for any $s > 0$, here $F(s) = \int_0^s f(\tau)\, d\tau$. However, in the superlinear case, this property of $(2+\theta)F(s) \le f(s)s$ with $\theta > 0$ play a crucial role in obtaining the uniform boundedness of $\varepsilon^{-N}\|u_\varepsilon\|_\varepsilon$ from (4-1), which will be used to derive the decay estimate of u_ε at infinity and the concentration of u_ε as $\varepsilon \to 0$ (one can see the details in [Fei and Yin 2010] and some references therein). To overcome this kind of difficulty, next we will use some different ingredients (motivated by the proofs of Lemmas 2.2–2.3) to treat the uniform boundedness of $\varepsilon^{-N}\|u_\varepsilon\|_\varepsilon$.

Lemma 4.3. *There exists a positive constant C independent of small ε such that*

$$\varepsilon^{-N}\int_{\mathbb{R}^N}\big(\varepsilon^2|\nabla u_\varepsilon|^2+V(x)|u_\varepsilon|^2\big)\,dx\le C,\tag{4-3}$$

namely,

$$\int_{\mathbb{R}^N}\big(|\nabla v_\varepsilon|^2+V(\varepsilon x+x_\varepsilon)|v_\varepsilon|^2\big)\,dx\le C,\tag{4-4}$$

where $v_\varepsilon(x)=u_\varepsilon(\varepsilon x+x_\varepsilon)$ and the meaning of x_ε is given in Lemma 4.2.

Proof. For convenience we will use the notation $\|v_\varepsilon\|$ with

$$\|v_\varepsilon\|=\bigg(\int_{\mathbb{R}^N}\big(|\nabla v_\varepsilon|^2+V(\varepsilon x+x_\varepsilon)|v_\varepsilon|^2\big)\,dx\bigg)^{1/2}.$$

If (4-4) does not hold, there exists a sequence of functions $v_n(x)\equiv u_{\varepsilon_n}(\varepsilon_n x+x_n)$ such that $\|v_n\|\to+\infty$ as $n\to\infty$ and $v_n(x)$ satisfies

$$\begin{aligned}-\Delta v_n+V(\varepsilon_n x+x_n)v_n\\ =\chi_{\Omega_n}(x)K(\varepsilon_n x+x_n)f(v_n)+(1-\chi_{\Omega_n}(x))g_{\varepsilon_n}(\varepsilon_n x+x_n,v_n),\end{aligned}\tag{4-5}$$

where $\Omega_n\equiv\varepsilon_n^{-1}(\Lambda-x_n)$ and $x_n\equiv x_{\varepsilon_n}\in\overline{\Lambda}$.

Set $\omega_n=v_n/\|v_n\|$, then $\|\omega_n\|=1$ and $\omega_n(x)$ satisfies

$$\begin{aligned}-\Delta\omega_n+V(\varepsilon_n x+x_n)\omega_n\\ =\chi_{\Omega_n}(x)K(\varepsilon_n x+x_n)\frac{f(v_n)}{v_n}\omega_n+(1-\chi_{\Omega_n}(x))\frac{g_{\varepsilon_n}(\varepsilon_n x+x_n,v_n)}{\|v_n\|}.\end{aligned}\tag{4-6}$$

We rewrite (4-6) as

$$-\Delta\omega_n=a_n(x)\omega_n,\tag{4-7}$$

where

$$a_n(x)=-V(\varepsilon_n x+x_n)+\chi_{\Omega_n}(x)K(\varepsilon_n x+x_n)\frac{f(v_n)}{v_n}+(1-\chi_{\Omega_n}(x))\frac{g_{\varepsilon_n}(\varepsilon_n x+x_n,v_n)}{v_n}.$$

For any fixed and bounded smooth domain $\Omega\subset\mathbb{R}^N$ and fixed $\alpha\in(0,1)$, due to $\|a_n(x)\|_{L^\infty(\Omega)}\le C(\Omega)$, it follows from $\|\omega_n\|=1$ and the elliptic equation (4-7) that $\|\omega_n\|_{C^{1,\alpha}(\bar\Omega)}\le C(\Omega,\alpha)$, where the positive constants $C(\Omega)$ and $C(\Omega,\alpha)$ depend on Ω and Ω,α respectively. Therefore, for fixed $\beta\in(0,\alpha)$, there exists a subsequence still denoted by $\{\omega_n\}$ and a function ω such that $\omega_n\to\omega$ in $C^{1,\beta}(\bar\Omega)$.

In particular, for a series of closed ball sequences $B_k(0)$, $k=1,2,\dots$, then there exists a subsequence $\{\omega_{1n}\}$ and a function ω_1 such that $\omega_{1n}\to\omega_1$ in $C^{1,\beta}(B_1(0))$, and there exists a subsequence $\{\omega_{(k+1)n}\}\subseteq\{\omega_{kn}\}$ and a function ω_{k+1} such that

$\omega_{(k+1)n} \to \omega_{k+1}$ in $C^{1,\beta}(B_{k+1}(0))$ as $n \to \infty$ for $k \geq 1$. By the diagonal process, one knows that there exists a subsequence still denoted by $\{\omega_n\}$ and a function ω such that $\omega_n \to \omega$ in $C^{1,\beta}_{\text{loc}}(\mathbb{R}^N)$ as $n \to +\infty$. Of course, $\lim_{n\to\infty} \omega_n(x) = \omega(x)$ holds for $x \in \mathbb{R}^N$.

Let $x_n \to x_0 \in \overline{\Lambda}$. We consider two cases.

Case I: $\lim_{n\to\infty} \operatorname{dist}(x_n, \partial\Lambda)/\varepsilon_n = +\infty$.

In this case, by taking a subsequence, we can assume $x_n \in \Lambda$. Hence $0 \in \Omega_n$ and $\lim_{n\to\infty} \operatorname{dist}(0, \partial\Omega_n) = \lim_{n\to\infty} \operatorname{dist}(x_n, \partial\Lambda)/\varepsilon_n = +\infty$, which leads to $\lim_{n\to\infty} \Omega_n = \mathbb{R}^N$.

For any fixed $\varphi \in C_0^\infty(\mathbb{R}^N)$, there holds $\operatorname{supp}\varphi \subseteq \Omega_n$ for lager n. Multiplying φ on two hand sides of (4-6) and integrating by parts yield, for large n,

$$\int [\nabla\omega_n \nabla\varphi + V(\varepsilon_n x + x_n)\omega_n\varphi]dx = \int K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega_n\varphi\, dx. \tag{4-8}$$

Note that

$$\lim_{n\to\infty} \int [\nabla\omega_n \nabla\varphi + V(\varepsilon_n x + x_n)\omega_n\varphi]dx = \int [\nabla\omega\nabla\varphi + V(x_0)\omega\varphi]dx. \tag{4-9}$$

Next we show that

$$\lim_{n\to\infty} \int K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega_n\varphi\, dx = \int K(x_0)l\omega\varphi\, dx. \tag{4-10}$$

Define the set $A = \{x \in \mathbb{R}^N : \lim_{n\to\infty} v_n(x) = +\infty\}$ and let $A^c = \mathbb{R}^N \setminus A$. If $x \in A$, then $\lim_{n\to\infty} f(v_n(x))/v_n(x) = l$. If $x \in A^c$, since $\lim_{n\to\infty} \|v_n\| = +\infty$, we have $\omega(x) = \lim_{n\to\infty} \omega_n(x) = \liminf_{n\to\infty} v_n(x)/\|v_n\| = 0$.

On the other hand, since $K(\varepsilon_n x + x_n)$ is uniformly bounded for $x \in \operatorname{supp}\varphi$ with respect to n and $f(s)/s$ is also bounded, we have

$$\lim_{n\to\infty} \int K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega_n\varphi\, dx = \lim_{n\to\infty} \int K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega\varphi\, dx. \tag{4-11}$$

Therefore,

$$\lim_{n\to\infty} \int_{\operatorname{supp}\varphi\cap A} K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega\varphi\, dx = \int_{\operatorname{supp}\varphi\cap A} K(x_0)l\omega\varphi\, dx. \tag{4-12}$$

In addition, obviously,

$$\lim_{n\to\infty} \int_{\operatorname{supp}\varphi\cap A^c} K(\varepsilon_n x + x_n)\frac{f(v_n)}{v_n}\omega\varphi\, dx = 0 = \int_{\operatorname{supp}\varphi\cap A^c} K(x_0)l\omega\varphi\, dx. \tag{4-13}$$

Collecting (4-11)–(4-13) yields (4-10).

From (4-8)–(4-10), we arrive at

$$\int_{\mathbb{R}^N} \nabla\omega\nabla\varphi + V(x_0)\omega\varphi = \int_{\mathbb{R}^N} K(x_0)l\omega\varphi, \tag{4-14}$$

which means that ω solves

$$-\Delta\omega + V(x_0)\omega = K(x_0)l\omega. \tag{4-15}$$

Case II: $\liminf_{n\to\infty} \operatorname{dist}(x_n, \partial\Lambda)/\varepsilon_n \le C$.

In this case, we can show that $x_0 \in \partial\Lambda$. Thus, up to a rotation, we can obtain $\lim_{n\to\infty} \Omega_n = \{x \in \mathbb{R}^N : x_1 < 0\}$. Similarly to Case I, we conclude that the function $\omega(x)$ satisfies

$$-\Delta\omega + V(x_0)\omega = K(x_0)l\omega\chi_{\{x_1<0\}}(x). \tag{4-16}$$

In Case I or Case II, for any fixed bounded domain $M \subset \mathbb{R}^N$ or $M \subset \{x \in \mathbb{R}^N : x_1 < 0\}$ we have

$$\begin{aligned}\int_M \big[|\nabla\omega|^2 + V(x_0)\omega^2\big]dx &= \lim_{n\to\infty}\int_M \big[|\nabla\omega_n|^2 + V(\varepsilon_n x + x_n)\omega_n^2\big]dx \\ &\le \int_{\mathbb{R}^N} \big[|\nabla\omega_n|^2 + V(\varepsilon_n x + x_n)\omega_n^2\big]dx = 1;\end{aligned}$$

then

$$\int_{\mathbb{R}^N} \big[|\nabla\omega|^2 + V(x_0)\omega^2\big]dx \le 1, \tag{4-17}$$

which means $\omega \in H^1(\mathbb{R}^N)$ due to $V(x_0) > 0$.

It follows the equations (4-15)–(4-16), together with (4-17), the fact that $\omega \ge 0$, regularity theory and the strong maximum principle for second-order elliptic equations, that we can get $\omega(x) \in C^{2,\gamma}(\mathbb{R}^N)$ in Case I and $\omega(x) \in C^{1,\alpha}(\mathbb{R}^N)$ for any $\alpha \in (0, 1)$ in Case II, and $\omega(x) > 0$ with $\omega(x) \to 0$ as $|x| \to \infty$. However, this is contradictory with the conclusion of Lemma 2.3. Thus (4-15) and (4-16) have no nontrivial nonnegative solutions. Lemma 4.3 is proved. □

Next we assert that the maximum point of u_ε on $\bar{\Lambda}$ must lie in the interior of Λ.

Lemma 4.4. $\lim_{\varepsilon\to 0} \max_{\partial\Lambda} u_\varepsilon = 0$.

Proof. To prove this, we argue by contradiction assuming that there exists a sequence $\varepsilon_n \to 0$ as $n \to \infty$ such that for each n,

$$\max_{\partial\Lambda} u_{\varepsilon_n} \ge C > 0. \tag{4-18}$$

Let $x_n \in \partial\Lambda$ such that $u_{\varepsilon_n}(x_n) = \max_{\partial\Lambda} u_{\varepsilon_n}$ and $x_n \to x_0 \in \partial\Lambda$ as $n \to \infty$. Define $v_n(x) = u_{\varepsilon_n}(\varepsilon_n x + x_n)$, then $v_n(0) \ge C$ and $v_n(x)$ satisfies

$$\begin{aligned}&-\Delta v_n + V(\varepsilon_n x + x_n)v_n \\ &\qquad = \chi_{\Omega_n}(x)K(\varepsilon_n x + x_n)f(v_n) + (1 - \chi_{\Omega_n}(x))g_{\varepsilon_n}(\varepsilon_n x + x_n, v_n),\end{aligned} \tag{4-19}$$

where $\Omega_n \equiv \varepsilon_n^{-1}(\Lambda - x_n)$.

By (4-4), there holds

$$\int_{\mathbb{R}^N} |\nabla v_n|^2 \, dx \le C,$$

which deduces that for large n, for any fixed $R > 0$, there exists a positive constant $C(R)$ depending on R such that

$$\int_{B_R(0)} \big(|\nabla v_n|^2 + v_n^2\big) \, dx \le C(R).$$

In terms of this and (4-19), as in the proof of Lemma 4.3, there exists some nonnegative function $v(x)$ such that $v_n \to v(x)$ in $C^2_{\text{loc}}(\mathbb{R}^N)$ and $v(x)$ satisfies

$$-\Delta v + V(x_0)v = K(x_0)\chi_{\{x_1<0\}} f(v), \quad x = (x_1, x') \in \mathbb{R}^N. \tag{4-20}$$

Note that $v_n(0) \ge C$, then $v(0) \ge C$ and further $v(x) > 0$ in $\mathbb{R}^N$ by the maximum principle and Equation (4-20).

On the other hand, acting the test function $\partial_{x_1} v$ on (4-20) yields

$$\int_{\mathbb{R}^{N-1}} F(v(0, x')) \, dx' = 0,$$

which leads to $v(0, x') = 0$. However, this is impossible due to $v(x) > 0$ in $\mathbb{R}^N$. Thus Lemma 4.4 is proved. □

Lemma 4.5. *For small ε, u_ε possesses at most one maximum point x_ε on $\overline{\Lambda}$ and $G(x_\varepsilon) \to c_0$ as $\varepsilon \to 0$.*

Proof. First, we prove $G(x_\varepsilon) \to c_0$ as $\varepsilon \to 0$.

If not, we have $\limsup_{\varepsilon\to 0} G(x_\varepsilon) > c_0$. Let $x_{\varepsilon_j} \to x_0 \in \overline{\Lambda}$; then $\lim_{j\to\infty} G(x_{\varepsilon_j}) = \limsup_{\varepsilon\to 0} G(x_\varepsilon) > c_0$, which means $G(x_0) > c_0$.

Set $v_j(x) = u_{\varepsilon_j}(\varepsilon_j x + x_{\varepsilon_j})$. Then v_j solves

$$\begin{aligned} -\Delta v_j &+ V(\varepsilon_j x + x_{\varepsilon_j}) v_j \\ &= \chi_{\Omega_j}(x) K(\varepsilon_j x + x_{\varepsilon_j}) f(v_j) + (1 - \chi_{\Omega_j}(x)) g_{\varepsilon_j}(\varepsilon_j x + x_{\varepsilon_j}, v_j). \end{aligned} \tag{4-21}$$

As before, we can show that v_j converges in $C^{1,\alpha}_{\text{loc}}(\mathbb{R}^N)$ for $\alpha \in (0, 1)$ to some function v_0 that satisfies

$$-\Delta v_0 + V(x_0)v_0 = K(x_0) f(v_0), \quad x \in \mathbb{R}^N \tag{4-22}$$

or

$$-\Delta v_0 + V(x_0)v_0 = K(x_0)\chi_{\{x_1<0\}} f(v_0), \quad x = (x_1, x') \in \mathbb{R}^N. \tag{4-23}$$

The case of (4-23) can be excluded by the same argument as in Lemma 4.4, so we focus on the case of (4-22).

Set

$$J_{\varepsilon_j}(v_j) = \frac{1}{2}\int_{\mathbb{R}^N} |\nabla v_j|^2\,dx + V(\varepsilon_j x + x_{\varepsilon_j})|v_j|^2\,dx \\ - \int_{(\Lambda - x_{\varepsilon_j})/\varepsilon_j} K(\varepsilon_j x + x_{\varepsilon_j})F(v_j)\,dx - \int_{\mathbb{R}^N\setminus(\Lambda - x_{\varepsilon_j})/\varepsilon_j} G(\varepsilon_j x + x_{\varepsilon_j}, v_j)\,dx. \tag{4-24}$$

By invoking Lemma 2.2 in [del Pino and Felmer 1996] together with $2F(s) \le f(s)s$, we conclude that

$$\liminf_{j\to\infty} J_{\varepsilon_j}(v_j) \ge I^{x_0}(v_0). \tag{4-25}$$

This, together with (4-1), yields

$$c_0 \ge \liminf_{j\to\infty} \varepsilon_j^{-N} I_{\varepsilon_j}(u_{\varepsilon_j}) = \liminf_{j\to\infty} J_{\varepsilon_j}(v_j) \ge I^{x_0}(v_0) \ge G(x_0) > c_0,$$

which leads to a contradiction.

In addition, using the arguments in [del Pino and Felmer 1996, p. 133], we can show that u_ε possesses at most one maximum point x_ε on $\overline{\Lambda}$. We omit the details. This concludes the proof of Lemma 4.5. □

Next we establish a compactness result for u_ε which will be crucial to derive the decay of $u_\varepsilon(x)$ as $|x| \to \infty$.

Lemma 4.6. *For any* $\nu > 0$, *there exist* $\rho_0(\nu), \varepsilon_0(\nu) > 0$ *such that for* $\rho > \rho_0(\nu)$, $\varepsilon < \varepsilon_0(\nu)$, *then*

$$\operatorname{dist}(x_\varepsilon, M) < \nu, \tag{4-26}$$

and

$$\varepsilon^{-N}\int_{\mathbb{R}^N\setminus B_{\varepsilon\rho}(x_\varepsilon)} \big(\varepsilon^2|\nabla u_\varepsilon|^2 + V(x)|u_\varepsilon|^2\big)\,dx < \nu, \tag{4-27}$$

where $M = \{\xi \in \Lambda : G(\xi) = c_0\}$, *and the meaning of* c_0 *is given in* (1-8).

Proof. Since the first conclusion can be directly derived from Lemma 4.5, then it suffices to prove (4-27).

As a consequence of Lemma 4.5 and the assumption on $G(x)$ in (H_4), we have $d = \inf_n \operatorname{dist}(x_n, \partial\Lambda) > 0$ and $\Lambda_n = (\Lambda - x_n)/\varepsilon_n \supset B_{d/\varepsilon_n} \equiv B_{\tilde{\rho}_n}$.

If (4-27) does not hold, then we can assume that there exist $\nu_0 > 0$, $\tilde{\rho}_n > \rho_n \to +\infty$, $\varepsilon_n \to 0$ as $n \to \infty$ such that

$$\mathcal{T} \equiv \varepsilon_n^{-N}\int_{\mathbb{R}^N\setminus B_{\varepsilon_n\rho_n}(x_n)} \big(\varepsilon^2|\nabla u_n|^2 + V(x)|u_n|^2\big)\,dx > \nu_0, \tag{4-28}$$

where $x_n \equiv x_{\varepsilon_n}$, $u_n \equiv u_{\varepsilon_n}$.

Set $v_n(x) = u_n(\varepsilon_n x + x_n)$, $V_n(x) = V(\varepsilon_n x + x_n)$ and $v_n \rightharpoonup v_0$, $x_n \to x_0 \in M$ as $n \to \infty$. Then, by (4-1) and (4-25) as $n \to \infty$,

$$\tfrac{1}{2}\mathcal{T} = \varepsilon_n^{-N}\frac{1}{2}\int_{\mathbb{R}^N}\big(\varepsilon_n^2|\nabla u_n|^2 + V(x)|u_n|^2\big)\,dx \\ - \varepsilon_n^{-N}\frac{1}{2}\int_{B_{\varepsilon_n\rho_n}(x_n)}\big(\varepsilon_n^2|\nabla u_n|^2 + V(x)|u_n|^2\big)\,dx \to 0,$$

which is contradictory with (4-28). We have completed the proof of Lemma 4.6. □

Before we treat the decay estimate of u_ε at infinity, we need to establish more integration estimates based on Lemma 4.6.

Note that by the assumptions in (H_2) and (H_3), then for any fixed $p > 1$, there exists a positive constant $C_1 = C_1(p)$ depending on p such that

$$f(s) \le \frac{1}{16}\max_{\xi\in\bar{\Lambda}}\frac{V(\xi)}{K(\xi)}s + C_1|s|^p. \tag{4-29}$$

Furthermore we have a relation between $\|u\|_\varepsilon$ and $\int_\Lambda K(x)|u|^{p+1}dx$ for any $1 < p < \frac{N+2}{N-2}$ as follows, which comes from Lemma 2.1 of [Yin and Zhang 2009].

Lemma 4.7. *Under the assumptions* (H_1) *and* (H_4), *for each* $\varepsilon \in (0, 1]$, *then there exists a positive constant* $C_2 = C_2(p)$ *depending only on* p *such that*

$$\int_\Lambda K(x)|u|^{p+1}dx \le C_2\varepsilon^{-N(p-1)/2}\|u\|_\varepsilon^{p+1} \quad \textit{for } u \in E_\varepsilon, \tag{4-30}$$

where the domain Λ *is defined in the assumption* (H_4).

For later use, we introduce two fixed positive numbers $K_0 > 128$ and $c > 0$ such that $c^2 \ge 128K_0^2/(d_0^2 V_1)$, where $d_0 = \operatorname{dist}(\partial\Lambda, M) > 0$ and $V_1 = \frac{1}{2}\min_{x\in\Lambda} V(x) > 0$.

Set $\nu_0 = \min\{d_0/K_0, (16C_1C_2)^{-2/(p-1)}\}$, where C_1 and C_2 are given in (4-29)–(4-30). Take $\varepsilon_1 = \min\{\varepsilon_0(\nu_0), d_0/(K_0\rho_0(\nu_0)), (\ln 2)/c\}$, where $\varepsilon_0(\nu_0)$ and $\rho_0(\nu_0)$ are given in Lemma 4.6. From now on, we always assume $\varepsilon < \varepsilon_1$ and $\nu < \nu_0$ in (4-26)–(4-27).

It follows from (4-26) that, for $\varepsilon < \varepsilon_1$ and $\nu < \nu_0$,

$$\operatorname{dist}(x_\varepsilon, \partial\Lambda) > \frac{d_0}{2} \quad \text{and} \quad \varepsilon\rho_0(\nu_0) < \frac{d_0}{K_0}. \tag{4-31}$$

Define $\Omega_{n,\varepsilon} = \mathbb{R}^N \setminus B_{R_{n,\varepsilon}}(x_\varepsilon)$ with $R_{n,\varepsilon} = e^{c\varepsilon n}$ and let $\tilde{n} > \hat{n}$ be integers such that

$$R_{\hat{n}-1,\varepsilon} < \frac{d_0}{K_0} \le R_{\hat{n},\varepsilon}, \quad R_{\tilde{n}+2,\varepsilon} \le \frac{d_0}{2} < R_{\tilde{n}+3,\varepsilon}. \tag{4-32}$$

By the second inequality in (4-31), one gets $R_{n,\varepsilon} \geq R_{\hat{n},\varepsilon} \geq d_0/K_0 > \varepsilon\rho_0(\nu_0)$ for $n \geq \hat{n}$ and $\varepsilon < \varepsilon_1$, and this also yields

$$\Omega_{n,\varepsilon} \cap B_{\varepsilon\rho_0(\nu_0)}(x_\varepsilon) = \varnothing. \tag{4-33}$$

Let $\chi_{n,\varepsilon}(x)$ be smooth cut-off functions such that $\chi_{n,\varepsilon}(x) = 0$ in $B_{R_{n,\varepsilon}}(x_\varepsilon)$, $\chi_{n,\varepsilon}(x) = 1$ in $\Omega_{n+1,\varepsilon}$, $0 \leq \chi_{n,\varepsilon} \leq 1$ and $|\nabla\chi_{n,\varepsilon}| \leq 2/(R_{n+1,\varepsilon} - R_{n,\varepsilon})$.

Lemma 4.8. *Under assumptions* (H_1) *and* (H_2), *if* $\varepsilon < \varepsilon_1$ *and* $\hat{n} \leq n \leq \tilde{n}$, *we have*

$$\int_{\mathbb{R}^N} A_{n,\varepsilon} dx \leq \frac{1}{2} \int_{\Omega_{n,\varepsilon}} \left(\varepsilon^2|\nabla u_\varepsilon|^2 + V(x)u_\varepsilon^2\right) dx, \tag{4-34}$$

where $A_{n,\varepsilon}(x) = \varepsilon^2|\nabla(\chi_{n,\varepsilon}u_\varepsilon)|^2 + V(x)(\chi_{n,\varepsilon}u_\varepsilon)^2$.

Proof. For $\varepsilon < \varepsilon_1$, it follows from a straightforward computation that

$$R_{n+1,\varepsilon} - R_{n,\varepsilon} \geq \frac{c\varepsilon R_{n+1,\varepsilon}}{2}.$$

This yields

$$\varepsilon^2|\nabla\chi_{n,\varepsilon}|^2 \leq \frac{4\varepsilon^2}{|R_{n+1,\varepsilon} - R_{n,\varepsilon}|^2} \leq \frac{16}{c^2 R_{n+1,\varepsilon}^2}. \tag{4-35}$$

From the choice of c, for $\varepsilon < \varepsilon_1$ and $\hat{n} \leq n \leq \tilde{n}$, we arrive at

$$\frac{128}{c^2 R_{n+1,\varepsilon}^2} \leq V(x) \quad \text{for } x \in \{x : R_{n,\varepsilon} \leq |x - x_\varepsilon| < R_{n+1,\varepsilon}\}. \tag{4-36}$$

Noting that $\nabla\chi_{n,\varepsilon}$ is supported in $\{x : R_{n,\varepsilon} \leq |x - x_\varepsilon| < R_{n+1,\varepsilon}\}$, then for $\varepsilon < \varepsilon_1$ and $\hat{n} \leq n \leq \tilde{n}$, by (4-35) and (4-36), we obtain

$$\varepsilon^2|\nabla\chi_{n,\varepsilon}|^2 \leq \tfrac{1}{8} V(x) \quad \text{in } \mathbb{R}^N. \tag{4-37}$$

Multiplying (2-1) by $\chi_{n,\varepsilon}^2 u_\varepsilon$ and integrating over $\mathbb{R}^N$ yields

$$\int_{\mathbb{R}^N} A_{n,\varepsilon} dx = I + II + III,$$

where

$$\begin{aligned}
I &= \int_{\Omega_{n,\varepsilon}} \varepsilon^2|\nabla\chi_{n,\varepsilon}|^2 u_\varepsilon^2\, dx,\\
II &= \int_{\Lambda\cap\Omega_{n,\varepsilon}} K(x) f(u_\varepsilon)\chi_{n,\varepsilon}^2 u_\varepsilon\\
&\leq \frac{1}{16}\int_{\Lambda\cap\Omega_{n,\varepsilon}} V(x)u_\varepsilon^2 dx + C_1 \int_{\Lambda\cap\Omega_{n,\varepsilon}} K(x)|u_\varepsilon|^{p+1} dx,\\
III &= \int_{(\mathbb{R}^N\setminus\Lambda)\cap\Omega_{n,\varepsilon}} g_\varepsilon(x, u_\varepsilon)\chi_{n,\varepsilon}^2 u_\varepsilon\, dx.
\end{aligned}$$

By (4-37), we have

$$|I| \le \frac{1}{8}\int_{\Omega_{n,\varepsilon}} V(x)u_\varepsilon^2\,dx. \tag{4-38}$$

Next we treat $|II|$.

Clearly, we only need to consider the case $\Lambda\cap\Omega_{n,\varepsilon}\neq\varnothing$. In this situation, there is a set $\Sigma_{n,\varepsilon}$ such that $\Sigma_{n,\varepsilon}\cap\Omega_{n,\varepsilon}$ has the uniform cone property and $\Lambda\subset\Sigma_{n,\varepsilon}\subset\Lambda_{r_0}=\{x:\operatorname{dist}(x,\Lambda)\le r_0\}$, where $r_0>0$ is a small constant such that $V(x)\ge V_1$ holds true for $x\in\Lambda_{2r_0}$.

By (4-30), one has

$$\int_{\Sigma_{n,\varepsilon}\cap\Omega_{n,\varepsilon}} K(x)|u_\varepsilon|^{p+1}dx \le C_2\varepsilon^{-N(p-1)/2}\biggl(\int_{\Sigma_{n,\varepsilon}\cap\Omega_{n,\varepsilon}}\bigl(\varepsilon^2|\nabla u_\varepsilon|^2+V(x)u_\varepsilon^2\bigr)\,dx\biggr)^{(p+1)/2}. \tag{4-39}$$

In addition, by (4-33), we arrive at $\Sigma_{n,\varepsilon}\cap\Omega_{n,\varepsilon}\subset\mathbb{R}^N\setminus B_{\varepsilon\rho_0(\nu_0)}(x_\varepsilon)$ for $\varepsilon<\varepsilon_1$ and $n\ge\hat{n}$. Thus, it follows from (4-27), (4-39) and the definition of ν_0 that

$$|II| \le \frac{1}{8}\int_{\Omega_{n,\varepsilon}}\bigl(\varepsilon^2|\nabla u_\varepsilon|^2+V(x)u_\varepsilon^2\bigr)\,dx. \tag{4-40}$$

Finally, we estimate $|III|$.

Similar to the proof of (2-3), for $\varepsilon<\varepsilon_1$, we have

$$|III| \le \int_{\Omega_{n,\varepsilon}}\frac{2\varepsilon^3}{1+|x|^{\theta_0}}u_\varepsilon^2\,dx \le \frac{1}{8}\int_{\Omega_{n,\varepsilon}}\bigl(\varepsilon^2|\nabla u_\varepsilon|^2+V(x)u_\varepsilon^2\bigr)\,dx. \tag{4-41}$$

Combining (4-38), (4-40) with (4-41) yields the conclusion of Lemma 4.8. □

From Lemma 4.8, repeating the same argument as in Lemma 3.3 of [Fei and Yin 2010] leads to the following result.

Lemma 4.9. *Under the assumptions of Lemma 4.8, for small $\varepsilon<\varepsilon_1$, one has*

$$\int_{\mathbb{R}^N}|\nabla(\chi_{\tilde{n},\varepsilon}u_\varepsilon)|^2\,dx \le C\varepsilon^{N-2}2^{-(\ln 2)/(c\varepsilon)}. \tag{4-42}$$

Next, we establish an estimate of $u_\varepsilon(x)$ for large $|x|$.

Lemma 4.10. *Under the assumptions of Lemma 4.8, for $x\in\mathbb{R}^N$ satisfying $|x-x_\varepsilon|\ge d_0/2$, where the meaning of x_ε is given in Lemma 4.2, we have*

$$u_\varepsilon(x)\le C2^{-(\ln 2)/(2c\varepsilon)}. \tag{4-43}$$

Proof. First we assert that

$$\max_{\overline{\Lambda}} u_\varepsilon \le C, \tag{4-44}$$

where $C > 0$ is independent of small ε.

In fact, for any fixed p with $1 < p < \frac{N+2}{N-2}$, it follows from (2-1) that $v_\varepsilon(x) = u_\varepsilon(\varepsilon x)$ satisfies

$$-\triangle v_\varepsilon + V(\varepsilon x) v_\varepsilon = K(\varepsilon x) f(v_\varepsilon) \le \tfrac{1}{16} V(\varepsilon x) v_\varepsilon + C(p) v_\varepsilon^p \quad \text{in } B_{d_0}(\varepsilon^{-1} x_\varepsilon)), \tag{4-45}$$

where $C(p)$ is a positive constant dependent of p.

Define $a_\varepsilon(x) = \frac{15}{16} V(\varepsilon x) - C(p) v_\varepsilon^{p-1}$; then $v_\varepsilon(x)$ is a weak subsolution of the equation

$$-\triangle v_\varepsilon + a_\varepsilon(x) v_\varepsilon = 0 \quad \text{in } B_{d_0}(\varepsilon^{-1} x_\varepsilon)). \tag{4-46}$$

By (4-3), then we obtain, for $\dfrac{N}{2} < q = \dfrac{2N}{(p-1)(N-2)}$ and small ε,

$$\left(\int_{B_{d_0}(\varepsilon^{-1} x_\varepsilon)} |a_\varepsilon|^q \, dx \right)^{1/q} \le C + C(\varepsilon^{-N/2} \|u_\varepsilon\|_\varepsilon)^{2N/(q(N-2))} \le C.$$

This, together with the weak Harnack inequality (see [Gilbarg and Trudinger 1983, p. 193]), yields that there is a positive constant C depending only on the space dimension N and the $L^q(B_{d_0}(\varepsilon^{-1} x_\varepsilon))$ norm of $a_\varepsilon(x)$ such that

$$\begin{aligned} \max_{\overline{\Lambda}} u_\varepsilon = u_\varepsilon(x_\varepsilon) = v_\varepsilon(\varepsilon^{-1} x_\varepsilon) &\le C \left(\int_{B_{d_0}(\varepsilon^{-1} x_\varepsilon)} v_\varepsilon^2 \, dx \right)^{1/2} \\ &= C \left(\varepsilon^{-N} \int_{B_{\varepsilon d_0}(x_\varepsilon)} u_\varepsilon^2 \, dx \right)^{1/2} \le C \varepsilon^{-N/2} \|u_\varepsilon\|_\varepsilon \le C, \end{aligned}$$

namely, (4-44) is proved.

In addition, as in (4-45)–(4-46), one knows that $v_\varepsilon(x) = u_\varepsilon(\varepsilon x)$ is also a weak subsolution of the equation

$$-\triangle v_\varepsilon + b_\varepsilon(x) v_\varepsilon = 0, \tag{4-47}$$

where $b_\varepsilon(x) = \frac{15}{16} V(\varepsilon x) - C(p) \chi_\varepsilon(x) v_\varepsilon^{p-1} - (1 - \chi_\varepsilon(x)) \varepsilon^3 / (1 + |\varepsilon x|^{\theta_0})$, and χ_ε is a characteristic function of $\Lambda^\varepsilon = \{\varepsilon^{-1} x : x \in \Lambda\}$. Moreover, $b_\varepsilon(x)$ has a uniform L^∞ bound independent of small ε by (4-44).

On the other hand, it is noted that for $x \in \mathbb{R}^N$ with $x \in \mathbb{R}^N \setminus B_{d_0/2}(x_\varepsilon)$, then $B_{\varepsilon c d_0}(x) \subset \Omega_{\tilde{n}+1,\varepsilon}$ holds true for small ε and a direct computation yields, for $2^* = 2N/(N-2)$,

$$\begin{aligned} \left(\int_{B_{c d_0}(\varepsilon^{-1} x)} |v_\varepsilon|^{2^*} dy \right)^{1/2^*} &\le C \varepsilon^{-(N-2)/2} \left(\int_{\mathbb{R}^N} |\nabla(\chi_{\tilde{n},\varepsilon} u_\varepsilon)|^2(z) \, dz \right)^{1/2} \\ &\le C 2^{-(\ln 2)/(2c\varepsilon)}. \end{aligned} \tag{4-48}$$

Subsequently, with the aid of Harnack inequality [Gilbarg and Trudinger 1983, Theorem 8.17] and (4-48), we arrive at

$$u_\varepsilon(x) = v_\varepsilon(\varepsilon^{-1}x) \le C\biggl(\int_{B_{cd_0}(\varepsilon^{-1}x)} |v_\varepsilon|^{2^*}\,dy\biggr)^{1/2^*} \le C2^{-(\ln 2)/(2c\varepsilon)}, \tag{4-49}$$

where $C > 0$ depends only on d_0, N and the uniform L^∞ bound of $b_\varepsilon(x)$.

Since the L^∞ norm of $b_\varepsilon(x)$ is uniformly bounded, the proof of Lemma 4.10 is complete. □

Remark 4.1. By Lemma 4.10, for $\theta \ge 1$, there exists an ε_0 such that for $\varepsilon < \varepsilon_0$,

$$|u_\varepsilon(x)| \le \varepsilon^\theta \quad \text{for } x \in \mathbb{R}^N \setminus B_{d_0/2}(x_\varepsilon). \tag{4-50}$$

Next, we show that the local maximum point x_ε of $u_\varepsilon(x)$ in the domain $\bar{\Lambda}$ is also a maximum point of $u_\varepsilon(x)$ in the whole space.

Lemma 4.11. *Under the assumptions of Lemma 4.8, x_ε is the maximum point of u_ε in $\mathbb{R}^N$.*

Proof. Let y_ε be the maximum point of u_ε in $\mathbb{R}^N$; then $u_\varepsilon(y_\varepsilon) = \max_{\mathbb{R}^N} u_\varepsilon \ge \max_{\bar{\Lambda}} u_\varepsilon \ge C$. According to (4-50), we have $y_\varepsilon \subset B_{d_0/2}(x_\varepsilon) \subset \Lambda$ for small ε. Hence $y_\varepsilon = x_\varepsilon$ for small ε by Lemma 4.5. Namely, the proof of Lemma 4.11 is completed. □

Proof of Theorem 1.1. It follows from the assumption (H_5) that there exist positive constants $\sigma_0, \theta_0, \theta_1$ and θ_2 such that

$$\beta < (\alpha - \theta_1)\sigma_0 - \theta_0 \quad \text{and} \quad 4 + 2(\alpha - \theta_1) \le (\theta_1 - 1)\theta_2, \tag{4-51}$$

where $N - \frac{9}{4} < \sigma_0 < N - 2$, $\theta_0 > 2$, $\theta_1 > 1$.

We define the comparison function

$$U(x) = \frac{1}{|x - x_\varepsilon|^{\sigma_0}} \quad \text{for } x \in \mathbb{R}^N \setminus B_{d_0/2}(x_\varepsilon).$$

It is easy to know that $Z(x) = U(x) - \varepsilon^2 u_\varepsilon(x) \ge 0$ on $\partial(B_{d_0/2}(x_\varepsilon))$ for small ε. Recalling that $v_\varepsilon(x) = u_\varepsilon(\varepsilon x)$ vanishes at infinity, this is also true for $Z(x)$.

On the other hand, using the expression for $h_\varepsilon(x, u_\varepsilon)$ and noting that $\sigma_0 < N-2$, we conclude from (4-50) that $\Delta Z = \Delta U - \varepsilon^2 \Delta u_\varepsilon \le 0$ holds for $x \in \mathbb{R}^N \setminus B_{d_0/2}(x_\varepsilon)$ and sufficiently small ε.

Thus, by the maximum principle, we deduce $u_\varepsilon \le U/\varepsilon^2$ in $x \in \mathbb{R}^N \setminus B_{d_0/2}(x_\varepsilon)$. This and the uniform boundedness of x_ε imply

$$u_\varepsilon(x) \le \frac{C}{\varepsilon^2(1 + |x|^{\sigma_0})} \quad \text{in } \mathbb{R}^N \setminus \Lambda. \tag{4-52}$$

Next we verify that u_ε actually solves Equation (1-1). Indeed, since $f(s) = O(s^\alpha)$ near $s = 0$, together with (4-50) we have, for small ε,

$$f(u_\varepsilon) \le C|u_\varepsilon|^\alpha \quad \text{in } \mathbb{R}^N \setminus \Lambda. \tag{4-53}$$

Combining (4-50)–(4-53), we have, for small ε,

$$K(x) f(u_\varepsilon) \le Ck(1 + |x|^\beta)|u_\varepsilon|^\alpha \le \frac{\varepsilon^3}{1 + |x|^{\theta_0}}|u_\varepsilon| \quad \text{in } \mathbb{R}^N \setminus \Lambda. \tag{4-54}$$

Choose two positive numbers θ_3 and θ_4 such that

$$\beta < (\alpha - \theta_3)\sigma_0 - N \quad \text{and} \quad 2 + 2(\alpha - \theta_3) \le \theta_3\theta_4. \tag{4-55}$$

Collecting (4-50), (4-52), (4-53), and (4-55) yields for small ε,

$$K(x) f(u_\varepsilon) \le Ck(1 + |x|^\beta)|u_\varepsilon|^{\alpha-\theta_3}|u_\varepsilon|^{\theta_3} \le \frac{\varepsilon}{1 + |x|^N} \quad \text{in } \mathbb{R}^N \setminus \Lambda. \tag{4-56}$$

Therefore, it follows from (4-54) and (4-56) that $g_\varepsilon(x, u_\varepsilon) \equiv K(x) f(u_\varepsilon)$ holds true in $\mathbb{R}^N \setminus \Lambda$ and subsequently u_ε solves the original equation (1-1). In addition, noting that $N - \frac{9}{4} < \sigma_0$, then the estimate (4-52) leads to $u_\varepsilon \in L^2(\mathbb{R}^N)$ for $N \ge 5$.

Finally, combining the conclusions in Lemma 4.2, Lemma 4.5 and Lemma 4.11, in order to finish the proof of Theorem 1.1, we only need to verify (1-12). Set $M = \{x_0\}$, due to (4-26), one has $x_\varepsilon \to x_0$ as $\varepsilon \to 0$. Let $v_\varepsilon(x) = u_\varepsilon(\varepsilon x + x_\varepsilon)$, then v_ε is uniformly bounded in $H^1_{\text{loc}}(\mathbb{R}^N)$ and satisfies the equation

$$-\triangle v_\varepsilon + V(\varepsilon x + x_\varepsilon)v_\varepsilon = K(\varepsilon x + x_\varepsilon) f(v_\varepsilon), \quad x \in \mathbb{R}^N. \tag{4-57}$$

As in the arguments of Lemma 4.3 or Lemma 4.5, we can show that v_ε converges to $v \in C^2(\mathbb{R}^N) \cap H^1(\mathbb{R}^N)$ in $C^2_{\text{loc}}(\mathbb{R}^N)$ as $\varepsilon \to 0$. With the aid of (4-43), v_ε converges to v in $L^\infty(\mathbb{R}^N)$ as $\varepsilon \to 0$. Therefore v is a solution of the equation

$$-\triangle v + V(x_0)v = K(x_0) f(v), \quad x \in \mathbb{R}^N; \tag{4-58}$$

moreover, by virtue of strong maximum principle, $v > 0$ can be derived. On the other hand, as a consequence of Theorem 2 [Gidas et al. 1981] and the subsequent remark, v is radially symmetric and decays exponentially.

Thus the proof of Theorem 1.1 is completed. □

References

[Ambrosetti and Malchiodi 2007] A. Ambrosetti and A. Malchiodi, "Concentration phenomena for nonlinear Schrödinger equations: recent results and new perspectives", pp. 19–30 in *Perspectives in nonlinear partial differential equations*, edited by H. Berestycki et al., Contemp. Math. **446**, Amer. Math. Soc., Providence, RI, 2007. MR 2008j:35037 Zbl 1200.35106

[Ambrosetti and Wang 2005] A. Ambrosetti and Z.-Q. Wang, “Nonlinear Schrödinger equations with vanishing and decaying potentials”, *Differential Integral Equations* **18**:12 (2005), 1321–1332. MR 2006k:35071 Zbl 1210.35087

[Ambrosetti et al. 2005] A. Ambrosetti, V. Felli, and A. Malchiodi, “Ground states of nonlinear Schrödinger equations with potentials vanishing at infinity”, *J. Eur. Math. Soc.* **7**:1 (2005), 117–144. MR 2006f:35049 Zbl 1064.35175

[Berestycki and Lions 1983] H. Berestycki and P.-L. Lions, “Nonlinear scalar field equations, I: Existence of a ground state”, *Arch. Rational Mech. Anal.* **82**:4 (1983), 313–345. MR 84h:35054a Zbl 0533.35029

[Bonheure and Van Schaftingen 2008] D. Bonheure and J. Van Schaftingen, “Bound state solutions for a class of nonlinear Schrödinger equations”, *Rev. Mat. Iberoam.* **24**:1 (2008), 297–351. MR 2009d:35069 Zbl 1156.35084

[Byeon and Wang 2006] J. Byeon and Z.-Q. Wang, “Spherical semiclassical states of a critical frequency for Schrödinger equations with decaying potentials”, *J. Eur. Math. Soc.* **8**:2 (2006), 217–228. MR 2007e:35073 Zbl 1245.35036

[Costa and Tehrani 2001] D. G. Costa and H. Tehrani, “On a class of asymptotically linear elliptic problems in $\mathbb{R}^N$”, *J. Differential Equations* **173**:2 (2001), 470–494. MR 2002g:35065 Zbl 1098.35526

[Dávila et al. 2007] J. Dávila, M. del Pino, M. Musso, and J. Wei, “Standing waves for supercritical nonlinear Schrödinger equations”, *J. Differential Equations* **236**:1 (2007), 164–198. MR 2009b:35389 Zbl 1124.35082

[Ekeland 1990] I. Ekeland, *Convexity methods in Hamiltonian mechanics*, Ergebnisse der Mathematik und ihrer Grenzgebiete, III **19**, Springer, Berlin, 1990. MR 91f:58027 Zbl 0707.70003

[Fei and Yin 2010] M. Fei and H. Yin, “Existence and concentration of bound states of nonlinear Schrödinger equations with compactly supported and competing potentials”, *Pacific J. Math.* **244**:2 (2010), 261–296. MR 2011j:35082 Zbl 1189.35304

[Gidas et al. 1981] B. Gidas, W. M. Ni, and L. Nirenberg, “Symmetry of positive solutions of nonlinear elliptic equations in $\mathbf{R}^n$”, pp. 369–402 in *Mathematical analysis and applications, Part A*, edited by L. Nachbin, Adv. in Math. Suppl. Stud. **7**, Academic, New York, 1981. MR 84a:35083 Zbl 0469.35052

[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

[Gui 1996] C. Gui, “Existence of multi-bump solutions for nonlinear Schrödinger equations via variational method”, *Comm. Partial Differential Equations* **21**:5-6 (1996), 787–820. MR 98a:35122 Zbl 0857.35116

[Jeanjean and Tanaka 2002] L. Jeanjean and K. Tanaka, “A positive solution for an asymptotically linear elliptic problem on $\mathbb{R}^N$ autonomous at infinity”, *ESAIM Control Optim. Calc. Var.* **7** (2002), 597–614. MR 2003f:35113 Zbl 1225.35088

[Liu and Wang 2004] Z. Liu and Z.-Q. Wang, “Existence of a positive solution of an elliptic equation on $\mathbb{R}^N$”, *Proc. Roy. Soc. Edinburgh Sect. A* **134**:1 (2004), 191–200. MR 2005c:35101 Zbl 1067.35029

[Liu et al. 2006] Z. Liu, J. Su, and T. Weth, “Compactness results for Schrödinger equations with asymptotically linear terms”, *J. Differential Equations* **231**:2 (2006), 501–512. MR 2009d:35095 Zbl 05115328

[Liu et al. 2008] C. Liu, Z. Wang, and H.-S. Zhou, "Asymptotically linear Schrödinger equation with potential vanishing at infinity", *J. Differential Equations* **245**:1 (2008), 201–222. MR 2009h:35144 Zbl 1188.35181

[del Pino and Felmer 1996] M. del Pino and P. L. Felmer, "Local mountain passes for semilinear elliptic problems in unbounded domains", *Calc. Var. Partial Differential Equations* **4**:2 (1996), 121–137. MR 97c:35057 Zbl 0844.35032

[del Pino et al. 2007] M. del Pino, M. Kowalczyk, and J.-C. Wei, "Concentration on curves for nonlinear Schrödinger equations", *Comm. Pure Appl. Math.* **60**:1 (2007), 113–146. MR 2007h:35113 Zbl 1123.35003

[Rabinowitz 1992] P. H. Rabinowitz, "On a class of nonlinear Schrödinger equations", *Z. Angew. Math. Phys.* **43**:2 (1992), 270–291. MR 93h:35194 Zbl 0763.35087

[Stuart and Zhou 1999] C. A. Stuart and H. S. Zhou, "Applying the mountain pass theorem to an asymptotically linear elliptic equation on $\mathbf{R}^N$", *Comm. Partial Differential Equations* **24**:9-10 (1999), 1731–1758. MR 2000g:35052 Zbl 0935.35043

[Wang and Zeng 1997] X. Wang and B. Zeng, "On concentration of positive bound states of nonlinear Schrödinger equations with competing potential functions", *SIAM J. Math. Anal.* **28**:3 (1997), 633–655. MR 98e:81032 Zbl 0879.35053

[Willem 1996] M. Willem, *Minimax theorems*, Progress in Nonlinear Differential Equations and their Applications **24**, Birkhäuser, Boston, MA, 1996. MR 97h:58037 Zbl 0856.49001

[Yin and Zhang 2009] H. Yin and P. Zhang, "Bound states of nonlinear Schrödinger equations with potentials tending to zero at infinity", *J. Differential Equations* **247**:2 (2009), 618–647. MR 2010m:35169 Zbl 1178.35353

Received August 25, 2011. Revised August 17, 2012.

MINGWEN FEI
DEPARTMENT OF MATHEMATICS & IMS
NANJING UNIVERSITY
210093 NANJING
CHINA

and

ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
100190 BEIJING
CHINA

mwfei@amss.ac.cn

HUICHENG YIN
DEPARTMENT OF MATHEMATICS & IMS
NANJING UNIVERSITY
210093 NANJING
CHINA

huicheng@nju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.369

TYPE I ALMOST HOMOGENEOUS MANIFOLDS OF COHOMOGENEITY ONE, III

DANIEL GUAN

This paper is one of a series in which we generalize our earlier results on the equivalence of existence of Calabi extremal metrics to the geodesic stability for any type I compact complex almost homogeneous manifolds of cohomogeneity one. In this paper, we actually carry all the earlier results to the type I cases. As requested by earlier referees of this series of papers, in this third part, we shall first give an updated description of the geodesic principles and the classification of compact almost homogeneous Kähler manifolds of cohomogeneity one. Then, we shall give a proof of the equivalence of the geodesic stability and the negativity of the integral in the first part. Finally, we shall address the relation of our result to Ross–Thomas version of Donaldson's K-stability. One should easily see that their result is a partial generalization of our integral condition in the first part. And we shall give some further comments on the Fano manifolds with the Ricci classes. In Theorem 14, we give a result of Nadel type. We define the strict slope stability. In our case, it is stronger than Ross–Thomas slope stability. We strengthen two Ross–Thomas results in Theorems 15 and 16. The similar proofs of the results other than the existence for the type II cases are more complicated and will be done elsewhere.

1. Introduction

This paper is one of a series of papers in which we finished the project of studying the existence (or not) of extremal metrics in any Kähler class on any compact almost homogeneous manifolds of cohomogeneity one.

In [Guan 2011a; 2011b] we proved that for the type I compact almost homogeneous Kähler manifolds of cohomogeneity one, the existence of Calabi extremal metrics is the same as the negativity of a topological integral. We also proved in [Guan 2011b] that for any two Kähler metrics in the Mabuchi moduli space of Kähler metrics there is a smooth geodesic connecting them. That is, the geodesic principle I is true for these manifolds.

Supported by DMS-0103282.
MSC2010: 14J45, 34B18, 53C10, 53C25, 53C55.
Keywords: Kähler–Einstein, stability, cohomogeneity one.

As in [Guan 2003], the major tool is from [Guan 1999]. Although the problem of existence of the extremal metrics can be reduced to an ordinary differential equation for our manifolds, the problem of the existence of the geodesics has two variables. Thanks to the Legendre transformation, we can carry it out for the type I manifolds. But for a general type II manifolds, this method does not work any more. And we need a new method, which will be carried out in [Guan $\geq$ 2013a].

Even for the Kähler–Einstein equation, our method in [Guan 2011a] is different from [Guan and Chen 2000]. We used a semisimple method in [Guan 2011a]. One notices that our exponential map there is not the one for the geodesics. No geodesic in that situation could have infinite length. It was well known for many years that there were many nonsmooth solutions for even a real homogeneous Monge–Ampère equations. In [Chen and Tian 2008] Professor Chen gave an example which looks like a nonsmooth solution for the one-dimensional toric case, that is, $\mathbb{C}P^1$. He also mentioned it earlier to me in 1999 at Princeton. Mabuchi also mentioned it to me in Pisa, Italy in 2004. However, we already solved the smoothness question for the toric manifolds in [Guan 1999]. In this simple case, the method of X. X. Chen should also produce the smooth solution; see [Guan and Phong 2012]. The content of this note was presented in the AMS meeting in Pomona California May 2008. Recently, L. Lempert and L. Vivas claimed (also mentioned by the referee) that they found a counterexample to our geodesic principle I on the torus. However, their examples are not very explicit and not published yet. We are not able to check their examples in this paper. As we know, there is no much equivariant geometry on the torus. The geodesic problem was trivial on the torus. However, see also [Feng 2012]. We checked that all the geodesic principles hold on compact cohomogeneity-one Kähler manifolds. We conjecture that the geodesic principles hold for all the spherical manifolds. We take them as working principles in our research. *For our safety, we just require that everything is analytic.* For example, for any analytic initial value in the tangent space of the equivariant Mabuchi moduli space at a given metric, there is a geodesic ray. That is, *the geodesic principle I is not really needed for the geodesic stability.* In [Guan and Chen 2000], some possible obstructions emerged that I eventually treated in [Guan 2002], which led to the strict slope stability. After a long run, we are able to overcome all the difficulties. To solve the extremal metrics cases, we have to deal with a fourth-order ordinary differential equation, which in our cases is fortunately reduced to a second-order nonlinear equation and is successfully treated.

All the solutions we find in the cohomogeneity-one cases are not explicit except those in [Guan 1995a; Guan 2007].

In this paper, we shall prove that the negativity of the integral is actually the same as the geodesic stability.

A classification which we refer to in this paper can be found in [Guan 2003, Section 12].

Here we shall describe our updated *geodesic stability principles*. We conclude these principles by following the cumulation of other people's observations and the evidence from our examples. See [Guan 2003]. We do not assume that these principles are due to us completely, in particular the first principle.

Motivated by the Donaldson's functional in the vector bundle case, Mabuchi [1986] defined a functional on the Mabuchi moduli space of the Kähler metrics (see also a conjecture therein). It was later modified independently by several people to fit the situation of Calabi extremal metrics (see [Guan 1999; Guan and Chen 2000], etc.) on the equivariant Mabuchi moduli space of Kähler metrics, which we call the modified Mabuchi functional.

Principle I. *For any two Kähler metrics in a given Kähler class, there is a unique (smooth) geodesic in the Mabuchi moduli space of Kähler metrics connecting them.*

This principle has been tested for toric bundles in [Guan 1999]. We also found that the same method applies to Kähler metrics on type-I and type-III compact almost homogeneous Kähler manifolds of cohomogeneity one in [Guan 2003; 2011b]; see also [Guan 2007]. It seems to us that there is not any complete geodesic except the ones induced by the holomorphic vector fields. X. X. Chen [2000] proved the existence of an unique $C^{1,1}$ solution in general.

We shall concentrate on the maximal geodesic rays. It turns out that the majority of the maximal geodesic rays are of finite length (this is different from holomorphic vector bundle theory on vector bundles; cf. [Kobayashi 1987, p. 197] and also the picture shown in [Semmes 1992, p. 544]). The maximal geodesic rays with infinite length are very special with some *strong convex* property, which we call "effective" maximal geodesic rays. The direction of the effective geodesic rays at each metric might form a *convex cone* $\mathscr{C}$.

Principle II. *The limit metrics of the maximal geodesics are concentrations*:

A. *Finite ray*: *cone concentration — partial concentration.*

B. *Infinite ray*: *blow up caused by some subvarieties outside a compact set — complete concentration outside the compact set, the metric on this compact set does not change.*

We call the limit of the ratio of the modified Mabuchi functional the ***generalized Futaki invariants*** *of the maximal geodesic rays. The generalized Futaki invariant is positive infinite for finite rays, that is, the only interesting generalized Futaki invariants come from the effective maximal geodesic rays.*

The second principle is based on our work on toric manifolds and cohomogeneity-one manifolds; see [Guan 2003; 2007] for examples.

For all the examples we consider in this paper, the Mabuchi equivariant moduli space is flat (see [Guan 1999]); this is similar to the vector bundle case and is not true in general (see [Mabuchi 1987]). For two maximal geodesic rays, the generalized Futaki invariants might be the same if there is a curve connecting the beginning points such that there is a parallel vector field along this curve which connects the two tangent vectors at these two points. This observation makes the definition of the generalized Futaki invariants independent of the initial Kähler metrics.

The generalized Futaki invariants define a function of the effective geodesic cone which is probably a linear function $F_{M,\omega}$, which is continuous on a certain given Banach space. Therefore, F can be defined on the closure $\bar{\mathscr{C}}$ of the effective cone $\mathscr{C}$ in the Banach space. We call $F|_{\bar{\mathscr{C}}}$ the *generalized Futaki invariant functional* or simply the *generalized Futaki invariant.* There is a seminorm $\|\cdot\|_*$, which is locally equivalent to the given norm except on some subvarieties and is zero on the functions induced by the holomorphic vector fields.

Principle III. *There is a unique extremal metric in a given Kähler class up to the automorphism group if and only if the Kähler class is geodesic stable*, *that is*, *with positive generalized Futaki invariant which is bounded below by the given seminorm.*

(Note: in many of our papers, this is called the fourth principle and the next principle is called the third, reflecting the order in which they were formulated.)

In general, the Mabuchi moduli space might not be flat. We might have some way to relate the Futaki invariants for two infinite maximal geodesic rays starting from different points. Let $\gamma_i(t)$, $i = 1, 2$ be two maximal geodesic rays. We say that they have the same infinite points if

$$d(\gamma_1, \gamma_2) = \sup_{t\in[0,+\infty)} d(\gamma_1(t), \gamma_2(t))$$

is finite. Then we have (see also [Guan 2007, Remark 4]):

Principle IV. *The Futaki invariants of two maximal geodesic rays with the same infinite point are the same.*

In the last section, we shall see that our stability in this case is the same as a version of the slope stability which is stronger than that in [Ross and Thomas 2006].

2. Preliminaries

Here we summarize some known results about the compact complex almost homogeneous manifolds of cohomogeneity one. In this paper, we only consider manifolds with a Kähler structure. For earlier results one might check [Ahiezer 1983; Huckleberry and Snow 1982].

We call a compact complex manifold an almost homogeneous manifold if its complex automorphism group has an open orbit. We say that a manifold is of cohomogeneity one if the maximal compact subgroup has a (real) hypersurface orbit. In [Guan and Chen 2000; Guan 2003], we reduced the compact complex almost homogeneous manifolds of cohomogeneity one into three types of manifolds.

We denote the manifold by M and let G be a complex subgroup of its automorphism group which has an open orbit on M.

Let us assume first that M is simply connected. Let the open orbit be G/H, K be the maximal connected compact subgroup of G, L be the generic isotropic subgroup of K, that is, K/L be a generic K-orbit. We have [Guan and Chen 2000, Theorem 1]:

Proposition 1. *If G is not semisimple, then M is a completion of a $\mathbb{C}^*$-bundle over a projective rational homogeneous space.*

If a compact almost homogeneous Kähler manifold is a completion of a $\mathbb{C}^*$-bundle over a product of a torus and a projective rational homogeneous space, we call it a *manifold of type III.* We dealt with this kind of manifold in our dissertation [Guan 1995a; 1995b]. There always exists an extremal metric in any Kähler class. In [Guan 2007], we generalized this existence result to a family of metrics connecting the extremal metric of [Guan 1995a] and the generalized quasi-Einstein metric of [Guan 1995b]; we called this family the *extremal-soliton metrics*. The existence of the extremal-soliton is the same as geodesic stability with respect to a generalized Mabuchi functional.

More recently in [Guan 2012], we even generalized the extremal-solitons to the generalized extremal solitons, which also include Nakagawa's [2011] generalized Kähler–Ricci solitons as a special case. We proved the existence of both generalized extremal solitons and the generalized Kähler–Ricci solitons on these manifolds. In a forthcoming paper [Guan $\geq$ 2013b], we proved the existence of the so called m-extremal metrics on these manifolds.

In general, if M is a compact almost homogeneous Kähler manifold and O is the open orbit, then $D = M - O$ is a proper closed submanifold. Moreover, D has at most two components. We call each component of D an end. If D has two components or one component, we say M is an almost homogeneous manifold with two ends or one end, respectively. We have [Huckleberry and Snow 1982, Theorem 3.2]:

Proposition 2. *If M is a compact almost homogeneous Kähler manifold with two ends, then M is a manifold of type III.*

Therefore, we only need to deal with the case with one end. In [Guan and Chen 2000], we treated the first example, that is, the blowup of the diagonal of the product of two copies of $\mathbb{C}P^n$. We treated another series in [Guan 2003]. We treated many more of them in [Guan 2009; 2011b; 2011c], etc. Again, in the case of M being simply connected, we only need to take care of the case in which G is semisimple.

If G is semisimple and M has two G orbits, one open and one closed, and moreover if the closed orbit is a complex hypersurface, there are two possibilities. Let $\mathscr{K}$, $\mathscr{L}$ be the Lie algebras of K, L. Then the centralizer of $\mathscr{L}$ in $\mathscr{K}$ is a direct sum of the center of $\mathscr{L}$ and a Lie subalgebra $\mathscr{A}$ with $\mathscr{A}$ being either one-dimensional or a 3-dimensional Lie algebra su(2). If $\mathscr{A}$ is one-dimensional, we call M a *manifold of type I.* If $\mathscr{A}$ is su(2), we call M a *manifold of type II.*

In general, if the closed orbit has a higher codimension, we can always blow up the closed orbit to obtain a manifold $\tilde{M}$ with a hypersurface end. We call the manifold M a *manifold of type I or II* if $\tilde{M}$ is of type I or II, respectively.

There is a special case of the type II manifolds. If the open orbit is a $\mathbb{C}^k$-bundle over a projective rational homogeneous manifold, we call M an *affine type manifold* (not to be confused with the closed complex submanifolds of $\mathbb{C}^m$).

Then we have (see [Guan 2003, Section 12]):

Proposition 3. *Any compact almost homogeneous Kähler manifold M of cohomogeneity one is an* $\mathrm{Aut}_0(M)$ *equivariant fibration over a product of a rational projective homogeneous manifold Q and a complex torus T with a fiber F. Therefore, M can be regarded as a fiber bundle over T with a simply connected fiber M_1. One of following holds*:

(i) *M is a manifold of type III.*

(ii) *M_1 is of type II but not affine.*

(iii) *M_1 is affine.*

(iv) *M_1 is of type I.*

We say that M is a *manifold of type I, or type II, affine,* if M_1 is, respectively, a manifold of type I or type II, affine.

We actually can also obtain the structure of an M_1-bundle over T from [Huckleberry and Snow 1982]. We only need to understand the bundle structure for the open orbit. By [ibid., Corollary 4.4] we have that the bundle structure is a product unless, when we apply Proposition 3 to $\tilde{M}$, $F = Q^k$. In the latter case, there is an unbranched double covering $\bar{M}$ of M such that the bundle structure of $\bar{M}$ is a product.

Proposition 4. *The M_1-bundle over T is a product except in the case where the open orbit is an F_0-bundle over $Q \times T$ such that F_0 is in the second, sixth and eighth cases in* [Ahiezer 1983, p. 67]. *In the latter cases, the M_1-bundle has an unbranched double covering which is a product of M_1 and T.*

In [Guan 2011a; 2011b], we dealt with the type I cases.

One updated remark is that since we are dealing with the Kähler metrics it is more convenient to separate the type II case into two cases in [Guan 2009] and [Guan 2011c]. We call the cases in [Guan 2009] (and the papers between [ibid.] and [Guan 2003]) the *type IV* cases. They are the affine cases such that the group

$\pi(G_F)$, the restriction of the subgroup G_F of G fixing a given fiber F, is not of type A. Therefore, one might also call them the *non-type-A type II* cases. All of them are Fano.

One might call the rest (in [Guan 2011c]) of the type II cases the new type II cases (or simply the type II cases). They are those type II cases such that $\pi(G_F)$ is of type A. Therefore, one might also call them the *type A type II* cases.

This note is a continuation of the first part and the second part of this paper [Guan 2011a; 2011b]. We shall retain all the notation from those papers here.

3. The complex structures of the type I almost homogeneous manifolds

In this section, we shall deal with the complex structure of the type I almost homogeneous manifolds. We retain the notation in [Guan 2011a; 2011b]. Let us recall some basic notation of the Lie algebras.

Let G be the complex Lie group action and S be the connected complex Lie subgroup acting on a given fiber. According to [Guan 2003, p. 283, Theorem 12.1(ii)], a compact complex almost homogeneous manifold of cohomogeneity one is type I if and only if the fiber F is one of (1) the second and third case with $n \geq 3$, (2) the fourth case, (3) the eight and ninth cases, (4) the fifth case in [Ahiezer 1983, p. 67].

The fiber F in (4) has $S = \pi(G_F) = F_4$, so $G = F_4 = S$, that is, $M = F$ is homogeneous. Therefore, every Kähler class of M has a metric with constant scalar curvature. So, we do not need to do anything with (4).

In [Guan 2011a], we look at three special possible fiber cases [Ahiezer 1983, p. 67] first:

(1) $F = F(\mathrm{OP}_n)$: The third case in [Ahiezer 1983, p. 67] with $n \geq 3$. We have $F = \mathbb{C}P^n$ and

$$S = \pi(G_F) = \mathrm{SO}(n, \mathbb{C}),$$

regarding $\mathbb{C}P^n$ as a completion of $\mathbb{C}^n$. The corresponding compact rank-one symmetric space is the real n-dimensional real projective space. It has an equivariant branched double covering Q^n of the second case. We denote the latter case by $F(\mathrm{OQ}_n)$.

(2) $F = F(\mathrm{Gr}_k)$: The fourth case with a standard $S = \mathrm{Sp}(k, \mathbb{C})$-action on the manifold $F = \mathrm{Gr}(2k, 2)$. The corresponding compact rank one symmetric space is the quaternionic projective space.

(3) $F = F(\mathrm{Sp}_7^p)$: The ninth case with an $S = \mathrm{Spin}(7, \mathbb{C})$-action on $F = \mathbb{C}P^7$. This is the restriction of (1) with $n+1 = 8$ to the complex Lie subgroup $\mathrm{Spin}(7, \mathbb{C})$. It has an equivariant branched double covering Q^7 of the eighth case. In [Guan

2011a], we also denote the latter case by $F(\mathrm{Sp}_7^q)$ and denote both of them by $F(\mathrm{Sp}_7)$ whenever there is no confusion.

In [ibid.], we defined a certain basis of the Lie algebra α, F_α and G_α for positive roots α. And, we considered a fixed point p_0 and its orbit p_s generated by a semisimple element $-iH$ in the Lie algebra. Let T be the tangent vector of p_s and p_∞ be the limit point in the closed orbit.

In the case (1), we obtained:

Proposition 5. *For $F(\mathrm{OP}_n)$ and $F(\mathrm{OQ}_n)$, along p_s we have*
$$J(F_{e_1+e_i} \pm F_{e_1-e_i}) = -(\tanh s)^{\mp 1}(G_{e_1+e_i} \pm G_{e_1-e_i})$$
(and $JF_{e_1} = -(\tanh s)G_{e_1}$). We also have that $F_{e_i \pm e_k} = G_{e_i \pm e_k} = 0$ (and $F_{e_i} = G_{e_i} = 0$) for $i > 1$. In particular, at p_∞, $JF_\alpha = -G_\alpha$ for $\alpha \neq e_i \pm e_k$ (and e_i), $1 < i < k$.

In the case of (2), we obtained:

Proposition 6. *For $F(\mathrm{Gr}_k)$, we have*
$$JF_{\alpha_1} = -(\tanh 2s)G_{\alpha_1},$$
$$J(F_{2e_1} \pm F_{2e_2}) = -(\tanh 2s)^{\mp 1}(G_{2e_1} \mp G_{2e_2}),$$
$$J(F_{e_1-e_k} \pm G_{e_2-e_k}) = -(\tanh s)^{\mp 1}(G_{e_1-e_k} \pm F_{e_2-e_k}),$$
$$J(F_{e_1+e_k} \pm G_{e_2+e_k}) = -(\tanh s)^{\mp 1}(G_{e_1+e_k} \pm F_{e_2+e_k}).$$
$F_\alpha = G_\alpha = 0$ for $\alpha = e_1 + e_2, e_i - e_k, 2e_i, e_i + e_k$ with $i > 2$.

At p_∞, we have $F_\alpha = G_\alpha = 0$ if $\alpha = e_1 + e_2, 2e_i, e_i \pm e_k, i > 2$, and $JF_\alpha = G_\alpha$ if $\alpha = 2e_2, e_2 \pm e_k$. Otherwise $JF_\alpha = -G_\alpha$.

Before we consider the isolated case (3), we can look at the general cases in which $G \neq S = \pi(G_F) \subset \mathrm{Aut}(F)$, where G_F is the subgroup that acts on the fiber F and $\pi : G_F \to \mathrm{Aut}(F)$ is the induced map from G_F to $\mathrm{Aut}(F)$. As in [Ahiezer 1983], G is semisimple, $U_G = H$ is the 1-subgroup. There is a parabolic subgroup $P = SS_1R$ with S, S_1 semisimple and R solvable such that $U_G = US_1R$ where $U = H \cap S$ is a 1-subgroup of S. The manifold is a fibration over G/P with the completion of $P/U_G = S/U$ as the isotropic open orbit of the almost homogeneous fiber. In this case, the root system of S is a subsystem of the root system of G. In the Lie algebra of G, we also have some other F_α, G_α outside $\mathcal{S}$. Let K be a maximal connected compact Lie subgroup of G and L be the isotropic subgroup of K at a generic orbit. Let $\mathcal{K}$, $\mathcal{L}$ be the corresponding Lie algebras. The tangent space of G/U_G along p_s is decomposed into irreducible $\mathcal{L}$-representations. These F_α, G_α are in the complement representation of the Lie algebra $\mathcal{S}$ of S. As it is in the tangent space of G/P, $JF_\alpha = -G_\alpha (\mathrm{mod}\, \mathcal{S})$. Therefore, we have $JF_\alpha = -G_\alpha$ for any α which is not in the root system of S.

If S is B_2, G can be B_n, C_n, F_4. If S is B_3, G can be B_n, F_4. If S is C_3, G can be C_n, F_4. If S is B_n with $n > 3$, G can only be B_{m+n}. If S is C_n with $n > 3$, then G can be C_{n+m}. The case of a B_2-action that has an isotropic group of $\mathrm{SO}(4, \mathbb{C})$ generated by roots $\pm e_1 \pm e_2$ is exactly the same as the case of an $\mathrm{Sp}(2, \mathbb{C})$-action, which has an isotropic subgroup of $\mathrm{Sp}(1, \mathbb{C}) \times \mathrm{Sp}(1, \mathbb{C})$ generated by $\pm 2e_1$, $\pm 2e_2$.

We have a few more possibilities. If $S = D_k$, $k > 3$, G can only be D_n, $n > 3$ or E_n, $n > k$. If $S = D_3$, that is an A_3, G can be A_n, $n > 2$, B_n, $n > 3$, C_n $n > 3$, D_n $n > 2$ and E_n. If $S = D_2$, G can be any simple group or product of simple groups other than G_2.

We then treated the isolated case (3) of the $\mathrm{Spin}(7, \mathbb{C})$-action on $\mathbb{C}P^7$ in [Guan 2011a]. This case is the restriction of the case (1) with an $G = S = \mathrm{SO}(8, \mathbb{C})$-action to the $\mathrm{Spin}(7, \mathbb{C})$-action induced by the spinor representation.

We obtained:

Proposition 7. *For $F(\mathrm{Sp}_7)$, we have*

$$J(\sqrt{2}F_{h_i} \pm F_{h_j+h_k}) = -\Big(\tanh \frac{\sqrt{3}}{2}s\Big)^{\mp 1}\big(\sqrt{2}G_{h_i} \pm G_{h_j+h_k}\big),$$
$$JH = -T,$$
$$F_{e_i-e_j} = G_{e_i-e_j} = 0 \quad \textit{for } 0 < i < j < 4.$$

At p_∞, $JF_{h_i} = -G_{h_i}$, $JF_{h_j+h_k} = -G_{h_j+h_k}$, $F_{h_i-h_k} = G_{h_i-h_k} = 0$.

However, in this case $S = B_3$, G can only be B_n or F_4.

4. The Kähler structures

In [Guan 2011a], we examined the Kähler structure for the $S = \mathrm{SO}(n, \mathbb{C})$-actions and obtained that for any possible G and $S = \mathrm{SO}(n, \mathbb{C})$ we always have a Kähler metric: $\omega([X, Y]) = (aH + I, [X, Y])$ with the I in the C center of $\mathscr{L}$ and a a nonpositive function of s.

See [Guan 2011a, Section 3].

Therefore, we have the volume formula

$$V = -Ma'a^{2(n-1)}\prod_1^r(a_i - a)\prod_1^s(b_j + a)$$
$$\Big(\text{or} \quad V = Ma'a^{2n-1}(\tanh s)\prod_1^r(a_i - a)\prod_1^s(b_j + a)\Big),$$

with some positive numbers a_i and b_j.

Then in [Guan 2011a], we dealt with the Kähler metrics with $\mathrm{Sp}(k, \mathbb{C})$ and $\mathrm{Spin}(7, \mathbb{C})$-actions. We have the volume form

$$V = Ma'a^{4k-5}(\tanh 2s)\prod_1^r(a_i - a)\prod_1^s(b_j + a)$$

for the $\mathrm{Sp}(k, \mathbb{C})$-actions.

For the $S = \mathrm{Spin}(7, \mathbb{C})$-action, we obtained the volume form

$$V = -Ma'a^6\prod_{i=1}^r(a_i - a)\prod_{j=1}^s(b_j + a).$$

We also observe that a_i and b_j come in pairs, and $b_{j(i)} = a_i$.

Altogether, we have:

Proposition 8. *For the type I case the volume is*

$$V = -Ma'a^{2m}\prod(a_i^2 - a^2)$$

for the cases $S = D_k$ or $\mathrm{Spin}(7, \mathbb{C})$ *and*

$$V = Ma'a^{2m+1}(\tanh bs)\prod(a_i^2 - a^2)$$

for the cases $S = B_k$ (or C_k) with $b = 1$ (or 2), where M and a_i are positive numbers, m are nonnegative integers. We also have that $2m+1$ (or $2m+2$) are the dimensions of the fiber. Moreover, the vectors in Propositions 5, 6 and 7 are orthogonal to each other.

Let $h = \log V$. In [Guan 2011a, Section 5, Theorem 2] we obtained:

Proposition 9. *If the fiber with the S-action is of type I of complex dimension n, then the function a for the Ricci form ρ is*

$$a_\rho = \tfrac{1}{2}\left(\left(\log\left(a'a^{n-1}\prod_1^r(a_i^2 - a^2)\right)\right)' - 2\sum_1^{n-1} N_i \coth 2N_i s\right).$$

Moreover, the N_i are (1) 1 *for $S = \mathrm{SO}(n+1, \mathbb{C})$ and* (2) 1 *except three of them being* 2 *for S of type C_k,* (3) $\sqrt{3}/2$ *for the case $S = \mathrm{Spin}(7, \mathbb{C})$. Other coefficients come from the Ricci curvature of G/P, which is $-(q_{G/P}, [X, Y])$ with $q_{G/P} = \sum_{\alpha\in\Delta^+-\Delta_P} H_\alpha$ with the standard inner product.*

Then we calculated the scalar curvature in [Guan 2011a, Section 6, Theorem 3]. We write

$$V = -Ma'\tilde{Q}(a) = -Ma'(-a)^{n-1}Q_1(a)g(s),$$

with $g(s) = 1$ for $S = D_k$ or $\mathrm{Spin}(7, \mathbb{C})$ and $g(s) = \tanh bs$ for $S = B_k$ or C_k. We write $Q(a) = (-a)^{n-1}Q_1(a)$ and obtained $\rho \wedge \omega^{N-1} = M((-a_\rho Q(a))' + p_0 a')$.

Proposition 10. *The scalar curvature is*

$$R = \frac{2(-a_\rho Q)' + pa'}{-a'Q}.$$

Moreover, $p(a) = (-a)^{n-1} p_1(a)$ *with* $p_1(a)$ *a polynomial of* a *and is a positive linear sum of* Q_1 *and product of* $\deg Q_1 - 1$ *factors of* Q_1. *The contribution of each constant factor* k_j *(that is, the vector* F_α *such that the corresponding metrics* $\omega(F_\alpha, JF_\alpha) = k_j$ *is a constant along* p_s*) is* $2k_{\rho,j}/k_j$ *for the* Q_1 *factor. The contribution of each* $a_i \pm a$ *is* $2a_{\rho,i} Q_1/q_i$.

Therefore, we have

$$R_0 = \frac{\int_0^{-l} [(2u_\rho Q)_x + p]\,dx}{\int_0^{-l} Q\,dx} = \frac{2u_\rho(-l)Q(-l) + \int_0^{-l} p\,dx}{\int_0^{-l} Q\,dx},$$

where we let $u = -a$ and $l = \lim_{s\to+\infty} a$. We also obtained in [Guan 2011a] that $a_\rho(0) = 0$.

5. Geodesic stability and existence of the Calabi extremal metrics

In [Guan 2011b, Section 2], for any metric we obtained a function $\Gamma(s)$ such that $-4a = 4u = \Gamma'$ and the geodesic equation is $\ddot{\Gamma}\Gamma'' = (\dot{\Gamma}')^2$, where $'$ is the derivative with respect to s, the parameter from the manifold, and $\dot{}$ is the derivative with respect to t, the parameter for the geodesic. We obtain the smooth geodesics and so the uniqueness. Therefore, we might regard $U = 4u$ as g in [Guan 2011a].

We also have

$$4u_s(+\infty) = \Gamma_{ss}(+\infty) = 0$$

since u is increasing and bounded by $-l$ (see the end of last section).

We shall apply the method in [Guan 2003] to prove the second and third *geodesic stability principles* for all the type I Kähler almost homogeneous manifolds of cohomogeneity one.

The proof is parallel to what we have in [ibid.] but even simpler (with our advanced notation).

Letting H be the Legendre transformation of Γ as in [ibid.], a path Γ_t represents a geodesic in the Mabuchi moduli space of the equivariant Kähler metrics in a given Kähler class is a geodesic if and only if H_t is linear on t. We denote $h = \dot{H}$.

Recall that R is the scalar curvature, HR its average, Q the volume function appeared right before Proposition 10. Applying the scalar curvature formula in Proposition 10, we have that with a positive constant C the derivative of Mabuchi functional is:

$$
\begin{aligned}
-\int_M \dot\Gamma(R-HR)\omega^{2n}
&= -C\int_0^{-l}\dot\Gamma(s,t)\Big(2u_\rho Q+\int(p-R_0Q)\,du\Big)_x\,dx\\
&= C\int_0^{-l}\dot H(x,t)\Big(2u_\rho Q-\int(R_0Q-p)\,du\Big)_x\,dx\\
&= C\Big(2h(-l)u_\rho(-l)Q(-l)-2h(0)u_\rho(0)Q(0)-R_0h(-l)\int_0^{-l}Q\,dx\\
&\quad + R_0\int_0^{-l}h'\Big(\int_0^x Q\,du\Big)\,dx+h(-l)\int_0^{-l}p\,dx-\int_0^{-l}h'\Big(\int_0^x p\,du\Big)\,dx\\
&\qquad -2\sum_1^{n-1}\int_0^{-l}N_i\coth(2N_is)h'Q\,dx+\int_0^{-l}h'(\log(Qu_s))_sQ\,dx\Big)\\
&= C\Big(R_0\int_0^{-l}h'\Big(\int_0^x Q\,du\Big)\,dx-\int_0^{-l}h'\Big(\int_0^x p\,du\Big)\,dx\\
&\qquad -2\sum_1^{n-1}N_i\int_0^{-l}\coth(2N_is)h'Q\,dx+\int_0^{-l}h'(Qu_s)_x\,dx\Big)\\
&= C\Big(R_0\int_0^{-l}h'\Big(\int_0^x Q\,du\Big)\,dx-\int_0^{-l}h'\Big(\int_0^x p\,du\Big)\,dx\\
&\qquad -2\sum_1^{n-1}N_i\int_0^{-l}\coth(2N_is)h'Q\,dx-\int_0^{-l}Qu_sh''\,dx\Big)\\
&= C\Big(R_0\int_0^{-l}h'\Big(\int_0^x Q\,du\Big)\,dx-\int_0^{-l}h'\Big(\int_0^x p\,du\Big)\,dx\\
&\qquad -2\sum_1^{n-1}N_i\int_0^{-l}\coth(2N_iH_x)h'Q\,dx-\int_0^{-l}Q(H_{xx})^{-1}h''\,dx\Big).
\end{aligned}
$$

The change of sign in the second equality comes from $\dot\Gamma(s,t)=-\dot H(x,t)$ for the Legendre transformation as in [Guan 2003].

If h'' is negative somewhere, then the geodesic is finite and the limit is a cone metric. The point $-l$ cannot be a singular point. At the singular points h'' is negative. Therefore, the last term of the right hand side is positive infinite. The second term from the right hand side is finite if 0 is not a singular point and positive if 0 is a singular point since in that case $h''(0)<0$ and $h'(0)=s(0)-s_0(0)=0$, $h'<0$ near 0.

If h'' is nonnegative, then the geodesic ray is infinite and h' is increasing. s becomes infinite at each point with $h' > 0$, so $\coth(2N_i s)$ is 1 at such points. It is not difficult to see that $(H_{xx})^{-1}$ is zero whenever h'' is not zero. The limit of the derivative is:

Theorem 11. *For type I compact Kähler almost homogeneous manifolds of cohomogeneity one, the generalized Futaki invariant of a maximal geodesic ray with a convex function h is*

$$C\left(\int_0^{-l} h'\left(\int_0^x (R_0 Q - p)\,du - 2\sum_1^{n-1} N_i Q\right) dx\right)$$

with a constant $C > 0$.

According to [Guan 2011a, (14)], this is proportional to the negative of

$$\int_0^{-l} h' g_l\,dx.$$

We notice that all the generalized Futaki invariants of the maximal geodesic rays do not depend on the initial metrics and they are positive if there is an extremal metric.

Moreover, if there is a Kähler metric with a constant scalar curvature, then at the corresponding H_0 we have that the slopes of Mabuchi functionals are zeros. Therefore, for any h,

$$\int_0^{-l}\left[h'\left[\int_0^x (R_0 Q - p)\,du - 2\sum_1^{n-1} N_i Q \coth(2N_i H_{0,x})\right] - Q(H_{0,xx})^{-1}h''\right] dx = 0.$$

In general, the slope of the Mabuchi functional is

$$C\int_0^{-l} Q\left(2\sum_1^{n-1} N_i(\coth(2N_i H_{0,x}) - \coth(2N_i H_x))h' + ((H_{0,xx})^{-1} - (H_{xx})^{-1})h''\right) dx$$

$$= C\int_0^{-l} Q\left(4\sum_1^{n-1} N_i \frac{e^{2N_i H_{0,x}}(e^{2N_i t h'} - 1)}{(e^{2N_i(H_{0,x}+th')} - 1)(e^{2N_i H_{0,x}} - 1)} h' + \big((H_{0,xx})^{-1} - (H_{xx})^{-1}\big)h''\right) dx.$$

It turns into

$$C\int_0^{-l} Q\left(\sum_1^{n-1} \frac{4N_i}{e^{2N_i H_{0,x}} - 1} h' + H_{0,xx}^{-1} h''\right) dx.$$

Therefore, using this formula as a hint, we can define

$$\|h\|_*^{2,1} = \int_0^{-l} Q\Big(\sum_1^{n-1} \frac{4N_i}{e^{2N_i H_{0,x}} - 1}|h'| + H_{0,xx}^{-1}|h''|\Big)\,dx$$

to be the norm of $W_*^{2,1}$. A calculation shows that this is related to

$$\int_0^{-l} |\Delta_0 h|\, Q\, dx \quad \text{and also} \quad \int_0^{-l} \sup\{|\partial^2 h(v)|_0/|v|_0\}\, dV,$$

with dV the volume element. The generalized Futaki functional is positive on the closure of the effective cone in $W_*^{2,1}$.

The generalized Futaki functional is positive if and only if it is positive for

$$h' = \begin{cases} 1 & \text{if } x > x_0, \\ 0 & \text{if } x \le x_0, \end{cases}$$

with $x_0 \in [0, -l)$. These functions h' correspond to functions of h in $W_*^{2,1}$ which are the extremal rays of the effective cone. As we see in the sentence right after Theorem 11, this is the same as the partial integral

$$\int_{x_0}^{-l} g_l\, du = \int_{x_0^2}^{l^2} f_l\, dx < 0$$

for the g_l, f_l in [Guan 2011a]. This is the same as the necessary and sufficient condition in [ibid.] (see (7) and (16) there) for the existence of the Kähler metrics with constant scalar curvatures.

Therefore, we obtain:

Theorem 12. *For type I Kähler compact almost homogeneous manifolds of cohomogeneity one*, *there is a unique extremal metric in a Kähler class on the manifold up to the automorphism group if and only if the Kähler class is geodesically stable.*

The same method works for some of Kähler classes on type II compact Kähler almost homogeneous manifolds of cohomogeneity one. But in general, we will use a different method. Theorem 12 and a result similar to Theorem 11 are true for general compact almost homogeneous manifolds of cohomogeneity one. But it will take us some more time to publish the related results and proofs. We also expect that Theorem 12 is true for any Kähler class on any compact Kähler manifold.

Theorem 11 also gives another proof for the stability (the necessary condition) in [ibid.]. However, the integral itself and its partial integrals *do not occur directly* as generalized Futaki invariants of any (smooth) geodesic.

A generalization of our argument is essential to prove the necessary condition for the type II cases (and the type IV case in [Guan 2009]). However, since we have not seen any example with a zero value of the integral for the Ricci classes, for all the known cases so far in [Guan 2009], etc., the corresponding result in the next section is enough for the necessary part for the Kähler–Einstein case.

6. Geodesic stability and strict slope stability

In this section, we shall discuss our result and the strict slope stability. This is something also similar to the holomorphic vector bundle case and can be defined on any Kähler class of any compact Kähler manifold.

6.1. To make the things simpler, first we assume that the Kähler class is the anticanonic class $-K_M$, N is a smooth subvariety and $M(N)$ is the blow-up of M along N. Let E be the exceptional divisor and e be the largest number such that $-K_M - aE > 0$ on $M(N)$, regarding K_M as the pullback line bundle for any a such that $0 < a < e$,

$$m(N) = \int_0^e (-K_M - (n - \dim N)E)(-K_M - aE)^{n-1}\, da,$$

$$m = \int_0^e (-K_M - aE)^n\, da.$$

We say that M is *strictly* slope stable if for any subvariety N (not necessary smooth) *that is not a component of the fixed point set of a holomorphic vector field* we have $m(N) < m$. That is

$$\int_0^e (a - (n - \dim N))E(-K_M - aE)^{n-1}\, da < 0.$$

Notice that there is only one possible zero for $a - (n - \dim N)$, we see that if $m(N) - m < 0$ then

$$\int_0^c (a - (n - \dim N))E(-K - aE)^{n-1}\, da < 0$$

for any $0 < c < e$. That is, when N is smooth, our stability is stronger than Ross–Thomas's *slope stability* in [Ross and Thomas 2006], which only requires the inequality for rational c with $0 < c < e$, while our inequality is true for any c with $0 < c \leq e$. If N is not smooth, we do not know whether the slope stability in [Ross and Thomas 2006] implies these inequalities or not.

A smooth N destabilizes M only if $-K_M - (n - \dim N)E$ is ample, therefore, $-K(E)$ is ample on E if E is smooth, and is kind of ample even if E is singular. When N is smooth, we see that E is Fano. By [Futaki 1987], we see that N is Fano also. This is quite similar to the calculation in [Guan 2003; 2011a].

Actually, when $F = \mathbb{C}P^k$ or $\mathrm{Gr}(2k, 2)$, we have $D(F) = 2$ by [Guan 2011b, Section 3, Theorem 15]. Therefore, for the closed orbit N, $e = -2^{-1}l_\rho$ and the codimension can only be 1; see [ibid., Section 3]. If $y = -l_\rho - 2a$, the integral above is

$$\int_0^{-l_\rho} (-2^{-1}(y+l_\rho)-1)E(\omega+2^{-1}(y+l_\rho)E)^{n-1}2^{-1}\,dy \\ = C\int_0^{-K(F)} (-K(F)-D(F)-y)Q\,dy,$$

with a positive number C. That is exactly the same condition as in Theorem 15 just cited.

When $F = Q^k$, $D(F) = 1$. Therefore, $e = -l_\rho$. Let

$$y = -l_\rho - a = -K(F) + m - 1 - a,$$

with $m = n - \dim N$. The integral above is

$$\int_0^{-l_\rho} (-l_\rho - y - m)E(\omega+(y+l_\rho)E)^{n-1}\,dy \\ = C\int_{-K(F)+m-1} (-K(F)-D(F)-y)Q\,dy,$$

with $C > 0$. Again, that is exactly Theorem 15 in [ibid.].

6.2. In general, for any given Kähler class ω we let

$$m_c(N) = \int_0^c (-K_M - (n-\dim N)E)(\omega - aE)^{n-1}\,da,$$
$$m_c = \int_0^c (\omega - aE)^n\,da,$$

with $0 < c \le e$ and e the largest number such that $\omega - aE > 0$ for $0 < a < e$. We let $\mu_c(N) = m_c(N)/m_c$. If $N = M$, we let $m(M) = (-K_M)\omega^{n-1}$ and $\mu = m(M)/\omega^n$. Then the strict slope stability says that $\mu_c(N) - \mu < 0$ for all $0 < c \le e$. Similar obstructions appeared in [Guan 2003]. At that time I was not able to understand the general meaning of this obstruction and related it to the Ding–Tian generalized Futaki invariant forcibly. But it was clear in [Guan 2003] it was not the Ding–Tian generalized Futaki invariant. I also talked on this at Pisa, Italy in 2004. Ross and Thomas [2006] partially generalized this obstruction but without the *strict* part for a smooth N, that is, they assume that $0 < c < e$. Also, they assume that c is rational, which makes their slope stability much weaker. For a nonsmooth subvariety N, I am not sure that their stability implies these inequalities or not. For our case, our strict slope stability is equivalent to the existence. But the Ross–Thomas slope stability is only a necessary condition. Therefore, *a Kähler class with the integral equal to zero when $c = e$ or c is irrational would give a counterexample for the equivalence between the Ross–Thomas slope stability or Donaldson K-stability and the existence.* See also [Guan 2003; 2007].

It is very easy to check that if K_M is the Kähler class and we replace $-K_M - aE$ by $K_M - aE$, and let

$$m_c(N) = \int_0^c (-K_M - (n - \dim N)E)(K_M - aE)^{n-1}\, da,$$

the strict slope stability means that $m_c(N) + m_c < 0$ holds automatically. Moreover, if $K_M = 0$, for any Kähler metric ω we replace $-K_M - aE$ by $\omega - aE$ and let

$$m_c(N) = \int_0^c (-n + \dim N)E(\omega - aE)^{n-1}\, da,$$

the strict slope stability means that $m_c(N) < 0$ holds automatically. These strengthen the Theorem 5.4 in [Ross and Thomas 2006], which is only concerned with when N is smooth and $0 < c < e$ is rational.

In the remainder of this section, we want to see that the strict slope stability is the same as the existence for type I manifolds.

To make things simpler, let us take care of the $F(\mathrm{OP}_n)$ fiber case first. In our setting, we only need to deal with the case in which N is the closed orbit. In this case, by [Guan 2011b, Section 3], we have $\dim N = n - 1$. Let us calculate the number e for our case. By [ibid., Section 3] we see that the curvature of the exceptional divisor has eigenvalues $D(\mathbb{C}P^n) = 2$ times the coefficient of u. Therefore, $\omega - aE$ has the first zero eigenvalues when $a = (D(F))^{-1}(-l)$. That is, $e = -2^{-1}l$.

$$\omega^n m_c(N) - m(M)m_c = \int_0^{-l} Q\, du \Big[\int_0^c (-K_M)((\omega - xE)^{n-1} - \omega^{n-1}) \\ -E(\omega - xE)^{n-1} - R_0((\omega - xE)^n - \omega^n)\, dx\Big].$$

This is proportional to

$$\int_0^c \Big[\int_0^x \big[(n-1)K_M E(\omega - uE)^{n-2} + nR_0 E(\omega - uE)^{n-1}\big]\, du - E(\omega - xE)^{n-1}\Big]\, dx.$$

Letting $y = -l - 2x$ and $v = -l - 2u$, $d = -l - 2c$, we obtain that the integral is proportional to

$$\int_d^{-l} \Big[\int_y^{-l} \big[(n-1)K_M E(\omega + 2^{-1}(v + l)E)^{n-2} + nR_0 E(\omega + 2^{-1}(v + l)E)^{n-1}\big]\, dv \\ -2E(\omega + 2^{-1}(y + l)E)^{n-1}\Big]\, dy = \int_d^{-l} h_l\, dy.$$

By taking the derivative twice we have

$$h_l' = -(n-1)K(E)E(\omega + 2^{-1}(y + l)E)^{n-2} - nR_0 E(\omega + 2^{-1}(y + l)E)^{n-1}.$$

By the argument in [Guan 2011a] after (14) and in the proof of Lemma 6, we see that h_l' is proportional to g_l' there. Therefore, we only need to check for a point 0, the function h_l is right. To prove our conclusion, we only need to check that

$$h_l(0) = \int_0^{-l} \big[(n-1)K_M E(\omega+2^{-1}(v+l)E)^{n-2} + nR_0 E(\omega+2^{-1}(v+l)E)^{n-1}\big]\, dv = 0,$$

since $g_l(0) = 0$. Notice that $nE(\omega + 2^{-1}(v+l)E)^{n-1}$ is related to ω^n there.

The exact same argument works for the case in which the fiber $F = \mathrm{Gr}(2k, 2)$.

For the case in which the fiber $F = Q^n$, we have $D(F) = 1$. Therefore, we could let $y = -l - x$, $v = -l - u$, $d = -l - c$ instead and we notice

$$-K(E) = -K_M - (n - \dim N)E.$$

The same proof goes through.

Theorem 13. *On a type I compact almost homogeneous manifold of cohomogeneity one there is a Kähler metric of constant scalar curvature in a given Kähler class if and only if the Kähler class is strictly slope stable with respect to the closed orbit.*

This is also true for general compact Kähler almost homogeneous manifolds of cohomogeneity one. But it will take some time for us to publish the detailed results and proofs.

6.3. In the case of Fano manifolds, our discussion in Section 6.1 shows:

Theorem 14. *Let M be any Fano manifold. If a smooth submanifold N destabilizes the Ricci class, then N, the blowing-up manifold $M(N)$ of M along N and the exceptional divisor E are all Fano manifolds.*

One could also consider the case where N is a union of smooth submanifolds. We expect that each of them should be Fano also. Similarly, it should be easy to obtain some results similar to those of Nadel [1990] and to check out the unstable Fano threefolds.

For the compact Kähler manifolds with a zero or negative first Chern class we showed at the beginning of Section 6.2 that:

Theorem 15. *Let M be any compact Kähler manifold with a negative first Chern class. Then the negative Ricci class is strictly slope stable.*

Theorem 16. *Let M be any compact Kähler manifold with a zero first Chern class. Then any Kähler class is strictly slope stable.*

Theorems 14, 15, 16 give a good reason why the Calabi conjecture is true for the negative and zero case but not true in general for the positive case.

References

[Ahiezer 1983] D. Ahiezer, "Equivariant completions of homogeneous algebraic varieties by homogeneous divisors", *Ann. Global Anal. Geom.* **1**:1 (1983), 49–78. MR 85j:32052 Zbl 0537.14033

[Chen 2000] X. Chen, "The space of Kähler metrics", *J. Differential Geom.* **56**:2 (2000), 189–234. MR 2003b:32031 Zbl 1041.58003

[Chen and Tian 2008] X. X. Chen and G. Tian, "Geometry of Kähler metrics and foliations by holomorphic discs", *Publ. Math. Inst. Hautes Études Sci.* **107**:1 (2008), 1–107. MR 2009g:32048 Zbl 1182.32009

[Feng 2012] R. Feng, "Bergman metrics and geodesics in the space of Kähler metrics on principally polarized abelian varieties", *J. Inst. Math. Jussieu* **11**:1 (2012), 1–25. MR 2862373 Zbl 1241.32018

[Futaki 1987] A. Futaki, "The Ricci curvature of symplectic quotients of Fano manifolds", *Tohoku Math. J.* (2) **39**:3 (1987), 329–339. MR 88m:53124 Zbl 0629.53057

[Guan 1995a] D. Guan, "Existence of extremal metrics on compact almost homogeneous Kähler manifolds with two ends", *Trans. Amer. Math. Soc.* **347**:6 (1995), 2255–2262. MR 96a:58059 Zbl 0853.53047

[Guan 1995b] D. Guan, "Quasi-Einstein metrics", *Internat. J. Math.* **6**:3 (1995), 371–379. MR 96e:53060 Zbl 0847.53032

[Guan 1999] D. Guan, "On modified Mabuchi functional and Mabuchi moduli space of Kähler metrics on toric bundles", *Math. Res. Lett.* **6**:5-6 (1999), 547–555. MR 2001b:32042 Zbl 0968.53050

[Guan 2002] D. Guan, "Existence of extremal metrics on almost homogeneous manifolds of cohomogeneity one, II", *J. Geom. Anal.* **12**:1 (2002), 63–79. MR 2002m:53068 Zbl 1030.58006

[Guan 2003] D. Guan, "Existence of extremal metrics on almost homogeneous manifolds of cohomogeneity one, III", *Internat. J. Math.* **14**:3 (2003), 259–287. MR 2004d:32029 Zbl 1048.32014

[Guan 2007] D. Guan, "Extremal solitons and exponential C^∞ convergence of the modified Calabi flow on certain $\mathbb{C}P^1$ bundles", *Pacific J. Math.* **233**:1 (2007), 91–124. MR 2009b:53108 Zbl 1154.53043

[Guan 2009] D. Guan, "Affine compact almost-homogeneous manifolds of cohomogeneity one", *Cent. Eur. J. Math.* **7**:1 (2009), 84–123. MR 2010b:53081 Zbl 1176.53073

[Guan 2011a] D. Guan, "Type I almost-homogeneous manifolds of cohomogeneity one, I", *Pac. J. Appl. Math.* **3**:1-2 (2011), 43–71. MR 2918555

[Guan 2011b] D. Guan, "Type I almost-homogeneous manifolds of cohomogeneity one, II", *Pac. J. Appl. Math.* **3**:3 (2011), 179–201. MR 2918591

[Guan 2011c] D. Guan, "Type II almost-homogeneous manifolds of cohomogeneity one", *Pacific J. Math.* **253**:2 (2011), 383–422. MR 2878816 Zbl 1241.32020

[Guan 2012] Z.-D. Guan, "Positive lemma, generalized extremal-solitons and second order linear equations", *Adv. Dev. Math. Sci.* **1**:2 (2012), 13–32.

[Guan $\geq$ 2013a] D. Guan, "Jacobi fields and geodesic stability", in preparation.

[Guan $\geq$ 2013b] D. Guan, "m-extremal metrics and m-Calabi flow", in preparation.

[Guan and Chen 2000] D. Guan and X. Chen, "Existence of extremal metrics on almost homogeneous manifolds of cohomogeneity one", *Asian J. Math.* **4**:4 (2000), 817–829. MR 2002j:32024 Zbl 1003.32003

[Guan and Phong 2012] P. Guan and D. H. Phong, "Partial Legendre transforms of non-linear equations", *Proc. Amer. Math. Soc.* **140**:11 (2012), 3831–3842. MR 2944724

[Huckleberry and Snow 1982] A. T. Huckleberry and D. M. Snow, "Almost-homogeneous Kähler manifolds with hypersurface orbits", *Osaka J. Math.* **19**:4 (1982), 763–786. MR 84i:32042 Zbl 0507.32023

[Kobayashi 1987] S. Kobayashi, *Differential geometry of complex vector bundles*, Publications of the Mathematical Society of Japan **15**, Princeton University Press, 1987. MR 89e:53100 Zbl 0708.53002

[Mabuchi 1986] T. Mabuchi, "K-energy maps integrating Futaki invariants", *Tohoku Math. J.* (2) **38**:4 (1986), 575–593. MR 88b:53060 Zbl 0619.53040

[Mabuchi 1987] T. Mabuchi, "Some symplectic geometry on compact Kähler manifolds, I", *Osaka J. Math.* **24**:2 (1987), 227–252. MR 88m:53126 Zbl 0645.53038

[Nadel 1990] A. M. Nadel, "Multiplier ideal sheaves and Kähler–Einstein metrics of positive scalar curvature", *Ann. of Math.* (2) **132**:3 (1990), 549–596. MR 92d:32038 Zbl 0731.53063

[Nakagawa 2011] Y. Nakagawa, "On generalized Kähler–Ricci solitons", *Osaka J. Math.* **48**:2 (2011), 497–513. MR 2012i:32032 Zbl 1234.32006

[Ross and Thomas 2006] J. Ross and R. Thomas, "An obstruction to the existence of constant scalar curvature Kähler metrics", *J. Differential Geom.* **72**:3 (2006), 429–466. MR 2007c:32028 Zbl 1125.53057

[Semmes 1992] S. Semmes, "Complex Monge–Ampère and symplectic manifolds", *Amer. J. Math.* **114**:3 (1992), 495–550. MR 94h:32022 Zbl 0790.32017

Received November 4, 2011. Revised September 6, 2012.

DANIEL GUAN
DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF CALIFORNIA AT RIVERSIDE
RIVERSIDE CA 92521
UNITED STATES

zguan@math.ucr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.389

THE SUBREPRESENTATION THEOREM FOR AUTOMORPHIC REPRESENTATIONS

MARCELA HANZER

We prove that every irreducible subrepresentation in the space of automorphic forms on $G(\mathbb{A})$, where G is a connected reductive group defined over a number field k, and $\mathbb{A}$ is the related ring of adeles, is a subrepresentation of the representation induced from a cuspidal automorphic representation of a Levi subgroup.

1. Introduction

In this note we prove the global (automorphic) version (over a number field k) of Casselman's subrepresentation theorem. We explain it in more detail: in the local theory (i.e., considering admissible representations of reductive groups over local fields) there is Harish-Chandra's subquotient theorem [1954], and then there is also Casselman's subrepresentation theorem [1980; 1995]; both of them state that every irreducible representation (in the appropriate category) of this given reductive group is a subquotient or (in the case of Casselman's theorem) a subrepresentation of a representation induced from a "simpler" one (of an appropriate subgroup). The global analog of the Harish-Chandra subquotient theorem would be Langlands' theorem which describes a general automorphic representation as a subquotient of a representation induced from a cuspidal representation of a Levi subgroup.

We prove the following global version of Casselman's subrepresentation theorem.

Theorem. *Let G be a connected reductive group defined over k. Let (Π, V) be an $((\mathfrak{g}_\infty, K_\infty)\times G(\mathbb{A}_f))$-irreducible subspace of automorphic forms in $A(G(k)\backslash G(\mathbb{A}))$. Then, there exists a parabolic subgroup $P = MU$ of G, an irreducible automorphic cuspidal representation π_0 of M (thus appearing in the space of cuspidal automorphic forms on M) such that, as abstract global representations, we have*

$$\Pi \hookrightarrow \mathrm{ind}_{P(\mathbb{A})}^{G(\mathbb{A})}\pi_0,$$

where we consider the normalized parabolic induction (so we extend π_0 trivially on $U(\mathbb{A})$) and we take $\mathbf{K}$*-finite vectors.*

MSC2010: 11F70.
Keywords: automorphic representation, subrepresentation, cuspidal representation.

We explain all the notation in the Preliminaries section.

We are sure that the experts in the field are aware of the above claim, but we were not able to find the reference for this statement, which is somewhat more precise than the aforementioned Langlands' result in his Corvallis lecture [Borel and Jacquet 1979]. The proof is a pretty straightforward application of the Langlands proof in his Corvallis lecture, with the decomposition results (on the spaces of automorphic forms) obtained (along with much stronger results) in [Mœglin and Waldspurger 1995]. We hope that this result will be very helpful for explicit calculations with automorphic forms, since it is explicitly applicable to the discrete (and **K**-finite) part of automorphic L^2 situation.

2. Preliminaries

Let k be a number field, and $\mathbb{A}$ its ring of adeles. Let G be a connected reductive group defined over k, and $G_\infty = \prod_v G(k_v)$, where the product is over archimedean places of k. We further denote $G(\mathbb{A}_f) = \prod'_{v<\infty} G(k_v)$. Let $\mathcal{U}$ be the enveloping algebra of the complexified Lie algebra $\mathfrak{g}$ of G_∞ (and $\mathfrak{g}_\infty$ is the Lie algebra of G_∞). We follow the notation of the first chapter of [Mœglin and Waldspurger 1995]. We denote by $\mathfrak{z}$ the center of $\mathcal{U}$ and by K_v a maximal compact subgroup of $G(k_v)$, where $K_v = G(O_{k_v})$ for almost all $v < \infty$. Here O_{k_v} is the ring of integers in k_v. We set $K_\infty = \prod_{v|\infty} K_v$ and $\mathbf{K} = \prod_v K_v$. We fix a minimal parabolic subgroup P_0 of G defined over k, and consequently, standard parabolic subgroups (defined over k) with respect to P_0. We denote by S a maximal k-split torus of G, chosen inside P_0 and by Δ the set of simple k-roots of G with respect to S (and P_0). We know that each standard k-parabolic subgroup corresponds to a subset θ of Δ. We denote this by putting $P = P_\theta$. We denote the modular function on P by δ_P. For a standard Levi k-subgroup M of G, we denote by $\mathfrak{z}^M$ the analogue of $\mathfrak{z}$ for group M. We denote by Z_M the center of M.

We use the following definition of an automorphic form: Let $P = MU$ be a standard k-parabolic subgroup of G and $\phi : U(\mathbb{A})M(k) \setminus G(\mathbb{A}) \to \mathbb{C}$ a function. We say that ϕ is automorphic if it satisfies the following conditions:

(1) ϕ has moderate growth (see [Mœglin and Waldspurger 1995, I.2.3]).

(2) ϕ is smooth (see [Mœglin and Waldspurger 1995, I.2.5]).

(3) ϕ is **K**-finite.

(4) ϕ is $\mathfrak{z}$-finite.

Note that the space $A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ of all automorphic forms as above can be related to the usual situation with the automorphic forms on $M(k) \setminus M(\mathbb{A})$ by attaching to each $k \in \mathbf{K}$ and ϕ as above a function $\phi_k : M(k) \setminus M(\mathbb{A}) \to \mathbb{C}$ defined by $\phi_k(m) = \delta_P^{-1/2}(m)\phi(mk)$ by noting that ϕ is automorphic if and only if it is smooth,

$\mathbf{K}$-finite, and for all $k \in \mathbf{K}$, ϕ_k is an automorphic form on $M(k) \setminus M(\mathbb{A})$. We denote by $A_0(U(\mathbb{A})M(k)\setminus G(\mathbb{A}))$ the cuspidal part of the space $A(U(\mathbb{A})M(k)\setminus G(\mathbb{A}))$; i.e., the space of all automorphic forms ϕ from $A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ with the property that for every standard k-parabolic subgroup $P' = M'U'$ such that $P_0 \subset P' \subsetneqq P$ we have $\phi_{P'} = 0$ (the constant term along P', defined by $\phi_{P'}(g) = \int_{U'(k)\setminus U'(\mathbb{A})} \phi(ug)\mathrm{d}u$).

The space $A(U(\mathbb{A})M(k)\setminus G(\mathbb{A}))$ is a module for the action of $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$, i.e., for the global idempotent Hecke algebra $\mathcal{H} = \mathcal{H}_\infty \otimes \mathcal{H}_f$, where $\mathcal{H}_\infty$ is related to $\mathcal{U}$ and finite measures on K_∞, and $\mathcal{H}_f = \otimes'_{v<\infty} \mathcal{H}_v$, where $\mathcal{H}_v,\ v < \infty$ is the Hecke algebra of compactly supported, locally constant functions on $G(k_v)$ (see [Borel and Jacquet 1979, Section 4]). Note that $A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ is a submodule of $A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ with this action. Note that the constant term (with respect to some standard k-parabolic subgroup $P = MU$) is an intertwining operator between $A(G(k) \setminus G(\mathbb{A}))$ and $A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ [Mœglin and Waldspurger 1995, I.2.6].

Let ξ be a character of $Z_M(k) \setminus Z_M(\mathbb{A})$, and let π be an irreducible submodule of $A(M(k) \setminus M(\mathbb{A}))$, for a standard k-Levi subgroup M of G. We denote by $A(M(k) \setminus M(\mathbb{A}))_\pi$ the isotypic submodule attached to π (in the theorem below we deal with cuspidal π, so the relevant subquotients are indeed subspaces). We set

$$\begin{aligned} A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\xi &= \big\{\phi \in A(U(\mathbb{A})M(k) \setminus G(\mathbb{A})) : \\ &\qquad \phi(zg) = \delta_P^{1/2}(z)\xi(z)\phi(g) \text{ for all } z \in Z_M(\mathbb{A}),\ g \in G(\mathbb{A})\big\}, \\ A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\pi &= \big\{\phi \in A(U(\mathbb{A})M(k) \setminus G(\mathbb{A})) : \\ &\qquad \phi_k \in A(M(k) \setminus M(\mathbb{A}))_\pi \text{ for all } k \in \mathbf{K}\big\}. \end{aligned}$$

Analogously, we define by $A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\xi$ and $A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\pi$ the cuspidal parts of the above spaces (i.e., the parts realized in the space of cuspidal automorphic forms).

Proposition 2.1. *Let ξ be a character of $Z_M(k) \setminus Z_M(\mathbb{A})$ and let $\Pi_0(M)_\xi$ denote the set of isomorphism classes of irreducible representations of $M(\mathbb{A})$ occurring as submodules in $A_0(M(k) \setminus M(\mathbb{A}))_\xi$. We have the decomposition*

$$A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\xi = \bigoplus_{\pi \in \Pi_0(M)_\xi} A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))_\pi.$$

Proof. This is explained in [Mœglin and Waldspurger 1995, p. 44]. □

Remark. By the proof of Lemma I.3.2 of [Mœglin and Waldspurger 1995], $\mathfrak{z}^M$ acts on $A(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ by left translations; every automorphic form there is $\mathfrak{z}^M$-finite; analogously every element of that space is $Z_M(\mathbb{A})$-finite, again here $Z_M(\mathbb{A})$ acts by left translations (because we examine $\mathbf{K}$-finite automorphic forms). Also, it is easy to see that $A_0(U(\mathbb{A})M(k) \setminus G(\mathbb{A}))$ is $Z_M(\mathbb{A})$-invariant subspace with this $Z_M(\mathbb{A})$-action.

3. The theorem

In this section we prove the main theorem stated in Section 1. The proof follows directly from the next theorem, so our embedding from the main theorem is realized through the calculation of the constant term.

Theorem 3.1. *Let* (Π, V) *be an* $((\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$*-) irreducible subspace of automorphic forms inside* $A(G(k) \setminus G(\mathbb{A}))$ *such that some constant term of a function from* V *does not vanish along a k-parabolic subgroup* P_θ *of* G; *assume that* θ *is minimal (set of simple roots) with this property. Then, there exists an irreducible automorphic representation* π_0 *of* $M_\theta(\mathbb{A})$ *(appearing in* $A_0(M_\theta(k) \setminus M_\theta(\mathbb{A}))$*) such that the space of constant terms of* V *along* P_θ*, denoted by* V_0*, belongs (up to a left translation by an element from* $Z_{M_\theta}(\mathbb{A})$*) to the space* $A_0(U_\theta(\mathbb{A})M_\theta(k) \setminus G(\mathbb{A}))_{\pi_0}$ *of cuspidal automorphic forms.*

Proof. Let $f \in V$. By definition, the constant term $f_{P_\theta}(g) = \int_{U_\theta(k)\setminus U_\theta(\mathbb{A})} f(ug)\mathrm{d}u$ belongs to $A(U_\theta(\mathbb{A})M_\theta(k)\setminus G(\mathbb{A}))$, more precisely, to the cuspidal part of this space (because of the minimality of θ; see [Mœglin and Waldspurger 1995, I.2.6, I.2.18]). By the remark above the Theorem, $Z_{M_\theta}(\mathbb{A})$ acts on $A_0(U_\theta(\mathbb{A})M_\theta(k) \setminus G(\mathbb{A}))$ by left translations, and every function from this space is $Z_{M_\theta}(\mathbb{A})$-finite. For every $z \in Z_{M_\theta}(\mathbb{A})$, let $V_0^z = l(z)V_0$ (the action by left translations). We know that taking the constant term is intertwining operator, so V_0 (and V_0^z) is (as an abstract $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$- representation) irreducible and isomorphic to V. Let $W = \sum_{z\in Z_{M_\theta}(\mathbb{A})} V_0^z$.

We prove that there exists $F \in W,\ F \neq 0$ such that $\dim_{\mathbb{C}} \mathrm{span}_{\mathbb{C}}\{l(z)F : z \in Z_{M_\theta}(\mathbb{A})\} = 1$. Firstly, let $F \neq 0$ be an element from W such that the dimension of the space $Y := \mathrm{span}_{\mathbb{C}}\{l(z)F : z \in Z_{M_\theta}(\mathbb{A})\}$ is minimal. We claim that this dimension is one. Indeed, let us assume that this dimension (of Y) is greater than one. If, for every $a \in Z_{M_\theta}(\mathbb{A})$ acting on Y, the whole space Y is an eigenspace for certain eigenvalue, it would mean that $l(a)$, for every a, acts as a scalar operator on Y, and then every one-dimensional subspace, (also the one spanned by F) would be $Z_{M_\theta}(\mathbb{A})$-invariant; a contradiction (this would mean that Y is one-dimensional). So, there exists $a \in Z_{M_\theta}(\mathbb{A})$ with a nonzero eigenspace strictly smaller than Y, attached to an eigenvalue $\alpha \neq 0$. This means that $Y_1 := (l(a) - \alpha)Y$ is a proper subspace of Y. Let $F_1 := (l(a) - \alpha)F \in Y_1$. F_1 is obviously nonzero; otherwise $l(b)F$ would be an eigenvector of $l(a)$ for eigenvalue α for every $b \in Z_{M_\theta}(\mathbb{A})$, so that the whole Y is an eigenspace for α; a contradiction. Now, we easily see that the span of the set $\{l(b)F_1 : b \in Z_{M_\theta}(\mathbb{A})\}$ is inside Y_1, which leads to contradiction with our choice of F.

So, we conclude that there exists a character ξ of $Z_M(k) \setminus Z_M(\mathbb{A})$ such that

$$l(z)F(g) = \delta_{P_\theta}^{1/2}(z)\xi(z)F(g) \quad \text{for all } g \in G(\mathbb{A}),\ z \in Z_{M_\theta}(\mathbb{A}). \tag{1}$$

Now, let W_0 denote the $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$-subspace of W generated by F. For every vector from this space, (1) holds. Now, since $W = \sum_{a \in Z_{M_\theta}(\mathbb{A})} V_0^a$, where V_0^a are irreducible subspaces, W is also a direct sum of irreducible subspaces (for example, [Lang 2002, Chapter XVII]), and every $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$-submodule of W is a direct summand. From this directly follows that W_0 has an irreducible submodule; indeed if $W = \bigoplus_{z \in I} V_0^z$, for some $I \subset Z_{M_\theta}(\mathbb{A})$, then some projection attached to this decomposition $p_z : W \to V_0^z$ is nonzero on W_0. Now $\operatorname{Ker} p_z \cap W_0$ has a direct (invariant) complement W_1 in W, and it is easy to see that $W_1 \cap W_0$ is an irreducible submodule of W_0. This means that we have found an irreducible subspace of W (so necessarily isomorphic to V i.e., to V_0) where the relation (1) holds. This realization of V inside $A_0(U_\theta(\mathbb{A}) M_\theta(k) \setminus G(\mathbb{A}))_\xi$ is thus obtained through taking of (maybe translated) constant term along P_θ. From Proposition 2.1 we have

$$A_0(U_\theta(\mathbb{A}) M_\theta(k) \setminus G(\mathbb{A}))_\xi = \bigoplus_{\pi \in \Pi_0(M_\theta)_\xi} A_0(U_\theta(\mathbb{A}) M_\theta(k) \setminus G(\mathbb{A}))_\pi,$$

and, combining our embedding with an appropriate projection, we have obtained an embedding

$$\Pi \hookrightarrow A_0(U_\theta(\mathbb{A}) M_\theta(k) \setminus G(\mathbb{A}))_{\pi_0},$$

for some automorphic (cuspidal) representation π_0 of $M_\theta(\mathbb{A})$. □

Note that the space $A_0(M_\theta(k) \setminus M_\theta(\mathbb{A}))_{\pi_0}$ is semisimple (Gelfand and Piatetski–Shapiro; see [Borel and Jacquet 1979, Section 4]); so there exists an irreducible subspace V_0' of automorphic forms in $A_0(M_\theta(k) \setminus M_\theta(\mathbb{A}))_{\pi_0}$ (thus isomorphic to π_0) such that there is an embedding

$$\Pi \hookrightarrow A_0(U_\theta(\mathbb{A}) M_\theta(k) \setminus G(\mathbb{A}))_{V_0'}$$

(the space on the right-hand side has an obvious meaning). We note that, as a $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbb{A}_f)$-module, the latter space is isomorphic to the global representation $\operatorname{ind}_{P_\theta(\mathbb{A})}^{G(\mathbb{A})} \pi_0$ (where we use normalized induction and **K**-finite vectors in this space) [Kim 2004, Section 4.5]. This isomorphism can also be given explicitly by $\phi \mapsto \phi'$, where $\phi'(g) = \phi_g$ and $\phi_g(m) = \delta_{P_\theta}(m)^{-1/2} \phi(mg)$. This is easily checked to be $G(\mathbb{A})$-isomorphism on the space of the smooth (not necessarily **K**-finite automorphic forms), but then taking **K**-finite vectors from both spaces, we get the claim (see the second and third lectures in [Cogdell 2004]). This, in turn, proves our main theorem from Section 1.

Acknowledgements

We want to thank Neven Grbac for the helpful discussions about automorphic representations, and to Goran Muić and Marko Tadić for their interest in our work.

Also, we would like to thank the anonymous referee for suggestions, which improved the style of the presentation and clarified some mathematical imprecisions.

References

[Borel and Jacquet 1979] A. Borel and H. Jacquet, "Automorphic forms and automorphic representations", pp. 189–207 in *Automorphic forms, representations and L-functions, 1* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR 81m:10055 Zbl 0414.22020

[Casselman 1980] W. Casselman, "Jacquet modules for real reductive groups", pp. 557–563 in *Proceedings of the International Congress of Mathematicians* (Helsinki, 1978), vol. 2, edited by O. Lehto, Acad. Sci. Fennica, Helsinki, 1980. MR 83h:22025 Zbl 0425.22019

[Casselman 1995] W. Casselman, "Introduction to the theory of admissible representations of p-adic reductive groups", preprint, 1995, http://www.math.ubc.ca/∼cass/research/pdf/p-adic-book.pdf.

[Cogdell 2004] J. W. Cogdell, "Lectures on L-functions, converse theorems, and functoriality for GL_n", pp. 1–96 in *Lectures on automorphic L-functions*, edited by J. W. Cogdell et al., Fields Inst. Monogr. **20**, Amer. Math. Soc., Providence, RI, 2004. MR 2071506 Zbl 1066.11021

[Harish-Chandra 1954] Harish-Chandra, "Representations of semisimple Lie groups, II", *Trans. Amer. Math. Soc.* **76** (1954), 26–65. MR 15,398a Zbl 0055.34002

[Kim 2004] H. H. Kim, "Automorphic L-functions", pp. 97–201 in *Lectures on automorphic L-functions*, edited by J. W. Cogdell et al., Fields Inst. Monogr. **20**, Amer. Math. Soc., Providence, RI, 2004. MR 2071507 Zbl 1066.11021

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, New York, 2002. MR 2003e:00003 Zbl 0984.00001

[Mœglin and Waldspurger 1995] C. Mœglin and J.-L. Waldspurger, *Spectral decomposition and Eisenstein series: a paraphrase of scripture*, Cambridge Tracts in Mathematics **113**, Cambridge University Press, Cambridge, 1995. MR 97d:11083 Zbl 0846.11032

Received October 8, 2011. Revised July 23, 2012.

MARCELA HANZER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ZAGREB
BIJENIČKA CESTA 30
10000 ZAGREB
CROATIA
hanmar@math.hr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.395

VARIATIONAL CHARACTERIZATIONS OF THE TOTAL SCALAR CURVATURE AND EIGENVALUES OF THE LAPLACIAN

SEUNGSU HWANG, JEONGWOOK CHANG AND GABJIN YUN

For the dual operator $s_g'^*$ of the linearization s_g' of the scalar curvature function, it is well-known that if $\ker s_g'^* \neq 0$, then s_g is a nonnegative constant. Moreover, if the Ricci curvature does not vanish, then $s_g/(n-1)$ is an eigenvalue of the Laplacian of the metric g. In this work, we give some variational characterizations for the space $\ker s_g'^*$. To accomplish this, we introduce a fourth-order elliptic differential operator $\mathscr{A}$ and a related geometric invariant ν. We prove that ν vanishes if and only if $\ker s_g'^* \neq 0$, and if the first eigenvalue of the Laplace operator is large compared to its scalar curvature, then ν is positive and $\ker s_g'^* = 0$. We calculate a lower bound for ν in the case of $\ker s_g'^* = 0$. We also show that if there exists a function which is $\mathscr{A}$-superharmonic and the Ricci curvature has a lower bound, then the first nonzero eigenvalue of the Laplace operator has an upper bound.

1. Introduction

Let M be a compact smooth n-manifold (without a boundary). The space of all Riemannian metrics, $\mathscr{M}$, on M is then open in the space of symmetric 2-tensors, $\mathscr{S}^2(M)$, for the compact-open topology or the $W^{k,p}$-topology, where $W^{k,p}$ denotes the Sobolev space. For a Riemannian metric g and a symmetric 2-tensor h, the differential $s_g'(h)$ of the scalar curvature at g in the direction h is given by

$$s_g'(h) = -\Delta_g \operatorname{tr}(h) + \delta_g(\delta_g h) - g(r_g, h), \tag{1-1}$$

where Δ_g is the negative Laplacian of g, and r_g and δ_g denote the Ricci curvature and divergence operator of g, respectively [Besse 1987]. In addition, the L^2-adjoint

Hwang was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST) grant 2011-0005211, Chang by MEST grant 2011-0005235, and Yun by MEST grant 2011-0007465. Yun is the corresponding author.
MSC2010: 53C21.
Keywords: critical point equation, fourth-order elliptic operator, eigenvalue, Einstein metric, Laplace operator, scalar curvature, total scalar curvature.

operator $s_g'^*$ of s_g' is given by

$$s_g'^*(f) = Ddf - (\Delta_g f)g - fr_g, \tag{1-2}$$

where Ddf denotes the Hessian of f with respect to the metric g. Note that both s_g' and $s_g'^*$ are linear second-order differential operators.

In this paper, we consider the fourth-order elliptic differential operator $\mathscr{A} = s_g' \circ s_g'^* : C^\infty(M) \to C^\infty(M)$. The existence of homogeneous or nonhomogeneous solutions to $\mathscr{A}$ is closely related to the kernel space of $s_g'^*$. For example, Bourguignon [1975] and Fischer and Marsden [1974] proved that if $\ker s_g'^* \neq 0$, then either (M, g) is Ricci-flat and $\ker s_g'^* = \mathbb{R} \cdot 1$, or the scalar curvature is a strictly positive constant and $s_g/(n-1)$ is an eigenvalue of the Laplacian. In particular, combined with the Lichnerowicz–Obata theorem [Lichnerowicz 1958; Obata 1962; Berger et al. 1971], it follows that if g is an Einstein metric with positive scalar curvature, then $\ker s_g'^* = 0$ or g is the standard round metric on the sphere.

On the other hand, if $\ker s_g'^* = 0$, then for any function $\eta \in C^\infty(M)$ there exists a unique function $u \in C^\infty(M)$ such that $\mathscr{A}(u) = \eta$ (Theorem 2.2). In fact, the condition $\ker s_g'^* = 0$ implies the injectivity of $s_g'^*$ and the surjectivity of s_g'. In order to perform variational characterizations of the condition $\ker s_g'^* \neq 0$, we introduce a geometric invariant ν which is defined by

$$\nu = \inf \left\{ \int_M \varphi \mathscr{A} \varphi \, dv_g \right\},$$

where the infimum is taken over all functions $\varphi \in H^2(M) = W^{2,2}(M)$ with $\int_M \varphi^2 = 1$. Here $H^2(M) = W^{2,2}(M)$ denotes the Sobolev space which is L^2 up to the second (weak) derivatives.

A basic result related to the invariant ν is the following.

Theorem A. *The invariant ν vanishes if and only if* $\ker s_g'^* = 0$.

For the case $\ker s_g'^* = 0$, we give a lower bound on ν and its relationship to the first nonzero eigenvalue of the Laplacian. We also show that if the first eigenvalue is large compared to the scalar curvature, then ν is positive and $\ker s_g'^* = 0$. In addition, if M is the product of two standard spheres of the same dimension, then ν is exactly equal to the dimension of the spheres.

Theorem B. *Let $M = S^m \times S^m$ $(m \geq 2)$ with the standard product metric. Then*

$$\nu = m = \frac{\dim(M)}{2}.$$

We also obtain upper bounds for the first nonzero eigenvalue of the Laplace operator when $\mathscr{A}$ satisfies a condition on evaluating functions. We say that a Riemannian manifold (M, g) satisfies the $\mathscr{A}$-superharmonic condition if there exists a smooth function φ such that $M_\varphi^+ \neq \varnothing$ and $\mathcal{A}\varphi \le 0$ on M_φ^+, and $\Delta\varphi = 0$ on the boundary ∂M_φ^+ of M_φ^+. Here $M_\varphi^+ = \{x \in M : \varphi(x) > 0\}$. For example, if M is the standard sphere, then the first eigenfunction of the Laplacian satisfies these conditions. In general, any compact Riemannian manifold (M, g) with positive scalar curvature and $\ker s_g'^* \neq 0$ satisfies the $\mathcal{A}$-superharmonic condition.

One of our main results is the following.

Theorem C. *Let (M^n, g) be a compact n-dimensional Riemannian manifold with a positive constant scalar curvature s_g. Suppose that (M, g) satisfies the $\mathcal{A}$-superharmonic condition. If* $\mathrm{Ric}_g \ge k \ge 0$, *then the first nonzero eigenvalue λ_1 of the Laplacian satisfies*

$$\lambda_1 \le \frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n}}{2(n-1)}. \tag{1-3}$$

Inequality (1-3) is sharp since the equality holds for the standard sphere. In performing analysis with the operator $\mathcal{A}$, the main difficulty is that we cannot apply the theory of second-order elliptic partial differential equations directly since $\mathcal{A}$ is a fourth-order differential operator.

The kernel space of $s_g'^*$ plays an important role in the critical point equation arising from the total scalar curvature functional. Let $\mathcal{M}_1$ be the set of all smooth Riemannian metrics of unit volume on M, and let $\mathcal{C} \subset \mathcal{M}_1$ be the set of all smooth Riemannian metrics on M with constant scalar curvature, i.e.,

$$\mathcal{C} = \{g \in \mathcal{M}_1 : s_g = \text{ constant}\}.$$

The total scalar curvature $\mathcal{S} : \mathcal{M}_1 \to \mathbb{R}$ is defined as

$$\mathcal{S}(g) = \int_M s_g \, dv_g.$$

It is well-known that the total scalar curvature functional $\mathcal{S}$ restricted to $\mathcal{C}$ will be critical at g if and only if there is a function f with $\int_M f = 0$ such that

$$z_g = s_g'^*(f), \tag{1-4}$$

where z_g is the traceless Ricci tensor defined as $z_g = r_g - (s_g/n)g$. We call (1-4) the critical point equation (CPE). Note that if $f = 0$, it follows from (1-4) that $z_g = 0$, and thus g is an Einstein metric. However, the existence of a nonzero solution is a very strong condition. The only known case with a nonzero solution is that of a standard sphere, and it has been conjectured that this is the only possible case [Besse 1987]. Namely, it is believed that if there exists a nonzero function f satisfying the CPE, then g must be an Einstein metric. We remark that a solution (g, f) to the CPE is a nontrivial example of the $\mathcal{A}$-superharmonic condition since $\mathcal{A}f = -|z_g|^2$ and $\Delta_g f = -(s_g/(n-1))f$.

Unless stated otherwise, we only consider Riemannian metrics on M whose scalar curvatures are positive constants.

2. Variational properties

Let (M, g) be a closed Riemannian n-manifold and δ be the adjoint operator of the differential d with respect to the metric g. Unless explicitly stated, we will use r rather than r_g as the Ricci tensor of the metric g, and s rather than s_g as the scalar curvature. The following expressions are well-known definitions and identities: for a function φ and any tensor T,

$$\delta D d\varphi = -d\Delta\varphi - r(d\varphi,\cdot), \quad \delta d\varphi = -\Delta\varphi, \quad \text{and} \quad \delta(\varphi T) = \varphi\delta T - T(d\varphi,\cdot).$$

Moreover, for any two functions φ, ψ,

$$\psi\langle Dd\varphi, r\rangle = -\delta(\psi r(d\varphi,\cdot)) - r(d\varphi, d\psi). \tag{2-1}$$

Lemma 2.1. *Let $\mathscr{A} = s_g' \circ s_g'^*$ and assume the scalar curvature $s_g = s$ is constant. Then, for any function φ,*

$$\mathscr{A}(\varphi) = (n-1)\Delta^2\varphi + 2s\Delta\varphi - \langle Dd\varphi, r\rangle + \varphi|r|^2.$$

Proof. It follows directly from (1-2) that

$$s_g'^*(\varphi) = Dd\varphi - (\Delta\varphi)g - \varphi r$$

and thus

$$\mathscr{A}(\varphi) = s_g' \circ s_g'^*(\varphi) = s_g'(Dd\varphi - (\Delta\varphi)g - \varphi r).$$

By (1-1), we have

$$s_g'(Dd\varphi) = -\delta(r(d\varphi,\cdot)) - \langle Dd\varphi, r\rangle.$$

Similarly, since $\delta g = 0$ and $\delta r = -\frac{1}{2}ds = 0$, we also obtain the following from (1-1):

$$s_g'((\Delta\varphi)g) = (1-n)\Delta^2\varphi - s\Delta\varphi \quad \text{and} \quad s_g'(\varphi r) = -s\Delta\varphi + \delta(-r(d\varphi,\cdot)) - \varphi|r|^2.$$

Combining these two expressions, we obtain

$$\mathscr{A}(\varphi) = (n-1)\Delta^2\varphi + 2s\Delta\varphi - \langle Dd\varphi, r\rangle + \varphi|r|^2. \qquad \square$$

Note that $\mathscr{A}$ is a fourth-order linear partial differential operator. The following theorem shows that $\mathscr{A}$ is elliptic and self-adjoint. We say that a fourth-order differential operator is elliptic if the symbol is injective.

Theorem 2.2. *The operator $\mathscr{A}$ is a self-adjoint, fourth-order elliptic linear operator. Furthermore, if $\ker s_g'^* = 0$, then for any $\psi \in C^\infty(M)$ there exists a unique function $u \in C^\infty(M)$ such that $\psi = \mathscr{A}(u)$.*

Proof. We first show that $s_g'^*$ has injective symbol. Recall that for any $p \in M$ and any cotangent vector $t \in T_p^* M$, there is a linear map $\sigma_t(s_g'^*) : T_p C^\infty(M) \to T_p C^\infty(S^2 M)$ called the symbol of $D = s_g'^*$, and the symbol of D is called injective if $\sigma_t(D)$ is injective for all nonzero t. Note that for $t \in T^* M$, $\psi \in C^\infty(M)$,

$$\sigma_t(s_g'^*) \cdot \psi = (-g(t,t)g + t \otimes t)\psi,$$

which is clearly injective for $n > 1$. Thus $s_g'^*$ is an operator of order 2 with injective symbol. By Lemma 4.4 of [Berger and Ebin 1969], $\mathscr{A} = s_g' \circ s_g'^*$ is an elliptic operator of order 4. It is clear from definition that $\mathscr{A}$ is self-adjoint.

Secondly, we show that $\mathscr{A}$ is surjective. Since s_g' is surjective, for any nontrivial $\psi \in C^\infty(M)$, there exists $\xi \in C^\infty(S^2 M)$ such that $s_g'(\xi) = \psi$. From the fact that s_g is constant and the proof of Theorem 5.2 in the same reference, $C^\infty(S^2 M) = \operatorname{im} s_g'^* \oplus \ker s_g'$. Thus, $\xi = \xi_1 + \xi_2$ with $\xi_1 \in \operatorname{im} s_g'^*$ and $\xi_2 \in \ker s_g'$. Therefore, for $\xi_1 = s_g'^*(u)$, we have $\mathscr{A}(u) = \psi$.

Finally uniqueness comes from the assumption that $\ker s_g'^* = 0$ since $\ker \mathscr{A} = \ker s_g'^*$; clearly $\ker s_g'^* \subset \ker s_g' \circ s_g'^*$, and $s_g \circ s_g'^*(u) = 0$ implies

$$0 = \big(u, s_g \circ s_g'^*(u)\big)_{L^2} = \big(s_g'^*(u), s_g'^*(u)\big)_{L^2},$$

where $(f,g)_{L^2} = \int_M fg \, dv_g$, and so $s_g'^*(u) = 0$. □

Given a smooth compact n-dimensional Riemannian manifold (M, g), we let $H^2(M) = W^{2,2}(M)$ be the Sobolev space defined as the completion of the space of smooth functions on M with respect to the norm

$$\|\varphi\|_{H^2(M)}^2 = \int_M |Dd\varphi|^2 \, dv_g + \int_M |\nabla\varphi|^2 \, dv_g + \int_M \varphi^2 \, dv_g.$$

To investigate the properties of operator $\mathscr{A}$ from the perspective of the calculus of variations, we define $E(\varphi)$ for any function $\varphi \in H^2(M)$ as

$$E(\varphi) = \frac{1}{2} \int_M \big[(n-1)(\Delta\varphi)^2 - 2s|d\varphi|^2 + r(d\varphi, d\varphi) + \varphi^2 |r|^2\big]. \tag{2-2}$$

Since $\varphi\langle Dd\varphi, r\rangle = \operatorname{div}(\varphi r(d\varphi, \cdot)) - r(d\varphi, d\varphi)$, and thus

$$\int_M \varphi\langle Dd\varphi, r\rangle = -\int_M r(d\varphi, d\varphi),$$

the Euler–Lagrange equation for the functional E is exactly

$$\mathscr{A}(\varphi) = (n-1)\Delta^2\varphi + 2s\Delta\varphi - \langle Dd\varphi, r\rangle + \varphi|r|^2 = 0.$$

Note that if $\varphi =$ constant and $\mathcal{A}(\varphi) = 0$, then $\varphi = 0$ if the Ricci curvature r does not identically vanish. Furthermore,

$$E(\varphi) = \frac{1}{2}\int_M \varphi\mathcal{A}(\varphi) = \frac{1}{2}\int_M |s_g'^*\varphi|^2 \geq 0 \tag{2-3}$$

for any function φ. In other words, E is the energy of $\mathcal{A}$.

A simple direct observation is as follows.

Lemma 2.3. *The kernel of $s_g'^*$ vanishes if and only if* $\ker \mathcal{A} = 0$.

Proof. The proof follows from the fact that

$$\int_M (s_g'^*\varphi)^2 = \int_M \varphi\mathcal{A}(\varphi)$$

for any function φ. In fact, assume that $\ker \mathcal{A} = 0$ and let $s_g'^* u = 0$. Then u realizes the infimum of $E(\varphi)$ among all smooth functions $C^\infty(M)$. That is, u is a critical point for E, and thus $\mathcal{A}(u) = 0$. □

Example 2.4. Let M be a round n-sphere S^n with a standard round metric. Also, let φ be the first nontrivial eigenfunction for the Laplacian so that

$$\Delta\varphi = -n\varphi, \quad \int_{S^n} |d\varphi|^2 = n\int_{S^n} \varphi^2.$$

Since $r_g = (n-1)g$, it is easy to see that $E(\varphi) = 0$. Thus the first eigenfunction φ realizes the infimum of the functional E and so

$$\mathcal{A}(\varphi) = 0 \quad \text{and} \quad \ker s_g'^* \neq 0.$$

On the other hand, consider $M = S^n \times S^{n+1}$ with the standard product metric. Then

$$s_g = 2n^2, \quad |r_g|^2 = n(2n^2 - n + 1), \tag{2-4}$$

and the first nonzero eigenvalue is given as

$$\lambda_1(M) = \lambda_1(S^n) = n.$$

Let φ be the first eigenfunction corresponding to $\lambda_1(M)$ so that

$$\Delta\varphi = -n\varphi, \quad r_g(d\varphi, d\varphi) = (n-1)|d\varphi|^2. \tag{2-5}$$

Substituting (2-4) and (2-5) into (2-2), we obtain $E(\varphi) = 0$. Therefore, we have $\mathcal{A}(\varphi) = 0$, and thus $\ker s_g'^* \neq 0$.

Recall that $H^2(M) = W^{2,2}(M)$ is the Sobolev space consisting of functions that are L^2 up to the second (weak) derivative. Let

$$\mathcal{W} = \left\{\varphi \in H^2(M) : \int_M \varphi^2 = 1\right\}$$

and define

$$\nu = \inf\left\{ \int_M \varphi \mathcal{A}(\varphi) : \varphi \in \mathcal{W} \right\}.$$

Note that $\nu \geq 0$, and $\ker \mathcal{A} \neq 0$ implies $\nu = 0$ by (2-3). The converse is also true.

Theorem 2.5. *Suppose that* $\nu = 0$. *Then* $\ker \mathcal{A} \neq 0$.

Proof. Since $\nu = 0$, there exists a sequence (φ_k) of functions in $H^2(M)$ with $\int_M \varphi_k^2 = 1$ such that

$$E(\varphi_k) \to 0 \quad \text{as } k \to \infty.$$

We now claim that (φ_k) is bounded in $H^2(M)$. On the contrary, suppose that the sequence (φ_k) is unbounded in $H^2(M)$. Defining $\widetilde{\varphi}_k$ as

$$\widetilde{\varphi}_k = \frac{\varphi_k}{\|\varphi_k\|_{H^2(M)}},$$

where $\|\varphi_k\|_{H^2(M)}$ denotes the Sobolev norm in $H^2(M)$, we have

$$\|\widetilde{\varphi}_k\|_{H^2(M)} = 1 \quad \text{and} \quad \int_M \widetilde{\varphi}_k^2 \to 0 \quad \text{as } k \to \infty.$$

Furthermore, $E(\widetilde{\varphi}_k) \to 0$ as $k \to \infty$. Thus the rescaled sequence $(\widetilde{\varphi}_k)$ is bounded in $H^2(M)$ and so $(\widetilde{\varphi}_k)$ converges weakly to a function $\widetilde{\varphi}_\infty \in H^2(M)$. Applying the Rellich–Kondrakov embedding theorem $H^2(M) \subset H^1(M) \subset L^2(M)$, $\widetilde{\varphi}_k$ converges strongly to $\widetilde{\varphi}_\infty$ in L^2, and thus, there exists a subsequence, say $(\widetilde{\varphi}_k)$, that converges almost everywhere. However, since $\|\widetilde{\varphi}_k\|_{L^2(M)} \to 0$, the limit function $\widetilde{\varphi}_\infty = 0$, which is contradictory to the fact that $\|d\widetilde{\varphi}\|_{L^2(M)} \neq 0$ or $\|Dd\widetilde{\varphi}\|_{L^2(M)} \neq 0$. Therefore, (φ_k) is bounded, and so φ_k converges weakly to a function φ in $H^2(M)$. By the Rellich–Kondrakov embedding theorem again, it is easy to see that φ_k converges strongly to φ in $L^2(M)$, and thus, there exists a subsequence, say (φ_k), that converges almost everywhere. Consequently, we have

$$E(\varphi) \leq \liminf_{k \to \infty} E(\varphi_k) = 0.$$

Hence since $E(\varphi) = 0$ and $\int_M \varphi^2 = 1$, φ is a nonconstant function and $\mathcal{A}(\varphi) = 0$. □

Corollary 2.6. *The invariant* ν *vanishes if and only if* $\ker s_g'^{*} \neq 0$ *or* $\ker \mathcal{A} \neq 0$.

Now we consider a special operator stemming from $\mathcal{A}$ that also plays a very important role in the kernel space of $s_g'^{*}$. For a function φ, define $P\varphi$ as

$$P\varphi = (n-1)\Delta^2\varphi + 2s_g\Delta\varphi - \langle Dd\varphi, r_g\rangle$$

and define

$$\mu = \inf_{\substack{\varphi \in H^2(M) \\ \varphi \neq 0}} \frac{\int \varphi P \varphi}{\int \varphi^2}.$$

Note that $\mu \leq 0$ since $P\varphi = 0$ when φ is a nonzero constant. Furthermore, it is easy to see that if $\mu = 0$, then either (M, g) is Ricci-flat or $\ker \mathscr{A} = 0$. In fact, if $u \in \ker \mathscr{A}$ and $r \neq 0$, then

$$\int_M uPu = -\int_M u^2 |r_g|^2 \leq 0.$$

Since $\mu = 0$ and $r_g \neq 0$, u must be zero because $\int_M u^2 |r_g|^2 = 0$. The following theorem shows that if $\ker \mathscr{A} \neq 0$, then μ must be nonpositive.

Theorem 2.7. *Assume that* $\ker \mathscr{A} \neq 0$ *and* $s = s_g$ *is constant. Then*

$$-\max_M |r_g|^2 \leq \mu \leq -\frac{s_g^2}{n}.$$

Proof. Let $u \in \ker \mathscr{A}$ be a nonconstant function and r be the Ricci tensor of the metric g. Since $s^2/n \leq |r|^2$, we have

$$\mu \int_M u^2 \leq \int_M uPu = -\int_M u^2 |r|^2 \leq -\frac{s^2}{n} \int_M u^2.$$

Thus

$$\mu \leq -\frac{s^2}{n}.$$

On the other hand, it follows from Lemma 2.1 that

$$\int_M (s_g'^* \varphi)^2 = \int_M \varphi \mathscr{A}(\varphi) = \int_M \{(n-1)(\Delta\varphi)^2 - 2s|d\varphi|^2 - \varphi\langle Dd\varphi, r\rangle + \varphi^2 |r|^2\}.$$

Thus,

$$\int_M \{(n-1)(\Delta\varphi)^2 - 2s|d\varphi|^2 - \varphi\langle Dd\varphi, r\rangle\} \geq -\int_M \varphi^2 |r|^2 \geq -\big(\max_M |r|^2\big) \int_M \varphi^2.$$

Therefore, since

$$\int_M \varphi P \varphi \geq -\big(\max_M |r|^2\big) \int_M \varphi^2$$

for any function φ, we conclude that

$$\mu \geq -\max_M |r|^2. \qquad \square$$

In view of Theorem 2.7, the invariant μ may designate a criteria for how close g is to an Einstein metric. In fact, when (M, g) is Einstein, it follows from Theorem 2.7 that, if $\ker \mathscr{A} \neq 0$,

$$\mu = -\frac{s^2}{n}.$$

In view of the operators $\mathscr{A}$ and P, for any real number α, we introduce an elliptic fourth-order partial differential operator $\mathscr{A}_\alpha$ defined by

$$\mathscr{A}_\alpha(\varphi) = (n-1)\Delta^2\varphi + 2s_g\Delta\varphi - \langle Dd\varphi, r_g\rangle + (1-\alpha)\varphi|r_g|^2,$$

where r_g is the Ricci tensor and s_g is the scalar curvature, which is assumed to be a positive constant. Note that $\mathscr{A}_0 = \mathscr{A}$ and $\mathscr{A}_1 = P$.

Theorem 2.8. *Assume that* $\ker\mathscr{A} = 0$ *and* $s = s_g$ *is constant. Then there exists a positive real number* $\alpha_0 > 0$ *such that* $\ker\mathscr{A}_\alpha = 0$ *for all* α, $0 \le \alpha \le \alpha_0$.

Proof. For $0 < \alpha < 1$, let $u \in \ker\mathscr{A}_\alpha$ be a nontrivial function. Then

$$\mathscr{A}(u) = \alpha u|r_g|^2 \le \big(\max_M |r_g|^2\big)\alpha u$$

and so $\nu \le \big(\max_M |r_g|^2\big)\alpha$. Since $\ker\mathscr{A} = 0$, Corollary 2.6 states that $\nu > 0$. Hence,

$$0 < \frac{\nu}{\max_M |r_g|^2} \le \alpha. \qquad \square$$

3. Case of $\nu > 0$

In this section, we consider the case in which ν is positive, or, equivalently, $\ker\mathscr{A} = 0$. We will investigate some necessary and sufficient conditions for ν to be positive and derive lower bounds on ν.

Lemma 3.1. *Assume* $\nu > 0$. *Then*

$$\inf_{\varphi\in\mathscr{W},\varphi\neq 1} \frac{E(\varphi)}{\|\varphi\|_{H^2(M)}} > 0.$$

Here $\|\varphi\|_{H^2(M)}$ *denotes the Sobolev norm in* $H^2(M)$.

Proof. Suppose that

$$\inf_{\varphi\in\mathscr{W},\varphi\neq 1} \frac{E(\varphi)}{\|\varphi\|_{H^2(M)}} = 0.$$

Then there exists a sequence (φ_k) in $\mathscr{W}$ such that $\|\varphi_k\|_{L^2(M)} = 1$ and

$$\frac{E(\varphi_k)}{\|\varphi_k\|_{H^2(M)}} \to 0 \quad \text{as } k \to \infty.$$

Since $\nu > 0$, we have $\|\varphi_k\|_{H^2(M)} \to \infty$ as $k \to \infty$. Defining $\widetilde{\varphi}_k$ as

$$\widetilde{\varphi}_k = \frac{\varphi_k}{\|\varphi_k\|_{H^2(M)}},$$

we can obtain a contradiction, as in the proof of Theorem 2.5. $\square$

Theorem 3.2. *Let (M, g) be a compact Riemannian n-manifold with positive constant scalar curvature s. If* $\ker \mathcal{A} = 0$, *then* $\nu > 0$ *is contained in the spectrum of* $\mathcal{A}$.

Proof. Recall that $\mathcal{W} = \{\varphi \in H^2(M) : \int_M \varphi^2 = 1\}$. Theorem 2.5 and Lemma 3.1 imply that

$$a := \inf_{\varphi \in \mathcal{W}, \varphi \neq 1} \frac{E(\varphi)}{\|\varphi\|_{H^2(M)}} > 0.$$

Then, for any function $\varphi \in \mathcal{W}$, we have $E(\varphi) \geq a\|\varphi\|_{H^2(M)}$, and thus,

$$E(\varphi) \to \infty \quad \text{as } \|\varphi\|_{H^2(M)} \to \infty.$$

In other words, the functional E is coercive on $\mathcal{W}$.

On the other hand, let (φ_k) be a sequence in $H^2(M)$ such that $\varphi_k \to \varphi$ weakly in $H^2(M)$. Then, according to the Rellich–Kondrakov theorem, $\varphi_k \to \varphi$ strongly in $L^2(M)$, and thus, a subsequence (φ_k) converges almost everywhere. This shows that the subspace $\mathcal{W}$ is weakly closed in $H^2(M)$. Furthermore, since M is compact, the subsequence (φ_k) uniformly converges to φ, and we obtain

$$E(\varphi) \leq \liminf_{k \to \infty} E(\varphi_k).$$

The functional E is bounded below and attains its minimum in $H^2(M)$ [Struwe 1990]. Letting

$$E(u) = \min\{E(\varphi) : \varphi \in \mathcal{W}\},$$

it is easy to see from the variational principle that

$$\mathcal{A}(u) = \nu u.$$ □

The properties of the operator $\mathcal{A}$ and the lower bound on ν are closely related to the first nonzero eigenvalue of the Laplacian. Let λ be the first nonzero eigenvalue of the Laplace operator Δ, which is characterized by

$$\lambda = \inf\left\{ \frac{\int_M |\nabla\varphi|^2}{\int_M \varphi^2} : \int_M \varphi = 0 \right\}.$$

It follows from the characterization of the first nonzero eigenvalue that, for any function φ with $\int_M \varphi = 0$,

$$\int_M \varphi^2 \leq \frac{1}{\lambda} \int_M |d\varphi|^2. \tag{3-1}$$

Lemma 3.3. *Let (M^n, g) be a compact Riemannian n-manifold. Then, for any function $\varphi \in C^\infty(M)$,*

$$\int_M |d\varphi|^2 \leq \frac{1}{\lambda} \int_M (\Delta\varphi)^2 \leq \frac{n}{\lambda} \int_M |Dd\varphi|^2, \tag{3-2}$$

where λ is the first nonzero eigenvalue of the Laplacian.

Proof. It follows from integration by parts and the Cauchy–Schwarz inequality that

$$\int_M |d\varphi|^2 \le \frac{1}{\lambda}\int_M (\Delta\varphi)^2.$$

The second inequality in (3-2) follows from the fact that $(\Delta\varphi)^2 \le n|Dd\varphi|^2$. □

Furthermore, for a function φ with $\int_M \varphi = 0$, we have

$$\int_M \varphi^2 \le \frac{1}{\lambda^2}\int_M (\Delta\varphi)^2 \quad \text{and} \quad \int_M \varphi^2 \le \frac{n}{\lambda^2}\int_M |Dd\varphi|^2.$$

A direct observation from the definition of $\mathcal{A}$ is the following theorem, which shows that if the first nonzero eigenvalue for the Laplacian is large compared to the sum of the scalar curvature and the norm of the Ricci tensor, then ν is positive.

Theorem 3.4. *Let (M^n, g) be a compact Riemannian n-manifold with positive constant scalar curvature s. If $(n-1)\lambda \ge 2s + \max_M |r_g|$, then $\nu \ge s^2/n$, and thus* $\ker \mathcal{A} = 0$, *or, equivalently*, $\ker s_g'^* = 0$.

Proof. Note that $|r_g|^2 \ge s^2/n$. It follows from Lemma 3.3 that

$$\int_M |d\varphi|^2 \le \frac{1}{\lambda}\int_M (\Delta\varphi)^2$$

for any function φ. Thus, for any function $\varphi \in \mathcal{W}$,

$$\begin{aligned} E(\varphi) &= \frac{1}{2}\int_M (n-1)(\Delta\varphi)^2 - 2s|d\varphi|^2 + r_g(d\varphi, d\varphi) + |r_g|^2\varphi^2 \\ &\ge \frac{1}{2}\{(n-1)\lambda - (2s + \max|r_g|)\}\int_M |d\varphi|^2 + \frac{s^2}{2n}\int_M \varphi^2. \end{aligned}$$

Hence, $\nu \ge s^2/n$. □

Remark 3.5. Assume $\nu > 0$ for a compact Riemannian n-manifold (M, g) with a positive constant scalar curvature. Then it follows from Theorem 3.2 that

$$\mathcal{A}(u) = \nu u$$

for some function $u \in \mathcal{W}$. In particular, we have

$$\int_M u|r|^2 = \nu \int_M u.$$

Since $\ker s_g'^* = 0$, by Theorem 2.2, there exists a unique function $\varphi \in C^\infty(M)$ such that $\mathcal{A}(\varphi) = u|r_g|^2$. Therefore,

$$\frac{s^2}{n} \le \int_M u^2|r_g|^2 = \int_M u\mathcal{A}\varphi = \int_M \varphi\mathcal{A}u = \nu\int_M \varphi u \le \nu\|\varphi\|_{L^2}.$$

On the other hand, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\nu\|\varphi\|_{L^2}^2 \le \int_M \varphi\mathscr{A}\varphi = \int_M \varphi u|r_g|^2 &\le \Big(\int_M \varphi^2|r_g|^2\Big)^{\frac{1}{2}}\Big(\int_M u^2|r_g|^2\Big)^{\frac{1}{2}}\\ &\le \big(\max_M |r_g|\big)\|\varphi\|_{L^2}\sqrt{\nu\|\varphi\|_{L^2}}.\end{aligned}$$

Therefore, we have $\nu\|\varphi\|_{L^2} \le \max_M |r_g|^2$, and so

$$\frac{s^2}{n} \le \nu\|\varphi\|_{L^2} \le \max_M |r_g|^2,$$

where φ is the function satisfying $\mathscr{A}(\varphi) = u|r_g|^2$.

Theorem 3.6. *Let $M = S^m \times S^m$ $(m \ge 2)$ with the standard product metric. Then*

$$\nu = m = \frac{\dim(M)}{2}.$$

Proof. First, we will examine the case $m = 2$ since key ingredients of the proof are contained in this setting. The cases of $m \ge 3$ will then be briefly explained.

For $M^4 = S^2 \times S^2$ with the standard product metric g, we obviously have $s_g = |r_g|^2 = 4$, $\lambda = 2$, and $r_g = g$. Thus, $\langle Dd\varphi, r\rangle = \Delta\varphi$ for any function φ, and so

$$\mathscr{A}(\varphi) = 3\Delta^2\varphi + 7\Delta\varphi + 4\varphi.$$

Let u be a first eigenfunction of S^2 so that $\Delta u = -2u$, $2\int_M u^2 = \int_M |du|^2$, and $r_g(du, du) = |du|^2$. Therefore,

$$\int_M u\mathscr{A}(u) = \int_M 3(\Delta u)^2 - 7|du|^2 + 4u^2 = 2\int_M u^2.$$

Hence $\nu \le 2$. To show the converse inequality $\nu \ge 2$, it is sufficient to prove that, for any C^∞ function φ,

$$F(\varphi) := \int_M \big[3(\Delta\varphi)^2 - 7|d\varphi|^2 + 2\varphi^2\big] \ge 0.$$

First, note that

$$F(\varphi) = \int_M (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi).$$

It follows from Lemma 3.3 that

$$2\int_M |d\varphi|^2 \le \int_M (\Delta\varphi)^2.$$

Thus, from the monotonicity of eigenvalues, it follows that, for any function φ that vanishes on the smooth boundary ∂D of a domain $D \subset M$, we have

$$2\int_D |d\varphi|^2 \le \int_D (\Delta\varphi)^2. \tag{3-3}$$

Assume for a moment that 0 is a regular value of φ. Let D_1 be a region on M such that

$$\Delta\varphi + 2\varphi \le 0 \quad \text{and} \quad \Delta\varphi + \tfrac{1}{3}\varphi \ge 0,$$

and D_2 be a region such that

$$\Delta\varphi + 2\varphi \ge 0 \quad \text{and} \quad \Delta\varphi + \tfrac{1}{3}\varphi \le 0.$$

Note that $\varphi \le 0$ on region D_1, and $\varphi \ge 0$ on region D_2. Thus, $\partial D_1 = \partial D_2$. On region D_1, we have

$$0 < -\tfrac{1}{3}\varphi \le \Delta\varphi \le -2\varphi. \tag{3-4}$$

Multiplying (3-4) by φ and integrating over D_1, we obtain

$$-2\int_{D_1} \varphi^2 \le \int_{D_1} \varphi\Delta\varphi \le -\frac{1}{3}\int_{D_1} \varphi^2.$$

Since $\varphi = 0$ on ∂D_1, we get

$$-2\int_{D_1} \varphi^2 \le -\int_{D_1} |d\varphi|^2 \le -\frac{1}{3}\int_{D_1} \varphi^2. \tag{3-5}$$

Similarly, on region D_2, we have

$$-2\int_{D_2} \varphi^2 \le \int_{D_2} \varphi\Delta\varphi \le -\frac{1}{3}\int_{D_2} \varphi^2. \tag{3-6}$$

Let $D = D_1 \cup D_2$. It follows from (3-5) and (3-6) that

$$\frac{1}{3}\int_D \varphi^2 \le \int_D |d\varphi|^2 \le 2\int_D \varphi^2. \tag{3-7}$$

Note that on $M - D$, we have

$$(\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi) \ge 0. \tag{3-8}$$

Furthermore, since the function φ vanishes on the boundary ∂D of D, we can apply integration by parts and Green's identity. Thus, it follows from (3-3), (3-7), and

(3-8) that

$$\begin{aligned}F(\varphi) &= \int_D (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi) + \int_{M-D} (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi) \\ &= 3\int_D \big((\Delta\varphi)^2 - 2|d\varphi|^2\big) + \int_D \big(2\varphi^2 - |d\varphi|^2\big) + \int_{M-D} (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi) \\ &\geq 0.\end{aligned}$$

Now, assume that 0 is a critical value of φ. By Sard's theorem, for any positive real number $\epsilon > 0$, there exists a real number a, $-\epsilon < a < 0$, such that a is a regular value of φ. Let $D_{1,a}$ be a region such that

$$\Delta\varphi + 2\varphi \leq \tfrac{5}{3}a \quad \text{and} \quad \Delta\varphi + \tfrac{1}{3}\varphi \geq 0. \tag{3-9}$$

Note that $\varphi \leq a < 0$ on region $D_{1,a}$, and $\varphi = a$ on the boundary $\partial D_{1,a}$. Multiplying (3-9) by φ and integrating it over $D_{1,a}$, we obtain

$$\tfrac{5}{3}a \int_{D_{1,a}} \varphi - a \int_{\partial D_{1,a}} \frac{\partial\varphi}{\partial n_1} \leq \int_{D_{1,a}} \big(2\varphi^2 - |d\varphi|^2\big), \tag{3-10}$$

where n_1 is the outward-pointing unit normal vector field to $\partial D_{1,a}$. Next, let $D_{2,a}$ be a region such that

$$\Delta\varphi + 2\varphi \geq -\tfrac{5}{3}a \quad \text{and} \quad \Delta\varphi + \tfrac{1}{3}\varphi \leq 0. \tag{3-11}$$

We may assume that $-a$ is also a regular value of φ. Note that $0 < -a \leq \varphi$ on region $D_{2,a}$, and $\varphi = -a$ on the boundary $\partial D_{2,a}$. Multiplying (3-11) by φ and integrating it over $D_{2,a}$, we obtain

$$a \int_{\partial D_{2,a}} \frac{\partial\varphi}{\partial n_2} - \tfrac{5}{3}a \int_{D_{2,a}} \varphi \leq \int_{D_{2,a}} \big(2\varphi^2 - |d\varphi|^2\big), \tag{3-12}$$

where n_2 is a unit normal vector field on $\partial D_{2,a}$. Decomposing M into three regions, we can write

$$\begin{aligned}F(\varphi) = 3\int_{D_{1,a}} \big[(\Delta\varphi)^2 - 2|d\varphi|^2\big] + \int_{D_{1,a}} \big(2\varphi^2 - |d\varphi|^2\big) + 3\int_{D_{2,a}} \big[(\Delta\varphi)^2 - 2|d\varphi|^2\big] \\ + \int_{D_{2,a}} \big(2\varphi^2 - |d\varphi|^2\big) + \int_{M-(D_{1,a}\cup D_{2,a})} (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi).\end{aligned}$$

Applying inequality (3-3) to $\varphi - a$, we have

$$\int_{D_{1,a}} \big[(\Delta\varphi)^2 - 2|d\varphi|^2\big] \geq 0 \quad \text{and} \quad \int_{D_{2,a}} \big[(\Delta\varphi)^2 - 2|d\varphi|^2\big] \geq 0.$$

Thus, from (3-10) and (3-12), we obtain

$$F(\varphi) \geq \tfrac{5}{3}|a| \int_{D_{1,a} \cup D_{2,a}} |\varphi| - a \int_{\partial D_{1,a}} \frac{\partial \varphi}{\partial n_1} + a \int_{\partial D_{2,a}} \frac{\partial \varphi}{\partial n_2} + \int_{M-(D_{1,a} \cup D_{2,a})} (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi).$$

Since $|\partial\varphi/\partial n_1| \leq |d\varphi|$ and $|\partial\varphi/\partial n_2| \leq |d\varphi|$, the first three terms on the right-hand side tend to 0 as $\epsilon \to 0$. Finally, let $E_{1,a}$ be a region such that $\Delta\varphi + 2\varphi > \frac{5}{3}a$ and $\Delta\varphi + \frac{1}{3}\varphi \geq 0$, and $E_{2,a}$ be a region such that $\Delta\varphi + 2\varphi < -\frac{5}{3}a$ and $\Delta\varphi + \frac{1}{3}\varphi \leq 0$. Then we have

$$\int_{M-(D_{1,a} \cup D_{2,a})} (\Delta\varphi + 2\varphi)(3\Delta\varphi + \varphi) \geq \tfrac{5}{3}a \int_{E_{1,a}} (3\Delta\varphi + \varphi) - \tfrac{5}{3}a \int_{E_{2,a}} (3\Delta\varphi + \varphi).$$

The right-hand side tends to 0 as $\epsilon \to 0$. Hence, $F(\varphi) \geq 0$.

In the general case, $M^{2m} = S^m \times S^m$ when $m \geq 2$, it is easy to see that

$$s_g = 2m(m-1), \quad |r_g|^2 = 2m(m-1)^2, \quad r_g = (m-1)g, \quad \lambda = m.$$

Thus,

$$\int_M \varphi \mathscr{A}(\varphi) = (2m-1) \int_M \big[(\Delta\varphi)^2 - m|d\varphi|^2\big] - (2m^2 - 4m + 1) \int_M |d\varphi|^2 + 2m(m-1)^2 \int_M \varphi^2.$$

Using a first eigenfunction u of S^m, $\Delta u = -mu$, we can demonstrate that $\nu \leq m$. To show that $\nu \geq m$, it is sufficient to prove that, for any function φ,

$$F(\varphi) := \int_M (\Delta\varphi + m\varphi)\big[(2m-1)\Delta\varphi + (2m^2 - 4m + 1)\varphi\big] \geq 0.$$

Note that

$$m \int_M |d\varphi|^2 \leq \int_M (\Delta\varphi)^2.$$

An argument identical to that used in the case $S^2 \times S^2$ shows that $F(\varphi) \geq 0$, and thus, $\nu = m$. □

Remark 3.7. For the case of $M = S^m \times S^{m+k}$ with $k \geq 2$, the first nonzero eigenfunction of S^m can be used to show that

$$\nu \leq \min\big\{(m+k)(k-1)^2,\, m(k+1)^2\big\}.$$

However, we do not know the exact value of ν.

4. The first eigenvalue of the Laplacian

As mentioned above, the first nonzero eigenvalue $\lambda = \lambda_1(M)$ of the Laplace operator for a Riemannian manifold (M, g) is related to the operator $\mathcal{A}$. For example, if $\ker \mathcal{A} \neq 0$ and g is an Einstein metric with positive scalar curvature, then $\lambda = s/(n-1)$, from the results obtained in [Berger et al. 1971] and [Bourguignon 1975]. We shall now see that, if there is a nontrivial function on which the action of $\mathcal{A}$ is nonpositive where the function is positive, then the first nonzero eigenvalue of the Laplacian is bounded above, and vice versa. Recall that we assumed that the scalar curvature $s_g = s$ of a Riemannian manifold (M, g) is always a positive constant.

For a function φ on a smooth manifold M, let us define

$$M_\varphi^+ = \{x \in M : \varphi(x) > 0\}.$$

We say that a Riemannian manifold (M, g) satisfies the *$\mathcal{A}$-superharmonic condition* if there exists a smooth function φ such that

(i) $M_\varphi^+ \neq \varnothing$ and $\mathcal{A}\varphi \leq 0$ on M_φ^+;

(ii) $\Delta\varphi = 0$ on the boundary ∂M_φ^+ of M_φ^+.

For example, if $M = S^n$ with the standard round metric g_0, and φ is the first nonzero eigenfunction of the Laplacian, i.e., $\Delta\varphi = -n\varphi$, then $\mathcal{A}\varphi = 0$ and (S^n, g_0) satisfies the $\mathcal{A}$-superharmonic condition. Furthermore, note that any eigenfunction of the Laplacian satisfies condition (ii). The following lemma shows that the $\mathcal{A}$-superharmonic condition is implied by $\ker \mathcal{A} \neq 0$.

Lemma 4.1. *Let (M^n, g) be a compact n-dimensional Riemannian manifold with a positive constant scalar curvature s_g. If* $\ker \mathcal{A} \neq 0$, *then (M, g) satisfies the $\mathcal{A}$-superharmonic condition.*

Proof. By Lemma 2.3, $\ker \mathcal{A} \neq 0$ is equivalent to $\ker s_g'^* \neq 0$. Let $s_g'^*\varphi = 0$ and $\varphi \neq 0$. Then

$$Dd\varphi - (\Delta\varphi)g - \varphi r_g = 0.$$

In particular, taking the trace yields

$$\Delta\varphi = -\frac{s_g}{n-1}\varphi,$$

and so $M_\varphi^+ \neq \varnothing$. Since $\mathcal{A}\varphi = 0$, the function φ satisfies conditions (i) and (ii) in the definition of the $\mathcal{A}$-superharmonic condition. □

Theorem 4.2. *Let (M^n, g) be a compact n-dimensional Riemannian manifold with a positive constant scalar curvature s_g. Suppose that (M, g) satisfies the*

$\mathscr{A}$-superharmonic condition. If $\mathrm{Ric}_g \ge k \ge 0$, *then the first nonzero eigenvalue* λ_1 *of the Laplacian satisfies*

$$\lambda_1 \le \frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n}}{2(n-1)}. \tag{4-1}$$

Proof. Let $s_g = s$ and $\mathrm{Ric}_g = r_g = r$. In addition, let φ be a smooth function satisfying $M_\varphi^+ \ne \varnothing$, $\mathscr{A}\varphi \le 0$ on M_φ^+ and $\Delta\varphi = 0$ on the boundary ∂M_φ^+. If φ is a constant function, then φ is a positive constant since $M_\varphi^+ \ne \varnothing$. However, we have $0 \ge \mathscr{A}\varphi = \varphi|r|^2$, which is a contradiction. Thus, we may assume that φ is a nonconstant function. By the above hypothesis, we have

$$\int_{M_\varphi^+} \varphi\mathscr{A}\varphi \le 0. \tag{4-2}$$

By the definition of $\mathscr{A}$ and integration by parts, together with the fact that $\Delta\varphi = 0$ on ∂M_φ^+, we obtain

$$\begin{aligned} \int_{M_\varphi^+} \varphi\mathscr{A}\varphi &= \int_{M_\varphi^+} (n-1)(\Delta\varphi)^2 - \int_{\partial M_\varphi^+} \Delta\varphi \frac{\partial\varphi}{\partial\nu} \\ &\qquad + \int_{M_\varphi^+} \big[2s\varphi\Delta\varphi + \varphi\langle Dd\varphi, r\rangle + |r|^2\varphi^2\big] \\ &\ge \int_{M_\varphi^+} \Big[(n-1)(\Delta\varphi)^2 + (2s-k)\varphi\Delta\varphi + \frac{s^2}{n}\varphi^2\Big]. \end{aligned} \tag{4-3}$$

Note that

$$(n-1)(\Delta\varphi)^2 + (2s-k)\varphi\Delta\varphi + \frac{s^2}{n}\varphi^2 = ((n-1)\Delta\varphi + \alpha\varphi)(\Delta\varphi + \beta\varphi), \tag{4-4}$$

where

$$\alpha = \frac{2s - k + \sqrt{k^2 - 4ks + 4s^2/n}}{2}, \quad \beta = \frac{2s - k - \sqrt{k^2 - 4ks + 4s^2/n}}{2(n-1)}.$$

Observe that $k^2 - 4ks + 4s^2/n > 0$ if and only if either

$$k < 2\left(1 - \sqrt{1 - \frac{1}{n}}\right) s_g \quad \text{or} \quad k > 2\left(1 + \sqrt{1 - \frac{1}{n}}\right) s_g,$$

and the first inequality always holds.

Claim. *If*

$$\lambda_1 > \frac{2s - k + \sqrt{k^2 - 4ks + 4s^2/n}}{2(n-1)} = \frac{\alpha}{n-1}, \tag{4-5}$$

then any subset Ω *of* M_φ^+ *with* C^1 *boundary on which* $(n-1)\Delta\varphi + \alpha\varphi \ge 0$ *and* $\Delta\varphi + \beta\varphi \le 0$ *has a measure of zero.*

Proof. Suppose that a subset Ω of M_φ^+ contains an open n-ball. Note that since $\Delta\varphi = \varphi = 0$ on $\partial\Omega$, we can apply the Dirichlet principle on the first nonzero eigenvalue of the Laplacian. By monotonicity, we have

$$\lambda_1 = \lambda_1(M) \le \lambda_1(\Omega).$$

Since $(n-1)\Delta\varphi + \alpha\varphi \ge 0$ and $\varphi > 0$ on Ω, we have

$$\varphi\Delta\varphi \ge -\frac{\alpha}{n-1}\varphi^2.$$

Integrating this over Ω, we obtain

$$\int_\Omega |d\varphi|^2 \le \frac{\alpha}{n-1}\int_\Omega \varphi^2 \le \frac{\alpha}{n-1}\cdot\frac{1}{\lambda_1(\Omega)}\int_\Omega |d\varphi|^2.$$

Thus,

$$1 \le \frac{\alpha}{n-1}\cdot\frac{1}{\lambda_1(\Omega)},$$

and so

$$\lambda_1 \le \lambda_1(\Omega) \le \frac{\alpha}{n-1},$$

which contradicts (4-5). This completes the proof of the claim. □

Now, suppose that $\lambda_1 > \frac{\alpha}{n-1}$. Since $\alpha > (n-1)\beta$, it follows from (4-4) and the above claim that

$$(n-1)(\Delta\varphi)^2 + (2s-k)\varphi\Delta\varphi + \frac{s^2}{n}\varphi^2 \ge 0 \quad \text{a.e. on } M_\varphi^+,$$

which implies that $\int_{M_\varphi^+} \varphi\mathscr{A}\varphi \ge 0$. Consequently, from (4-2), we have

$$\int_{M_\varphi^+} \varphi\mathscr{A}\varphi = 0.$$

Thus, on the set M_φ^+, we have $\mathscr{A}\varphi = 0$ and

$$(n-1)(\Delta\varphi)^2 + (2s-k)\varphi\Delta\varphi + \frac{s^2}{n}\varphi^2 = ((n-1)\Delta\varphi + \alpha\varphi)(\Delta\varphi + \beta\varphi) = 0$$

by (4-3). Since $\alpha > (n-1)\beta$, either $(n-1)\Delta\varphi + \alpha\varphi = 0$ or $\Delta\varphi + \beta\varphi = 0$ on the entire set M_φ^+. Therefore, we obtain

$$\lambda_1 \le \lambda_1(M_\varphi^+) \le \max\left\{\frac{\alpha}{n-1}, \beta\right\} = \frac{\alpha}{n-1},$$

which contradicts the assumption $\lambda_1 > \frac{\alpha}{n-1}$. Hence,

$$\lambda_1 = \lambda_1(M) \le \frac{\alpha}{n-1}.$$

This completes the proof of Theorem 4.2. □

Remark 4.3. If $M = S^n$ with the standard round metric, then taking $k = n-1$, the right-hand side in inequality (4-1) becomes

$$\frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n}}{2(n-1)} = n,$$

and so the result in Theorem 4.2 is optimal.

In fact, in the case $\mathrm{Ric}_g \geq k$ and $s_g = nk$ — corresponding to the assumption that g is Einstein — the conclusion of Theorem 4.2 is that $\lambda_1 \leq \frac{nk}{n-1}$. Thus, by the Lichnerowicz–Obata theorem [Lichnerowicz 1958; Obata 1962; Berger et al. 1971], the only Einstein metric with positive constant scalar curvature which is $\mathcal{A}$-superharmonic is the standard metric on the sphere. This fact also shows that the assumption $\ker \mathcal{A} \neq 0$ cannot be removed from Lemma 4.1.

Remark 4.4. Let (M^n, g) be a compact n-dimensional Riemannian manifold such that $\mathrm{Ric}_g \geq k \geq 0$, where the scalar curvature s_g is a positive constant. In addition, suppose that there exists a function φ such that $M_\varphi^- = \{x \in M : \varphi(x) < 0\} \neq \varnothing$ and $\mathcal{A}\varphi \geq 0$ on M_φ^-. Then, by simply applying Theorem 4.2 to the function $\bar{\varphi} = -\varphi$, we can see that the first nonzero eigenvalue λ_1 of the Laplacian satisfies

$$\lambda_1 \leq \frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n}}{2(n-1)}.$$

In particular, if $k = 0$, then

$$\lambda_1 \leq \frac{s}{n-1}\left(1 + \frac{1}{\sqrt{n}}\right).$$

Finally, we consider the relationship of ν to the first nonzero eigenvalue of the Laplace operator. In the case of $\nu > 0$, it follows from Theorem 3.2 that a minimizer u for the functional E satisfies $\mathcal{A}u = \nu u$. In particular, since $\ker s_g'^* = 0$ when $\nu > 0$, we cannot, in general, expect that $s_g/(n-1)$ is contained in the spectrum of the Laplace operator.

Theorem 4.5. *Let (M, g) be a compact n-dimensional Riemannian manifold such that $\mathrm{Ric}_g \geq k \geq 0$ and assume that $\nu > s_g^2/n$, where the scalar curvature s_g is a positive constant. In addition, suppose that $M_u^+ \neq \varnothing$ for a function u satisfying $\mathcal{A}u = \nu u$. Then the first nonzero eigenvalue of the Laplacian satisfies*

$$\lambda_1(M) \leq \frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n + 4(n-1)\nu}}{2(n-1)},$$

unless (M, g) is Einstein.

Proof. We shall denote s_g by s and $\mathrm{Ric}_g = r_g$ by r. From $\int_M u\mathscr{A}u = \nu \int_M u^2$,

$$0 = \int_M u\mathscr{A}u - \nu u^2 = \int_M (n-1)(\Delta u)^2 + 2su\Delta u + r(du, du) + (|r|^2 - \nu)u^2$$
$$> \int_M (n-1)(\Delta u)^2 + (2s-k)u\Delta u + \Big(\frac{s^2}{n} - \nu\Big)u^2.$$

The third inequality is strict since (M, g) is not Einstein. We may factor the integrand as follows:

$$(n-1)(\Delta u)^2 + (2s-k)u\Delta u + \Big(\frac{s^2}{n} - \nu\Big)u^2 = ((n-1)\Delta u + \alpha u)\Big(\Delta u + \frac{\beta}{n-1}u\Big),$$

where

$$\alpha = \frac{1}{2}\bigg(2s - k + \sqrt{k^2 - 4ks + \frac{4s^2}{n} + 4(n-1)\nu}\,\bigg),$$
$$\beta = \frac{1}{2}\bigg(2s - k - \sqrt{k^2 - 4ks + \frac{4s^2}{n} + 4(n-1)\nu}\,\bigg).$$

Note that if $\nu > s^2/n$, the radicand is positive for any $k \geq 0$.

The remainder of the proof is similar to that of Theorem 4.2. Hence, if g is not an Einstein metric and $\lambda > \frac{\alpha}{n-1}$, then

$$0 \geq \int_M u\mathscr{A}u - \nu u^2 > \int_M (n-1)(\Delta u)^2 + (2s-k)u\Delta u + \Big(\frac{s^2}{n} - \nu\Big)u^2 \geq 0,$$

which is a contradiction. □

Theorem 4.6. *Let (M, g) be a compact n-dimensional Riemannian manifold such that* $\mathrm{Ric}_g \geq k$ *with*

$$0 \leq k \leq 2s_g\bigg(1 - \sqrt{1 - \frac{1}{n} - (n-1)\frac{\nu}{s_g^2}}\,\bigg). \tag{4-6}$$

Suppose that $0 < \nu \leq s_g^2/n$. In addition, assume that $M_u^+ \neq \varnothing$ for a function u satisfying $\mathscr{A}u = \nu u$. Then the first nonzero eigenvalue λ_1 of the Laplacian satisfies

$$\lambda_1 \leq \frac{2s_g - k + \sqrt{k^2 - 4ks_g + 4s_g^2/n + 4(n-1)\nu}}{2(n-1)},$$

unless (M, g) is Einstein.

Proof. Note that if $\nu \leq s_g^2/n$ and (4-6) is satisfied,

$$k^2 - 4ks_g + \frac{4s_g^2}{n} + 4(n-1)\nu \geq 0.$$

The remainder of the proof proceeds in the same manner as that of Theorem 4.5. □

Acknowledgement

The authors would like to express their gratitude to the referee for several valuable comments, which allowed us to express some arguments more clearly and finely.

References

[Berger and Ebin 1969] M. Berger and D. Ebin, "Some decompositions of the space of symmetric tensors on a Riemannian manifold", *J. Differential Geometry* **3** (1969), 379–392. MR 42 #993 Zbl 0194.53103

[Berger et al. 1971] M. Berger, P. Gauduchon, and E. Mazet, *Le spectre d'une variété riemannienne*, Lecture Notes in Mathematics **194**, Springer, Berlin, 1971. MR 43 #8025 Zbl 0223.53034

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Math. (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Bourguignon 1975] J.-P. Bourguignon, "Une stratification de l'espace des structures riemanniennes", *Compositio Math.* **30** (1975), 1–41. MR 54 #6189 Zbl 0301.58015

[Fischer and Marsden 1974] A. E. Fischer and J. E. Marsden, "Manifolds of Riemannian metrics with prescribed scalar curvature", *Bull. Amer. Math. Soc.* **80** (1974), 479–484. MR 49 #11561 Zbl 0288.53040

[Lichnerowicz 1958] A. Lichnerowicz, *Géométrie des groupes de transformations*, Travaux et Recherches Mathématiques **3**, Dunod, Paris, 1958. MR 23 #A1329 Zbl 0096.16001

[Obata 1962] M. Obata, "Certain conditions for a Riemannian manifold to be isometric with a sphere", *J. Math. Soc. Japan* **14** (1962), 333–340. MR 25 #5479 Zbl 0115.39302

[Struwe 1990] M. Struwe, *Variational methods: Applications to nonlinear partial differential equations and Hamiltonian systems*, Springer, Berlin, 1990. MR 92b:49002 Zbl 0746.49010

Received December 2, 2011. Revised September 26, 2012.

SEUNGSU HWANG
DEPARTMENT OF MATHEMATICS
CHUNG-ANG UNIVERSITY
SEOUL 156-756
SOUTH KOREA
seungsu@cau.ac.kr

JEONGWOOK CHANG
DEPARTMENT OF MATHEMATICS EDUCATION
DANKOOK UNIVERSITY
GYEONG-GI 448-701
SOUTH KOREA
jchang@dankook.ac.kr

GABJIN YUN
DEPARTMENT OF MATHEMATICS
MYONG JI UNIVERSITY
GYEONG-GI 449-728
SOUTH KOREA
gabjin@mju.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.417

FILL-INS OF NONNEGATIVE SCALAR CURVATURE, STATIC METRICS, AND QUASI-LOCAL MASS

JEFFREY L. JAUREGUI

Consider a triple of "Bartnik data" (Σ, γ, H), where Σ is a topological 2-sphere with Riemannian metric γ and positive function H. We view Bartnik data as a boundary condition for the problem of finding a compact Riemannian 3-manifold (Ω, g) of nonnegative scalar curvature whose boundary is isometric to (Σ, γ) with mean curvature H. Considering the perturbed data $(\Sigma, \gamma, \lambda H)$ for a positive real parameter λ, we find that such a "fill-in" (Ω, g) must exist for λ small and cannot exist for λ large; moreover, we prove there exists an intermediate threshold value.

The main application is the construction of a new quasi-local mass, a concept of interest in general relativity. This mass has a nonnegativity property and is bounded above by the Brown–York mass. However, our definition differs from many others in that it tends to vanish on static vacuum (as opposed to flat) regions. We also recognize this mass as a special case of a type of twisted product of quasi-local mass functionals.

1. Introduction

Riemannian 3-manifolds of nonnegative scalar curvature arise naturally in general relativity as totally geodesic spacelike submanifolds of spacetimes obeying Einstein's equation and the dominant energy condition. In this setting, scalar curvature plays the role of energy density. Black holes in this setting are manifested as connected minimal surfaces that minimize area to the outside. If S is a disjoint union of such surfaces of total area A, the number $\sqrt{A/16\pi}$ is interpreted to encode the total mass of the collection of black holes, possibly accounting for potential energy between them [Bray 2001].

A fundamental question in general relativity is to quantify how much mass is contained in a compact region Ω in a spacelike slice of a spacetime [Penrose 1982]. Constructing examples of such *quasi-local mass* has led to a very active field of research (we mention here a small number of possible references: [Szabados 2009; Wang and Yau 2009; Huisken and Ilmanen 2001]). For most definitions,

MSC2010: primary 53C20; secondary 83C99.
Keywords: general relativity, quasi-local mass, scalar curvature, static vacuum.

the quasi-local mass of Ω depends only boundary data of Ω: namely the induced 2-metric and induced mean curvature function. We reference pioneering work of Bartnik [1997; 1989], whose name is given in the following definition.

All metrics and functions in this paper are assumed to be smooth, unless otherwise stated.

Definition 1. A triple $\mathcal{B} = (\Sigma, \gamma, H)$, where Σ is a topological 2-sphere, γ is a Riemannian metric on Σ of positive Gauss curvature, and H is a positive function on Σ is called *Bartnik data*.

While not always necessary, it is often customary to restrict to positive Gauss curvature and positive functions H, as we do here. A typical problem involving Bartnik data (Σ, γ, H) is to construct a Riemannian 3-manifold (M, g) satisfying some nice geometric properties such that the boundary ∂M is isometric to (Σ, γ), and the mean curvature of ∂M agrees with H. For instance, one might require (M, g) to be asymptotically flat with nonnegative or zero scalar curvature (see [Bartnik 1993], for instance). Such a manifold is called an *extension* of the Bartnik data.

We focus on the dual problem of constructing compact *fill-ins* of the Bartnik data, realizing (Σ, γ, H) as the boundary of a compact 3-manifold. This problem was considered by Bray in the construction of the Bartnik inner mass [Bray 2001] (see Section 2.3 below).

Definition 2. A *fill-in* of Bartnik data (Σ, γ, H) is a compact, connected Riemannian 3-manifold (Ω, g) with boundary such that there exists an isometric embedding $\iota : (\Sigma, \gamma) \to (\Omega, g)$ with the following properties:

(1) the image $\iota(\Sigma)$ is some connected component S_0 of $\partial\Omega$, and

(2) $H = H_{S_0} \circ \iota$ on Σ, where H_{S_0} is the mean curvature of S_0 in (Ω, g).

We adopt the sign convention that the mean curvature equals $-g(\vec{H}, \vec{n})$, where $\vec{H}$ is the mean curvature vector and $\vec{n}$ is the unit normal pointing out of Ω (e.g., the boundary of a ball in $\mathbb{R}^n$ has positive mean curvature).

Without loss of generality, if (Ω, g) is a fill-in of (Σ, γ, H), we shall henceforth identify Σ with $\iota(\Sigma)$ and H with the mean curvature of $\iota(\Sigma)$.

We will primarily be concerned with fill-ins satisfying the following geometric constraints.

Definition 3. A fill-in (Ω, g) of (Σ, γ, H) is *valid* if the metric g has nonnegative scalar curvature and either

(1) $\partial\Omega = \Sigma$, or

(2) $\partial\Omega \setminus \Sigma$ is a minimal (zero mean curvature) surface, possibly disconnected.

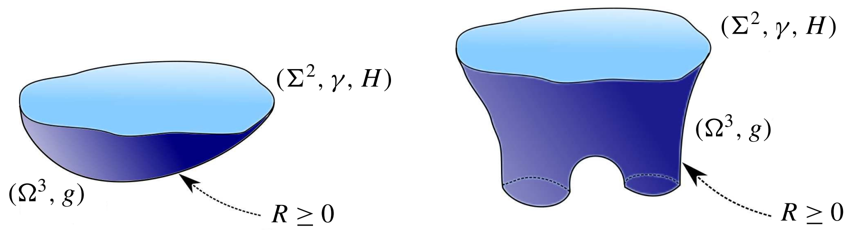


Figure 1. Left: a valid fill-in of (Σ, γ, H) of the first type (i.e., $\partial\Omega = \Sigma$). Right: a valid fill-in of the second type ($\partial\Omega \setminus \Sigma$ is minimal). R denotes the scalar curvature of g.

Figure 1 provides a graphical depiction. In physical terms, a valid fill-in is a compact region in a slice of a spacetime that has nonnegative energy density and possibly contains black holes. Another characterization of the second class of valid fill-ins is a cobordism of nonnegative scalar curvature that joins the given Bartnik data to a minimal surface. Note that we require $\partial\Omega \setminus \Sigma$ to be minimal, but not necessarily area-minimizing.

Interestingly, Bartnik data falls into one of three types. Although trivial to prove, the following fact motivates much of the present paper.

Observation 4 (trichotomy of Bartnik data). Bartnik data (Σ, γ, H) belongs to exactly one of the following three classes:

(1) Negative type: (Σ, γ, H) admits no valid fill-in.

(2) Zero type: (Σ, γ, H) admits a valid fill-in, but every valid fill-in (Ω, g) has $\partial\Omega = \Sigma$.

(3) Positive type: (Σ, γ, H) admits a valid fill-in (Ω, g) with nonempty minimal boundary $\partial\Omega \setminus \Sigma$.

Outline. In Section 2, we give some geometric characterizations of valid fill-ins of Bartnik data of zero and positive type, making connections with static vacuum metrics. We also recall in Section 2.3 the Bartnik inner mass, which explains the use of the words positive, zero, and negative in the trichotomy.

The essential idea of this paper, presented in Section 3, is to study the behavior of Bartnik data $(\Sigma, \gamma, \lambda H)$, where the real parameter $\lambda > 0$ is allowed to vary. We show in Theorem 11, the main result, that the data passes through all three classes of the trichotomy, with interesting behavior at some unique borderline value $\lambda = \lambda_0$. In Section 3.1, we introduce a function that probes the geometry of valid fill-ins of $(\Sigma, \gamma, \lambda H)$.

The main application occurs in Section 4, where we use the number λ_0 to define a quasi-local mass for regions in 3-manifolds of nonnegative scalar curvature (Definition 14). Several properties are shown to hold, including nonnegativity. What distinguishes this definition from most others is its tendency to vanish on static vacuum, as opposed to flat, data. We give a brief physical argument for why such a property may be desirable in Section 4.1.

Section 5 consists of examples of Bartnik data of all three types, and compares our definition with the Hawking mass and Brown–York mass. In Section 6 we introduce a general construction for "twisting" two quasi-local mass functionals together, of which the above quasi-local mass is a special case. The final section is a discussion of some potentially interesting open problems.

2. Fill-ins of nonnegative type and the inner mass

2.1. ***Zero type data and static vacuum metrics.*** First, we classify the geometry of valid fill-ins of Bartnik data of zero type. Recall that a Riemannian 3-manifold (Ω, g) is *static vacuum* if there exists a function $u \geq 0$ (called the *static potential*), with $u > 0$ on the interior of Ω, such that the Lorentzian metric

$$h = -u^2 dt^2 + g$$

on $\mathbb{R} \times \text{int}(\Omega)$ has zero Ricci curvature. This condition is equivalent to the system of equations

$$\Delta u = 0, \tag{1}$$

$$u \,\text{Ric} = \text{Hess}\, u, \tag{2}$$

where Δ, Ric and Hess are the Laplacian, Ricci curvature, and Hessian with respect to g. Equation (1) together with the trace of (2) shows that static vacuum metrics have zero scalar curvature. The following result is primarily a consequence of Corvino's work [2000] on local scalar curvature deformation.

Proposition 5. *If $\mathcal{B}$ is Bartnik data of zero type, then any valid fill-in is static vacuum.*

The idea of the proof is to use a valid fill-in that is not static vacuum to construct a valid fill-in that contains a black hole. By a very rough analogy, one might think of this physically as taking some of the energy content in a fill-in and squeezing it down into a black hole. The delicate issue is that we must preserve the boundary data in the process.

Proof. Let (Ω, g) be a valid fill-in of zero type data (Σ, γ, H). By definition, $\partial\Omega = \Sigma$.

We claim g has identically zero scalar curvature. If not, there exists $p \in \text{int}(\Omega)$ and $r > 0$ such that on the closed metric ball $\overline{B}(p, r)$, the scalar curvature of g is bounded below by some $\epsilon > 0$. On the set $\overline{B}(p, r/2) \setminus \{p\}$, let G be a Green's function for the Laplacian that blows up at p and vanishes on $\partial \overline{B}(p, r/2)$ (see

[Aubin 1998, Theorem 4.17]). By the maximum principle, G is positive, except on $\partial \bar{B}(p, r/2)$. Extend G by zero to the rest of $\Omega \setminus \{p\}$, so that G Lipschitz, smooth away from $\partial \bar{B}(p, r/2)$. Perturb G to a smooth, nonnegative function $\tilde{G}$ on $\Omega \setminus \{p\}$ that agrees with G except possibly on the annular region $\bar{B}(p, 3r/4) \setminus \bar{B}(p, r/4)$. For a parameter $\delta > 0$ to be determined, define the conformal metric

$$\tilde{g} = (1 + \delta \tilde{G})^4 g$$

on $\Omega \setminus \{p\}$. By construction, $\tilde{g} = g$ outside $\bar{B}(p, r)$ and thus has nonnegative scalar curvature outside this ball. For points inside $\bar{B}(p, r)$, we apply the rule for the change in scalar curvature under a conformal deformation (see Appendix A):

$$\begin{aligned} \tilde{R} &= (1 + \delta \tilde{G})^{-5}(-8\Delta(1 + \delta \tilde{G}) + (1 + \delta \tilde{G})R) \\ &\geq (1 + \delta \tilde{G})^{-5}(-8\delta \Delta \tilde{G} + \epsilon), \end{aligned}$$

Here, $\tilde{R}$ and R are the scalar curvatures of $\tilde{g}$ and g. Since $\Delta \tilde{G}$ has compact support, we may choose $\delta > 0$ sufficiently small so that the above is strictly positive. In particular, $\tilde{R} \geq 0$ on $\Omega \setminus \{p\}$.

Now, suppose s is the distance function with respect to g from the point p. For s sufficiently small, G is of the form $c/s + O(1)$ for some constant $c > 0$. The normal derivative of G to the sphere of radius s about p in the outward direction is $-c/s^2 + O(s^{-1})$ (see [Aubin 1998, Proposition 4.12 and Theorem 4.13]). The mean curvature of the sphere of radius s with respect to g is $2/s + O(1)$ (by [Fan et al. 2009, Lemma 3.4]), and so the mean curvature of this sphere with respect to $\tilde{g}$ is (using Appendix A):

$$(1 + \delta \tilde{G})^{-3}\big((2s^{-1} + O(1))(1 + \delta c s^{-1} + O(1)) - 4\delta c s^{-2} + O(s^{-1})\big).$$

The dominant term is $-2\delta c s^{-2}$, so that for some $s > 0$ sufficiently small, $\partial \bar{B}(p, s)$ has negative mean curvature (with respect to $\tilde{g}$) in direction pointing away from p. Let $\tilde{\Omega}$ be $\Omega \setminus B(p, s)$, and restrict $\tilde{g}$ to $\tilde{\Omega}$.

The manifold $(\tilde{\Omega}, \tilde{g})$ has boundary with two connected components, both of positive mean curvature in the *outward* direction. By Lemma 6 below, $(\tilde{\Omega}, \tilde{g})$ contains a subset that is a valid fill-in of (Σ, γ, H) with a minimal boundary component. This contradicts the assumption that the Bartnik data is of zero type, and so we have proved the claim that g is scalar-flat.

Finally, if (Ω, g) is not static vacuum, then Corvino proves the existence of a metric $\bar{g}$ on Ω with nonnegative, scalar curvature, positive at some interior point p, such that $g - \bar{g}$ is supported away from $\partial\Omega$ [Corvino 2000]. In particular, $(\Omega, \bar{g})$ is a valid fill-in for the type-zero data (Σ, γ, H), and the above argument leads to a contradiction. □

To complete the proof of the proposition, we need also the following lemma:

Lemma 6. *Suppose $\mathcal{B} = (\Sigma, \gamma, H)$ admits a fill-in (Ω, g) with nonnegative scalar curvature, such that $\partial\Omega \setminus \Sigma$ has positive mean curvature in the outward direction. Then a subset Ω' of Ω is a valid fill-in of $\mathcal{B}$ with metric $g|_{\Omega'}$. Moreover, Ω' has at least one minimal boundary component.*

Proof. By assumption Σ has positive mean curvature H and $\partial\Omega \setminus \Sigma$ has positive mean curvature. By Theorem 19 in Appendix B, there exists a smooth, embedded minimal surface S homologous to Σ. The closure of the region bounded between Σ and S is the desired valid fill-in. □

2.2. ***Data of positive type.***

Proposition 7. *Given Bartnik data $\mathcal{B}$, the following are equivalent:*

(1) *$\mathcal{B}$ is of positive type.*

(2) *$\mathcal{B}$ admits a valid fill-in that has positive scalar curvature at some point.*

(3) *$\mathcal{B}$ admits a valid fill-in that has positive scalar curvature everywhere.*

The idea of proving the proposition is to create positive energy density at some interior points at the expense of decreasing the size of the minimal surface. As in the previous section, the delicate issue is preserving the boundary data in the process.

Proof. If $\mathcal{B}$ admits a valid fill-in with positive scalar curvature at a point, then $\mathcal{B}$ is of nonnegative type and Proposition 5 rules out the case of zero type (since static vacuum metrics have zero scalar curvature). This shows (2) implies (1); (3) trivially implies (2).

Last, we show (1) implies (3). Suppose $\mathcal{B} = (\Sigma, \gamma, H)$ has positive type, so there exists some valid fill-in (Ω, g) of $\mathcal{B}$ with boundary $\Sigma \dot{\cup} S$, where S is a nonempty minimal surface. If (Ω, g) is not static vacuum, we may complete the proof by again using the work of Corvino [2000] to perturb (Ω, g) to a valid fill-in with positive scalar curvature at a point. Thus, assume (Ω, g) is static vacuum, and so in particular it is scalar-flat.

Replace (Ω, g) with its double across the minimal surface S. Now, (Ω, g) has two boundary components Σ and Σ' (its reflected copy), and contains a minimal surface S that is fixed by the $\mathbb{Z}_2$ reflection symmetry. Moreover, g is Lipschitz continuous across S and smooth elsewhere[1]. For simplicity of exposition, we separately treat the cases in which g is smooth and nonsmooth across S.

[1] This doubling trick across a minimal surface was used by Bunting and Masood-ul-Alam [1987] to classify static vacuum metrics with compact minimal boundary that are asymptotically flat. Because of the asymptotic condition, their theorem does not apply to the present case. We also mention the fact that because of minimality and the static vacuum condition, S is totally geodesic, which implies that $\tilde{g}$ is $C^{1,1}$ across S [Bunting and Masood-ul Alam 1987; Corvino 2000]. However, we do not need this fact.

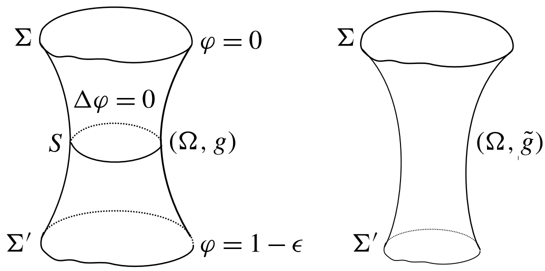


Figure 2. Construction in proof of Proposition 7. Left: the double of (Ω, g), which we also refer to as (Ω, g), abusing notation. The function φ is harmonic, with the given prescribed Dirichlet boundary values. Right: Ω equipped with the metric $\tilde{g}$, obtained from g by applying the conformal factor φ^4.

Smooth case. For $\epsilon \in (0, 1)$, let φ be the function on Ω solving the following Dirichlet problem:

$$\begin{cases} \Delta\varphi = 0 & \text{on } \Omega, \\ \varphi = 1 & \text{on } \Sigma, \\ \varphi = 1 - \epsilon & \text{on } \Sigma'. \end{cases}$$

Consider the conformal metric $\tilde{g} = \varphi^4 g$, which is smooth with zero scalar curvature. Moreover, the mean curvature $\tilde{H}$ of Σ with respect to $\tilde{g}$ strictly exceeds H (for all choices of ϵ), since φ has positive outward normal derivative on Σ (see Appendix A). The mean curvature of Σ' remains positive for $\epsilon > 0$ sufficiently small. Fix such an ϵ. This construction is demonstrated in Figure 2.

Fix any smooth function $\rho > 0$ on Ω. For all $\delta \geq 0$ small, let u_δ be the unique solution to the elliptic problem

$$\begin{cases} \tilde{L} u_\delta = \delta\rho & \text{in } \tilde{\Omega}, \\ u_\delta = 1 & \text{on } \Sigma, \\ \partial_\nu(u_\delta) = 0 & \text{on } \Sigma', \end{cases}$$

where $\tilde{L} = -8\tilde{\Delta}$ is the conformal Laplacian of $\tilde{g}$. Clearly $u_0 \equiv 1$, and u_δ converges in C^2 to 1 as $\delta \to 0^+$. For $\delta > 0$ small enough to ensure $u_\delta > 0$, the conformal metric $u_\delta^4 \tilde{g}$ has:

- positive scalar curvature (equal to $\delta\rho u_\delta^{-5}$);
- induced metric on Σ equal to γ (by the boundary condition $u_\delta|_\Sigma = 1$);
- positive mean curvature on Σ' (by the boundary condition $\partial_\nu(u_\delta)|_{\Sigma'} = 0$);
- mean curvature on Σ converging uniformly to $\tilde{H}$ as $\delta \to 0^+$.

Fix a particular value of δ such that the mean curvature of Σ' is positive and the mean curvature $\tilde{H}_\delta$ of Σ is pointwise greater than H (which is possible, since $\tilde{H} > H$). By Lemma 6, there is a valid fill-in of $(\Sigma, \gamma, \tilde{H}_\delta)$ that contains a minimal surface. By Lemma 20 in Appendix C, this valid fill-in can be perturbed to a valid fill-in of (Σ, γ, H) so that the latter fill-in still has positive scalar curvature.

Lipschitz case. In general we must carry out an extra step to deal with the lack of smoothness across S. Define φ analogously by first solving $\Delta\varphi_1 = 0$ with boundary conditions of 1 on Σ and $1-\epsilon/2$ on S, then defining $\varphi_2 = 2-\epsilon-\varphi_1$ in the reflected copy. The function φ obtained by gluing φ_1 and φ_2 is $C^{1,1}$ on Ω, and smooth and harmonic away from S. Again, let $\tilde{g} = \varphi^4 g$, which has zero scalar curvature (away from S), is Lipschitz across S, and induces the same mean curvature on both sides of S. Fix $\epsilon > 0$ so that $\tilde{H} > H$ and the $\tilde{g}$-mean curvature of Σ' is positive.

By the work of Miao [2002], the fact that both sides of S have the same mean curvature implies the existence of a family of C^2 metrics $\{\tilde{g}_\delta\}_{0<\delta<\delta_0}$ such that

(1) $\tilde{g}_\delta$ converges to $\tilde{g}$ in C^0 as $\delta \to 0^+$,

(2) $\tilde{g}_\delta$ agrees with $\tilde{g}$ outside a δ-neighborhood of S, and

(3) the scalar curvature $\tilde{R}_\delta$ of $\tilde{g}_\delta$ is bounded below by a constant independent of δ.

In particular, the L^p norm of $\tilde{R}_\delta$ (taken with respect to $\tilde{g}$ or $\tilde{g}_\delta$) for any $1 \le p < \infty$ converges to zero as $\delta \to 0$. We mimic arguments of Schoen and Yau [1979] to prove:

Lemma 8. *For each $\delta > 0$ sufficiently small, the conformal Laplacian*

$$\tilde{L}_\delta = -8\tilde{\Delta}_\delta + \tilde{R}_\delta$$

of $\tilde{g}_\delta$ has trivial kernel on the space of functions v with boundary conditions of $v = 0$ on Σ and $\partial_\nu v = 0$ on Σ'.

Proof. Let v belong to the kernel of $\tilde{L}_\delta$ with the above boundary conditions. Multiplying $\tilde{L}_\delta v$ by v and integrating by parts gives

$$0 = \int_\Omega (8|\nabla v|^2_{\tilde{g}_\delta} + \tilde{R}_\delta v^2)\, d\tilde{V}_\delta.$$

Let $\tilde{R}^-_\delta = -\min(\tilde{R}_\delta, 0)$, so that

$$\begin{aligned}\int_\Omega 8|\nabla v|^2_{\tilde{g}_\delta} d\tilde{V}_\delta &\le \int_\Omega \tilde{R}^-_\delta v^2 d\tilde{V}_\delta \\ &\le \left(\int_\Omega (\tilde{R}^-_\delta)^{3/2} d\tilde{V}_\delta\right)^{2/3} \left(\int_\Omega v^6 d\tilde{V}_\delta\right)^{1/3} \\ &\le c\left(\int_\Omega (\tilde{R}^-_\delta)^{3/2} d\tilde{V}_\delta\right)^{2/3} \left(\int_\Omega |\nabla v|^2_{\tilde{g}_\delta} d\tilde{V}_\delta\right),\end{aligned}$$

having used the Hölder and Sobolev inequalities (where $c > 0$ is a constant). Thus, for δ sufficiently small, a nonzero v may not exist, since the $L^{3/2}$ norm of $\tilde{R}_\delta^-$ converges to zero. □

Fix a smooth function $\rho > 0$ on Ω. By the lemma and standard elliptic theory, for $\delta > 0$ small there exists unique solution u_δ to the problem:

$$\begin{cases} \tilde{L}_\delta u_\delta = \delta\rho & \text{in } \tilde{\Omega}, \\ u_\delta = 1 & \text{on } \Sigma, \\ \partial_\nu(u_\delta) = 0 & \text{on } \Sigma'. \end{cases}$$

A key fact is that u_δ converges to 1 in C^0 as $\delta \to 0^+$, and this convergence is C^2 away from S (see the proof of Proposition 4.1 of [Miao 2002]).

At this point, the proof follows nearly the same steps as in the smooth case, where we work with the metric $u_\delta^4 \tilde{g}_\delta$ (which has positive scalar curvature and induces the metric γ on Σ). We pick $\delta > 0$ sufficiently small so that $\tilde{H}_\delta > H$ and Σ' has positive mean curvature with respect to $u_\delta^4 \tilde{g}_\delta$. Now, if necessary, perturb the C^2 metric $u_\delta^4 \tilde{g}_\delta$ on a neighborhood of S to a C^∞ metric, preserving the above properties. The proof now goes as in the smooth case, making use of Lemmas 6 and 20. □

We remark that our assumption of positive Gauss curvature of (Σ, γ) is not necessary in Propositions 5 and 7.

2.3. ***Bartnik inner mass.*** One source of inspiration for the problem of considering valid fill-ins with minimal boundary is Bray's definition of the Bartnik inner mass [Bray 2001], an example of a quasi-local mass (see Section 4 for more on quasi-local mass). The Bartnik inner mass aims to measure the size of the largest black hole that could be placed inside a valid fill-in of given Bartnik data.

Definition 9. The *Bartnik inner mass* of Bartnik data $\mathcal{B}$ is the real number

$$m_{\text{inner}}(\mathcal{B}) = \sup_{(\Omega,g)} \left\{ \sqrt{\frac{A}{16\pi}} \right\}$$

where the supremum is taken over the class of all valid fill-ins (Ω, g) of $\mathcal{B}$, and A is the minimum area in the homology class of Σ in (Ω, g).

This definition, though formulated differently, is equivalent to Bray's. The purpose of using the minimum area in the homology class of Σ is to ignore any large minimal surfaces "hidden behind" a smaller minimal surface.

We observe that the sign of $m_{\text{inner}}(\mathcal{B})$ corresponds directly to the type of the Bartnik data $\mathcal{B}$. To see this, first note that for fill-ins with a minimal boundary, the minimum area of A in the homology class of Σ in (Ω, g) is always attained by a smooth minimal surface, and so A is positive (see Theorem 19). For fill-ins without boundary, Σ is homologically trivial, and so $A = 0$. Thus, $m_{\text{inner}}(\mathcal{B})$ is positive if $\mathcal{B}$ is of positive type; zero if $\mathcal{B}$ is of zero type; and $-\infty$ if $\mathcal{B}$ is of negative type.

3. The interval of positivity

The following idea was suggested by Bray: as a function of a parameter $\lambda > 0$, consider the Bartnik data $(\Sigma, \gamma, \lambda H)$. The main purpose of this section is to state and prove Theorem 11, which partially answers the question of how the type of the data depends on λ.

One key ingredient is the following well-known theorem.[2]

Theorem 10 [Shi and Tam 2002]. *If Bartnik data (Σ, γ, H) has a valid fill-in (Ω, g), then*

$$\int_\Sigma (H_0 - H)\, dA_\gamma \geq 0, \tag{3}$$

where H_0 is the mean curvature of an isometric embedding of (Σ, γ) into Euclidean space $\mathbb{R}^3$, and dA_γ is the area form on Σ with respect to the metric γ. Moreover, equality holds if and only if (Ω, g) is isometric to a subdomain of $\mathbb{R}^3$.

Recall that we assume γ to have positive Gauss curvature, which is necessary for the theorem: H_0 is well-defined, since an isometric embedding of a positive Gauss curvature surface into $\mathbb{R}^3$ exists and is unique up to rigid motions (see references [13] and [19] in [Shi and Tam 2002]).

In our case, inequality (3), which depends only on the Bartnik data, must be satisfied for data that admits a valid fill-in. In particular, by increasing H (while keeping γ, and therefore H_0, fixed), it is clear that some Bartnik data do not possess fill-ins (i.e., are of negative type). Hence, the Shi–Tam theorem gives an obstruction to Bartnik data being of nonnegative type.

The following main theorem demonstrates that there exists a unique interval of values of λ for which this data $(\Sigma, \gamma, \lambda H)$ is of positive type.

Theorem 11. *Fix Bartnik data (Σ, γ, H). There exists a unique number $\lambda_0 > 0$ such that $(\Sigma, \gamma, \lambda H)$ is of positive type if and only if $\lambda \in (0, \lambda_0)$. Moreover, $(\Sigma, \gamma, \lambda H)$ is of negative type if $\lambda > \lambda_0$.*

As a consequence, $(\Sigma, \gamma, \lambda H)$ is zero type for at most one value of λ, namely λ_0.

[2] We remark that the Shi–Tam theorem was originally stated for the case in which every component of $\partial\Omega$ has positive Gauss and mean curvatures. However, one can allow additional minimal surface components (as we have done here) by observing the positive mass theorem is true for manifolds with compact minimal boundary. Alternatively, one could employ a reflection argument to eliminate any minimal surface boundary components.

Proof. Define

$$\begin{aligned} I_+ &= \{\lambda \in \mathbb{R}^+ : (\Sigma, \gamma, \lambda H) \text{ is of positive type}\}, \\ I_0 &= \{\lambda \in \mathbb{R}^+ : (\Sigma, \gamma, \lambda H) \text{ is of zero type}\}, \\ I_{\geq 0} &= I_+ \cup I_0. \end{aligned}$$

Step 1: We first show I_+ is nonempty. Consider the space $\Omega = \Sigma \times [-1, 0]$ with product metric g, and identify Σ with $\Sigma \times \{0\}$. Let S be the other boundary component of Ω, namely $\Sigma \times \{-1\}$. Observe that 1) Ω has positive scalar curvature since Σ has positive Gauss curvature, and 2) all leaves $\Sigma \times \{t\}$ are minimal surfaces.

Choose a smooth function v on Ω satisfying the following properties: $v \leq 0$, v vanishes on Σ and in a neighborhood of S, and $\partial_\nu v = \frac{1}{4} H$ on Σ. For $\epsilon > 0$, let $u_\epsilon = 1 + \epsilon v$. In particular, u_ϵ is positive for $\epsilon > 0$ sufficiently small. Consider the conformal metric $g_\epsilon = u_\epsilon^4 g$. Note that g_ϵ induces the metric γ on Σ, and assigns the following value to the mean curvature of Σ:

$$H_\epsilon = 4\partial_\nu(u_\epsilon) = 4\epsilon \partial_\nu v = \epsilon H,$$

by our choice of v. Moreover, the scalar curvature of g_ϵ is

$$R_{g_\epsilon} = u_\epsilon^{-5}(-8\Delta_g u_\epsilon + R_g u_\epsilon) = u_\epsilon^{-5}(-8\epsilon\Delta_g v + R_g u_\epsilon),$$

which is positive for ϵ sufficiently small, since $R_g > 0$ and u_ϵ is uniformly bounded below as $\epsilon \to 0^+$. Fix such an ϵ. We can see (Ω, g_ϵ) is a valid fill-in of $(\Sigma, \gamma, \epsilon H)$, since this fill-in has positive scalar curvature, induces the correct boundary geometry on Σ, and S is minimal (since $g_\epsilon = g$ near S). In particular, ϵ belongs to I_+, so $I_+ \neq \varnothing$.

Step 2: The next step is to show that $I_{\geq 0}$ is connected, and I_0 contains at most one point. To accomplish this, we show that for every number in $I_{\geq 0}$, every smaller positive number belongs to I_+. It suffices to show that if (Σ, γ, H) is of nonnegative type, then $(\Sigma, \gamma, \lambda H)$ is of positive type for all $\lambda \in (0, 1)$. This fact follows from the next lemma, by Proposition 7.

Lemma 12. *Let (Ω, g) be a fill-in of arbitrary Bartnik data (Σ, γ, H). Fix $\lambda \in (0, 1)$ and a neighborhood U of Σ in Ω. There exists a metric $\tilde{g}$ on Ω such that*

(a) *$\tilde{g}$ is a fill-in of $(\Sigma, \gamma, \lambda H)$,*

(b) *$\tilde{g} \geq g$, with equality outside U, and*

(c) *$R_{\tilde{g}} \geq \min(0, R_g)$ pointwise, with strict inequality on a neighborhood of Σ, where R and $R_{\tilde{g}}$ are the scalar curvature of g and $\tilde{g}$.*

In particular, if (Ω, g) is a valid fill-in, so is $(\Omega, \tilde{g})$.

Proof. Working in a neighborhood of Σ in Ω diffeomorphic to $(-t_0, 0] \times \Sigma$ and contained in U, we may assume g takes the form

$$g = dt^2 + G_t,$$

where t is the negative of g-distance to Σ, and G_t is a Riemannian metric on the surface $\Sigma_t = \Sigma \times \{t\}$. Shrinking t_0 if necessary, we may assume that every (Σ_t, G_t) has positive Gauss curvature K_t and positive mean curvature H_t (in the outward direction ∂_t). Let $\rho : (-t_0, 0] \to \mathbb{R}$ be a smooth function equal to 1 in a neighborhood of $-t_0$, and satisfying

$$\rho(0) = \lambda^{-1} > 1, \quad \rho'(t) \geq 0, \quad \rho'(0) > 0.$$

Define a new metric $\tilde{g}$ on Ω by setting

$$\tilde{g} = \rho(t)^2 dt^2 + G_t \tag{4}$$

on the neighborhood of Σ, and extending smoothly by g to the rest of Ω; claim (b) is satisfied. A straightforward calculation shows that Σ has mean curvature λH in the metric $\tilde{g}$; moreover $\tilde{g}$ induces the metric γ on Σ, so claim (a) holds. Last, we must study the scalar curvature of $\tilde{g}$ on the neighborhood $(-t_0, 0] \times \Sigma$. The following well-known formula, obtained from computing the variation of mean curvature under a unit normal flow, gives the scalar curvature of g as:

$$R_g = -2\frac{\partial H_t}{\partial t} + 2K_t - H_t^2 - \|h_t\|^2, \tag{5}$$

where h_t is the second fundamental form of Σ_t in (Ω, g), and its norm $\|\cdot\|^2$ is taken with respect to G_t. Applying this formula to the metric $\tilde{g}$ yields

$$R_{\tilde{g}} = \frac{1}{\rho(t)^2} R_g + 2K_t(1 - \rho(t)^{-2}) + 2\frac{\rho'(t)}{\rho(t)^3} H_t. \tag{6}$$

Now, $K_t > 0$, $\rho(t) \geq 1$, $\rho'(t) \geq 0$ and $H_t > 0$, so we see $R_{\tilde{g}}(x) \geq 0$ if $R_g(x) \geq 0$ and $R_{\tilde{g}}(x) \geq R_g(x)$ if $R_g(x) < 0$; both are strict inequalities near $t = 0$, proving claim (c). □

We conclude that $I_{\geq 0}$ is a convex subset of $\mathbb{R}^+$, containing all arbitrarily small positive numbers. Moreover, I_0 contains at most a single point.

Step 3: We prove that $I_{\geq 0}$ is bounded above. This follows immediately from the work of Shi and Tam. More precisely, if $\lambda \in I_{\geq 0}$, then

$$\lambda \leq \frac{\int_\Sigma H_0 dA_\gamma}{\int_\Sigma H dA_\gamma}.$$

Together with step 2, we see $I_{\geq 0}$ and I_+ are intervals of the form $(0, \lambda_0]$ or $(0, \lambda_0)$.

Step 4: Here we prove that λ_0 does not belong to I_+. If $\lambda_0 \in I_+$, then by Proposition 7, there exists a valid fill-in (Ω, g) of $(\Sigma, \gamma, \lambda_0 H)$ with positive scalar curvature at some point and boundary $\Sigma \dot{\cup} S_0$, with S_0 minimal and nonempty. Solve the mixed Dirichlet–Neumann problem:

$$\begin{cases} \Delta u = \frac{1}{8} R_g u & \text{in } \Omega, \\ u = 1 & \text{on } \Sigma, \\ \partial_\nu(u) = 0 & \text{on } S_0. \end{cases} \tag{7}$$

Here, ν is the unit normal, always chosen to point out of Ω. Note that a solution exists because $R_g \geq 0$. By the maximum principle, $u > 0$ in Ω and $\partial_\nu(u) > 0$ on Σ. Let $g' = u^4 g$. Note that g' has zero scalar curvature, induces the metric γ on Σ and assigns zero mean curvature to S_0. In particular, if we let H' be the mean curvature of Σ with respect to g', then (Σ, γ, H') has a valid fill-in with minimal boundary, namely (Ω, g'), and is therefore of positive type. Observe that $H' > \lambda_0 H$. Choose $\beta > 1$ so that $H' > \beta\lambda_0 H$. By Lemma 20 in Appendix C, we see that $(\Sigma, \gamma, \beta\lambda_0 H)$ is of positive type. Therefore $\beta\lambda_0 \in I_+$, which contradicts $\lambda_0 = \sup I_+$. We conclude $I_+ = (0, \lambda_0)$, and either $I_{\geq 0} = (0, \lambda_0)$ or $(0, \lambda_0]$. It follows that if $\lambda > \lambda_0$, then $(\Sigma, \gamma, \lambda H)$ must be of negative type. □

To emphasize the picture, the data $(\Sigma, \gamma, \lambda H)$ is of positive type for λ small. As we increase λ, this behavior persists until $\lambda = \lambda_0$. At this point, the data is zero or negative, and for $\lambda > \lambda_0$, the data is negative. See Section 7 for further discussion of the behavior near $\lambda = \lambda_0$.

3.1. ***Inner mass function.*** In the rest of this section we will study the function

$$m(\lambda) = m_{\text{inner}}(\Sigma, \gamma, \lambda H) \tag{8}$$

defined for $\lambda \in (0, \lambda_0)$. Intuitively, one would expect the following behavior of the function $m(\lambda)$. For λ small, the mean curvature λH is close to zero, so one might anticipate the existence of a valid fill-in with minimal boundary of approximately the same area as Σ.

As λ increases, one would expect the class of valid fill-ins to shrink; one reason is that the Shi–Tam inequality is more difficult to satisfy. Consequently, the Bartnik inner mass ought to decrease as well. The following statement supports this intuition.

Proposition 13. *Given Bartnik data (Σ, γ, H), the function $m : (0, \lambda_0) \to \mathbb{R}^+$ is continuous and decreasing, with the limiting behavior*

$$\lim_{\lambda \to 0^+} m(\lambda) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}},$$

where $|\Sigma|_\gamma$ is the area of Σ with respect to γ.

Proof. Monotonicity: Given $0 < \lambda_1 < \lambda_2 < \lambda_0$, we showed in Lemma 12 that any valid fill-in of $(\Sigma, \gamma, \lambda_2 H)$ gives rise to a valid fill-in of $(\Sigma, \gamma, \lambda_1 H)$ with a metric that is pointwise at least as large (see (4)). From the definition of the Bartnik inner mass, this shows that $m(\lambda_1) \geq m(\lambda_2)$.

Continuity: Suppose $0 < \lambda_1 < \lambda_0$, and let $\epsilon > 0$. From the definition of the Bartnik inner mass, there exists a valid fill-in (Ω, g) of $(\Sigma, \gamma, \lambda_1 H)$ whose minimum area A in the homology class of Σ satisfies

$$m(\lambda_1) - \sqrt{\frac{A}{16\pi}} < \frac{\epsilon}{3}.$$

From Proposition 7, there exists a valid fill-in $(\tilde{\Omega}, \tilde{g})$ of $(\Sigma, \gamma, \lambda_1 H)$ that has strictly positive scalar curvature, and whose minimum area $\tilde{A}$ in the homology class of Σ is close to A:

$$\sqrt{\frac{A}{16\pi}} - \sqrt{\frac{\tilde{A}}{16\pi}} < \frac{\epsilon}{3}.$$

Now, for $\lambda_0 > \lambda > \lambda_1$, $(\tilde{\Omega}, \tilde{g})$ can be perturbed to a fill-in $(\tilde{\Omega}, \tilde{g}_\lambda)$ of $(\Sigma, \gamma, \lambda H)$ using a metric of the form (4). The scalar curvature of $\tilde{g}_\lambda$ has potentially decreased relative to that of $\tilde{g}$, but remains positive for $\lambda > \lambda_1$ sufficiently close to λ_1. Since $\tilde{g}_\lambda \to \tilde{g}$ in C^0, we may assume $\lambda - \lambda_1$ is small enough so that

$$\sqrt{\frac{\tilde{A}}{16\pi}} - \sqrt{\frac{\tilde{A}_\lambda}{16\pi}} < \frac{\epsilon}{3},$$

where $\tilde{A}_\lambda$ is the minimum $\tilde{g}_\lambda$-area in the homology class of Σ. Adding the last three inequalities and using the definition of the Bartnik inner mass gives

$$m(\lambda_1) < \epsilon + \sqrt{\frac{\tilde{A}_\lambda}{16\pi}} \leq \epsilon + m(\lambda),$$

for $\lambda - \lambda_1$ sufficiently small. Together with the fact that $m(\cdot)$ is decreasing, we have shown $m(\cdot)$ is continuous at λ_1.

Lower limit behavior: To study the behavior of $m(\epsilon)$ for ϵ small, recall that in Step 1 of the proof of Theorem 11 we constructed a valid fill-in of $(\Sigma, \gamma, \epsilon H)$ by a metric g_ϵ uniformly close (controlled by ϵ) to a cylindrical product metric g over (Σ, γ). As $\epsilon \to 0^+$, the minimum g_ϵ area in the homology class of Σ converges to the minimum g-area in the same homology class, which is $|\Sigma|_\gamma$. On the other hand, the Bartnik inner mass of (Σ, γ, H') (for any H') never exceeds $\sqrt{|\Sigma|_\gamma/16\pi}$ by definition. This proves

$$\lim_{\lambda \to 0+} m(\lambda) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}. \qquad \square$$

In Section 7 we conjecture that $m(\lambda)$ limits to zero as $\lambda \to \lambda_0^-$, behavior supported by the explicit computation of $m(\lambda)$ in a spherically symmetric case in Section 5.1.

4. Quasilocal mass

Recall from the introduction the problem of assigning a "quasi-local mass" to a bounded region Ω in a totally geodesic spacelike slice (M, g) of a spacetime. By most definitions, the quasi-local mass of Ω depends only on the Bartnik data (Σ, γ, H) of the boundary, and we adopt this perspective here. That is, we define a *quasi-local mass functional* to be a map from (a subspace of) the set of Bartnik data to the real numbers. We refer the reader to [Szabados 2009] for a recent comprehensive survey of quasi-local mass.

We begin by recalling some well-known examples of quasi-local mass. First, the *Hawking mass* of (Σ, γ, H) is defined to be

$$m_H(\Sigma, \gamma, H) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}} \left(1 - \frac{1}{16\pi} \int_\Sigma H^2 dA_\gamma\right).$$

There is no correlation between the sign of the Bartnik data and the sign of the Hawking mass. That is, the Hawking mass can be negative for positive Bartnik data, and vice versa (see Section 5).

Next, the *Brown–York mass* is defined for Bartnik data (Σ, γ, H) (assuming as we do that $K_\gamma > 0$ and $H > 0$) by

$$m_{\mathrm{BY}}(\Sigma, \gamma, H) = \frac{1}{8\pi} \int_\Sigma (H_0 - H) dA_\gamma,$$

where H_0 is the mean curvature of an isometric embedding of (Σ, γ) into $\mathbb{R}^3$. Theorem 10 establishes that the Brown–York mass is nonnegative for Bartnik data of nonnegative type. However, there exist Bartnik data of both negative and zero type for which the Brown–York mass is strictly positive (see Section 5).

A third example is the Bartnik inner mass, defined in Section 2.3.

A key observation is that Theorem 11 canonically associates to any Bartnik data (with $H > 0$ and $K_\gamma > 0$) a positive number λ_0, which we call the *critical parameter*. In this section we use λ_0 to construct a new example of a quasi-local mass functional.

To motivate this definition, we will compute the number λ_0 for concentric round spheres Σ_r in the Schwarzschild manifold of mass m, with induced metric γ_r and mean curvature H_r. For our purposes the Schwarzschild manifold of mass m is $\mathbb{R}^3$ minus the open Euclidean ball of radius $m/2$, where $m > 0$, equipped with the metric

$$g = \left(1 + \frac{m}{2r}\right)^4 \delta, \tag{9}$$

where δ is the Euclidean metric. Note that g is scalar-flat and its boundary is a minimal 2-sphere, called the *horizon*.

Straightforward computations show that (Σ_r, γ_r) is a round sphere of area $4\pi r^2(1+m/2r)^4$, and that

$$H_r = \frac{2}{r}\left(1+\frac{m}{2r}\right)^{-2} - \frac{2m}{r^2}\left(1+\frac{m}{2r}\right)^{-3},$$

where we have used (19). The mean curvature H_r^0 of (Σ_r, γ_r) embedded in $\mathbb{R}^3$ is

$$H_r^0 = \frac{2}{r}\left(1+\frac{m}{2r}\right)^{-2}.$$

Therefore, if we let $\lambda_r = H_r^0/H_r$, then $(\Sigma_r, \gamma_r, \lambda_r H_r)$ admits a valid fill-in — namely a closed ball in flat-space of boundary area $4\pi r^2(1+m/2r)^4$. On the other hand, if λ belongs to the interval of positivity for $(\Sigma_r, \gamma_r, H_r)$, then by Shi–Tam

$$\lambda \leq \frac{\int_\Sigma H_r^0 dA_\gamma}{\int_\Sigma H_r dA_\gamma} = \lambda_r.$$

Thus, λ_r is the critical parameter for the Bartnik data. Some simplifications show

$$\lambda_r = \frac{1+m/2r}{1-m/2r}. \tag{10}$$

In particular, we have the identity in Schwarzschild space:

$$m = \sqrt{\frac{|\Sigma_r|_g}{16\pi}}\left(1-\frac{1}{\lambda_r^2}\right),$$

for all values of r, motivating the following definition of quasi-local mass.

Definition 14. Let $\mathcal{B} = (\Sigma, \gamma, H)$ be Bartnik data with critical parameter λ_0 (from Theorem 11). Define

$$m(\mathcal{B}) = m(\Sigma, \gamma, H) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(1-\frac{1}{\lambda_0^2}\right).$$

Recall that we assume γ has positive Gauss curvature and $H > 0$.

Theorem 15. *Definition 14 of quasi-local mass satisfies the following properties*:

(1) (*nonnegativity*) *If Bartnik data $\mathcal{B}$ admits a valid fill-in, then its mass $m(\mathcal{B})$ is nonnegative and is zero only if every valid fill-in is static vacuum.*

(2) (*spherical symmetry*) *If Bartnik data $\mathcal{B}$ arises from a coordinate sphere in a Schwarzschild metric of mass m, then $m(\mathcal{B}) = m$.*

(3) (*black hole limit*) *If* $\mathcal{B}_n = (\Sigma, \gamma, H_n)$ *is a sequence of Bartnik data and* $H_n \to 0$ *uniformly*, *then*

$$\lim_{n\to\infty} m(\mathcal{B}_n) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}.$$

(4) (*ADM-sublimit*) *If* (M, g) *is an asymptotically flat manifold with nonnegative scalar curvature*, *and if* S_r *is a coordinate sphere of radius* r *with induced metric* γ_r *and mean curvature* H_r, *then*

$$m_{\mathrm{ADM}}(M, g) \geq \limsup_{r\to\infty} m(S_r, \gamma_r, H_r). \tag{11}$$

Remarks. The proof of Theorem 15 uses the positive mass theorem [Schoen and Yau 1979] implicitly, via Lemma 16 below, which relies on the theorem of Shi and Tam. On the other hand Theorem 15 also recovers the positive mass theorem: if (M, g) is asymptotically flat, has nonnegative scalar curvature, with ∂M empty or consisting of minimal surfaces, then by property (1), $m(S_r) \geq 0$ for all S_r. From this, inequality (11) gives $m_{\mathrm{ADM}} \geq 0$.

Proof. Nonnegativity: Observe the following four statements are equivalent, using Theorem 11: $m(\mathcal{B}) > 0$; $\lambda_0 > 1$; the number 1 belongs to the interval of positivity I_+; $\mathcal{B}$ is of positive type. Also, if (Σ, γ, H) is of zero type, then $\lambda_0 = 1$ (as follows from Theorem 11), so $m(\mathcal{B})$ vanishes. On the other hand, if $m(\mathcal{B})$ vanishes, then $\lambda_0 = 1$, so the data is either negative or zero (again, by Theorem 11). But if it is given that the data admits a fill-in, then the data must be of zero type. By Proposition 5, any such fill-in is static vacuum.

Spherical symmetry: This is clear from the construction at the beginning of this section; we defined quasi-local mass so that it has this property.

Black hole limit: It is straightforward to check that if $H_n \to 0$ uniformly, then the sequence of critical parameters λ_n diverges to infinity.

ADM-sublimit: For all r sufficiently large, the coordinate spheres S_r have positive mean and Gauss curvatures. To prove (11), recall that the Brown–York mass limits to the ADM mass in the sense that

$$m_{\mathrm{ADM}}(M, g) = \lim_{r\to\infty} m_{\mathrm{BY}}(S_r).$$

(See Theorem 1.1 of [Fan et al. 2009] and the references therein.) Since we assume (M, g) has nonnegative scalar curvature, S_r is of positive or zero type for all r for which the coordinate sphere is defined. We invoke Lemma 16 below, which states $m(S_r) \leq m_{\mathrm{BY}}(S_r)$, completing the proof. □

Lemma 16. *For Bartnik data* $\mathcal{B} = (\Sigma, \gamma, H)$ *of nonnegative type*,

$$m(\mathcal{B}) \leq m_{\mathrm{BY}}(\mathcal{B}).$$

Proof. Let H_0 be the mean curvature of an isometric embedding of (Σ, γ) in $\mathbb{R}^3$, which is well-defined because $K_\gamma > 0$. By Shi–Tam, we have

$$\lambda_0 \leq \frac{\int_\Sigma H_0 dA_\gamma}{\int_\Sigma H dA_\gamma}.$$

In particular,

$$(12)\qquad m(\mathcal{B}) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(1 - \frac{1}{\lambda_0^2}\right) \leq \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(1 - \left(\frac{\int_\Sigma H dA_\gamma}{\int_\Sigma H_0 dA_\gamma}\right)^2\right)$$
$$\leq \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(\frac{\left(\int_\Sigma H_0 dA_\gamma + \int_\Sigma H dA_\gamma\right)\left(\int_\Sigma H_0 dA_\gamma - \int_\Sigma H dA_\gamma\right)}{\left(\int_\Sigma H_0 dA_\gamma\right)^2}\right)$$
$$\leq \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\frac{16\pi m_{\mathrm{BY}}(\mathcal{B})}{\int_\Sigma H_0 dA_\gamma},$$

where again we have used Shi–Tam and the fact that the data is of nonnegative type. The Minkowski inequality for convex regions in $\mathbb{R}^3$ [Pólya and Szegö 1951] states that

$$\left(\int_\Sigma H_0 dA_\gamma\right)^2 \geq 16\pi|\Sigma|_\gamma.$$

Together with the above, this completes the proof. □

The right-hand side of (12) is a definition of quasi-local mass proposed by Miao, which he observed is bounded above by the Brown–York mass using the same argument [Miao 2009].

4.1. *Physical remarks.* It has been suggested in the literature (see [Bartnik 2002], for instance) that if the quasi-local mass of the boundary of a region Ω vanishes, then Ω ought to be flat. The Brown–York mass and Bartnik mass both satisfy this property (see [Bartnik 1989; Huisken and Ilmanen 2001]). Definition 14 suggests an alternative viewpoint that such Ω ought to be *static vacuum*, which includes flat metrics as a special case. Indeed, one could make a physical argument that in a region of a spacetime that is static vacuum, quasi-local mass should vanish since there is no matter content and no gravitational dynamics (cf. [Anderson 2010], which also discusses the vanishing of quasi-local mass on static vacuum regions).

5. Examples

Let (M, g) be a Riemannian 3-manifold. If Ω is a subset of M with boundary $\partial\Omega$ homeomorphic to S^2, and if $\partial\Omega$ has positive mean curvature H (with respect to some chosen normal direction), define

$$m(\Omega) = m(\partial\Omega, g_{T\partial\Omega}, H),$$

where $T\partial\Omega$ is the tangent bundle of $\partial\Omega$. If g has nonnegative scalar curvature, then $m(\Omega) \geq 0$ by Theorem 15.

Euclidean space. Consider $\mathbb{R}^3$ with the standard flat metric. Let $\Omega \subset \mathbb{R}^3$ be a strictly convex open set with smooth boundary that is not round, with mean curvature H_0 and induced metric γ_0. There is no valid fill-in of $(\partial\Omega, \gamma_0, H_0)$ with mean curvature $H > H_0$; this statement follows from the Shi–Tam inequality (3) or alternatively by Miao's positive mass theorem with corners [Miao 2002]. This implies $\lambda_0 = 1$, and so $m(\Omega) = 0$. The Brown–York mass of Ω also vanishes, as $H_0 = H$. A straightforward computation shows that the Hawking mass of Ω is strictly negative.

Schwarzschild, positive mass. Next let (M, g) be a Schwarzschild manifold of mass $m > 0$; see (9). Suppose $\Omega \subset M$ is topologically an open 3-ball with boundary Σ disjoint from the horizon.

Lemma 17. *For the Bartnik data induced on* $\partial\Omega$, $\lambda_0 = 1$. *Equivalently*, $m(\Omega) = 0$.

Figure 3 gives a depiction of the Bartnik data in question.

Proof. Certainly $\lambda_0 \geq 1$, since Ω is tautologically a valid fill-in. If $\lambda_0 > 1$, there exists a valid fill-in Ω' of Σ (with the same boundary metric and mean curvature), such that $\partial\Omega' \setminus \Sigma$ is nonempty and consists of minimal surfaces. Glue Ω' to $M \setminus \Omega$ along Σ, obtaining a manifold (M', g') that is smooth and has nonnegative scalar curvature away from Σ. Moreover, g' is Lipschitz across S, and $\partial M'$ consists of minimal surfaces (including the Schwarzschild horizon). Let A and A' be the minimum areas in the homology class of the boundary for the respective manifolds (M, g) and (M', g'). A is attained uniquely by the horizon S in M, and by a similar consideration A' is attained by a surface S' that includes S as a proper subset. (To see this, observe that the Schwarzschild manifold minus its horizon is foliated by the $\{r = \text{const.}\}$ spheres, which are convex; thus S' may not intersect the interior of $M' \setminus \Omega$ yet must intersect $M' \setminus \Omega$ to be homologous to the boundary of M'.) Thus

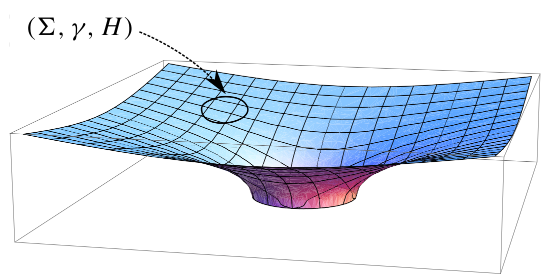


Figure 3. Off-center ball in Schwarzschild. The Bartnik data (Σ, γ, H) arises from the boundary of a small ball away from the horizon in a Schwarzschild manifold.

$A' > A$. By direct computation,

$$m = \sqrt{\frac{A}{16\pi}},$$

and so

$$m' < \sqrt{\frac{A'}{16\pi}}, \tag{13}$$

where $m' = m$ is the ADM mass of (M', g') (equal because g and g' agree outside a compact set). Using an argument similar to that of [Miao 2002], one can mollify (M', g') to a smooth, asymptotically flat metric of nonnegative scalar curvature and minimal boundary that gives strict inequality in (13). This violates the well-known Riemannian Penrose inequality [Huisken and Ilmanen 2001; Bray 2001]. This contradiction implies that $\lambda_0 = 1$, so $m(\Omega) = 0$. □

Thus, we have examples of Bartnik data of zero type that do not arise as the boundaries of regions in flat space. In other words, we have nonflat domains Ω for which $m(\Omega) = 0$. Of course, by Theorem 15, such Ω must be static vacuum (as is the case for the Schwarzschild metric).

To further extend this example, Huisken and Ilmanen [2001] show that there exist small balls Ω away from the horizon in the Schwarzschild manifold whose Bartnik data (Σ, γ, H) have strictly positive Hawking mass. Moreover, the case of equality in Theorem 10 shows that the Brown–York mass of Ω is also strictly positive. It follows that for $\lambda > 1$ sufficiently close to 1, the data $(\Sigma, \gamma, \lambda H)$ is of negative type, yet still has strictly positive m_H and m_{BY}.

Note that we have not stated Ω being static vacuum implies $m(\Omega) = 0$. Counterexamples are unknown to the author.

Schwarzschild, negative mass. Let (M, g) be the Schwarzschild metric of mass $m < 0$ (defined by (9) on $\mathbb{R}^3$ minus the closed ball of radius $|m|/2$). The Bartnik data induced on spheres $\{r = \text{const.}\}$ is of negative type because the critical parameter λ_r is less than one by (10).

5.1. *Example $m(\lambda)$ function.* Here we give an explicit computation of the inner mass function $m(\lambda)$ defined in Section 3.1 for Bartnik data $\mathcal{B}$ corresponding to the coordinate sphere S_r of radius $r > m/2$ in the Schwarzschild metric of mass $m > 0$, with induced metric γ and mean curvature H. The Riemannian Penrose inequality [Bray 2001] shows that[3] the Bartnik inner mass of $\mathcal{B}$ equals m. Let $\lambda > 0$; the data $(S_r, \gamma, \lambda H)$ embeds uniquely as a coordinate sphere $S_{r'}$ of some radius r' in a

[3] In a Schwarzschild manifold, $m = \sqrt{A/16\pi}$, where A is the area of the horizon. Now, $m_{\text{inner}}(\mathcal{B}) \geq m$ follows from the definition. If $m_{\text{inner}}(\mathcal{B}) > m$, there exists a fill-in of $\mathcal{B}$ with minimum area $A' > A$ attained by a minimal surface. One can then arrange a strict violation of the Penrose inequality by gluing the exterior Schwarzschild region of $\mathcal{B}$ to the fill-in. The gluing is

Schwarzschild metric of some mass m'. Equating the areas of S_r and $S_{r'}$ in their respective metrics, we have

$$4\pi r^2\left(1+\frac{m}{2r}\right)^4 = 4\pi (r')^2\left(1+\frac{m'}{2r'}\right)^4. \tag{14}$$

Equating λH with the mean curvature of $S_{r'}$ leads to

$$\lambda\left(\frac{2}{r}\left(1+\frac{m}{2r}\right)^{-2}-\frac{2m}{r^2}\left(1+\frac{m}{2r}\right)^{-3}\right)=\frac{2}{r'}\left(1+\frac{m'}{2r'}\right)^{-2}-\frac{2m'}{(r')^2}\left(1+\frac{m'}{2r'}\right)^{-3}. \tag{15}$$

With some calculations, one can compute r' and m' explicitly. For $\lambda \in (0, \lambda_0)$, we know $m(\lambda)$, the Bartnik inner mass of $S_{r'}$, simply equals m' (again, by the Riemannian Penrose inequality). Omitting some details, we give the formula:

$$m(\lambda) = \frac{r}{2}\left(\left(1+\frac{m}{2r}\right)^2 - \lambda^2\left(1-\frac{m}{2r}\right)^2\right). \tag{16}$$

As anticipated by Proposition 13, $m(\lambda)$ is continuous, decreasing, and $m(0) = \sqrt{A/16\pi}$, where A is the area of S_r in the Schwarzschild metric of mass m. Moreover, $m(\lambda)$ vanishes at the critical value $\lambda_0 = (1+m/2r)/(1-m/2r)$ (computed in Section 4), a property conjectured to hold in general (see the paragraph following Problem 2 in Section 7).

6. An algebraic operation on quasi-local mass functionals

For a quasi-local mass functional m_i (i.e., a map from the set of Bartnik data to the real numbers), define the following quantity in $[-\infty, \infty]$:

$$\lambda_i(\Sigma, \gamma, H) = \sup\{\lambda > 0 : m_i(\Sigma, \gamma, \lambda H) \geq 0\}.$$

In other words, λ_i measures how much one can scale the boundary mean curvature until the mass m_i becomes negative. Up to this point, we have studied this quantity for the case in which m_i is the Bartnik inner mass (since $m_{\text{inner}}(\Sigma, \gamma, H) \geq 0$ if and only if (Σ, γ, H) has a valid fill-in). Theorem 11 implies that λ_i is a positive, finite number for the case $m_i = m_{\text{inner}}$.

Here we use the number λ_i to construct an algebraic product of two quasi-local mass functionals, of which that constructed in Section 4 is a special case. We restrict to quasi-local mass functionals m_i satisfying the following mild assumptions on all Bartnik data:

(1) $\lambda_i(\Sigma, \gamma, H)$ is a positive real number, and

(2) $m_i(\Sigma, \gamma, \lambda H)$ is decreasing as a function of λ.

---

only Lipschitz across $\mathcal{B}$, but the smoothing and conformal techniques in [Miao 2002] can be used to produce a smooth example, leading to a contradiction.

For example the Hawking mass, Brown–York mass, and Bartnik inner mass (see Proposition 13) satisfy these properties.

Define the following binary operation on the set of quasi-local mass functionals. Given m_1 and m_2, let

$$(m_1 * m_2)(\Sigma, \gamma, H) = m_1\Big(\Sigma, \gamma, \frac{\lambda_1}{\lambda_2} H\Big), \tag{17}$$

where $\lambda_i = \lambda_i(\Sigma, \gamma, H)$ for $i = 1, 2$. This operation satisfies a number of properties.

Proposition 18. *Let m_1, m_2, and m_3 be quasi-local mass functionals.*

(1) $m_1 * m_1 = m_1$.

(2) $(m_1 * m_2) * m_3 = m_1 * m_3 = m_1 * (m_2 * m_3)$. *In particular*, $*$ *is associative.*

(3) *m_2 controls the sign of $m_1 * m_2$ in the following sense*:

   (a) $m_1 * m_2(\Sigma, \gamma, H) > 0$ *if and only if* $m_2(\Sigma, \gamma, H) > 0$.

   (b) $m_1 * m_2(\Sigma, \gamma, H) = 0$ *if and only if* $m_2(\Sigma, \gamma, H) = 0$.

(4) *If m_1 has the black hole limit property* (*see Theorem 15*), *so does $m_1 * m_2$.*

(5) *If both m_1 and m_2 produce the value m on concentric round spheres in the Schwarzschild metric of mass m, then so does $m_1 * m_2$.*

(6) *If $m_2 \le m_3$* (*as functions*), *then $m_1 * m_2 \le m_1 * m_3$.*

Sketch of proof. These properties all follow easily from the definitions, so we omit detailed proofs. We sketch some of the steps as a sample.

We compute $(m_1 * m_2) * m_3$. First, $(m_1 * m_2)(\Sigma, \gamma, H) = m_1\left(\Sigma, \gamma, \frac{\lambda_1}{\lambda_2} H\right)$ has critical parameter λ_2. Then

$$((m_1 * m_2) * m_3(\Sigma, \gamma, H)) = (m_1 * m_2)\left(\Sigma, \gamma, \frac{\lambda_2}{\lambda_3} H\right) = m_1\left(\Sigma, \gamma, \frac{\lambda_1}{\lambda_2}\frac{\lambda_2}{\lambda_3} H\right),$$

which equals $(m_1 * m_3)(\Sigma, \gamma, H)$.

We also demonstrate property (3a). Note

$$(m_1 * m_2)(\Sigma, \gamma, H) = m_1\left(\Sigma, \gamma, \frac{\lambda_1}{\lambda_2} H\right)$$

is positive if and only if $\lambda_1/\lambda_2 < \lambda_1$; that is, $\lambda_2 > 1$. However, $\lambda_2 > 1$ if and only if $m_2(\Sigma, \gamma, H) > 0$. □

6.1. ***Examples of $m_1 * m_2$.*** In this section, we demonstrate $m_1 * m_2$ generally does not equal $m_2 * m_1$.

Hawking mass and Bartnik inner mass. The quasi-local mass of Definition 14 is equal to $m_H * m_{\text{inner}}$, where m_H is the Hawking mass. To see this, note that

$$\lambda_H = \sqrt{\frac{16\pi}{\int_\Sigma H^2 dA_\gamma}}$$

and $\lambda_{\text{inner}} = \lambda_0$. Then, by definition,

$$m_H * m_{\text{inner}}(\Sigma, \gamma, H) = m_H\left(\Sigma, \gamma, \frac{\lambda_H}{\lambda_0} H\right) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(1 - \frac{1}{\lambda_0^2}\right).$$

We reiterate that $m_H * m_{\text{inner}}$ inherits the following property from m_{inner}: vanishing precisely on Bartnik data of zero type.

Hawking mass and Brown–York mass. To compute $m_H * m_{\text{BY}}$, we note that λ_H was found in the last example, and

$$\lambda_{\text{BY}} = \frac{\int_\Sigma H_0 dA_\gamma}{\int_\Sigma H dA_\gamma}.$$

Using the definition,

$$m_H * m_{\text{BY}}(\Sigma, \gamma, H) = \sqrt{\frac{|\Sigma|_\gamma}{16\pi}}\left(1 - \left(\frac{\int_\Sigma H dA_\gamma}{\int_\Sigma H_0 dA_\gamma}\right)^2\right).$$

This quasi-local mass was written down in a different context by Miao [2009].

Brown–York mass and Hawking mass. The steps from the last example show

$$m_{\text{BY}} * m_H(\Sigma, \gamma, H) = \int_\Sigma H_0 dA_\gamma \left(1 - \sqrt{\frac{\int_\Sigma H^2 dA_\gamma}{16\pi}}\right),$$

illustrating concretely the noncommutativity of $*$.

7. Concluding remarks and open problems

We conclude by mentioning some questions raised in this paper.

Problem 1. Determine whether the quasi-local mass of Definition 14 is monotone under some flow.

Monotonicity means that if $\{(\Sigma_t, \gamma_t, H_t)\}_{t\in[0,\epsilon)}$ is some family of surfaces (together with their Bartnik data) moving outward in a manifold of nonnegative scalar curvature, then $m(\Sigma_t, \gamma_t, H_t)$ is nondecreasing. Monotonicity is often (but not universally) suggested as a desirable property of quasi-local mass [Bartnik 2002].

Problem 2. Determine whether the Bartnik data $(\Sigma, \gamma, \lambda_0 H)$ is of zero type. Equivalently, construct a static vacuum fill-in of $(\Sigma, \gamma, \lambda_0 H)$.

That the two statements above are equivalent follows from Proposition 5 and Theorem 11. The precise nature of the Bartnik data rescaled with the critical parameter λ_0 is perhaps the biggest open question of this paper. An affirmative answer to Problem 2 would imply that $(\Sigma, \gamma, \lambda_0 H)$ admits a static vacuum fill-in. In general, constructing static vacuum metrics with prescribed boundary data is a very difficult problem (cf. the work of Anderson and Khuri [2011] on static vacuum asymptotically flat "extensions" of Bartnik data).

More generally, one could ask what happens to the geometry of the class valid fill-ins of $(\Sigma, \gamma, \lambda H)$ in the limit $\lambda \nearrow \lambda_0$. An optimistic conjecture would be that in the limit $\lambda \nearrow \lambda_0$, any valid fill-in $(\Omega_\lambda, g_\lambda)$ of $(\Sigma, \gamma, \lambda H)$ satisfies:

- the black holes (area-minimizing minimal surfaces) in $(\Omega_\lambda, g_\lambda)$ are shrinking to zero size (i.e., $\lim_{\lambda \to \lambda_0^-} m(\lambda) = 0$), and
- the metric g_λ is approaching a static vacuum metric in an appropriate sense.

There may be a connection between the first point and Miao's localized Riemannian Penrose inequality [Miao 2009].

The above discussion is basically a localization of the near-equality case of the positive mass theorem [Schoen and Yau 1979]. In such a global setting, the question is: what happens to the geometry of a sequence of asymptotically flat manifolds (M_i, g_i) of nonnegative scalar curvature whose total mass is approaching zero? The Riemannian Penrose inequality [Huisken and Ilmanen 2001; Bray 2001] shows that any black holes in (M_i, g_i) must be approaching zero, and some partial results exist for proving that g_i is approaching a flat metric [Bartnik 1997; Lee 2009; Bray and Finster 2002; Lee and Sormani 2011].

Appendix A. Conformal transformation of curvatures

We repeatedly used the following formulas that relate the scalar curvature and mean curvature of conformal metrics. Suppose g and $\bar{g}$ are Riemannian metrics on a 3-manifold for which

$$\bar{g} = u^4 g$$

for some smooth function $u > 0$. If R and $\bar{R}$ are the scalar curvatures of g and $\bar{g}$, then

$$\bar{R} = u^{-5}(-8\Delta u + Ru), \tag{18}$$

where Δ is the Laplacian with respect to g. Next, suppose S is a hypersurface with unit normal field ν with respect to g. Then the mean curvatures H and $\overline{H}$ (in the direction defined by ν) with respect to g and $\bar{g}$ satisfy

$$\overline{H} = u^{-2} H + 4u^{-3}\partial_\nu(u). \tag{19}$$

Appendix B. Geometric measure theory

Here is an extremely useful result from geometric measure theory on the existence and regularity of area-minimizing surfaces.

Theorem 19. *Let (M, g) be a smooth, compact Riemannian manifold of dimension $2 \leq n \leq 7$ with boundary ∂M. Suppose ∂M has positive mean curvature (i.e., its mean curvature vector points inward). Given a connected component S of ∂M, there exists a smooth, embedded hypersurface $\tilde{S}$ of zero mean curvature that minimizes area among surfaces homologous to S. Moreover, $\tilde{S}$ does not intersect ∂M.*

These results are essentially due to Federer and Fleming [1960; Fleming 1962; Federer 1970]. The rough idea of the proof of Theorem 19 is to take a minimizing sequence of surfaces $\{S_i\}$ (viewed as integral currents) in $[S]$, the homology class of S. By the Federer–Fleming compactness theorem, some subsequence converges to a surface $\tilde{S}$. Standard arguments show that $\tilde{S}$ remains in $[S]$ and indeed has the desired minimum of area. Regularity theory (requiring $n \leq 7$) proves that $\tilde{S}$ is a smooth, embedded hypersurface. By the first variation of area formula, $\tilde{S}$ has zero mean curvature and may not touch the positive mean curvature boundary (which acts as a barrier). See the appendix of [Schoen and Yau 1979] for a careful proof of the last fact.

Appendix C. Deformations of scalar curvature near a boundary

Here we prove the following useful lemma.

Lemma 20. *Suppose that (Σ, γ, H_1) admits a valid fill-in. If $0 < H_2 < H_1$, then (Σ, γ, H_2) admits a valid fill-in with positive scalar curvature at a point. In particular, (Σ, γ, H_2) is of positive type.*

Although we only prove the case $K_\gamma > 0$ here, Lemma 20 is true without this hypothesis. The proof is an application of techniques developed recently by Brendle, Marques, and Neves [2011].

Proof. Step 1: We construct a valid fill-in of (Σ, γ) with mean curvature strictly greater than H_2 and with positive scalar curvature in a neighborhood of Σ.

Since Σ is compact, we may choose $\alpha \in (0, 1)$ so that $\alpha H_1 > H_2$. We proved in step 2 of Theorem 11 that $(\Sigma, \gamma, \alpha H_1)$ is of positive type and moreover admits a valid fill-in (Ω, g_1) whose scalar curvature is strictly positive in a neighborhood U of Σ. (For the latter statement, refer to Equation (6) and note that $\rho'(0) > 0$.)

Step 2: We define a metric g_2 on Ω as follows, with the goal of making the boundary mean curvature of g_2 equal to H_2. First, consider a neighborhood of Σ contained in U that is diffeomorphic to $\Sigma \times (-t_0, 0]$ (where $t = 0$ corresponds to Σ). Define for $x \in \Sigma$ and $t \in (-t_0, 0]$:

$$g_2(x,t) = \rho(t)^2 dt^2 + (1 + t H_2(x))\gamma(x),$$

where $\rho(t)$ is a function satisfying $\rho(0) = 1$ and will be specified later. It is readily checked that g_2 induces on Σ the metric γ and mean curvature H_2. Shrinking t_0 if necessary and choosing $\rho(t)$ bounded below by a positive constant with $\rho'(t) > 0$ sufficiently large, we may arrange g_2 to have strictly positive scalar curvature on $\Sigma \times (-t_0, 0]$. This is readily checked using (5). Now, extend g_2 arbitrarily to a smooth metric on Ω (not necessarily preserving nonnegative scalar curvature). Replace U with the smaller neighborhood $\Sigma \times (-t_0, 0]$

To summarize, we have two metrics g_1 and g_2 on the compact manifold Ω, inducing boundary data $(\Sigma, \gamma, \alpha H_1)$ and (Σ, γ, H_2), respectively, each with positive scalar curvature on the neighborhood U of Σ. By compactness, the scalar curvatures of $g_1|_U$ and $g_2|_U$ are bounded below by a constant $R_0 > 0$.

Step 3: Apply Theorem 5 of [Brendle et al. 2011] to produce a metric $\hat{g}$ on Ω satisfying the following properties[4]:

(i) $R_{\hat{g}}(x) \geq \min\{R_{g_1}(x), R_{g_2}(x)\} - R_0/2$.

(ii) $\hat{g}$ agrees with g_1 outside of U.

(iii) $\hat{g}$ agrees with g_2 in some neighborhood of Σ.

(To apply the theorem, it is crucial that $\alpha H_1 > H_2$.)

By the third condition, $(\Omega, \hat{g})$ is a fill-in of (Σ, γ, H_2). By the first and second conditions, $\hat{g}$ has nonnegative (but not identically zero) scalar curvature and $\partial\Omega \setminus \Sigma$ (if nonempty) is a minimal surface. In particular, $(\Omega, \hat{g})$ is a valid fill-in with positive scalar curvature at some point.

Finally, the last statement in the lemma follows from Proposition 7. □

Acknowledgements

The author is indebted to Hubert Bray for suggesting the main idea of varying the parameter λ. He would like to thank the referee for carefully reading the work and offering a number of thoughtful suggestions.

References

[Anderson 2010] M. Anderson, "Quasilocal Hamiltonians in general relativity", *Phys. Rev. D* **82**:8 (2010), 1–6.

[Anderson and Khuri 2011] M. Anderson and M. Khuri, "The static extension problem in general relativity", preprint, 2011. arXiv 0909.4550

[4]Due to the local nature of the construction, it is clear that we can ignore any connected components of $\partial\Omega$ that are not Σ.

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 99i:58001 Zbl 0896.53003

[Bartnik 1989] R. Bartnik, "A new definition of quasi-local mass", pp. 399–401 in *Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity, Part A, B* (Perth, 1988), edited by D. G. Blair et al., World Science, Teaneck, NJ, 1989. MR 91j:83032

[Bartnik 1993] R. Bartnik, "Quasi-spherical metrics and prescribed scalar curvature", *J. Differential Geom.* **37**:1 (1993), 31–71. MR 93i:53041 Zbl 0786.53019

[Bartnik 1997] R. Bartnik, "Energy in general relativity", pp. 5–27 in *Tsing Hua lectures on geometry & analysis* (Hsinchu, 1990–1991), edited by S.-T. Yau, International Press, Cambridge, MA, 1997. MR 99a:83028 Zbl 0884.53065

[Bartnik 2002] R. Bartnik, "Mass and 3-metrics of non-negative scalar curvature", pp. 231–240 in *Proceedings of the International Congress of Mathematicians, II* (Beijing, 2002), edited by T. Li, Higher Ed. Press, Beijing, 2002. MR 2003k:53034 Zbl 1009.53055

[Bray 2001] H. L. Bray, "Proof of the Riemannian Penrose inequality using the positive mass theorem", *J. Differential Geom.* **59**:2 (2001), 177–267. MR 2004j:53046 Zbl 1039.53034

[Bray and Finster 2002] H. Bray and F. Finster, "Curvature estimates and the positive mass theorem", *Comm. Anal. Geom.* **10**:2 (2002), 291–306. MR 2003c:53047 Zbl 1030.53041

[Brendle et al. 2011] S. Brendle, F. C. Marques, and A. Neves, "Deformations of the hemisphere that increase scalar curvature", *Invent. Math.* **185**:1 (2011), 175–197. MR 2012h:53094 Zbl 1227.53048

[Bunting and Masood-ul Alam 1987] G. L. Bunting and A. K. M. Masood-ul Alam, "Nonexistence of multiple black holes in asymptotically Euclidean static vacuum space-time", *Gen. Relativity Gravitation* **19**:2 (1987), 147–154. MR 88e:83031 Zbl 0615.53055

[Corvino 2000] J. Corvino, "Scalar curvature deformation and a gluing construction for the Einstein constraint equations", *Comm. Math. Phys.* **214**:1 (2000), 137–189. MR 2002b:53050 Zbl 1031.53064

[Fan et al. 2009] X.-Q. Fan, Y. Shi, and L.-F. Tam, "Large-sphere and small-sphere limits of the Brown–York mass", *Comm. Anal. Geom.* **17**:1 (2009), 37–72. MR 2010e:53132 Zbl 1175.53083

[Federer 1970] H. Federer, "The singular sets of area minimizing rectifiable currents with codimension one and of area minimizing flat chains modulo two with arbitrary codimension", *Bull. Amer. Math. Soc.* **76** (1970), 767–771. MR 41 #5601 Zbl 0194.35803

[Federer and Fleming 1960] H. Federer and W. H. Fleming, "Normal and integral currents", *Ann. of Math.* (2) **72** (1960), 458–520. MR 23 #A588 Zbl 0187.31301

[Fleming 1962] W. H. Fleming, "On the oriented Plateau problem", *Rend. Circ. Mat. Palermo* (2) **11** (1962), 69–90. MR 28 #499 Zbl 0107.31304

[Huisken and Ilmanen 2001] G. Huisken and T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose inequality", *J. Differential Geom.* **59**:3 (2001), 353–437. MR 2003h:53091 Zbl 1055.53052

[Lee 2009] D. A. Lee, "On the near-equality case of the positive mass theorem", *Duke Math. J.* **148**:1 (2009), 63–80. MR 2010f:53054 Zbl 1168.53018

[Lee and Sormani 2011] D. A. Lee and C. Sormani, "Stability of the Positive Mass Theorem for rotationally symmetric Riemannian manifolds", preprint, 2011. arXiv 1104.2657

[Miao 2002] P. Miao, "Positive mass theorem on manifolds admitting corners along a hypersurface", *Adv. Theor. Math. Phys.* **6**:6 (2002), 1163–1182. MR 2005a:53065

[Miao 2009] P. Miao, "On a localized Riemannian Penrose inequality", *Comm. Math. Phys.* **292**:1 (2009), 271–284. MR 2010h:53046 Zbl 1182.53064

[Penrose 1982] R. Penrose, "Some unsolved problems in classical general relativity", pp. 631–668 in *Seminar on Differential Geometry*, edited by S. T. Yau, Annals of Mathematics Studies **102**, Princeton University Press, 1982. MR 83c:83001 Zbl 0481.53053

[Pólya and Szegö 1951] G. Pólya and G. Szegö, *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies **27**, Princeton University Press, 1951. MR 13,270d Zbl 0044.38301

[Schoen and Yau 1979] R. Schoen and S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Comm. Math. Phys.* **65**:1 (1979), 45–76. MR 80j:83024 Zbl 0405.53045

[Shi and Tam 2002] Y. Shi and L.-F. Tam, "Positive mass theorem and the boundary behaviors of compact manifolds with nonnegative scalar curvature", *J. Differential Geom.* **62**:1 (2002), 79–125. MR 2005b:53046 Zbl 1071.53018

[Szabados 2009] L. Szabados, "Quasi-local energy-momentum and angular momentum in general relativity", *Living Rev. Relativity* **12**:4 (2009), 1–163. Zbl 1215.83010

[Wang and Yau 2009] M.-T. Wang and S.-T. Yau, "Isometric embeddings into the Minkowski space and new quasi-local mass", *Comm. Math. Phys.* **288**:3 (2009), 919–942. MR 2010d:53077 Zbl 1195.53039

Received November 28, 2011. Revised October 12, 2012.

JEFFREY L. JAUREGUI
DAVID RITTENHOUSE LAB
209 SOUTH 33RD STREET
PHILADELPHIA, PA 19104
UNITED STATES
jauregui@math.upenn.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.445

OPERATOR ALGEBRAS AND CONJUGACY PROBLEM FOR THE PSEUDO-ANOSOV AUTOMORPHISMS OF A SURFACE

IGOR NIKOLAEV

In memory of W. P. Thurston

The conjugacy problem for the pseudo-Anosov automorphisms of a compact surface is studied. To each pseudo-Anosov automorphism ϕ, we assign an AF C^*-algebra $\mathbb{A}_\phi$ (an operator algebra). It is proved that the assignment is functorial, i.e., every ϕ', conjugate to ϕ, maps to an AF C^*-algebra $\mathbb{A}_{\phi'}$, which is stably isomorphic to $\mathbb{A}_\phi$. The new invariants of the conjugacy of the pseudo-Anosov automorphisms are obtained from the known invariants of the stable isomorphisms of the AF C^*-algebras. Namely, the main invariant is a triple $(\Lambda, [I], K)$, where Λ is an order in the ring of integers in a real algebraic number field K and $[I]$ an equivalence class of the ideals in Λ. The numerical invariants include the determinant Δ and the signature Σ, which we compute for the case of the Anosov automorphisms. A question concerning the p-adic invariants of the pseudo-Anosov automorphism is formulated.

Introduction

A. ***Conjugacy problem.*** Let $\mathrm{Mod}(X)$ be the mapping class group of a compact surface X, i.e., the group of orientation preserving automorphisms of X modulo the trivial ones. Recall that $\phi, \phi' \in \mathrm{Mod}(X)$ are conjugate automorphisms whenever $\phi' = h \circ \phi \circ h^{-1}$ for an $h \in \mathrm{Mod}(X)$. It is not hard to see that conjugation is an equivalence relation which splits the mapping class group into disjoint classes of conjugate automorphisms. The construction of invariants of the conjugacy

Partially supported by NSERC.
MSC2010: primary 46L85; secondary 57M27.
Keywords: mapping class group, AF C^*-algebras.

classes in $\mathrm{Mod}(X)$ is an important and difficult problem studied by Hemion [1979], Mosher [1986], and others. Any knowledge of such invariants leads to a topological classification of three-dimensional manifolds, which fiber over the circle with monodromy $\phi \in \mathrm{Mod}(X)$ [Thurston 1982].

B. *Pseudo-Anosov automorphisms.* It is known that any $\phi \in \mathrm{Mod}(X)$ is isotopic to an automorphism ϕ', such that either (i) ϕ' has a finite order, or (ii) ϕ' is a pseudo-Anosov (aperiodic) automorphism, or else (iii) ϕ' is reducible by a system of curves Γ surrounded by the small tubular neighborhoods $N(\Gamma)$, such that on $X \setminus N(\Gamma)$, ϕ' satisfies either (i) or (ii). Let ϕ be a representative of the equivalence class of a pseudo-Anosov automorphism. Then there exist a pair consisting of the stable $\mathcal{F}_s$ and unstable $\mathcal{F}_u$ mutually orthogonal measured foliations on the surface X, such that $\phi(\mathcal{F}_s) = (1/\lambda_\phi)\mathcal{F}_s$ and $\phi(\mathcal{F}_u) = \lambda_\phi \mathcal{F}_u$, where $\lambda_\phi > 1$ is called a dilatation of ϕ. The foliations $\mathcal{F}_s$, $\mathcal{F}_u$ are minimal, uniquely ergodic and describe the automorphism ϕ up to a power. In the sequel, we shall focus on the conjugacy problem for the pseudo-Anosov automorphisms of a surface X.

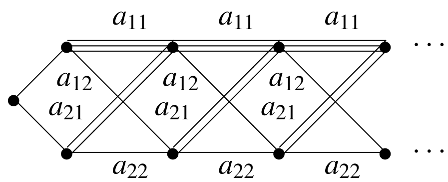
C. *AF C*-algebras.* A C^*-algebra is an algebra $\mathbb{A}$ over $\mathbb{C}$ with a norm $a \mapsto \|a\|$ and an involution $a \mapsto a^*$ such that it is complete with respect to the norm and $\|ab\| \le \|a\|\|b\|$ and $\|a^*a\| = \|a^2\|$ for all $a, b \in \mathbb{A}$. The C^*-algebras have been introduced by Murray and von Neumann as rings of bounded operators on a Hilbert space and are strongly connected with the geometry and topology of manifolds [Blackadar 1986, Section 24]. Any simple finite-dimensional C^*-algebra is isomorphic to the algebra $M_n(\mathbb{C})$ of the complex $n \times n$ matrices. A natural completion of the finite-dimensional semisimple C^*-algebras (as $n \to \infty$) is known as an *AF C^*-algebra* [Effros 1981]. An AF C^*-algebra is most conveniently given by an infinite graph, which records the inclusion of the finite-dimensional subalgebras into the AF C^*-algebra. The graph is called a *Bratteli diagram*. When the diagram is periodic, the AF C^*-algebra is *stationary*; this is an important special case. In addition to the usual isomorphism $\cong$, the C^*-algebras $\mathbb{A}$, $\mathbb{A}'$ are called *stably isomorphic* whenever $\mathbb{A} \otimes \mathcal{K} \cong \mathbb{A}' \otimes \mathcal{K}$, where $\mathcal{K}$ is the C^*-algebra of compact operators.

D. *Motivation.* Let $\phi \in \mathrm{Mod}(X)$ be a pseudo-Anosov automorphism. The main idea of the present paper is to assign to ϕ an AF C^*-algebra, $\mathbb{A}_\phi$, so that for every $h \in \mathrm{Mod}(X)$ the following diagram commutes:

$$\begin{array}{ccc}
\phi & \xrightarrow{\text{conjugacy}} & \phi' = h \circ \phi \circ h^{-1} \\
\downarrow & & \downarrow \\
\mathbb{A}_\phi & \xrightarrow[\text{isomorphism}]{\text{stable}} & \mathbb{A}_{\phi'}
\end{array}$$

(In other words, if ϕ and ϕ' are conjugate pseudo-Anosov automorphisms, then the AF C^*-algebras $\mathbb{A}_\phi$ and $\mathbb{A}_{\phi'}$ are stably isomorphic.) For the sake of clarity, we shall consider an example illustrating the idea in the case $X = T^2$ (a torus).

E. ***Model example.*** Let $\phi \in \mathrm{Mod}(T^2)$ be the Anosov automorphism given by a nonnegative matrix $A_\phi \in \mathrm{SL}_2(\mathbb{Z})$. (The assumption is not restrictive; each A_ϕ with $\mathrm{Tr}(A_\phi) > 0$ is similar to a nonnegative matrix. The case $\mathrm{Tr}(A_\phi) < 0$ is treated likewise — by reduction to a nonpositive matrix; then the absolute value of all entries must be taken.) Consider a stationary AF C^*-algebra, $\mathbb{A}_\phi$, given by the following periodic Bratteli diagram:



$$A_\phi = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

Figure 1. The AF C^*-algebra $\mathbb{A}_\phi$.

where a_{ij} indicate the multiplicity of the respective edges of the graph. We encourage the reader to verify that $F : \phi \mapsto \mathbb{A}_\phi$ is a well-defined function on the set of Anosov automorphisms given by the hyperbolic matrices with nonnegative entries. Let us show that if $\phi, \phi' \in \mathrm{Mod}(T^2)$ are conjugate Anosov automorphisms, then $\mathbb{A}_\phi$, $\mathbb{A}_{\phi'}$ are stably isomorphic AF C^*-algebras. Indeed, let $\phi' = h \circ \phi \circ h^{-1}$ for an $h \in \mathrm{Mod}(X)$. Then $A_{\phi'} = T A_\phi T^{-1}$ for a matrix $T \in \mathrm{SL}_2(\mathbb{Z})$. Note that

$$(A'_\phi)^n = (T A_\phi T^{-1})^n = T A_\phi^n T^{-1},$$

where $n \in \mathbb{N}$. We shall use the following criterion: the AF C^*-algebras $\mathbb{A}$, $\mathbb{A}'$ are stably isomorphic if and only if their Bratteli diagrams contain a common block of an arbitrary length (compare with [Effros 1981, Theorem 2.3]; recall that an order-isomorphism mentioned in the theorem is equivalent to the condition that the corresponding Bratteli diagrams have the same infinite tails — i.e., a common block of infinite length). Consider two sequences of matrices:

$$\underbrace{A_\phi A_\phi \cdots A_\phi}_{n},$$

which mimics the Bratteli diagram of $\mathbb{A}_\phi$, and

$$T \underbrace{A_\phi A_\phi \cdots A_\phi}_{n} T^{-1},$$

which mimics that of $\mathbb{A}_{\phi'}$. Letting $n \to \infty$, we conclude that $\mathbb{A}_\phi \otimes \mathcal{K} \cong \mathbb{A}_{\phi'} \otimes \mathcal{K}$.

F. ***Invariants of torus automorphisms obtained from the operator algebras.*** The conjugacy problem for the Anosov automorphisms can now be recast in terms of AF C^*-algebras: find invariants of stable isomorphism classes of the stationary AF C^*-algebras. One such invariant is due to Handelman [1981]. Consider an eigenvalue problem for the hyperbolic matrix $A_\phi \in SL_2(\mathbb{Z})$: $A_\phi v_A = \lambda_A v_A$, where $\lambda_A > 1$ is the Perron–Frobenius eigenvalue and $v_A = (v_A^{(1)}, v_A^{(2)})$ the corresponding eigenvector with the positive entries normalized so that $v_A^{(i)} \in K = \mathbb{Q}(\lambda_A)$. Denote by $\mathfrak{m} = \mathbb{Z}v_A^{(1)} + \mathbb{Z}v_A^{(2)}$ the $\mathbb{Z}$-module in the number field K. Recall that the coefficient ring, Λ, of module $\mathfrak{m}$ consists of the elements $\alpha \in K$ such that $\alpha\mathfrak{m} \subseteq \mathfrak{m}$. It is known that Λ is an order in K (i.e., a subring of K containing 1) and, with no restriction, one can assume that $\mathfrak{m} \subseteq \Lambda$. It follows from the definition that $\mathfrak{m}$ coincides with an ideal, I, whose equivalence class in Λ we shall denote by $[I]$. It has been proved by Handelman that the triple $(\Lambda, [I], K)$ is an arithmetic invariant of the stable isomorphism class of $\mathbb{A}_\phi$: the $\mathbb{A}_\phi$, $\mathbb{A}_{\phi'}$ are stably isomorphic AF C^*-algebras if and only if $\Lambda = \Lambda'$, $[I] = [I']$ and $K = K'$. It is interesting to compare the operator algebra invariants with the matrix invariants obtained in [Latimer and MacDuffee 1933] and [Wallace 1984].

G. ***AF C*-algebra $\mathbb{A}_\phi$ (pseudo-Anosov case).*** Denote by $\mathcal{F}_\phi$ the stable foliation of a pseudo-Anosov automorphism $\phi \in \mathrm{Mod}(X)$. For brevity, we assume that $\mathcal{F}_\phi$ is an oriented foliation given by the trajectories of a closed 1-form $\omega \in H^1(X; \mathbb{R})$. Let $v^{(i)} = \int_{\gamma_i} \omega$, where $\{\gamma_1, \dots, \gamma_n\}$ is a basis in the relative homology $H_1(X, \mathrm{Sing}\,\mathcal{F}_\phi; \mathbb{Z})$, such that $\theta = (\theta_1, \dots, \theta_{n-1})$ is a vector with positive coordinates $\theta_i = v^{(i+1)}/v^{(1)}$. (Note that the θ_i depend on a basis in the homology group, but a $\mathbb{Z}$-module generated by the θ_i does not — see Lemma 5.) Consider the (infinite) Jacobi–Perron continued fraction [Bernstein 1971] of θ:

$$\begin{pmatrix}1\\ \theta\end{pmatrix} = \lim_{k\to\infty} \begin{pmatrix}0 & 1\\ I & b_1\end{pmatrix} \cdots \begin{pmatrix}0 & 1\\ I & b_k\end{pmatrix} \begin{pmatrix}0\\ \mathbb{I}\end{pmatrix},$$

where $b_i = (b_1^{(i)}, \dots, b_{n-1}^{(i)})^T$ is a vector of nonnegative integers, I the unit matrix and $\mathbb{I} = (0, \dots, 0, 1)^T$. By definition, $\mathbb{A}_\phi$ is an (isomorphism class of the) AF C^*-algebra given by the Bratteli diagram whose incidence matrices coincide with $B_k = \left(\begin{smallmatrix}0 & 1\\ I & b_k\end{smallmatrix}\right)$ for $k = 1, \dots, \infty$. Note that this yields the Bratteli diagram derived in the model example (the Anosov case).

H. ***Main results.*** For a matrix $A \in GL_n(\mathbb{Z})$ with positive entries, we denote by λ_A the Perron–Frobenius eigenvalue and let $(v_A^{(1)}, \dots, v_A^{(n)})$ denote the corresponding normalized eigenvector with $v_A^{(i)} \in K = \mathbb{Q}(\lambda_A)$. The coefficient (endomorphism) ring of the module $\mathfrak{m} = \mathbb{Z}v_A^{(1)} + \cdots + \mathbb{Z}v_A^{(n)}$ will be denoted by Λ. The equivalence class of ideal I in Λ will be denoted $[I]$. Finally, we denote by $\Delta = \mathrm{Det}(a_{ij})$ and Σ

the determinant and signature of the symmetric bilinear form $q(x, y) = \sum_{i,j}^{n} a_{ij} x_i x_j$, where $a_{ij} = \mathrm{Tr}(v_A^{(i)} v_A^{(j)})$, with $\mathrm{Tr}(\cdot)$ the trace function. Our main results can be expressed as follows.

Theorem 1. *$\mathbb{A}_\phi$ is a stationary AF C^*-algebra.*

Let Φ be a category of all pseudo-Anosov (Anosov, respectively) automorphisms of a surface of the genus $g \geq 2$ ($g = 1$, respectively); the arrows (morphisms) are conjugations between the automorphisms. Likewise, let $\mathscr{A}$ be the category of all stationary AF C^*-algebras $\mathbb{A}_\phi$, where ϕ runs over the set Φ; the arrows of $\mathscr{A}$ are stable isomorphisms among the algebras $\mathbb{A}_\phi$.

Theorem 2. *Let $F : \Phi \to \mathscr{A}$ be a map given by the formula $\phi \mapsto \mathbb{A}_\phi$. Then*:

(i) *F is a functor*; *it maps conjugate pseudo-Anosov automorphisms to stably isomorphic AF C^*-algebras.*

(ii) Ker $F = [\phi]$, *where* $[\phi] = \{\phi' \in \Phi \mid (\phi')^m = \phi^n, m, n \in \mathbb{N}\}$ *is the commensurability class of the pseudo-Anoov automorphism ϕ.*

Corollary 3. *The triple $(\Lambda, [I], K)$ and the integers Δ and Σ are invariants of the conjugacy classes of the pseudo-Anosov automorphisms.*

I. ***How can the invariants $(\Lambda, [I], K)$, Δ and Σ be calculated?*** There is no easy way; the problem is comparable to that of numerical invariants of the fundamental group of a knot. A step in this direction would be computation of the matrix A; the latter is similar to the matrix $\rho(\phi)$, where $\rho : \mathrm{Mod}(X) \to \mathrm{PIL}$ is a faithful representation of the mapping class group as a group of the piecewise-integral-linear transformations [Penner 1984, p. 45]. The entries of $\rho(\phi)$ are the linear combinations of the Dehn twists along the $(3g - 1)$ (Lickorish) curves on the surface X. Then one can effectively determine whether $\rho(\phi)$ and A are similar matrices (over $\mathbb{Z}$) by bringing the polynomial matrices $\rho(\phi) - xI$ and $A - xI$ to the Smith normal form; when the similarity is established, the numerical invariants Δ and Σ become the polynomials in the Dehn twists. A tabulation of the simplest elements of $\mathrm{Mod}(X)$ is possible in terms of Δ and Σ (see the Examples section, page 459); however, this task lies beyond the scope of present paper.

J. ***Structure of the paper.*** Proofs of the main results can be found in Section 3. Sections 1 and 2 consist of lemmas used to prove the main results. Section 4 includes some examples, open problems and conjectures. Since the paper does not include a formal section on the preliminaries, we encourage the reader to consult [Blackadar 1986; Effros 1981; Krieger 1980] (operator algebras and dynamics), [Hubbard and Masur 1979; Thurston 1988] (measured foliations) and [Bernstein 1971; Perron 1907] (Jacobi–Perron continued fractions).

1. The jacobian of a measured foliation

Let $\mathcal{F}$ be a measured foliation on a compact surface X [Thurston 1988]. For the sake of brevity, we shall always assume that $\mathcal{F}$ is an oriented foliation, i.e., given by the trajectories of a closed 1-form ω on X. (The assumption is not a restriction; by [Hubbard and Masur 1979], every measured foliation is oriented on a double cover $\widetilde{X}$ of X ramified at the singular points of the half-integer index of the nonoriented foliation.) Let $\{\gamma_1, \dots, \gamma_n\}$ be a basis in the relative homology group $H_1(X, \operatorname{Sing}\mathcal{F}; \mathbb{Z})$, where $\operatorname{Sing}\mathcal{F}$ is the set of singular points of the foliation $\mathcal{F}$. It is well known that $n = 2g + m - 1$, where g is the genus of X and $m = |\operatorname{Sing}(\mathcal{F})|$. The periods of ω in this basis will be written

$$\lambda_i = \int_{\gamma_i} \omega.$$

The real numbers λ_i are coordinates of $\mathcal{F}$ in the space of all measured foliations on X (with the fixed set of singular points) [Douady and Hubbard 1975].

Definition 4. By the jacobian $\operatorname{Jac}(\mathcal{F})$ of the measured foliation $\mathcal{F}$, we understand the $\mathbb{Z}$-module $\mathfrak{m} = \mathbb{Z}\lambda_1 + \cdots + \mathbb{Z}\lambda_n$ regarded as a subset of the real line $\mathbb{R}$.

The importance of the jacobian stems from the observation that although the periods, λ_i, depend on the choice of a basis in $H_1(X, \operatorname{Sing}\mathcal{F}; \mathbb{Z})$, the jacobian does not. Moreover, up to a scalar multiple, the jacobian is an invariant of the equivalence class of the foliation $\mathcal{F}$. We formalize these observations in the following two results.

Lemma 5 (invariance of the jacobian). *The $\mathbb{Z}$-module $\mathfrak{m}$ is independent of the choice of a basis in $H_1(X, \operatorname{Sing}\mathcal{F}; \mathbb{Z})$ and depends solely on the foliation $\mathcal{F}$.*

Proof. Indeed, let $A = (a_{ij}) \in GL_n(\mathbb{Z})$ and let

$$\gamma_i' = \sum_{j=1}^{n} a_{ij}\gamma_j$$

be a new basis in $H_1(X, \operatorname{Sing}\mathcal{F}; \mathbb{Z})$. Then using the integration rules,

$$\lambda_i' = \int_{\gamma_i'} \omega = \int_{\sum_{j=1}^{n} a_{ij}\gamma_j} \omega = \sum_{j=1}^{n} \int_{\gamma_j} \omega = \sum_{j=1}^{n} a_{ij}\lambda_j.$$

To prove that $\mathfrak{m} = \mathfrak{m}'$, consider the following equations:

$$\mathfrak{m}' = \sum_{i=1}^{n} \mathbb{Z}\lambda_i' = \sum_{i=1}^{n} \mathbb{Z} \sum_{j=1}^{n} a_{ij}\lambda_j = \sum_{j=1}^{n} \Big(\sum_{i=1}^{n} a_{ij}\mathbb{Z} \Big) \lambda_j \subseteq \mathfrak{m}.$$

Let $A^{-1} = (b_{ij}) \in \mathrm{GL}_n(\mathbb{Z})$ be an inverse to the matrix A. Then $\lambda_i = \sum_{j=1}^{n} b_{ij}\lambda'_j$ and

$$\mathfrak{m} = \sum_{i=1}^{n} \mathbb{Z}\lambda_i = \sum_{i=1}^{n} \mathbb{Z} \sum_{j=1}^{n} b_{ij}\lambda'_j = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} b_{ij}\mathbb{Z}\right)\lambda'_j \subseteq \mathfrak{m}'.$$

Since both $\mathfrak{m}' \subseteq \mathfrak{m}$ and $\mathfrak{m} \subseteq \mathfrak{m}'$, we conclude that $\mathfrak{m}' = \mathfrak{m}$. Lemma 5 follows. □

Now recall that two measured foliations $\mathcal{F}$ and $\mathcal{F}'$ are *equivalent* if there exists an automorphism $h \in \mathrm{Mod}(X)$ that sends the leaves of the foliation $\mathcal{F}$ to the leaves of the foliation $\mathcal{F}'$. This equivalence deals with topological foliations, i.e., projective classes of measured foliations; see [Thurston 1988] for an explanation.

Lemma 6 (projective invariance). *Let $\mathcal{F}$, $\mathcal{F}'$ be the equivalent measured foliations on a surface X. Then*

$$\mathrm{Jac}(\mathcal{F}') = \mu\, \mathrm{Jac}(\mathcal{F}),$$

where $\mu > 0$ is a real number.

Proof. Let $h : X \to X$ be an automorphism of the surface X. Denote by h_* its action on $H_1(X, \mathrm{Sing}(\mathcal{F}); \mathbb{Z})$ and by h^* on $H^1(X; \mathbb{R})$ connected by the formula

$$\int_{h_*(\gamma)} \omega = \int_{\gamma} h^*(\omega), \qquad \text{for all } \gamma \in H_1(X, \mathrm{Sing}(\mathcal{F}); \mathbb{Z}) \text{ and } \omega \in H^1(X; \mathbb{R}).$$

Let $\omega, \omega' \in H^1(X; \mathbb{R})$ be the closed 1-forms whose trajectories define the foliations $\mathcal{F}$ and $\mathcal{F}'$, respectively. Since $\mathcal{F}$, $\mathcal{F}'$ are equivalent measured foliations,

$$\omega' = \mu h^*(\omega)$$

for a $\mu > 0$.

Let $\mathrm{Jac}(\mathcal{F}) = \mathbb{Z}\lambda_1 + \cdots + \mathbb{Z}\lambda_n$ and $\mathrm{Jac}(\mathcal{F}') = \mathbb{Z}\lambda'_1 + \cdots + \mathbb{Z}\lambda'_n$. Then

$$\lambda'_i = \int_{\gamma_i} \omega' = \mu \int_{\gamma_i} h^*(\omega) = \mu \int_{h_*(\gamma_i)} \omega, \qquad 1 \le i \le n.$$

By Lemma 5, we have

$$\mathrm{Jac}(\mathcal{F}) = \sum_{i=1}^{n} \mathbb{Z} \int_{\gamma_i} \omega = \sum_{i=1}^{n} \mathbb{Z} \int_{h_*(\gamma_i)} \omega.$$

Therefore

$$\mathrm{Jac}(\mathcal{F}') = \sum_{i=1}^{n} \mathbb{Z} \int_{\gamma_i} \omega' = \mu \sum_{i=1}^{n} \mathbb{Z} \int_{h_*(\gamma_i)} \omega = \mu\, \mathrm{Jac}(\mathcal{F}).$$

Lemma 6 follows. □

2. Equivalent foliations are stably isomorphic

Let $\mathcal{F}$ be a measured foliation on the surface X. We introduce an AF C^*-algebra, $\mathbb{A}_{\mathcal{F}}$, corresponding to the foliation $\mathcal{F}$ as explained in Section G of the Introduction (for the foliation $\mathcal{F}_\phi$). The goal of this section is to prove the commutativity of the following diagram:

$$\begin{array}{ccc} \mathcal{F} & \xrightarrow{\text{equivalence}} & \mathcal{F}' \\ \downarrow & & \downarrow \\ \mathbb{A}_{\mathcal{F}} & \xrightarrow[\text{isomorphism}]{\text{stable}} & \mathbb{A}_{\mathcal{F}'} \end{array}$$

We start with a simple property of Jacobi–Perron fractions [Bernstein 1971].

Lemma 7 (modules and continued fractions). *Let $\mathfrak{m} = \mathbb{Z}\lambda_1 + \cdots + \mathbb{Z}\lambda_n$ and $\mathfrak{m}' = \mathbb{Z}\lambda_1' + \cdots + \mathbb{Z}\lambda_n'$ be two $\mathbb{Z}$-modules, such that $\mathfrak{m}' = \mu\mathfrak{m}$ for a $\mu > 0$. Then the Jacobi–Perron continued fractions of the vectors λ and λ' coincide except, possibly, at a finite number of terms.*

Proof. Let $\mathfrak{m} = \mathbb{Z}\lambda_1 + \cdots + \mathbb{Z}\lambda_n$ and $\mathfrak{m}' = \mathbb{Z}\lambda_1' + \cdots + \mathbb{Z}\lambda_n'$. Since $\mathfrak{m}' = \mu\mathfrak{m}$, where μ is a positive real, one gets the following identity of the $\mathbb{Z}$-modules:

$$\mathbb{Z}\lambda_1' + \cdots + \mathbb{Z}\lambda_n' = \mathbb{Z}(\mu\lambda_1) + \cdots + \mathbb{Z}(\mu\lambda_n).$$

One can always assume that λ_i and λ_i' are positive reals. For obvious reasons, there exists a basis $\{\lambda_1'', \ldots, \lambda_n''\}$ of the module $\mathfrak{m}'$, such that

$$\begin{cases} \lambda'' = A(\mu\lambda), \\ \lambda'' = A'\lambda', \end{cases}$$

where $A, A' \in \mathrm{GL}_n^+(\mathbb{Z})$ are the matrices whose entries are nonnegative integers. In view of Proposition 3 of [Bauer 1996], we have

$$A = \begin{pmatrix} 0 & 1 \\ I & b_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_k \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} 0 & 1 \\ I & b_1' \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_l' \end{pmatrix},$$

where b_i, b_i' are nonnegative integer vectors. Since the (Jacobi–Perron) continued fraction for the vectors λ and $\mu\lambda$ coincide for any $\mu > 0$ [Bernstein 1971], we conclude that

$$\begin{pmatrix} 1 \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ I & b_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_k \end{pmatrix} \begin{pmatrix} 0 & 1 \\ I & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ I & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ \vdots \end{pmatrix},$$

$$\begin{pmatrix} 1 \\ \theta' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ I & b_1' \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_l' \end{pmatrix} \begin{pmatrix} 0 & 1 \\ I & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ I & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ \vdots \end{pmatrix},$$

where

$$\begin{pmatrix} 1 \\ \theta'' \end{pmatrix} = \lim_{i\to\infty} \begin{pmatrix} 0 & 1 \\ I & a_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & a_i \end{pmatrix} \begin{pmatrix} 0 \\ \mathbb{I} \end{pmatrix}.$$

In other words, the continued fractions of the vectors λ and λ' coincide except at a finite number of terms. □

Lemma 8 (main lemma). *Let $\mathcal{F}$ and $\mathcal{F}'$ be equivalent measured foliations on a surface X. Then the AF C^*-algebras $\mathbb{A}_{\mathcal{F}}$ and $\mathbb{A}_{\mathcal{F}'}$ are stably isomorphic.*

Proof. Notice that Lemma 6 implies that equivalent measured foliations $\mathcal{F}$, $\mathcal{F}'$ have proportional jacobians, i.e., $\mathfrak{m}' = \mu\mathfrak{m}$ for a $\mu > 0$. On the other hand, by Lemma 7 the continued fraction expansion of the basis vectors of the proportional jacobians must coincide, except a finite number of terms. Thus, the AF C^*-algebras $\mathbb{A}_{\mathcal{F}}$ and $\mathbb{A}_{\mathcal{F}'}$ are given by the Bratteli diagrams, which are identical, except a finite part of the diagram. It is well known [Effros 1981, Theorem 2.3] that AF C^*-algebras that have such a property are stably isomorphic. □

3. Proofs

Proof of Theorem 1. Let $\phi \in \mathrm{Mod}(X)$ be a pseudo-Anosov automorphism of the surface X. Denote by $\mathcal{F}_\phi$ the invariant foliation of ϕ. By definition of such a foliation, $\phi(\mathcal{F}_\phi) = \lambda_\phi \mathcal{F}_\phi$, where $\lambda_\phi > 1$ is the dilatation of ϕ.

Consider the jacobian $\mathrm{Jac}(\mathcal{F}_\phi) = \mathfrak{m}_\phi$ of $\mathcal{F}_\phi$. Since $\mathcal{F}_\phi$ is an invariant foliation of the pseudo-Anosov automorphism ϕ, one gets the following equality of the $\mathbb{Z}$-modules:

$$\mathfrak{m}_\phi = \lambda_\phi \mathfrak{m}_\phi, \quad \lambda_\phi \neq \pm 1. \tag{1}$$

Let $\{v^{(1)}, \dots, v^{(n)}\}$ be a basis in module $\mathfrak{m}_\phi$, such that $v^{(i)} > 0$. In view of (1), one obtains the following system of linear equations:

$$\begin{cases} \lambda_\phi v^{(1)} = a_{11} v^{(1)} + a_{12} v^{(2)} + \cdots + a_{1n} v^{(n)}, \\ \lambda_\phi v^{(2)} = a_{21} v^{(1)} + a_{22} v^{(2)} + \cdots + a_{2n} v^{(n)}, \\ \quad\vdots \\ \lambda_\phi v^{(n)} = a_{n1} v^{(1)} + a_{n2} v^{(2)} + \cdots + a_{nn} v^{(n)}, \end{cases} \tag{2}$$

where $a_{ij} \in \mathbb{Z}$. The matrix $A = (a_{ij})$ is invertible. Indeed, since the foliation $\mathcal{F}_\phi$ is minimal, the real numbers $v^{(1)}, \dots, v^{(n)}$ are linearly independent over $\mathbb{Q}$. So are the numbers $\lambda_\phi v^{(1)}, \dots, \lambda_\phi v^{(n)}$, which therefore can be taken for a basis of the module $\mathfrak{m}_\phi$. Thus, there exists an integer matrix $B = (b_{ij})$, such that $v^{(j)} = \sum_{i,j} w^{(i)}$, where $w^{(i)} = \lambda_\phi v^{(i)}$. Clearly, B is an inverse to matrix A. Therefore, $A \in \mathrm{GL}_n(\mathbb{Z})$.

Moreover, without loss of generality one can assume that $a_{ij} \geq 0$. Indeed, if this is not yet the case, consider the conjugacy class $[A]$ of the matrix A. Since

$v^{(i)} > 0$, there exists a matrix $A^+ \in [A]$ whose entries are nonnegative integers. One has to replace basis $v = (v^{(1)}, \dots, v^{(n)})$ in the module $\mathfrak{m}_\phi$ by a basis Tv, where $A^+ = TAT^{-1}$. It will be further assumed that $A = A^+$.

Lemma 9. *The vector* $(v^{(1)}, \dots, v^{(n)})$ *is the limit of a periodic Jacobi–Perron continued fraction.*

Proof. It follows from the discussion above that there exists a nonnegative integer matrix A, such that $Av = \lambda_\phi v$. In view of [Bauer 1996, Proposition 3], matrix A admits a unique factorization:

$$A = \begin{pmatrix} 0 & 1 \\ I & b_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_k \end{pmatrix}, \tag{3}$$

where $b_i = (b_1^{(i)}, \dots, b_n^{(i)})^T$ are vectors of nonnegative integers. Let us consider the periodic Jacobi–Perron continued fraction:

$$\mathrm{Per}\, \overline{\begin{pmatrix} 0 & 1 \\ I & b_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ I & b_k \end{pmatrix}} \begin{pmatrix} 0 \\ \mathbb{I} \end{pmatrix}. \tag{4}$$

According to [Perron 1907, Satz XII], the above fraction converges to a vector

$$w = (w^{(1)}, \dots, w^{(n)})$$

satisfying the equation $(B_1 B_2 \cdots B_k)w = Aw = \lambda_\phi w$. In view of the equation $Av = \lambda_\phi v$, we conclude that vectors v and w are collinear. Therefore, the Jacobi–Perron continued fractions of v and w must coincide. □

It is now straightforward to prove that the AF C^*-algebra attached to foliation $\mathscr{F}_\phi$ is stationary. Indeed, by Lemma 9, the vector of periods $v^{(i)} = \int_{\gamma_i} \omega$ unfolds into a periodic Jacobi–Perron continued fraction. By definition, the Bratteli diagram of the AF C^*-algebra $\mathbb{A}_\phi$ is periodic as well. In other words, the AF C^*-algebra $\mathbb{A}_\phi$ is stationary. □

Proof of Theorem 2. (i) For completeness, we give a proof of the following well-known lemma.

Lemma 10. *If* ϕ *and* ϕ' *are conjugate pseudo-Anosov automorphisms of a surface* X, *their invariant foliations* $\mathscr{F}_\phi$ *and* $\mathscr{F}_{\phi'}$ *are equivalent as measured foliations.*

Proof. Let $\phi, \phi' \in \mathrm{Mod}(X)$ be conjugate, i.e., $\phi' = h \circ \phi \circ h^{-1}$ for an automorphism $h \in \mathrm{Mod}(X)$. Since ϕ is the pseudo-Anosov automorphism, there exists a measured foliation $\mathscr{F}_\phi$, such that $\phi(\mathscr{F}_\phi) = \lambda_\phi \mathscr{F}_\phi$. Let us evaluate the automorphism ϕ' on the foliation $h(\mathscr{F}_\phi)$:

$$\phi'(h(\mathscr{F}_\phi)) = h\phi h^{-1}(h(\mathscr{F}_\phi)) = h\phi(\mathscr{F}_\phi) = h\lambda_\phi \mathscr{F}_\phi = \lambda_\phi(h(\mathscr{F}_\phi)). \tag{5}$$

Thus, $\mathscr{F}_{\phi'} = h(\mathscr{F}_\phi)$ is the invariant foliation for the pseudo-Anosov automorphism ϕ' and $\mathscr{F}_\phi$, $\mathscr{F}_{\phi'}$ are equivalent foliations. Note also that the pseudo-Anosov automorphism ϕ' has the same dilatation as the automorphism ϕ. □

Suppose that ϕ and ϕ' are conjugate pseudo-Anosov automorphisms. The functor F acts by the formulas $\phi \mapsto \mathbb{A}_\phi$ and $\phi' \mapsto \mathbb{A}_{\phi'}$, where $\mathbb{A}_\phi$, $\mathbb{A}_{\phi'}$ are the AF C^*-algebras corresponding to the invariant foliations $\mathscr{F}_\phi$, $\mathscr{F}_{\phi'}$. In view of Lemma 10, $\mathscr{F}_\phi$ and $\mathscr{F}_{\phi'}$ are equivalent measured foliations. Then, by Lemma 8, the AF C^*-algebras $\mathbb{A}_\phi$ and $\mathbb{A}_{\phi'}$ are stably isomorphic AF C^*-algebras. Item (i) follows.

(ii) We start with an elementary observation. Let $\phi \in \mathrm{Mod}(X)$ be a pseudo-Anosov automorphism. Then there exists a unique measured foliation, $\mathscr{F}_\phi$, such that $\phi(\mathscr{F}_\phi) = \lambda_\phi \mathscr{F}_\phi$, where $\lambda_\phi > 1$ is an algebraic integer. Let us evaluate automorphism $\phi^2 \in \mathrm{Mod}(X)$ on the foliation $\mathscr{F}_\phi$:

$$\phi^2(\mathscr{F}_\phi) = \phi(\phi(\mathscr{F}_\phi)) = \phi(\lambda_\phi \mathscr{F}_\phi) = \lambda_\phi \phi(\mathscr{F}_\phi) = \lambda_\phi^2 \mathscr{F}_\phi = \lambda_{\phi^2} \mathscr{F}_\phi, \tag{6}$$

where $\lambda_{\phi^2} := \lambda_\phi^2$. Thus, foliation $\mathscr{F}_\phi$ is an invariant foliation for the automorphism ϕ^2 as well. By induction, one concludes that $\mathscr{F}_\phi$ is an invariant foliation of the automorphism ϕ^n for any $n \geq 1$.

Even more is true. Suppose that $\psi \in \mathrm{Mod}(X)$ is a pseudo-Anosov automorphism, such that $\psi^m = \phi^n$ for some $m \geq 1$ and $\psi \neq \phi$. Then $\mathscr{F}_\phi$ is an invariant foliation for the automorphism ψ. Indeed, $\mathscr{F}_\phi$ is invariant foliation of the automorphism ψ^m. If there exists $\mathscr{F}' \neq \mathscr{F}_\phi$ such that the foliation $\mathscr{F}'$ is an invariant foliation of ψ, then the foliation $\mathscr{F}'$ is also an invariant foliation of the pseudo-Anosov automorphism ψ^m. Thus, by uniqueness, $\mathscr{F}' = \mathscr{F}_\phi$. We have just proved the following lemma.

Lemma 11. *Let ϕ be the pseudo-Anosov automorphism of a surface X. Denote by $[\phi]$ a set of the pseudo-Anosov automorphisms ψ of X, such that $\psi^m = \phi^n$ for some positive integers m and n. Then the pseudo-Anosov foliation $\mathscr{F}_\phi$ is an invariant foliation for every pseudo-Anosov automorphism $\psi \in [\phi]$.*

In view of Lemma 11, one arrives at the following identities among the AF C^*-algebras:

$$\mathbb{A}_\phi = \mathbb{A}_{\phi^2} = \cdots = \mathbb{A}_{\phi^n} = \mathbb{A}_{\psi^m} = \cdots = \mathbb{A}_{\psi^2} = \mathbb{A}_\psi. \tag{7}$$

Thus, functor F is not an injective functor: the preimage, $\mathrm{Ker}\, F$, of algbera $\mathbb{A}_\phi$ consists of a countable set of the pseudo-Anosov automorphisms $\psi \in [\phi]$, commensurable with the automorphism ϕ. This proves Theorem 2(ii). □

Proof of Corollary 3.

Proof that $(\Lambda, [I], K)$ *is an invariant.* (i) It follows from Theorem 1 that $\mathbb{A}_\phi$ is a stationary AF C^*-algebra. An arithmetic invariant of the stable isomorphism classes of the stationary AF C^*-algebras has been found by D. Handelman [1981]. Summing up his results, the invariant is as follows.

Let $A \in GL_n(\mathbb{Z})$ be a matrix with strictly positive entries, such that A is equal to the minimal period of the Bratteli diagram of the stationary AF C^*-algebra. (In case the matrix A has zero entries, it is necessary to take a proper minimal power of the matrix A.) By the Perron–Frobenius theory, matrix A has a real eigenvalue $\lambda_A > 1$, which exceeds the absolute values of other roots of the characteristic polynomial of A. Note that λ_A is an invertible algebraic integer (the unit). Consider the real algebraic number field $K = \mathbb{Q}(\lambda_A)$ obtained as an extension of the field of the rational numbers by the algebraic number λ_A. Let $(v_A^{(1)}, \dots, v_A^{(n)})$ be the eigenvector corresponding to the eigenvalue λ_A. One can normalize the eigenvector so that $v_A^{(i)} \in K$.

The departure point of Handelman's invariant is the $\mathbb{Z}$-module

$$\mathfrak{m} = \mathbb{Z}v_A^{(1)} + \cdots + \mathbb{Z}v_A^{(n)}.$$

The module $\mathfrak{m}$ brings in two new arithmetic objects: (i) the ring Λ of the endomorphisms of $\mathfrak{m}$ and (ii) an ideal I in the ring Λ, such that $I = \mathfrak{m}$ after a scaling [Borevich and Shafarevich 1966, Lemma 1, p. 88]. The ring Λ is an order in the algebraic number field K and therefore one can talk about the ideal classes in Λ. The ideal class of I is denoted by $[I]$. Omitting the embedding question for the field K, the triple $(\Lambda, [I], K)$ is an invariant of the stable isomorphism class of the stationary AF C^*-algebra $\mathbb{A}_\phi$ [Handelman 1981, Section 5]. □

Proof that Δ *and* Σ *ae invariants.* Numerical invariants of the stable isomorphism classes of the stationary AF C^*-algebras can be derived from the triple $(\Lambda, [I], K)$. These invariants are rational integers — called the determinant and signature — and can be obtained as follows.

Let $\mathfrak{m}$, $\mathfrak{m}'$ be the full $\mathbb{Z}$-modules in an algebraic number field K. It follows from (i) that if $\mathfrak{m} \neq \mathfrak{m}'$ are distinct as the $\mathbb{Z}$-modules, then the corresponding AF C^*-algebras cannot be stably isomorphic. We wish to find the numerical invariants, which discern the case $\mathfrak{m} \neq \mathfrak{m}'$. It is assumed that a $\mathbb{Z}$-module is given by the set of generators $\{\lambda_1, \dots, \lambda_n\}$. Therefore, the problem can be formulated as follows: find a number attached to the set of generators $\{\lambda_1, \dots, \lambda_n\}$, which does not change on the set of generators $\{\lambda_1', \dots, \lambda_n'\}$ of the same $\mathbb{Z}$-module.

One such invariant is associated with the trace function on the algebraic number field K. Recall that $\mathrm{Tr} : K \to \mathbb{Q}$ is a linear function on K, that is, $\mathrm{Tr}(\alpha + \beta) = \mathrm{Tr}(\alpha) + \mathrm{Tr}(\beta)$ and $\mathrm{Tr}(a\alpha) = a\,\mathrm{Tr}(\alpha)$ for all $\alpha, \beta \in K$ and all $a \in \mathbb{Q}$.

Let $\mathfrak{m}$ be a full $\mathbb{Z}$-module in the field K. The trace function defines a symmetric bilinear form $q(x, y) : \mathfrak{m} \times \mathfrak{m} \to \mathbb{Q}$ by the formula

$$(x, y) \longmapsto \mathrm{Tr}(xy) \quad \text{for all } x, y \in \mathfrak{m}. \tag{8}$$

The form $q(x, y)$ depends on the basis $\{\lambda_1, \dots, \lambda_n\}$ in the module $\mathfrak{m}$:

$$q(x, y) = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i y_j, \quad \text{where } a_{ij} = \mathrm{Tr}(\lambda_i \lambda_j). \tag{9}$$

However, the general theory of bilinear forms (over the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or the ring of rational integers $\mathbb{Z}$) tells us that certain numerical quantities will not depend on the choice of such a basis.

Namely, one such invariant is as follows. Consider a symmetric matrix A corresponding to the bilinear form $q(x, y)$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}. \tag{10}$$

It is known that the matrix A, written in a new basis, will take the form $A' = U^T A U$, where $U \in \mathrm{GL}_n(\mathbb{Z})$. Then $\mathrm{Det}(A') = \mathrm{Det}(U^T A U) = \mathrm{Det}(U^T)\,\mathrm{Det}(A)\,\mathrm{Det}(U) = \mathrm{Det}(A)$. Therefore, the rational integer number

$$\Delta = \mathrm{Det}(\mathrm{Tr}(\lambda_i \lambda_j)), \tag{11}$$

called a *determinant* of the bilinear form $q(x, y)$, does not depend on the choice of the basis $\{\lambda_1, \dots, \lambda_n\}$ in the module $\mathfrak{m}$. We conclude that the determinant Δ discerns[1] the modules $\mathfrak{m} \neq \mathfrak{m}'$.

Finally, recall that the form $q(x, y)$ can be brought by an integer linear transformation to the diagonal form:

$$a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2, \tag{12}$$

where $a_i \in \mathbb{Z} \setminus \{0\}$. We let a_i^+ be the positive and a_i^- the negative entries in the diagonal form. In view of the law of inertia for bilinear forms, the integer number $\Sigma = (\# a_i^+) - (\# a_i^-)$, called a *signature*, does not depend on a particular choice of the basis in the module $\mathfrak{m}$. Thus, Σ discerns the modules $\mathfrak{m} \neq \mathfrak{m}'$. Corollary 3 follows. □

[1] Note that if $\Delta = \Delta'$ for the modules $\mathfrak{m}$, $\mathfrak{m}'$, one cannot conclude that $\mathfrak{m} = \mathfrak{m}'$. The problem of equivalence of symmetric bilinear forms over $\mathbb{Q}$ (i.e., the existence of a linear substitution over $\mathbb{Q}$ that transforms one form to the other), is a fundamental question of number theory. The Minkowski–Hasse theorem says that two such forms are equivalent if and only if they are equivalent over the field $\mathbb{Q}_p$ for every prime number p and over the field $\mathbb{R}$. Clearly, the resulting p-adic quantities will give new invariants of the stable isomorphism classes of the AF C^*-algebras. The question is similar to the Minkowski units attached to knots; see, e.g., [Reidemeister 1932]. We will not pursue this topic here and refer the reader to the section on open problems, on page 460.

4. Examples, open problems and conjectures

In the present section we shall calculate invariants Δ and Σ for the Anosov automorphisms of the two-dimensional torus. Examples of two nonconjugate Anosov automorphisms with the same Alexander polynomial, but different determinants Δ are constructed. Recall that isotopy classes of the orientation-preserving diffeomorphisms of the torus T^2 are bijective with the 2×2 matrices with integer entries and determinant $+1$, i.e., $\mathrm{Mod}(T^2)\cong \mathrm{SL}(2,\mathbb{Z})$. Under the identification, the nonperiodic automorphisms correspond to the matrices $A\in \mathrm{SL}(2,\mathbb{Z})$ with $|\mathrm{Tr}\,A|>2$.

Full modules and orders in the quadratic field. Let $K=\mathbb{Q}(\sqrt{d})$ be a quadratic extension of the field of rational numbers $\mathbb{Q}$. Further we suppose that d is a positive square free integer. Let

$$\omega=\begin{cases}\dfrac{1+\sqrt{d}}{2} & \text{if } d\equiv 1 \mod 4,\\ \sqrt{d} & \text{if } d\equiv 2,3 \mod 4.\end{cases} \tag{13}$$

Proposition 12. *Let f be a positive integer. Every order in K has form $\Lambda_f=\mathbb{Z}+(f\omega)\mathbb{Z}$, where f is the conductor of Λ_f.*

Proof. See [Borevich and Shafarevich 1966, pp. 130–132]. □

Proposition 12 allows to classify the similarity classes of the full modules in the field K. Indeed, there exists a finite number of $\mathfrak{m}_f^{(1)},\dots,\mathfrak{m}_f^{(s)}$ of the nonsimilar full modules in the field K, whose coefficient ring is the order Λ_f; cf. [Borevich and Shafarevich 1966, Theorem 3, Chapter 2.7]. Thus, Proposition 12 gives a finite-to-one classification of the similarity classes of full modules in the field K.

Numerical invariants of Anosov automorphisms. Let Λ_f be an order in K with the conductor f. Under the addition operation, the order Λ_f is a full module, which we denote by $\mathfrak{m}_f$. Let us evaluate the invariants $q(x,y)$, Δ and Σ on the module $\mathfrak{m}_f$. To calculate $(a_{ij})=\mathrm{Tr}(\lambda_i\lambda_j)$, we let $\lambda_1=1$, $\lambda_2=f\omega$. Then

$$\begin{array}{llll} a_{11}=2, & a_{12}=a_{21}=f, & a_{22}=\frac{1}{2}f^2(d+1) & \text{if } d\equiv 1 \mod 4,\\ a_{11}=2, & a_{12}=a_{21}=0, & a_{22}=2f^2d & \text{if } d\equiv 2,3 \mod 4,\end{array} \tag{14}$$

and

$$\begin{array}{ll} q(x,y)=2x^2+2fxy+\frac{1}{2}f^2(d+1)y^2 & \text{if } d\equiv 1 \mod 4,\\ q(x,y)=2x^2+2f^2dy^2 & \text{if } d\equiv 2,3 \mod 4.\end{array} \tag{15}$$

Therefore

$$\Delta=\begin{cases} f^2d & \text{if } d\equiv 1 \mod 4,\\ 4f^2d & \text{if } d\equiv 2,3 \mod 4,\end{cases} \tag{16}$$

and $\Sigma = +2$ in both cases, where $\Sigma = \#(\text{positive}) - \#(\text{negative})$ entries in the diagonal normal form of $q(x, y)$.

Examples. Let us consider some numerical examples, which illustrate advantages of our invariants in comparison to the classical Alexander polynomials.

Example 13. Denote by M_A and M_B the hyperbolic 3-dimensional manifolds obtained as a torus bundle over the circle with the monodromies

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 1 \\ 4 & 1 \end{pmatrix}, \tag{17}$$

respectively. The Alexander polynomials of M_A and M_B are identical: $\Delta_A(t) = \Delta_B(t) = t^2 - 6t + 1$. However, the manifolds M_A and M_B are *not* homotopy equivalent. Indeed, the Perron–Frobenius eigenvector of matrix A is $v_A = (1, \sqrt{2} - 1)$ while of the matrix B is $v_B = (1, 2\sqrt{2} - 2)$. The bilinear forms for the modules $\mathfrak{m}_A = \mathbb{Z} + (\sqrt{2} - 1)\mathbb{Z}$ and $\mathfrak{m}_B = \mathbb{Z} + (2\sqrt{2} - 2)\mathbb{Z}$ can be written as

$$q_A(x, y) = 2x^2 - 4xy + 6y^2, \quad q_B(x, y) = 2x^2 - 8xy + 24y^2, \tag{18}$$

respectively. The modules $\mathfrak{m}_A$, $\mathfrak{m}_B$ are not similar in the number field $K = \mathbb{Q}(\sqrt{2})$, since their determinants $\Delta(\mathfrak{m}_A) = 8$ and $\Delta(\mathfrak{m}_B) = 32$ are not equal. Therefore, matrices A and B are not conjugate[2] in the group $SL(2, \mathbb{Z})$. Note that the class number $h_K = 1$ for the field K.

Example 14 [Handelman 2009, p. 12]. Let M_A and M_B be 3-dimensional manifolds corresponding to matrices

$$A = \begin{pmatrix} 4 & 3 \\ 5 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 4 & 15 \\ 1 & 4 \end{pmatrix}, \tag{19}$$

respectively. The Alexander polynomials of M_A and M_B are identical: $\Delta_A(t) = \Delta_B(t) = t^2 - 8t + 1$. Yet the manifolds M_A and M_B are not homotopy equivalent. Indeed, the Perron–Frobenius eigenvector of matrix A is $v_A = (1, \frac{1}{3}\sqrt{15})$ while of the matrix B is $v_B = (1, \frac{1}{15}\sqrt{15})$. The corresponding modules are $\mathfrak{m}_A = \mathbb{Z} + (\frac{1}{3}\sqrt{15})\mathbb{Z}$ and $\mathfrak{m}_B = \mathbb{Z} + (\frac{1}{15}\sqrt{15})\mathbb{Z}$; note that $d = 15 \equiv 3 \mod 4$ in both cases, but the corresponding conductors are $f_A = 3$ and $f_B = 15$. Using formulas (15) one finds

$$q_A(x, y) = 2x^2 + 18y^2, \quad q_B(x, y) = 2x^2 + 450y^2, \tag{20}$$

[2]The reader may verify this fact using the method of periods, which dates back to Gauss. First we have to find the fixed points $Ax = x$ and $Bx = x$, which gives us $x_A = 1 + \sqrt{2}$ and $x_B = (1 + \sqrt{2})/2$, respectively. Then one unfolds the fixed points into a periodic continued fraction, which gives us $x_A = [2, 2, 2, \dots]$ and $x_B = [1, 4, 1, 4, \dots]$. Since the period (2) of x_A differs from the period (1, 4) of B, the matrices A and B belong to different conjugacy classes in $SL(2, \mathbb{Z})$.

respectively. The modules $\mathfrak{m}_A$, $\mathfrak{m}_B$ are not similar in the number field $K = \mathbb{Q}(\sqrt{15})$, since formulas (16) imply that their determinants $\Delta(\mathfrak{m}_A) = 36$ and $\Delta(\mathfrak{m}_B) = 900$ are not equal. Therefore, matrices A and B are not conjugate in the group $\mathrm{SL}(2, \mathbb{Z})$.

Example 15 [Handelman 2009, p. 12]. Let a, b be positive integers satisfying the Pell equation $a^2 - 8b^2 = 1$; the latter has infinitely many solutions, e.g., $a = 3$, $b = 1$, etc. Denote by M_A and M_B the 3-dimensional manifolds corresponding to matrices

$$A = \begin{pmatrix} a & 4b \\ 2b & a \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} a & 8b \\ b & a \end{pmatrix}. \tag{21}$$

M_A and M_B have the same Alexander polynomial, $\Delta_A(t) = \Delta_B(t) = t^2 - 2at + 1$, yet they are not homotopy equivalent. Indeed, the Perron–Frobenius eigenvector of matrix A is $v_A = (1, \frac{1}{4b}\sqrt{a^2-1})$ while of the matrix B is $v_B = (1, \frac{1}{8b}\sqrt{a^2-1})$. The corresponding modules are $\mathfrak{m}_A = \mathbb{Z} + (\frac{1}{4b}\sqrt{a^2-1})\mathbb{Z}$ and $\mathfrak{m}_B = \mathbb{Z} + (\frac{1}{8b}\sqrt{a^2-1})\mathbb{Z}$. It is easy to see that the discriminant $d = a^2 - 1 \equiv 3 \mod 4$ for all $a \geq 2$. Indeed, $d = (a-1)(a+1)$, so the integer a satisfies $a \not\equiv 1; 3 \mod 4$; hence $a \equiv 2 \mod 4$, so that $a - 1 \equiv 1 \mod 4$ and $a + 1 \equiv 3 \mod 4$ and, thus, $d = a^2 - 1 \equiv 3 \mod 4$. Therefore the corresponding conductors are $f_A = 4b$ and $f_B = 8b$, and

$$q_A(x, y) = 2x^2 + 32b^2(a^2-1)y^2, \quad q_B(x, y) = 2x^2 + 128b^2(a^2-1)y^2. \tag{22}$$

The modules $\mathfrak{m}_A$, $\mathfrak{m}_B$ are not similar in the number field $K = \mathbb{Q}(\sqrt{a^2-1})$, because their determinants $\Delta(\mathfrak{m}_A) = 64b^2(a^2-1)$ and $\Delta(\mathfrak{m}_B) = 256b^2(a^2-1)$ are not equal. Therefore, the matrices A and B are not conjugate in $\mathrm{SL}(2, \mathbb{Z})$.

Open problems and conjectures. This section is devoted to some questions and conjectures in connection with the invariants $(\Lambda, [I], K)$, $q(x, y)$, Δ and Σ.

1. *P-adic invariants of pseudo-Anosov automorphisms*

A. Let $\phi \in \mathrm{Mod}(X)$ be a pseudo-Anosov automorphism of a surface X. If λ_ϕ is the dilatation of ϕ, then one can consider a $\mathbb{Z}$-module $\mathfrak{m} = \mathbb{Z}v^{(1)} + \cdots + \mathbb{Z}v^{(n)}$ in the number field $K = \mathbb{Q}(\lambda_\phi)$ generated by the normalized eigenvector $(v^{(1)}, \ldots, v^{(n)})$ corresponding to the eigenvalue λ_ϕ. The trace function on the number field K gives rise to a symmetric bilinear form $q(x, y)$ on the module $\mathfrak{m}$. The form is defined over the field $\mathbb{Q}$. It has been shown that a pseudo-Anosov automorphism ϕ', conjugate to ϕ, yields a form $q'(x, y)$, equivalent to $q(x, y)$, i.e., $q(x, y)$ can be transformed to $q'(x, y)$ by an invertible linear substitution with the coefficients in $\mathbb{Z}$.

B. Recall that two rational bilinear forms $q(x, y)$ and $q'(x, y)$ are equivalent whenever the following conditions are met:

(i) $\Delta = \Delta'$, where Δ is the determinant of the form.

(ii) For each prime number p (including $p = \infty$), certain p-adic equations between the coefficients of forms q, q' must be satisfied; see, e.g., [Borevich and Shafarevich 1966, Chapter 1, Section 7.5]. (In fact, only a *finite* number of such equations have to be verified.)

Condition (i) has already been used to discern between the conjugacy classes of the pseudo-Anosov automorphisms. One can use condition (ii) to discern between the pseudo-Anosov automorphisms with $\Delta = \Delta'$. The following question can be posed: *find the p-adic invariants of the pseudo-Anosov automorphisms.*

2. *Signature of pseudo-Anosov automorphism*

The signature is an important and well-known invariant connected to the chirality and knotting number of knots and links [Reidemeister 1932]. It will be interesting to find a geometric interpretation of the signature Σ for the pseudo-Anosov automorphisms. One can ask the following question: *find a geometric meaning of the invariant Σ.*

3. *Number of conjugacy classes of pseudo-Anosov automorphisms with the same dilatation*

The dilatation λ_ϕ is an invariant of the conjugacy class of the pseudo-Anosov automorphism $\phi \in \mathrm{Mod}(X)$. On the other hand, it is known that there exist nonconjugate pseudo-Anosov's with the same dilatation and the number of such classes is finite [Thurston 1988]. It is natural to expect that the invariants of operator algebras can be used to evaluate the number. We conclude with the following conjecture.

Conjecture 16. Let $(\Lambda, [I], K)$ be the triple corresponding to a pseudo-Anosov automorphism $\phi \in \mathrm{Mod}(X)$. Then the number of the conjugacy classes of the pseudo-Anosov automorphisms with the dilatation λ_ϕ is equal to the class number $h_\Lambda = |\Lambda/[I]|$ of the integral order Λ.

Acknowledgments

I thank the referee for helpful comments and Daniel Silver and Susan Williams for their interest and hospitality.

References

[Bauer 1996] M. Bauer, "A characterization of uniquely ergodic interval exchange maps in terms of the Jacobi-pPerron algorithm", *Bol. Soc. Brasil. Mat.* (*N.S.*) **27**:2 (1996), 109–128. MR 98a:58100 Zbl 0877.11044

[Bernstein 1971] L. Bernstein, *The Jacobi–Perron algorithm: Its theory and application*, Lecture Notes in Mathematics **207**, Springer, Berlin, 1971. MR 44 #2696 Zbl 0213.05201

[Blackadar 1986] B. Blackadar, *K-theory for operator algebras*, Mathematical Sciences Research Institute Publications **5**, Springer, New York, 1986. MR 88g:46082 Zbl 0597.46072

[Borevich and Shafarevich 1966] A. I. Borevich and I. R. Shafarevich, *Number theory*, Pure and Applied Mathematics **20**, Academic, New York, 1966. MR 33 #4001 Zbl 0145.04902

[Douady and Hubbard 1975] A. Douady and J. Hubbard, "On the density of Strebel differentials", *Invent. Math.* **30**:2 (1975), 175–179. MR 53 #796 Zbl 0371.30017

[Effros 1981] E. G. Effros, *Dimensions and C^*-algebras*, CBMS Regional Conference Series in Mathematics **46**, Conference Board of the Mathematical Sciences, Washington, DC, 1981. MR 84k:46042 Zbl 0475.46050

[Handelman 1981] D. Handelman, "Positive matrices and dimension groups affiliated to C^*-algebras and topological Markov chains", *J. Operator Theory* **6**:1 (1981), 55–74. MR 84i:46058 Zbl 0495.06011

[Handelman 2009] D. Handelman, "Matrices of positive polynomials", *Electron. J. Linear Algebra* **19** (2009), 2–89. MR 2011g:46105 Zbl 1195.37010

[Hemion 1979] G. Hemion, "On the classification of homeomorphisms of 2-manifolds and the classification of 3-manifolds", *Acta Math.* **142**:1-2 (1979), 123–155. MR 80f:57003 Zbl 0402.57027

[Hubbard and Masur 1979] J. Hubbard and H. Masur, "Quadratic differentials and foliations", *Acta Math.* **142**:3-4 (1979), 221–274. MR 80h:30047 Zbl 0415.30038

[Krieger 1980] W. Krieger, "On dimension functions and topological Markov chains", *Invent. Math.* **56**:3 (1980), 239–250. MR 81m:28018 Zbl 0431.54024

[Latimer and MacDuffee 1933] C. G. Latimer and C. C. MacDuffee, "A correspondence between classes of ideals and classes of matrices", *Ann. of Math.* (2) **34**:2 (1933), 313–316. MR 1503108 Zbl 0006.29002

[Mosher 1986] L. Mosher, "The classification of pseudo–Anosovs", pp. 13–75 in *Low-dimensional topology and Kleinian groups* (Coventry/Durham, 1984), edited by D. B. A. Epstein, London Math. Soc. Lecture Note Ser. **112**, Cambridge Univ. Press, 1986. MR 89f:57016 Zbl 0623.57004

[Penner 1984] R. C. Penner, "The action of the mapping class group on curves in surfaces", *Enseign. Math.* (2) **30**:1-2 (1984), 39–55. MR 85j:57012 Zbl 0554.57007

[Perron 1907] O. Perron, "Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus", *Math. Ann.* **64**:1 (1907), 1–76. MR 1511422 JFM 38.0262.01

[Reidemeister 1932] K. Reidemeister, *Knotentheorie*, Ergebnisse der Mathematik (1) **1**, Springer, Berlin, 1932. In German. Zbl 0005.12001

[Thurston 1982] W. P. Thurston, "Three-dimensional manifolds, Kleinian groups and hyperbolic geometry", *Bull. Amer. Math. Soc.* (*N.S.*) **6**:3 (1982), 357–381. MR 83h:57019 Zbl 0496.57005

[Thurston 1988] W. P. Thurston, "On the geometry and dynamics of diffeomorphisms of surfaces", *Bull. Amer. Math. Soc.* (*N.S.*) **19**:2 (1988), 417–431. MR 89k:57023 Zbl 0674.57008

[Wallace 1984] D. I. Wallace, "Conjugacy classes of hyperbolic matrices in $\mathrm{Sl}(n, \mathbf{Z})$ and ideal classes in an order", *Trans. Amer. Math. Soc.* **283**:1 (1984), 177–184. MR 85h:11024 Zbl 0562.12010

Received February 22, 2010. Revised September 11, 2012.

IGOR NIKOLAEV
THE FIELDS INSTITUTE
TORONTO, ON M5T 3J1
CANADA
Current address:
616-315 HOLMWOOD AVENUE
OTTAWA, ON K1S 2R2
CANADA
igor.v.nikolaev@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.463

CONNECTED SUMS OF CLOSED RIEMANNIAN MANIFOLDS AND FOURTH-ORDER CONFORMAL INVARIANTS

DAVID RASKE

In this note we take some initial steps in the investigation of a fourth-order analogue of the Yamabe problem in conformal geometry. The Paneitz constants and the Paneitz invariants considered are believed to be very helpful to understand the topology of the underlying manifolds. We calculate how those quantities change, analogous to how the Yamabe constants and the Yamabe invariants do, under the connected sum operations.

1. Introduction

Let (M, g) be a connected compact Riemannian manifold without boundary of dimension $n \geq 5$. Let

(1-1)
$$Q[g] = -\frac{n-4}{4(n-1)}\Delta R + \frac{(n-4)(n^3-4n^2+16n-16)}{16(n-1)^2(n-2)^2}R^2 - \frac{2(n-4)}{(n-2)^2}|\mathrm{Ric}|^2$$

be the so-called Q-curvature, where R is the scalar curvature, Ric is the Ricci curvature. And let

$$P[g] = (-\Delta)^2 - \mathrm{div}_g\left(\left(\frac{(n-2)^2+4}{2(n-1)(n-2)}Rg - \frac{4}{n-2}\,\mathrm{Ric}_g\right)d\right) + Q[g] \tag{1-2}$$

be the so-called the Paneitz–Branson operator. It is known that

$$P[g]u = Q[g_u]u^{\frac{n+4}{n-4}} \tag{1-3}$$

which is called the Paneitz–Branson equation, where $g_u = u^{\frac{4}{n-4}}g$ (see [Paneitz 1983; Branson 1987; Xu and Yang 2001; Djadli et al. 2000]). We consider the equation (1-3) as a fourth-order analogue of the well-known scalar curvature equation

$$L[g]v = R[g_v]v^{\frac{n+2}{n-2}}, \tag{1-4}$$

where

$$L[g] = -\frac{4(n-1)}{n-2}\Delta + R \tag{1-5}$$

MSC2010: primary 57R65; secondary 57R99.
Keywords: Paneitz–Branson operator, Q-curvature.

is the so-called conformal Laplacian and $g_v = v^{\frac{4}{n-2}} g$. The well-known Yamabe problem in conformal geometry is to find a metric, in a given class of conformal metrics, which is of constant scalar curvature, i.e., to solve

$$L[g]v = Y v^{\frac{n+2}{n-2}}$$

on a given manifold (M, g) for some positive function v and a constant Y. The affirmative resolution to the Yamabe problem was given in [Schoen 1984] after other notable works [Yamabe 1960; Trudinger 1968; Aubin 1976]. In fact, it was proven that there exists a so-called Yamabe metric g_v in the class $[g]$ which is a minimizer for the so-called Yamabe functional

$$Y(v) = \frac{\int_M (vL[g]v)\, dv_g}{\left(\int_M v^{\frac{2n}{n-2}} dv_g\right)^{\frac{n-2}{n}}}.$$

In this paper we investigate a fourth-order analogue of the Yamabe problem. Let $C_+^\infty(M)$ be the space of smooth positive functions on M. Similar to the Yamabe problem, we define the Paneitz functional

$$\wp_g(u) = \frac{\int_M (uP[g]u)\, dv_g}{\left(\int_M u^{\frac{2n}{n-4}} dv_g\right)^{\frac{n-4}{n}}} \tag{1-6}$$

for $u \in C_+^\infty(M)$ and the *Paneitz constant* associated with $(M, [g])$

$$\lambda(M, [g]) = \inf_{u \in C_+^\infty(M)} \wp_g(u). \tag{1-7}$$

It is clear that $\lambda(M, [g])$ is a conformal invariant of the conformal class $[g]$ because of the conformally covariant property of the Paneitz–Branson operator:

$$P[g_w]u = w^{-\frac{n+4}{n-4}} P[g](w \cdot u) \tag{1-8}$$

where $g_w = w^{\frac{4}{n-4}} g \in [g]$. To describe the differential structure of M, we define

$$\lambda(M) = \sup_{[g]} \lambda(M, [g]). \tag{1-9}$$

We will refer to $\lambda(M)$ as the *Paneitz invariant* of the manifold M as the counterpart of Yamabe invariant. In [1986], Gil-Medrano studied the Yamabe constant for a connected sum of two closed manifolds. One interesting consequence of connected sum results in [Gil-Medrano 1986] is that every compact manifold without boundary admits a conformal class of metrics whose Yamabe constant is very negative. In Section 3 we calculate as Gil-Medrano did in [1986] to verify that

Theorem 1.1. *Let (M_1, g_1) and (M_2, g_2) be two compact Riemannian manifolds of dimension $n \geq 5$. Then, for each $\epsilon > 0$, there is a conformal class $[g]$ of metrics* *on* $M_1 \# M_2$ *such that*

$$\lambda(M_1 \# M_2, [g]) < \min\{\lambda(M_1, [g_1]), \lambda(M_2, [g_2])\} + \epsilon \tag{1-10}$$

and there exists a conformal class $[h]$ *of metrics on* $M_1 \# M_2$ *such that*

$$\lambda(M_1 \# M_2, [h]) < 2^{-\frac{n-4}{n}} \big(\lambda(M_1, [g_1]) + \lambda(M_2, [g_2])\big) + \epsilon. \tag{1-11}$$

Due to the works of Schoen and Yau [1979] (see also [Gromov and Lawson 1980]), one knows that there is some topological constraint for a manifold to possess a metric of positive Yamabe constant. Therefore it is interesting to see how the Yamabe invariant is effected by connected sum. It was proven in [Kobayashi 1987], [Schoen and Yau 1979], and [Gromov and Lawson 1980] that the Yamabe invariant of connected sum of two manifolds with positive Yamabe invariants is still positive. More precisely, Kobayashi in [1987] showed that the Yamabe invariant of connected sum of two manifolds is greater than or equal to the smaller of the Yamabe invariants of the two. In Section 4 we obtain an analogue for the Paneitz invariant.

Theorem 1.2. *If* M_1 *and* M_2 *are compact manifolds of dimension* $n \geq 5$, *then*

$$\lambda(M_1 \# M_2) \geq \min\{\lambda(M_1), \lambda(M_2)\}. \tag{1-12}$$

The positivity of Paneitz invariant in dimension higher than 4 should be a topological constraint, as indicated by successful researches in [Chang and Yang 2002] (references therein) for a fourth-order analogue of how Gaussian curvature influences the geometry of surfaces in dimension 2. Another testing ground is to consider closed locally conformally flat manifolds. Then the recent works in [Chang et al. 2004] and [González 2005] indicate to us that the positivity of fourth-order curvature is indeed very informative about the topology of the underlying manifolds. We would also like to mention the work by Xu and Yang in [2001] where they demonstrated that positivity of the Paneitz–Branson operator is stable under the process of taking connected sums of two closed Riemannian manifolds.

In Section 2 we discuss some preliminary facts about the Paneitz functional. In Section 3 we calculate and verify Theorem 1.1. In Section 4 we prove Theorem 1.2.

2. Preliminaries

Recall that the Yamabe constant of any closed manifold of dimension greater than 2 is a finite number and the largest possible Yamabe constant is realized and only realized by the Yamabe constant of the standard round sphere in each dimension. The difficult part is to show that the round sphere is the only one that has the largest Yamabe constant, which was the last step in the resolution of Yamabe problem solved by Schoen in [1984] based on a positive mass theorem of Schoen and Yau.

We observe that, by (1-3),

$$\int_M (uP[g]u)\,dv_g = \int_M uQ[g_u]u^{\frac{n+4}{n-4}}\,dv_g = \int_M Q[g_u]u^{\frac{2n}{n-4}}\,dv_g = \int_M Q[g_u]dv_{g_u}, \tag{2-1}$$

where $g_u = u^{\frac{4}{n-4}} g \in [g]$. Hence

$$\begin{aligned}\int_M (uP[g]u)\,dv_g &= \int_M \left(\left(\frac{(n-4)(n^3-4n^2+16n-16)}{16(n-1)^2(n-2)^2}R^2 - \frac{2(n-4)}{(n-2)^2}|\mathrm{Ric}|^2\right)dv\right)[g_u]\\ &\le c_n \int_M ((R^2)\,dv)[g_u],\end{aligned}$$

where

$$c_n = \frac{(n-4)(n^3-4n^2+16n-1g)}{16(n-1)^2(n-2)^2} - \frac{2(n-4)}{n(n-2)^2}.$$

When we consider a Yamabe metric g_u, we have

$$\frac{\int_M (Rdv)[g_u]}{\mathrm{vol}(M,g_u)^{\frac{n-2}{n}}} = Y\,\mathrm{vol}(M,g_u)^{\frac{2}{n}} \le n(n-1)\,\mathrm{vol}(S^n,g_0)^{\frac{2}{n}}, \tag{2-2}$$

and since Y and c_n are nonnegative by hypothesis, we have

$$\begin{aligned}\frac{\int_M (uP[g]u)\,dv_g}{\mathrm{vol}(M,g_u)^{\frac{n-4}{n}}} &\le c_n Y^2\,\mathrm{vol}(M,g_u)^{\frac{4}{n}} \le c_n(n(n-1))^2\,\mathrm{vol}(S^n,g_0)^{\frac{4}{n}}\\ &= \frac{\int_{S^n}(Qdv)[g_0]}{\mathrm{vol}(S^n,g_0)^{\frac{n-4}{n}}} = \lambda(S^n,[g_0]).\end{aligned} \tag{2-3}$$

Consequently we obtain:

Lemma 2.1. *Let (M^n, g) be a closed Riemannian manifold of dimension greater than 5 with nonnegative Yamabe constant. Then*

$$\lambda(M^n,[g]) \le \lambda(S^n,[g_0]) \tag{2-4}$$

and the equality holds if and only if (M, g) is conformally equivalent to the standard round sphere (S^n, g_0).

On the other hand, by some choices of testing functions similar to the ones used to estimate the Yamabe functional, we get:

Lemma 2.2. *Let (M^n, g) be a closed Riemannian manifold of dimension greater than* 4. *Then*

$$-\infty < \lambda(M^n, [g]) \le \lambda(S^n, [g_0]), \tag{2-5}$$

where g_0 is the standard round metric on the sphere S^n.

Proof. The Paneitz constant is easily seen to be bounded from below, because, by (1-2),

$$\int_M (uP[g]u)\,dv = \int_M |\Delta u|^2 dv + a_n \int_M R|\nabla u|^2 dv - \frac{4}{n-4}\int_M \mathrm{Ric}(\nabla u, \nabla u)\,dv + \int_M Qu^2 dv, \tag{2-6}$$

where

$$a_n = \frac{(n-2)^2+4}{2(n-1)(n-2)}.$$

It suffices to estimate (2-3) for nonnegative functions such that

$$\int_M u^{\frac{2n}{n-4}} dv = 1.$$

Hence, by Hölder's inequality,

$$\begin{aligned}\int_M (uP[g]u)\,dv &\ge \int_M |\Delta u|^2 dv - C_1\int_M |\nabla u|^2 dv - C_2 \int_M u^2 dv \\ &\ge \int_M |\Delta u|^2 dv - C_1 \int_M (-\Delta u)u\,dv - C_2\int_M u^2 dv \\ &\ge \frac{1}{2}\int_M |\Delta u|^2 dv - \tfrac{1}{2}C_1^2 \int_M u^2 dv - C_2 \int_M u^2 dv \\ &\ge -\left(\tfrac{1}{2}C_1^2 + C_2\right)\left(\int_M u^{\frac{2n}{n-4}} dv\right)^{\frac{n-4}{n}} \mathrm{vol}(M, g)^{\frac{4}{n}} \\ &\ge -\left(\tfrac{1}{2}C_1^2 + C_2\right)\mathrm{vol}(M, g)^{\frac{4}{n}},\end{aligned} \tag{2-7}$$

for some constants $C_1, C_2 > 0$ depending on (M^n, g).

To estimate the upper bound we choose to work in geodesic normal coordinates in a very small geodesic ball $B_{2\epsilon} \subset M$ and transplant the rescaled round sphere metric. Let $B_{2\epsilon}(0) \subset R^n$ and

$$g_{ij}(x) = \delta_{ij} + O(|x|^2) \quad \text{for all } x \in B_{2\epsilon}(0). \tag{2-8}$$

Define a smooth nonnegative function u_ϵ on M by

$$u_\epsilon(x) = \begin{cases} \left(\dfrac{2\epsilon^3}{\epsilon^6 + |x|^2}\right)^{\frac{n-4}{2}} & \text{for } x \in B_\epsilon(0), \\ 0 & \text{for } x \notin B_{2\epsilon}(0) \end{cases}. \tag{2-9}$$

It is easily calculated that

$$\begin{aligned} \int_M (u_\epsilon P[g] u_\epsilon)\, dv &= \int_{B_\epsilon(0)} |\Delta u_\epsilon|^2 dx + o(1) \\ &= \int_{R^n} \left| \Delta \left(\frac{2\epsilon^3}{\epsilon^6 + |x|^2} \right)^{\frac{n-4}{2}} \right|^2 dx + o(1) \\ &= \int_{R^n} \left| \Delta \left(\frac{2}{1 + |x|^2} \right)^{\frac{n-4}{2}} \right|^2 dx + o(1) \end{aligned} \tag{2-10}$$

and

$$\begin{aligned} \int_M u_\epsilon^{\frac{2n}{n-4}} dv &= \int_{B_\epsilon(0)} u_\epsilon^{\frac{2n}{n-4}} dx + o(1) \\ &= \int_{R^n} \left(\frac{2\epsilon^3}{\epsilon^6 + |x|^2} \right)^n dx + o(1) \\ &= \int_{R^n} \left(\frac{2}{1 + |x|^2} \right)^n dx + o(1). \end{aligned} \tag{2-11}$$

Therefore

$$\wp(u_\epsilon) = \frac{\int_M (u_\epsilon P[g] u_\epsilon)\, dv}{\left(\int_M u_\epsilon^{\frac{2n}{n-4}} dv\right)^{\frac{n-4}{n}}} = \frac{\int_{R^n} |\Delta s|^2 dx}{\left(\int_{R^n} s^{\frac{2n}{n-4}} dx\right)^{\frac{n-4}{n}}} + o(1), \tag{2-12}$$

where $s = \left(\frac{2}{1+|x|^2}\right)^{\frac{n-4}{2}}$. Thus, taking $\epsilon \to 0$, we arrive at

$$\lambda(M, [g]) \leq \lambda(S^n, [g_0]). \tag{2-13}$$

□

One interesting question would be whether (M, g) is conformally equivalent to (S^n, g_0) when $\lambda(M, [g]) = \lambda(S^n, [g_0])$ without assuming the Yamabe constant of (M, g) is nonnegative. In other words one would be interested in searching for some analogue of a positive mass theorem of Schoen and Yau here if it make any sense.

3. Connected sums and the Paneitz constant

In this section we will calculate the Paneitz functional on a connected sum of two closed manifolds and verify Theorem 1.1. Let (M, g) be a closed manifold of dimension higher than 4. Fix a point $p \in M$ and let

$$f_\delta = \begin{cases} 0 & \text{for } x \in B_\delta(p), \\ 1 & \text{for } x \in M \setminus B_{2\delta}(p), \end{cases} \tag{3-1}$$

be a family of smooth functions. We may ask that

$$0 \le f_\delta \le 1, \quad |\nabla f_\delta| < \frac{C_0}{\delta}, \quad |\Delta f_\delta| < \frac{C_0}{\delta^2} \tag{3-2}$$

for some number $C_0 > 0$.

Lemma 3.1. *Let (M, g) be a closed manifold of dimension greater than 4. Let $u \in C_+^\infty(M)$ be given. Then $u_\delta = f_\delta u \in C_+^\infty(M)$ and*

$$\wp_g(u_\delta) = \wp_g(u) + o(1) \tag{3-3}$$

as $\delta \to 0$.

Proof. We simply calculate, for a fixed $\delta > 0$, by (2-6) and (3-2),

$$\begin{aligned} &\int_M (u_\delta P[g] u_\delta)\, dv \\ &= \int_M |\Delta u_\delta|^2 dv + a_n \int_M R |\nabla u_\delta|^2 dv - \frac{4}{n-4} \int_M \mathrm{Ric}(\nabla u_\delta, \nabla u_\delta)\, dv + \int_M Q u_\delta^2 dv \\ &= \int_M (u P[g] u)\, dv + o(1) \end{aligned} \tag{3-4}$$

and

$$\int_M u_\delta^{\frac{2n}{n-4}} dv = \int_M u^{\frac{2n}{n-4}} dv + o(1), \tag{3-5}$$

as $\delta \to 0$. □

Now let us consider the connected sum of two closed Riemannian manifolds. Let (M_1, g_1) and (M_2, g_2) be two compact Riemannian manifolds without boundary of dimension $n \ge 5$. For $x_1 \in M_1$ and $x_2 \in M_2$, let $B_{\delta_1}(x_1) \subset M_1$ and $B_{\delta_2}(x_2) \subset M_2$ be geodesic balls respectively. To make the connected sum one simply takes off the open balls $B_{\frac{1}{2}\delta_1}(x_1)$ and $B_{\frac{1}{2}\delta_2}(x_2)$ from M_1 and M_2, identify $\partial B_{\frac{1}{2}\delta_1}(x_1)$ with $\partial B_{\frac{1}{2}\delta_2}(x_2)$. Hence

$$M_1 \# M_2 = \big[\big(M_1 \setminus B_{\frac{1}{2}\delta_1}(x_1)\big) \cup \big(M_2 \setminus B_{\frac{1}{2}\delta_2}(x_2)\big)\big] / \big\{\partial B_{\frac{1}{2}\delta_1}(x_1) \sim \partial B_{\frac{1}{2}\delta_2}(x_2)\big\}. \tag{3-6}$$

We may construct a metric g on the connected sum $M_1 \# M_2$ such that g agrees with g_1 on $M_1 \setminus B_{\delta_1}(x_1)$ and g_2 on $M_2 \setminus B_{\delta_2}(x_2)$. Notice that topologically $M_1 \# M_2$ does not depend on the value of δ_i when they are sufficiently small. Now let us calculate and estimate the Paneitz functional on the connected sum.

Theorem 3.2. *Let* (M_1, g_1) *and* (M_2, g_2) *be two closed Riemannian manifolds of dimension* $n \geq 5$. *Then for each* $\epsilon > 0$, *there is a conformal structure* $[g]$ *on* $M_1 \# M_2$ *such that*

$$\lambda(M_1 \# M_2, [g]) < \min\{\lambda(M_1, [g_1]), \lambda(M_2, [g_2])\} + \epsilon. \tag{3-7}$$

Alternatively, we may find a conformal structure $[g]$ *on* $M_1 \# M_2$ *such that*

$$\lambda(M, [g]) < \lambda(M_1, [g_1]) + \lambda(M_2, [g_2]) 2^{-\frac{n-4}{n}} + \epsilon. \tag{3-8}$$

Proof. Let us assume that $\lambda(M_1, [g_1]) \leq \lambda(M_2, [g_2])$ and $\epsilon > 0$ fixed. By the definition of the Paneitz constant, we know that there is a real number $\delta > 0$ and a smooth function $u_\delta \in C_+^\infty(M)$ such that u_δ vanishes on a geodesic ball $B_\delta(x_1)$ of radius δ and centered at $x_1 \in M_1$ and such that

$$\wp_g(u_\delta) < \lambda(M_1, [g_1]) + \epsilon.$$

Let g be a metric on $M = M_1 \# M_2$ which agrees with g_1, when restricted to $M_1 \setminus B_\delta(x_1)$. And define the function $\tilde{u}_\delta$ on $M_1 \# M_2$ as follows:

$$\begin{cases} \tilde{u}_\delta = u_\delta & \text{on } M_1 \setminus B_\delta(x_1), \\ \tilde{u}_\delta = 0 & \text{elsewhere.} \end{cases}$$

We then have

$$\wp_g(\tilde{u}_\delta) = \frac{\int_M \left(\Delta \tilde{u}_\delta{}^2 + a_n R |\nabla \tilde{u}_\delta|^2 - \frac{4}{n-2} \operatorname{Ric}(\nabla \tilde{u}_\delta, \nabla \tilde{u}_\delta) + Q \tilde{u}_\delta^2\right) dv}{\left(\int_M \tilde{u}_\delta^{\frac{2n}{n-4}} dv\right)^{\frac{n}{n-4}}}.$$

Recalling that u_δ vanishes on $B_\delta(x_1)$ we see that

$$\wp_g(\tilde{u}_\delta) = \wp_{g_1}(u_\delta) < \lambda(M_1, [g_1]) + \epsilon.$$

Consequently,

$$\lambda(M, [g]) < \lambda(M_1, [g_1]) + \epsilon = \min(\lambda(M_1, [g_1]), \lambda(M_2, [g_2])) + \epsilon.$$

We now proceed to prove (3-8). First, Lemma 3.1 can be used to say that for any fixed $\epsilon > 0, x_1 \in M_1, x_2 \in M_2$, we can find two positive reals δ_1, δ_2 and smooth functions $u_{\delta_1}, u_{\delta_2}$, where $u_{\delta_i} \in C^\infty(M_i)$, with the following properties:

$$\begin{aligned} u_{\delta_1} &= 0 \quad \text{on } B_{\delta_1}(x_1), \quad \wp_{g_1}(u_{\delta_1}) < \lambda(M_1, [g_1]) + \epsilon_1, \\ u_{\delta_2} &= 0 \quad \text{on } B_{\delta_2}(x_2), \quad \wp_{g_2}(u_{\delta_2}) < \lambda(M_2, [g_2]) + \epsilon_1, \end{aligned}$$

where $\epsilon_1 = 2^{\frac{-n+4}{n}} \epsilon$. Also, notice that we can assume without loss of generality that the $L^{\frac{2n}{n-4}}(M)$ norms of u_{δ_1} and u_{δ_2} are normalized. Using the same reasoning as in the proof of (3-7), a metric g on $M_1 \# M_2$ can be constructed such that g

agrees with g_i when restricted to $M_i \setminus B_{\delta_i}(x_i)$. Let us consider now the function $\tilde{u}$ on $M = M_1 \# M_2$ given by

$$\tilde{u} = \begin{cases} u_{\delta_1} & \text{on } M_1 \setminus B_{\delta_1}(x_1), \\ u_{\delta_2} & \text{on } M_2 \setminus B_{\delta_2}(x_1), \\ 0 & \text{elsewhere,} \end{cases} \tag{3-9}$$

then

$$\wp_g(\tilde{u}) = \frac{\int_{M_1\setminus B_{\delta_1}(x_1)}\big((\Delta\tilde{u})^2 + a_n R|\nabla\tilde{u}|^2 - \frac{4}{n-4}\operatorname{Ric}(\nabla\tilde{u},\nabla\tilde{u}) + Q\tilde{u}^2\big)\,dv}{\big(\int_{M_1\setminus B_{\delta_1}(x_1)}\tilde{u}^{\frac{2n}{n-4}}dv + \int_{M_2\setminus B_{\delta_2}(x_2)}\tilde{u}^{\frac{2n}{n-4}}dv\big)^{\frac{n}{n-4}}} + \frac{\int_{M_2\setminus B_{\delta_2}(x_2)}\big((\Delta\tilde{u})^2 + a_n R|\nabla\tilde{u}|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla\tilde{u},\nabla\tilde{u}) + Q\tilde{u}^2\big)\,dv}{\big(\int_{M_1\setminus B_{\delta_1}(x_1)}\tilde{u}^{\frac{2n}{n-4}}dv + \int_{M_2\setminus B_{\delta_2}(x_2)}\tilde{u}^{\frac{2n}{n-4}}dv\big)^{\frac{n}{n-4}}}.$$

Using (3-9) we then obtain

$$\wp_g(\tilde{u}) = \frac{\int_{M_1\setminus B_{\delta_1}(x_1)}\big((\Delta\tilde{u}_{\delta_1})^2 + a_n R|\nabla\tilde{u}_{\delta_1}|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla\tilde{u}_{\delta_1},\nabla\tilde{u}_{\delta_1}) + Q\tilde{u}_{\delta_1}^2\big)\,dv}{\big(\int_{M_1\setminus B_{\delta_1}(x_1)}\tilde{u}_{\delta_1}^{\frac{2n}{n-4}}dv + \int_{M_2\setminus B_{\delta_2}(x_2)}\tilde{u}_{\delta_2}^{\frac{2n}{n-4}}dv\big)^{\frac{n}{n-4}}} + \frac{\int_{M_2\setminus B_{\delta_2}(x_2)}\big((\Delta\tilde{u}_{\delta_2})^2 + a_n R|\nabla\tilde{u}_{\delta_2}|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla\tilde{u}_{\delta_2},\nabla\tilde{u}_{\delta_2}) + Q\tilde{u}_{\delta_2}^2\big)\,dv}{\big(\int_{M_1\setminus B_{\delta_1}(x_1)}\tilde{u}_{\delta_1}^{\frac{2n}{n-4}}dv + \int_{M_2\setminus B_{\delta_2}(x_2)}\tilde{u}_{\delta_2}^{\frac{2n}{n-4}}dv\big)^{\frac{n}{n-4}}}.$$

Now, recalling the above stated properties of u_{δ_1} and u_{δ_2}, we may also assume

$$\int_{M_i\setminus B_{\delta_i}(x_i)} u_{\delta_i}{}^{\frac{2n}{n-4}}\,dv = 1,$$

and

$$\begin{aligned}\wp_{g_i}(u_{\delta_i}) &= \int_{M_i\setminus B_{\delta_i}(x_i)}\Big(\Delta\tilde{u}_{\delta_i}{}^2 + a_n R|\nabla\tilde{u}_{\delta_i}|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla\tilde{u}_{\delta_i},\nabla\tilde{u}_{\delta_i}) + Q\tilde{u}_{\delta_i}^2\Big)\,dv \\ &< \lambda(M_i,[g_i]) + \epsilon_1.\end{aligned}$$

Thus

$$\begin{aligned}\lambda(M,[g]) \le \wp_g(\tilde{u}) &< \big(\lambda(M_1,[g_1]) + \lambda(M_2,[g_2]) + 2\epsilon_1\big)2^{-\frac{n-4}{n}} \\ &= \big(\lambda(M_1,[g_1]) + \lambda(M_2,[g_2])\big)2^{-\frac{n-4}{n}} + \epsilon. \qquad \square\end{aligned}$$

4. Connected sums and the Paneitz invariants

Kobayashi in [1987] showed that the Yamabe invariant of connected sum of two manifolds is greater than or equal to the smaller of the Yamabe invariants of the two. The aim of this section is to generalize this result of Kobayashi to the case of compact manifolds of dimension $n \geq 5$, and with the Yamabe invariant $Y(M)$ replaced by it's fourth-order analogue the Paneitz invariant $\lambda(M)$. Namely, we have

Theorem 4.1. *Let M_1 and M_2 be closed manifolds of dimension $n \geq 5$. If $\lambda(M_1) > 0$ and $\lambda(M_2) > 0$ then*

$$\lambda(M_1 \# M_2) \geq \min\{\lambda(M_1), \lambda(M_2)\}. \tag{4-1}$$

We will basically follow the approach taken in [Kobayashi 1987]. First we consider the Paneitz invariant on the disjoint union of compact manifolds. Take two n-manifolds with conformal structures, say $(M_1, [g_1])$ and $(M_2, [g_2])$. We write $(M, [g]) = (M_1, [g_1]) \sqcup (M_2, [g_2])$ if M is the disjoint union of M_1 and M_2, and $g_i = \{g|_{M_i}; g \in [g]\}$ for $i = 1, 2$. Let u be a smooth nonnegative function on M. Since M is the disjoint union of M_1 and M_2 it follows that we can write $u = u_1 + u_2$, where $u_i = 0$ on M_j, where $i \neq j$ and where u_i is a nonnegative smooth function on M_i. If we assume that $\lambda(M_i, [g_i]) \geq 0$ for $i = 1, 2$, then it can easily be seen that

$$\lambda(M, [g]) = \min\big\{\lambda(M_1, [g_1]), \lambda(M_2, [g_2])\big\}.$$

Due to Lemma 2.2, we can assume that $\lambda(M_1)$ and $\lambda(M_2)$ are finite; and we can use the above equation to conclude that

$$\lambda(M) = \min\{\lambda(M_1), \lambda(M_2)\}.$$

Let M be a compact manifold of dimension $n \geq 5$, and p_1 and p_2 two points of M. We take off two small balls around p_1 and p_2, and then attach a handle instead, the handle being topologically the product of a line segment and S^{n-1}. The new manifold obtained in this way will be denoted by $\overline{M}$. Let M_1 and M_2 be Riemannian manifolds and let $M_1 \sqcup M_2$ denote the disjoint union of M_1 and M_2. If $M = M_1 \sqcup M_2$ and p_1 and p_2 are taken from M_1 and M_2 respectively, then $\overline{M} = M_1 \# M_2$. Therefore we see that in order to prove Theorem 4.1 it suffices to show

$$\lambda(\overline{M}) \geq \lambda(M).$$

Proof of Theorem 4.1. Let ϵ be an arbitrary positive number, which will be fixed throughout. First, we take a metric g on M such that

$$\lambda(M, [g]) > \lambda(M) - \epsilon. \tag{4-2}$$

Due to continuity considerations we may assume that $[g]$ is conformally flat around the points p_1 and p_2. Then there is a function $\gamma \in C^\infty(M \setminus \{p_1, p_2\})$ and $g \in [g]$ such that $\tilde{g} = e^\gamma g$ is a complete metric of $M \setminus \{p_1, p_2\}$ and that each of the two ends is isometric to the half-infinite cylinder $[0,\infty)\times S^{n-1}(1)$. For convenience, we write

$$(M\setminus\{p_1,p_2\},\tilde{g})=[0,\infty)\times S^{n-1}(1)\cup(\widetilde{M},\tilde{g})\cup[0,\infty)\times S^{n-1}(1),$$

where $\widetilde{M}$ is the complement of the two cylinders. We can glue $(\widetilde{M},\tilde{g})$ and $[0,l]\times S^{n-1}(1)$, along their boundaries to get a smooth Riemannian manifold $(\overline{M},g_l)$, where $\overline{M}$ is as mentioned in the beginning of the section:

$$(\overline{M},\bar{g}_l)=(\widetilde{M},\tilde{g})\cup[0,l]\times S^{n-1}(1). \tag{4-3}$$

We then have

$$\lambda(\overline{M},[g_l])=\inf_{f>0}\frac{\int_{\overline{M}}\big((\Delta f)^2+a_nR|\nabla f|^2-\frac{4}{n-2}\operatorname{Ric}(\nabla f,\nabla f)+Qf^2\big)dv}{\big(\int_{\overline{M}}f^{\frac{2n}{n-4}}dv\big)^{\frac{n}{n-4}}},$$

So, take a positive function $f_l\in C^\infty(\overline{M})$ such that

$$\int_{\overline{M}}\left((\Delta f_l)^2+a_nR|\nabla f_l|^2-\frac{4}{n-2}\operatorname{Ric}(\nabla f_l,\nabla f_l)+Qf^2\right)dv<\lambda(\overline{M},[g_l])+\frac{1}{l+1} \tag{4-4}$$

and

$$\int_{\overline{M}}f_l^{\frac{2n}{n-4}}dv=1. \tag{4-5}$$

Lemma 4.2. *There is a section, say $\{t_l\}\times S^{n-1}$, in the cylindrical part of $\overline{M}$ such that*

$$\int_{\{t_l\}\times S^{n-1}}\left((\Delta f_l)^2+a_nR|\nabla f_l|^2-\frac{4}{n-2}\operatorname{Ric}(\nabla f_l,\nabla f_l)+Qf^2\right)dv<\frac{B}{l},$$

where B is a constant independent of l.

Proof. Using (4-4) we have

$$\begin{aligned}\int_{S^{n-1}\times[0,l]}&\left((\Delta f_l)^2+a_nR|\nabla f_l|^2-\frac{4}{n-2}\operatorname{Ric}(\nabla f_l,\nabla f_l)+Qf_l^2\right)dv\\&<\lambda(\overline{M},[g_l])+\frac{1}{1+l}\\&\qquad-\int_{\widetilde{M}}\left((\Delta f_l)^2+a_nR|\nabla f_l|^2-\frac{4}{n-2}\operatorname{Ric}(\nabla f_l,\nabla f_l)+Qf_l^2\right)dv.\end{aligned}$$

Now suppose that $\int_{\overline{M}}|\nabla f_l|^2dv$ goes to infinity as $l\to\infty$. It would follow that $\int_{\overline{M}}(\Delta f_l)^2\to\infty$ as $l\to\infty$ and that this divergence is much faster than the divergence of $\int_{\overline{M}}|\nabla f_l|^2dv$. But this implies that $\int_{\overline{M}}f_lP_lf_ldv>\lambda(\overline{M},[g_l])+\frac{1}{l+1}$ for large l, which forces a contradiction (here P_l is the Paneitz–Branson operator of the metric g_l.) It follows that there exists a constant D independent of l such that

$$\int_{\overline{M}} a_n R|\nabla f_l|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla f_l, \nabla f_l)\, dv \le D.$$

Note as well that there exists a constant E such that $-\int_{\overline{M}} Q f_l^2 dv \le E$. Putting this together we conclude that there exists a $t_1 \in [0, l]$ such that

$$l \int_{t_1 \times S^{n-1}} \Big((\Delta f_l)^2 + a_n R|\nabla f_l|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla f_l, \nabla f_l) + Q f_l^2 \Big)\, dv \\ < \lambda(\overline{M}, [g_l]) + \frac{1}{1+l} + D + E.$$

The lemma follows. □

Now we cut off $\overline{M}$ on the section $\{t_1 \times S^{n-1}\}$, and attach two half-infinite cylinders to it, so $(M, \backslash\{p_1, p_2\}, \bar{g})$ reappears. But this time we describe it as follows:

$$(M, \backslash\{p_1, p_2\}, \bar{g}) = [0,\infty)\times S^{n-1}(1) \cup (\overline{M} - \{t_1\}\times S^{n-1}, g_l) \cup [0,\infty)\times S^{n-1}(1).$$

We think of the function f_l as defined on $\overline{M} - \{\{t_l\}\times S^{n-1}\}$, and extend it to the whole space $M - \{p_1, p_2\}$ as follows: Let F_l be $W^{2,\infty}$ function of $\overline{M} - \{p_1, p_2\}$ such that

$$F_l = f_l \qquad \text{on } \overline{M} - \{t_l\}\times S^{n-1}$$

and

$$F_l(t,x) = \begin{cases} g(t)\tilde{f}_l(x) & \text{for } (t,x) \in [0,1]\times S^{n-1}, \\ 0 & \text{for } (t,x) \in [1,\infty]\times S^{n-1}, \end{cases}$$

where $\tilde{f}_l = f_l|_{\{t_l\}\times S^{n-1}} \in C^\infty(S^{n-1})$ and where g is a smooth function on [0, 1] that goes from a value of 1 to a value of 0, and whose derivative vanishes at 1. Now it easy to see from (4-4) and the above lemma that

$$\int_{M\backslash\{p_1,p_2\}} \Big((\Delta F_l)^2 + a_n R|\nabla F_l|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla F_l, \nabla F_l) + Q F^2 \Big)\, dv \\ < \lambda(\overline{M}, [g_l]) + \frac{B}{l},$$

where B is a constant independent of l. Obviously from (4-5) we get

$$\int_{\overline{M}\backslash\{p_1,p_2\}} F_l^{\frac{2n}{n-4}} dv > 1.$$

Therefore, we have

$$\inf \frac{\int_{M\backslash\{p_1,p_2\}} \big((\Delta F)^2 + a_n R|\nabla F|^2 - \frac{4}{n-2}\operatorname{Ric}(\nabla F, \nabla F) + Q F^2 \big)\, dv}{\big(\int_{M\backslash\{p_1,p_2\}} F^{\frac{2n}{n-4}} dv \big)^{\frac{n}{n-4}}} \le \lambda(\overline{M}),$$

where the infimum is taken over all nonnegative $W^{2,\infty}$ functions F with compact support. It follows from the choice of the metric $\tilde{g}$ that the left side of the preceding equation is equal to $\lambda(M, [g])$. Since ϵ can be chosen arbitrarily in (4-2), we conclude $\lambda(M) \le \lambda(\overline{M})$, which completes the proof. □

References

[Aubin 1976] T. Aubin, "The scalar curvature", pp. 5–18 in *Differential geometry and relativity*, edited by M. Cahen and M. Flato, Mathematical Phys. and Appl. Math. **3**, Reidel, Dordrecht, 1976. MR 55 #6476 Zbl 0345.53029

[Branson 1987] T. P. Branson, "Group representations arising from Lorentz conformal geometry", *J. Funct. Anal.* **74**:2 (1987), 199–291. MR 90b:22016 Zbl 0643.58036

[Chang and Yang 2002] S.-Y. A. Chang and P. C. Yang, "Non-linear partial differential equations in conformal geometry", pp. 189–207 in *Proceedings of the International Congress of Mathematicians, I* (Beijing, 2002), edited by T. Li, Higher Ed. Press, Beijing, 2002. MR 2004d:53031

[Chang et al. 2004] S.-Y. A. Chang, F. Hang, and P. C. Yang, "On a class of locally conformally flat manifolds", *Int. Math. Res. Not.* **2004**:4 (2004), 185–209. MR 2005d:53051 Zbl 1137.53327

[Djadli et al. 2000] Z. Djadli, E. Hebey, and M. Ledoux, "Paneitz-type operators and applications", *Duke Math. J.* **104**:1 (2000), 129–169. MR 2002f:58061 Zbl 0998.58009

[Gil-Medrano 1986] O. Gil-Medrano, "Connected sums and the infimum of the Yamabe functional", pp. 160–167 in *Differential geometry* (Peñíscola, 1985), edited by A. M. Naveira et al., Lecture Notes in Math. **1209**, Springer, Berlin, 1986. MR 88a:58206 Zbl 0604.58023

[González 2005] M. d. M. González, "Singular sets of a class of locally conformally flat manifolds", *Duke Math. J.* **129**:3 (2005), 551–572. MR 2006d:53034 Zbl 1088.53023

[Gromov and Lawson 1980] M. Gromov and H. B. Lawson, Jr., "The classification of simply connected manifolds of positive scalar curvature", *Ann. of Math.* (2) **111**:3 (1980), 423–434. MR 81h:53036 Zbl 0463.53025

[Kobayashi 1987] O. Kobayashi, "Scalar curvature of a metric with unit volume", *Math. Ann.* **279**:2 (1987), 253–265. MR 89a:53048 Zbl 0611.53037

[Paneitz 1983] S. M. Paneitz, "A quartic conformally covariant differential operator for arbitrary pseudo-Riemannian manifolds (summary)", preprint, 1983. arXiv 0803.4331

[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 and Yau 1979] R. Schoen and S. T. Yau, "On the structure of manifolds with positive scalar curvature", *Manuscripta Math.* **28**:1-3 (1979), 159–183. MR 80k:53064 Zbl 0423.53032

[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

[Xu and Yang 2001] X. Xu and P. C. Yang, "Positivity of Paneitz operators", *Discrete Contin. Dynam. Systems* **7**:2 (2001), 329–342. MR 2002d:58043 Zbl 1032.58018

[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 September 1, 2010. Revised September 17, 2012.

DAVID RASKE
UNIVERSITY OF MICHIGAN
ANN ARBOR, MI 48104
UNITED STATES
david.t.raske@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.477

RULED MINIMAL SURFACES IN THE THREE-DIMENSIONAL HEISENBERG GROUP

HEAYONG SHIN, YOUNG WOOK KIM, SUNG-EUN KOH,
HYUNG YONG LEE AND SEONG-DEOG YANG

To the memory of Professor Seok Woo Kim

It is shown that parts of planes, helicoids and hyperbolic paraboloids are the only minimal surfaces ruled by geodesics in the three-dimensional Riemannian Heisenberg group. It is also shown that they are the only surfaces in the three-dimensional Heisenberg group whose mean curvature is zero with respect to both the standard Riemannian metric and the standard Lorentzian metric.

1. Introduction

The three-dimensional Heisenberg group $\mathbb{H}_3$ is the two-step nilpotent Lie group $(\mathbb{R}^3, \star)$ where

$$(x, y, z) \star (x', y', z') := \big(x + x', y + y', z + z' + \tfrac{1}{2}(xy' - x'y)\big).$$

It is in general identified with a subgroup of $GL_3(\mathbb{R})$ by

$$(x, y, z) \leftrightarrow \begin{bmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}.$$

We consider in this paper two left-invariant metrics on $\mathbb{H}_3$: one is Riemannian and the other Lorentzian. Let us denote by Nil^3 the 3-dimensional Heisenberg group $\mathbb{H}_3$ endowed with the left-invariant Riemannian metric

$$g = dx^2 + dy^2 + \big(dz + \tfrac{1}{2}(y\,dx - x\,dy)\big)^2$$

Kim was supported by NRF 2009-0086794, Koh was supported by NRF 2009-0086794 and NRF 2011-0001565 and Yang was supported by NRF 2012R1A1A2042530.
MSC2010: 53A35.
Keywords: Heisenberg group, ruled surface, minimal surface.

on $\mathbb{R}^3$. The Riemannian Heisenberg group Nil^3 is a three-dimensional homogeneous manifold with a 4-dimensional isometry group; hence it is the most simple 3-manifold apart from the space-forms. Moreover, it is a Riemannian fibration over the Euclidean plane $\mathbb{R}^2$, with the projection $(x, y, z) \mapsto (x, y)$.

In the first part of this paper, we give a classification of all ruled minimal surfaces in Nil^3. In order to do this, we first show in Lemma 2.1 that if a ruled surface is minimal and if a ruling geodesic is not tangent to the fiber, then the ruled surface should be horizontally ruled. That is, its ruling geodesics are orthogonal to the fibers. In fact, it was one of the key observations in classifying the ruled minimal surfaces in $\mathbb{S}^2 \times \mathbb{R}$ or in $\mathbb{H}^2 \times \mathbb{R}$ in our previous paper [Kim et al. 2009a]. It turns out that this fact simplifies the nonlinear partial differential equations describing ruled minimal surfaces. Then we show in Theorem 2.3 that any ruled minimal surface in Nil^3 is, up to isometries, a part of the horizontal plane $z = 0$, the vertical plane $y = 0$, a helicoid $\tan(\lambda z) = y/x$, $\lambda \neq 0$ or a hyperbolic paraboloid $z = -xy/2$; see page 480 for the definition of planes. Moreover, we show on pages 488–489 that all of them can be regarded as helicoids or the limits of sequences of helicoids in the Hausdorff sense.

In fact, it was shown in [Bekkar and Sari 1992] that, up to isometries, parts of planes, the helicoids and the hyperbolic paraboloids are the only minimal surfaces in Nil^3 ruled by straight lines that are geodesics. According to Lemma 2.1, any ruling geodesic of a ruled minimal surface is either parallel or orthogonal to the fibers. We then note in Proposition 2.4 that geodesics parallel or orthogonal to the fibers everywhere are straight lines (in the Euclidean sense), and thereby show that "straight line" condition may be deleted in the aforementioned claim. For the properties of the Gauss map and representation formulae of the minimal surfaces in Nil^3, see for example [Bekkar et al. 2007; Daniel 2011; Inoguchi 2005; 2008; Mercuri et al. 2006; Sanini 1997].

In the second part, we consider the natural left-invariant Lorentzian metric

$$g_L = dx^2 + dy^2 - \left(dz + \tfrac{1}{2}(y\,dx - x\,dy)\right)^2$$

on $\mathbb{H}_3$. (Lorentzian metrics on $\mathbb{H}_3$ are discussed in [Rahmani 1992; Rahmani and Rahmani 2006].) Then we consider surfaces in $\mathbb{H}_3$ whose mean curvature is zero with respect to both metrics g and g_L and show in Theorem 3.2 that they must be one of the above mentioned surfaces, that is, a part of planes, helicoids or hyperbolic paraboloids. It can be considered as a generalization of the fact that the helicoids are the only surfaces except the planes in $\mathbb{R}^3$ whose mean curvature is zero with respect to both the standard Riemannian metric and the standard Lorentzian metric [Kobayashi 1983] and the fact that the helicoids (surfaces invariant under the screw motion) are the only surfaces except the trivial ones in $\mathbb{S}^2 \times \mathbb{R}$ or $\mathbb{H}^2 \times \mathbb{R}$ whose mean curvature is zero with respect to both the standard Riemannian metric and the standard Lorentzian metric [Kim et al. 2009a]. For this we derive the equation for the mean curvature of a graph in $\mathbb{H}_3$ to be zero with respect to the Lorentzian metric g_L and compare it with the minimal surface equation. The idea of considering these two equations at the same time is not new — see [Albujer and Alías 2009; Alías and Palmer 2001; Kobayashi 1983]; also, the "dualities" between minimal surfaces in Nil^3 and maximal surfaces in the Lorentzian Nil^3 were studied in [Lee 2011].

2. Ruled minimal surfaces in Nil^3

We first state several facts on the geometry of Nil^3, necessary for the proof of the main result in this section. For their proofs, one may refer, for example, to [Inoguchi et al. 1999].

A frame field. It can be easily seen that

$$\boldsymbol{e}_1 = \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}, \quad \boldsymbol{e}_2 = \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}, \quad \boldsymbol{e}_3 = \frac{\partial}{\partial z}$$

is a left-invariant orthonormal frame field on Nil^3 and in particular, $\boldsymbol{e}_3$ is tangent to the fibers. Letting ∇ be the Levi-Civita connection on Nil^3, we have for this frame field

$$\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_i = 0, \quad i = 1, 2, 3,$$
$$\nabla_{\boldsymbol{e}_1}\boldsymbol{e}_2 = -\nabla_{\boldsymbol{e}_2}\boldsymbol{e}_1 = \tfrac{1}{2}\boldsymbol{e}_3, \quad \nabla_{\boldsymbol{e}_1}\boldsymbol{e}_3 = \nabla_{\boldsymbol{e}_3}\boldsymbol{e}_1 = -\tfrac{1}{2}\boldsymbol{e}_2, \quad \nabla_{\boldsymbol{e}_2}\boldsymbol{e}_3 = \nabla_{\boldsymbol{e}_3}\boldsymbol{e}_2 = \tfrac{1}{2}\boldsymbol{e}_1.$$

Isometries. The isometry group of Nil^3 has two connected components: an isometry either preserves the orientation of both the fibers and the base of the fibration, or reverses both orientations. The identity component of the isometry group of Nil^3 is isomorphic to $\mathrm{SO}(2) \ltimes \mathbb{R}^3$ whose action is given by

$$\left(\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$
$$= \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ \tfrac{1}{2}(a\sin\theta - b\cos\theta) & \tfrac{1}{2}(a\cos\theta + b\sin\theta) & 1 \end{bmatrix} \begin{bmatrix} cx \\ y \\ z \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

which shows that Nil^3 is a homogeneous space. In fact, one can see that, for any point $p \in \mathbb{H}_3$ and a unit tangent vector $\boldsymbol{v}$ orthogonal to $\boldsymbol{e}_3(p)$, there exists a unique isometry φ such that $\varphi(p) = \mathbf{0}$, $d\varphi(\boldsymbol{v}) = \boldsymbol{e}_1(\mathbf{0})$ and $d\varphi(\boldsymbol{e}_3(p)) = \boldsymbol{e}_3(\mathbf{0})$. Note also that the translations along the z-axis (in the Euclidean sense) are isometries belonging to the identity component.

Euclidean planes. A *Euclidean plane* or simply a *plane* is a set of points $(x, y, z) \in \mathbb{H}_3$ satisfying a linear equation $ax + by + cz + d = 0$. It is easy to see that all the planes except the "vertical" planes $ax + by + d = 0$ are congruent. In fact, every nonvertical plane $ax + by + z + d = 0$ is congruent to the "horizontal" plane $z = 0$ under the left translation by $(-2b, 2a, -d)$, $(x, y, z) \mapsto (-2b, 2a, -d) \star (x, y, z)$, that is,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -a & -b & 1 \end{bmatrix} \begin{bmatrix} cx \\ y \\ z \end{bmatrix} + \begin{bmatrix} r - 2b \\ 2a \\ -d \end{bmatrix}.$$

Moreover, a vertical plane is not congruent to a nonvertical plane since every isometric image of a fiber is a fiber. In fact, one can check that a vertical plane is not isometric to a nonvertical plane by computing their curvatures.

A parametrization of ruled surfaces. Let Σ be a ruled surface in Nil^3 and let $p \in \Sigma$ be a point at which $T_p\Sigma$ is transversal to the fiber. Assume, furthermore, that the direction of the ruling geodesic at p is not perpendicular to the fibers. Then, in a neighborhood of p, we can take a tangent vector field V to Σ in the direction of the ruling (everywhere on the neighborhood) as

$$V = \eta(\cos\theta\, \boldsymbol{e}_1 - \sin\theta\, \boldsymbol{e}_2) + \boldsymbol{e}_3$$

for some functions η and θ on Σ. Since $T_p\Sigma$ is transversal to the fiber, the unit normal vector field $\boldsymbol{n}$ of Σ is not perpendicular to $\boldsymbol{e}_3$: $\langle \boldsymbol{n}, \boldsymbol{e}_3 \rangle \neq 0$. Then

$$W = \sin\theta\, \boldsymbol{e}_1 + \cos\theta\, \boldsymbol{e}_2 - \frac{\langle \boldsymbol{n}, \sin\theta\, \boldsymbol{e}_1 + \cos\theta\, \boldsymbol{e}_2 \rangle}{\langle \boldsymbol{n}, \boldsymbol{e}_3 \rangle} \boldsymbol{e}_3$$

gives another tangent vector field on Σ which is transversal to V. Now we take a parametrization $X(s, t)$ of Σ in the neighborhood of p such that $X(s, 0)$ is the integral curve of W with $X(0, 0) = p$ and such that t parameter curves are the ruling geodesics with $X_t(s, 0) = V(X(s, 0))$. Then $X(s, t)$ is a parametrization of the ruled surface Σ in the neighborhood of p satisfying

$$\begin{aligned} & X_s(s, 0) = \sin\alpha(s)\, \boldsymbol{e}_1 + \cos\alpha(s)\, \boldsymbol{e}_2 + g(s)\boldsymbol{e}_3, \\ (1) \qquad & X_t(s, 0) = h(s)(\cos\alpha(s)\, \boldsymbol{e}_1 - \sin\alpha(s)\, \boldsymbol{e}_2) + \boldsymbol{e}_3, \\ & \nabla_{X_t} X_t = 0, \end{aligned}$$

for some smooth functions $h(s)$, $\alpha(s)$ and $g(s)$.

For the parametrization X satisfying the condition (1), we are to compute the functions X_{si} and X_{ti} defined by

$$\begin{aligned} X_s(s, t) &= X_{s1}(s, t)\boldsymbol{e}_1 + X_{s2}(s, t)\boldsymbol{e}_2 + X_{s3}(s, t)\boldsymbol{e}_3, \\ X_t(s, t) &= X_{t1}(s, t)\boldsymbol{e}_1 + X_{t2}(s, t)\boldsymbol{e}_2 + X_{t3}(s, t)\boldsymbol{e}_3. \end{aligned}$$

Now, since t parameter curves are geodesics, we have

$$\nabla_{X_t}X_t = \sum_i \frac{\partial X_{ti}}{\partial t}\boldsymbol{e}_i + \sum_{i,j} X_{ti}X_{tj}\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j$$
$$= \Big(\frac{\partial X_{t1}}{\partial t} + X_{t2}X_{t3}\Big)\boldsymbol{e}_1 + \Big(\frac{\partial X_{t2}}{\partial t} - X_{t1}X_{t3}\Big)\boldsymbol{e}_2 + \frac{\partial X_{t3}}{\partial t}\boldsymbol{e}_3 = 0.$$

By solving the system of equations

$$\frac{\partial X_{t1}}{\partial t} + X_{t2}X_{t3} = 0, \quad \frac{\partial X_{t2}}{\partial t} - X_{t1}X_{t3} = 0, \quad \frac{\partial X_{t3}}{\partial t} = 0$$

with the initial condition

$$X_{t1}(s,0) = h(s)\cos\alpha(s), \quad X_{t2}(s,0) = -h(s)\sin\alpha(s), \quad X_{t3}(s,0) = 1,$$

we have

$$X_{t1}(s,t) = h(s)\cos(t-\alpha(s)), \quad X_{t2}(s,t) = h(s)\sin(t-\alpha(s)), \quad X_{t3}(s,t) = 1.$$

On the other hand, since the Levi-Civita connection ∇ is torsion-free, one has

$$\nabla_{X_t}X_s = \nabla_{X_s}X_t.$$

Hence we have

$$\Big(\frac{\partial X_{s1}}{\partial t} + \tfrac{1}{2}(X_{t2}X_{s3} + X_{t3}X_{s2})\Big)\boldsymbol{e}_1 + \Big(\frac{\partial X_{s2}}{\partial t} - \tfrac{1}{2}(X_{t1}X_{s3} + X_{t3}X_{s1})\Big)\boldsymbol{e}_2$$
$$+\Big(\frac{\partial X_{s3}}{\partial t} + \tfrac{1}{2}(X_{t1}X_{s2} - X_{t2}X_{s1})\Big)\boldsymbol{e}_3$$
$$= \Big(\frac{\partial X_{t1}}{\partial s} + \tfrac{1}{2}(X_{s2}X_{t3} + X_{s3}X_{t2})\Big)\boldsymbol{e}_1 + \Big(\frac{\partial X_{t2}}{\partial s} - \tfrac{1}{2}(X_{s1}X_{t3} + X_{s3}X_{t1})\Big)\boldsymbol{e}_2$$
$$+\Big(\frac{\partial X_{t3}}{\partial s} + \tfrac{1}{2}(X_{s1}X_{t2} - X_{s2}X_{t1})\Big)\boldsymbol{e}_3,$$

and X_{si} satisfies the equations

$$\frac{\partial X_{s1}}{\partial t} = \frac{\partial X_{t1}}{\partial s} = h'(s)\cos(t-\alpha(s)) + h(s)\alpha'(s)\sin(t-\alpha(s)),$$
$$\frac{\partial X_{s2}}{\partial t} = \frac{\partial X_{t2}}{\partial s} = h'(s)\sin(t-\alpha(s)) - h(s)\alpha'(s)\cos(t-\alpha(s)),$$
$$\frac{\partial X_{s3}}{\partial t} = \frac{\partial X_{t3}}{\partial s} + (X_{s1}X_{t2} - X_{s2}X_{t1})$$
$$= h(s)\sin(t-\alpha(s))X_{s1} - h(s)\cos(t-\alpha(s))X_{s2}$$

with the initial condition

$$X_{s1}(s,0) = \sin\alpha(s), \quad X_{s2}(s,0) = \cos\alpha(s), \quad X_{s3}(s,0) = g(s).$$

By solving these equations, we get

$$X_{s1}(s,t) = \sin\alpha(s) + h'(s)\sin(t-\alpha(s)) + h'(s)\sin\alpha(s) \\ - h(s)\alpha'(s)\cos(t-\alpha(s)) + h(s)\alpha'(s)\cos\alpha(s),$$

$$X_{s2}(s,t) = \cos\alpha(s) - h'(s)\cos(t-\alpha(s)) + h'(s)\cos\alpha(s) \\ - h(s)\alpha'(s)\sin(t-\alpha(s)) - h(s)\alpha'(s)\sin\alpha(s),$$

$$X_{s3}(s,t) = g(s) - h(s)\sin t + th(s)h'(s) - h(s)h'(s)\sin t \\ + h(s)^2\alpha'(s) - h(s)^2\alpha'(s)\cos t.$$

The second derivatives of X. We will compute the derivatives

$$\nabla_{X_t}X_t, \quad \nabla_{X_s}X_t = \nabla_{X_t}X_s \quad \text{and} \quad \nabla_{X_s}X_s.$$

For notational simplicity, let us set

$$X_{t;t} := \nabla_{X_t}X_t = X_{tt1}\boldsymbol{e}_1 + X_{tt2}\boldsymbol{e}_2 + X_{tt3}\boldsymbol{e}_3,$$
$$X_{s;t} := \nabla_{X_t}X_s = X_{st1}\boldsymbol{e}_1 + X_{st2}\boldsymbol{e}_2 + X_{st3}\boldsymbol{e}_3,$$
$$X_{s;s} := \nabla_{X_s}X_s = X_{ss1}\boldsymbol{e}_1 + X_{ss2}\boldsymbol{e}_2 + X_{ss3}\boldsymbol{e}_3.$$

Since t parameter curves are geodesics, we have $X_{t;t} = 0$, that is,

$$X_{tt1} = X_{tt2} = X_{tt3} = 0.$$

From the equalities

$$X_{s;t} = X_{t;s} = \Big(\frac{\partial X_{s1}}{\partial t} + \tfrac{1}{2}(X_{t2}X_{s3} + X_{t3}X_{s2})\Big)\boldsymbol{e}_1 \\ + \Big(\frac{\partial X_{s2}}{\partial t} - \tfrac{1}{2}(X_{t1}X_{s3} + X_{t3}X_{s1})\Big)\boldsymbol{e}_2 + \Big(\frac{\partial X_{s3}}{\partial t} + \tfrac{1}{2}(X_{t1}X_{s2} - X_{t2}X_{s1})\Big)\boldsymbol{e}_3,$$

$$X_{s;s} = \Big(\frac{\partial X_{s1}}{\partial s} + X_{s2}X_{s3}\Big)\boldsymbol{e}_1 + \Big(\frac{\partial X_{s2}}{\partial s} - X_{s1}X_{s3}\Big)\boldsymbol{e}_2 + \frac{\partial X_{s3}}{\partial s}\boldsymbol{e}_3$$

we have

$$X_{st1} = \tfrac{1}{2}\big[\cos\alpha(s) + h'(s)\cos(t-\alpha(s)) + h'(s)\cos\alpha(s) \\ + h(s)\alpha'(s)\sin(t-\alpha(s)) - h(s)\alpha'(s)\sin\alpha(s) + h(s)\sin(t-\alpha(s))\big(g(s) \\ + h(s)\big(-\sin t + h'(s)(t-\sin t) + 2h(s)\alpha'(s)\sin^2 t/2\big)\big)\big],$$

$$X_{st2} = \tfrac{1}{2}\big[-\sin\alpha(s) + h'(s)\sin(t-\alpha(s)) - h'(s)\sin\alpha(s) - h(s)\alpha'(s)\cos(t-\alpha(s)) \\ - h(s)\alpha'(s)\cos\alpha(s) + h(s)\cos(t-\alpha(s))\big(-g(s) \\ + h(s)\big(\sin t - h'(s)(t-\sin t) - 2h(s)\alpha'(s)\sin^2 t/2\big)\big)\big],$$

$$X_{st3} = \tfrac{1}{2}h(s)\big[-\cos t - h'(s)(\cos t - 1) + h(s)\alpha'(s)\sin t\big],$$

$$
\begin{aligned}
X_{ss1} = {} & \alpha'(s)\cos\alpha(s) - 2h'(s)\alpha'(s)\cos(t-\alpha(s)) + 2h'(s)\alpha'(s)\cos\alpha(s) \\
& - h(s)\alpha'(s)^2\sin(t-\alpha(s)) - h(s)\alpha'(s)^2\sin\alpha(s) \\
& + \big(-\cos\alpha(s) + \cos(t-\alpha(s)) - h'(s)\cos\alpha(s) \\
& \qquad + h(s)\sin(t-\alpha(s)) + \alpha'(s)\sin\alpha(s)\big) \\
& \quad \times \big(-g(s) + h(s)\big(\sin t + h'(s)(\sin t - t) - 2h(s)\alpha'(s)\sin^2 t/2\big)\big) \\
& + h''(s)\sin(t-\alpha(s)) + h''(s)\sin\alpha(s) \\
& - h(s)\alpha''(s)\cos(t-\alpha(s)) + h(s)\alpha''(s)\cos\alpha(s),
\end{aligned}
$$

$$
\begin{aligned}
X_{ss2} = {} & -\alpha'(s)\sin\alpha(s) - 2h'(s)\alpha'(s)\sin(t-\alpha(s)) - 2h'(s)\alpha'(s)\sin\alpha(s) \\
& + h(s)\alpha'(s)^2\cos(t-\alpha(s)) - h(s)\alpha'(s)^2\cos\alpha(s) \\
& + \big(\sin\alpha(s) + h'(s)\big(\sin(t-\alpha(s)) + \sin\alpha(s)\big) \\
& \qquad + 2h(s)\alpha'(s)\sin t/2\sin(t/2-\alpha(s))\big) \\
& \quad \times \big(-g(s) + h(s)\big(\sin t + h'(s)(\sin t - t) - 2h(s)\alpha'(s)\sin^2 t/2\big)\big) \\
& - h''(s)\cos(t-\alpha(s)) + h''(s)\cos\alpha(s) \\
& - h(s)\alpha''(s)\sin(t-\alpha(s)) - h(s)\alpha''(s)\sin\alpha(s),
\end{aligned}
$$

$$
\begin{aligned}
X_{ss3} = {} & g'(s) + h'(s)^2(t-\sin t) - h'(s)\big(\sin t - 4h(s)\alpha'(s)\sin^2 t/2\big) \\
& \qquad + h(s)\big(h''(s)(t-\sin t) - h(s)\alpha''(s)(\cos t - 1)\big).
\end{aligned}
$$

Mean curvature. We give a condition for the ruled surface Σ to be minimal in terms of the parametrization X. Let E, F, G and l, m, n, respectively, be the coefficients of the first and second fundamental forms of the surface Σ whose parametrization satisfies (1). Then the mean curvature of Σ in a neighborhood of p is given by

$$
H = \frac{1}{2}\,\frac{Gl - 2Fm + En}{EG - F^2} = \frac{1}{2}\,\frac{\langle X_t, X_t\rangle\langle X_{s;s}, X_s \times X_t\rangle - 2\langle X_s, X_t\rangle\langle X_{s;t}, X_s \times X_t\rangle}{\|X_s \times X_t\|^3}.
$$

Since

$$
X_s \times X_t = (X_{s2}X_{t3} - X_{s3}X_{t2})\boldsymbol{e}_1 + (X_{s3}X_{t1} - X_{s1}X_{t3})\boldsymbol{e}_2 + (X_{s1}X_{t2} - X_{s2}X_{t1})\boldsymbol{e}_3,
$$

X is a parametrization of a minimal surface if and only if

$$
\begin{aligned}
(2) \quad \tilde{H} := {} & \langle X_t, X_t\rangle\langle X_{s;s}, X_s \times X_t\rangle - 2\langle X_s, X_t\rangle\langle X_{s;t}, X_s \times X_t\rangle \\
= {} & \Big(\sum_i X_{ti}^2\Big)\big((X_{s2}X_{t3} - X_{s3}X_{t2})X_{ss1} + (X_{s3}X_{t1} - X_{s1}X_{t3})X_{ss2} \\
& \qquad + (X_{s1}X_{t2} - X_{s2}X_{t1})X_{ss3}\big) \\
& - 2\Big(\sum_i X_{si}X_{ti}\Big)\big((X_{s2}X_{t3} - X_{s3}X_{t2})X_{st1} \\
& \qquad + (X_{s3}X_{t1} - X_{s1}X_{t3})X_{st2} + (X_{s1}X_{t2} - X_{s2}X_{t1})X_{st3}\big) \\
= {} & 0.
\end{aligned}
$$

***Ruled minimal surfaces in* Nil^3.** Now we will find all ruled minimal surfaces in Nil^3.

Lemma 2.1. *If the surface whose parametrization X satisfies* (1) *is minimal*, *then $h(s)=0$ for all s.*

Proof. Considering the parametrizations $\tilde{X}(s,t) := X(s-s_0,t)$ if necessary, we need only to prove $h(0)=0$. By rotating the surface in Nil^3 if necessary, we may assume that $\alpha(0)=0$. Since we have explicit formulae for all X_s, X_t, $X_{s;s}$, $X_{s;t}$, $X_{t;t}$, we can compute $\tilde{H}$ directly. In particular, since X is minimal, we have $\tilde{H}(0,t)=0$ for all t. Since $\alpha(0)=0$, $\tilde{H}(0,t)$ becomes

$$\begin{aligned}\tilde{H}(0,t) = A_0 + A_1 t + A_2 t^2 + A_3 t^3 &\\ + B_0 \cos t + B_1 t \cos t + B_2 t^2 \cos t + B_3 \cos 2t + B_4 t \cos 2t + B_5 \cos 3t &\\ + C_0 \sin t + C_1 t \sin t + C_2 t^2 \sin t + C_3 \sin 2t + C_4 t \sin 2t + C_5 \sin 3t &\end{aligned}$$

where the constants A_i, B_i, C_i are functions of $h(0)$, $h'(0)$, $h''(0)$, $\alpha'(0)$, $\alpha''(0)$ and $g(0)$, $g'(0)$. In the following computation, we are to use only the following terms:

$$\begin{aligned}A_3 &= h(0)^5 h'(0)^3,\\ B_1 &= -3h(0)h'(0)^2 - h(0)^3h'(0)^2 - 3h(0)h'(0)^3 - h(0)^3h'(0)^3\\ &\quad - 2h(0)^3g(0)h'(0)\alpha'(0) - 6g(0)h(0)^5h'(0)\alpha'(0) - 3h(0)^3h'(0)\alpha'(0)^2\\ &\quad - 9h'(0)h(0)^5\alpha'(0)^2 - 6h(0)^7h'(0)\alpha'(0)^2 - h(0)^4h''(0) - h(0)^2h''(0),\\ B_5 &= \tfrac{1}{4}\big(3h(0)^4\alpha'(0) + 3h(0)^6\alpha'(0) + 6h(0)^4h'(0)\alpha'(0) + 6h(0)^6h'(0)\alpha'(0)\\ &\quad + 3h(0)^4h'(0)^2\alpha'(0) + 3h'(0)^2\alpha'(0)h(0)^6 - h(0)^6\alpha'(0)^3 - h(0)^8\alpha'(0)^3\big),\\ C_5 &= \tfrac{1}{4}\big(h(0)^3 + h(0)^5 + 3h(0)^3h'(0) + 3h(0)^5h'(0) + 3h(0)^3h'(0)^2\\ &\quad + 3h(0)^5h'(0)^2 + h(0)^3h'(0)^3 + h(0)^5h'(0)^3 - 3h(0)^5\alpha'(0)^2\\ &\quad - 3h(0)^7\alpha'(0)^2 - 3h(0)^5h'(0)\alpha'(0)^2 - 3h'(0)h(0)^7\alpha'(0)^2\big).\end{aligned}$$

Since $\tilde{H}(0,t)=0$ for all t and since the above expression is a linear combination of linearly independent functions of t, all of A_i, B_i, C_i must be 0. Since $A_3 = h(0)^5h'(0)^3 = 0$, we have either $h(0)=0$ or $h'(0)=0$. Now suppose $h(0)\neq 0$. Then $h'(0)=0$ and B_1 becomes

$$B_1 = -h''(0)h(0)^4 - h''(0)h(0)^2 = -h''(0)h(0)^2(h(0)^2+1) = 0.$$

Hence we have $h''(0)=0$ and in addition

$$\begin{aligned}4B_5 &= -\alpha'(0)^3h(0)^8 - \alpha'(0)^3h(0)^6 + 3\alpha'(0)h(0)^6 + 3\alpha'(0)h(0)^4 = 0,\\ 4C_5 &= -3\alpha'(0)^2h(0)^7 - 3\alpha'(0)^2h(0)^5 + h(0)^5 + h(0)^3 = 0.\end{aligned}$$

Then, since $3B_5 - h(0)\alpha'(0)C_5 = 2\alpha'(0)h(0)^4(h(0)^2+1) = 0$, we have $\alpha'(0) = 0$ and C_5 becomes $4C_5 = h(0)^3(h(0)^2+1) = 0$. This contradicts the assumption $h(0) \neq 0$. Hence we must have $h(0) = 0$ if X is a parametrization of a minimal surface. □

If p is a point in a ruled surface Σ at which $T_p\Sigma$ is transversal to the fiber and the direction of the ruling is not perpendicular to the fibers, then Σ has the parametrization of the type given in (1) in a neighborhood of p. If, in addition, Σ is minimal then the above lemma implies that the direction of the ruling at p is parallel to the fibers. This contradicts the fact that $T_p\Sigma$ is transversal to the fibers. Therefore we can conclude that in a ruled minimal surface Σ the directions of the rulings are horizontal, that is, perpendicular to the fibers wherever $T_p\Sigma$ is transversal to the fibers.

Now we consider the minimal surfaces which are ruled by horizontal geodesics.

Lemma 2.2. *If Σ is a minimal surface in* Nil^3 *ruled by geodesics perpendicular to the fibers, then up to the isometries in* Nil^3, *Σ is a part of the horizontal plane $z = 0$, the vertical plane $y = 0$, a helicoid* $\tan(\lambda z) = y/x$, $\lambda \neq 0$ *or a hyperbolic paraboloid $z = -xy/2$.*

Proof. One can see that the surface Σ has a local parametrization $Y(s,t)$ satisfying

$$
\begin{aligned}
&Y_s(s,0) = \cos\beta(s)(-\sin\alpha(s)\boldsymbol{e}_1 + \cos\alpha(s)\boldsymbol{e}_2) + \sin\beta(s)\boldsymbol{e}_3,\\
&Y_t(s,0) = \cos\alpha(s)\boldsymbol{e}_1 + \sin\alpha(s)\boldsymbol{e}_2,\\
&\nabla_{Y_t}Y_t = 0.
\end{aligned}
\tag{3}
$$

If we set

$$
\begin{aligned}
Y_s(s,t) &= Y_{s1}(s,t)\boldsymbol{e}_1 + Y_{s2}(s,t)\boldsymbol{e}_2 + Y_{s3}(s,t)\boldsymbol{e}_3,\\
Y_t(s,t) &= Y_{t1}(s,t)\boldsymbol{e}_1 + Y_{t2}(s,t)\boldsymbol{e}_2 + Y_{t3}(s,t)\boldsymbol{e}_3,
\end{aligned}
$$

by solving the equation $\nabla_{Y_t}Y_t = 0$ with the initial condition

$$Y_t(s,0) = \cos\alpha(s)\boldsymbol{e}_1 + \sin\alpha(s)\boldsymbol{e}_2,$$

we have

$$Y_{t1}(s,t) = \cos\alpha(s), \quad Y_{t2}(s,t) = \sin\alpha(s), \quad Y_{t3}(s,t) = 0.$$

Moreover, from $\nabla_{Y_t}Y_s = \nabla_{Y_s}Y_t$, we can see that Y_{si} satisfies the equations

$$
\begin{aligned}
\frac{\partial Y_{s1}}{\partial t} &= \frac{\partial Y_{t1}}{\partial s} = -\alpha'(s)\sin\alpha(s),\\
\frac{\partial Y_{s2}}{\partial t} &= \frac{\partial Y_{t2}}{\partial s} = \alpha'(s)\cos\alpha(s),\\
\frac{\partial Y_{s3}}{\partial t} &= \frac{\partial Y_{t3}}{\partial s} + (Y_{s1}Y_{t2} - Y_{s2}Y_{t1}) = \sin\alpha(s)Y_{s1} - \cos\alpha(s)Y_{s2}
\end{aligned}
$$

with the initial condition

$$Y_{s1}(s,0) = -\cos\beta(s)\sin\alpha(s),\ \ Y_{s2}(s,0) = \cos\beta(s)\cos\alpha(s),\ \ Y_{s3}(s,0) = \sin\beta(s).$$

By solving this system of equations, we get

$$\begin{aligned} Y_{s1}(s,t) &= -\cos\beta(s)\sin\alpha(s) - t\alpha'(s)\sin\alpha(s),\\ Y_{s2}(s,t) &= \cos\beta(s)\cos\alpha(s) + t\alpha'(s)\cos\alpha(s),\\ Y_{s3}(s,t) &= \sin\beta(s) - t\cos\beta(s) - \tfrac{1}{2}t^2\alpha'(s). \end{aligned}$$

By direct computations, we can see that the minimal surface (2) can be written as

$$\begin{aligned} \beta'(s) + t\big(\alpha'(s)\beta'(s)\cos\beta(s) - \alpha''(s)\sin\beta(s)\big)& \\ + \frac{t^2}{2}\big(\alpha'(s)\beta'(s)\sin\beta(s) + \alpha''(s)\cos\beta(s)\big) &= 0. \end{aligned}$$

Therefore we have $\beta'(s) = 0$ and $\alpha''(s) = 0$, that is, $\beta(s) = b$ and $\alpha(s) = as + c$ for some constants a, b, c.

When $a \neq 0$, relocating the surface Σ by an isometry in Nil^3, we may assume that

$$\alpha(s) = as \quad \text{and} \quad Y(0,0) = \Big(\frac{\cos b}{a}, 0, 0\Big).$$

Then, since

$$\boldsymbol{e}_1 = \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}, \qquad \boldsymbol{e}_2 = \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}, \qquad \boldsymbol{e}_3 = \frac{\partial}{\partial z},$$

we have

$$\begin{aligned} Y_s(s,0) &= -\cos b\sin(as)\,\boldsymbol{e}_1 + \cos b\cos(as)\,\boldsymbol{e}_2 + \sin b\,\boldsymbol{e}_3\\ &= -\cos b\sin(as)\frac{\partial}{\partial x} + \cos b\cos(as)\frac{\partial}{\partial y}\\ &\qquad + \Big(\sin b + \frac{y}{2}\cos b\sin(as) + \frac{x}{2}\cos b\cos(as)\Big)\frac{\partial}{\partial z},\\ Y_t(s,t) &= \cos(as)\,\boldsymbol{e}_1 + \sin(as)\,\boldsymbol{e}_2\\ &= \cos(as)\frac{\partial}{\partial x} + \sin(as)\frac{\partial}{\partial y} + \Big(-\frac{y}{2}\cos(as) + \frac{x}{2}\sin(as)\Big)\frac{\partial}{\partial z}. \end{aligned}$$

Integrating the components of $Y_s(s,0)$ with initial data $Y(0,0) = ((\cos b)/a, 0, 0)$, we have

$$Y(s,0) = \Big(\frac{1}{a}\cos b\cos(as), \frac{1}{a}\cos b\sin(as), m\frac{s}{4a}(1 + \cos(2b) + 4a\sin b)\Big).$$

Then integrating the components of $Y_t(s, t)$ with initial data $Y(s, 0)$, we have

$$Y(s,t) = \Big(t\cos(as) + \frac{1}{a}\cos b\cos(as),\\ t\sin(as) + \frac{1}{a}\cos b\sin(as), \frac{s}{4a}(1+\cos(2b)+4a\sin b)\Big).$$

Noting that

$$Y(s,t) = \Big(t\cos(as), t\sin(as), \frac{s}{4a}(1+\cos(2b)+4\sin b)\Big),$$

we can see that Y is a parametrization of either the helicoid

$$\tan\lambda z = \frac{y}{x} \quad \text{where } \lambda = \frac{4a^2}{1+\cos(2b)+4a\sin b}$$

if $1+\cos(2b)+4a\sin b \neq 0$, or the plane $z=0$ if $1+\cos(2b)+4a\sin b = 0$.

When $a=0$ and $\cos b \neq 0$, we may assume up to isometries that $\alpha(s)=0$ and $Y(0,0) = (-\tan b, 0, 0)$. Then

$$Y_s(s,0) = \cos b\,\boldsymbol{e}_2 + \sin b\,\boldsymbol{e}_3 = \cos b\frac{\partial}{\partial y} + \Big(\sin b + \frac{x}{2}\cos b\Big)\frac{\partial}{\partial z},$$
$$Y_t(s,t) = \boldsymbol{e}_1 = \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z},$$

and a similar computation as above gives

$$Y(s,t) = \big(t - \tan b, s\cos b, -\tfrac{1}{2}st\cos b + \tfrac{1}{2}s\sin b\big),$$

which is a parametrization of the hyperbolic paraboloid $z = -xy/2$. When $a=0$ and $\cos b = 0$, we have

$$Y_s(s,0) = \boldsymbol{e}_3, \quad Y_t(s,t) = \boldsymbol{e}_1 = \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}$$

and $Y(s,t)$ is a parametrization of the xz-plane if we set $Y(0,0) = (0,0,0)$. □

Theorem 2.3. *If Σ is a minimal surface in* Nil^3 *ruled by geodesics, then up to the isometries in* Nil^3, *Σ is a part of the horizontal plane $z=0$, the vertical plane $y=0$, a helicoid $\tan(\lambda z) = y/x$, $\lambda \neq 0$ or a hyperbolic paraboloid $z = -xy/2$.*

Proof. If there is a point $p \in \Sigma$ at which $T_p\Sigma$ is transversal to the fibers, then Σ is transversal to the fibers in a neighborhood of p. Therefore, from the argument following Lemma 2.1, the ruling geodesics through any points in the neighborhood must be horizontal. Then by Lemma 2.2 the neighborhood coincides with a part of the helicoids, the hyperbolic paraboloid or the xy-plane up to the isometries in Nil^3. Now since the tangent spaces at every point of these surfaces are transversal to fibers, the whole Σ must be a part of one of these surfaces.

On the other hand, if the tangent space $T_p\Sigma$ is tangent to the fibers at every point $p \in \Sigma$, then $\boldsymbol{e}_3$ is tangent to Σ. Relocating Σ by an isometry of Nil^3, we may assume that $(0, 0, 0) \in \Sigma$ and that Σ is tangent to the plane $y = 0$ at $(0, 0, 0)$. So Σ is ruled by the fibers and has a ruled parametrization $X(s, t) = (x(s), y(s), t)$ satisfying $x(0) = y(0) = 0$, $y'(0) = 0$ and $x'(0) = 1$. The mean curvature of this parametrized surface can be easily computed to be

$$\frac{x''(s)y'(s) - x'(s)y''(s)}{(x'(s)^2 + y'(s)^2)^{3/2}}.$$

Solving the equation $x''(s)y'(s) - x'(s)y''(s) = 0$ with the above initial conditions, we have $y(s) = 0$, which implies that Σ is a part of the vertical plane $y = 0$. □

We remark that the mean curvature formula of the cylinder over curves in the xy-plane is given in p. 22 of [Inoguchi et al. 2000], and that characterizations of these cylinders in terms of the harmonicity of the tangential Gauss maps are given in [Sanini 1997].

By the above theorem, we know that the ruled minimal surfaces in Nil^3 are congruent to the surfaces given in the theorem, which are all ruled by horizontal geodesics. In fact, the vertical plane $y = 0$ is also ruled by vertical geodesics, that is, fibers, and this is the only doubly ruled surface among the surfaces in Theorem 2.3. Noting that isometries in Nil^3 always move fibers to fibers, we can see that the ruled minimal surfaces in Nil^3 always have horizontal ruling geodesics.

Ruled minimal surfaces as limits of helicoids. Consider the (generic) helicoids

$$H_\lambda : y - x\tan(\lambda z) = 0$$

and the point $p_\lambda(r_\lambda, 0, 0)$ on the x-axis, where $r_\lambda = \sqrt{2/\lambda}$. The isometry which sends x-axis to itself and sends the origin to p_λ is given by the left translation by $(r_\lambda, 0, 0)$, that is,

$$(x, y, z) \mapsto (r_\lambda, 0, 0) \star (x, y, z) = \Bigl(x + r_\lambda, y, z + \frac{r_\lambda}{2} y\Bigr).$$

If we pull back H_λ via this isometry, then p_λ is moved to the origin and the equation of the pullback of H_λ becomes

$$y - (x + r_\lambda)\tan\Bigl(\lambda z + \frac{\sqrt{2\lambda}}{2} y\Bigr) = 0.$$

Now let $\mu = 1/r_\lambda = \sqrt{\lambda/2}$. Then the above equation can be written as

$$z = -\frac{y}{2\mu} + \frac{1}{2\mu^2}\tan^{-1}\Bigl(\frac{\mu y}{\mu x + 1}\Bigr).$$

Using the Taylor expansion of $\tan^{-1}(x)$, we see that this is equivalent to

$$z = -\frac{xy}{2} + O(\mu)$$

in a fixed-size box around the origin when μ is sufficiently small. This function converges uniformly to $z = -xy/2$ as μ goes to 0, which shows that the pointed helicoids (H_λ, p_λ) converge (in the Hausdorff sense) to the exceptional ruled minimal surface $z + xy/2 = 0$:

$$(H_\lambda, p_\lambda) \to \{z + xy/2 = 0\} \quad \text{as } \lambda \to 0+.$$

On the other hand, one can easily check that

$$\begin{aligned} (H_\lambda, 0) &\to \text{horizontal plane} && \text{as } \lambda \to \infty, \\ (H_\lambda, 0) &\to \text{vertical plane} && \text{as } \lambda \to 0. \end{aligned}$$

Therefore all ruled minimal surfaces in Nil^3 are either helicoids or limits of sequences of them.

Straight line geodesics. We characterize the geodesics that are straight lines in the Euclidean sense and give another proof of the result in [Bekkar and Sari 1992] mentioned in Section 1.

Proposition 2.4. *Let $\gamma(t) = (x(t), y(t), z(t))$ be a geodesic in* Nil^3.

(1) *If $\gamma'(0)$ is perpendicular to the fiber, then $\gamma(t)$ is a straight line everywhere perpendicular to the fibers.*
(2) *If $\gamma'(0)$ is parallel to the fiber, then $\gamma(t)$ is a straight line everywhere parallel to the fibers.*

Proof. The following equation of geodesics is given in [Inoguchi et al. 1999], but we derive it here again for self-completeness. Note first that

$$\gamma' = x'\frac{\partial}{\partial x} + y'\frac{\partial}{\partial y} + z'\frac{\partial}{\partial z} = x'\boldsymbol{e}_1 + y'\boldsymbol{e}_2 + \big(z' + \tfrac{1}{2}(x'y - xy')\big)\boldsymbol{e}_3.$$

Then we have

$$\begin{aligned} \nabla_{\gamma'}\gamma' &= x''\boldsymbol{e}_1 + y''\boldsymbol{e}_2 + \big(z' + \tfrac{1}{2}(x'y - xy')\big)'\boldsymbol{e}_3 \\ &\qquad + x'\nabla_{\gamma'}\boldsymbol{e}_1 + y'\nabla_{\gamma'}\boldsymbol{e}_2 + \big(z' + \tfrac{1}{2}(x'y - xy')\big)\nabla_{\gamma'}\boldsymbol{e}_3 \\ &= \big(x'' + y'(z' + \tfrac{1}{2}(x'y - xy'))\big)\boldsymbol{e}_1 + \big(y'' - x'(z' + \tfrac{1}{2}(x'y - xy'))\big)\boldsymbol{e}_2 \\ &\qquad + \big(z' + \tfrac{1}{2}(x'y - xy')\big)'\boldsymbol{e}_3. \end{aligned}$$

Hence $\gamma(t) = (x(t), y(t), z(t))$ is a geodesic if and only if

$$
\begin{aligned}
&x'' + y'\big(z' + \tfrac{1}{2}(x'y - xy')\big) = 0,\\
&y'' - x'\big(z' + \tfrac{1}{2}(x'y - xy')\big) = 0,\\
&\big(z' + \tfrac{1}{2}(x'y - xy')\big)' = 0.
\end{aligned}
\tag{4}
$$

Note that the straight line $(a, b, ct + d)$ parallel to the fiber is a geodesic. Now, suppose $\langle\gamma'(0), \boldsymbol{e}_3\rangle = 0$. Then, since $\langle\gamma'(0), \boldsymbol{e}_3\rangle = \big(z' + \frac{1}{2}(x'y - xy')\big)(0) = 0$ and since $z'+\frac{1}{2}(x'y-xy')$ is a constant function, from the geodesic equation (4) or by the so-called conservation lemma [O'Neill 1983, p. 152], we have $z'+\frac{1}{2}(x'y-xy')=0$ for all t. Moreover, the geodesic equation (4) gives $x''(t) = y''(t) = 0$, that is, $x(t)$ and $y(t)$ are linear functions of t and consequently from the geodesic equation (4) again, we have

$$z(t) = -\tfrac{1}{2}(x'(0)y(0) - x(0)y'(0))t + c$$

for a constant c. Now it is easy to see that $\gamma(t)$ is perpendicular to the fibers everywhere.

If $\gamma'(0)$ is parallel to the fiber, then the fiber through $\gamma(0)$ is an image of a geodesic, and from the uniqueness of the geodesic, we have $\gamma(t) = (x(0), y(0), at+b)$ for constants a, b which is parallel to the fiber everywhere. □

Proposition 2.5. *Suppose the straight line $\delta(t) = (a_1t + b_1, a_2t + b_2, a_3t + b_3)$ is a geodesic in* Nil^3. *Then $\delta'(0) = (a_1, a_2, a_3)$ is either perpendicular or parallel to the fiber. Moreover, if $\delta'(0)$ is perpendicular to the fiber, then $\delta(t)$ is perpendicular to the fiber everywhere and if $\delta'(0)$ is parallel to the fiber, then $\delta(t)$ is parallel to the fiber everywhere.*

Proof. In the proof of the above Proposition 2.4, one can see that in order for the straight line $\delta(t)$ to be a geodesic, it should be that $a_3 = -\frac{1}{2}(a_1b_2 - a_2b_1)$. The claims follow easily from this fact. □

Now we can also say that every ruled minimal surface in Nil^3 is ruled by geodesics which are also straight lines. We remark that it was shown in [Bekkar and Sari 1992] that if the surface is ruled by geodesics that are also straight lines then the surface must be a part of the planes, helicoids or hyperbolic paraboloids. However, in view of Theorem 2.3, we can see that the "straight line" condition is redundant. On the other hand, one may get Theorem 2.3 by applying the aforementioned result together with Lemma 2.1 and Proposition 2.4.

3. Another characterization of ruled minimal surfaces in $\mathbb{H}_3$

We consider surfaces in $\mathbb{H}_3$ whose mean curvature is zero with respect to both metrics g and g_L and show that they must be one of (a part of) the above mentioned surfaces, that is, planes, helicoids and hyperbolic paraboloids.

A Lorentzian connection. Let us consider the left-invariant Lorentzian metric

$$g_L = dx^2 + dy^2 - \big(dz + \tfrac{1}{2}(y\,dx - x\,dy)\big)^2$$

on $\mathbb{H}_3$ and let $\langle\,\cdot\,,\cdot\,\rangle$ be the Lorentzian inner product. Let $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ be the same as the ones given in Section 2. It is easy to show that they are orthonormal with respect to the Lorentzian metric g_L as well, that is, $\langle \boldsymbol{e}_i, \boldsymbol{e}_j\rangle = 0$ if $i \neq j$ and

$$\langle \boldsymbol{e}_1, \boldsymbol{e}_1\rangle = \langle \boldsymbol{e}_2, \boldsymbol{e}_2\rangle = 1, \quad \langle \boldsymbol{e}_3, \boldsymbol{e}_3\rangle = -1.$$

Let D be the Levi-Civita connection for the metric g_L.

Proposition 3.1. *We have* $D_{\boldsymbol{e}_i}\boldsymbol{e}_i = 0$ *for* $i = 1, 2, 3,$ *and*

$$D_{\boldsymbol{e}_1}\boldsymbol{e}_2 = -D_{\boldsymbol{e}_2}\boldsymbol{e}_1 = \tfrac{1}{2}\boldsymbol{e}_3, \quad D_{\boldsymbol{e}_1}\boldsymbol{e}_3 = D_{\boldsymbol{e}_3}\boldsymbol{e}_1 = \tfrac{1}{2}\boldsymbol{e}_2, \quad D_{\boldsymbol{e}_2}\boldsymbol{e}_3 = D_{\boldsymbol{e}_3}\boldsymbol{e}_2 = -\tfrac{1}{2}\boldsymbol{e}_1.$$

Proof. It is known that the Koszul formula

$$\begin{aligned}2\langle \nabla_V W, X\rangle = V\langle W, X\rangle + W\langle X, V\rangle - X\langle V, W\rangle \\ -\langle V, [W, X]\rangle + \langle W, [X, V]\rangle + \langle X, [V, W]\rangle\end{aligned}$$

holds; see, for instance, [O'Neill 1983]. Since $[\boldsymbol{e}_1, \boldsymbol{e}_2] = \boldsymbol{e}_3$, $[\boldsymbol{e}_2, \boldsymbol{e}_3] = [\boldsymbol{e}_3, \boldsymbol{e}_1] = 0$, one has $\langle D_{\boldsymbol{e}_1}\boldsymbol{e}_2, \boldsymbol{e}_1\rangle = 0$, $\langle D_{\boldsymbol{e}_1}\boldsymbol{e}_2, \boldsymbol{e}_2\rangle = 0$, $2\langle D_{\boldsymbol{e}_1}\boldsymbol{e}_2, \boldsymbol{e}_3\rangle = \langle \boldsymbol{e}_3, [\boldsymbol{e}_1, \boldsymbol{e}_2]\rangle = \langle \boldsymbol{e}_3, \boldsymbol{e}_3\rangle = -1$ and $D_{\boldsymbol{e}_1}\boldsymbol{e}_2 = \tfrac{1}{2}\boldsymbol{e}_3$. Since

$$\langle D_{\boldsymbol{e}_1}\boldsymbol{e}_3, \boldsymbol{e}_1\rangle = 0, \quad \langle D_{\boldsymbol{e}_1}\boldsymbol{e}_3, \boldsymbol{e}_3\rangle = 0, \quad 2\langle D_{\boldsymbol{e}_1}\boldsymbol{e}_3, \boldsymbol{e}_2\rangle = \langle \boldsymbol{e}_3, [\boldsymbol{e}_2, \boldsymbol{e}_1]\rangle = \langle \boldsymbol{e}_3, -\boldsymbol{e}_3\rangle = 1,$$

one has $D_{\boldsymbol{e}_1}\boldsymbol{e}_3 = \tfrac{1}{2}\boldsymbol{e}_2$. One can check the others in the same manner. □

Lorentzian exterior product. For tangent vectors

$$\boldsymbol{v} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + a_3\boldsymbol{e}_3, \quad \boldsymbol{w} = b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2 + b_3\boldsymbol{e}_3$$

in Nil_1^3, the Lorentzian exterior product $\boldsymbol{v} \times_L \boldsymbol{w}$ is computed as

$$\boldsymbol{v} \times_L \boldsymbol{w} = \begin{vmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & -\boldsymbol{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = (a_2b_3 - a_3b_2)\boldsymbol{e}_1 + (a_3b_1 - a_1b_3)\boldsymbol{e}_2 + (a_2b_1 - a_1b_2)\boldsymbol{e}_3,$$

which is orthogonal to both $\boldsymbol{v}$ and $\boldsymbol{w}$. One can easily see that $\boldsymbol{v} \times_L \boldsymbol{w} = \boldsymbol{0}$ if and only if $\boldsymbol{v}$ and $\boldsymbol{w}$ are linearly dependent.

Zero mean curvature equation. Let Σ be a graph of a function $z = f(x, y)$ in $\mathbb{H}_3$ and consider the parametrization $\boldsymbol{r}(x, y) = (x, y, f(x, y))$ of Σ. Set

$$p = f_x + \frac{y}{2}, \quad q = f_y - \frac{x}{2}.$$

If Σ is minimal, that is, the mean curvature is zero in Nil^3, the function f satisfies the minimal surface equation

$$(1+q^2)f_{xx} - 2pqf_{xy} + (1+p^2)f_{yy} = 0.$$

For the derivation of this equation, see for example [Inoguchi et al. 2000].

In this section, we will derive an equation for the mean curvature of the graph Σ to be zero with respect to the Lorentzian metric g_L. First, let us recall some definitions. A point $z \in \Sigma$ is called *spacelike* if the induced metric on $T_z\Sigma$ is Riemannian, *timelike* if the induced metric is Lorentzian and *lightlike* if the induced metric has rank 1. We will derive the equation when Σ is spacelike, that is, every point of Σ is a spacelike point. The case when Σ is timelike is almost identical. Note that when $z \in \Sigma$ is lightlike, one cannot define the mean curvature.

Now let Σ be a spacelike graph of a function $z = f(x, y)$. Note first that $p^2 + q^2 < 1$ since the graph is spacelike. We now compute the first fundamental form I and the second fundamental form II of Σ. Since

$$\boldsymbol{r}_x = (1, 0, f_x) = \boldsymbol{e}_1 + p\boldsymbol{e}_3, \quad \boldsymbol{r}_y = (0, 1, f_y) = \boldsymbol{e}_2 + q\boldsymbol{e}_3$$

and

$$\begin{aligned} \langle \boldsymbol{r}_x, \boldsymbol{r}_x \rangle &= 1 - p^2, \\ \langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle &= -pq, \\ \langle \boldsymbol{r}_y, \boldsymbol{r}_y \rangle &= 1 - q^2, \end{aligned}$$

one has

$$\begin{aligned} E &= \langle \boldsymbol{r}_x, \boldsymbol{r}_x \rangle = 1 - p^2, \\ F &= \langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle = -pq, \\ G &= \langle \boldsymbol{r}_y, \boldsymbol{r}_y \rangle = 1 - q^2. \end{aligned}$$

Since $\boldsymbol{r}_x \times_L \boldsymbol{r}_y = -p\boldsymbol{e}_1 - q\boldsymbol{e}_2 - \boldsymbol{e}_3$, the unit normal vector field $\boldsymbol{n}$ to the graph is

$$\boldsymbol{n} = \frac{1}{W}(-p\boldsymbol{e}_1 - q\boldsymbol{e}_2 - \boldsymbol{e}_3), \quad W = \sqrt{1 - (p^2 + q^2)}\,.$$

Since the directional derivatives of p and q, $\boldsymbol{e}_i(p)$ and $\boldsymbol{e}_i(q)$, are computed as

$$\begin{aligned} \boldsymbol{e}_1(p) &= \Big(\frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}\Big)\Big(f_x + \frac{y}{2}\Big) = f_{xx}, \\ \boldsymbol{e}_1(q) &= \Big(\frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}\Big)\Big(f_y - \frac{x}{2}\Big) = f_{xy} - \tfrac{1}{2}, \\ \boldsymbol{e}_2(p) &= \Big(\frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}\Big)\Big(f_x + \frac{y}{2}\Big) = f_{xy} + \tfrac{1}{2}, \\ \boldsymbol{e}_2(q) &= \Big(\frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}\Big)\Big(f_y - \frac{x}{2}\Big) = f_{yy}, \end{aligned}$$

one has

$$\begin{aligned}
D_{\boldsymbol{r}_x}\boldsymbol{r}_x &= D_{(\boldsymbol{e}_1+p\boldsymbol{e}_3)}(\boldsymbol{e}_1+p\boldsymbol{e}_3) = p\boldsymbol{e}_2 + f_{xx}\boldsymbol{e}_3,\\
D_{\boldsymbol{r}_y}\boldsymbol{r}_x &= -\frac{p}{2}\boldsymbol{e}_1 + \frac{q}{2}\boldsymbol{e}_2 + f_{xy}\boldsymbol{e}_3,\\
D_{\boldsymbol{r}_y}\boldsymbol{r}_y &= -q\boldsymbol{e}_1 + f_{yy}\boldsymbol{e}_3.
\end{aligned}$$

Then one has the following coefficients of the second fundamental form II:

$$\begin{aligned}
l = \langle D_{\boldsymbol{r}_x}\boldsymbol{r}_x, \boldsymbol{n}\rangle &= \frac{1}{W}(-pq + f_{xx}),\\
m = \langle D_{\boldsymbol{r}_y}\boldsymbol{r}_x, \boldsymbol{n}\rangle &= \frac{1}{W}\Big(\frac{p^2}{2} - \frac{q^2}{2} + f_{xy}\Big),\\
n = \langle D_{\boldsymbol{r}_y}\boldsymbol{r}_y, \boldsymbol{n}\rangle &= \frac{1}{W}(pq + f_{yy}).
\end{aligned}$$

Now the mean curvature H of the spacelike graph Σ is computed as

$$H = \tfrac{1}{2}\,\frac{lG - 2mF + nE}{EG - F^2}.$$

Then, since

$$\begin{aligned}
&lG - 2mF + nE\\
&\quad = \frac{1}{W}\Big[(-pq + f_{xx})(1-q^2) + \Big(\frac{p^2-q^2}{2} + f_{xy}\Big)pq + (pq + f_{yy})(1-p^2)\Big]\\
&\quad = \frac{1}{W}\big[(1-q^2)f_{xx} + 2pqf_{xy} + (1-p^2)f_{yy}\big],
\end{aligned}$$

one can see that the mean curvature of the graph $z = f(x, y)$ of a function $f(x, y)$ is zero if and only if

$$(1-q^2)f_{xx} + 2pqf_{xy} + (1-p^2)f_{yy} = 0.$$

When the graph Σ is timelike, one has the same equation.

Zero mean curvature surface.

Theorem 3.2. *Let Σ be a surface in $\mathbb{H}_3$. If the mean curvature of Σ is zero with respect to both metrics g and g_L, then up to the isometries in* Nil^3, *Σ is contained in one of these surfaces*:

- *the horizontal plane* $z = 0$;
- *the vertical plane* $y = 0$;
- *a helicoid* $\tan(\lambda z) = y/x$, $\lambda \neq 0$;
- *a hyperbolic paraboloid* $z = -xy/2$.

Proof. Suppose first that Σ has a point around which it can be represented as a graph of a function of (x, y), say, $z = f(x, y)$. Consider the vector field $X = -q\boldsymbol{e}_1 + p\boldsymbol{e}_2$. Since $X = -q\boldsymbol{e}_1 + p\boldsymbol{e}_2 = -q\boldsymbol{r}_x + p\boldsymbol{r}_y$, it is tangent to Σ. Since the vector $N = \boldsymbol{r}_x \times \boldsymbol{r}_y = -p\boldsymbol{e}_1 - q\boldsymbol{e}_2 - \boldsymbol{e}_3$ is orthogonal to Σ and since $N \times \boldsymbol{e}_3 = -q\boldsymbol{e}_1 + p\boldsymbol{e}_2 = X$, X is orthogonal to both N and $\boldsymbol{e}_3$. Then one has

$$\nabla_X X = \big(q\big(f_{xy} - \tfrac{1}{2}\big) - pf_{yy}\big)\boldsymbol{e}_1 + \big(p\big(f_{xy} + \tfrac{1}{2}\big) - qf_{xx}\big)\boldsymbol{e}_2.$$

Now, since the mean curvature of $\Sigma \subset \mathbb{H}_3$ is zero with respect to both g and g_L, one has

$$(1 + q^2) f_{xx} - 2pq f_{xy} + (1 + p^2) f_{yy} = 0, \tag{5}$$

$$(1 - q^2) f_{xx} + 2pq f_{xy} + (1 - p^2) f_{yy} = 0. \tag{6}$$

Subtracting the two equations, one has

$$q^2 f_{xx} - 2pq f_{xy} + p^2 f_{yy} = 0 \tag{7}$$

and then one has finally by (7),

$$\begin{aligned} X \times \nabla_X X &= (-q\boldsymbol{e}_1 + p\boldsymbol{e}_2) \times \big[\big(q\big(f_{xy} - \tfrac{1}{2}\big) - pf_{yy}\big)\boldsymbol{e}_1 + \big(p\big(f_{xy} + \tfrac{1}{2}\big) - qf_{xx}\big)\boldsymbol{e}_2\big] \\ &= (q^2 f_{xx} - 2pq f_{xy} + p^2 f_{yy})\boldsymbol{e}_3 = 0. \end{aligned}$$

Now, since X and $\nabla_X X$ have the same direction, the integral curve of X passing through a point in Σ is a geodesic, and since X is orthogonal to $\boldsymbol{e}_3$, the geodesic is orthogonal to the fiber. Hence the surface Σ is a horizontally ruled minimal surface in Nil^3.

If the surface Σ has no point around which Σ is represented as the graph of $f(x, y)$, then it is a vertical cylinder over a curve in the xy-plane and has a parametrization

$$X(s, t) = (x(s), y(s), t),$$

with $x(0) = y(0) = 0$. By repeating the arguments in Theorem 2.3, one can show that the surface is isometric to the vertical plane $y = 0$. Now this completes the proof. □

Remark. If we add (5) and (6), we have $f_{xx} + f_{yy} = 0$, that is, if a graph of a function $z = f(x, y)$ in $\mathbb{H}_3$ satisfies the condition of Theorem 3.2, f must be a harmonic function. This fact is true for the three-dimensional Lorentzian space $\mathbb{L}^3$ and is the motivation of [Kim et al. 2009b]. We think it is a nontrivial fact and would like to find applications of this fact in future study.

Acknowledgments

We would like to express our gratitude to the anonymous referee for many valuable suggestions.

References

[Albujer and Alías 2009] A. L. Albujer and L. J. Alías, "Calabi–Bernstein results for maximal surfaces in Lorentzian product spaces", *J. Geom. Phys.* **59**:5 (2009), 620–631. MR 2010i:47088 Zbl 1173.53025

[Alías and Palmer 2001] L. J. Alías and B. Palmer, "A duality result between the minimal surface equation and the maximal surface equation", *An. Acad. Brasil. Ciênc.* **73**:2 (2001), 161–164. MR 2002c:53007 Zbl 0999.53007

[Bekkar and Sari 1992] M. Bekkar and T. Sari, "Surfaces minimales réglées dans l'espace de Heisenberg $\mathbf{H}_3$", *Rend. Sem. Mat. Univ. Politec. Torino* **50**:3 (1992), 243–254. MR 94h:53009 Zbl 0810.53012

[Bekkar et al. 2007] M. Bekkar, F. Bouziani, Y. Boukhatem, and J. Inoguchi, "Helicoids and axially symmetric minimal surfaces in 3-dimensional homogeneous spaces", *Differ. Geom. Dyn. Syst.* **9** (2007), 21–39. MR 2008e:53009 Zbl 1159.53335

[Daniel 2011] B. Daniel, "The Gauss map of minimal surfaces in the Heisenberg group", *Int. Math. Res. Not.* **2011**:3 (2011), 674–695. MR 2012b:53117 Zbl 1209.53048

[Inoguchi 2005] J.-i. Inoguchi, "Flat translation invariant surfaces in the 3-dimensional Heisenberg group", *J. Geom.* **82**:1-2 (2005), 83–90. MR 2006d:53070 Zbl 1082.53064

[Inoguchi 2008] J.-i. Inoguchi, "Minimal surfaces in the 3-dimensional Heisenberg group", *Differ. Geom. Dyn. Syst.* **10** (2008), 163–169. MR 2009a:53009 Zbl 1160.53364

[Inoguchi et al. 1999] J.-i. Inoguchi, T. Kumamoto, N. Ohsugi, and Y. Suyama, "Differential geometry of curves and surfaces in 3-dimensional homogeneous spaces, I", *Fukuoka Univ. Sci. Rep.* **29**:2 (1999), 155–182. MR 2000h:53018 Zbl 0962.53032

[Inoguchi et al. 2000] J.-I. Inoguchi, T. Kumamoto, N. Ohsugi, and Y. Suyama, "Differential geometry of curves and surfaces in 3-dimensional homogeneous spaces, II", *Fukuoka Univ. Sci. Rep.* **30**:1 (2000), 17–47. MR 2001c:53085 Zbl 0974.53039

[Kim et al. 2009a] Y. W. Kim, S.-E. Koh, H. Shin, and S.-D. Yang, "Helicoids in $\mathbb{S}^2 \times \mathbb{R}$ and $\mathbb{H}^2 \times \mathbb{R}$", *Pacific J. Math.* **242**:2 (2009), 281–297. MR 2010h:53008 Zbl 1172.53037

[Kim et al. 2009b] Y. W. Kim, H. Y. Lee, and S.-D. Yang, "Minimal harmonic graphs and their Lorentzian cousins", *J. Math. Anal. Appl.* **353**:2 (2009), 666–670. MR 2010i:53110 Zbl 1161.53051

[Kobayashi 1983] O. Kobayashi, "Maximal surfaces in the 3-dimensional Minkowski space L^3", *Tokyo J. Math.* **6**:2 (1983), 297–309. MR 85d:53003 Zbl 0535.53052

[Lee 2011] H. Lee, "Maximal surfaces in Lorentzian Heisenberg space", *Differential Geom. Appl.* **29**:1 (2011), 73–84. MR 2012e:53117 Zbl 1216.53055

[Mercuri et al. 2006] F. Mercuri, S. Montaldo, and P. Piu, "A Weierstrass representation formula for minimal surfaces in $\mathbb{H}_3$ and $\mathbb{H}^2 \times \mathbb{R}$", *Acta Math. Sin.* (*Engl. Ser.*) **22**:6 (2006), 1603–1612. MR 2007g:53007 Zbl 1119.53041

[O'Neill 1983] B. O'Neill, *Semi-Riemannian geometry: with applications to relativity*, Pure and Applied Mathematics **103**, Academic Press, New York, 1983. MR 85f:53002 Zbl 0531.53051

[Rahmani 1992] S. Rahmani, "Métriques de Lorentz sur les groupes de Lie unimodulaires, de dimension trois", *J. Geom. Phys.* **9**:3 (1992), 295–302. MR 93f:53061 Zbl 0752.53036

[Rahmani and Rahmani 2006] N. Rahmani and S. Rahmani, "Lorentzian geometry of the Heisenberg group", *Geom. Dedicata* **118** (2006), 133–140. MR 2007h:53112 Zbl 1094.53065

[Sanini 1997] A. Sanini, "Gauss map of a surface of the Heisenberg group", *Boll. Un. Mat. Ital. B* (7) **11**:2, suppl. (1997), 79–93. MR 98e:53009 Zbl 0886.53018

Received October 6, 2011. Revised September 14, 2012.

HEAYONG SHIN
DEPARTMENT OF MATHEMATICS
CHUNG-ANG UNIVERSITY
SEOUL 156-756
SOUTH KOREA
hshin@cau.ac.kr

YOUNG WOOK KIM
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEOUL 136-701
SOUTH KOREA
ywkim@korea.ac.kr

SUNG-EUN KOH
DEPARTMENT OF MATHEMATICS
KONKUK UNIVERSITY
SEOUL 143-701
SOUTH KOREA
sekoh@konkuk.ac.kr

HYUNG YONG LEE
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEOUL 136-713
SOUTH KOREA
distgeo@korea.ac.kr

SEONG-DEOG YANG
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEOUL 136-713
SOUTH KOREA
sdyang@korea.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 261, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.261.497

G-BUNDLES OVER ELLIPTIC CURVES FOR NON-SIMPLY LACED LIE GROUPS AND CONFIGURATIONS OF LINES IN RATIONAL SURFACES

MANG XU AND JIAJIN ZHANG

We study the relation between the moduli space of flat G-bundles over a fixed elliptic curve Σ and the moduli space of rational surfaces with G-configurations containing Σ as a fixed anticanonical curve, where G is a non-simply laced, compact, simple and simply connected Lie group. Our method is to reduce G to a simply laced maximal subgroup G'.

1. Introduction

This paper is a continuation of our earlier study, briefly recapitulated below, on the identification between the moduli space of flat G-bundles over a fixed elliptic curve Σ and the moduli space of rational surfaces with G-configurations containing Σ as an anticanonical curve. For the case of $G = E_n$, the rational surfaces are exactly del Pezzo surfaces, and the identification was predicted by a duality argument in physics and proved in [Looijenga 1976; Donagi 1997; 1998; Friedman et al. 1997]. The essential reason for this identification in this case is the existence of an E_n-structure on del Pezzo surfaces [Demazure et al. 1980; Manin 1974], which turns out to be related to Gosset polytopes [Lee 2010; 2012].

This structure on rational surfaces was extended to the cases A_n and D_n in [Leung 2000]. Starting from Leung's result, we obtained in [Leung and Zhang 2009a] an analogous identification for all simply laced Lie groups G. In [Leung et al. 2012; Leung and Zhang 2009b], we extended this identification further to the non-simply laced cases and the affine Kac–Moody $\widetilde{E}_n$ case. The method in that last paper consists in reducing non-simply laced cases to simply laced cases, by considering a non-simply laced Lie group G as the fixed subgroup of a bigger simply laced group G', under the action of the outer automorphism group of G'.

In this paper, we consider another reduction. From Lie theory (see [Bourbaki 2005], for example), a non-simply laced Lie group G is uniquely determined by

Xu is partially supported by National Natural Science Foundation of China (grant 11126264). Zhang is partially supported by Doctoral Fund of Ministry of Education of China (grant 20110181120091).
MSC2010: primary 14J26; secondary 17B10, 14H60.
Keywords: G-bundle, rational surface, elliptic curve.

a simply laced maximal subgroup G' determined by the long roots of G. Hence it is natural to apply our earlier results for the simply laced cases in [Leung and Zhang 2009a] to the current situation. In this way, we establish the identification between flat G-bundles over a fixed elliptic curve Σ and rational surfaces with Σ as an anticanonical curve for non-simply laced Lie groups G ($G \neq F_4$), by considering the maximal simply laced subgroup G' determined by the long roots of G. Unfortunately, this method is not very effective for the case $G = F_4$. In the following, we assume that $G \neq F_4$. Similar to the simply laced cases, we define *G-surfaces* and *rational surfaces with G-configurations* (see Definitions 5, 12, and 16). Let $\mathrm{Out}(G')$ be the finite group defined in Proposition 2. Our result is this:

Theorem 1 (Propositions 10, 14 and 19). *Let Σ be an elliptic curve with identity element $0 \in \Sigma$, and let G be any simple, compact, simply connected Lie group of B_n, C_n or G_2 type. Denote by $\mathcal{S}_\Sigma^G$ the moduli space of the pairs (S, Σ), where S is a G-surface such that $\Sigma \in |-K_S|$. Denote by $\mathcal{M}_\Sigma^G$ the moduli space of flat G-bundles over Σ.*

(i) *$\mathcal{S}_\Sigma^G$ can be embedded into $\mathcal{M}_\Sigma^G$ as an open dense subset.*

(ii) *This embedding can extend to an isomorphism from $\overline{\mathcal{S}_\Sigma^G}$ onto $\mathcal{M}_\Sigma^G$ by including all rational surfaces with G-configurations, and this gives us a natural and explicit compactification $\overline{\mathcal{S}_\Sigma^G}$ of $\mathcal{S}_\Sigma^G$.*

This study is motivated by a certain duality in physics. When $G = E_n$ is considered as a simple subgroup of $E_8 \times E_8$, these G-bundles are related to the duality between F-theory and string theory. Among other things, this duality predicts the identification between the moduli of flat E_n-bundles over a fixed elliptic curve Σ and the moduli of del Pezzo surfaces with the fixed anticanonical curve Σ. For more details, one can see [Donagi 1997; 1998; Friedman et al. 1997]. Our result can be considered as a test of this duality for other Lie groups.

As an application, this identification provides us with an intuitive explanation for $\mathcal{M}_\Sigma^G$. We also provide an interesting geometric realization of root system theory, and we can see very clearly how the Weyl group acts on the moduli space of (marked) flat G-bundles over Σ.

Notation. Let G be a compact, simple and simply connected Lie group. We preserve the notation of in [Leung and Zhang 2009a], which is as follows.

$r(G)$	the rank of G	$\Lambda(G)$	the root lattice
$R(G)$	the root system	$\Lambda_c(G)$	the coroot lattice
$R_c(G)$	the coroot system	$\Lambda_w(G)$	the weight lattice
$W(G)$	the Weyl group	$\mathrm{ad}(G)$	the adjoint group of G (= $G/C(G)$)
$T(G)$	a maximal torus	$\Delta(G)$	the set of simple roots of G
$C(G)$	the center of G	$\mathrm{Out}(G)$	the outer automorphism group of G

Recall that the outer automorphism group of G is defined as the quotient of the automorphism group of G by its inner automorphism group. As is well-known, it is isomorphic to the diagram automorphism group of the Dynkin diagram of G.

When there is no danger of confusion, we can omit the letter G.

2. Reductions to the simply laced cases

Let G be a simple, compact and simply connected Lie group. Then G is classified into the following 7 types according to its Lie algebra.

(1) A_n-type, $G = \mathrm{SU}(n+1)$;

(2) B_n-type, $G = \mathrm{Spin}(2n+1)$;

(3) C_n-type, $G = \mathrm{Sp}(n)$;

(4) D_n-type, $G = \mathrm{Spin}(2n)$;

(5) E_n-type, $n = 6, 7, 8$;

(6) F_4-type;

(7) G_2-type.

Among these, A_n, D_n and E_n are called of simply laced type, while B_n, C_n, F_4 and G_2 are called of non-simply laced type. A_n, B_n, C_n, D_n are called classical Lie groups, while E_n, F_4 and G_2 are called exceptional Lie groups.

From now on, we always assume that G is a compact, simple, simply connected Lie group of non-simply laced type, that is, of type B_n, C_n, F_4, G_2. There are two natural approaches to reduce these situations to the simply laced cases. One is embedding G into a simply laced Lie group G'' such that G is the subgroup fixed by the outer automorphism group of G''. Another is taking the simply laced subgroup G' of maximal rank.

In [Leung and Zhang 2009b] we explained the first reduction. In this paper we concentrate on the second.

Proposition 2 [Bourbaki 2005]. *There exists canonically a simply laced Lie subgroup G' of maximal rank of G determined by the long roots of G, such that G' and G share a common maximal torus. There is a short exact sequence*

$$1 \to W(G') \to W(G) \to \mathrm{Out}(G') \to 1.$$

This exact sequence is split, that is,

$$W(G) \cong W(G') \ltimes \mathrm{Out}(G').$$

We write the moduli space of flat G-bundles on Σ as $\mathcal{M}_\Sigma^G$.

Corollary 3. $$\mathcal{M}_\Sigma^G \cong \mathcal{M}_\Sigma^{G'} / \mathrm{Out}(G').$$

Proof. Let T be the common maximal torus of G and G'. Then

$$\mathcal{M}_{\Sigma}^{G} \cong \mathrm{Hom}(\pi_1(\Sigma), G)/\mathrm{ad}(G) \cong \mathrm{Hom}(\pi_1(\Sigma), T)/W(G) \cong T \times T/W(G).$$

Similarly, $\mathcal{M}_{\Sigma}^{G'} \cong T \times T/W(G')$. Therefore

$$\mathcal{M}_{\Sigma}^{G} \cong T \times T/W(G) \cong (T \times T/W(G'))/(W(G)/W(G')) \cong \mathcal{M}_{\Sigma}^{G'}/\mathrm{Out}(G'). \quad \square$$

We defined in [Leung and Zhang 2009a] (rational) G'-surfaces and rational surfaces with G'-configurations. Let $\mathcal{S}_{\Sigma}^{G'}$ be the moduli space of G'-surfaces containing Σ as a fixed anticanonical curve. As shows in the same paper, we have the following identification of moduli spaces

$$\mathcal{S}_{\Sigma}^{G'} \cong \mathcal{M}_{\Sigma}^{G'}.$$

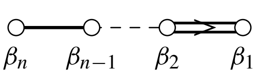
Let $\mathrm{Out}(G')$ act on $\mathcal{S}_{\Sigma}^{G'}$ via the above isomorphism. In the next section, we shall see explicitly how $\mathrm{Out}(G')$ acts on $\mathcal{S}_{\Sigma}^{G'}$.

Thus we have a natural question: How can we define G-configurations on rational surfaces when G is non-simply laced, in such a way that $\mathcal{S}_{\Sigma}^{G} \cong \mathcal{S}_{\Sigma}^{G'}/\mathrm{Out}(G')$? We answer this question in the next section.

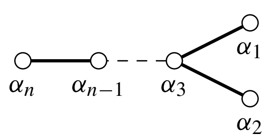
Remark 4 [Bourbaki 2005; Humphreys 1978]. We give the construction, the root system, and the finite group $\mathrm{Out}(G')$ of G' for non-simply laced Lie group G in each case. We also give the Dynkin diagrams of G and G'.

(1) For $G = \mathrm{Spin}(2n+1)$, we take $G' = \mathrm{Spin}(2n)$.
$\Delta(G') = \{\alpha_i, i = 1, \dots, n\}$.
$\Delta(G) = \{\beta_i, i = 1, \dots, n\}$, where $\beta_1 = \frac{1}{2}(\alpha_2 - \alpha_1)$, $\beta_2 = \alpha_1$, $\beta_i = \alpha_i$, $i = 3, \dots, n$.
$\mathrm{Out}(G')$ is the group $\mathbb{Z}_2$ that exchanges the two spin representations of $\mathrm{Spin}(2n)$. In fact, $\mathrm{Out}(G') = \{1, \rho\}$, where $\rho(\alpha_i) = \alpha_i$, $i = 3, \dots, n$, $\rho(\alpha_1) = \alpha_2$, and $\rho(\alpha_2) = \alpha_1$.

B_n

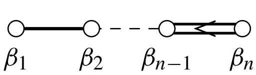


D_n

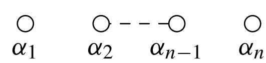


(2) For $G = \mathrm{Sp}(n)$, we take $G' = \mathrm{SU}(2)^n$.
$\Delta(G') = \{\alpha_i, i = 1, \dots, n\}$.
$\Delta(G) = \{\beta_i, i = 1, \dots, n\}$, where $\beta_i = \frac{1}{2}(\alpha_i - \alpha_{i+1})$, $i = 1, \dots, n-1$, $\beta_n = \alpha_n$.
$\mathrm{Out}(G')$ is the symmetry group S_n of the n copies of $\mathrm{SU}(2)$ in G'.

C_n

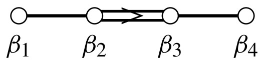


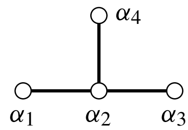
A_1^n



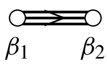
(3) For $G = F_4$, we take $G' = \mathrm{Spin}(8)$.
$\Delta(G') = \{\alpha_i, i = 1, \dots, 4\}$.
$\Delta(G) = \{\beta_i, i = 1, \dots, 4\}$, where $\beta_1 = \alpha_2$, $\beta_2 = \alpha_3$, $\beta_3 = \frac{1}{2}(\alpha_4 - \alpha_3)$, $\beta_4 = \frac{1}{2}(\alpha_1 - \alpha_4)$.

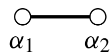
$\mathrm{Out}(G')$ is the triality group S_3 that permutes the three 8-dimensional representations of Spin(8).

F_4


D_4


(4) For $G = G_2$, we take $G' = \mathrm{SU}(3)$.
$\Delta(G') = \{\alpha_i, i = 1, 2\}$.
$\Delta(G) = \{\beta_i, i = 1, 2\}$, where $\beta_1 = \alpha_1$, $\beta_2 = -1/3(\alpha_1 + \alpha_2)$.
$\mathrm{Out}(G')$ is the group $\mathbb{Z}_2$ that exchanges the 3-dimensional representation of SU(3) with its dual. In fact, $\mathrm{Out}(G')$ is generated by $-1 \in \mathrm{Aut}(\Lambda(G'))$.

G_2


A_2


In the following we let Σ be a fixed elliptic curve with the identity element 0, and we fix a primitive d-th root of $\mathrm{Jac}(\Sigma) \cong \Sigma$ (equivalently, a level d structure on Σ), where $d = 2$ for $G = D_n, B_n$, $d = 9 - n$ for $G = E_n$, and $d = n + 1$ for $G = A_n, C_n, G_2$, respectively; see [Leung and Zhang 2009a] for the ADE cases. Recall from the same reference (for instance) that for any compact, simple and simply connected Lie group H, we have

$$\mathcal{M}_\Sigma^H \cong (\Lambda_c(H) \otimes \Sigma)/W(H),$$

where $\mathcal{M}_\Sigma^H$ is the moduli space of flat H-bundles over Σ.

3. Flat G-bundles over elliptic curves and rational surfaces: the non-simply laced cases

In this section, we study case by case the G-bundles over elliptic curves and corresponding rational surfaces for a non-simply laced Lie group G ($G \neq F_4$).

3.1. B_n-bundles ($n \geq 2$). According to the last section, for $G = \mathrm{Spin}(2n+1)$ we take $G' = \mathrm{Spin}(2n) \subseteq G$.

Let S be a D_n surface containing Σ as a smooth anticanonical curve. Recall from [Leung and Zhang 2009a] that S is a blow-up of $\mathbb{F}_1$ at n points $x_1, \ldots, x_n$ on Σ that are in general position,[1] with corresponding exceptional classes $l_1, \ldots, l_n$. Let f and s be the classes of fibers and the section in $\mathbb{F}_1$. The Picard group of S is isomorphic to $H^2(S, \mathbb{Z})$, which is a lattice with basis $s, f, l_1, \ldots, l_n$. The canonical class is $K = -\big(2s + 3f - \sum_{i=1}^n l_i\big)$.

We know from [ibid.] that the set

$$\{x \in H^2(S, \mathbb{Z}) \mid x \cdot K = x \cdot f = 0\}$$

[1] This means that the x_i are all distinct and that $x_i + x_j \neq 0$ for all i, j.

is a root lattice of D_n type. We take a simple root system of $G' = D_n$ as

$$\Delta(D_n) = \{\alpha_1 = l_1 - l_2, \alpha_2 = f - l_1 - l_2, \alpha_3 = l_2 - l_3, \ldots, \alpha_n = l_{n-1} - l_n\}.$$

Let ρ be the generator of $\mathrm{Out}(G') \cong \mathbb{Z}_2$, such that $\rho(\alpha_1) = \alpha_2$, $\rho(\alpha_2) = \alpha_1$ and $\rho(\alpha_i) = \alpha_i$ for $i = 3, \ldots, n$.

Recall that a D_n-configuration on S is an n-tuple $\zeta = (e_1, \ldots, e_n)$ where $e_i = l_{\sigma(i)}$ or $f - l_{\sigma(i)}$ such that $\sum e_i \cdot s \equiv 0 \pmod 2$. Equivalently, a D_n-configuration on S is an n-tuple $\zeta = (e_1, \ldots, e_n)$ such that after blowing down $e_n, \ldots, e_1$ successively, we obtain $\mathbb{F}^1$ with a fibration $\mathbb{F}^1 \to \mathbb{P}^1$ defined by the fiber f.

On the other hand, the exceptional system $\zeta' = (e'_1, \ldots, e'_n)$ where $e'_i = l_{\sigma(i)}$ or $f - l_{\sigma(i)}$ such that $\sum e'_i \cdot s \equiv 1 \pmod 2$ also determines $\Lambda(D_n)$. The condition $\sum e'_i \cdot s \equiv 1 \pmod 2$ is equivalent to the fact that after blowing down $e'_n, \ldots, e'_1$ successively, we obtain $\mathbb{P}^1 \times \mathbb{P}^1$ with a fibration $\mathbb{P}^1 \times \mathbb{P}^1 \to \mathbb{P}^1$ defined by f. It is easy to see that the map which interchanges l_1 and $f - l_1$, and preserves all other l_i and $f - l_i$, plays the role of the generator of $\mathrm{Out}(D_n) \cong \mathbb{Z}_2$. Therefore we have the following natural definition of B_n-configurations.

Let S be a rational surface with a ruling $f : S \to \mathbb{P}^1$ [ibid.], and $\Sigma \in |-K_S|$, such that $f|_\Sigma : \Sigma \to \mathbb{P}^1$ is a double covering with $0 \in \Sigma$ as a ramification point. Recall that an *exceptional system of length n* on S is an n-tuple $\zeta = (e_1, e_2, \ldots, e_n)$ where the e_i's are exceptional divisors such that $e_i \cdot e_j = -\delta_{ij}$, $e_i \cdot K_S = -1$, $1 \le i, j \le n$. A divisor defining the ruling $f : S \to \mathbb{P}^1$ is still denoted by f, which is effective of arithmetic genus 0.

Definition 5. A *B_n-configuration* on S is an exceptional system of length n (if exists) $\zeta = (e_1, e_2, \ldots, e_n)$ with $e_i \cdot f = 0$ for all i, such that we can consider S as a blow-up of $\mathbb{F}_1$ or $\mathbb{P}^1 \times \mathbb{P}^1$ at n points $x_1, x_2, \ldots, x_n$ on Σ, with corresponding exceptional divisors $e_1, e_2, \ldots, e_n$. When such a ζ exists, we call S a (*rational*) *surface with a B_n-configuration.* Let $\rho \in \mathrm{Out}(D_n)$ be the diagram automorphism. Define $\rho(\zeta) := (f - e_1, e_2, \ldots, e_n)$.

Lemma 6. *Let $\zeta = (e_1, e_2, \ldots, e_n)$ be a B_n-configuration. Then*

$$\rho(\zeta) = (f - e_1, e_2, \ldots, e_n)$$

is also a B_n-configuration.

Proof. By [Leung and Zhang 2009a], if after blowing down $e_n, \ldots, e_1$ successively we obtain $\mathbb{F}_1$, then after blowing down $e_n, \ldots, e_2, f - e_1$ we shall obtain $\mathbb{P}^1 \times \mathbb{P}^1$. Conversely, if after blowing down $e_n, \ldots, e_1$ successively we obtain $\mathbb{P}^1 \times \mathbb{P}^1$, then after blowing down $e_n, \ldots, e_2, f - e_1$ we shall obtain $\mathbb{F}_1$. The result follows. □

When $x_1, \ldots, x_n \in \Sigma$ are in general position (footnote 1), the surface S in Definition 5 is called a *B_n-surface*.

Lemma 7. *Let S be a B_n-surface.*

(i) *Any B_n-configuration on S consists of exceptional curves.*

(ii) *The Weyl group $W(D_n)$ acts on all B_n-configurations with two orbits and acts on each orbit simply transitively.*

(iii) *These two orbits are exchanged by* $\mathrm{Out}(D_n)$.

(iv) *The group $W(D_n) \ltimes \mathrm{Out}(D_n)$ acts on all B_n-exceptional systems simply transitively*

Proof. Let S be a B_n-surface with a ruling $f : S \to \mathbb{P}^1$. Then by definition, S is a blow-up of $\mathbb{F}_1$ or $\mathbb{P}^1 \times \mathbb{P}^1$ at n points $x_1, x_2, \ldots, x_n \in \Sigma$. Let $l_1, \ldots, l_n$ be the corresponding exceptional divisors. Then we have

$$\begin{aligned} &\{\, x \in \mathrm{Pic}(S) \mid x^2 = xK = -1, xf = 0 \,\} \\ &= \{\, l_1, \ldots, l_n, f - l_1, \ldots, f - l_n \,\}. \end{aligned}$$

Thus a B_n-configuration must be of the form: $\zeta = (e_1, \ldots, e_n)$ where $e_i = l_{\sigma(i)}$ or $e_i = f - l_{\sigma(i)}$ and σ is a permutation of $1, \ldots, n$. Obviously, $x_1, \ldots, x_n$ are in general position if and only if all the l_i and $f - l_i$ are exceptional curves. Therefore, (i) is true.

(iii) This follows from Definition 5.

(iv) This is a consequence of (ii) and (iii).

(ii) Let $(e_1, e_2, \ldots, e_n)$ be a B_n-configuration on S. Then $e_i = l_{\sigma(i)}$ or $f - l_{\sigma(i)}$ for $1 \le i \le n$, where σ is a permutation of $\{1, \ldots, n\}$. The Weyl group $W(D_n)$ acts as the group generated by permutations of the n pairs $\{(l_i, f - l_i) \mid i = 1, \ldots, n\}$ and interchanges of l_i and $f - l_i$ simultaneously in two pairs in $\{(l_i, f - l_i) \mid 1 \le i \le n\}$. Therefore $W(D_n)$ acts on the set $\{(e_1, \ldots, e_n) \mid \sum e_i \cdot s \equiv 0 \pmod 2\}$ simply transitively. Similarly the condition $\sum e_i \cdot s \equiv 1 \pmod 2$ determines another orbit on which $W(D_n)$ acts simply transitively. □

Remark 8. Although we know the B_n-configurations on S, unfortunately, we can not single out the B_n-root system within the Picard lattice $\mathrm{Pic}(S) \cong H^2(S, \mathbb{Z})$. However, according to Section 2, we have a root system of B_n type consisting of $\mathbb{Q}$-divisors on S:

$$R(B_n) \triangleq \left\{\pm\left(\tfrac{1}{2} f - l_i\right), \pm(l_i - l_j), \pm(f - l_i - l_j) \,\middle|\, i \ne j, 1 \le i, j \le n\right\}.$$

It is easy to see that the corresponding root lattice is

$$\Lambda(B_n) \triangleq \left\{x \in \mathbb{Z}\left(\tfrac{1}{2} f\right) \oplus \bigoplus_{i=1}^{n} \mathbb{Z}(l_i) \,\middle|\, xf = xK = 0\right\}$$

and

$$R(B_n) = \{x \in \Lambda(B_n) \mid x^2 = -2 \text{ or } x^2 = -1\}.$$

The set of simple roots of B_n is

$$\Delta(B_n) = \left\{\beta_1 = \tfrac{1}{2} f - l_1,\, \beta_i = l_{i-1} - l_i,\, i = 2, \ldots, n\right\}.$$

Recall that the Weyl group $W(B_n)$ is the subgroup of $\mathrm{Aut}(\Lambda(B_n))$ generated by the reflections σ_α with $\alpha \in R(B_n)$.

Corollary 9. *Let $R(B_n)$ be defined as above. Then $W(B_n)$ acts on the set of all B_n-configurations simply transitively.*

Let $\mathcal{S}_\Sigma^{B_n}$ be the moduli space of pairs (S, Σ) where S is a B_n-surface (so the blown-up points $x_1, x_2, \ldots, x_n$ are in general position), and $\Sigma \in |-K_S|$, where two pairs (S, Σ) and (S', Σ) are said to be isomorphic to each other if there is an isomorphism $f : S \xrightarrow{\sim} S'$ such that $f|_\Sigma = \mathrm{id}_\Sigma$. Denote $\mathcal{M}_\Sigma^{B_n}$ the moduli space of flat B_n-bundles over Σ. Let $\underline{\mathcal{S}}_\Sigma^{B_n}$ be the (marked) moduli space of the triples $(S, \Sigma, \zeta = (l_1, \ldots, l_n))$. By Lemma 7, we have

$$\mathcal{S}_\Sigma^{B_n} \cong \underline{\mathcal{S}}_\Sigma^{B_n} / W(B_n) \cong \underline{\mathcal{S}}_\Sigma^{B_n} / (W(D_n) \ltimes \mathrm{Out}(D_n)).$$

Let $(S, \Sigma, \zeta = (l_1, \ldots, l_n)) \in \underline{\mathcal{S}}_\Sigma^{B_n}$ be as above. For all $\alpha = \frac{a_0}{2} f + \sum a_i l_i \in \Lambda(B_n) \subseteq \mathrm{Pic}(S)_\mathbb{Q} = \mathrm{Pic}(S) \otimes \mathbb{Q}$ with $a_i \in \mathbb{Z}, i = 0, \ldots, n$, the invertible sheaf induced by restriction to Σ

$$\mathcal{O}_\Sigma(\alpha) := \mathcal{O}_\Sigma(a_0(0)) \otimes \mathcal{O}\big(\sum a_i l_i\big)|_\Sigma$$

is well-defined. Moreover, $\deg(\mathcal{O}_\Sigma(\alpha)) = \alpha \cdot (-K_S) = 0$. Then

$$\mathcal{O}_\Sigma(\alpha) \in \mathrm{Jac}(\Sigma) \cong \Sigma.$$

Thus there is a morphism

$$\underline{\phi} : \underline{\mathcal{S}}_\Sigma^{B_n} \to \mathrm{Hom}(\Lambda(B_n), \Sigma),$$

which is induced by the restriction: for all $\alpha \in \Lambda(B_n) \subseteq \mathrm{Pic}(S)_\mathbb{Q}$,

$$\underline{\phi}((S, \Sigma, \zeta))(\alpha) := \mathcal{O}_\Sigma(\alpha) \in \mathrm{Jac}(\Sigma) \cong \Sigma.$$

Proposition 10. (i) *$\mathcal{S}_\Sigma^{B_n}$ is embedded into $\mathcal{M}_\Sigma^{B_n}$ as an open dense subset.*

(ii) *This embedding can be extended naturally to an isomorphism*

$$\overline{\mathcal{S}_\Sigma^{B_n}} \cong \mathcal{M}_\Sigma^{B_n},$$

by including all rational surfaces with B_n-configurations.

Proof. Similarly as in [Leung and Zhang 2009a], we have

$$\mathcal{M}_\Sigma^{B_n} \cong \mathrm{Hom}(\Lambda(B_n), \Sigma) / W(B_n).$$

Then by Lemma 7 or Corollary 9, since two different sets of simple roots differ by a $W(B_n)$-action, we just need to show that the map

$$\underline{\phi}: \underline{\mathcal{S}}_{\Sigma}^{B_n} \hookrightarrow \mathrm{Hom}(\Lambda(B_n), \Sigma)$$

is an open dense embedding and that $\underline{\phi}$ can be extended to an isomorphism $\overline{\underline{\phi}}$ from the natural compactification $\overline{\underline{\mathcal{S}}_{\Sigma}^{B_n}}$ of $\underline{\mathcal{S}}_{\Sigma}^{B_n}$ to $\mathrm{Hom}(\Lambda(B_n), \Sigma)$:

$$\overline{\underline{\phi}}: \overline{\underline{\mathcal{S}}_{\Sigma}^{B_n}} \xrightarrow{\sim} \mathrm{Hom}(\Lambda(B_n), \Sigma).$$

The map $\underline{\phi}$ is injective. For this, we take a simple root system of D_n as

$$\beta_1 = \tfrac{1}{2}f - l_1, \qquad \beta_i = l_{i-1} - l_i \quad \text{for } 2 \le i \le n.$$

Then the restriction induces an element $u \in \mathrm{Hom}(\Lambda(B_n), \Sigma)$. For

$$\beta = a_0\big(\tfrac{1}{2}f\big) + \sum a_i l_i \in \Lambda(B_n),$$

let $x_i = l_i \cap \Sigma$ and $p = u(\beta) \in \Sigma$. Then we have an equation

$$\sum a_i x_i = p,$$

where $+$ is the addition on the elliptic curve Σ. Taking $\beta = \beta_i, i = 1, \dots, n$ respectively, and setting $p_i = u(\beta_i)$ accordingly, we obtain the following system of linear equations

$$\begin{cases} -x_1 = p_1, \\ x_{i-1} - x_i = p_i, \ i = 2, \dots, n. \end{cases}$$

Obviously, the solution of this system of linear equations exists uniquely for given p_i with $1 \le i \le n$.

The open dense property of the image of the embedding $\underline{\phi}$ is obvious.

Finally, the statement (ii) comes from the existence of the solutions to the above system of linear equations. □

3.2. *C_n-bundles.* We take $G' = A_1^n \subseteq G = C_n$, where $C_n = \mathrm{Sp}(n)$ and $A_1 = \mathrm{SU}(2)$. Note that $\mathrm{Out}(A_1^n) \cong S_n$.

Let S be a rational surface with an A_1^n-configuration that contains Σ as a smooth anticanonical curve. Recall from [Leung and Zhang 2009a] that S is a (successive) blow-up of $\mathbb{P}^2$ at $2n$ points $x_1, y_1, \dots, x_n, y_n$ on Σ, with corresponding exceptional classes $l_1, l_1', \dots, l_n, l_n'$, where $x_i + y_i = 0 \in \Sigma$. The Picard group of S is $H^2(S, \mathbb{Z})$, which is a lattice with basis $h, l_1, l_1', \dots, l_n, l_n'$. The canonical divisor is $K = -\big(3h - \sum_{i=1}^n (l_i + l_i')\big)$.

A simple root system of A_1^n can be taken as

$$\Delta(A_n^1) = \{\alpha_i = l_i - l_i' \mid 1 \le i \le n\}.$$

When the above simple root system is chosen, the pair (S, Σ) determines a homomorphism $u \in \mathrm{Hom}(\Lambda(G'), \Sigma)$ which is given by the restriction map

$$u(\alpha) := \mathbb{O}(\alpha)|_{\Sigma}.$$

Lemma 11. *Let $u \in \mathrm{Hom}(\Lambda(G'), \Sigma)$ be an element corresponding to a triple (S, Σ, ζ), where S is a surface with an A_1^n-configuration $\zeta = (l_1, l_1', \dots, l_n, l_n')$. Let $\rho \in \mathrm{Out}(G') \cong S_n$. Then $\rho \cdot u$ corresponds to the triple $(S, \Sigma, \rho(\zeta))$, where $\rho(\zeta) = (l_{\rho(1)}, l'_{\rho(1)}, \dots, l_{\rho(n)}, l'_{\rho(n)})$.*

Proof. Since u is the restriction map: $\alpha_i \mapsto \mathbb{O}(\alpha_i)|_{\Sigma}$, $u(\alpha_i) = \mathbb{O}(l_i - l_i')|_{\Sigma} = x_i - y_i$ for $i = 1, \dots, n$. Hence $\rho \cdot u(\alpha_i) = u(\alpha_{\rho(i)}) = x_{\rho(i)} - y_{\rho(i)}$. Therefore we have the result, since $x_{\rho(i)} + y_{\rho(i)} = 0$. □

Thus, it is natural to define a C_n-configuration on S to be the form

$$\zeta = ((l_1, l_1'), \dots, (l_n, l_n')).$$

More precisely, denote S the blow-up of $\mathbb{P}_2$ at n pairs of points $(x_1, -x_1), \dots, (x_n, -x_n)$ on Σ, with n pairs of corresponding exceptional divisors $(l_1, l_1'), \dots, (l_n, l_n')$, where l_i and l_i' are the exceptional divisors corresponding to the blowing up at x_i and $-x_i$, respectively.

Definition 12. A *C_n-exceptional system* on S is an n-tuple of pairs

$$((e_1, e_1'), \dots, (e_n, e_n'))$$

where $(e_i, e_i') = (l_{\sigma(i)}, l'_{\sigma(i)})$ or $(l'_{\sigma(i)}, l_{\sigma(i)})$, $i = 1, \dots, n$, and σ is a permutation of $1, \dots, n$. A *C_n-configuration* on S is a C_n-exceptional system $\zeta_{C_n} = ((e_1, e_1'), \dots, (e_n, e_n'))$ such that after blowing down successively $e_n', e_n, \dots, e_1', e_1$, we obtain the surface $\mathbb{P}^2$.

It can be shown that $x_1, x_2, \dots, x_n \in \Sigma \subseteq \mathbb{P}^2$ are in general position (in the sense of footnote 1) if and only if any C_n-exceptional system on S consists of smooth exceptional curves. Such a surface is called a *C_n-surface*.

Lemma 13. (i) *Let S be a surface with a C_n-configuration. The group $W(A_1^n) \ltimes S_n$ acts on all C_n-exceptional systems on S simply transitively.*

(ii) *Let S be a C_n-surface. The group $W(A_1^n) \ltimes S_n$ acts on all C_n-configurations on S simply transitively.*

Proof. It suffices to prove (i). The Weyl group $W(A_1^n) \ltimes S_n$ acts as the group generated by permutations of the n pairs $\{(l_i, l_i') \mid i = 1, \dots, n\}$ and interchanging of l_i and l_i' for each i. From this, we see that $W(A_1^n) \ltimes S_n$ acts on all C_n-configurations simply transitively. □

Denote by $\mathcal{S}_{\Sigma}^{G'}$ the moduli space of $G' = A_1^n$-surfaces with a fixed anticanonical curve Σ, and by $\overline{\mathcal{S}_{\Sigma}^{G'}}$ the natural compactification by including all rational surfaces with A_1^n-configurations. From [Leung and Zhang 2009a] we know that there is an isomorphism $\phi : \overline{\mathcal{S}_{\Sigma}^{G'}} \xrightarrow{\sim} \mathcal{M}_{\Sigma}^{G'}$.

Denote by $\mathcal{S}_{\Sigma}^{C_n}$ the moduli space of pairs (S, Σ), where S is a C_n-surface, that is, S is the blow-up of $\mathbb{P}^2$ at $2n$ points $\pm x_1, \dots, \pm x_n \in \Sigma$ such that $x_1, \dots, x_n$ are in general position, and two pairs (S, Σ) and (S', Σ) are said to be isomorphic to each other if there is an isomorphism $f : S \xrightarrow{\sim} S'$ such that $f|_{\Sigma} = \mathrm{id}_{\Sigma}$. Denote by $\mathcal{M}_{\Sigma}^{C_n}$ the moduli space of flat C_n-bundles over Σ.

Proposition 14. (i) *$\mathcal{S}_{\Sigma}^{C_n}$ is embedded into $\mathcal{M}_{\Sigma}^{C_n}$ as an open dense subset.*

(ii) *This embedding can be extended naturally to an isomorphism*

$$\overline{\mathcal{S}_{\Sigma}^{C_n}} \cong \mathcal{M}_{\Sigma}^{C_n},$$

by including all rational surfaces with C_n-configurations.

Proof. By Corollary 3, $\mathcal{M}_{\Sigma}^{C_n} \cong \mathcal{M}_{\Sigma}^{A_1^n}/S_n \cong \overline{\mathcal{S}_{\Sigma}^{A_1^n}}/S_n$. Therefore it is sufficient to show that $\mathcal{S}_{\Sigma}^{C_n} \cong \mathcal{S}_{\Sigma}^{A_1^n}/S_n$. This follows from Lemma 13. □

Remark 15. Obviously, this description in Proposition 14 coincides with the well-known description of flat C_n-bundles over elliptic curves [Friedman et al. 1997]. A flat $C_n = \mathrm{Sp}(n)$-bundle over Σ corresponds to n pairs (unordered) of points $(x_i, -x_i)$, $i = 1, \dots, n$ on Σ, uniquely up to isomorphism. One pair $(x_i, -x_i)$ will determine exactly one point on $\mathbb{CP}^1$, since the rational map determined by the linear system $|2(0)|$ induces a double covering from Σ onto $\mathbb{CP}^1$. The moduli space of flat $\mathrm{SU}(2)$-bundles over Σ is isomorphic to $\mathbb{P}^1$. So the moduli space of flat C_n-bundles over Σ is precisely isomorphic to $S^n(\mathbb{CP}^1) = \mathbb{CP}^n$, the ordinary projective n space.

3.3. ***G_2-bundles.*** For $G = G_2$, we take $G' = A_2 = \mathrm{SU}(3)$.

Let S be a rational surface with an A_2-configuration (see [Leung and Zhang 2009a]) containing Σ as a smooth anticanonical curve. Recall [ibid.] that S is a (successive) blow-up of $\mathbb{P}^2$ at 3 points x_1, x_2, x_3 on Σ, with corresponding exceptional classes l_1, l_2, l_3, where $x_1 + x_2 + x_3 = 0 \in \Sigma$. Let h be the class of lines in $\mathbb{P}^2$. The Picard group of S is $Pic(S) \cong H^2(S, \mathbb{Z})$, which is a lattice with basis h, l_1, l_2, l_3. The canonical line bundle $K = -\big(3h - \sum_{i=1}^{3} l_i\big)$.

Recall that

$$\{x \in H^2(S, \mathbb{Z}) \mid x \cdot K = x \cdot h = 0\}$$

is a root lattice of A_2 type. We can take a simple root system of A_2 as

$$\Delta(A_2) = \{\alpha_1 = l_1 - l_2, \alpha_2 = l_2 - l_3\}.$$

Let $\rho \in \mathrm{Out}(A_2) \cong \mathbb{Z}_2$ be the generator of order 2 (we can take $\rho = -1$, that is, $\rho(\alpha_i) = -\alpha_i$).

Denote by $\mathcal{S}_\Sigma^{A_2}$ the moduli space of A_2-surfaces with a fixed anticanonical curve Σ, and $\overline{\mathcal{S}_\Sigma^{A_2}}$ the natural compactification by including all rational surfaces with A_2-configurations. From [Leung and Zhang 2009a] we know that $\overline{\mathcal{S}_\Sigma^{A_2}} \xrightarrow{\sim} \mathcal{M}_\Sigma^{A_2}$. Let ϕ be the isomorphism.

Definition 16. Let S be as immediately above. A *G_2-exceptional system* on S is an ordered triple (e_1, e_2, e_3) of exceptional divisors such that $e_i \cdot e_j = 0 = e_i \cdot h$, $i \neq j$ and $y_1 + y_2 + y_3 = 0$ where $y_i = e_i \cap \Sigma$. A *G_2-configuration* on S is a G_2-exceptional system $\zeta_{G_2} = (e_1, e_2, e_3)$ such that we can consider S as a blow-up of $\mathbb{P}^2$ at these 3 points y_1, y_2, y_3 on Σ, with corresponding exceptional divisors e_1, e_2, e_3. When S has a G_2-configuration (of course $\Sigma \in |-K_S|$), we call S a *(rational) surface with a G_2-configuration*.

When x_1, x_2, x_3 are nonzero distinct points on Σ, any G_2-exceptional system on S consists of exceptional curves. Such a surface is called a *G_2-surface*. These 3 points $x_1, x_2, x_3 \in \Sigma$ are said to be *in general position*.

Let S, S' be two surfaces with G_2-configurations ζ, ζ' respectively. We say that $(S, \Sigma, \zeta) \cong (S', \Sigma, \zeta')$ if there exists an isomorphism $f : S \xrightarrow{\sim} S'$ such that $f|_\Sigma : \Sigma \to \Sigma$ is the identity or the involution of Σ.

A triple (S, Σ, ζ) determines an element u of $\mathrm{Hom}(\Lambda(A_2), \Sigma)$ by the restriction

$$u(\alpha) := \mathcal{O}(\alpha)|_\Sigma.$$

Lemma 17. *Let $u \in \mathrm{Hom}(\Lambda(A_2), \Sigma)$ correspond to the triple (S, Σ, ζ), where S is a surface with a G_2-configuration $\zeta = \{l_1, l_2, l_3\}$. Then $\rho \cdot u$ corresponds to (S', Σ, ζ'), where S' is another surface with a G_2-configuration $\zeta' = (l'_1, l'_2, l'_3)$ with $l'_i \cap \Sigma = -x_i$. Moreover, we have $(S, \Sigma, \zeta) \cong (S', \Sigma, \zeta')$.*

Proof. Since u is the restriction map: $\alpha_i \mapsto \mathcal{O}(\alpha_i)|_\Sigma$, $u(\alpha_1) = \mathcal{O}(l_1 - l_2)|_\Sigma = x_1 - x_2$, $u(\alpha_2) = x_2 - x_3$. Hence $\rho \cdot u = v \Leftrightarrow v(\alpha_i) = -u(\alpha_i) \Leftrightarrow x_1 - x_2 = y_2 - y_1$, $x_2 - x_3 = y_3 - y_2 \Leftrightarrow y_i = -x_i$.

Next we prove the second assertion. We first fix an embedding $\iota : \Sigma \hookrightarrow \mathbb{P}^2$ such that (the image of) Σ is defined by the equation $zy^2 = 4x^3 + axz^2 + bz^3$ and $0 = [0, 1, 0] \in \Sigma$, where $[x, y, z]$ is the coordinate system of $\mathbb{P}^2$. Then the automorphism of $\mathbb{P}^2$ defined by $[x, y, z] \mapsto [x, -y, z]$ induces an isomorphism f of the triple $(0, \Sigma, \mathbb{P}^2)$, which is the involution on Σ that maps $x \in \Sigma$ to $-x$. On the other hand, for $x_1, x_2, x_3 \in \Sigma$, we have obviously $(-x_1) + (-x_2) + (-x_3) = 0$. Thus we have the isomorphism ϕ defined by f. □

Lemma 18. (i) *Let S be a surface with a G_2-configuration. The Weyl group $W(A_2)$ acts on all G_2-exceptional systems on S simply transitively.*

(ii) *Let S be a G_2-surface. The Weyl group $W(A_2)$ acts on all G_2-configurations on S simply transitively.*

(iii) *Let $[(S, \Sigma, \zeta)]$ be the isomorphism class of (S, Σ, ζ). Then $W(A_2) \ltimes \mathbb{Z}_2$ acts on the set $[(S, \Sigma, \zeta)]$ simply transitively.*

Proof. Let $f : (S', \Sigma, \zeta') \xrightarrow{\sim} (S, \Sigma, \zeta)$. If $f|_\Sigma = \mathrm{id}_\Sigma$, then $S = S'$ and $f = \mathrm{id}_S$. In this case, $W(A_2)$ acts on the G_2-configurations on S simply transitively. On the other hand, by Lemma 17, the involution on Σ can be extended to an isomorphism from S' onto S. In this case the involution $-\mathrm{id}_\Sigma$ acts on the set $[(S, \Sigma, \zeta)]$. Thus the result follows. □

Proposition 19. *Let $\mathcal{S}_\Sigma^{G_2}$ be the moduli space of pairs (S, Σ) where S is a G_2-surface, and $\mathcal{M}_\Sigma^{G_2}$ be the moduli space of flat G_2-bundles over Σ. Then we have*

(i) *$\mathcal{S}_\Sigma^{G_2}$ is embedded into $\mathcal{M}_\Sigma^{G_2}$ as an open dense subset.*

(ii) *This embedding can be extended naturally to an isomorphism*

$$\overline{\mathcal{S}_\Sigma^{G_2}} \cong \mathcal{M}_\Sigma^{G_2},$$

by including all rational surfaces with G_2-configurations.

Proof. By Corollary 3 we have $\mathcal{M}_\Sigma^{G_2} \cong \mathcal{M}_\Sigma^{A_2} / \operatorname{Out}(A_2) \cong \overline{\mathcal{S}_\Sigma^{A_2}}/\mathbb{Z}_2$. Thus it suffices to show that $\mathcal{S}_\Sigma^{G_2} \cong \mathcal{S}_\Sigma^{A_2}/\mathbb{Z}_2$. This follows from Lemma 18. □

Remark 20 [Friedman et al. 1997]. A SU(3)-bundles over Σ is determined by a section of $H^0(\mathbb{O}_\Sigma(3(0)))$, which is a meromorphic function with the only pole 0 of order at most 3. Let x, y be the local coordinates of Σ around 0, then this function is $a_0 + a_1 x + a_2 y$ up to nonzero constant. Thus the moduli space $\mathcal{M}_\Sigma^{A_2}$ is isomorphic to $\mathbb{P}^2$. By the proof of Lemma 17, the function $a_0 + a_1 x + (-a_2) y$ defines the same G_2-bundle over Σ. Thus we have $\mathcal{M}_\Sigma^{G_2} \cong \mathbb{WP}^2_{1,1,2}$.

Remark 21. For the F_4 case, unfortunately, the method used in this paper is not very effective. We can not find a suitable definition for F_4-configurations. Thus in this case, the method used in [Leung and Zhang 2009b] is the better one.

Acknowledgments

We would like to express our heartfelt gratitude to the referee for a very careful reading of our paper and valuable suggestions that improved the paper greatly.

References

[Bourbaki 2005] N. Bourbaki, *Lie groups and Lie algebras*, Chapter 7–9, Springer, Berlin, 2005. MR 2005h:17001 Zbl 1139.17002

[Demazure et al. 1980] M. Demazure, H. C. Pinkham, and B. Teissier (editors), *Séminaire sur les Singularités des Surfaces* (Palaiseau, 1976–1977), Lecture Notes in Mathematics **777**, Springer, Berlin, 1980. MR 82d:14021 Zbl 0415.00010

[Donagi 1997] R. Y. Donagi, "Principal bundles on elliptic fibrations", *Asian J. Math.* **1**:2 (1997), 214–223. MR 99d:14010 Zbl 0927.14006 arXiv alg-geom/9702002

[Donagi 1998] R. Y. Donagi, "Taniguchi lectures on principal bundles on elliptic fibrations", pp. 33–46 in *Integrable systems and algebraic geometry* (Kobe/Kyoto, 1997), edited by M.-H. Saito et al., World Scientific, River Edge, NJ, 1998. MR 2000a:14015 Zbl 0963.14004

[Friedman et al. 1997] R. Friedman, J. W. Morgan, and E. Witten, "Vector bundles and F theory", *Comm. Math. Phys.* **187**:3 (1997), 679–743. MR 99g:14052 Zbl 0919.14010

[Humphreys 1978] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics **9**, Springer, New York, 1978. MR 81b:17007 Zbl 0447.17001

[Lee 2010] J.-H. Lee, "Configuration of lines in del Pezzo surfaces with Gosset polytopes", preprint, 2010. arXiv 1001.4174

[Lee 2012] J.-H. Lee, "Gosset polytopes in Picard groups of del Pezzo surfaces", *Canad. J. Math.* **64**:1 (2012), 123–150. MR 2932172 Zbl 06023029

[Leung 2000] N. C. Leung, "ADE-bundle over rational surfaces, configuration of lines and rulings", preprint, 2000. arXiv math/0009192

[Leung and Zhang 2009a] N. C. Leung and J. Zhang, "Moduli of bundles over rational surfaces and elliptic curves, I: Simply laced cases", *J. Lond. Math. Soc.* (2) **80**:3 (2009), 750–770. MR 2011e:14018 Zbl 1188.14025

[Leung and Zhang 2009b] N. C. Leung and J. Zhang, "Moduli of bundles over rational surfaces and elliptic curves, II: Nonsimply laced cases", *Int. Math. Res. Not.* **2009**:24 (2009), 4597–4625. MR 2011e:14019 Zbl 1222.14023

[Leung et al. 2012] N. C. Leung, M. Xu, and J. Zhang, "Kac–Moody $\widetilde{E}_k$-bundles over elliptic curves and del Pezzo surfaces with singularities of type A", *Math. Ann.* **352**:4 (2012), 805–828. MR 2892453 Zbl 1242.14036

[Looijenga 1976] E. Looijenga, "Root systems and elliptic curves", *Invent. Math.* **38**:1 (1976), 17–32. MR 57 #6015 Zbl 0358.17016

[Manin 1974] Y. I. Manin, *Cubic forms: algebra, geometry, arithmetic*, North-Holland Mathematical Library **4**, North-Holland, Amsterdam, 1974. MR 57 #343 Zbl 0277.14014

Received September 25, 2011.

MANG XU
DEPARTMENT OF MATHEMATICS
SOUTHWEST JIAOTONG UNIVERSITY
CHENGDU, 610031
CHINA
xumang@home.swjtu.edu.cn

JIAJIN ZHANG
DEPARTMENT OF MATHEMATICS
SICHUAN UNIVERSITY
CHENGDU, 610065
CHINA
jjzhang@scu.edu.cn

CONTENTS

Volume 261, 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 261 No. 2 February 2013

0030-8730(201302)261:2;1-B