# Pacific Journal of Mathematics

Volume 295 No. 2 August 2018

# PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm





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

The subscription price for 2018 is US $475/year for the electronic version, and $640/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber 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 Web of Knowledge (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 twelve times a year. 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. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.257

# NONSMOOTH CONVEX CAUSTICS FOR BIRKHOFF BILLIARDS

MAXIM ARNOLD AND MISHA BIALY

**This paper is devoted to the examination of the properties of the string construction for the Birkhoff billiard. Based on purely geometric considerations, string construction is suited to providing a table for the Birkhoff billiard, having the prescribed caustic. Exploiting this framework together with the properties of convex caustics, we give a geometric proof of a result by Innami first proved in 2002 by means of Aubry–Mather theory. In the second part of the paper we show that applying the string construction one can find a new collection of examples of $C^2$-smooth convex billiard tables with a nonsmooth convex caustic.**

## 1. Introduction

Let $\Gamma$ be a simple closed $C^1$-smooth convex curve in the Euclidean plane. We consider a Birkhoff billiard inside $\Gamma$. This simple dynamical system creates many geometric and dynamical questions and reflects many difficulties appearing in general Hamiltonian systems. Readers may refer to any textbook among the wide variety written on the subject (e.g., [Katok et al. 1986; Kozlov and Treshchëv 1991; Mather and Forni 1994; Tabachnikov 2005]).

We will use the following nonstandard notations: the interior of the set bounded by the simple closed curve $\gamma$ will be denoted by $\gamma^\circ$, while $\bar{\gamma}$ denotes the compact $\gamma^\circ \cup \gamma$. The length of the curve is denoted by $\mathsf{Length}(\gamma)$. The convex hull of $\gamma$ is denoted by $\mathsf{Conv}(\gamma)$.

**Definition 1.** A simple closed curve $\gamma \subset \Gamma^\circ$ is called a *convex caustic* for $\Gamma$ if $\bar{\gamma}$ is a convex set and any supporting line for $\bar{\gamma}$ remains a supporting line for $\bar{\gamma}$ after billiard reflection in $\Gamma$.

Every convex caustic $\gamma$ corresponds to an invariant curve $r_\gamma$ of the billiard ball map. The curve $r_\gamma \subset \mathbb{R}_+ \times \mathbb{S}^1$ consists of all supporting lines to $\gamma$. This curve winds once around the phase cylinder and therefore is called rotational. We shall denote its rotation number by $\rho_\gamma$.

*MSC2010:* primary 37E30, 37E40; secondary 78A05.
*Keywords:* string construction, convex caustics, Birkhoff billiard.

In the original Birkhoff paper [1917] there was posed a conjecture that the existence of a continuous set of caustics, being a very restrictive property, actually provides an extreme rigidity on the shape of the curve $\Gamma$. The first result in this direction was achieved in [Bialy 1993]. Our paper is motivated by recent progress in the Birkhoff conjecture solution achieved in [Avila et al. 2016; Kaloshin and Sorrentino 2016]. The crucial assumption in these papers consists in the existence of convex caustics such that the rotation numbers of the corresponding invariant curves form a rational sequence in the interval $\left(0; \frac{1}{3}\right]$, converging to 0. It seems natural to compare such a result with one proved by N. Innami [2002].

**Theorem 2** [Innami 2002]. *Assume that there exists a sequence of convex caustics $\gamma_n$ inside $\Gamma$ such that the rotation numbers $\rho_n$ of the corresponding invariant curves tend to $\frac{1}{2}$. Then $\Gamma$ is an ellipse.*

Originally, Innami's arguments were based on the Aubry–Mather variational theory. In the next section we present a simple geometric proof using string construction. Yet, it remains a challenging question whether one can prove a more general statement relaxing the requirement of convexity of the caustics.

Let us recall the string construction framework. Given a convex compact set $\bar{\gamma}$ bounded by $\gamma$, and a number $S > \mathsf{Length}(\gamma)$, define the curve $\Gamma$ as a union of those points $P$ such that the *cap-body* $\mathsf{Conv}(P \cup \bar{\gamma})$ has boundary of length $S$. Geometrically such a construction gives the set of all points traversed by the tip of a nonelastic string of length $S > \mathsf{Length}(\gamma)$ wrapped around $\gamma$ and stretched to its full extent. The curve $\Gamma$ provided by such construction has $\gamma$ as its billiard caustic. We shall refer to $S$ as a string parameter of the caustic. A closely related so-called Lazutkin parameter is defined as $L = S - \mathsf{Length}(\gamma)$.

The string construction is widely known and can be easily proved to provide $\Gamma$ for smooth enough $\gamma$. In fact it remains valid also in the more general case as it is stated in the following theorem.

**Theorem 3** [Stoll 1930; Turner 1982].

(1) *For a given compact convex set $\bar{\gamma}$ and for every $S > \mathsf{Length}(\gamma)$ the string construction determines a $C^1$-smooth convex closed curve $\Gamma$ such that $\gamma$ is a billiard caustic for $\Gamma$.*

(2) *If $\gamma$ is a convex billiard caustic for a $C^1$ curve $\Gamma$ then $\Gamma$ can be obtained from $\gamma$ by the string construction for some $S$.*

Let us emphasize that the string construction is highly nonexplicit and makes calculations difficult. A very important consequence of KAM theory, proved by Lazutkin [1973; 1981] and Douady [1982], states the existence of convex caustics near the boundary of a sufficiently smooth (at least $C^6$) billiard table. On the other hand, applying string construction to the triangle, one gets a billiard table which is

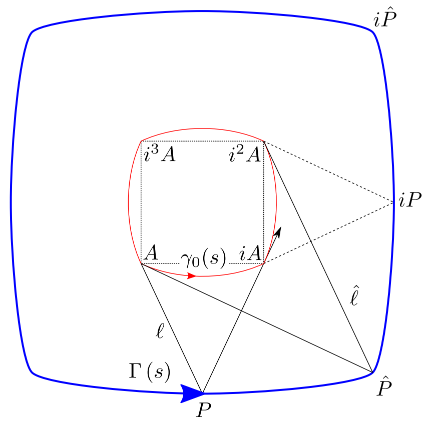



**Figure 1.** A switched caustic string construction.

piecewise $C^2$ with jumps of the curvature and hence by [Hubacher 1987] cannot have caustics near the boundary.

The scenario of destruction of caustics when one moves away from the boundary towards the interior could be understood in principle by the analogy with wave front propagation inside a convex curve [Mather and Forni 1994]. For example, take the ellipse and consider the wave fronts as in the famous picture [Arnold 1990, Figure 36]. For small distances the fronts remain smooth, but starting from some critical value they start to develop singularities. However, nobody has observed such a bifurcation in practice for caustics of convex billiards due to the lack of integrable examples. On the other hand, nonconvex caustics exist, for instance, for convex bodies of constant width, and were studied in [Knill 1998].

Motivated by the above discussion, the natural question about the existence of nonsmooth convex caustics arises. More generally, it is natural to study how irregular the convex caustic can be. In [Fetter 2012] a billiard table of class $C^2$ was constructed which has a caustic of a regular hexagon. In this paper we were able to construct the whole functional family of the examples of $C^2$ billiard tables having nonsmooth convex caustics.

**Theorem 4.** *There exist a one-parametric family of strictly convex nonsmooth compact sets $\bar{\gamma}$ and values of the string parameter $S$ such that the curves $\Gamma$ obtained by the string construction are $C^2$-smooth.*

We will use the following geometric idea (we use the complex notation $x+iy$ for points $(x, y)$ in the plane). Start with a curve $\gamma_0(t) : [-1, 1] \to \mathbb{C}$ such that $\gamma_0(-1) = A = -1-i$, $\gamma_0(1) = iA = 1-i$ and $\gamma_0(t)$ is symmetric with respect to the vertical axis (i.e., $i\gamma_0(-t) = \overline{i\gamma_0(t)}$) (see Figure 1). Construct $\gamma$ as a concatenation of $\{i^k\gamma_0\}_{k=0}^{3}$. Parametrize $\gamma$ by the arc-length parameter $s$ and choose the initial point in such a way that $\gamma(0) = A$. We will denote the total length of $\gamma$ by $4S$. Then $\gamma(S) = iA$.

The main idea is to choose the curve $\gamma$ and string parameter $S$ in such a way that the string construction will have the following properties:

- At the beginning (point $P$ in Figure 1), the left part $AP$ of the string remains fixed at point $A$ while the right part of the string unwinds from the arc $(\widehat{iA, i^2A})$.
- At the moment when the left part of the string becomes tangent to $\gamma$ at the point $A$ (this corresponds to the point $\hat{P}$ on $\Gamma$) the right part reaches the point $i^2A$ and remains fixed after that. We will call this moment the *switching of the first kind*.
- While the left part of the string winds around the arc $(\widehat{A, iA})$ the right part remains fixed at $i^2A$ (see Figure 1) until the moment when the vertex of the string reaches the point $iP$. We will call this the *switching of the second kind*.
- $D_4$ symmetry provides the whole picture.

Let us reemphasize, that the string construction, being a nonexplicit procedure, typically does not provide any analytic expression for the table $\Gamma$ from a given $\gamma$. In the example [Fetter 2012], the construction is made explicit by fixing two end-points on the string. The disadvantage of such a situation is the complete loss of any flexibility, since the corresponding table may consist only of the elliptic arcs. We propose another, more flexible yet explicit construction, fixing only one end-point of the string and allowing another point to slide along the given curve $\gamma$.

***Structure of the paper.*** In the next section we will provide geometric arguments for the proof of Theorem 2. Section 3 is devoted to the construction of the $C^2$ tables with nonsmooth caustics. In Section 4 we will pose some open questions arising in our considerations.

## 2. Geometric proof of Innami's result

We will start with the following simple remarks.

**Remark 5.** If the billiard in $\Gamma$ has a convex caustic $\gamma$ with $\gamma^\circ = \varnothing$ then $\Gamma$ is either an ellipse or a circle.

Indeed, the condition $\gamma^\circ = \varnothing$ for convex $\gamma$ means that $\gamma$ is either a point or a segment. The rest follows from the string construction.

**Remark 6.** Recall that for any point $P$ and for any convex body with nonempty interior there exist exactly two supporting lines to the body passing through $P$. Moreover if the convex caustic $\gamma$ has nonempty interior, then every supporting line to $\bar{\gamma}$ after reflection in $\Gamma$ at point $P$ becomes the second supporting line to $\bar{\gamma}$ from $P$.

Indeed, assume that there exists a supporting line $l$ to $\bar{\gamma}$ which is reflected to itself at a point $P \in \Gamma$. This means that $l$ is orthogonal to $\Gamma$ at $P$. Let $l'$ be the other supporting line to $\bar{\gamma}$ passing through $P$. Then by the definition of convex caustic, the line $l'$ is also reflected to itself at the point $P$ and hence is also orthogonal to $\Gamma$ at $P$. Thus $l$ and $l'$ coincide, which contradicts the assumption that $l$ and $l'$ are two different supporting lines to $\bar{\gamma}$.

**Lemma 7.** *Let $\gamma$ be a convex caustic for $\Gamma$. Then $\gamma^\circ \neq \varnothing$ if and only if the rotation number of the corresponding invariant curve is strictly less then $\frac{1}{2}$.*

*Proof.* If a convex caustic $\gamma$ has empty interior then, by the Remark 5, $\Gamma$ is necessarily an ellipse (or a circle) and the invariant curve corresponding to $\gamma$ has rotation number $\frac{1}{2}$ since it contains a diameter. Vice versa, any convex caustic with nonempty interior has a rotation number strictly less than $\frac{1}{2}$, since otherwise the invariant curve corresponding to the caustic would have a 2-periodic orbit, i.e., a diameter, which is not possible due to Remark 6. □

Let $\gamma_n$ be a sequence of convex caustics for $\Gamma$ with the rotation numbers $\rho_n \in \left(0; \frac{1}{2}\right]$ of corresponding invariant curves. By Lemma 7 we may assume that $\rho_n < \frac{1}{2}$ since otherwise $\gamma_n$ has empty interior and then $\Gamma$ must be an ellipse by the Remark 5. Passing to a subsequence we can assume with no loss of generality that $\rho_n$ is strictly increasing, $\rho_n \nearrow \frac{1}{2}$.

**Lemma 8.** *Let $\gamma_1$ and $\gamma_2$ be two convex caustics for $\Gamma$. If the corresponding invariant curves have rotation numbers $\rho_1 < \rho_2$, then $\bar{\gamma}_2 \subset \gamma_1^\circ$.*

*Proof.* Assume that $\bar{\gamma}_2$ is not a subset of $\gamma_1^\circ$. Then there are only three possibilities: (1): $\bar{\gamma}_1 \cap \bar{\gamma}_2 = \varnothing$; (2): $\gamma_1 \cap \gamma_2 \neq \varnothing$ or (3): $\bar{\gamma}_1 \subset \gamma_2^\circ$.

In the third case one obviously has $\rho_1 \geq \rho_2$ contrary to the assumption of the lemma. In the first and the second cases there necessarily exists a supporting line to both $\bar{\gamma}_1$ and $\bar{\gamma}_2$. Therefore, all billiard reflections in $\Gamma$ of this line are also supporting lines for both $\bar{\gamma}_1$ and $\bar{\gamma}_2$. This means that there exists a whole infinite orbit lying in the intersection of the invariant curves $r_1$ and $r_2$ corresponding to $\gamma_1$ and $\gamma_2$. But then $\rho_1$ must be equal to $\rho_2$, since the rotation number is completely determined by one orbit. □

**Remark 9.** The statement of Lemma 8 holds true also in the opposite direction which will not be used below. Namely, $\bar{\gamma}_2 \subset \gamma_1^\circ$ implies $\rho_1 < \rho_2$. As we already mentioned in the proof, it is obvious that $\rho_1 \leq \rho_2$. In addition $\rho_1$ cannot be equal to $\rho_2$. Otherwise there exist two disjoint graphs of $r_1$ and $r_2$ with the same rotation number, invariant under the billiard map of the cylinder, which is impossible since a billiard map is a twist map (see [Katok and Hasselblatt 1995, p. 428]).

Let $\{S_n\}$ be the sequence of string parameters corresponding to the caustics $\gamma_n$. Then by Lemma 8, $S_n$ is decreasing. Denote $S = \lim_{n\to\infty} S_n$.

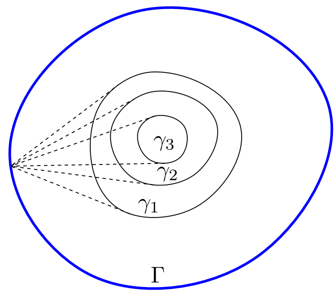



**Figure 2.** A family of nested convex caustics with decreasing string parameter.

**Lemma 10.** *The boundary of the intersection set*

$$C = \bigcap_{n=1}^{\infty} \bar{\gamma}_n$$

*is a convex caustic for* $\Gamma$ *with string parameter* $S$.

*Proof.* The intersection set $C$ is compact and convex. Moreover, it is easy to see that $\partial_C$ is also a caustic with string parameter $S$. Indeed, this follows from the following geometric consideration (see Figure 2). Fix a point $P$ on $\Gamma$ and consider the cap-bodies

$$K_n = \mathsf{Conv}(P \cup \bar{\gamma}_n), \qquad K = \mathsf{Conv}(P \cup C).$$

Then, obviously,

$$K_n \subseteq K, \qquad K = \bigcap_{n=1}^{\infty} K_n,$$

and moreover

$$\mathsf{Length}(\partial_{K_n}) = S_n \to S = \mathsf{Length}(\partial_K).$$

In addition, since $\gamma_n$ is a caustic then $S_n$ does not depend on $P \in \Gamma$ (by Theorem 3). Therefore, $S$ also does not depend on $P$, and hence $C$ reconstructs $\Gamma$ via string construction. Thus $\partial_C$ is a caustic by Theorem 3. □

The last step in the proof of Theorem 2 consists in the following Lemma.

**Lemma 11.** *The limit caustic* $\partial_C$ *has empty interior.*

*Proof.* First notice that it follows from continuity of the invariant curves and their rotation numbers that the invariant curve corresponding to $C$ has rotation number $\frac{1}{2}$. Then from Lemma 7 we conclude that $\partial_C$ has empty interior. □

## 3. Nonsmooth caustic

The main idea of the proof of our result is to carefully choose the Lazutkin parameter and the germ of the function $\gamma$ at the point $A$. While a vertex of the string slides in

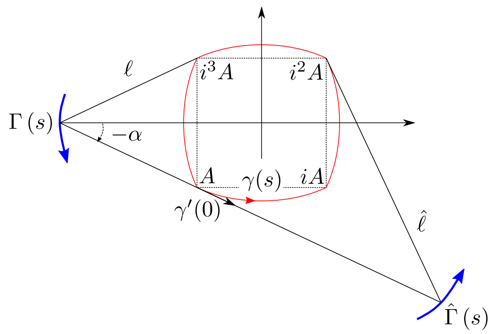



**Figure 3.** A switched caustic string construction.

the regime corresponding to the unwinding from $\gamma(s)$, its trajectory corresponds to the smooth curve. Thus we have to take care of the smoothness of $\Gamma$ near only two points corresponding to the switching moments of the first and second kinds respectively. We will denote by $\Gamma(s)$ the part of $\Gamma$ corresponding to the switching of the second kind about the point $A$. The part of $\Gamma$ corresponding to the switching of the first kind about the point $A$ will be denoted by $\hat{\Gamma}$. The smoothness conditions read as follows: all odd terms in the germs of $\Gamma$ and $\hat{\Gamma}$ have to be orthogonal to the axis of the symmetry while all the even terms must be collinear with the axis of symmetry. Indeed, let $\Gamma(s)$ be the curve symmetric with respect to the line $l$ and intersecting $l$ at the point $\Gamma(0)$. Let $R_l$ be the reflection of the plane in the line $l$. Differentiating the identity

$$R_l\Gamma(s) = \Gamma(-s)$$

$n$ times, at $s = 0$, we get

$$R_l(\Gamma^{(n)}(0)) = (-1)^n\Gamma^{(n)}(0).$$

***Coordinate formulation.*** Parametrize the curve $\gamma$ by the arc-length parameter $s$, so that $|\gamma'(s)| = 1$. Choose the initial point such that $\gamma(0) = A$. Denote by $\alpha$ the angle between $\gamma'(0)$ and the horizontal axis. Then one easily obtains a parametrization for $\Gamma$ and $\hat{\Gamma}$ (see Figure 3):

$$\begin{aligned} \Gamma(s) &= \gamma(s) - t(s)\gamma'(s), \\ \hat{\Gamma}(s) &= \gamma(s) + \hat{t}(s)\gamma'(s), \end{aligned} \tag{1}$$

where $t(s)$ and $\hat{t}(s)$ are some functions of $s$ denoting the length of the right part of the string near the point $\Gamma(s)$ and the left part of the string near the point $\hat{\Gamma}(s)$ correspondingly. Functions $t$ and $\hat{t}$ can be found from the condition of the string to be unstretchable. We will denote $iA = B$.

$$\begin{aligned} |\Gamma(s) + B| + |t\gamma'(s)| - s &= 2\ell, \\ |\hat{\Gamma}(s) + A| + |\hat{t}\gamma'(s)| + s &= 2\hat{\ell}, \end{aligned} \tag{2}$$

where $\ell = 1/\sin\alpha$ and $\hat{\ell} = \sqrt{2}/\sin(\pi/4-\alpha)$. Simple computations yield:

$$(3)\qquad \begin{aligned} t(s) &= \frac{p(s)}{p'(s)}, \ \text{with } p(s) = \tfrac{1}{2}\big((s+2\ell)^2 - |\gamma(s)+B|^2\big),\\ \hat{t}(s) &= -\frac{\hat{p}(s)}{\hat{p}'(s)}, \ \text{with } \hat{p}(s) = \tfrac{1}{2}\big((s-2\hat{\ell})^2 - |\gamma(s)+A|^2\big). \end{aligned}$$

Finally, introducing (3) into (1) we get

$$(4)\qquad \Gamma(s) = \gamma(s) - \frac{p(s)}{p'(s)}\gamma'(s), \qquad \hat{\Gamma}(s) = \gamma(s) - \frac{\hat{p}(s)}{\hat{p}'(s)}\gamma'(s).$$

Orient the curve $\gamma$ as it is shown in Figure 3. We will use the complex notation for the coordinates of the points. Then smoothness conditions for the $n$-th derivative of $\Gamma$ read

$$(5)\qquad \Re(i^{n-1}\Gamma^{(n)}(0)) = 0, \qquad \Re(i^{n-1}\hat{\Gamma}^{(n)}(0)) = \Im(i^{n-1}\hat{\Gamma}^{(n)}(0)).$$

Here $\Re$ and $\Im$ stand for the real and imaginary part of the complex number. For the curve $\gamma(s)$ we get the following parametrization:

$$(6)\qquad \gamma(s) = A + \int_0^s \exp\{i(\varphi(t)-\alpha)\}\,dt, \qquad \text{where } \varphi(t) = \sum_{n=0}^{\infty}\varphi_n t^n.$$

Thus $\varphi_0 = 0$, and $\varphi_n$ corresponds to the $(n-1)$-st derivative of the curvature $\kappa$.

**Lemma 12.** *The smoothness conditions in* (5) *for $n = 1$ are always satisfied.*

This lemma follows from the fact that any $C^0$ caustic produces a $C^1$ table via string construction. However, we present a more analytic proof of this result for the sake of completeness.

*Proof.* Switching of the second kind. From (4) we get

$$\Gamma' = \Big(1 - \Big(\frac{p}{p'}\Big)'\Big)\gamma' - \frac{p}{p'}\gamma''.$$

Therefore the conditions in (5) read $\Re(p''\gamma' - p'\gamma'') = 0$. We will denote $z_1 \cdot z_2 := \frac{1}{2}\Re(z_1\bar{z}_2)$. Using (3) we get

$$p' = -(A+B)\cdot\gamma' + 2\ell, \qquad p'' = -(A+B)\cdot\gamma''.$$

From (6) it follows that $\gamma'' = i\kappa\gamma'$ thus $p''\gamma' - p'\gamma''$ can be written as

$$\begin{aligned} p''\gamma' - p'\gamma'' &= \tfrac{1}{2}\big(-\Re((A+B)\overline{i\kappa\gamma'})\gamma' + \Re((A+B)\bar{\gamma}')(i\kappa\gamma') - 4\ell i\kappa\gamma'\big)\\ &= i\kappa(A+B-2\ell\gamma'). \end{aligned}$$

Thus

$$\Re(p''\gamma' - p'\gamma'') = \kappa\Im(A+B-2\ell\gamma').$$

The latter is identically zero since $\ell\gamma'(0) = \Gamma(0) - \gamma(0)$ and so $\Im(\ell\gamma') = \Im(A)$ (see Figure 3).

Switching of the first kind. Similarly, the smoothness conditions in (5) read

$$\Re(\hat{p}''\gamma' - \hat{p}'\gamma'') = \Im(\hat{p}''\gamma' - \hat{p}'\gamma''),$$

where

$$\hat{p}' = -(2A)\cdot\gamma' - 2\hat{\ell}, \qquad \hat{p}'' = -(2A)\cdot\gamma''$$

and so

$$\hat{p}''\gamma' - \hat{p}'\gamma'' = \big(\Re(Ai\kappa\bar{\gamma}')\gamma' + \Re(A\bar{\gamma}')(i\kappa\gamma') + 2\hat{\ell}i\kappa\gamma'\big) = 2i\kappa(A + \hat{\ell}\gamma').$$

The real part of the right-hand side of the latter is always equal to the imaginary part by the definition of $\hat{\ell}$. □

The two conditions in (5) for $n = 2$ provide, via computations similar to the above, two equations for parameters $\varphi_1$ and $\varphi_2$ with coefficients depending on $\alpha$:

$$\frac{\varphi_1^2 \sin\alpha - \varphi_1 \sin\alpha\cos\alpha - \varphi_2\cos\alpha}{\sin\alpha\cos^2\alpha} = 0,$$

$$\frac{\varphi_1(\cos 2\alpha + 2(\sin\alpha - \cos\alpha)\varphi_1) - 2(\cos\alpha + \sin\alpha)\varphi_2}{(\cos\alpha - \sin\alpha)(1 + \sin 2\alpha)} = 0.$$

The latter system has a solution,

$$\varphi_1 = \tfrac{1}{2}\cos\alpha(1 + \sin 2\alpha), \qquad \varphi_2 = -\tfrac{1}{8}\cos^2 2\alpha \sin 2\alpha, \tag{7}$$

which provides a family of germs for $\gamma$, depending on the parameter $\alpha$, guaranteeing the $C^2$-smoothness for the table $\Gamma$.

Next we will need to construct the whole curve $\gamma$ providing the needed phenomenon in the string construction. Recall that our geometric idea was based on the construction of the curve $\gamma_0$ (see Figure 1). Thus we need to present a convex curve of length $\boldsymbol{S}$, starting at $A$ and ending at $iA$, having tangent slope $-\alpha$ at the left end and being symmetric with respect to the vertical axis. We define $\gamma$ from $\varphi$ through (6). In order to finish the construction we have to prove the following theorem.

**Theorem 13.** *There exists a strictly monotonically increasing function $\varphi(s)$ satisfying the following three conditions*: (1) *$\varphi(s)$ has the given germ* (7) *at $s = 0$*, (2) $\varphi_0(\boldsymbol{S}/2) = \alpha$ *and* $\varphi_{2n}(\boldsymbol{S}/2) = 0$ *for* $n \geqslant 1$, *and* (3) $\int_0^{\boldsymbol{S}/2} \cos\varphi(s)\,ds = 1$.

*Proof.* The Borel theorem states that every power series is the Taylor series of some smooth function. Obviously, using cutting off, one can find a smooth function having a given Taylor series at two given points. Thus there exists a nonempty set $\Psi$ of $C^\infty$ functions having given germs at $s = 0$ and $s = \boldsymbol{S}/2$. Since for $\alpha < \frac{\pi}{2}$ the term $\varphi_1$ in (7) is positive, one may assume without loss of generality that $\Psi$ consists of strictly monotonically increasing functions. Therefore the only condition which

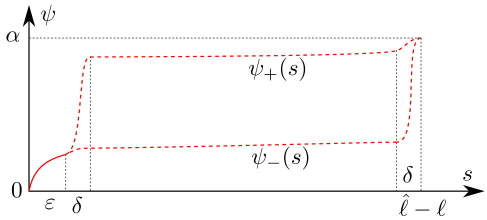



**Figure 4.** Construction of the solution.

has to be satisfied is Theorem 13(3). Taking a small enough $\varepsilon$-step in $s$ we can ensure $\psi(\varepsilon) < \frac{\alpha}{100}$ for all $\psi \in \Psi$. Next we choose two functions $\psi_-$ and $\psi_+$ from the set $\Psi$ as in Figure 4. That is, $\psi_+(s)$ is almost equal to $\alpha$ for $s \in (\varepsilon+\delta, \boldsymbol{S}/2-\delta)$ and $\psi_-(s)$ is almost equal to $\psi(\varepsilon)$ for $s \in (\varepsilon, \boldsymbol{S}/2-\delta)$ for small enough $\delta$. We will look for $\varphi$ as a convex combination $\varphi(s) = l\psi_-(s) + (1-l)\psi_+(s)$. Therefore $\varphi(s)$ obviously satisfies conditions 1 and 2. If we may choose $\psi_\pm$ in such a way that

$$(\boldsymbol{S}/2)\cos\alpha < \int_0^{\boldsymbol{S}/2} \cos(\psi_-(s)-\alpha)\,ds < 1 \quad \text{and} \quad \boldsymbol{S}/2 > \int_0^{\boldsymbol{S}/2} \cos(\psi_+(s)-\alpha)\,ds > 1 \tag{8}$$

then there exists $l$ such that $\int_0^{\boldsymbol{S}/2} \cos(\varphi(s))\,ds = 1$, thus satisfying condition Theorem 13(3). Hence it is sufficient to check that the conditions in (8) can be satisfied for an open set of parameters $\alpha$. Recall that by the construction $\boldsymbol{S} = 2\hat{\ell} - 2\ell$. From the first inequality in (8) we obtain, since $\alpha < \frac{\pi}{4}$,

$$\hat{\ell} - \ell = \frac{2}{\cos\alpha - \sin\alpha} - \frac{1}{\sin\alpha} < \frac{1}{\cos\alpha}.$$

This condition can be interpreted as follows: *the length of the curve $\gamma$ cannot exceed the sum of the lengths of the segments of the two tangent lines from point $P$ to $\gamma$* (*see* Figure 1). The latter inequality is satisfied whenever $\tan 2\alpha < 1$ or

$$\alpha < \frac{\pi}{8}. \tag{9}$$

The second condition in (8) has the following geometric interpretation: *the length of $\gamma$ cannot be less than the distance between points $A$ and $B$.* This yields:

$$3\sin\alpha - \cos\alpha > \cos\alpha\sin\alpha - \sin^2\alpha.$$

Since the latter is satisfied for $\alpha = \frac{\pi}{8}$ we have found an open set of $\alpha$ for which one can find appropriate functions $\psi_-$ and $\psi_+$ shown in Figure 4. □

**Remark 14.** Since the conditions in (5) provide two conditions on $\varphi_n$ to obtain $C^3$ of $\Gamma$ one gets four equations for $\varphi_1$, $\varphi_2$, $\varphi_3$ and $\alpha$. Although the number of parameters matches the number of equations, the corresponding value of $\alpha$ violates (9). Since (9) arises from the construction based on square symmetry, there

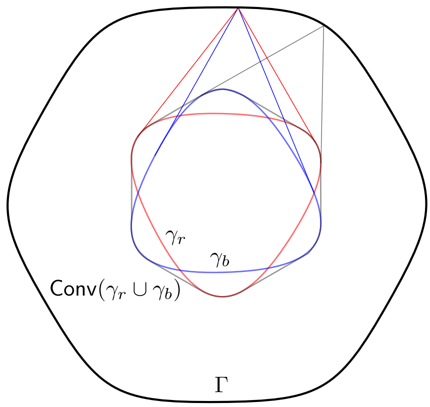



**Figure 5.** The convex hull of two intersecting caustics is also a caustic.

is a hope that starting from other regular polygons one can obtain an inequality which can be satisfied. However, we haven't found any such examples.

## 4. Open problems

Here we want to highlight some general questions which are ultimately related to the string construction. Since the string construction is implicit these questions turn out to be nontrivial.

**Question 15.** Is it possible to have two convex caustics $\gamma_1$ and $\gamma_2$ of $\Gamma$ such that neither of them is a subset of the interior of the other?

In such a case $\gamma_1$ and $\gamma_2$ must have the same rotation number since there is a line tangent to both of the caustics. Moreover it is obvious that $\bar{\gamma}_1$ and $\bar{\gamma}_2$ cannot be disjoint. So the question is if it is possible that two convex caustics have nontrivial intersection. In such a case their convex hull is also a caustic. One can strengthen the question:

**Question 16.** Is it possible for a $\Gamma$ which is symmetric with respect to a certain axis to have a convex caustic $C$ which is not symmetric with respect to this axis?

For example one could imagine two caustics forming a rounded Star of David (Figure 5). The answer to the quantum analog of this question is positive: for a symmetric domain the Dirichlet eigenfunction can be nonsymmetric. We could not however decide if such a counterexample would be possible in the original setting.

**Question 17.** How irregular a convex caustic can be compared to a regular boundary curve $\Gamma$?

**Question 18.** Let $\Gamma$ be a billiard table different from a circle and having a convex caustic $\gamma$. For every point $P \in \Gamma$, denote by $P_-$, and $P_+$ the tangency points of the caustic $\gamma$ with tangent lines to $\gamma$ passing through $P$. Is it possible that the length of the arc of $\gamma$ between $P_-$ and $P_+$ does not depend on $P$?

## Acknowledgements.

The authors are thankful to the anonymous referee for the useful remarks which greatly improved this manuscript. Bialy is also thankful to the participants of the course "Billiards" given in Tel Aviv University for very useful discussions and ideas. He was supported by ISF 162/15.


## References


[Arnold 1990] V. I. Arnold, *Singularities of caustics and wave fronts*, Math. Appl. **62**, Kluwer, Dordrecht, 1990. MR Zbl

[Avila et al. 2016] A. Avila, J. De Simoi, and V. Kaloshin, "An integrable deformation of an ellipse of small eccentricity is an ellipse", *Ann. Math.* (2) **184**:2 (2016), 527–558. MR Zbl

[Bialy 1993] M. Bialy, "Convex billiards and a theorem by E. Hopf", *Math. Z.* **214**:1 (1993), 147–154. MR Zbl

[Birkhoff 1917] G. D. Birkhoff, "Dynamical systems with two degrees of freedom", *Trans. Amer. Math. Soc.* **18**:2 (1917), 199–300. MR Zbl

[Douady 1982] R. Douady, *Applications du théoréme des tores invariants*, thèse de 3ème cycle, Paris Diderot Univ., 1982.

[Fetter 2012] H. L. Fetter, "Numerical exploration of a hexagonal string billiard", *Phys. D* **241**:8 (2012), 830–846. MR Zbl

[Hubacher 1987] A. Hubacher, "Instability of the boundary in the billiard ball problem", *Comm. Math. Phys.* **108**:3 (1987), 483–488. MR Zbl

[Innami 2002] N. Innami, "Geometry of geodesics for convex billiards and circular billiards", *Nihonkai Math. J.* **13**:1 (2002), 73–120. MR Zbl

[Kaloshin and Sorrentino 2016] V. Kaloshin and A. Sorrentino, "On local Birkhoff conjecture for convex billiards", preprint, 2016. arXiv

[Katok and Hasselblatt 1995] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Math. Appl. **54**, Cambridge Univ. Press, 1995. MR Zbl

[Katok et al. 1986] A. Katok, J.-M. Strelcyn, F. Ledrappier, and F. Przytycki, *Invariant manifolds, entropy and billiards: smooth maps with singularities*, Lecture Notes in Math. **1222**, Springer, 1986. MR Zbl

[Knill 1998] O. Knill, "On nonconvex caustics of convex billiards", *Elem. Math.* **53**:3 (1998), 89–106. MR Zbl

[Kozlov and Treshchëv 1991] V. V. Kozlov and D. V. Treshchëv, *Billiards: a genetic introduction to the dynamics of systems with impacts*, Translations of Mathematical Monographs **89**, Amer. Math. Soc., Providence, RI, 1991. MR Zbl

[Lazutkin 1973] V. F. Lazutkin, "Existence of caustics for the billiard problem in a convex domain", *Izv. Akad. Nauk SSSR Ser. Mat.* **37** (1973), 186–216. In Russian; translated in *Math. USSR Izv.* **7**:1 (1973), 185–214. MR Zbl

[Lazutkin 1981] V. F. Lazutkin, Выпуклый биллиард и собственные функции оператора Лапласа, Leningrad. Univ., 1981. MR Zbl

[Mather and Forni 1994] J. N. Mather and G. Forni, "Action minimizing orbits in Hamiltonian systems", pp. 92–186 in *Transition to chaos in classical and quantum mechanics* (Montecatini Terme, Italy, 1991), edited by S. Graffi, Lecture Notes in Math. **1589**, Springer, 1994. MR Zbl

[Stoll 1930] A. Stoll, "Über den Kappenkörper eines konvexen Körpers", *Comment. Math. Helv.* **2**:1 (1930), 35–68. MR Zbl

[Tabachnikov 2005] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library **30**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Turner 1982] P. H. Turner, "Convex caustics for billiards in $\mathbb{R}^2$ and $\mathbb{R}^3$", pp. 85–106 in *Convexity and related combinatorial geometry* (Norman, OK, 1980), edited by D. C. Kay and M. Breen, Lecture Notes in Pure and Appl. Math. **76**, Dekker, New York, 1982. MR Zbl

Received August 14, 2017. Revised December 29, 2017.

MAXIM ARNOLD
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
RICHARDSON, TX
UNITED STATES
maxim.arnold@utdallas.edu

MISHA BIALY
TEL AVIV UNIVERSITY
TEL AVIV
ISRAEL
bialy@post.tau.ac.il

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.271

# CERTAIN CHARACTER SUMS AND HYPERGEOMETRIC SERIES

RUPAM BARMAN AND NEELAM SAIKIA

**We prove two transformations for the $p$-adic hypergeometric series which can be described as $p$-adic analogues of a Kummer's linear transformation and a transformation of Clausen. We first evaluate two character sums, and then relate them to the $p$-adic hypergeometric series to deduce the transformations. We also find another transformation for the $p$-adic hypergeometric series from which many special values of the $p$-adic hypergeometric series as well as finite field hypergeometric functions are obtained.**

## 1. Introduction and statement of results

For a complex number $a$, the rising factorial or the Pochhammer symbol is defined as $(a)_0 = 1$ and $(a)_k = a(a+1)\cdots(a+k-1)$, $k \geq 1$. For a nonnegative integer $r$, and $a_i, b_i \in \mathbb{C}$ with $b_i \notin \{\dots, -3, -2, -1\}$, the classical hypergeometric series ${}_{r+1}F_r$ is defined by

$$ {}_{r+1}F_r\left(\begin{matrix} a_1, & a_2, & \dots, & a_{r+1} \\ & b_1, & \dots, & b_r \end{matrix} \;\middle|\; \lambda\right) := \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{r+1})_k}{(b_1)_k \cdots (b_r)_k} \cdot \frac{\lambda^k}{k!}, $$

which converges for $|\lambda| < 1$. Throughout the paper, $p$ denotes an odd prime and $\mathbb{F}_q$ denotes the finite field with $q$ elements, where $q = p^r$, $r \geq 1$. Greene [1987] introduced the notion of hypergeometric functions over finite fields analogous to the classical hypergeometric series. Finite field hypergeometric series were developed mainly to simplify character sum evaluations. Let $\widehat{\mathbb{F}_q^\times}$ be the group of all multiplicative characters on $\mathbb{F}_q^\times$. We extend the domain of each $\chi \in \widehat{\mathbb{F}_q^\times}$ to $\mathbb{F}_q$ by setting $\chi(0) = 0$ including the trivial character $\varepsilon$. For multiplicative characters $A$ and $B$ on $\mathbb{F}_q$, the binomial coefficient $\binom{A}{B}$ is defined by

$$ \binom{A}{B} := \frac{B(-1)}{q} J(A, \bar{B}) = \frac{B(-1)}{q} \sum_{x \in \mathbb{F}_q} A(x)\bar{B}(1-x), \tag{1-1} $$

*MSC2010:* 11S80, 11T24, 33E50, 33C99.
*Keywords:* character sum, hypergeometric series, $p$-adic gamma function.

where $J(A, B)$ denotes the usual Jacobi sum and $\bar{B}$ is the character inverse of $B$. Let $n$ be a positive integer. For characters $A_0, A_1, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ on $\mathbb{F}_q$, Greene defined the ${}_{n+1}F_n$ finite field hypergeometric functions over $\mathbb{F}_q$ by

$$ {}_{n+1}F_n\left(\begin{matrix} A_0, & A_1, & \ldots, & A_n \\ & B_1, & \ldots, & B_n \end{matrix} \,\middle|\, x\right)_q = \frac{q}{q-1} \sum_{\chi \in \widehat{\mathbb{F}_q^\times}} \binom{A_0\chi}{\chi}\binom{A_1\chi}{B_1\chi} \cdots \binom{A_n\chi}{B_n\chi} \chi(x). $$

Some of the biggest motivations for studying finite field hypergeometric functions have been their connections with Fourier coefficients and eigenvalues of modular forms and with counting points on certain kinds of algebraic varieties. Their links to Fourier coefficients and eigenvalues of modular forms are established by many authors, for example, see [Ahlgren and Ono 2000; Evans 2010; Frechette et al. 2004; Fuselier 2010; Fuselier and McCarthy 2016; Lennon 2011b; McCarthy 2012b; Mortenson 2005]. Very recently, McCarthy and Papanikolas [2015] linked the finite field hypergeometric functions to Siegel modular forms. It is well known that finite field hypergeometric functions can be used to count points on varieties over finite fields. For example, see [Barman and Kalita 2013a; 2013b; Fuselier 2010; Koike 1992; Lennon 2011a; Ono 1998; Salerno 2013; Vega 2011].

Since the multiplicative characters on $\mathbb{F}_q$ form a cyclic group of order $q-1$, a condition like $q \equiv 1 \pmod{\ell}$ must be satisfied where $\ell$ is the least common multiple of the orders of the characters appearing in the hypergeometric function. Consequently, many results involving these functions are restricted to primes in certain congruence classes. To overcome these restrictions, McCarthy [2012a; 2013] defined a function ${}_nG_n[\cdots]_q$ in terms of quotients of the $p$-adic gamma function which can best be described as an analogue of hypergeometric series in the $p$-adic setting (defined in Section 2).

Many transformations exist for finite field hypergeometric functions which are analogues of certain classical results [Greene 1987; McCarthy 2012c]. Results involving finite field hypergeometric functions can readily be converted to expressions involving ${}_nG_n[\cdots]$. However these new expressions in ${}_nG_n[\cdots]$ will be valid for the same set of primes for which the original expressions involving finite field hypergeometric functions existed. It is a nontrivial exercise to then extend these results to almost all primes. There are very few identities and transformations for the $p$-adic hypergeometric series ${}_nG_n[\cdots]_q$ which exist for all but finitely many primes (see for example [Barman and Saikia 2014; 2015; Barman et al. 2015]. Recently, Fuselier and McCarthy [2016] proved certain transformations for ${}_nG_n[\cdots]_q$, and used them to establish a supercongruence conjecture of Rodriguez-Villegas between a truncated ${}_4F_3$ hypergeometric series and the Fourier coefficients of a certain weight four modular form.

Let $\chi_4$ be a character of order 4. Then a finite field analogue of ${}_2F_1\left(\begin{matrix} 1/4, & 3/4 \\ & 1 \end{matrix} \,\middle|\, x\right)$ is the function ${}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \,\middle|\, x\right)$. Using the relation between finite field hypergeometric functions and ${}_nG_n$-functions as given in Proposition 3.5 in Section 3, the function ${}_2G_2\left[\begin{smallmatrix}1/4, & 3/4\\ 0, & 0\end{smallmatrix} \mid \frac{1}{x}\right]_q$ can be described as a $p$-adic analogue of the classical hypergeometric series ${}_2F_1\left(\begin{smallmatrix}1/4, & 3/4\\ & 1\end{smallmatrix} \mid x\right)$. In this article, we prove the following transformation for the $p$-adic hypergeometric series which can be described as a $p$-adic analogue of the Kummer's linear transformation [Bailey 1935, p. 4, Equation (1)]. Let $\varphi$ be the quadratic character on $\mathbb{F}_q$.

**Theorem 1.1.** *Let $p$ be an odd prime and $x \in \mathbb{F}_q$. Then, for $x \neq 0, 1$, we have*

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix} \;\middle|\; \frac{1}{x}\right]_q = \varphi(-2)\,{}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix} \;\middle|\; \frac{1}{1-x}\right]_q. $$

We note that the finite field analogue of Kummer's linear transformation was discussed by Greene [1984, p. 109, Equation (7.7)] when $q \equiv 1 \pmod 4$.

We have $\varphi(-2) = -1$ if and only if $p \equiv 5, 7 \pmod 8$. Hence, using Theorem 1.1 for $x = \frac{1}{2}$, we obtain the following special value of the ${}_2G_2$-function.

**Corollary 1.2.** *Let $p$ be a prime such that $p \equiv 5, 7 \pmod 8$. Then we have*

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix} \;\middle|\; 2\right]_p = 0. \tag{1-2} $$

If we convert the ${}_2G_2$-function given in (1-2) using Proposition 3.5 in Section 3, then we have ${}_2F_1\left(\begin{smallmatrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{smallmatrix} \mid \frac{1}{2}\right)_p = 0$ for $p \equiv 5 \pmod 8$ which also follows from [Greene 1987, Equation (4.15)]. The value of ${}_2G_2\left[\begin{smallmatrix}1/4, & 3/4\\ 0, & 0\end{smallmatrix} \mid 2\right]_p$ can be deduced from [Greene 1987, Equation (4.15)] when $p \equiv 1 \pmod 8$. It would be interesting to know the value of ${}_2G_2\left[\begin{smallmatrix}1/4, & 3/4\\ 0, & 0\end{smallmatrix} \mid 2\right]_p$ when $p \equiv 3 \pmod 8$.

The following transformation for classical hypergeometric series is a special case of Clausen's famous classical identity [Bailey 1935, p. 86, Equation (4)]:

$$ {}_3F_2\left(\begin{matrix}\frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}\\ & 1, & 1\end{matrix} \;\middle|\; x\right) = (1-x)^{-1/2}\,{}_2F_1\left(\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ & 1\end{matrix} \;\middle|\; \frac{x}{x-1}\right)^2. \tag{1-3} $$

A finite field analogue of (1-3) was studied by Greene [1984, p. 94, Proposition 6.14]. Evans and Greene [2009a] gave a finite field analogue of the Clausen's classical identity. We prove the following transformation for the ${}_nG_n$-function which can be described as a $p$-adic analogue of (1-3). Let $\delta$ be the function defined on $\mathbb{F}_q$ by

$$ \delta(x) = \begin{cases} 1 & \text{if } x = 0;\\ 0 & \text{if } x \neq 0.\end{cases} $$

**Theorem 1.3.** *Let $p$ be an odd prime and $x \in \mathbb{F}_p$. Then, for $x \neq 0, 1$, we have*

$$ {}_3G_3\left[\begin{matrix}\frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}\\ 0, & 0, & 0\end{matrix} \;\middle|\; \frac{1}{x}\right]_p = \varphi(1-x)\cdot{}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix} \;\middle|\; \frac{x-1}{x}\right]_p^2 - p\cdot\varphi(1-x). $$

We also prove the following transformation using Theorem 1.1 and [Greene 1987, Theorem 4.16].

**Theorem 1.4.** *Let $p$ be an odd prime and $x \in \mathbb{F}_q$. Then, for $x \neq 0, \pm 1$, we have*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \,\Big|\, \frac{(1+x)^2}{(1-x)^2}\right]_q = \varphi(-2)\varphi(1+x)\,{}_2G_2\left[\begin{matrix} \frac{1}{2}, & \frac{1}{2} \\ 0, & 0 \end{matrix} \,\Big|\, x^{-1}\right]_q. \tag{1-4}$$

The following transformation is a finite field analogue of (1-4).

**Theorem 1.5.** *Let $p$ be an odd prime and $q = p^r$ for some $r \geq 1$ such that $q \equiv 1 \pmod 4$. Then, for $x \neq 0, \pm 1$, we have*

$$ {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \,\Big|\, \frac{(1-x)^2}{(1+x)^2}\right)_q = \varphi(-2)\varphi(1+x)\,{}_2F_1\left(\begin{matrix} \varphi, & \varphi \\ & \varepsilon \end{matrix} \,\Big|\, x\right)_q.$$

Using Theorems 1.4 and 1.5, one can deduce many special values of the $p$-adic hypergeometric series as well as the finite field hypergeometric functions. For example, we have the following special values of a ${}_2G_2$-function and its finite field analogue.

**Theorem 1.6.** *For any odd prime $p$, we have*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \,\Big|\, 9\right]_p = \begin{cases} 0 & \text{if } p \equiv 3 \pmod 4; \\ -2x\varphi(6)(-1)^{\frac{x+y+1}{2}} & \text{if } p \equiv 1 \pmod 4,\ x^2+y^2 = p, \text{ and } x \text{ odd.} \end{cases}$$

*For $p \equiv 1 \pmod 4$, we have*

$$ {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \,\Big|\, \frac{1}{9}\right)_p = \frac{2x\varphi(6)(-1)^{\frac{x+y+1}{2}}}{p},$$

*where $x^2 + y^2 = p$ and $x$ is odd.*

We also find special values of the following ${}_2G_2$-function.

**Theorem 1.7.** *For $q \equiv 1 \pmod 8$ we have*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \,\Big|\, \left(\frac{6\sqrt{2}\pm 3}{-2\sqrt{2}\pm 3}\right)^2\right]_q = -q\varphi(6\pm 12\sqrt{2})\left\{\binom{\chi_4}{\varphi} + \binom{\chi_4^3}{\varphi}\right\}. \tag{1-5}$$

*For $q \equiv 11 \pmod{12}$ we have*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \,\Big|\, \left(\frac{6\pm\sqrt{3}}{-2\pm\sqrt{3}}\right)^2\right]_q = 0. \tag{1-6}$$

*For* $q \equiv 1 \pmod{12}$ *we have*

$$(1\text{-}7)\qquad {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix}\;\middle|\;\left(\frac{6\pm\sqrt{3}}{-2\pm\sqrt{3}}\right)^2\right]_q = -q\varphi\left(\frac{8\pm5\sqrt{3}}{12\pm6\sqrt{3}}\right)\left\{\binom{\varphi}{\chi_3}+\binom{\varphi}{\chi_3^2}\right\}.$$

The following theorem is a finite field analogue of Theorem 1.7.

**Theorem 1.8.** *For* $q \equiv 1 \pmod 8$ *we have*

$$(1\text{-}8)\qquad {}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\middle|\;\left(\frac{-2\sqrt{2}\pm3}{6\sqrt{2}\pm3}\right)^2\right)_q = \varphi(6\pm12\sqrt{2})\left\{\binom{\chi_4}{\varphi}+\binom{\chi_4^3}{\varphi}\right\}.$$

*For* $q \equiv 1 \pmod{12}$ *we have*

$$(1\text{-}9)\qquad {}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\middle|\;\left(\frac{-2\pm\sqrt{3}}{6\pm\sqrt{3}}\right)^2\right)_q = \varphi\left(\frac{8\pm5\sqrt{3}}{12\pm6\sqrt{3}}\right)\left\{\binom{\varphi}{\chi_3}+\binom{\varphi}{\chi_3^2}\right\}.$$

In Section 3 we prove two character sum identities and then use them to prove Theorems 1.1, 1.3, and 1.4. We also prove Theorem 1.5 in Section 3. In Section 4 we prove Theorems 1.6, 1.7 and 1.8.

## 2. Notations and preliminaries

Let $\mathbb{Z}_p$ and $\mathbb{Q}_p$ denote the ring of $p$-adic integers and the field of $p$-adic numbers, respectively. Let $\overline{\mathbb{Q}}_p$ be the algebraic closure of $\mathbb{Q}_p$ and $\mathbb{C}_p$ the completion of $\overline{\mathbb{Q}}_p$. Let $\mathbb{Z}_q$ be the ring of integers in the unique unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$. We know that $\chi \in \widehat{\mathbb{F}_q^\times}$ takes values in $\mu_{q-1}$, where $\mu_{q-1}$ is the group of $(q-1)$-th roots of unity in $\mathbb{C}^\times$. Since $\mathbb{Z}_q^\times$ contains all $(q-1)$-th roots of unity, we can consider multiplicative characters on $\mathbb{F}_q^\times$ to be maps $\chi : \mathbb{F}_q^\times \to \mathbb{Z}_q^\times$. Let $\omega : \mathbb{F}_q^\times \to \mathbb{Z}_q^\times$ be the Teichmüller character. For $a \in \mathbb{F}_q^\times$, the value $\omega(a)$ is just the $(q-1)$-th root of unity in $\mathbb{Z}_q$ such that $\omega(a) \equiv a \pmod p$.

We now introduce some properties of Gauss sums. For further details, see [Berndt et al. 1998]. Let $\zeta_p$ be a fixed primitive $p$-th root of unity in $\overline{\mathbb{Q}}_p$. The trace map $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ is given by

$$\mathrm{tr}(\alpha) = \alpha + \alpha^p + \alpha^{p^2} + \cdots + \alpha^{p^{r-1}}.$$

For $\chi \in \widehat{\mathbb{F}_q^\times}$, the *Gauss sum* is defined by

$$g(\chi) := \sum_{x\in\mathbb{F}_q} \chi(x)\zeta_p^{\mathrm{tr}(x)}.$$

Now, we will see some elementary properties of Gauss and Jacobi sums. We let $T$ denote a fixed generator of $\widehat{\mathbb{F}_q^\times}$.

**Lemma 2.1** [Greene 1987, Equation 1.12]. *If* $k \in \mathbb{Z}$ *and* $T^k \neq \varepsilon$, *then*

$$g(T^k)g(T^{-k}) = qT^k(-1).$$

Let $\delta$ denote the function on multiplicative characters defined by

$$\delta(A)=\begin{cases}1 & \text{if } A \text{ is the trivial character;}\\ 0 & \text{otherwise.}\end{cases}$$

**Lemma 2.2** [Greene 1987, Equation 1.14]. *For $A, B \in \widehat{\mathbb{F}_q^\times}$ we have*

$$J(A,B)=\frac{g(A)g(B)}{g(AB)}+(q-1)B(-1)\delta(AB).$$

The following are character sum analogues of the binomial theorem [Greene 1987]. For any $A \in \widehat{\mathbb{F}_q^\times}$ and $x \in \mathbb{F}_q$ we have

$$\overline{A}(1-x)=\delta(x)+\frac{q}{q-1}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{A\chi}{\chi}\chi(x), \tag{2-1}$$

$$A(1+x)=\delta(x)+\frac{q}{q-1}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{A}{\chi}\chi(x). \tag{2-2}$$

We recall some properties of the binomial coefficients from [Greene 1987]:

$$\binom{A}{B}=\binom{A}{A\overline{B}}, \tag{2-3}$$

$$\binom{A}{\varepsilon}=\binom{A}{A}=\frac{-1}{q}+\frac{q-1}{q}\delta(A). \tag{2-4}$$

**Theorem 2.3** [Berndt et al. 1998, Davenport–Hasse relation]. *Let $m$ be a positive integer and let $q=p^r$ be a prime power such that $q\equiv 1 \pmod{m}$. For multiplicative characters $\chi$ and $\psi$ in $\widehat{\mathbb{F}_q^\times}$, we have*

$$\prod_{\chi^m=\varepsilon} g(\chi\psi)=-g(\psi^m)\psi(m^{-m})\prod_{\chi^m=\varepsilon} g(\chi).$$

Now, we recall the $p$-adic gamma function. For further details, see [Koblitz 1980]. For a positive integer $n$, the $p$-adic gamma function $\Gamma_p(n)$ is defined as

$$\Gamma_p(n):=(-1)^n\prod_{0<j<n,\, p\nmid j} j$$

and one extends it to all $x\in\mathbb{Z}_p$ by setting $\Gamma_p(0):=1$ and

$$\Gamma_p(x):=\lim_{x_n\to x}\Gamma_p(x_n)$$

for $x\neq 0$, where $x_n$ runs through any sequence of positive integers $p$-adically approaching $x$. This limit exists, is independent of how $x_n$ approaches $x$, and determines a continuous function on $\mathbb{Z}_p$ with values in $\mathbb{Z}_p^\times$. For $x\in\mathbb{Q}$ we let $\lfloor x\rfloor$ denote the greatest integer less than or equal to $x$ and $\langle x\rangle$ denote the fractional part of $x$, i.e., $x - \lfloor x \rfloor$, satisfying $0 \le \langle x \rangle < 1$. We now recall the McCarthy's $p$-adic hypergeometric series ${}_nG_n[\cdots]$ as follows.

**Definition 2.4** [McCarthy 2013, Definition 5.1]. Let $p$ be an odd prime and $q = p^r$, $r \ge 1$. Let $t \in \mathbb{F}_q$. For a positive integer $n$ and $1 \le k \le n$, let $a_k,\ b_k \in \mathbb{Q} \cap \mathbb{Z}_p$. Then the function ${}_nG_n[\cdots]$ is defined by

$$
{}_nG_n\left[\begin{matrix} a_1, & a_2, & \dots, & a_n \\ b_1, & b_2, & \dots, & b_n \end{matrix} \,\middle|\, t\right]_q :=
\frac{-1}{q-1}\sum_{a=0}^{q-2}(-1)^{an}\,\overline{\omega}^a(t) \times \prod_{k=1}^{n}\prod_{i=0}^{r-1}(-p)^{-\lfloor\langle a_k p^i\rangle - \frac{ap^i}{q-1}\rfloor - \lfloor\langle -b_k p^i\rangle + \frac{ap^i}{q-1}\rfloor}
\times \frac{\Gamma_p(\langle(a_k - \frac{a}{q-1})p^i\rangle)}{\Gamma_p(\langle a_k p^i\rangle)} \cdot \frac{\Gamma_p(\langle(-b_k + \frac{a}{q-1})p^i\rangle)}{\Gamma_p(\langle -b_k p^i\rangle)}.
$$

Let $\pi \in \mathbb{C}_p$ be the fixed root of $x^{p-1} + p = 0$ which satisfies

$$\pi \equiv \zeta_p - 1 \pmod{(\zeta_p - 1)^2}.$$

Then the Gross–Koblitz formula relates Gauss sums and the $p$-adic gamma function as follows.

**Theorem 2.5** [Gross and Koblitz 1979]. *For $a \in \mathbb{Z}$ and $q = p^r$,*

$$g(\overline{\omega}^a) = -\pi^{(p-1)\sum_{i=0}^{r-1}\langle\frac{ap^i}{q-1}\rangle}\prod_{i=0}^{r-1}\Gamma_p\Big(\langle\frac{ap^i}{q-1}\rangle\Big).$$

The following lemma relates products of values of $p$-adic gamma function.

**Lemma 2.6** [Barman and Saikia 2014, Lemma 3.1]. *Let $p$ be a prime and $q = p^r$. For $0 \le a \le q-2$ and $t \ge 1$ with $p \nmid t$, we have*

$$\omega(t^{-ta})\prod_{i=0}^{r-1}\Gamma_p\Big(\langle\frac{-tp^i a}{q-1}\rangle\Big)\prod_{h=1}^{t-1}\Gamma_p\Big(\langle\frac{hp^i}{t}\rangle\Big) = \prod_{i=0}^{r-1}\prod_{h=0}^{t-1}\Gamma_p\Big(\langle\frac{p^i(1+h)}{t} - \frac{p^i a}{q-1}\rangle\Big).$$

We now prove a lemma that will be used to prove our results.

**Lemma 2.7.** *Let $p$ be an odd prime and $q = p^r$. Then for $0 \le a \le q-2$ and $0 \le i \le r-1$ we have*

$$-\left\lfloor\frac{-4ap^i}{q-1}\right\rfloor + \left\lfloor\frac{-2ap^i}{q-1}\right\rfloor = -\left\lfloor\left\langle\frac{p^i}{4}\right\rangle - \frac{ap^i}{q-1}\right\rfloor - \left\lfloor\left\langle\frac{3p^i}{4}\right\rangle - \frac{ap^i}{q-1}\right\rfloor. \tag{2-5}$$

*Proof.* Let

$$\left\lfloor \frac{-4ap^i}{q-1} \right\rfloor = 4k+s,$$

where $k, s \in \mathbb{Z}$ satisfy $0 \le s \le 3$. Then

$$4k+s \le \frac{-4ap^i}{q-1} < 4k+s+1. \tag{2-6}$$

If $p^i \equiv 1 \pmod 4$, then (2-6) yields

$$\left\lfloor \frac{-2ap^i}{q-1} \right\rfloor = \begin{cases} 2k & \text{if } s = 0, 1; \\ 2k+1 & \text{if } s = 2, 3, \end{cases} \tag{2-7}$$

$$\left\lfloor \left\langle \frac{p^i}{4} \right\rangle - \frac{ap^i}{q-1} \right\rfloor = \begin{cases} k & \text{if } s = 0, 1, 2; \\ k+1 & \text{if } s = 3, \end{cases} \tag{2-8}$$

$$\left\lfloor \left\langle \frac{3p^i}{4} \right\rangle - \frac{ap^i}{q-1} \right\rfloor = \begin{cases} k & \text{if } s = 0; \\ k+1 & \text{if } s = 1, 2, 3. \end{cases} \tag{2-9}$$

Putting the above values for different values of $s$ we readily obtain (2-5). The proof of (2-5) is similar when $p^i \equiv 3 \pmod 4$. □

## 3. Proofs of the main results

We first prove two propositions which enable us to express certain character sums in terms of the $p$-adic hypergeometric series.

**Proposition 3.1.** *Let $p$ be an odd prime and $x \in \mathbb{F}_q^\times$. Then we have*

$$\begin{aligned}\sum_{y \in \mathbb{F}_q} \varphi(y)\varphi(1-2y+xy^2) &= \varphi(2x) + \frac{q^2\varphi(-2)}{q-1} \sum_{\chi \in \widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\left(\frac{x}{4}\right) \\ &= -\varphi(-2)\,{}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \,\middle|\, \frac{1}{x}\right]_q.\end{aligned}$$

*Proof.* Applying (2-3) and then (1-1) we have

$$\begin{aligned}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{4}\Big)&=\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{4}\Big)\binom{\varphi\chi^2}{\varphi\chi}\\&=\frac{\varphi(-1)}{q}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi}{\chi}\chi\Big(\frac{-x}{4}\Big)J(\varphi\chi^2,\varphi\overline{\chi})\\&=\frac{\varphi(-1)}{q}\sum_{\substack{\chi\in\widehat{\mathbb{F}_q^\times}\\ y\in\mathbb{F}_q}}\binom{\varphi\chi}{\chi}\chi\Big(\frac{-x}{4}\Big)\varphi\chi^2(y)\varphi\overline{\chi}(1-y)\\&=\frac{\varphi(-1)}{q}\sum_{\substack{\chi\in\widehat{\mathbb{F}_q^\times}\\ y\in\mathbb{F}_q,y\neq1}}\varphi(y)\varphi(1-y)\binom{\varphi\chi}{\chi}\chi\Big(-\frac{xy^2}{4(1-y)}\Big).\end{aligned}$$

Now, (2-1) yields

$$\begin{aligned}&\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{4}\Big)\\&\quad=\frac{\varphi(-1)(q-1)}{q^2}\sum_{y\in\mathbb{F}_q,y\neq1}\varphi(y)\varphi(1-y)\bigg(\varphi\Big(1+\frac{xy^2}{4(1-y)}\Big)-\delta\Big(\frac{xy^2}{4(1-y)}\Big)\bigg)\\&\quad=\frac{(q-1)\varphi(-1)}{q^2}\sum_{y\in\mathbb{F}_q,y\neq1}\varphi(y)\varphi(1-y)\varphi\Big(1+\frac{xy^2}{4(1-y)}\Big).\end{aligned}$$

Since $p$ is an odd prime, taking the transformation $y\mapsto 2y$ we get

$$\begin{aligned}&\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{4}\Big)\\&\qquad=\frac{(q-1)\varphi(-2)}{q^2}\sum_{\substack{y\in\mathbb{F}_q\\ y\neq\frac12}}\varphi(y)\varphi(1-2y)\varphi\Big(1+\frac{xy^2}{1-2y}\Big)\\&\qquad=\frac{(q-1)\varphi(-2)}{q^2}\sum_{\substack{y\in\mathbb{F}_q\\ y\neq\frac12}}\varphi(y)\varphi(1-2y+xy^2)\\&\qquad=\frac{(q-1)\varphi(-2)}{q^2}\sum_{y\in\mathbb{F}_q}\varphi(y)\varphi(1-2y+xy^2)-\frac{\varphi(-x)(q-1)}{q^2},\end{aligned}$$

from which we readily obtain the first identity of the proposition.

To complete the proof of the proposition, we relate the above character sums to the $p$-adic hypergeometric series. From (1-1), Lemma 2.2, and then using the facts that $\delta(\chi)=0$ for $\chi\neq\varepsilon$, $\delta(\varepsilon)=1$ and $g(\varepsilon)=-1$, we deduce that

$$\begin{aligned}
A :=& \sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\left(\frac{x}{4}\right)\\
=&\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} J(\varphi\chi^2,\overline{\chi})J(\varphi\chi,\overline{\chi})\chi\left(\frac{x}{4}\right)\\
=&\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \frac{g(\varphi\chi^2)g^2(\overline{\chi})}{g(\varphi)}\chi\left(\frac{x}{4}\right)+\frac{q-1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\frac{g(\varphi\chi)g(\overline{\chi})}{g(\varphi)}\chi\left(-\frac{x}{4}\right)\delta(\varphi\chi)\\
=&\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \frac{g(\varphi\chi^2)g^2(\overline{\chi})}{g(\varphi)}\chi\left(\frac{x}{4}\right)-\frac{q-1}{q^2}\varphi(-x).
\end{aligned}$$

Now, taking $\chi=\omega^a$ we have

$$A=\frac{1}{q^2}\sum_{a=0}^{q-2}\frac{g(\varphi\omega^{2a})g^2(\overline{\omega}^a)}{g(\varphi)}\omega^a\left(\frac{x}{4}\right)-\frac{q-1}{q^2}\varphi(-x).$$

Using the Davenport–Hasse relation for $m=2$ and $\psi=\omega^{2a}$ we obtain

$$g(\varphi\omega^{2a})=\frac{g(\omega^{4a})\overline{\omega}^{2a}(4)g(\varphi)}{g(\omega^{2a})}.$$

Thus,

$$A=\frac{1}{q^2}\sum_{a=0}^{q-2}\omega^a(x)\overline{\omega}^{3a}(4)\frac{g(\omega^{4a})g^2(\overline{\omega}^a)}{g(\omega^{2a})}-\frac{q-1}{q^2}\varphi(-x).$$

Applying the Gross–Koblitz formula we deduce that

$$A=\frac{1}{q^2}\sum_{a=0}^{q-2}\omega^a(x)\overline{\omega}^{3a}(4)\pi^{(p-1)\alpha}\prod_{i=0}^{r-1}\frac{\Gamma_p\left(\left\langle\frac{-4ap^i}{q-1}\right\rangle\right)\Gamma_p^2\left(\left\langle\frac{ap^i}{q-1}\right\rangle\right)}{\Gamma_p\left(\left\langle\frac{-2ap^i}{q-1}\right\rangle\right)}-\frac{q-1}{q^2}\varphi(-x),$$

where

$$\alpha=\sum_{i=0}^{r-1}\left\{\left\langle\frac{-4ap^i}{q-1}\right\rangle+2\left\langle\frac{ap^i}{q-1}\right\rangle-\left\langle\frac{-2ap^i}{q-1}\right\rangle\right\}.$$

Using Lemma 2.6 for $t = 4$ and $t = 2$, we deduce that

$$A = \frac{1}{q^2}\sum_{a=0}^{q-2}\omega^a(x)\pi^{(p-1)\alpha}\prod_{i=0}^{r-1}\frac{\Gamma_p\left(\left\langle\left(\frac{1}{4}-\frac{a}{q-1}\right)p^i\right\rangle\right)\Gamma_p\left(\left\langle\left(\frac{3}{4}-\frac{a}{q-1}\right)p^i\right\rangle\right)\Gamma_p^2\left(\left\langle\frac{ap^i}{q-1}\right\rangle\right)}{\Gamma_p\left(\left\langle\frac{p^i}{4}\right\rangle\right)\Gamma_p\left(\left\langle\frac{3p^i}{4}\right\rangle\right)} - \frac{q-1}{q^2}\varphi(-x).$$

Finally, using Lemma 2.7 we have

$$A = -\frac{q-1}{q^2}\cdot {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix}\;\middle|\;\frac{1}{x}\right]_q - \frac{q-1}{q^2}\varphi(-x). \qquad \square$$

**Proposition 3.2.** *Let $p$ be an odd prime and $x \in \mathbb{F}_q$. Then, for $x \neq 1$, we have*

$$\sum_{y\in\mathbb{F}_q}\varphi(y)\varphi(1-2y+xy^2) = 2\varphi(x-1) + \frac{q^2}{q-1}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi^2}\chi(x-1)$$
$$= -{}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix}\;\middle|\;\frac{1}{1-x}\right]_q.$$

*Proof.* From (1-1) and then using Lemma 2.2, we have

$$\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi^2} = \frac{\chi(-1)}{q^2}J(\varphi\chi^2,\overline{\chi})J(\varphi\chi,\overline{\chi}^2)$$
$$= \frac{\chi(-1)}{q^2}\left[\frac{g(\varphi\chi^2)g(\overline{\chi})}{g(\varphi\chi)} + (q-1)\chi(-1)\delta(\varphi\chi)\right] \times\left[\frac{g(\varphi\chi)g(\overline{\chi}^2)}{g(\varphi\overline{\chi})} + (q-1)\delta(\varphi\overline{\chi})\right]. \tag{3-1}$$

From Lemma 2.1, we have $g(\varphi)^2 = q\varphi(-1)$. Since $\delta(\chi) = 0$ for $\chi \neq \varepsilon$, $\delta(\varepsilon) = 1$ and $g(\varepsilon) = -1$, (3-1) yields

$$B := \sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi^2}\chi(x-1)$$
$$= \frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\frac{g(\varphi\chi^2)g(\overline{\chi})g(\overline{\chi}^2)}{g(\varphi\overline{\chi})}\chi(1-x) - 2\frac{q-1}{q^2}\varphi(x-1). \tag{3-2}$$

Using Lemma 2.2 and then (1-1) we obtain

$$\frac{g(\varphi\chi^2)g(\overline{\chi}^2)}{g(\varphi)} = q\binom{\varphi\chi^2}{\chi^2}, \tag{3-3}$$

and

$$\frac{g(\varphi)g(\overline{\chi})}{g(\varphi\overline{\chi})} = q\chi(-1)\binom{\varphi}{\chi} - (q-1)\chi(-1)\delta(\varphi\overline{\chi}). \tag{3-4}$$

From (2-4), we have $\binom{\varphi}{\varepsilon} = -\frac{1}{q}$. Hence, (3-3) and (3-4) yield

$$\begin{aligned}\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} &\frac{g(\varphi\chi^2)g(\overline{\chi})g(\overline{\chi}^2)}{g(\varphi\overline{\chi})}\chi(1-x) \\ &= \sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi^2}\binom{\varphi}{\chi}\chi(x-1) - \frac{q-1}{q}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\chi(x-1)\binom{\varphi\chi^2}{\chi^2}\delta(\varphi\overline{\chi}) \\ &= \sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi^2}\binom{\varphi}{\chi}\chi(x-1) - \frac{q-1}{q}\binom{\varphi}{\varepsilon}\varphi(x-1) \\ &= \sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi^2}\binom{\varphi}{\chi}\chi(x-1) + \frac{q-1}{q^2}\varphi(x-1).\end{aligned} \tag{3-5}$$

Applying (1-1) on the right-hand side of (3-5), and then (2-2) we have

$$\begin{aligned}\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} &\frac{g(\varphi\chi^2)g(\overline{\chi})g(\overline{\chi}^2)}{g(\varphi\overline{\chi})}\chi(1-x) \\ &= \frac{1}{q}\sum_{\substack{\chi\in\widehat{\mathbb{F}_q^\times}\\ y\in\mathbb{F}_q}} \binom{\varphi}{\chi}\chi(x-1)\varphi\chi^2(y)\overline{\chi}^2(1-y) + \frac{q-1}{q^2}\varphi(x-1) \\ &= \frac{1}{q}\sum_{\substack{\chi\in\widehat{\mathbb{F}_q^\times}\\ y\in\mathbb{F}_q,y\neq 1}} \varphi(y)\binom{\varphi}{\chi}\chi\left(\frac{(x-1)y^2}{(1-y)^2}\right) + \frac{q-1}{q^2}\varphi(x-1) \\ &= \frac{q-1}{q^2}\sum_{y\in\mathbb{F}_q,y\neq 1}\varphi(y)\left[\varphi\left(1+\frac{(x-1)y^2}{(1-y)^2}\right) - \delta\left(\frac{(x-1)y^2}{(1-y)^2}\right)\right] + \frac{q-1}{q^2}\varphi(x-1) \\ &= \frac{q-1}{q^2}\sum_{\substack{y\in\mathbb{F}_q\\ y\neq 1}}\varphi(y)\varphi(1-2y+xy^2) + \frac{q-1}{q^2}\varphi(x-1).\end{aligned}$$

Adding and subtracting the term under summation for $y = 1$, we have

$$\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \frac{g(\varphi\chi^2)g(\overline{\chi})g(\overline{\chi}^2)}{g(\varphi\overline{\chi})}\chi(1-x) = \frac{q-1}{q^2}\sum_{y\in\mathbb{F}_q}\varphi(y)\varphi(1-2y+xy^2). \tag{3-6}$$

Combining (3-2) and (3-6) we readily obtain the first equality of the proposition.

To complete the proof of the proposition, we relate the character sums given in (3-2) to the $p$-adic hypergeometric series. Using the Davenport–Hasse relation for $m=2,\ \psi=\chi^2$ and $m=2,\ \psi=\overline{\chi}$, we have

$$g(\varphi\chi^2)=\frac{g(\chi^4)g(\varphi)\overline{\chi}^2(4)}{g(\chi^2)}\quad\text{and}\quad g(\varphi\overline{\chi})=\frac{g(\overline{\chi}^2)g(\varphi)\chi(4)}{g(\overline{\chi})},$$

respectively. Plugging these two expressions into (3-2) we obtain

$$B=\frac{1}{q^2}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\frac{g(\chi^4)g^2(\overline{\chi})}{g(\chi^2)}\overline{\chi}^3(4)\chi(1-x)-2\frac{(q-1)}{q^2}\varphi(x-1).$$

Now, considering $\chi=\omega^a$ and then applying the Gross–Koblitz formula we obtain

$$B=\frac{1}{q^2}\sum_{a=0}^{q-2}\omega^a(1-x)\overline{\omega}^{3a}(4)\pi^{(p-1)\alpha}\prod_{i=0}^{r-1}\frac{\Gamma_p\big(\big\langle\frac{-4ap^i}{q-1}\big\rangle\big)\Gamma_p^2\big(\big\langle\frac{ap^i}{q-1}\big\rangle\big)}{\Gamma_p\big(\big\langle\frac{-2ap^i}{q-1}\big\rangle\big)}-2\frac{(q-1)}{q^2}\varphi(x-1),$$

where

$$\alpha=\sum_{i=0}^{r-1}\Big\{\Big\langle\frac{-4ap^i}{q-1}\Big\rangle+2\Big\langle\frac{ap^i}{q-1}\Big\rangle-\Big\langle\frac{-2ap^i}{q-1}\Big\rangle\Big\}.$$

Proceeding in a similar way to that shown in the proof of Proposition 3.1, we deduce:

$$B=-\frac{q-1}{q^2}\cdot{}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\,\Big|\,\frac{1}{1-x}\right]_q-2\frac{q-1}{q^2}\varphi(x-1).\qquad\square$$

Before we prove our main results, we now recall the following definition of a finite field hypergeometric function introduced by McCarthy [2012c].

**Definition 3.3** [McCarthy 2012c, Definition 1.4]. Let $A_0,A_1,\dots,A_n,B_1,B_2,\dots,B_n$ be in $\widehat{\mathbb{F}_q^\times}$. Then the ${}_{n+1}F_n(\cdots)^*$ finite field hypergeometric function over $\mathbb{F}_q$ is defined by

$${}_{n+1}F_n\left(\begin{matrix}A_0, & A_1, & \dots, & A_n\\ & B_1, & \dots, & B_n\end{matrix}\,\Big|\,x\right)_q^*=$$
$$\frac{1}{q-1}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\prod_{i=0}^{n}\frac{g(A_i\chi)}{g(A_i)}\prod_{j=1}^{n}\frac{g(\overline{B_j\chi})}{g(\overline{B_j})}g(\overline{\chi})\chi(-1)^{n+1}\chi(x).$$

The following proposition gives a relation between McCarthy's and Greene's finite field hypergeometric functions when certain conditions on the parameters are satisfied.

**Proposition 3.4** [McCarthy 2012c, Proposition 2.5]. *If $A_0 \neq \varepsilon$ and $A_i \neq B_i$ for $1 \le i \le n$, then*

$$ {}_{n+1}F_n\left(\begin{matrix} A_0, & A_1, & \dots, & A_n \\ & B_1, & \dots, & B_n \end{matrix} \Bigg| x\right)^*_q = \left[\prod_{i=1}^{n} \binom{A_i}{B_i}^{-1}\right] {}_{n+1}F_n\left(\begin{matrix} A_0, & A_1, & \dots, & A_n \\ & B_1, & \dots, & B_n \end{matrix} \Bigg| x\right)_q. $$

McCarthy [2013, Lemma 3.3] proved a relation between ${}_{n+1}F_n(\cdots)^*$ and the $p$-adic hypergeometric series ${}_nG_n[\cdots]$. We note that the relation is true for $\mathbb{F}_q$ though it was proved for $\mathbb{F}_p$ in [McCarthy 2013]. Hence, we obtain a relation between ${}_nG_n[\cdots]$ and the Greene's finite field hypergeometric functions due to Proposition 3.4. In the following proposition, we list three such identities which will be used to prove our main results.

**Proposition 3.5.** *Let $x \neq 0$. Then*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \Bigg| x\right]_q = -q \cdot {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)_q; \tag{3-7} $$

$$ {}_2G_2\left[\begin{matrix} \frac{1}{2}, & \frac{1}{2} \\ 0, & 0 \end{matrix} \Bigg| x\right]_q = -q \cdot {}_2F_1\left(\begin{matrix} \varphi, & \varphi \\ & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)_q; \tag{3-8} $$

$$ {}_3G_3\left[\begin{matrix} \frac{1}{2}, & \frac{1}{2} & \frac{1}{2} \\ 0, & 0, & 0 \end{matrix} \Bigg| x\right]_q = q^2 \cdot {}_3F_2\left(\begin{matrix} \varphi, & \varphi, & \varphi \\ & \varepsilon, & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)_q. \tag{3-9} $$

*We note that* (3-7) *is valid when* $q \equiv 1 \pmod 4$.

*Proof.* Applying [McCarthy 2013, Lemma 3.3] we have

$$ {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)^*_q = {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \Bigg| x\right]_q. \tag{3-10} $$

From (2-4), we have $\binom{\chi_4^3}{\varepsilon} = \frac{-1}{q}$. Using this value and Proposition 3.4 we find that

$$ {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)_q = -\frac{1}{q} {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \Bigg| \frac{1}{x}\right)^*_q. \tag{3-11} $$

Now, combining (3-10) and (3-11) we readily obtain (3-7). Proceeding similarly we deduce (3-8) and (3-9). This completes the proof. □

We now prove our main results.

*Proof of Theorem 1.1.* From Proposition 3.1 and Proposition 3.2 we have

$$ \sum_{y \in \mathbb{F}_q} \varphi(y)\varphi(1-2y+xy^2) = -\varphi(-2) \cdot {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \Bigg| \frac{1}{x}\right]_q = -{}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \Bigg| \frac{1}{1-x}\right]_q, $$

which readily gives the desired transformation. □

*Proof of Theorem 1.3.* From [Greene and Stanton 1986, Equation 4.5] we have

$$\varphi\Big(\frac{1-u}{u}\Big)\,{}_3F_2\Big(\begin{matrix}\varphi, & \varphi, & \varphi\\ & \varepsilon, & \varepsilon\end{matrix}\ \Big|\ \frac{u}{u-1}\Big)_p = \varphi(u)f(u)^2 + 2\frac{\varphi(-1)}{p}f(u) - \frac{p-1}{p^2}\varphi(u) + \frac{p-1}{p^2}\delta(1-u), \tag{3-12}$$

where $u = x/(x-1),\ x \neq 1$ and

$$f(u) := \frac{p}{p-1}\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{u}{4}\Big).$$

From (3-9) and (3-12), we have

$$\frac{\varphi((1-u)/u)}{p^2}\cdot{}_3G_3\Big[\begin{matrix}\frac12, & \frac12, & \frac12\\ 0, & 0, & 0\end{matrix}\ \Big|\ \frac{u-1}{u}\Big]_p = \varphi(u)f(u)^2 + 2\frac{\varphi(-1)}{p}f(u) - \frac{p-1}{p^2}\varphi(u) + \frac{p-1}{p^2}\delta(1-u). \tag{3-13}$$

Now, Proposition 3.1 gives

$$f(u) = \frac{-\varphi(-u)}{p} - \frac{1}{p}\cdot{}_2G_2\Big[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\ \Big|\ \frac{1}{u}\Big]_p. \tag{3-14}$$

Finally, combining (3-13) and (3-14) and then putting $u = \frac{x}{x-1}$ we obtain the desired result. This completes the proof of the theorem. □

*Proof of Theorem 1.4.* Let $A = B = \varphi$ and $x \neq 0, \pm1$. Then [Greene 1987, Theorem 4.16] yields

$${}_2F_1\Big(\begin{matrix}\varphi, & \varphi\\ & \varepsilon\end{matrix}\ \Big|\ x\Big)_q = \frac{\varphi(-1)}{q}\varphi(x(1+x)) + \varphi(1+x)\frac{q}{q-1}\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{(1+x)^2}\Big). \tag{3-15}$$

Now, using Proposition 3.1 we have

$$\sum_{\chi\in\widehat{\mathbb{F}_q^\times}}\binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{(1+x)^2}\Big) = -\frac{q-1}{q^2}\varphi\Big(\frac{-4x}{(1+x)^2}\Big) - \frac{q-1}{q^2}\cdot{}_2G_2\Big[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\ \Big|\ \frac{(1+x)^2}{4x}\Big]_q. \tag{3-16}$$

Applying Theorem 1.1 on the right-hand side of (3-16) we obtain

$$\sum_{\chi\in\widehat{\mathbb{F}_q^\times}} \binom{\varphi\chi^2}{\chi}\binom{\varphi\chi}{\chi}\chi\Big(\frac{x}{(1+x)^2}\Big) = -\frac{q-1}{q^2}\varphi(-x)-\frac{q-1}{q^2}\varphi(-2)\cdot {}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\;\middle|\;\frac{(1+x)^2}{(1-x)^2}\right]_q. \tag{3-17}$$

Combining (3-15) and (3-17) we have

$${}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\;\middle|\;\frac{(1+x)^2}{(1-x)^2}\right]_q = -q\varphi(-2)\varphi(1+x)\cdot {}_2F_1\left(\begin{matrix}\varphi, & \varphi\\ & \varepsilon\end{matrix}\;\middle|\;x\right)_q, \tag{3-18}$$

which completes the proof of the theorem due to (3-8). □

*Proof of Theorem 1.5.* Let $q\equiv 1 \pmod 4$. Then we readily obtain the desired transformation for the finite field hypergeometric functions from (1-4) using (3-7) and (3-8). □

## 4. Special values of ${}_2G_2[\cdots]$

Finding special values of hypergeometric function is an important and interesting problem. Only a few special values of the ${}_nG_n$-functions are known; see for example [Barman et al. 2015]. Therein, we obtained some special values of ${}_nG_n[\cdots]$ when $n=2,3,4$. From (3-18), for any odd prime $p$ and $x\neq 0,\pm1$, we have

$${}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\;\middle|\;\frac{(1+x)^2}{(1-x)^2}\right]_q = -q\varphi(-2)\varphi(1+x)\cdot {}_2F_1\left(\begin{matrix}\varphi, & \varphi\\ & \varepsilon\end{matrix}\;\middle|\;x\right)_q. \tag{4-1}$$

Values of the finite field hypergeometric function ${}_2F_1\left(\begin{smallmatrix}\varphi, & \varphi\\ & \varepsilon\end{smallmatrix}\mid x\right)_q$ are obtained for many values of $x$. For example, see [Barman and Kalita 2012; 2013a; Evans and Greene 2009b; Greene 1987; Kalita 2018; Ono 1998].

*Proof of Theorem 1.6.* Let $\lambda\in\{-1,\frac12,2\}$. If $p$ is an odd prime, then from [Ono 1998, Theorem 2] we have

$${}_2F_1\left(\begin{matrix}\varphi, & \varphi\\ & \varepsilon\end{matrix}\;\middle|\;\lambda\right)_p = \begin{cases} 0 & \text{if } p\equiv 3 \pmod 4;\\ \dfrac{2x}{p}(-1)^{\frac{x+y+1}{2}} & \text{if } p\equiv 1 \pmod 4,\ x^2+y^2=p, \text{ and } x \text{ odd.}\end{cases}$$

Putting the above values for $\lambda=\frac12, 2$ into (4-1) we readily obtain the required values of the ${}_2G_2$-function.

Let $q\equiv 1 \pmod 4$. Then from (3-7) we have

$${}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\middle|\;\frac19\right)_q = -\frac1q\,{}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix}\;\middle|\;9\right]_q.$$

From the above identity we readily obtain the required value of the finite field hypergeometric function. This completes the proof of the theorem. □

**Corollary 4.1.** *Let* $p \equiv 1 \pmod 4$. *We have*

$$\binom{\chi_4}{\varphi} + \binom{\chi_4^3}{\varphi} = \frac{2x(-1)^{\frac{x+y+1}{2}}}{p},$$

*where* $x^2 + y^2 = p$ *and* $x$ *is odd.*

*Proof.* From Theorem 1.6 and [Barman and Kalita 2013a, Theorem 1.4(i)] we have

$$\binom{\chi_4}{\varphi} + \binom{\chi_4^3}{\varphi} = \frac{2x\varphi(2)\chi_4(-1)(-1)^{\frac{x+y+1}{2}}}{p},$$

where $x^2 + y^2 = p$ and $x$ is odd. Let $m$ be the order of $\chi \in \widehat{\mathbb{F}_q^\times}$. We know that $\chi(-1) = -1$ if and only if $m$ is even and $(q-1)/m$ is odd. Since $p \equiv 1 \pmod 4$, therefore, either $p \equiv 1 \pmod 8$ or $p \equiv 5 \pmod 8$. If $p \equiv 1 \pmod 8$, then $\varphi(2) = \chi_4(-1) = 1$. Also, if $p \equiv 5 \pmod 8$, then $\varphi(2) = \chi_4(-1) = -1$. Hence, in both the cases, $\varphi(2) \cdot \chi_4(-1) = 1$. This completes the proof. ☐

*Proof of Theorem 1.7.* From [Kalita 2018, Theorem 1.1], for $q \equiv 1 \pmod 8$, we have

$$(4\text{-}2) \qquad {}_2F_1\left(\begin{matrix}\varphi, & \varphi\\ & \varepsilon\end{matrix} \;\middle|\; \frac{4\sqrt{2}}{2\sqrt{2}\pm3}\right)_q = \varphi(3\pm2\sqrt{2})\left\{\binom{\chi_4}{\varphi} + \binom{\chi_4^3}{\varphi}\right\}.$$

Now, comparing (3-18) and (4-2) for $x = 4\sqrt{2}/(2\sqrt{2}\pm3)$, we obtain (1-5). Similarly, using [Kalita 2018, Theorem 1.1] and (3-18) for $x = 4/(2\pm\sqrt{3})$ we derive (1-6) and (1-7). ☐

*Proof of Theorem 1.8.* From (3-7), we have

$$(4\text{-}3) \qquad {}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix} \;\middle|\; \left(\frac{-2\sqrt{2}\pm3}{6\sqrt{2}\pm3}\right)^2\right)_q = -\frac{1}{q}\cdot {}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix} \;\middle|\; \left(\frac{6\sqrt{2}\pm3}{-2\sqrt{2}\pm3}\right)^2\right]_q.$$

Comparing (1-5) and (4-3) we readily obtain (1-8). Again, we have

$$(4\text{-}4) \qquad {}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix} \;\middle|\; \left(\frac{-2\pm\sqrt{3}}{6\pm\sqrt{3}}\right)^2\right)_q = -\frac{1}{q}\cdot {}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix} \;\middle|\; \left(\frac{6\pm\sqrt{3}}{-2\pm\sqrt{3}}\right)^2\right]_q.$$

Now, comparing (1-7) and (4-4) we deduce (1-9). ☐

Applying Corollary 4.1, from (1-5) and (1-8) we have the following corollary.

**Corollary 4.2.** *Let* $p \equiv 1 \pmod 8$. *Then*

$${}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & 0\end{matrix} \;\middle|\; \left(\frac{6\sqrt{2}\pm3}{-2\sqrt{2}\pm3}\right)^2\right]_p = -2x\varphi(6\pm12\sqrt{2})(-1)^{\frac{x+y+1}{2}},$$

*where* $x^2 + y^2 = p$ *and* $x$ *is odd.*

***Acknowledgements.*** We thank the referee for valuable comments. This work is partially supported by a start up grant awarded to Barman by the Indian Institute of Technology Guwahati. Saikia acknowledges the financial support of the Department of Science and Technology, Government of India for supporting a part of this work under the INSPIRE Faculty Fellowship.

## References

[Ahlgren and Ono 2000] S. Ahlgren and K. Ono, "A Gaussian hypergeometric series evaluation and Apéry number congruences", *J. Reine Angew. Math.* **518** (2000), 187–212. MR Zbl

[Bailey 1935] W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Math. and Math. Phys. **32**, Cambridge Univ. Press, 1935. Zbl

[Barman and Kalita 2012] R. Barman and G. Kalita, "Certain values of Gaussian hypergeometric series and a family of algebraic curves", *Int. J. Number Theory* **8**:4 (2012), 945–961. MR Zbl

[Barman and Kalita 2013a] R. Barman and G. Kalita, "Elliptic curves and special values of Gaussian hypergeometric series", *J. Number Theory* **133**:9 (2013), 3099–3111. MR Zbl

[Barman and Kalita 2013b] R. Barman and G. Kalita, "Hypergeometric functions over $\mathbb{F}_q$ and traces of Frobenius for elliptic curves", *Proc. Amer. Math. Soc.* **141**:10 (2013), 3403–3410. MR Zbl

[Barman and Saikia 2014] R. Barman and N. Saikia, "$p$-adic gamma function and the trace of Frobenius of elliptic curves", *J. Number Theory* **140** (2014), 181–195. MR Zbl

[Barman and Saikia 2015] R. Barman and N. Saikia, "Certain transformations for hypergeometric series in the $p$-adic setting", *Int. J. Number Theory* **11**:2 (2015), 645–660. MR Zbl

[Barman et al. 2015] R. Barman, N. Saikia, and D. McCarthy, "Summation identities and special values of hypergeometric series in the $p$-adic setting", *J. Number Theory* **153** (2015), 63–84. MR Zbl

[Berndt et al. 1998] B. C. Berndt, R. J. Evans, and K. S. Williams, *Gauss and Jacobi sums*, Wiley, New York, 1998. MR Zbl

[Evans 2010] R. Evans, "Hypergeometric ${}_3F_2(1/4)$ evaluations over finite fields and Hecke eigenforms", *Proc. Amer. Math. Soc.* **138**:2 (2010), 517–531. MR Zbl

[Evans and Greene 2009a] R. Evans and J. Greene, "Clausen's theorem and hypergeometric functions over finite fields", *Finite Fields Appl.* **15**:1 (2009), 97–109. MR Zbl

[Evans and Greene 2009b] R. Evans and J. Greene, "Evaluations of hypergeometric functions over finite fields", *Hiroshima Math. J.* **39**:2 (2009), 217–235. MR Zbl

[Frechette et al. 2004] S. Frechette, K. Ono, and M. Papanikolas, "Gaussian hypergeometric functions and traces of Hecke operators", *Int. Math. Res. Not.* **2004**:60 (2004), 3233–3262. MR Zbl

[Fuselier 2010] J. G. Fuselier, "Hypergeometric functions over $\mathbb{F}_p$ and relations to elliptic curves and modular forms", *Proc. Amer. Math. Soc.* **138**:1 (2010), 109–123. MR Zbl

[Fuselier and McCarthy 2016] J. G. Fuselier and D. McCarthy, "Hypergeometric type identities in the $p$-adic setting and modular forms", *Proc. Amer. Math. Soc.* **144**:4 (2016), 1493–1508. MR Zbl

[Greene 1984] J. R. Greene, *Character sum analogues for hypergeometric and generalized hypergeometric functions over finite fields*, Ph.D. thesis, University of Minnesota, Minneapolis, 1984.

[Greene 1987] J. Greene, "Hypergeometric functions over finite fields", *Trans. Amer. Math. Soc.* **301**:1 (1987), 77–101. MR Zbl

[Greene and Stanton 1986] J. Greene and D. Stanton, "A character sum evaluation and Gaussian hypergeometric series", *J. Number Theory* **23**:1 (1986), 136–148. MR Zbl

[Gross and Koblitz 1979] B. H. Gross and N. Koblitz, "Gauss sums and the $p$-adic $\Gamma$-function", *Ann. of Math.* (2) **109**:3 (1979), 569–581. MR Zbl

[Kalita 2018] G. Kalita, "Values of Gaussian hypergeometric series and their connections to algebraic curves", *Int. J. Number Theory* **14**:1 (2018), 1–18. MR Zbl

[Koblitz 1980] N. Koblitz, *p-adic analysis: a short course on recent work*, London Math. Soc. Lecture Note Series **46**, Cambridge Univ. Press, 1980. MR Zbl

[Koike 1992] M. Koike, "Hypergeometric series over finite fields and Apéry numbers", *Hiroshima Math. J.* **22**:3 (1992), 461–467. MR Zbl

[Lennon 2011a] C. Lennon, "Gaussian hypergeometric evaluations of traces of Frobenius for elliptic curves", *Proc. Amer. Math. Soc.* **139**:6 (2011), 1931–1938. MR Zbl

[Lennon 2011b] C. Lennon, "Trace formulas for Hecke operators, Gaussian hypergeometric functions, and the modularity of a threefold", *J. Number Theory* **131**:12 (2011), 2320–2351. MR Zbl

[McCarthy 2012a] D. McCarthy, "Extending Gaussian hypergeometric series to the $p$-adic setting", *Int. J. Number Theory* **8**:7 (2012), 1581–1612. MR Zbl

[McCarthy 2012b] D. McCarthy, "On a supercongruence conjecture of Rodriguez-Villegas", *Proc. Amer. Math. Soc.* **140**:7 (2012), 2241–2254. MR Zbl

[McCarthy 2012c] D. McCarthy, "Transformations of well-poised hypergeometric functions over finite fields", *Finite Fields Appl.* **18**:6 (2012), 1133–1147. MR Zbl

[McCarthy 2013] D. McCarthy, "The trace of Frobenius of elliptic curves and the $p$-adic gamma function", *Pacific J. Math.* **261**:1 (2013), 219–236. MR Zbl

[McCarthy and Papanikolas 2015] D. McCarthy and M. A. Papanikolas, "A finite field hypergeometric function associated to eigenvalues of a Siegel eigenform", *Int. J. Number Theory* **11**:8 (2015), 2431–2450. MR Zbl

[Mortenson 2005] E. Mortenson, "Supercongruences for truncated ${}_{n+1}F_n$ hypergeometric series with applications to certain weight three newforms", *Proc. Amer. Math. Soc.* **133**:2 (2005), 321–330. MR Zbl

[Ono 1998] K. Ono, "Values of Gaussian hypergeometric series", *Trans. Amer. Math. Soc.* **350**:3 (1998), 1205–1223. MR Zbl

[Salerno 2013] A. Salerno, "Counting points over finite fields and hypergeometric functions", *Funct. Approx. Comment. Math.* **49**:1 (2013), 137–157. MR Zbl

[Vega 2011] M. V. Vega, "Hypergeometric functions over finite fields and their relations to algebraic curves", *Int. J. Number Theory* **7**:8 (2011), 2171–2195. MR Zbl

Received June 23, 2017. Revised February 6, 2018.

RUPAM BARMAN
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI
ASSAM
INDIA
rupam@iitg.ac.in

NEELAM SAIKIA
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI
ASSAM
INDIA
neelam16@iitg.ernet.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.291

# ON THE STRUCTURE OF HOLOMORPHIC ISOMETRIC EMBEDDINGS OF COMPLEX UNIT BALLS INTO BOUNDED SYMMETRIC DOMAINS

SHAN TAI CHAN

**We study general properties of holomorphic isometric embeddings of complex unit balls $\mathbb{B}^n$ into bounded symmetric domains of rank $\geq 2$. In the first part, we study holomorphic isometries from $(\mathbb{B}^n, kg_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ with nonminimal isometric constants $k$ for any irreducible bounded symmetric domain $\Omega$ of rank $\geq 2$, where $g_D$ denotes the canonical Kähler–Einstein metric on any irreducible bounded symmetric domain $D$ normalized so that minimal disks of $D$ are of constant Gaussian curvature $-2$. In particular, results concerning the upper bound of the dimension of isometrically embedded $\mathbb{B}^n$ in $\Omega$ and the structure of the images of such holomorphic isometries are obtained.**

**In the second part, we study holomorphic isometries from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ for any irreducible bounded symmetric domains $\Omega \Subset \mathbb{C}^N$ of rank equal to 2 with $2N > N'+1$, where $N'$ is an integer such that $\iota : X_c \hookrightarrow \mathbb{P}^{N'}$ is the minimal embedding (i.e., the first canonical embedding) of the compact dual Hermitian symmetric space $X_c$ of $\Omega$. We completely classify images of all holomorphic isometries from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ for $1 \leq n \leq n_0(\Omega)$, where $n_0(\Omega) := 2N - N' > 1$. In particular, for $1 \leq n \leq n_0(\Omega) - 1$ we prove that any holomorphic isometry from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ extends to some holomorphic isometry from $(\mathbb{B}^{n_0(\Omega)}, g_{\mathbb{B}^{n_0(\Omega)}})$ to $(\Omega, g_\Omega)$.**

## 1. Introduction

Calabi [1953] studied local holomorphic isometries from Kähler manifolds endowed with real-analytic metrics into complex space forms and their local rigidity. Many results concerning local holomorphic isometric embeddings between bounded symmetric domains were obtained by Mok [2002b; 2011; 2012; 2016] and by Ng [2010; 2011]. In [Chan and Mok 2017], henceforth abbreviated [CM], Mok and the author obtained a general result concerning general properties of the images of holomorphic isometric embeddings from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$, where $g_D$ denotes

*MSC2010:* 32M15, 53C55, 53C42.
*Keywords:* Bergman metrics, holomorphic isometric embeddings, bounded symmetric domains, Borel embedding, complex unit balls.

the canonical Kähler–Einstein metric on $D$ normalized so that minimal disks of $D$ are of constant Gaussian curvature $-2$ for any irreducible bounded symmetric domain $D \Subset \mathbb{C}^N$ in its Harish-Chandra realization. In addition, Mok and the author [CM] classified images of all holomorphic isometric embeddings from $(\mathbb{B}^m, g_{\mathbb{B}^m})$ to $(D_n^{\mathrm{IV}}, g_{D_n^{\mathrm{IV}}})$ for $1 \le m \le n-1$ and $n \ge 3$, where $D_n^{\mathrm{IV}}$ denotes the type-IV domain (i.e., the Lie ball) of complex dimension $n$ (see Section 2). On the other hand, Xiao and Yuan [2016] and Upmeier, Wang and Zhang [Upmeier et al. 2016] classified all holomorphic isometric embeddings from $(\mathbb{B}^{n-1}, g_{\mathbb{B}^{n-1}})$ to $(D_n^{\mathrm{IV}}, g_{D_n^{\mathrm{IV}}})$, $n \ge 3$, independently with explicit parametrizations. Moreover, Xiao and Yuan [2016, Theorem 1.1] proved that any proper holomorphic map from the complex unit $m$-ball $\mathbb{B}^m$ to $D_n^{\mathrm{IV}}$, $n \ge 3$ and $m \le n-1$, with certain boundary regularities is a holomorphic isometric embedding provided that the codimension $n-m$ of the image of the $m$-ball is sufficiently small and $m \ge 4$.

In the present article, we also denote by $ds_U^2$ the Bergman metric of any bounded domain $U \Subset \mathbb{C}^N$ and we will simply use the term "*holomorphic isometries*" for holomorphic isometric embeddings. In what follows, we will assume that any bounded symmetric domain in a complex Euclidean space is in its Harish-Chandra realization.

Let $f : (\mathbb{B}^n, \lambda' g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ be a holomorphic isometry for some positive real constant $\lambda'$, where $\Omega$ is an irreducible bounded symmetric domain. It is well known that any bounded symmetric domain is equivalently a Hermitian symmetric space of the noncompact type and vice versa by the Harish-Chandra embedding theorem; see [Wolf 1972; Mok 1989]. Then, it follows from [CM, Lemma 3] that $\lambda'$ is a positive integer satisfying $1 \le \lambda' \le r$, where $r := \mathrm{rank}(\Omega)$ is the rank of $\Omega$ as a Hermitian symmetric space of the noncompact type. Throughout the present article, we will call $\lambda'$ the *isometric constant* of any given holomorphic isometry from $(\mathbb{B}^n, \lambda' g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$. In addition, given any holomorphic isometry $F : (\Delta, k\, ds_\Delta^2) \to (\Delta^p, ds_{\Delta^p}^2)$, we will call $k$ the *isometric constant* of $F$, where $\Delta \Subset \mathbb{C}$ (resp. $\Delta^p \Subset \mathbb{C}^p$) denotes the open unit disk (resp. open unit polydisk) in the complex plane $\mathbb{C}$ (resp. the complex $p$-dimensional Euclidean space $\mathbb{C}^p$).

In the present article, we denote by $\widehat{\mathrm{HI}}_k(\mathbb{B}^n, \Omega)$ the space of all holomorphic isometries from $(\mathbb{B}^n, k g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$, where $k$ is any positive integer satisfying $1 \le k \le \mathrm{rank}(\Omega)$. Motivated by [Mok 2016] and [CM], we continue to study the structure of holomorphic isometries from $(\mathbb{B}^n, k g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ for any irreducible bounded symmetric domain $\Omega$ of rank $r \ge 2$ and any positive integer $k$ such that $1 \le k \le r$.

In the first part, we consider the case where $k \ge 2$ is not the minimal isometric constant and obtain a result similar to [CM, Theorem 1] when the isometric constant $k$ is equal to 2. As a corollary of this result, we will also show that given any irreducible bounded symmetric domain $\Omega$ of rank at most 3, all holomorphic isometries from $(\mathbb{B}^n, k g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ arise from linear sections of the minimal embedding of the compact dual Hermitian symmetric space $X_c$ of $\Omega$.

In the second part, the aim is to generalize our results in [CM] for type-IV domains to more general irreducible bounded symmetric domains $\Omega$ of rank 2. Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$. Mok [2016] proved that if $f : (\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ is a holomorphic isometry, then $n \leq p(\Omega) + 1$, where $p(\Omega) := p(X_c) = p$ is defined by $c_1(X_c) = (p+2)\delta$ for the compact dual Hermitian symmetric space $X_c$ of $\Omega$ and the positive generator $\delta$ of $H^2(X_c, \mathbb{Z}) \cong \mathbb{Z}$; see [Mok 2016] and [CM]. By slicing the complex unit ball $\mathbb{B}^{p(\Omega)+1}$ with affine linear subspaces $L$ of $\mathbb{C}^{p(\Omega)+1}$ such that $L \cap \mathbb{B}^{p(\Omega)+1}$ is nonempty, we obtain many holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ from any given holomorphic isometry $F \in \widehat{\mathrm{HI}}_1(\mathbb{B}^{p(\Omega)+1}, \Omega)$ for $n \leq p(\Omega)$. It is natural to ask whether all holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ were obtained in that way for each $n \leq p(\Omega)$. In the case where $\Omega = D^{\mathrm{IV}}_N$ is the type-IV domain for some integer $N \geq 3$, the author and Mok [CM, Theorem 2] have shown that the answer is affirmative. In general, this problem remains open. In [CM], we showed that holomorphic isometries from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ arise from linear sections of the compact dual $X_c$ of $\Omega$, where $\Omega$ is an irreducible bounded symmetric domain of rank $\geq 2$. In general, we do not know whether this gives any relation between the spaces $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ and $\widehat{\mathrm{HI}}_1(\mathbb{B}^m, \Omega)$ for $1 \leq n < m \leq p(\Omega) + 1$, except in the case where $\Omega = D^{\mathrm{IV}}_N$, $N \geq 3$, is the type-IV domain; see [CM]. Recall that a type-IV domain is of rank 2. On the other hand, for a rank-$r$ irreducible bounded symmetric domain $\Omega$, any holomorphic isometry from $(\mathbb{B}^n, r g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ is totally geodesic by the Ahlfors–Schwarz lemma; see [CM, Proposition 1]. In particular, we only need to consider the space $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ if $\Omega$ is of rank 2. Therefore, it is natural to study the problem when the target bounded symmetric domain $\Omega$ is of rank 2.

In short, we will generalize the method in [CM] for classifying images of all holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, D^{\mathrm{IV}}_N)$ for $N \geq 3$ and $n \geq 1$ to the study of images of holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ for $1 \leq n \leq n_0$ and certain irreducible bounded symmetric domains $\Omega \Subset \mathbb{C}^N$ of rank 2, where $n_0 = n_0(\Omega) > 1$ is some integer depending on $\Omega$. One of the key ingredients is the use of the explicit form of the polynomial $h_\Omega(z, z)$, as mentioned in [CM, Remark 1]. On the other hand, the author has found that the relation between $h_\Omega(z, \xi)$ and $\iota|_{\mathbb{C}^N}$ obtained from [Loos 1977] has been written down explicitly by Fang, Huang and Xiao [Fang et al. 2016] for each irreducible bounded symmetric domain $\Omega$, where $\iota : X_c \hookrightarrow \mathbb{P}\big(\Gamma(X_c, \mathcal{O}(1))^*\big) \cong \mathbb{P}^{N'}$ is the minimal embedding, i.e., the first canonical embedding; see [Nakagawa and Takagi 1976]. Here $\mathcal{O}(1)$ is the positive generator of the Picard group $\mathrm{Pic}(X_c) \cong \mathbb{Z}$ of the compact dual $X_c$ of $\Omega$, and $\mathbb{C}^N \subset X_c$ is identified as a dense open subset of $X_c$ by the Harish-Chandra embedding theorem; see [Mok 1989; 2016] and [CM]. In addition, $\Gamma(X_c, \mathcal{O}(1))^*$ denotes the dual of the space $\Gamma(X_c, \mathcal{O}(1))$ of all holomorphic sections of the holomorphic line bundle $\mathcal{O}(1)$ over $X_c$; see [Mok 2016] and [CM]. We refer the readers to [CM, Section 2.1] for the background of bounded symmetric domains and their compact dual Hermitian symmetric spaces. We will identify $\mathbb{P}\big(\Gamma(X_c, \mathcal{O}(1))^*\big) = \mathbb{P}^{N'}$ and write $N' := \dim_{\mathbb{C}} \mathbb{P}\big(\Gamma(X_c, \mathcal{O}(1))^*\big)$ throughout the present article, where $X_c$ is the compact dual Hermitian symmetric space of the irreducible bounded symmetric domain $\Omega$.

The main results in the first part of the present article are as follows.

**Theorem 1.1.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$ and $\lambda' \geq 2$ be an integer. If $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$, then we have $n \leq n_{\lambda'-1}(\Omega)$, where $n_{\lambda'-1}(\Omega)$ is the $(\lambda'-1)$-th null dimension of $\Omega$* (see [Mok 1989, p. 253] and Section 2A).

**Theorem 1.2.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain with* $\mathrm{rank}(\Omega) =: r \geq 2$ *and $f \in \widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ for some real constant $\lambda' > 0$. We have the standard embeddings $\Omega \Subset \mathbb{C}^N \subset X_c$ of $\Omega$ as a bounded domain and its Borel embedding $\Omega \subset X_c$ as an open subset of its compact dual Hermitian symmetric space $X_c$* (see [CM, Theorem 1]). *Suppose that either $\lambda' = 2$ or $2 \leq r \leq 3$. Then, $f(\mathbb{B}^n)$ is an irreducible component of $\mathscr{V} := \mathscr{V}' \cap \Omega$ for some affine-algebraic subvariety $\mathscr{V}' \subset \mathbb{C}^N$ such that $\iota(\mathscr{V}) = P \cap \iota(\Omega)$, where $P \subseteq \mathbb{P}^{N'}$ is some projective linear subspace and $\iota : X_c \hookrightarrow \mathbb{P}^{N'}$ is the minimal embedding.*

The main result of the second part is the following.

**Theorem 1.3.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank* 2 *satisfying $2N > N' + 1$, where $N' := \dim_{\mathbb{C}} \mathbb{P}\big(\Gamma(X_c, \mathcal{O}(1))^*\big)$ and $X_c$ is the compact dual Hermitian symmetric space of $\Omega$. Set $n_0(\Omega) := 2N - N'$. For $1 \leq n \leq n_0(\Omega) - 1$, if $f : (\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ is a holomorphic isometric embedding, then $f = F \circ \rho$ for some holomorphic isometric embeddings $F : (\mathbb{B}^{n_0(\Omega)}, g_{\mathbb{B}^{n_0(\Omega)}}) \to (\Omega, g_\Omega)$ and $\rho : (\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\mathbb{B}^{n_0(\Omega)}, g_{\mathbb{B}^{n_0(\Omega)}})$.*

**Remark 1.4.** (1) Theorem 1.3 actually asserts that any holomorphic isometric embedding $f \in \widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$, $1 \leq n \leq n_0(\Omega) - 1$, extends to a holomorphic isometric embedding $F \in \widehat{\mathrm{HI}}_1(\mathbb{B}^{n_0(\Omega)}, \Omega)$, where $\Omega \Subset \mathbb{C}^N$ is a rank-2 irreducible bounded symmetric domain satisfying $2N > N' + 1$.

(2) We will show that for such irreducible bounded symmetric domains $\Omega$, we have $n_0(\Omega) = p(\Omega) + 1$ only if $\Omega \cong D^{\mathrm{IV}}_N$ is the type-IV domain for some $N \geq 3$. Therefore, one may regard this theorem as a generalization of Theorem 2 in [CM] to holomorphic isometric embeddings from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ for any rank-2 irreducible bounded symmetric domain $\Omega$ satisfying $n_0(\Omega) > 1$ and $1 \leq n \leq n_0(\Omega) - 1$.

## 2. Preliminaries

Denote by $\|\boldsymbol{v}\|_{\mathbb{C}^n}$ the standard complex Euclidean norm of any vector $\boldsymbol{v}$ in $\mathbb{C}^n$. The following lemma is a special case of a well-known result of Calabi [1953, Theorem 2 (local rigidity)]:

**Lemma 2.1** [Calabi 1953; Ng 2011, Lemma 3.3]. *Let $g, f : B \to \mathbb{C}^N$ be holomorphic maps defined on some open subset $B \subset \mathbb{C}^n$ such that $\|f(w)\|^2_{\mathbb{C}^N} = \|g(w)\|^2_{\mathbb{C}^N}$ for any $w \in B$. Then, there exists a unitary transformation $U$ in $\mathbb{C}^N$ such that $f = U \circ g$.*

**Remark 2.2.** Writing $f = (f^1, \ldots, f^N)$ and $g = (g^1, \ldots, g^N)$, there exists an $N \times N$ unitary matrix $\boldsymbol{U}'$ such that

$$\boldsymbol{U}'(f^1(w), \ldots, f^N(w))^T = (g^1(w), \ldots, g^N(w))^T \quad \text{for all } w \in B.$$

Moreover, we have the following fact from linear algebra.

**Lemma 2.3** [CM, Lemma 5]. *Let $m'$ and $n'$ be integers such that $1 \le m' < n'$ and let $\boldsymbol{A}'' \in M(m', n'; \mathbb{C})$ be such that $\boldsymbol{A}''\overline{\boldsymbol{A}}''^T = \boldsymbol{I}_{m'}$. Then, there exists $\boldsymbol{U}' \in M(n'-m', n'; \mathbb{C})$ such that*

$$\begin{bmatrix} \boldsymbol{U}' \\ \boldsymbol{A}'' \end{bmatrix} \in U(n').$$

For the complex unit ball $\mathbb{B}^n \subset \mathbb{C}^n$, the Kähler form $\omega_{g_{\mathbb{B}^n}}$ of $(\mathbb{B}^n, g_{\mathbb{B}^n})$ is given by

$$\omega_{g_{\mathbb{B}^n}} = -\sqrt{-1}\partial\bar\partial \log(1 - \|w\|^2_{\mathbb{C}^n})$$

so that $(\mathbb{B}^n, g_{\mathbb{B}^n})$ is of constant holomorphic sectional curvature $-2$. Note that the Bergman metric $K_\Omega(z, \xi)$ of $\Omega$ can be expressed as

$$K_\Omega(z, \xi) = \frac{1}{\mathrm{Vol}(\Omega)} h_\Omega(z, \xi)^{-(p(\Omega)+2)},$$

where $\mathrm{Vol}(\Omega)$ is the Euclidean volume of $\Omega \Subset \mathbb{C}^N$, $h_\Omega(z, \xi)$ is some polynomial in $(z, \bar\xi)$ such that $h_\Omega(z, \mathbf{0}) \equiv 1$ and $p(\Omega)$ is defined as in Section 1. It follows from [CM] that the Kähler form $\omega_{g_\Omega}$ of $(\Omega, g_\Omega)$ is given by

$$\omega_{g_\Omega} = -\sqrt{-1}\partial\bar\partial \log h_\Omega(z, z)$$

in terms of the Harish-Chandra coordinates $z \in \Omega \Subset \mathbb{C}^N$. The type-IV domain $D^{\mathrm{IV}}_N$, $N \ge 3$, is given by

$$D^{\mathrm{IV}}_N = \Big\{z = (z_1, \ldots, z_N) \in \mathbb{C}^N : \textstyle\sum_{j=1}^N |z_j|^2 < 2,\ \sum_{j=1}^N |z_j|^2 < 1 + \big|\frac{1}{2}\sum_{j=1}^N z_j^2\big|^2\Big\};$$

see [Mok 1989, p. 83]. Then, the Kähler form $\omega_{g_{D^{\mathrm{IV}}_N}}$ of $(D^{\mathrm{IV}}_N, g_{D^{\mathrm{IV}}_N})$ is given by

$$\omega_{g_{D^{\mathrm{IV}}_N}} = -\sqrt{-1}\partial\bar\partial \log\bigg(1 - \sum_{j=1}^N |z_j|^2 + \bigg|\frac{1}{2}\sum_{j=1}^N z_j^2\bigg|^2\bigg).$$

As mentioned in Section 1, we have the following: for any irreducible bounded symmetric domain $\Omega \Subset \mathbb{C}^N$ of rank $r \ge 2$, we may suppose that the Harish-Chandra coordinates $z = (z_1, \ldots, z_N)$ on $\Omega \Subset \mathbb{C}^N$ are chosen so that there are homogeneous polynomials $G_l(z)$ in $z$ of degree $\deg(G_l)$, $1 \le l \le N'$, such that

(i) $2 \le \deg(G_l) \le r$ for $N+1 \le l \le N'$ and $G_j(z) = z_j$ for $1 \le j \le N$,

(ii) $h_\Omega(z,\xi) = 1 + \sum_{j=1}^{N'} (-1)^{\deg(G_l)} G_l(z)\overline{G_l(\xi)}$ and the restriction of the minimal embedding $\iota : X_c \hookrightarrow \mathbb{P}^{N'}$ to the dense open subset $\mathbb{C}^N \subset X_c$ may be written as

$$\iota(z) = [1, G_1(z), \dots, G_{N'}(z)]$$

in terms of the Harish-Chandra coordinates $z = (z_1, \dots, z_N) \in \mathbb{C}^N$,

(iii) For any integer $\mu$, $2 \le \mu \le r$, there exists $l$, $N+1 \le l \le N'$, such that $\deg(G_l) = \mu$.

For instance, if $\Omega = D_N^{\mathrm{IV}} \Subset \mathbb{C}^N$, $N \ge 3$, is the type-IV domain, then

$$h_\Omega(z,z) = 1 - \sum_{j=1}^{N} |z_j|^2 + \left| \frac{1}{2} \sum_{j=1}^{N} z_j^2 \right|^2 \quad \text{and} \quad \iota(z) = \left[ z_1, \dots, z_N, 1, \frac{1}{2} \sum_{j=1}^{N} z_j^2 \right]$$

for $z = (z_1, \dots, z_N) \in \mathbb{C}^N$; see [Mok 1989, p. 83]. We refer the readers to [Loos 1977; Fang et al. 2016] for details of the above facts.

Let $f : (\mathbb{B}^n, k g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ be a holomorphic isometry such that $f(\mathbf{0}) = \mathbf{0}$, where $\Omega$ is an irreducible bounded symmetric domain of rank $r \ge 2$ and $k$ is an integer such that $1 \le k \le r$. Then, we have the functional equation

$$h_\Omega(f(w), f(w)) = (1 - \|w\|_{\mathbb{C}^n}^2)^k$$

for $w \in \mathbb{B}^n$; see [Mok 2012] and [CM].

**2A.** ***On higher-characteristic bundles over irreducible bounded symmetric domains.*** Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $r$ and $X_c$ be the compact dual of $\Omega$. Throughout this section, we follow [Wolf 1972; Mok 1989, pp. 251–253]. We always identify the base point $o \in X_0$ with $\mathbf{0} \in \Omega = \xi^{-1}(X_0)$, where $\xi : \mathfrak{m}^+ \cong \mathbb{C}^N \to G^{\mathbb{C}}/P \cong X_c$ is the embedding defined by $\xi(v) = \exp(v) \cdot P$; see [Wolf 1972; Mok 1989, p. 94]. Let $\Psi = \{\psi_1, \dots, \psi_r\} \subset \Delta_M^+$ be a maximal strongly orthogonal set of noncompact positive roots; see [Wolf 1972]. Then, we have the corresponding root vectors $e_{\psi_j}$, $1 \le j \le r$. Moreover, we have $\mathfrak{g}_{\psi_j} = \mathbb{C} e_{\psi_j}$ for $1 \le j \le r$ and the maximal polydisk $\Delta^r \cong \Pi \subset \Omega$ is given by $\Pi = \left( \bigoplus_{j=1}^r \mathfrak{g}_{\psi_j} \right) \cap \Omega$; see [Wolf 1972; Mok 2014]. From [Mok 1989, p. 252], for any $v \in \mathfrak{m}^+ \cong T_{\mathbf{0}}(\Omega)$, there exists $k \in \mathfrak{k}$ such that $\mathrm{ad}(k) \cdot v = \sum_{j=1}^r a_j e_{\psi_j}$ with $a_j \in \mathbb{R}$ $(1 \le j \le r)$ and $a_1 \ge \cdots \ge a_r \ge 0$. Then, $\eta = \sum_{j=1}^r a_j e_{\psi_j}$ is said to be the normal form of $v$ and is uniquely determined by $v$. The cardinality of the set $\{j \in \{1, \dots, r\} : a_j \ne 0\}$ is called the rank of $v$, which is denoted by $r(v)$. For $1 \le k \le r = \mathrm{rank}(\Omega)$, we define

$$\mathcal{S}_{k,x}(\Omega) := \{[v] \in \mathbb{P}(T_x(\Omega)) : 1 \le r(v) \le k\} \subseteq \mathbb{P}(T_x(\Omega)),$$

called the $k$-th characteristic projective subvariety at $x \in \Omega$. Then, $\mathcal{S}_k(\Omega) := \bigcup_{x \in \Omega} \mathcal{S}_{k,x}(\Omega) \subset \mathbb{P}T(\Omega)$ is called the $k$-th characteristic bundle over $\Omega$. We simply call $\mathcal{S}_x(\Omega) := \mathcal{S}_{1,x}(\Omega)$ the characteristic variety at $x \in \Omega$. From [Mok 1989], $\mathcal{S}_x(\Omega) \subset \mathbb{P}(T_x(\Omega))$ is a connected complex submanifold, while $\mathcal{S}_{k,x}(\Omega) \subset \mathbb{P}(T_x(\Omega))$ is singular for $2 \le k \le r-1$ provided that $r = \text{rank}(\Omega) \ge 3$. In addition, $\mathcal{S}_{r,x}(\Omega) = \mathbb{P}(T_x(\Omega))$ for $x \in \Omega$ and we have the inclusions $\mathcal{S}_{1,x}(\Omega) \subset \cdots \subset \mathcal{S}_{r,x}(\Omega)$. Furthermore, for $r \ge 2$, $k \ge 2$ and $x \in \Omega$, we know $\mathcal{S}_{k,x}(\Omega) \subseteq \mathbb{P}(T_x(\Omega))$ is an irreducible projective subvariety because $\mathcal{S}_{k,x}(\Omega) \setminus \mathcal{S}_{k-1,x}(\Omega) = P \cdot [v]$ is an orbit for any $[v]$ such that $v \in T_x(\Omega) \setminus \{\mathbf{0}\}$ is a rank-$k$ vector, see [Mok 2002a], and $\mathcal{S}_{k,x}(\Omega) \setminus \mathcal{S}_{k-1,x}(\Omega)$ is dense in $\mathcal{S}_{k,x}(\Omega)$.

**Proposition 2.4** [Mok 1989, p. 252]. *The $k$-th characteristic bundle $\mathcal{S}_k(\Omega) \to \Omega$ is holomorphic. In addition, in terms of the Harish-Chandra embedding $\Omega \hookrightarrow \mathbb{C}^N$, $\mathcal{S}_k(\Omega)$ is parallel on $\Omega$ in the Euclidean sense*; *i.e., identifying $\mathbb{P}T(\Omega)$ with $\Omega \times \mathbb{P}^{N-1}$ using the Harish-Chandra coordinates, we have $\mathcal{S}_k(\Omega) \cong \Omega \times \mathcal{S}_{k,\mathbf{0}}(\Omega)$.*

**Remark 2.5.** For any nonzero vector $v \in T_{\mathbf{0}}(\Omega)$, we let $\mathcal{N}_v := \{\xi \in T_{\mathbf{0}}(\Omega) : R_{v\bar{v}\xi\bar{\xi}}(\Omega, g_\Omega) = 0\}$ be the null space of $v$. From [Mok 1989], the $k$-th null dimension of $\Omega$ is defined by $n_k(\Omega) := \dim_{\mathbb{C}} \mathcal{N}_v = \dim_{\mathbb{C}} \mathcal{N}_\eta$, where $\eta = \sum_{j=1}^k a_j e_{\psi_j}$ ($a_j > 0$ for $1 \le j \le k$) is the normal form of some vector $v \in T_{\mathbf{0}}(\Omega)$ of rank $k$. Here $n_k(\Omega) := \dim_{\mathbb{C}} \mathcal{N}_v$ only depends on the rank $k = r(v)$ of $v$. Then, Mok [1989] proved that $\dim_{\mathbb{C}} \mathcal{S}_k(\Omega) = 2N - n_k(\Omega) - 1$. In particular, $\mathcal{S}_{k,x}(\Omega)$ is of dimension $N - n_k(\Omega) - 1$ as an irreducible projective subvariety of $\mathbb{P}(T_x(\Omega))$ for any $x \in \Omega$. Moreover, we have $n(\Omega) := n_1(\Omega) \ge \cdots \ge n_r(\Omega) = 0$ and $n(\Omega)$ is called the null dimension of $\Omega$. From [Mok 1989], we define $p(\Omega) = \dim_{\mathbb{C}} \mathcal{S}_{\mathbf{0}}(\Omega)$. Then, we have $\dim_{\mathbb{C}} \Omega = N = p(\Omega) + n(\Omega) + 1$.

For $x \in \Omega$, under the identification $T_x(\Omega) = T_x(X_c)$, we have $\mathcal{S}_x(\Omega) = \mathscr{C}_x(X_c)$, where $\mathscr{C}_y(X_c) \subset \mathbb{P}(T_y(X_c))$ is the variety of minimal rational tangents (VMRT) of the compact dual $X_c$ of $\Omega$ at $y \in X_c$. We define $p(X_c) := \dim_{\mathbb{C}} \mathscr{C}_o(X_c)$ for the base point $o \in X_c$, which is identified with $\mathbf{0} \in \mathfrak{m}^+$, i.e., $\xi(\mathbf{0}) = o \in X_c \cong G^{\mathbb{C}}/P$. For the notion of the VMRTs of Hermitian symmetric spaces of the compact type, we refer the reader to [Hwang and Mok 1999]. Note that $\dim_{\mathbb{C}} \mathscr{C}_y(X_c)$ does not depend on the choice of $y \in X_c$. Then, we have $p(X_c) = p(\Omega) = \dim_{\mathbb{C}} \mathscr{C}_x(X_c)$ for any $x \in \Omega \subset X_c$.

**2A1.** *Holomorphic sectional curvature.* Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $r$ and $X_c$ be its compact dual Hermitian symmetric space. Recall that $g_\Omega$ is the canonical Kähler–Einstein metric on $\Omega$ normalized so that minimal disks are of constant Gaussian curvature $-2$. Then, the Bergman kernel on $\Omega$ is given by

$$K_\Omega(z, \xi) = \frac{1}{\text{Vol}(\Omega)} h_\Omega(z, \xi)^{-(p(\Omega)+2)},$$

where $\text{Vol}(\Omega)$ is the Euclidean volume of $\Omega$ in $\mathbb{C}^N$, $h_\Omega(z, \xi)$ is a polynomial in $(z, \bar{\xi})$ and $p(\Omega) := p(X_c)$ is the complex dimension of the VMRT of $X_c$ at the base point $o\in X_c$; see [Mok 2016]. For $z\in\Omega\cong G_0/K$, there exists $k\in K$ such that $k\cdot z=\sum_{j=1}^r a_j e_{\psi_j}\in\big(\bigoplus_{j=1}^r \mathfrak{g}_{\psi_j}\big)\cap\Omega=\Pi$ for $|a_j|^2<1,\ 1\le j\le r$, and

$$h_\Omega(z,z)=\prod_{j=1}^r(1-|a_j|^2),$$

where $r$ is the rank of the irreducible bounded symmetric domain $\Omega$, $\Pi\cong\Delta^r$ is a maximal polydisk in $\Omega$ which satisfies $(\Pi, g_\Omega|_\Pi)\cong\big(\Delta^r,\frac{1}{2}ds^2_{\Delta^r}\big)$; see [Mok 2014]. In particular, it follows from the polydisk theorem, see [Mok 1989, p. 88], that

$$-2\le R_{\alpha\bar\alpha\alpha\bar\alpha}(\Omega,g_\Omega)\le-\frac{2}{r}$$

for any unit vector $\alpha\in T_x(\Omega)$ and $x\in\Omega$. Let $x\in\Omega$ and $\beta\in T_x(\Omega)$ be such that $\|\beta\|^2_{g_\Omega}=1$. If $\beta$ is of rank $r(\beta)=s$, then we have $R_{\beta\bar\beta\beta\bar\beta}(\Omega,g_\Omega)\le-2/s$ because there exists $g\in G_0\cong\mathrm{Aut}_0(\Omega)$ such that $g\cdot\beta\in T_{\mathbf{0}}(\Pi_s)$ for some totally geodesic submanifold $(\Pi_s, g_\Omega|_{\Pi_s})\subset(\Pi, g_\Omega|_\Pi)$ which is holomorphically isometric to $\big(\Delta^s,\frac{1}{2}ds^2_{\Delta^s}\big)$.

## 3. On holomorphic isometries of complex unit balls into bounded symmetric domains with nonminimal isometric constants

Let $\Omega\Subset\mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\ge 2$. Mok [2016] studied the space $\widehat{\mathrm{HI}}_1(\mathbb{B}^n,\Omega)$ and provided a sharp upper bound on dimensions of isometrically embedded complex unit balls $(\mathbb{B}^n, g_{\mathbb{B}^n})$ in the irreducible bounded symmetric domain $(\Omega, g_\Omega)$ equipped with the canonical Kähler–Einstein metric $g_\Omega$. Recall that given any $f\in\widehat{\mathrm{HI}}_k(\mathbb{B}^n,\Omega)$ with $k>0$ being a real constant, $k$ is a positive integer satisfying $1\le k\le\mathrm{rank}(\Omega)$; see [CM]. It is natural to ask whether some results in Mok's study [2016] could be generalized to the study of the space $\widehat{\mathrm{HI}}_k(\mathbb{B}^n,\Omega)$ for $k\ge 2$.

In the first part of this section (see Section 3A), we provide an upper bound of $n$ whenever $\widehat{\mathrm{HI}}_k(\mathbb{B}^n,\Omega)\ne\varnothing$, where $k\ge 2$. Note that such an upper bound is not sharp in general. For instance, if $\Omega=D^{\mathrm{I}}_{p,q}$ with $q\ge p\ge 2$ and $k=\mathrm{rank}(\Omega)=p$, then $\widehat{\mathrm{HI}}_k(\mathbb{B}^n,\Omega)\ne\varnothing$ implies $n\le q/p$; see [Koziarz and Maubon 2008, Proposition 3.2]. On the other hand, our general result will imply that $n\le n_{p-1}(D^{\mathrm{I}}_{p,q})=q-p+1$ whenever $\widehat{\mathrm{HI}}_p(\mathbb{B}^n,D^{\mathrm{I}}_{p,q})\ne\varnothing$ with $q\ge p\ge 2$. In the case where $q=3$ and $p=2$, we have $n\le 2$ from our general result. But then it follows from [Koziarz and Maubon 2008, Proposition 3.2] that $n=1$ whenever $\widehat{\mathrm{HI}}_2(\mathbb{B}^n,D^{\mathrm{I}}_{2,3})\ne\varnothing$. This explains that the upper bound obtained in our general result is not sharp in general. However, one of the applications of our general result is that if $\Omega$ satisfies certain conditions and $\widehat{\mathrm{HI}}_k(\mathbb{B}^n,\Omega)\ne\varnothing$ for some fixed real constant $k>1$, then $n\le p(\Omega)$. In the second part of this section (see Section 3B), we continue our study in [CM] to the study of the space $\widehat{\mathrm{HI}}_2(\mathbb{B}^n,\Omega)$. In particular, we will obtain an analogue of [CM, Theorem 1] for holomorphic isometries in the space $\widehat{\mathrm{HI}}_2(\mathbb{B}^n, \Omega)$ without using the system of functional equations introduced in [Mok 2012].

**3A.** ***Upper bounds on dimensions of isometrically embedded complex unit balls in an irreducible bounded symmetric domain.*** Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$. Motivated by Mok's study [2016], one may continue to study the space $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ for $\lambda' > 1$. In this section, we study the upper bound on dimensions of isometrically embedded complex unit balls in an irreducible bounded symmetric domain of rank $\geq 2$ when the isometric constant is equal to $\lambda' > 1$. It is natural to ask whether the upper bound $p(\Omega) + 1$ obtained in [Mok 2016] is optimal in the sense that $n \leq p(\Omega) + 1$ whenever $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$ for some real constant $\lambda' > 0$. More specifically, we may ask whether $n \leq p(\Omega)$ whenever $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$ for some real constant $\lambda' > 1$.

For any given integer $\lambda' \geq 2$, in order to obtain a sharp upper bound of $n$ such that $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$, one should construct a holomorphic isometry $f \in \widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^{n_0}, \Omega)$ for some integer $n_0 \geq 1$ such that $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$ only if $n \leq n_0$. Note that this problem remains unsolved, but we can provide a (rough) upper bound of $n$ by using the $k$-th characteristic bundle on $\Omega$. More precisely, for any integer $\lambda'$ satisfying $2 \leq \lambda' \leq \mathrm{rank}(\Omega)$, we prove that if $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$, then $n \leq n_{\lambda'-1}(\Omega)$, where $n_k(\Omega)$ is the $k$-th null dimension of $\Omega$; see [Mok 1989]. This is precisely the assertion of Theorem 1.1. Moreover, for certain irreducible bounded symmetric domains $\Omega$ of rank $\geq 2$ (including the two irreducible bounded symmetric domains of the exceptional type) we will show that $n \leq p(\Omega)$ whenever $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$ for some integer $\lambda' \geq 2$. Now, we are ready to prove Theorem 1.1.

*Proof of Theorem 1.1.* Let $f \in \widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ be a holomorphic isometry. Write $S := f(\mathbb{B}^n)$. If $\mathbb{P}(T_y(S)) \cap \mathcal{S}_{\lambda'-1,y}(\Omega) \neq \varnothing$ for some $y \in S$, then there exists a vector $\alpha \in T_y(S) \subset T_y(\Omega)$ of unit length with respect to $g_\Omega$ and of rank $k \leq \lambda' - 1$ such that

$$R_{\alpha\bar{\alpha}\alpha\bar{\alpha}}(\Omega, g_\Omega) \leq -\frac{2}{k} \leq -\frac{2}{\lambda'-1}$$

(see Section 2A1). But then we have

$$-\frac{2}{\lambda'} = R_{\alpha\bar{\alpha}\alpha\bar{\alpha}}(S, g_\Omega|_S) \leq R_{\alpha\bar{\alpha}\alpha\bar{\alpha}}(\Omega, g_\Omega) \leq -\frac{2}{\lambda'-1}$$

from the Gauss equation, which is a contradiction. Hence, we have $\mathbb{P}(T_y(S)) \cap \mathcal{S}_{\lambda'-1,y}(\Omega) = \varnothing$ for any $y \in S$. Recall from Section 2A that $\mathcal{S}_{\lambda'-1,y}(\Omega) \subseteq \mathbb{P}(T_y(\Omega))$ is an irreducible projective subvariety of complex dimension $N - n_{\lambda'-1}(\Omega) - 1$. Then, it follows from the inequality

$$\dim_{\mathbb{C}}\big(\mathbb{P}(T_y(S)) \cap \mathcal{S}_{\lambda'-1,y}(\Omega)\big) \geq \dim_{\mathbb{C}} \mathbb{P}(T_y(S)) + \dim_{\mathbb{C}} \mathcal{S}_{\lambda'-1,y}(\Omega) - \dim_{\mathbb{C}} \mathbb{P}(T_y(\Omega))$$

that $n \leq n_{\lambda'-1}(\Omega)$; see [Mumford 1976, p. 57]. □

**Lemma 3.1.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$. Then, $n(\Omega) \leq p(\Omega)$ if and only if $\Omega$ is biholomorphic to one of the following*:

(1) $D^{\mathrm{I}}_{p',q'}$, *where $p'$ and $q'$ are integers satisfying $2 = p' < q'$ or $p' = q' = 3$.*

(2) $D^{\mathrm{II}}_m$ *for some integer $m$ satisfying $5 \leq m \leq 7$.*

(3) $D^{\mathrm{IV}}_n$ *for some integer $n \geq 3$.*

(4) $D^{\mathrm{V}}$.

(5) $D^{\mathrm{VI}}$.

*Proof.* From [Mok 1989, pp. 105–106], we have $n(\Omega) + p(\Omega) + 1 = N$. Then, the result follows from direct computations by the explicit data provided in [Mok 1989, pp. 249–251]. □

**Remark 3.2.** We observe that if $\Omega$ satisfies $n(\Omega) \leq p(\Omega)$, then $\mathrm{rank}(\Omega) \leq 3$. In addition, Lemma 3.1 implies that any irreducible bounded symmetric domain $\Omega$ of rank 2 satisfies $n(\Omega) \leq p(\Omega)$. From [Mok 1989], it is clear that the condition $n(\Omega) \leq p(\Omega)$ is equivalent to $\dim_{\mathbb{C}} \mathbb{P}(T_o(X_c)) \leq 2 \cdot \dim_{\mathbb{C}} \mathscr{C}_o(X_c)$, where $X_c$ is the compact dual Hermitian symmetric space of $\Omega$ and $o \in X_c$ is a fixed base point.

The following corollary shows that for certain irreducible bounded symmetric domains $\Omega$ of rank $\geq 2$ and a fixed real constant $\lambda' > 0$, we have $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^{p(\Omega)+1}, \Omega) \neq \varnothing$ only if $\lambda' = 1$. On the other hand, Mok [2016, Main Theorem] proved that $\widehat{\mathrm{HI}}_1(\mathbb{B}^{p(\Omega)+1}, \Omega) \neq \varnothing$ for any irreducible bounded symmetric domain $\Omega$ of rank $\geq 2$. Therefore, combining with [Mok 2016, Main Theorem], we actually have $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^{p(\Omega)+1}, \Omega) \neq \varnothing$ if and only if $\lambda' = 1$ for certain irreducible bounded symmetric domains $\Omega$ of rank $\geq 2$.

**Corollary 3.3.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$ such that $n(\Omega) \leq p(\Omega)$. If $f \in \widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ for some real constant $\lambda' \geq 2$, then $n \leq p(\Omega)$.*

*Proof.* Note that $\lambda'$ is an integer satisfying $2 \leq \lambda' \leq \mathrm{rank}(\Omega)$. By the assumption, it follows from Theorem 1.1 that $n \leq n_{\lambda'-1}(\Omega) \leq n(\Omega) \leq p(\Omega)$. □

**Remark 3.4.** Actually, Corollary 3.3 together with [Mok 2016, Main Theorem] implies that the upper bound $p(\Omega) + 1$ is optimal when the bounded symmetric domain $\Omega$ satisfies $n(\Omega) \leq p(\Omega)$. Moreover, the statement of Corollary 3.3 holds true for any irreducible bounded symmetric domain $\Omega$ of rank 2.

**3A1.** *Holomorphic isometries with the maximal isometric constant and applications.* Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $r \geq 2$. Recall that if $f \in \widehat{\mathrm{HI}}_r(\mathbb{B}^n, \Omega)$, then $f$ is totally geodesic by the Ahlfors–Schwarz lemma. The results obtained in Section 3A can be applied so that we may prove $n \leq p(\Omega)$ without using the total geodesy of holomorphic isometries lying in the space $\widehat{\mathrm{HI}}_r(\mathbb{B}^n, \Omega)$.

**Proposition 3.5.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $r \geq 2$ such that $\Omega \not\cong D_3^{\mathrm{IV}}$ and let $f \in \widehat{\mathrm{HI}}_r(\mathbb{B}^n, \Omega)$. Then, we have $n < p(\Omega)$. If $F \in \widehat{\mathrm{HI}}_r(\mathbb{B}^n, \Omega)$, where $\Omega$ is an irreducible bounded symmetric domain of rank $r \geq 2$ and of tube type, then we have $n = 1$.*

*Proof.* Under the assumptions, Theorem 1.1 asserts that $n \leq n_{r-1}(\Omega)$, so it remains to check that $n_{r-1}(\Omega) < p(\Omega)$ for any irreducible bounded symmetric domain $\Omega$ of rank $r \geq 2$ and $\Omega \not\cong D_3^{\mathrm{IV}}$. Note that if $\Omega \cong D_3^{\mathrm{IV}}$, then $r = 2$ and $n_{r-1}(\Omega) = 1 = p(\Omega)$. It follows from [Mok 2002a] that $\Omega$ is of tube type if and only if $n_{r-1}(\Omega) = 1$ due to the dimension formula $\dim_{\mathbb{C}} \mathcal{S}_{r-1,x}(\Omega) = \dim_{\mathbb{C}} \mathbb{P}(T_x(\Omega)) - n_{r-1}(\Omega)$ of [Mok 1989]. It is clear that if $\Omega$ is of tube type and $\Omega \not\cong D_3^{\mathrm{IV}}$, then $p(\Omega) > 1$ so that $n_{r-1}(\Omega) = 1 < p(\Omega)$. If $\Omega$ is not of tube type, then $\Omega$ is biholomorphic to one of the following:

(1) $D^{\mathrm{I}}_{p',q'}$ for some integers $p', q'$ satisfying $2 \leq p' < q'$.
(2) $D^{\mathrm{II}}_{2m+1}$ for some integer $m \geq 2$.
(3) $D^{\mathrm{V}}$.

From the classification of the boundary components of bounded symmetric domains and the fact that $n_{r-1}(\Omega)$ is precisely the dimension of rank-1 boundary components of $\Omega$, see [Wolf 1972; Mok 2002a, p. 298], we have

$$\begin{aligned} n_{p'-1}(D^{\mathrm{I}}_{p',q'}) &= q' - p' + 1 < p(D^{\mathrm{I}}_{p',q'}) = p' + q' - 2 \quad && \text{for } 2 \leq p' < q', \\ n_{m-1}(D^{\mathrm{II}}_{2m+1}) &= 3 < p(D^{\mathrm{II}}_{2m+1}) = 2(2m-1) && \text{for } m \geq 2, \\ n_1(D^{\mathrm{V}}) &= 5 < p(D^{\mathrm{V}}) = 10. \end{aligned}$$

Hence, we have $n < p(\Omega)$. On the other hand, given an irreducible bounded symmetric domain $\Omega$ of rank $r \geq 2$ and of tube type, if $F \in \widehat{\mathrm{HI}}_r(\mathbb{B}^n, \Omega)$, then we have $n \leq n_{r-1}(\Omega) = 1$, i.e., $n = 1$. □

From the proof of Proposition 3.5, we have $n_{r-1}(\Omega) \leq p(\Omega)$ for any irreducible bounded symmetric domain $\Omega$ of rank $r \geq 2$. Given any irreducible bounded symmetric domain $\Omega$ of rank $r \geq 2$, we define

$$\lambda_0(\Omega) := \min\{\lambda \in \mathbb{Z} : 1 \leq \lambda \leq r,\ n_\lambda(\Omega) \leq p(\Omega)\}.$$

Then, we have $\lambda_0(\Omega) \leq r - 1$. Note that $\Omega$ satisfies $n(\Omega) \leq p(\Omega)$ if and only if $\lambda_0(\Omega) = 1$. Combining with Corollary 3.3, we have the following:

**Theorem 3.6.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $r \geq 2$ and $\lambda' \geq 2$ be an integer. If $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega) \neq \varnothing$, then $n \leq p(\Omega)$ provided that one of the following holds true:*

(1) *$\Omega$ satisfies $n(\Omega) \leq p(\Omega)$.*
(2) *$\lambda'$ satisfies $\lambda_0(\Omega) + 1 \leq \lambda' \leq r$.*

*Proof.* If the bounded symmetric domain $\Omega$ satisfies $n(\Omega) \le p(\Omega)$, then the result follows from Corollary 3.3. If $\lambda'$ satisfies $\lambda_0(\Omega) + 1 \le \lambda' \le r$, then we have $n_{\lambda'-1}(\Omega) \le n_{\lambda_0(\Omega)}(\Omega) \le p(\Omega)$. By Theorem 1.1, we have $n \le n_{\lambda'-1}(\Omega) \le p(\Omega)$. $\square$

**Remark 3.7.** If $\Omega$ satisfies $n(\Omega) \le p(\Omega)$, then $\lambda_0(\Omega) = 1$ so that the condition (2) does not provide an additional restriction on the given isometric constant $\lambda'$.

In general, let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\ge 2$ such that $n(\Omega) > p(\Omega)$. Then, Lemma 3.1 asserts that $\Omega$ is biholomorphic to one of the following:

(1) $D^{\mathrm{I}}_{p,q}$ for some integers $p, q$ satisfying $3 \le p \le q$ and $(p, q) \ne (3, 3)$.

(2) $D^{\mathrm{II}}_m$ for some integer $m \ge 8$.

(3) $D^{\mathrm{III}}_m$ for some integer $m \ge 3$.

In particular, we are able to compute $\lambda_0(\Omega)$ explicitly for each case.

| type | $\Omega$ | $\lambda_0(\Omega)$ |
|---|---|---|
| $\mathrm{I}_{p,q}$ $(3 \le p \le q,\ (p,q) \ne (3,3))$ | $D^{\mathrm{I}}_{p,q}$ | $\lceil \frac{1}{2}((p+q)-\sqrt{(q-p)^2+4(p+q-2)}) \rceil$ |
| $\mathrm{II}_m$ $(m \ge 8)$ | $D^{\mathrm{II}}_m$ | $\lceil \frac{1}{4}((2m-1)-\sqrt{16m-31}) \rceil$ |
| $\mathrm{III}_m$ $(m \ge 3)$ | $D^{\mathrm{III}}_m$ | $\lceil \frac{1}{2}((2m+1)-\sqrt{8m-7}) \rceil$ |

Here $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x$ for any real number $x$.

**Example 3.8.** If $\Omega = D^{\mathrm{III}}_7$, then $\Omega$ is of rank 7, $n_k(\Omega) = \frac{1}{2}(7-k)(7-k+1)$ and $p(\Omega) = 6$, see [Mok 1989, p. 86, p. 250], so that $\lambda_0(\Omega) = 4 = \mathrm{rank}(\Omega) - 3$. Given any integer $\lambda'$ satisfying $5 \le \lambda' \le 7$, Theorem 3.6 asserts that $n \le p(\Omega) = 6$ whenever $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, D^{\mathrm{III}}_7) \ne \varnothing$.

In general, by using the expression of $\lambda_0(D^{\mathrm{III}}_{m+2})$ in terms of $m$ for any integer $m \ge 1$ (see the table above), one observes that the sequence

$$\{\mathrm{rank}(D^{\mathrm{III}}_{m+2}) - (\lambda_0(D^{\mathrm{III}}_{m+2}) + 1)\}_{m=1}^{+\infty}$$

is monotonic increasing and $a_m := \mathrm{rank}(D^{\mathrm{III}}_{m+2}) - (\lambda_0(D^{\mathrm{III}}_{m+2}) + 1) \to +\infty$ as $m \to +\infty$. Moreover, $a_m/\mathrm{rank}(D^{\mathrm{III}}_{m+2}) \to 0$ as $m \to +\infty$. That means $\mathrm{rank}(D^{\mathrm{III}}_{m+2})$ grows much faster than $a_m$ as $m$ is increasing. This shows that in general the range of the isometric constants $\lambda'$ mentioned in condition (2) of Theorem 3.6 is quite restrictive for a rank-$r$ irreducible bounded symmetric domain $\Omega$, $r \ge 2$, such that $n(\Omega) > p(\Omega)$.

**3B.** ***Holomorphic isometries with the isometric constant equal to 2 and applications.*** Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\geq 2$ and $X_c$ be the compact dual Hermitian symmetric space of $\Omega$. Then, it follows from the observation in Section 2 that the polynomial $h_\Omega(z, z)$ can be written as

$$h_\Omega(z,z) = 1 - \sum_{l=1}^{m_1(\Omega)} |G_l^{(1)}(z)|^2 + \sum_{l'=1}^{m_2(\Omega)} |G_{l'}^{(2)}(z)|^2,$$

where $G_l^{(1)}(z)$, $G_{l'}^{(2)}(z)$ are homogeneous polynomials in $z$ and $m_1(\Omega)$, $m_2(\Omega)$ are positive integers depending on $\Omega$ such that

(1) $m_1(\Omega) + m_2(\Omega) = N'$ and $m_1(\Omega) \geq N$,

(2) $\deg(G_l^{(1)})$ $(1 \leq l \leq m_1(\Omega))$ is odd, while $\deg(G_{l'}^{(2)}) \geq 2$ $(1 \leq l' \leq m_2(\Omega))$ is even,

(3) $G_j^{(1)}(z) = z_j$ for $1 \leq j \leq N$,

(4) when $\Omega$ is of rank $\geq 3$, we have $m_1(\Omega) > N$ and $\deg(G_l^{(1)}) \geq 3$ for $N + 1 \leq l \leq m_1(\Omega)$.

Moreover, in terms of the Harish-Chandra coordinates $z = (z_1, \dots, z_N) \in \mathbb{C}^N$, the restriction of $\iota$ to the dense open subset $\mathbb{C}^N \subset X_c$ may be written as

$$\iota(z_1, \dots, z_N) = [1, G_1^{(1)}(z), \dots, G_{m_1(\Omega)}^{(1)}(z), G_1^{(2)}(z), \dots, G_{m_2(\Omega)}^{(2)}(z)]$$

up to reparametrizations, where $\iota : X_c \hookrightarrow \mathbb{P}^{N'}$ is the minimal embedding.

**Remark 3.9.** As mentioned in Section 2, the above observation can be obtained from [Loos 1977] and has been written down explicitly by Fang, Huang and Xiao [Fang et al. 2016].

In [CM], we studied images of holomorphic isometries in $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ when $\lambda' = 1$. However, it is not obvious how the method in [CM] could be generalized to the study of images of holomorphic isometries in $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ for $\lambda' > 1$ so as to obtain an analogue of Theorem 1 in [CM] for all holomorphic isometries in $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ and for any $\lambda' > 0$. After that, we observe that the above explicit form of $h_\Omega(z, z)$ is useful for continuing the study of images of holomorphic isometries in $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ when the isometric constant $\lambda'$ equals 2. Recall that the case where $\lambda' = 2$ in Theorem 1.2 is exactly an analogue of Theorem 1 in [CM] for all holomorphic isometries in $\widehat{\mathrm{HI}}_2(\mathbb{B}^n, \Omega)$. We are now ready to prove Theorem 1.2 for the case where $\lambda' = 2$.

*Proof of Theorem 1.2 for the case where $\lambda' = 2$.* Let $f : (\mathbb{B}^n, 2g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ be a holomorphic isometric embedding, where $\Omega \Subset \mathbb{C}^N$ is an irreducible bounded symmetric domain of rank $\geq 2$. Assume without loss of generality that $f(\mathbf{0}) = \mathbf{0}$.

Then, we have the functional equation

$$1-\sum_{l=1}^{m_1(\Omega)}|G_l^{(1)}(f(w))|^2+\sum_{l=1}^{m_2(\Omega)}|G_l^{(2)}(f(w))|^2 = \Big(1-\sum_{\mu=1}^{n}|w_\mu|^2\Big)^2 = 1-\sum_{\mu=1}^{n}|\sqrt{2}w_\mu|^2+\sum_{1\le\mu,\mu'\le n}|w_\mu w_{\mu'}|^2 \tag{3-1}$$

for $w\in\mathbb{B}^n$ and the polarized functional equation

$$1-\sum_{l=1}^{m_1(\Omega)}G_l^{(1)}(f(w))\overline{G_l^{(1)}(f(\zeta))}+\sum_{l=1}^{m_2(\Omega)}G_l^{(2)}(f(w))\overline{G_l^{(2)}(f(\zeta))} = \Big(1-\sum_{\mu=1}^{n}w_\mu\bar{\zeta}_\mu\Big)^2 \tag{3-2}$$

for $w,\zeta\in\mathbb{B}^n$; see equation (14) in [CM, p. 688]. We write

$$\sum_{1\le\mu,\mu'\le n}|w_\mu w_{\mu'}|^2=\sum_{l=1}^{m_0}|\Xi_l(w)|^2$$

for some homogeneous polynomials $\Xi_l(w)$ of degree 2 and $m_0:=\frac{1}{2}n(n+1)$. Moreover, we write $\boldsymbol{G}^{(j)}(z)=(G_1^{(j)}(z),\dots,G_{m_j(\Omega)}^{(j)}(z))^T$ for $j=1,2$. Let $N_0:=\max\{n+m_2(\Omega),m_0+m_1(\Omega)\}$. Then, there exists $\boldsymbol{U}\in U(N_0)$ such that

$$\boldsymbol{U}\cdot\begin{pmatrix}\sqrt{2}w_1\\ \vdots\\ \sqrt{2}w_n\\ \boldsymbol{G}^{(2)}(f(w))\\ \boldsymbol{0}_{(N_0-n-m_2(\Omega))\times 1}\end{pmatrix}=\begin{pmatrix}\Xi_1(w)\\ \vdots\\ \Xi_{m_0}(w)\\ \boldsymbol{G}^{(1)}(f(w))\\ \boldsymbol{0}_{(N_0-m_1(\Omega)-m_0)\times 1}\end{pmatrix} \tag{3-3}$$

by Lemma 2.1 and (3-1). We write

$$\boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1\\ \boldsymbol{U}_2\end{bmatrix}$$

with $\boldsymbol{U}_1\in M(m_0,N_0;\mathbb{C})$ and $\boldsymbol{U}_2\in M(N_0-m_0,N_0;\mathbb{C})$. We also write $\boldsymbol{U}_2=\begin{bmatrix}\boldsymbol{U}_{21} & \boldsymbol{U}_{22}\end{bmatrix}$ with $\boldsymbol{U}_{21}\in M(N_0-m_0,n;\mathbb{C})$ and $\boldsymbol{U}_{22}\in M(N_0-m_0,N_0-n;\mathbb{C})$. Denote by $(Jf)(w)$ the complex Jacobian matrix of the holomorphic map $f:\mathbb{B}^n\to\Omega\Subset\mathbb{C}^N$ at $w\in\mathbb{B}^n$. Recall that $G_j^{(1)}(z)=z_j$ for $1\le j\le N$, $G_l^{(2)}(z)$, $1\le l\le m_2(\Omega)$, are homogeneous polynomials of degree $\ge 2$ in $z$ so that $\frac{\partial}{\partial z_j}G_l^{(2)}(z)|_{z=\boldsymbol{0}}=0$ for $1\le j\le N$, $1\le l\le m_2(\Omega)$. In addition, if the rank of $\Omega$ is at least 3 so that $m_1(\Omega)>N$, then $G_l^{(1)}(z)$, $N+1\le l\le m_1(\Omega)$, are homogeneous polynomials of degree $\ge 3$ in $z$, so that $\frac{\partial}{\partial z_j}G_l^{(1)}(z)|_{z=\boldsymbol{0}}=0$ for $1\le j\le N$, $N+1\le l\le m_1(\Omega)$. Then, we have

$$\overline{(Jf)(\boldsymbol{0})}^T(f^1(w),\dots,f^N(w))^T=2(w_1,\dots,w_n)^T \tag{3-4}$$

by differentiating both sides of (3-2) with respect to $\bar{\zeta}_\mu$ at $\zeta = \mathbf{0}$ for each $\mu$, $1 \le \mu \le n$. In addition, $(Jf)(\mathbf{0}) \in M(N, n; \mathbb{C})$ is of rank $n$. Moreover, from the above settings and (3-3) we have

$$\boldsymbol{U}_{21}\begin{pmatrix}\sqrt{2}w_1\\ \vdots\\ \sqrt{2}w_n\end{pmatrix} + \boldsymbol{U}_{22}\begin{pmatrix}\boldsymbol{G}^{(2)}(f(w))\\ \mathbf{0}_{(N_0-n-m_2(\Omega))\times 1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G}^{(1)}(f(w))\\ \mathbf{0}_{(N_0-m_1(\Omega)-m_0)\times 1}\end{pmatrix}. \tag{3-5}$$

Differentiating both sides of (3-5) with respect to $w_\mu$ at $w = \mathbf{0}$ for each $\mu$, $1 \le \mu \le n$, we obtain

$$\sqrt{2}\boldsymbol{U}_{21} = \begin{pmatrix}(Jf)(\mathbf{0})\\ \mathbf{0}_{(N_0-m_0-N)\times n}\end{pmatrix}.$$

In addition, by differentiating both sides of (3-4) with respect to $w_\mu$ at $w = \mathbf{0}$ for each $\mu$, $1 \le \mu \le n$, we have $\overline{(Jf)(\mathbf{0})}^T (Jf)(\mathbf{0}) = 2\boldsymbol{I}_n$. Therefore, it follows from (3-5) and (3-4) that

$$\left[\begin{bmatrix}\frac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T\\ \mathbf{0}_{(N_0-m_0-N)\times N}\end{bmatrix} \;\boldsymbol{U}_{22}\right]\begin{pmatrix}\boldsymbol{f}(w)\\ \boldsymbol{G}^{(2)}(f(w))\\ \mathbf{0}_{(N_0-n-m_2(\Omega))\times 1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G}^{(1)}(f(w))\\ \mathbf{0}_{(N_0-m_0-m_1(\Omega))\times 1}\end{pmatrix} \tag{3-6}$$

for any $w \in \mathbb{B}^n$, where $\boldsymbol{f}(w) := (f^1(w), \dots, f^N(w))^T$. Writing $\boldsymbol{B} := \left[\widehat{\boldsymbol{U}}_{21} \;\; \boldsymbol{U}_{22}\right]$ with

$$\widehat{\boldsymbol{U}}_{21} = \begin{bmatrix}\frac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T\\ \mathbf{0}_{(N_0-m_0-N)\times N}\end{bmatrix},$$

we define

$$\mathscr{V}' := \left\{ z \in \mathbb{C}^N : \boldsymbol{B}\begin{pmatrix}z^T\\ \boldsymbol{G}^{(2)}(z)\\ \mathbf{0}_{(N_0-n-m_2(\Omega))\times 1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G}^{(1)}(z)\\ \mathbf{0}_{(N_0-m_0-m_1(\Omega))\times 1}\end{pmatrix}\right\} \tag{3-7}$$

and $\mathscr{V} := \mathscr{V}' \cap \Omega$. Then, we have $f(\mathbb{B}^n) \subseteq \mathscr{V}$ by (3-6). Note that the tangential dimension $\mathrm{tdim}_{\mathbf{0}}\, \mathscr{V}$ of $\mathscr{V}$ at $\mathbf{0}$ is less than or equal to $N - \mathrm{rank}\big(\frac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big)$. Here we refer the readers to [Gunning 1990] for the notion of the tangential dimension $\mathrm{tdim}_x\, V$ of a complex-analytic variety $V$ at a point $x \in V$. From [Zhang 1999, p. 49], we have

$$\mathrm{rank}\big(\tfrac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big) \ge \big|\mathrm{rank}\big(\tfrac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T\big) - \mathrm{rank}\ \boldsymbol{I}_N\big| = N - n.$$

On the other hand, $\big(\frac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big) \cdot (Jf)(\mathbf{0}) = \mathbf{0}$ so that

$$0 \ge \mathrm{rank}\big(\tfrac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big) + \mathrm{rank}(Jf)(\mathbf{0}) - N$$

and thus $\mathrm{rank}\big(\frac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big) \le N - n$. Therefore, we have

$$\mathrm{rank}\big(\tfrac{1}{2}(Jf)(\mathbf{0})\overline{(Jf)(\mathbf{0})}^T - \boldsymbol{I}_N\big) = N - n.$$

Moreover, $\mathscr{V}$ contains $f(\mathbb{B}^n)$ and $\mathbf{0} \in f(\mathbb{B}^n)$, thus $\dim_{\mathbf{0}} \mathscr{V} \geq n \geq \operatorname{tdim}_{\mathbf{0}} \mathscr{V}$. Note that $\dim_{\mathbf{0}} \mathscr{V} \leq \operatorname{tdim}_{\mathbf{0}} \mathscr{V}$; see [Gunning 1990]. Hence, we have $\dim_{\mathbf{0}} \mathscr{V} = \operatorname{tdim}_{\mathbf{0}} \mathscr{V} = n$ and thus $\mathscr{V}$ is smooth at $\mathbf{0}$. Let $S$ be the irreducible component of $\mathscr{V}$ containing $f(\mathbb{B}^n)$. Then, we have $\dim S = n = \dim f(\mathbb{B}^n)$ and thus $S = f(\mathbb{B}^n)$ because both $S$ and $f(\mathbb{B}^n)$ are irreducible complex-analytic subvarieties of $\mathscr{V}$ containing the smooth point $\mathbf{0} \in \mathscr{V}$ of $\mathscr{V}$. In particular, $f(\mathbb{B}^n)$ is the irreducible component of $\mathscr{V}$ containing $\mathbf{0}$. Moreover, it is clear that $\mathscr{V}' \subset \mathbb{C}^N$ is an affine-algebraic subvariety and $\iota(\mathscr{V}) = P \cap \iota(\Omega)$, where

$$P := \left\{[\xi_0, \xi_1, \dots, \xi_{N'}] \in \mathbb{P}^{N'} : \boldsymbol{B}\boldsymbol{x} = \boldsymbol{y}\right\}, \tag{3-8}$$

with

$$\boldsymbol{x} = \left(\xi_1, \dots, \xi_N, \xi_{m_1(\Omega)+1}, \dots, \xi_{N'}, \mathbf{0}_{1\times(N_0-n-m_2(\Omega))}\right)^T,$$
$$\boldsymbol{y} = \left(\xi_1, \dots, \xi_{m_1(\Omega)}, \mathbf{0}_{1\times(N_0-m_0-m_1(\Omega))}\right)^T,$$

is a projective linear subspace of $\mathbb{P}^{N'}$. □

**3B1.** *On holomorphic isometries from the Poincaré disk into polydisks.* The author [Chan 2016] and Ng [2010] studied the classification problem of all holomorphic isometries from the Poincaré disk into the $p$-disk with any isometric constant $k$, $1 \leq k \leq p$, and $p \geq 2$. The classification problem remains unsolved when $p \geq 5$. In this section, we consider the structure of images of such holomorphic isometries for $k \leq 2$ and obtain an analogue of Theorem 1.2 when the domain is the Poincaré disk and the target is the $p$-disk for some $p \geq 2$.

Note that the restriction $\varrho$ of the Segre embedding $\varsigma : (\mathbb{P}^1)^p \hookrightarrow \mathbb{P}^{2^p-1}$ to the dense open subset $\mathbb{C}^p \subset (\mathbb{P}^1)^p$ is given by

$$\varrho(z_1, \dots, z_p) = \varsigma([1, z_1], \dots, [1, z_p])$$

in terms of the standard holomorphic coordinates $z = (z_1, \dots, z_p) \in \mathbb{C}^p$. Here $\mathbb{C}^p$ is identified with its image $\xi(\mathbb{C}^p)$ in $(\mathbb{P}^1)^p$, where the map $\xi : \mathbb{C}^p \hookrightarrow (\mathbb{P}^1)^p$ is defined by $\xi(z_1, \dots, z_p) := ([1, z_1], \dots, [1, z_p])$.

Actually, the author [Chan 2016] observed that the following can be proved by the same method as the proof of Theorem 1 in [CM].

**Proposition 3.10** [Chan 2016, Proposition 5.2.4]. *Let $f : (\Delta, ds^2_\Delta) \to (\Delta^p, ds^2_{\Delta^p})$ be a holomorphic isometric embedding, where $p \geq 2$ is an integer. Then, $f(\Delta)$ is an irreducible component of $\mathscr{V} \cap \Delta^p$ for some affine-algebraic subvariety $\mathscr{V} \subset \mathbb{C}^p$ such that $\varrho(\mathscr{V} \cap \Delta^p) = \varrho(\Delta^p) \cap P$, where $P \subseteq \mathbb{P}^{2^p-1}$ is a projective linear subspace.*

Similarly, we observe that the method in the proof of Theorem 1.2 is also valid for any holomorphic isometry from $(\Delta, 2ds^2_\Delta)$ to $(\Delta^p, ds^2_{\Delta^p})$, where $p \geq 2$.

**Proposition 3.11.** *Let $f : (\Delta, 2ds^2_\Delta) \to (\Delta^p, ds^2_{\Delta^p})$ be a holomorphic isometric embedding, where $p \geq 2$ is an integer. Then, $f(\Delta)$ is an irreducible component* *of $\mathscr{V}\cap\Delta^p$ for some affine-algebraic subvariety $\mathscr{V}\subset\mathbb{C}^p$ such that $\varrho(\mathscr{V}\cap\Delta^p)=\varrho(\Delta^p)\cap P$, where $P\subseteq\mathbb{P}^{2^p-1}$ is a projective linear subspace.*

*Proof.* Assume without loss of generality that $f(\mathbf{0})=\mathbf{0}$. Note that

$$\begin{aligned}h_{\Delta^p}(z,z)&=\prod_{j=1}^{p}(1-|z_j|^2)\\&=1-\sum_{n=1}^{\lfloor(p+1)/2\rfloor}\sum_{1\le i_1<\cdots<i_{2n-1}\le p}|z_{i_1}\cdots z_{i_{2n-1}}|^2+\sum_{n=1}^{\lfloor p/2\rfloor}\sum_{1\le j_1<\cdots<j_{2n}\le p}|z_{j_1}\cdots z_{j_{2n}}|^2.\end{aligned}$$

In the proof of Theorem 1.2, we put $n=1$ and replace the term $\sum_{l=1}^{m_1(\Omega)}|G_l^{(1)}(z)|^2$ (resp. $\sum_{l=1}^{m_2(\Omega)}|G_l^{(2)}(z)|^2$) by

$$\sum_{n=1}^{\lfloor(p+1)/2\rfloor}\sum_{1\le i_1<\cdots<i_{2n-1}\le p}\left|\prod_{\mu=1}^{2n-1}z_{i_\mu}\right|^2\quad\left(\text{resp. }\sum_{n=1}^{\lfloor p/2\rfloor}\sum_{1\le j_1<\cdots<j_{2n}\le p}\left|\prod_{\mu=1}^{2n}z_{j_\mu}\right|^2\right).\tag{3-9}$$

Indeed, we may define $m_1(\Delta^p)$ and $m_2(\Delta^p)$. Then, we compute $m_1(\Delta^p)=m_2(\Delta^p)+1=2^{p-1}$. In this situation, the integer $N_0$ defined in the proof of Theorem 1.2 is equal to $m_1(\Delta^p)+1=2^{p-1}+1$. Then, the result follows directly from the arguments in the proof of Theorem 1.2. □

**3B2.** *On holomorphic isometries of complex unit balls into irreducible bounded symmetric domains of rank at most* 3. Given an irreducible bounded symmetric domain $\Omega\Subset\mathbb{C}^N$ of rank $\ge 2$, it is natural to ask whether all holomorphic isometries in $\widehat{\mathrm{HI}}(\mathbb{B}^n,\Omega)$ arise from linear sections of the minimal embedding of the compact dual $X_c$ of $\Omega$ in general. In [CM], we showed that the answer is affirmative for all holomorphic isometries in $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n,\Omega)$ whenever $\widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n,\Omega)\ne\varnothing$ and $\lambda'\in\{1,\mathrm{rank}(\Omega)\}$. On the other hand, Theorem 1.2 asserts that the answer is also affirmative for all holomorphic isometries in $\widehat{\mathrm{HI}}_2(\mathbb{B}^n,\Omega)$ whenever $\widehat{\mathrm{HI}}_2(\mathbb{B}^n,\Omega)\ne\varnothing$. In other words, we may prove Theorem 1.2 for the case where $2\le\mathrm{rank}(\Omega)\le 3$ as follows.

*Proof of Theorem 1.2 for the case where* $2\le\mathrm{rank}(\Omega)\le 3$. Recall that $\lambda'$ is an integer satisfying $1\le\lambda'\le r$; see [CM, Lemma 3]. If $r=2$, then $\lambda'=1$ or $\lambda'=2$. In the case of $\lambda'=1$, the result follows from [CM, Theorem 1]. When $\lambda'=2$, we may suppose that $f(\mathbf{0})=\mathbf{0}$. Then, $f$ is totally geodesic by [CM, Proposition 1] and $f(\mathbb{B}^n)$ is indeed an affine linear section of $\Omega$ in $\mathbb{C}^N$; see [Mok 2012]. Therefore, the result follows when $r=2$. Now, we suppose that $r=3$. If $\lambda'=1$ or $\lambda'=3$, then the result follows from Proposition 1 and Theorem 1 in [CM]. If $\lambda'=2$, then the result follows from the proof of Theorem 1.2 for the case where $\lambda'=2$. □

The proof of Theorem 1.2 is complete.

**Remark 3.12.** In general, we expect that Theorem 1 in [CM] holds true for any holomorphic isometry from $(\mathbb{B}^n, kg_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ for $1 \le k \le \mathrm{rank}(\Omega)$. Actually, the case where $2 \le \mathrm{rank}(\Omega) \le 3$ in Theorem 1.2 asserts that our expectation is true when $\Omega$ is an irreducible bounded symmetric domain of rank at most 3. Moreover, the statement of Theorem 1.2 for the case where $2 \le \mathrm{rank}(\Omega) \le 3$ also holds true for any holomorphic isometry from $(\Delta, kds_\Delta^2)$ to $(\Delta^p, ds_{\Delta^p}^2)$ for any positive integer $k$ and any integer $p$ such that $2 \le p \le 3$. However, for $2 \le p \le 3$ one may make use of Ng's classification of all holomorphic isometries from $(\Delta, kds_\Delta^2)$ to $(\Delta^p, ds_{\Delta^p}^2)$, see [Ng 2010], to prove such an analogue of Theorem 1.2 for the case where $2 \le \mathrm{rank}(\Omega) \le 3$.

On the other hand, when $\Omega \Subset \mathbb{C}^N$ is an irreducible bounded symmetric domain of rank $r \ge 4$, it is not known whether all holomorphic isometries in $\widehat{\mathrm{HI}}_k(\mathbb{B}^n, \Omega)$ arise from linear sections of the minimal embedding of the compact dual $X_c$ of $\Omega$ for $3 \le k \le r-1$. In other words, the problem remains open for the space $\widehat{\mathrm{HI}}_k(\mathbb{B}^n, \Omega)$ when $\Omega$ is of rank $r \ge 4$ and $3 \le k \le r-1$.

Now, we would like to emphasize the following consequence of both Theorem 3.6 and Theorem 1.2.

**Corollary 3.13.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank $\ge 2$ such that $n(\Omega) \le p(\Omega)$. If $f \in \widehat{\mathrm{HI}}_{\lambda'}(\mathbb{B}^n, \Omega)$ for some real constant $\lambda' > 0$, then we have the following*:

(1) $n \le p(\Omega)$ *when* $\lambda' \ge 2$; $n \le p(\Omega) + 1$ *when* $\lambda' = 1$.

(2) $f(\mathbb{B}^n)$ *is an irreducible component of some complex-analytic subvariety $\mathscr{V} \subset \Omega$ satisfying $\iota(\mathscr{V}) = P \cap \iota(\Omega)$, where $\iota : X_c \hookrightarrow \mathbb{P}^{N'}$ is the minimal embedding and $P \subseteq \mathbb{P}^{N'}$ is some projective linear subspace.*

*Proof.* Note that (1) follows from Theorem 3.6 when $\lambda' \ge 2$. On the other hand, (1) follows from Theorem 2 in [Mok 2016] when $\lambda' = 1$. Moreover, (2) follows from Theorem 1.2 because $\Omega$ is of rank at most 3 whenever $\Omega$ satisfies $n(\Omega) \le p(\Omega)$. □

**Remark 3.14.** (1) In particular, Corollary 3.13 holds true when $\Omega$ is either of type IV or of the exceptional type by Lemma 3.1. From the method used in this section, it is not known whether both parts (1) and (2) of Corollary 3.13 still hold true in general when the assumption $n(\Omega) \le p(\Omega)$ is removed.

(2) Recently, Yuan (personal communication, 2017) pointed out to the author that one may obtain upper bounds on dimensions of isometrically embedded complex unit balls into irreducible bounded symmetric domains $\Omega$ of rank $\ge 2$ by using the functional equation for any holomorphic isometry $f : (\mathbb{B}^n, kg_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$, $k \ge 2$, with $f(\mathbf{0}) = \mathbf{0}$ and the signature of the sum of squares; see [Xiao and Yuan 2016, Proposition 2.11]. When the target is $\Omega = D^{\mathrm{I}}_{3,4}$, it suffices to consider the case where $k = 2$ and we compute $m_2(D^{\mathrm{I}}_{3,4}) = \binom{3}{2}\binom{4}{2} = 18$ by [Fang et al. 2016] (noting

that $\Omega = D^{\mathrm{I}}_{3,4}$ does not satisfy $n(\Omega) \le p(\Omega)$). Moreover, one may make use of the signature of the sum of squares, see [Xiao and Yuan 2016, Proposition 2.11], to conclude that $\frac{1}{2}n(n+1) \le m_2(D^{\mathrm{I}}_{3,4}) = \binom{3}{2}\binom{4}{2} = 18$, i.e., $n \le 5 = p(D^{\mathrm{I}}_{3,4})$. In other words, combining with the results of the present article, both parts (1) and (2) of Corollary 3.13 hold true for $\Omega = D^{\mathrm{I}}_{3,4}$. Moreover, in general this method does not imply that $n \le p(\Omega)$ if there exists a holomorphic isometry $f : (\mathbb{B}^n, kg_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ with $k \ge 2$, where $\Omega$ is any irreducible bounded symmetric domain of rank $\ge 2$.

## 4. On holomorphic isometries of complex unit balls into certain irreducible bounded symmetric domains of rank 2

**4A.** ***Characterization of images of holomorphic isometries.*** We start with the following lemma which identifies those irreducible bounded symmetric domains $\Omega \Subset \mathbb{C}^N$ of rank 2 which carry extra properties.

**Lemma 4.1.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank* 2. *Then*, $2N > N' + 1$ *provided that $\Omega$ is not biholomorphic to $D^{\mathrm{I}}_{2,q}$ for any $q \ge 5$.*

*Proof.* The proof follows from direct computation for any irreducible bounded symmetric domain $\Omega$ of rank 2 by using results in [Nakagawa and Takagi 1976, p. 663]. Actually, we obtain from that paper the value of $N' := N(1)$ for any irreducible Hermitian symmetric space $X_c$ of the compact type.

Case 1: When $\Omega$ is not biholomorphic to any type-I domains $D^{\mathrm{I}}_{2,q}$ for $q \ge 3$, $\Omega$ is either biholomorphic to $D^{\mathrm{IV}}_m$ (for some $m \ge 3$), $D^{\mathrm{II}}_5$ or $D^{\mathrm{V}}$ because of $D^{\mathrm{IV}}_4 \cong D^{\mathrm{I}}_{2,2}$, $D^{\mathrm{IV}}_6 \cong D^{\mathrm{II}}_4$ and $D^{\mathrm{III}}_2 \cong D^{\mathrm{IV}}_3$. If $\Omega \cong D^{\mathrm{IV}}_m$, $m \ge 3$, then it is clear that $2m > N' + 1 = m + 2$. If $\Omega \cong D^{\mathrm{II}}_5$, then $2 \dim_{\mathbb{C}} D^{\mathrm{II}}_5 = 20 > N' + 1 = 2^{5-1} = 16$. If $\Omega \cong D^{\mathrm{V}}$, then $2 \dim_{\mathbb{C}} D^{\mathrm{V}} = 32 > N' + 1 = 26 + 1 = 27$, where $X_c$ is the compact dual of $D^{\mathrm{V}}$. Thus, any such $\Omega$ satisfies the desired property.

Case 2: When $\Omega \cong D^{\mathrm{I}}_{2,q}$ for some $q \ge 3$, we have

$$4q = 2N > N' + 1 = \binom{2+q}{q} = \tfrac{1}{2}(q+1)(q+2)$$

if and only if $0 > q^2 - 5q + 2 = \left(q - \frac{5}{2}\right)^2 - \frac{17}{4}$, which is equivalent to $q = 3$ or $q = 4$ because $q \ge 3$ is an integer and $\left(q - \frac{5}{2}\right)^2 \ge \frac{25}{4} > \frac{17}{4}$ for $q \ge 5$. The result follows. □

**Remark 4.2.** We consider rank-2 irreducible bounded symmetric domains $\Omega$ because the functional equations of holomorphic isometries from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(\Omega, g_\Omega)$ are similar to those of holomorphic isometries from $(\mathbb{B}^n, g_{\mathbb{B}^n})$ to $(D^{\mathrm{IV}}_m, g_{D^{\mathrm{IV}}_m})$ for $m \ge 3$ under the assumption that the isometries map $\mathbf{0}$ to $\mathbf{0}$. This is related to the study in [CM]. In addition, we will assume that such a bounded symmetric domain $\Omega$ satisfies $2 \cdot \dim_{\mathbb{C}} \Omega > N' + 1$.

Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank 2 satisfying $2N > N'+1$, where $N'$ is defined in Section 1. Recall that $g_\Omega$ is the canonical Kähler–Einstein metric on $\Omega$ normalized so that minimal disks are of constant Gaussian curvature $-2$. In terms of the Harish-Chandra coordinates $z = (z_1, \dots, z_N) \in \Omega \subset \mathbb{C}^N$, the Kähler form with respect to $g_\Omega$ is equal to $\omega_{g_\Omega} = -\sqrt{-1}\partial\bar\partial \log h_\Omega(z, z)$, where

$$h_\Omega(z,\xi) = 1 - \sum_{j=1}^{N} z_j \bar\xi_j + \sum_{l=1}^{N'-N} \widehat{G}_l(z)\overline{\widehat{G}_l(\xi)}$$

such that each $\widehat{G}_l(z)$ is a homogeneous polynomial of degree 2 in $z$ so that $\widehat{G}_l(\lambda z) = \lambda^2 \widehat{G}_l(z)$ for any $\lambda \in \mathbb{C}^*$. Note that from Section 2, we have $G_{l+N}(z) = \widehat{G}_l(z)$ for $l = 1, \dots, N'-N$. Write $\boldsymbol{G}(z) := (\widehat{G}_1(z), \dots, \widehat{G}_{N'-N}(z))^T$. Let $n$, $N$ and $N'$ be positive integers satisfying $N' - N + n \le N$. We also let $\boldsymbol{U}' \in M(N-n, N; \mathbb{C})$ be such that $\operatorname{rank}(\boldsymbol{U}') = N - n$. Then, we define

$$W'_{\boldsymbol{U}} := \left\{ z = (z_1, \dots, z_N) \in \Omega : \boldsymbol{U}' z^T = \begin{pmatrix} \boldsymbol{G}(z) \\ \boldsymbol{0}_{(2N-n-N')\times 1} \end{pmatrix} \right\}. \tag{4-1}$$

The following generalizes the study of $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, D_N^{\mathrm{IV}})$, $N \ge 3$, in [CM]. Moreover, in the following proposition, the reason of assuming $n \le 2N - N' =: n_0(\Omega)$ is that there is a certain explicitly defined class of complex-analytic subvarieties of $\Omega$ which contains the images of all holomorphic isometries $(\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$ up to composing with elements in $\mathrm{Aut}(\Omega)$, and each of them is contained entirely in $W_{\boldsymbol{U}''}$ for some matrix $\boldsymbol{U}'' \in M(N - n_0(\Omega), N; \mathbb{C})$ satisfying $\boldsymbol{U}''\overline{\boldsymbol{U}}''^T = \boldsymbol{I}_{N-n_0(\Omega)}$. We will show that this gives a relation between the spaces $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$, $1 \le n \le n_0(\Omega) - 1$, and $\widehat{\mathrm{HI}}_1(\mathbb{B}^{n_0(\Omega)}, \Omega)$.

**Proposition 4.3.** *Let $\Omega \Subset \mathbb{C}^N$ be an irreducible bounded symmetric domain of rank 2 such that $2N > N' + 1$, where $N'$ is defined in Section 1. Let $n$ be an integer satisfying $1 \le n \le 2N - N'$. If $f \in \widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$, then $\Psi(f(\mathbb{B}^n))$ is the irreducible component of $W'_{\boldsymbol{U}}$ containing $\boldsymbol{0}$ for some matrix $\boldsymbol{U}' \in M(N-n, N; \mathbb{C})$ satisfying $\boldsymbol{U}'\overline{\boldsymbol{U}}'^T = \boldsymbol{I}_{N-n}$ and some $\Psi \in \mathrm{Aut}(\Omega)$ satisfying $\Psi(f(\boldsymbol{0})) = \boldsymbol{0}$. Conversely, given any matrix $\boldsymbol{U}'' \in M(N-n, N; \mathbb{C})$ satisfying $\boldsymbol{U}''\overline{\boldsymbol{U}}''^T = \boldsymbol{I}_{N-n}$, the irreducible component of $W_{\boldsymbol{U}''}$ containing $\boldsymbol{0}$ is the image of some $\tilde f \in \widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$.*

*Proof.* Let $f \in \widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$. Assume without loss of generality that $f(\boldsymbol{0}) = \boldsymbol{0}$. Then, we have

$$1 - \sum_{j=1}^{N} |f^j(w)|^2 + \sum_{l=1}^{N'-N} |\widehat{G}_l(f(w))|^2 = 1 - \sum_{l=1}^{n} |w_l|^2.$$

Note that $2N - 1 \ge N'$ and $2N - N' \ge n$. By Lemma 2.1, there exists $\boldsymbol{U} \in U(N)$ such that

$$\boldsymbol{U}(f^1(w), \dots, f^N(w))^T = (w_1, \dots, w_n, \boldsymbol{G}(f(w))^T, \boldsymbol{0}_{1\times(2N-n-N')})^T. \tag{4-2}$$

We write $\boldsymbol{U} = [\boldsymbol{A}'\ \boldsymbol{U}']^T$, where $\boldsymbol{U}' \in M(N-n, N; \mathbb{C})$ is a matrix which satisfies $\boldsymbol{U}'\overline{\boldsymbol{U}}'^T = \boldsymbol{I}_{N-n}$. Then, we have $f(\mathbb{B}^n) \subseteq W'_{\boldsymbol{U}}$ by (4-2). It is clear that the Jacobian matrix of $W'_{\boldsymbol{U}}$ at $\mathbf{0}$ is equal to $\boldsymbol{U}'$, which is of full rank $N-n$ so that $W'_{\boldsymbol{U}}$ is smooth at $\mathbf{0}$ and of dimension $n$ at $\mathbf{0}$. Let $S$ be the irreducible component of $W'_{\boldsymbol{U}}$ containing $f(\mathbb{B}^n)$, which also contains $\mathbf{0}$. Then, we have $\dim S = n$. Since both $S$ and $f(\mathbb{B}^n)$ are irreducible complex-analytic subvarieties of $\Omega$, $f(\mathbb{B}^n) \subseteq S$ and $\dim S = \dim f(\mathbb{B}^n) = n$, we have $S = f(\mathbb{B}^n)$. Thus, the irreducible component of $W'_{\boldsymbol{U}}$ containing $\mathbf{0}$ is the image of some holomorphic isometric embedding $f : (\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$.

Conversely, let $n$ be an integer satisfying $1 \le n \le 2N - N'$ and let $\boldsymbol{U}'' \in M(N-n, N; \mathbb{C})$ be a matrix satisfying $\boldsymbol{U}''\overline{\boldsymbol{U}}''^T = \boldsymbol{I}_{N-n}$. By Lemma 2.3, there exists $\boldsymbol{A}'' \in M(n, N; \mathbb{C})$ such that $[\boldsymbol{A}''\ \boldsymbol{U}'']^T \in U(N)$ so that

$$\begin{bmatrix}\boldsymbol{A}''\\ \boldsymbol{U}''\end{bmatrix}\begin{pmatrix}z_1\\ \vdots\\ z_N\end{pmatrix} = \begin{pmatrix}\boldsymbol{w}(z)\\ \boldsymbol{G}(z)\\ \mathbf{0}_{(2N-n-N')\times 1}\end{pmatrix} \quad \text{for all } z = (z_1, \dots, z_n) \in W_{\boldsymbol{U}''}, \tag{4-3}$$

where $\boldsymbol{w}(z) = (w_1(z), \dots, w_n(z))^T := \boldsymbol{A}''(z_1, \dots, z_N)^T$. Note that the Jacobian matrix of $W_{\boldsymbol{U}''}$ at $\mathbf{0}$ is equal to $\boldsymbol{U}''$, which is of full rank $N-n$ so that $W_{\boldsymbol{U}''}$ is smooth at $\mathbf{0}$ and of dimension $n$ at $\mathbf{0}$. Let $S'$ be the irreducible component of $W_{\boldsymbol{U}''}$ containing $\mathbf{0}$. Then, we have $\dim S' = n$. Actually $S'$ is precisely the point set closure of the connected component of $\mathrm{Reg}(W_{\boldsymbol{U}''})$ containing $\mathbf{0}$ in $\Omega$. Denote by $\mathrm{Reg}(S')$ the regular locus of $S'$. Then, $\mathrm{Reg}(S')$ is a connected complex manifold lying inside $\Omega$ and $\mathbf{0} \in \mathrm{Reg}(S')$. Let $\varphi : B(\mathbf{0}) \to \mathrm{Reg}(S')$ be a biholomorphism onto an open neighborhood of $\mathbf{0}$ in $\mathrm{Reg}(S')$ such that $\varphi(\mathbf{0}) = \mathbf{0}$, where $B(\mathbf{0})$ is some open neighborhood of $\mathbf{0}$ in $\mathbb{C}^n$. Here the image $\varphi(B(\mathbf{0}))$ is a germ of complex submanifold of $\Omega$ at $\mathbf{0}$, i.e., a complex submanifold of some open neighborhood of $\mathbf{0}$ in $\Omega$. Note that $h_\Omega(z, z) = 1 - \sum_{l=1}^n |w_l(z)|^2$ for any $z \in S'$ and $\zeta = (\zeta_1, \dots, \zeta_n)$ can be regarded as local holomorphic coordinates on $\mathrm{Reg}(S')$ around $\mathbf{0} \in \mathrm{Reg}(S')$. Then, it follows from (4-3) that for $\zeta \in B(\mathbf{0})$, we have

$$h_\Omega(\varphi(\zeta), \varphi(\zeta)) = 1 - \sum_{l=1}^n |w_l(\varphi(\zeta))|^2 \tag{4-4}$$

and $-\log h_\Omega(\varphi(\zeta), \varphi(\zeta)) = -\log\big(1 - \sum_{l=1}^n |w_l(\varphi(\zeta))|^2\big)$ is a local Kähler potential on $\mathrm{Reg}(S')$ which is the restriction of the Kähler potential on $(\Omega, g_\Omega)$ to an open neighborhood of $\mathbf{0}$ in $\mathrm{Reg}(S')$. It follows from (4-4) that the germ of $S'$ at $\mathbf{0}$ is the image of a germ of holomorphic isometry $\tilde f : (\mathbb{B}^n, g_{\mathbb{B}^n}; \mathbf{0}) \to (\Omega, g_\Omega; \mathbf{0})$. By the extension theorem of [Mok 2012], $\tilde f$ extends to a holomorphic isometric embedding $\tilde f : (\mathbb{B}^n, g_{\mathbb{B}^n}) \to (\Omega, g_\Omega)$. Since both $\tilde f(\mathbb{B}^n)$ and $S'$ are $n$-dimensional irreducible complex-analytic subvarieties of $\Omega$ and $\tilde f(B^n(\mathbf{0}, \varepsilon)) \subset \tilde f(\mathbb{B}^n) \cap S'$ for some real number $\varepsilon \in (0, 1)$. It follows that $S' = \tilde f(\mathbb{B}^n)$. Hence, the irreducible component of $W_{\boldsymbol{U}''}$ containing $\mathbf{0}$ is the image of some holomorphic isometric embedding $\tilde{f} \in \widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$. □

**Remark 4.4.** From the proof of Lemma 4.1, we see that Proposition 4.3 precisely holds true for the space $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega)$ whenever the integer $n$ and the bounded symmetric domain $\Omega$ satisfy one of the following:

(1) $\Omega \cong D^{\mathrm{I}}_{2,3},\ 1 \le n \le 3 = p(D^{\mathrm{I}}_{2,3})$.

(2) $\Omega \cong D^{\mathrm{I}}_{2,4},\ 1 \le n \le 2$.

(3) $\Omega \cong D^{\mathrm{II}}_{5},\ 1 \le n \le 5 = p(D^{\mathrm{II}}_{5}) - 1$.

(4) $\Omega \cong D^{\mathrm{IV}}_{m}$ for some integer $m \ge 3,\ 1 \le n \le m-1 = p(D^{\mathrm{IV}}_{m}) + 1$.

(5) $\Omega \cong D^{\mathrm{V}},\ 1 \le n \le 6$.

Moreover, Proposition 4.3 actually provides the classification of images of all $f \in \widehat{\mathrm{HI}}_1(\Delta, \Omega)$ whenever $\Omega$ is a rank-2 irreducible bounded symmetric domain which is not biholomorphic to $D^{\mathrm{I}}_{2,q}$ for any $q \ge 5$. This also solves part of Problem 3 in [Mok and Ng 2009, p. 2645] theoretically. It is expected that there are many incongruent holomorphic isometries in $\widehat{\mathrm{HI}}_1(\Delta, \Omega)$. However, Proposition 4.3 at least provides a source of constructing explicit examples of holomorphic isometries in $\widehat{\mathrm{HI}}_1(\Delta, \Omega)$. In particular, for the case where the target is an irreducible bounded symmetric domain of rank 2, Problem 3 in [Mok and Ng 2009, p. 2645] remains unsolved precisely in the case where the target $\Omega$ is $D^{\mathrm{I}}_{2,q}$ for some $q \ge 5$.

**4B. *Proof of Theorem 1.3.*** As we have mentioned in Section 4A, Proposition 4.3 actually gives a relation between the spaces $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, \Omega),\ 1 \le n \le n_0(\Omega) - 1$, and $\widehat{\mathrm{HI}}_1(\mathbb{B}^{n_0(\Omega)}, \Omega)$. In other words, this yields Theorem 1.3.

*Proof of Theorem 1.3.* We follow the setting in the proof of Proposition 4.3. Assume without loss of generality that $f(\mathbf{0}) = \mathbf{0}$. Note that $N' - N + n < N$ and thus $f(\mathbb{B}^n)$ is the irreducible component of $W'_{\boldsymbol{U}}$ containing $\mathbf{0}$ for some matrix $\boldsymbol{U}' \in M(N-n, N; \mathbb{C})$ satisfying $\boldsymbol{U}'\overline{\boldsymbol{U}}'^T = \boldsymbol{I}_{N-n}$ by Proposition 4.3. Moreover, we have

$$\begin{bmatrix} \boldsymbol{A}' \\ \boldsymbol{U}' \end{bmatrix} (f^1(w), \ldots, f^N(w))^T = \big(w_1, \ldots, w_n, \boldsymbol{G}(f(w))^T, \mathbf{0}_{1\times(2N-N'-n)}\big)^T$$

for some $\boldsymbol{A}' \in M(n, N; \mathbb{C})$ such that $[\boldsymbol{A}'\ \boldsymbol{U}']^T \in U(N)$ after composing with some element in the isotropy subgroup of $\mathrm{Aut}(\mathbb{B}^n)$ at $\mathbf{0}$ if necessary (by Lemma 2.3). We write

$$\boldsymbol{U}' = \begin{bmatrix} \boldsymbol{U}'_1 \\ \boldsymbol{U}'_2 \end{bmatrix} \quad \text{for some } \boldsymbol{U}'_1 \in M(N'-N, N; \mathbb{C}),\ \boldsymbol{U}'_2 \in M(2N-N'-n, N; \mathbb{C}).$$

Moreover, we have $\boldsymbol{U}'_1(z_1, \ldots, z_N)^T = \boldsymbol{G}(z)$ and $\boldsymbol{U}'_1\overline{\boldsymbol{U}}'^T_1 = \boldsymbol{I}_{N'-N}$ for any $z \in W'_{\boldsymbol{U}}$. It follows from Proposition 4.3 that the irreducible component of $W_{\boldsymbol{U}'_1}$ containing $\mathbf{0}$

is the image of some holomorphic isometric embedding $F : (\mathbb{B}^{n_0}, g_{\mathbb{B}^{n_0}}) \to (\Omega, g_\Omega)$, where $n_0 = n_0(\Omega) := 2N - N'$. We may suppose that $F(\mathbf{0}) = \mathbf{0}$ without loss of generality. Since $f(\mathbb{B}^n) \subset \Omega$ is irreducible and $f(\mathbb{B}^n) \subset W_{U_1'}$, we know $S := f(\mathbb{B}^n)$ lies inside the irreducible component $S' := F(\mathbb{B}^{n_0})$ of $W_{U_1'}$ containing $\mathbf{0}$. Since $(S, g_\Omega|_S) \cong (\mathbb{B}^n, g_{\mathbb{B}^n})$ and $(S', g_\Omega|_{S'}) \cong (\mathbb{B}^{n_0}, g_{\mathbb{B}^{n_0}})$ are of constant holomorphic sectional curvature $-2$, we have $(S, g_\Omega|_S) \subset (S', g_\Omega|_{S'})$ is totally geodesic and the result follows; see the proof of [CM, Theorem 2]. □

**Remark 4.5.** (1) It follows from Lemma 4.1 that Theorem 1.3 holds true when the pair $(\Omega, n_0(\Omega))$ is one of the following:

(a) $\Omega \cong D^{\mathrm{I}}_{2,3}$, $n_0(\Omega) = 3$.

(b) $\Omega \cong D^{\mathrm{I}}_{2,4}$, $n_0(\Omega) = 2$.

(c) $\Omega \cong D^{\mathrm{II}}_{5}$, $n_0(\Omega) = 5$.

(d) $\Omega \cong D^{\mathrm{IV}}_{m}$ $(m \geq 3)$, $n_0(\Omega) = m - 1$.

(e) $\Omega \cong D^{\mathrm{V}}$, $n_0(\Omega) = 6$.

(2) It is not known whether Theorem 1.3 still holds true when $n_0(\Omega)$ is replaced by $p(\Omega) + 1$ and $\Omega \ncong D^{\mathrm{IV}}_m$ for any integer $m \geq 3$.

(3) For the particular case where $\Omega = D^{\mathrm{I}}_{2,3}$, it follows from [Mok 2016] that if the space $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, D^{\mathrm{I}}_{2,3})$ is nonempty, then $n \leq p(D^{\mathrm{I}}_{2,3}) + 1 = 4$. In this case, it is motivated by our study in the present article to consider the following problem in order to classify all holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, D^{\mathrm{I}}_{2,3})$:

Given any $f \in \widehat{\mathrm{HI}}_1(\mathbb{B}^3, D^{\mathrm{I}}_{2,3})$, can $f$ be factorized as $f = F \circ \rho$ for some $F \in \widehat{\mathrm{HI}}_1(\mathbb{B}^4, D^{\mathrm{I}}_{2,3})$ and $\rho \in \widehat{\mathrm{HI}}_1(\mathbb{B}^3, \mathbb{B}^4)$?

If the problem were solved and the answer were affirmative, then the classification of all holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^n, D^{\mathrm{I}}_{2,3})$ would be reduced to the uniqueness problem for nonstandard (i.e., not totally geodesic) holomorphic isometries in $\widehat{\mathrm{HI}}_1(\mathbb{B}^4, D^{\mathrm{I}}_{2,3})$ constructed by Mok [2016].

## Acknowledgements

This work is part of the author's Ph.D. thesis [Chan 2016] at the University of Hong Kong, except for item (2) of Remark 3.14. He would like to express his gratitude to his supervisor, Professor Ngaiming Mok, for his guidance and encouragement. The author would also like to thank Dr. Yuan Yuan for his interest in the research which lead to item (2) of Remark 3.14, and thank the anonymous referees for their helpful suggestions.

## References

[Calabi 1953] E. Calabi, "Isometric imbedding of complex manifolds", *Ann. of Math.* (2) **58** (1953), 1–23. MR Zbl

[Chan 2016] S.-T. Chan, *On holomorphic isometric embeddings of complex unit balls into bounded symmetric domains*, Ph.D. thesis, University of Hong Kong, 2016, available at http://hdl.handle.net/10722/235865.

[Chan and Mok 2017] S. T. Chan and N. Mok, "Holomorphic isometries of $\mathbb{B}^m$ into bounded symmetric domains arising from linear sections of minimal embeddings of their compact duals", *Math. Z.* **286**:1-2 (2017), 679–700. MR Zbl

[Fang et al. 2016] H. Fang, X. Huang, and M. Xiao, "Volume-preserving maps between Hermitian symmetric spaces of compact type", preprint, 2016. arXiv

[Gunning 1990] R. C. Gunning, *Introduction to holomorphic functions of several variables, I: Function theory*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1990. MR Zbl

[Hwang and Mok 1999] J.-M. Hwang and N. Mok, "Varieties of minimal rational tangents on uniruled projective manifolds", pp. 351–389 in *Several complex variables* (Berkeley, CA, 1995–1996), edited by M. Schneider and Y.-T. Siu, Math. Sci. Res. Inst. Publ. **37**, Cambridge Univ. Press, 1999. MR Zbl

[Koziarz and Maubon 2008] V. Koziarz and J. Maubon, "Representations of complex hyperbolic lattices into rank 2 classical Lie groups of Hermitian type", *Geom. Dedicata* **137** (2008), 85–111. MR Zbl

[Loos 1977] O. Loos, "Bounded symmetric domains and Jordan pairs", mathematical lectures, University of California, Irvine, 1977.

[Mok 1989] N. Mok, *Metric rigidity theorems on Hermitian locally symmetric manifolds*, Series in Pure Mathematics **6**, World Scientific, Teaneck, NJ, 1989. MR Zbl

[Mok 2002a] N. Mok, "Characterization of certain holomorphic geodesic cycles on quotients of bounded symmetric domains in terms of tangent subspaces", *Compositio Math.* **132**:3 (2002), 289–309. MR Zbl

[Mok 2002b] N. Mok, "Local holomorphic isometric embeddings arising from correspondences in the rank-1 case", pp. 155–165 in *Contemporary trends in algebraic geometry and algebraic topology* (Tianjin, 2000), edited by S.-S. Chern et al., Nankai Tracts Math. **5**, World Scientific, River Edge, NJ, 2002. MR Zbl

[Mok 2011] N. Mok, "Geometry of holomorphic isometries and related maps between bounded domains", pp. 225–270 in *Geometry and analysis, II*, edited by L. Ji, Adv. Lect. Math. (ALM) **18**, International Press, Somerville, MA, 2011. MR Zbl

[Mok 2012] N. Mok, "Extension of germs of holomorphic isometries up to normalizing constants with respect to the Bergman metric", *J. Eur. Math. Soc.* (*JEMS*) **14**:5 (2012), 1617–1656. MR Zbl

[Mok 2014] N. Mok, "Local holomorphic curves on a bounded symmetric domain in its Harish-Chandra realization exiting at regular points of the boundary", *Pure Appl. Math. Q.* **10**:2 (2014), 259–288. MR Zbl

[Mok 2016] N. Mok, "Holomorphic isometries of the complex unit ball into irreducible bounded symmetric domains", *Proc. Amer. Math. Soc.* **144**:10 (2016), 4515–4525. MR Zbl

[Mok and Ng 2009] N. Mok and S. C. Ng, "Second fundamental forms of holomorphic isometries of the Poincaré disk into bounded symmetric domains and their boundary behavior along the unit circle", *Sci. China Ser. A* **52**:12 (2009), 2628–2646. MR Zbl

[Mumford 1976] D. Mumford, *Algebraic geometry, I: Complex projective varieties*, Grundlehren der Mathematischen Wissenschaften **221**, Springer, 1976. MR Zbl

[Nakagawa and Takagi 1976] H. Nakagawa and R. Takagi, "On locally symmetric Kaehler submanifolds in a complex projective space", *J. Math. Soc. Japan* **28**:4 (1976), 638–667. MR Zbl

[Ng 2010] S.-C. Ng, "On holomorphic isometric embeddings of the unit disk into polydisks", *Proc. Amer. Math. Soc.* **138**:8 (2010), 2907–2922. MR Zbl

[Ng 2011] S.-C. Ng, "On holomorphic isometric embeddings of the unit $n$-ball into products of two unit $m$-balls", *Math. Z.* **268**:1-2 (2011), 347–354. MR Zbl

[Upmeier et al. 2016] H. Upmeier, K. Wang, and G. Zhang, "Holomorphic isometries from the unit ball into symmetric domains", preprint, 2016. arXiv

[Wolf 1972] J. A. Wolf, "Fine structure of Hermitian symmetric spaces", pp. 271–357 in *Symmetric spaces* (St. Louis, MO 1969–1970), edited by W. M. Boothby and G. L. Weiss, Pure and App. Math. **8**, Dekker, New York, 1972. MR Zbl

[Xiao and Yuan 2016] M. Xiao and Y. Yuan, "Holomorphic maps from the complex unit ball to type IV classical domains", preprint, 2016. arXiv

[Zhang 1999] F. Zhang, *Matrix theory: basic results and techniques*, Springer, 1999. MR Zbl

Received February 3, 2017. Revised November 1, 2017.

SHAN TAI CHAN
DEPARTMENT OF MATHEMATICS
SYRACUSE UNIVERSITY
SYRACUSE, NY
UNITED STATES
schan08@syr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.317

# HAMILTONIAN STATIONARY CONES WITH ISOTROPIC LINKS

JINGYI CHEN AND YU YUAN

*In memory of Professor Wei-Yue Ding*

**We show that any closed oriented immersed Hamiltonian stationary isotropic surface $\Sigma$ with genus $g_\Sigma$ in $S^5 \subset \mathbb{C}^3$ is (1) Legendrian and minimal if $g_\Sigma = 0$; (2) either Legendrian or with exactly $2g_\Sigma - 2$ Legendrian points if $g_\Sigma \geq 1$. In general, every compact oriented immersed isotropic submanifold $L^{n-1} \subset S^{2n-1} \subset \mathbb{C}^n$ such that the cone $C(L^{n-1})$ is Hamiltonian stationary must be Legendrian and minimal if its first Betti number is zero. Corresponding results for nonorientable links are also provided.**

## 1. Introduction

In this note we study the problem of when a Hamiltonian stationary cone $C(L)$ with isotropic link $L$ on $S^{2n-1}$ in $\mathbb{C}^n$ becomes special Lagrangian. A submanifold $M \subset \mathbb{C}^n$, not necessarily a Lagrangian submanifold, is *Hamiltonian stationary* if

$$\operatorname{div}_M(JH) = 0,$$

where $J$ is the complex structure in $\mathbb{C}^n$ and $H$ is the mean curvature vector of $M$ in $\mathbb{C}^n$. In fact this is the variational equation of the volume of $M$, when one makes an arbitrary deformation $J\nabla_M\varphi$ with $\varphi \in C_0^\infty(M)$ for $M$:

$$\int_M \langle H, J\nabla_M\varphi\rangle = \int_M \varphi \operatorname{div}_M(JH) - \operatorname{div}_M(\varphi JH) = \int_M \varphi \operatorname{div}_M(JH).$$

The notion of Hamiltonian stationary Lagrangian submanifolds in a Kähler manifold was introduced in [Oh 1993] as critical points of the volume functional under Hamiltonian variations (known to A. Weinstein, as noted there). Chen and Morvan [1994] generalized it to the isotropic deformations.

As in [Harvey and Lawson 1982], a submanifold $M$ in $\mathbb{C}^n$ is *isotropic* at $p \in M$ if

$$J(T_pM) \perp T_pM,$$

Chen is partially supported by NSERC, and Yuan is partially supported by an NSF grant .
*MSC2010:* 58J05.
*Keywords:* Hamiltonian stationary cone, minimal Legendrian link.

and it is isotropic if it is isotropic for every $p$. A submanifold $M$ being isotropic is equivalent to the standard symplectic 2-form on $\mathbb{R}^{2n}$ vanishing on $M$. The dimension of an isotropic submanifold is at most $n$, the half real dimension of $\mathbb{C}^n$, and when it is $n$, the submanifold is Lagrangian.

For an immersed $(n-1)$-dimensional submanifold $L$ in the unit sphere $S^{2n-1}$, let $u : L \to S^{2n-1}$ be the restriction of the coordinate functions in $\mathbb{R}^{2n}$ to $L$. A point $u \in L$ is *Legendrian* if $T_u L$ is isotropic in $\mathbb{R}^{2n}$ and

$$J(T_u L) \perp u.$$

$L$ is Legendrian if all the points $u$ are Legendrian. This is equivalent to $L$ being an $(n-1)$-dimensional integral submanifold of the standard contact distribution on $S^{2n-1}$. The cone

$$C(L) = \{rx : r \geq 0, x \in L\}$$

is said to have *link* $L$. In this article, all links $L$are assumed to be connected, and we shall use $\Sigma$ for the 2-dimensional link $L$.

The Hamiltonian stationary condition is a third-order constraint on the submanifold $M$, as seen when $M$ is locally written as a graph over its tangent space at a point. The minimal submanifolds, a second-order constraint on the local graphical representation of $M$, are automatically Hamiltonian stationary. We are particularly interested in the case when $M$ is a Lagrangian submanifold. The existence of (many) compact Hamiltonian stationary Lagrangian submanifolds in $\mathbb{C}^n$ versus the nonexistence of compact minimal submanifolds makes the study of Hamiltonian stationary ones interesting. In this note, we shall not be concerned with the existence of Hamiltonian stationary ones; instead, we shall concentrate on the rigidity property, namely, when the Hamiltonian stationary ones reduce to special Lagrangians, in the case when the submanifold is a cone over a spherical link in $\mathbb{C}^n$.

A well-known fact about a link $L^m \subset S^n$ and the cone $C(L)$ over it is that $L$ is minimal in $S^n$ if and only if $C(L) \setminus \{0\}$ is minimal in $\mathbb{R}^{n+1}$. When $C(L)$ is Hamiltonian stationary and isotropic, possibly away from the cone vertex $0 \in \mathbb{R}^{2n}$, we observe that the Hamiltonian stationary equation for $C(L)$ splits into two equations:

$$\operatorname{div}_L(JH_L) = 0,$$

i.e., the link $L$ is Hamiltonian stationary in $\mathbb{R}^{2n}$ as well, and

$$\langle JH_L, u\rangle = 0,$$

where $H_L$ is the mean curvature vector of $L$ in $\mathbb{R}^{2n}$ and $u$ is the position vector of $L$. Moreover, if the link $L$ is isotropic in $\mathbb{C}^n$, then

$$\operatorname{div}_L(J\overline{H}_L) = 0,$$

where $\overline{H}_L = H_L - mu$ is the mean curvature vector of $L$ in $\mathbb{S}^{2n-1}$; in fact,

$$\operatorname{div}_L(Ju) = \sum_{i=1}^{m} \langle D_{E_i}(Ju), E_i \rangle = \sum_{i=1}^{m} \langle J D_{E_i} u, E_i \rangle = 0$$

as $D_{E_i} u$ is tangent to $L$, where $D$ is the derivative in $\mathbb{R}^{2n}$ and $\{E_1, \ldots, E_m\}$ is an orthonormal local frame on $TL$.

Our observation is that the rigidity statements in [Chen and Yuan 2006] for minimal links generalize to the Hamiltonian stationary setting.

**Theorem 1.1.** *Let $\Sigma$ be a closed oriented immersed isotropic surface with genus $g_\Sigma$ in $S^5 \subset \mathbb{C}^3$ such that the cone $C(\Sigma)$ is Hamiltonian stationary away from its vertex. Then*

(1) *if $g_\Sigma = 0$, the surface $\Sigma$ is Legendrian and minimal* (*in fact*, *totally geodesic*);

(2) *if $g_\Sigma \geq 1$, the surface $\Sigma$ is either Legendrian or has exactly $2g_\Sigma - 2$ Legendrian points counting the multiplicity.*

It is known that the immersed minimal Legendrian sphere ($g_\Sigma = 0$) must be a great two-sphere in $S^5$; see, for example, [Haskins 2004, Theorem 2.7]. Simple isotropic tori ($g_\Sigma = 1$) can be constructed so that the Hamiltonian stationary cone $C(\Sigma)$ is nowhere Lagrangian. A family of Hamiltonian stationary (nonminimal) Lagrangian cones $C(\Sigma)$ with $g_\Sigma = 1$ are presented in [Iriyeh 2005]. Bryant's classification [1985, p. 269] of minimal surfaces with constant curvature in spheres provides examples of flat Legendrian minimal tori, as well as flat non-Legendrian isotropic minimal tori ($g_\Sigma = 1$) in $S^5$. The constructions of [Haskins 2004; Haskins and Kapouleas 2007] show that there are infinitely many immersed (embedded if $g_\Sigma = 1$) minimal Legendrian surfaces for each odd genus in $S^5$.

In general dimensions and codimensions, we have:

**Theorem 1.2.** *Let $L^m$ be a compact isotropic immersed oriented submanifold in the unit sphere $S^{2n-1} \subset \mathbb{C}^n$ such that the cone $C(L^m)$ is Hamiltonian stationary away from its vertex. Suppose that the first Betti number of $L^m$ is* 0. *Then*, *away from its vertex*,

(1) *when $m$ is the top dimension $n-1$, the cone $C(L^{n-1})$ is Lagrangian and minimal* (*or equivalently $L^{n-1}$ is Legendrian and minimal*);

(2) *for $m < n-1$, the cone $C(L^m)$ is isotropic*, *and if the differential* 1*-form $\langle JH_{C(L^m)}, \cdot \rangle$ is closed then the mean curvature $H_{C(L^m)}$ of $C(L^m)$ vanishes on the normal subbundle $JTC(L^m)$.*

We make two remarks when the dimension $m$ of the link is two. First, Theorem 1.2 also implies Theorem 1.1(1). Second, if the first Betti number of $L^2$ is not zero ($g_{L^2} > 0$) and $L$ is isotropically immersed in $S^{2n-1}$, with $2n-1 \geq 5$, and $C(L)$ is

Hamiltonian stationary away from its cone vertex, the same argument as in the proof of Theorem 1.1 leads to the same conclusion as in part (2) of Theorem 1.1, that the cone $C(L^2)$ is isotropic either everywhere or along exactly $2g_{L^2}-2=-\chi(L^2)$ lines.

Theorems 1.2 and 1.1 (except the totally geodesic part) remain valid for nonorientable links (note that $\chi(\Sigma)=2-g_\Sigma$ for a compact nonorientable surface $\Sigma$); see Remarks 2.1 and 3.1. The nonorientable version of Theorem 1.2 implies that one cannot immerse a compact nonorientable $L^{n-1}$ with first Betti number zero Hamiltonian stationarily and isotropically into $S^{2n-1}\subset\mathbb{C}^n$. Otherwise, the cone $C(L^{n-1})$ would be a special Lagrangian cone; then $C(L^{n-1})$ would be orientable, and $L^{n-1}$ would also be orientable. In particular, there exists no isotropic Hamiltonian stationary immersion of a real projective sphere $\mathbb{R}P^2$ into $S^5\subset\mathbb{C}^3$. In passing, we mention that Lê and Wang [2001] showed that minimal link $L^{n-1}\subset S^{2n-1}$ is Legendrian if and only if $f=\langle Au, Ju\rangle$ satisfies $\triangle_L f=-2nf$ for any $A\in su(n)$.

It is interesting to find out whether there exists an isotropic Hamiltonian stationary surface in $S^5$ with exactly $2g_\Sigma-2$ Legendrian points for $g_\Sigma>1$.

## 2. Hopf differentials and proof of Theorem 1.1

To measure how far the isotropic $\Sigma$ is from being Legendrian, or the deviation of the corresponding is cone from being Lagrangian, we project $Ju$ onto the tangent space of $\Sigma$ in $\mathbb{C}^3$, where $J$ is the complex structure in $\mathbb{C}^3$. Denote the length of the projection by

$$f=|\operatorname{Pr} Ju|^2.$$

To compute the length, we need some preparation. Locally, take an isothermal coordinate system $(t^1,t^2)$ on the isotropic surface

$$u:\Sigma\to S^5\subset\mathbb{C}^3.$$

Set the complex variable

$$z=t^1+\sqrt{-1}t^2.$$

Then the induced metric has the local expression with the conformal factor $\varphi$

$$g=\varphi^2[(dt^1)^2+(dt^2)^2]=\varphi^2\,dz\,d\bar{z}.$$

We project $Ju$ to each of the orthonormal bases $\varphi^{-1}u_1$, $\varphi^{-1}u_2$ with $u_i=\partial u/\partial t^i$. Then the sum of the squares of each projection is

$$f=\frac{|\langle Ju,u_1\rangle|^2+|\langle Ju,u_2\rangle|^2}{\varphi^2}=\frac{4|\langle Ju,u_z\rangle|^2}{\varphi^2},$$

where $u_z=\partial u/\partial z$ and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product on $\mathbb{R}^6$, and in particular $0\le f\le 1$. In fact, $f$ is the square of the norm of the symplectic form $\omega$ in $\mathbb{C}^3$

restricted on the cone $C(\Sigma)$ with link $\Sigma$:

$$\omega|_{C(\Sigma)} \wedge *\omega|_{C(\Sigma)} = f \cdot \text{volume form of } C(\Sigma).$$

The Hamiltonian stationary condition for the cone $C(\Sigma) = ru(t^1, t^2)$ is

$$\begin{aligned} 0 &= \operatorname{div}_{C(\Sigma)}(JH_{C(\Sigma)}) \\ &= \langle \partial_r(JH_{C(\Sigma)}), \partial_r \rangle + \frac{1}{r^2} \operatorname{div}_\Sigma\Big(J\frac{1}{r}H_\Sigma\Big) \\ &= -\frac{1}{r^2}\langle JH_\Sigma, u\rangle + \frac{1}{r^3}\operatorname{div}_\Sigma(JH_\Sigma). \end{aligned}$$

It follows that

$$\operatorname{div}_\Sigma(JH_\Sigma) = 0$$

and

$$0 = \langle JH_\Sigma, u\rangle = -\Big\langle \frac{4}{\varphi^2}u_{z\bar{z}}, Ju\Big\rangle.$$

Coupled with the isotropy condition

$$\langle Ju_i, u_j\rangle = 0,$$

we have the holomorphic condition

$$\langle Ju, u_z\rangle_{\bar{z}} = \langle Ju_{\bar{z}}, u_z\rangle + \langle Ju, u_{z\bar{z}}\rangle = \big\langle Ju, -\tfrac{1}{2}\varphi^2 u\big\rangle = 0.$$

The induced metric $g$ yields a compatible conformal structure on the oriented surface $\Sigma$, which makes $\Sigma$ a Riemann surface. We shall consider two cases according to the genus $g_\Sigma$.

Case 1: $g_\Sigma = 0$. By the uniformization theorem for Riemann surfaces, see, for example, [Ahlfors and Sario 1960, p. 125, p. 181], there exists a holomorphic covering map

$$\Phi : (S^2, g_{\text{canonical}}) \to (\Sigma, g),$$

or locally

$$\Phi : \Big(\mathbb{C}^1, \frac{1}{(1+|w|^2)^2}\,dw\,d\bar{w}\Big) \to (\Sigma, g).$$

For $z = \Phi(w)$ one has

$$\frac{1}{(1+|w|^2)^2}\,dw\,d\bar{w} = \Phi^*(\psi^2 g) = \Phi^*(\psi^2\varphi^2\,dz\,d\bar{z}) = \psi^2\varphi^2|z_w|^2\,dw\,d\bar{w},$$

where $\psi$ is a positive (real analytic) function on $\Sigma$. In particular

$$|z_w|^2 = \frac{1}{\psi^2\varphi^2(1+|w|^2)^2}.$$

Note that

$$\langle Ju, u_w\rangle = \langle Ju, u_z\rangle z_w = \langle Ju, u_z\rangle \frac{1}{w_z}$$

is a holomorphic function of $z$; in turn it is a holomorphic function of $w$. Also $\langle Ju, u_w\rangle$ is bounded, approaching 0 as $w$ goes to $\infty$, because

$$|\langle Ju, u_w\rangle|^2 = \frac{|\langle Ju, u_z\rangle|^2}{\varphi^2}\frac{1}{\psi^2(1+|w|^2)^2}.$$

So $\langle Ju, u_w\rangle \equiv 0$. Therefore $f\equiv 0$ and $\Sigma$ is Legendrian. We conclude that $C(\Sigma)\setminus\{0\}$ is Lagrangian.

The 1-form $\langle JH_{C(\Sigma)}, \cdot\rangle$ on the Lagrangian submanifold $C(\Sigma)\setminus\{0\}$ is closed. (This follows directly either from Theorem 3.4 of [Dazord 1981], or can be verified by local exactness via the local expression

$$H_{C(\Sigma)} = -J\nabla_{C(\Sigma)}\theta$$

given in [Harvey and Lawson 1982]; this will be done in next section.) Its restriction along $\Sigma$ is therefore a closed 1-form $i^*\langle JH_{C(\Sigma)}, \cdot\rangle$ as the pullback by the inclusion $i:\Sigma\to C(\Sigma)$ of a closed 1-form. Since the first Betti number of $\Sigma$ is zero ($g_\Sigma = 0$), there is a smooth function $\theta_\Sigma$ on $\Sigma$ such that

$$d\theta_\Sigma = i^*\langle JH_{C(\Sigma)}, \cdot\rangle.$$

Then

$$\langle\nabla_\Sigma\theta_\Sigma, \cdot\rangle = d\theta_\Sigma = \langle JH_\Sigma, \cdot\rangle.$$

As we have seen, the Hamiltonian stationary condition on $C(\Sigma)$ implies

$$0 = \operatorname{div}_\Sigma(JH_\Sigma) = \operatorname{div}_\Sigma(\nabla_\Sigma\theta_\Sigma) = \Delta_g\theta_\Sigma.$$

On the closed surface $\Sigma$, we have $\theta_\Sigma$ is constant, and in turn, $\Sigma$ is minimal.

An immersed minimal Legendrian 2-sphere in $\mathbb{S}^5$ is totally geodesic. This is a known fact; for a proof, see, for example, [Chen and Yuan 2006].

Case 2: $g_\Sigma \geq 1$. As in Case 1, where $g_\Sigma = 0$, the isotropic and Hamiltonian stationary condition gives us a local holomorphic function $\langle Ju, u_z\rangle$ and global holomorphic Hopf 1-differential $\langle Ju, u_z\rangle\, dz$. We only consider the case where $\langle Ju, u_z\rangle\, dz$ is not identically zero. The zeros of $\langle Ju, u_z\rangle$ are therefore isolated and near each of the zeros, we can write

$$\langle Ju, u_z\rangle = h(z)z^k,$$

where $h$ is a local holomorphic function, nonvanishing at the zero point $z=0$ and $k$ is a positive integer. One can also view

$$\langle Ju, u_z\rangle = \tfrac{1}{2}\big(\langle Ju, u_1\rangle - \sqrt{-1}\langle Ju, u_2\rangle\big)$$

as the tangent vector

$$\tfrac{1}{2}\langle Ju, u_1\rangle u_1 - \tfrac{1}{2}\langle Ju, u_2\rangle u_2 = \tfrac{1}{2}\langle Ju, u_1\rangle\,\partial_1 - \tfrac{1}{2}\langle Ju, u_2\rangle\,\partial_2$$

along the tangent space $T\Sigma$, where $\partial_i = \partial u/\partial t^i$. The projection $\Pr Ju$ on the tangent space of $T\Sigma$ is locally represented as

$$\Pr Ju = \frac{\langle Ju, u_1\rangle\, \partial_1 + \langle Ju, u_2\rangle\, \partial_2}{\varphi^2}.$$

The index of the globally defined vector field $\Pr Ju$ at each of its singular points, i.e., where $\Pr Ju = 0$, is the negative of that for the vector field $\frac{1}{2}\langle Ju, u_1\rangle\, \partial_1 - \frac{1}{2}\langle Ju, u_2\rangle\, \partial_2$. Note that the index of the latter is $k$.

From the Poincaré–Hopf index theorem, for any vector field $V$ with isolated singularities on $\Sigma$, one has

$$\sum_{V=0} \operatorname{index}(V) = \chi(\Sigma) = 2 - 2g_\Sigma \le 0.$$

The zeros of $\Pr Ju$ are just the Legendrian points on $\Sigma$. So we conclude that the number of Legendrian points is $2g_\Sigma - 2$ counting the multiplicity. This completes the proof of Theorem 1.1.

**Remark 2.1.** As mentioned in the Introduction, Theorem 1.1 (except the totally geodesic part) and its generalization to higher codimensions can be extended for the nonorientable links. This can be seen as follows. The Poincaré–Hopf index theorem holds on compact nonorientable surfaces, our count of the indices of the still globally defined $\Pr Ju$ via *local* holomorphic functions is valid too, and the index of a singular point of a vector field is independent of local orientations. Moreover, this index-counting argument yields an alternative proof for Theorem 1.1(1) (except the totally geodesic part) and its generalization.

## 3. Harmonic forms and proof of Theorem 1.2

Consider an immersed isotropic Hamiltonian stationary submanifold in $S^{2n-1}$

$$u : L^m \to S^{2n-1} \subset \mathbb{C}^n.$$

The isotropy condition for any local coordinates $(t^1, \ldots, t^m)$ on $L^m$ is given by

$$\langle Ju_i, u_j\rangle = 0,$$

where $J$ is the complex structure of $\mathbb{C}^n$ and $u_i = \partial u/\partial t^i$.

The Hamiltonian stationary condition for the cone $C(\Sigma) = ru(t)$ is

$$\begin{aligned} 0 &= \operatorname{div}_{C(L)}(JH_{C(L)}) \\ &= \langle \partial_r(JH_{C(L)}), \partial_r\rangle + \frac{1}{r^2}\operatorname{div}_L\Big(J\Big(\frac{1}{r}H_L\Big)\Big) \\ &= -\frac{1}{r^2}\langle JH_L, u\rangle + \frac{1}{r^3}\operatorname{div}_L(JH_L). \end{aligned}$$

Notice that $\langle JH_L, u\rangle$ and $\mathrm{div}_L(JH_L)$ are independent of $r$. Therefore, the equation above splits into two equations

$$\mathrm{div}_L(JH_L) = 0$$

and

$$0 = \langle JH_L, u\rangle = -\langle \Delta_g u, Ju\rangle,$$

where $g$ is the induced metric on $L$ and $\Delta_g$ is the Laplace–Beltrami operator of $(L, g)$.

To measure the deviation of the corresponding cone $C(u(L^m))$ from being isotropic, we project $Ju$ onto the tangent space of $u(L^m)$ in $\mathbb{C}^n$. Note that the projection is the vector field along $u(L)$

$$\mathrm{Pr}\, Ju = \sum_{i,j=1}^{m} g^{ij} \langle Ju, u_i\rangle u_j,$$

where $g_{ij} = \langle u_i, u_j\rangle,\ 1 \le i, j \le m$. The corresponding 1-form

$$\alpha = \sum_{i=1}^{m} \langle Ju, u_i\rangle\, dt^i$$

is of course globally defined on $L^m$. In fact it is a harmonic 1-form, because $\alpha$ is closed and coclosed as verified as follows:

$$\begin{aligned} d\alpha &= \sum_{i,j=1}^{m} \langle Ju, u_i\rangle_j\, dt^j \wedge dt^i \\ &= \sum_{i,j=1}^{m} (\langle Ju_j, u_i\rangle + \langle Ju, u_{ij}\rangle)\, dt^j \wedge dt^i \\ &= \sum_{i,j=1}^{m} \langle Ju, u_{ij}\rangle\, dt^j \wedge dt^i = 0, \end{aligned}$$

and

$$\begin{aligned} \delta\alpha &= (-1)^{m\cdot 1+m+1} * d * \alpha \\ &= -* d\Bigl(\sum_{i,j=1}^{m} (-1)^{j+1} \sqrt{g} g^{ij} \langle Ju, u_i\rangle\, dt^1 \wedge \cdots \wedge \widehat{dt^j} \wedge \cdots \wedge dt^m\Bigr) \\ &= -* \sum_{i,j=1}^{m} \partial_j(\sqrt{g} g^{ij} \langle Ju, u_i\rangle)\, dt^1 \wedge \cdots \wedge dt^j \wedge \cdots \wedge dt^m \\ &= -\frac{1}{\sqrt{g}} \sum_{i,j=1}^{m} \partial_j(\sqrt{g} g^{ij} \langle Ju, u_i\rangle) \\ &= -\sum_{i,j=1}^{m} \Bigl(\langle Ju_j, g^{ij} u_i\rangle + \Bigl\langle Ju, \frac{1}{\sqrt{g}} \partial_j(\sqrt{g} g^{ij} u_i)\Bigr\rangle\Bigr) = -\langle Ju, \Delta_g u\rangle = 0, \end{aligned}$$

where we have used the isotropy condition and the consequence of Hamiltonian stationary condition in the last two steps, respectively.

The Hodge–de Rham theorem implies that the harmonic 1-form $\alpha$ must vanish because the first Betti number of $L^m$ is zero by assumption. It follows that $\Pr Ju$ must vanish. Therefore, the cone $C(L^m)$ is isotropic.

Next, we claim that the differential 1-form

$$\beta = \langle JH_L, \cdot \rangle$$

on $L^m$ is closed. When $m = n - 1$, the isotropic cone $C(L^{n-1})$ is Lagrangian. By [Harvey and Lawson 1982], around each point of $C(L^{n-1}) \setminus \{0\}$, there is a locally defined Lagrangian angle $\theta$ such that

$$H_{C(L)} = -J\nabla_{C(L)}\theta.$$

Now the globally defined 1-form $\beta$ on the link $L$ can be expressed locally as

$$\beta = \langle \nabla_{C(L)}\theta, \cdot \rangle = \langle \nabla_L\theta, \cdot \rangle = d_L\theta$$

by noticing that $H_{C(L)} = H_L$ as $r = 1$, where the second equality holds as the two 1-forms are on $TL$ and the tangent vectors to $L$ are orthogonal to $\partial_r$, and $d_L$ stands for the exterior differentiation on $L$. We conclude that $\beta$ is a closed 1-form on $L$. When $m < n - 1$, the 1-form $\langle JH_{C(L)}, \cdot \rangle$ is closed by assumption, so its restriction $\beta$ on $L$ is closed.

Since the first Betti number of $L$ is zero, there is a smooth function $\theta_L$ on $L$ such that $\langle JH_L, \cdot \rangle = d_L\theta_L$. This implies that the projection of $JH_L$ onto $TL$ satisfies

$$\sum_{i=1}^{m} \langle JH_L, E_i \rangle E_i = \nabla_L\theta_L,$$

where $\{E_1, \ldots, E_m\}$ is a local orthonormal frame of $TL$. The Hamiltonian stationary condition on $C(L)$ asserts, as we have seen earlier, that

$$\Delta_L\theta_L = \operatorname{div}_L \nabla_L\theta_L = \operatorname{div}_L(JH_L) = 0.$$

On the closed submanifold $L$, we know $\theta_L$ is constant. In turn, for $m = n - 1$, $C(L^{n-1})$ is minimal, and for $m < n - 1$, $C(L^m)$ is partially minimal, namely $H_{C(L^m)}$ vanishes on the normal subbundle $JTC(L^m)$. The proof of Theorem 1.2 is complete.

**Remark 3.1.** As the projection $\Pr Ju$ and the adjoint operator $\delta$ are independent of the local orientations and the Hodge–de Rham theorem holds for compact nonorientable manifolds, see, for example, [Lawson and Michelsohn 1994, p. 125–126], we see that Theorem 1.2 remains true for nonorientable links $L^m$.

**Remark 3.2.** For a surface link $L^2 \subset \mathbb{S}^{2n-1}$ with $g_L = 0$ for the case $n > 3$, if it is isotropic and $C(L^2)$ is Hamiltonian stationary, the same argument as in [Chen and Yuan 2006] leads to the conclusion that the second fundamental form of $L$ in $\mathbb{S}^{2n-1}$ vanishes in the normal subbundle $Ju \oplus JTL$. When $n = 3$, $L$ is totally geodesic in $\mathbb{S}^5$ as noted before.

**Corollary 3.3.** *Let $L^m$ be a compact immersed isotropic submanifold in the unit sphere $S^{2n-1} \subset \mathbb{C}^n$. If the Ricci curvature of $L^m$ is nonnegative, and it is positive somewhere or the Euler characteristic $\chi(L^m)$ is not zero, then the Hamiltonian stationary cone $C(L^m)$ is isotropic; in particular, $C(L^{n-1})$ is Lagrangian (or equivalently $L^{n-1}$ is Legendrian) and minimal when $m$ is the top dimension $n-1$.*

Under the above condition, from [Bochner 1948, p. 381], it follows immediately that the first Betti number of $L^m$ is zero. Then Theorem 1.2 and its nonorientable version imply the corollary.

## References

[Ahlfors and Sario 1960] L. V. Ahlfors and L. Sario, *Riemann surfaces*, Princeton Mathematical Series **26**, Princeton University Press, 1960. MR Zbl

[Bochner 1948] S. Bochner, "Curvature and Betti numbers", *Ann. of Math.* (2) **49** (1948), 379–390. MR Zbl

[Bryant 1985] R. L. Bryant, "Minimal surfaces of constant curvature in $S^n$", *Trans. Amer. Math. Soc.* **290**:1 (1985), 259–271. MR Zbl

[Chen and Morvan 1994] B.-Y. Chen and J.-M. Morvan, "Deformations of isotropic submanifolds in Kähler manifolds", *J. Geom. Phys.* **13**:1 (1994), 79–104. MR Zbl

[Chen and Yuan 2006] J. Chen and Y. Yuan, "Minimal cones with isotropic links", *Int. Math. Res. Not.* **2006** (2006), art. id. 69284. MR Zbl

[Dazord 1981] P. Dazord, "Sur la géométrie des sous-fibrés et des feuilletages lagrangiens", *Ann. Sci. École Norm. Sup.* (4) **14**:4 (1981), 465–480. MR Zbl

[Harvey and Lawson 1982] R. Harvey and H. B. Lawson, Jr., "Calibrated geometries", *Acta Math.* **148** (1982), 47–157. MR Zbl

[Haskins 2004] M. Haskins, "Special Lagrangian cones", *Amer. J. Math.* **126**:4 (2004), 845–871. MR Zbl

[Haskins and Kapouleas 2007] M. Haskins and N. Kapouleas, "Special Lagrangian cones with higher genus links", *Invent. Math.* **167**:2 (2007), 223–294. MR Zbl

[Iriyeh 2005] H. Iriyeh, "Hamiltonian minimal Lagrangian cones in $\mathbb{C}^m$", *Tokyo J. Math.* **28**:1 (2005), 91–107. MR Zbl

[Lawson and Michelsohn 1994] H. B. Lawson, Jr. and M.-L. Michelsohn, *Spin geometry*, 2nd ed., Princeton Mathematical Series **38**, Princeton University Press, 1994. Zbl

[Lê and Wang 2001] H. Lê and G. Wang, "A characterization of minimal Legendrian submanifolds in $\mathbb{S}^{2n+1}$", *Compositio Math.* **129**:1 (2001), 87–93. MR Zbl

[Oh 1993] Y.-G. Oh, "Volume minimization of Lagrangian submanifolds under Hamiltonian deformations", *Math. Z.* **212**:2 (1993), 175–192. MR Zbl

Received May 2, 2017. Revised October 10, 2017.

JINGYI CHEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC
CANADA
jychen@math.ubc.ca

YU YUAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WA
UNITED STATES
yuan@math.washington.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.329

# QUANDLE THEORY AND THE OPTIMISTIC LIMITS OF THE REPRESENTATIONS OF LINK GROUPS

JINSEOK CHO

**For a given boundary-parabolic representation of a link group to PSL(2, $\mathbb{C}$), Inoue and Kabaya suggested a combinatorial method to obtain the developing map of the representation using the octahedral triangulation and the shadow-coloring of certain quandles. A quandle is an algebraic system closely related to the Reidemeister moves, so their method changes quite naturally under the Reidemeister moves.**

**We apply their method to the potential function, which was used to define the optimistic limit, and construct a saddle point of the function. This construction works for any boundary-parabolic representation, and it shows that the octahedral triangulation is good enough to study all possible boundary-parabolic representations of the link group. Furthermore, the evaluation of the potential function at the saddle point becomes the complex volume of the representation, and this saddle point changes naturally under the Reidemeister moves because it is constructed using the quandle.**

## 1. Introduction

A link $L$ has the hyperbolic structure when there exists a discrete faithful representation $\rho : \pi_1(L) \to \mathrm{PSL}(2, \mathbb{C})$, where the *link group* $\pi_1(L)$ is the fundamental group of the link complement $\mathbb{S}^3 \setminus L$. The standard method to find the hyperbolic structure of $L$ is to consider some triangulation of $\mathbb{S}^3 \setminus L$ and solve certain sets of equations. (These equations are called the *hyperbolicity equations*.) Each solution determines a boundary-parabolic representation[1] and one of them is the *geometric representation*, which means the determined boundary-parabolic representation is discrete and faithful. Due to Mostow's rigidity theorem, the hyperbolic structure of a link is a topological property. Therefore, it is natural to expect the invariance of the hyperbolic structure under the Reidemeister moves. However, this cannot be seen easily, because even a small change on the triangulation changes the solution radically.

*MSC2010:* primary 57M27; secondary 51M25, 58J28.
*Keywords:* optimistic limit, quandle, hyperbolic volume, boundary-parabolic representation, link group.

[1] *boundary-parabolic* means the image of the peripheral subgroup $\pi_1(\partial(\mathbb{S}^3 \setminus L))$ is a parabolic subgroup of PSL(2, $\mathbb{C}$). Note that the geometric representation is boundary-parabolic.

Recently, Inoue and Kabaya [2014] developed a method to construct the hyperbolic structure of $L$ using the link diagram and the geometric representation. More generally, for a given boundary-parabolic representation $\rho$, they constructed the explicit geometric shapes of the tetrahedra of certain triangulations using $\rho$. Their main method is to construct the geometric shapes using certain quandle homology, which is defined directly from the link diagram $D$ and the representation $\rho$. Here, a quandle is an algebraic system whose axioms are closely related to the Reidemeister moves of link diagrams, so their construction changes quite naturally under the Reidemeister moves. (The definition of the quandle is in Section 2A. A good survey of quandles is the book [Elhamdadi and Nelson 2015].) A result of Inoue and Kabaya [2014] suggests a combinatorial method to obtain the hyperbolic structure of the link complement.

Interestingly, the triangulation used in [Inoue and Kabaya 2014] was also used to define the optimistic limit of the Kashaev invariant in [Cho et al. 2014]. As a matter of fact, this triangulation arises naturally from the link diagram. (See Section 3 of [Weeks 2005] and Section 2C of this article for the definition.) We call this triangulation *octahedral triangulation* of $\mathbb{S}^3 \setminus (L \cup \{\text{two points}\})$ associated with the link diagram $D$.

The optimistic limit first appeared in [Kashaev 1995] where the volume conjecture was proposed. This conjecture relates certain limits of link invariants, called Kashaev invariants, with the hyperbolic volumes. The optimistic limit, which was first defined in [Murakami 2000], is the value of a certain potential function evaluated at a saddle point, where the function and the value are expected to be an analytic continuation of the Kashaev invariant and the limit of the invariant, respectively. As a matter of fact, physicists usually call the evaluation the *classical limit* and consider it the *actual limit* of the invariant. A mathematically rigorous definition of the optimistic limit was proposed in [Yokota 2011] and the value was proved to coincide with the hyperbolic volume. Several versions of the optimistic limit have been developed, in a number of articles, but we will modify the version of [Cho et al. 2014] so as to construct a solution without the need to solve equations.

The optimistic limit is defined by the potential function $V(z_1, \ldots, z_n, w_k^j, \ldots)$. Previously, in [Cho et al. 2014], this function was defined purely by the link diagram, but here we modify it using the information of the representation $\rho$. (The definition is in Section 3.) We consider a solution of the set

$$\mathcal{H} := \left\{ \exp\left( z_k \frac{\partial V}{\partial z_k} \right) = 1, \exp\left( w_k^j \frac{\partial V}{\partial w_k^j} \right) = 1 \;\middle|\; j : \text{degenerate crossings}, k = 1, \ldots, n \right\},$$

which is a saddle-point of the potential function $V$. Then Proposition 3.1 will show that $\mathcal{H}$ becomes the hyperbolicity equations of the octahedral triangulation.

Solving the equations in $\mathcal{H}$ is not easy because there are infinitely many solutions.

The standard way to avoid this difficulty is to deform the octahedral triangulation of $\mathbb{S}^3 \setminus (L \cup \{\text{two points}\})$ to the triangulation of $\mathbb{S}^3 \setminus L$, as in [Yokota 2011]. However, this deformation produces the problem of the existence of solutions because some triangulations constructed from a link diagram may have no solution. (Sakuma and Yokota [2016] proved the existence of solutions for the alternating links.) Furthermore, the author believes these deformations of the triangulation lose the combinatorial properties of link diagrams. Therefore, we will use the octahedral triangulation without any deformation and do not solve the equations in $\mathcal{H}$. Instead, we will construct an explicit solution $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ of $\mathcal{H}$.

**Theorem 1.1.** *Using the quandle associated with the representation $\rho$, there exists a formula to construct a solution $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ of $\mathcal{H}$. (The exact formulas are in Theorem 3.2.)*

The evaluation of the potential function $V$ depends on the choice of log-branch. To obtain a well-defined value, modify the potential function to

$$(1)\quad V_0(z_1, \dots, z_n, (w_k^j), \dots) := V(z_1, \dots, z_n, (w_k^j), \dots) - \sum_k \Big(z_k \frac{\partial V}{\partial z_k}\Big) \log z_k - \sum_{j,k} \Big(w_k^j \frac{\partial V}{\partial w_k^j}\Big) \log w_k^j.$$

**Theorem 1.2.** *For the constructed solution $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ of $\mathcal{H}$ and the modified potential function $V_0$ above, the following holds:*

$$(2)\qquad V_0(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots) \equiv i(\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)) \pmod{\pi^2},$$

*where* $\mathrm{vol}(\rho)$ *and* $\mathrm{cs}(\rho)$ *are the hyperbolic volume and the Chern–Simons invariant of $\rho$ defined in [Zickert 2009], respectively.*

The proof will be in Theorem 3.3. The left-hand side of (2) is called the *optimistic limit of $\rho$*, and $\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)$ in the right-hand side is called *the complex volume of $\rho$*.

Note that for any boundary-parabolic representation $\rho$, we can always construct the solution associated with $\rho$. This implies that the octahedral triangulation is good enough for the study of all possible boundary-parabolic representations from the link group to $\mathrm{PSL}(2, \mathbb{C})$. The set of all possible representations can be regarded as the *Ptolemy variety* (see [Garoufalidis et al. 2015] for detail) and we expect the octahedral triangulation will be very useful to the study of the Ptolemy variety. (An actual application to the Ptolemy variety is in preparation now.)

Furthermore, the construction of the solution is based on the quandle in [Inoue and Kabaya 2014]. Therefore, this solution changes locally under the Reidemeister moves. This implies that we can explore the hyperbolic structure of a link by finding the solution and keeping track of the changes of the solution under the Reidemeister moves. As a matter of fact, after the appearance of the first draft of this article, this idea was successfully used in [Cho 2016a; Cho and Murakami 2017] and more applications are in preparation.

Among the applications, we remark that [Cho 2016a] contains very similar results to this article. Both articles construct the solution associated with $\rho$ using the same quandle. However, the major differences are the triangulations. Both use the same *octahedral decomposition* of $\mathbb{S}^3 \setminus (L \cup \{\text{two points}\})$, but this article uses the subdivision of each octahedron into four tetrahedra and call the result *four-term* (*or octahedral*) *triangulation*, whereas [Cho 2016a] uses the subdivision of the same octahedron into five tetrahedra and calls the result *five-term triangulation*. Some tetrahedra in the four-term triangulation can be degenerate and this introduces technical difficulties. However, the five-term triangulation used in [Cho 2016a] does not contain any degenerate tetrahedra, so it is far easier and more convenient. In conclusion, this article contains the original idea of using a quandle to construct the solution and [Cho 2016a] improved the idea.

The layout of this article is as follows. In Section 2, we will summarize some results from [Inoue and Kabaya 2014]. In particular, the definition of the quandle and the octahedral triangulation will appear. Section 3 will define the optimistic limit and the hyperbolicity equations. The main formula (Theorem 3.3) of the solution associated with the given representation $\rho$ will appear. Section 4 will discuss two simple examples, the figure-eight knot $4_1$ and the trefoil knot $3_1$.

## 2. Quandles

In this section, we will survey some results of [Inoue and Kabaya 2014]. We remark that all formulas in this section come from that article, and the author learned them from the series of lectures given by Ayumu Inoue at Seoul National University during the spring of 2012.

### 2A. *Conjugation quandle of parabolic elements.*

**Definition 2.1.** A *quandle* is a set $X$ with a binary operation $*$ satisfying the following three conditions:

(1) $a * a = a$ for any $a \in X$.

(2) The map $*b : X \to X$ $(a \mapsto a * b)$ is bijective for any $b \in X$.

(3) $(a * b) * c = (a * c) * (b * c)$ for any $a, b, c \in X$.

The inverse of $*b$ is notated by $*^{-1}b$. In other words, the equation $a *^{-1} b = c$ is equivalent to $c * b = a$.

**Definition 2.2.** Let $G$ be a group and $X$ be a subset of $G$ satisfying

$$g^{-1} X g = X \quad \text{for any } g \in G.$$

Define the binary operation $*$ on $X$ by

$$a * b = b^{-1}ab \tag{3}$$

for any $a, b \in X$. Then $(X, *)$ becomes a quandle and is called the *conjugation quandle*.

As an example, let $\mathcal{P}$ be the set of parabolic elements of $\mathrm{PSL}(2, \mathbb{C}) = \mathrm{Isom}^+(\mathbb{H}^3)$. Then,

$$g^{-1}\mathcal{P}g = \mathcal{P}$$

holds for any $g \in \mathrm{PSL}(2, \mathbb{C})$. Therefore, $(\mathcal{P}, *)$ is a conjugation quandle, and this is the only quandle we use in this article.

To perform concrete calculations, an explicit expression of $(\mathcal{P}, *)$ was introduced in [Inoue and Kabaya 2014]. First, note that

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} 1+rs & s^2 \\ -r^2 & 1-rs \end{pmatrix},$$

for $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{C})$. Therefore, we can identify $(\mathbb{C}^2 \setminus \{0\})/\pm$ with $\mathcal{P}$ by

$$\begin{pmatrix} \alpha & \beta \end{pmatrix} \longleftrightarrow \begin{pmatrix} 1+\alpha\beta & \beta^2 \\ -\alpha^2 & 1-\alpha\beta \end{pmatrix}, \tag{4}$$

where $\pm$ means the equivalence relation $\begin{pmatrix} \alpha & \beta \end{pmatrix} \sim \begin{pmatrix} -\alpha & -\beta \end{pmatrix}$. We define the operation $*$ on $\mathcal{P}$ by

$$\begin{pmatrix} \alpha & \beta \end{pmatrix} * \begin{pmatrix} \gamma & \delta \end{pmatrix} := \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} 1+\gamma\delta & \delta^2 \\ -\gamma^2 & 1-\gamma\delta \end{pmatrix} \in (\mathbb{C}^2 \setminus \{0\})/\pm,$$

where the matrix multiplication on the right-hand side is the standard multiplication. (This definition is the transpose of the one used in [Inoue and Kabaya 2014] and [Cho 2016a].) Note that this definition coincides with the operation of the conjugation quandle $(\mathcal{P}, *)$ by

$$\begin{aligned} \begin{pmatrix} \alpha & \beta \end{pmatrix} * \begin{pmatrix} \gamma & \delta \end{pmatrix} &= \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} 1+\gamma\delta & \delta^2 \\ -\gamma^2 & 1-\gamma\delta \end{pmatrix} \in (\mathbb{C}^2 \setminus \{0\})/\pm \\ &\longleftrightarrow \begin{pmatrix} 1+\gamma\delta & \delta^2 \\ -\gamma^2 & 1-\gamma\delta \end{pmatrix}^{-1} \begin{pmatrix} 1+\alpha\beta & -\alpha^2 \\ \beta^2 & 1-\alpha\beta \end{pmatrix} \begin{pmatrix} 1+\gamma\delta & \delta^2 \\ -\gamma^2 & 1-\gamma\delta \end{pmatrix} \\ &= \begin{pmatrix} \gamma & \delta \end{pmatrix}^{-1} \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} \gamma & \delta \end{pmatrix} \in \mathrm{PSL}(2, \mathbb{C}). \end{aligned}$$

The inverse operation is given by

$$\begin{pmatrix} \alpha & \beta \end{pmatrix} *^{-1} \begin{pmatrix} \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} 1-\gamma\delta & -\gamma^2 \\ \delta^2 & 1+\gamma\delta \end{pmatrix}.$$

From now on, we use the notation $\mathcal{P}$ instead of $(\mathbb{C}^2 \setminus \{0\})/\pm$.

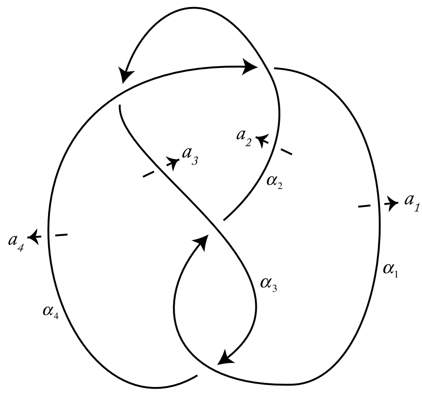



**Figure 1.** The figure-eight knot $4_1$.

**2B.** ***Link group and shadow-coloring.*** Consider a representation $\rho : \pi_1(L) \to \mathrm{PSL}(2, \mathbb{C})$ of a hyperbolic link $L$. We call $\rho$ *boundary-parabolic* when the peripheral subgroup $\pi_1(\partial(\mathbb{S}^3 \setminus L))$ of $\pi_1(L)$ maps to a subgroup of $\mathrm{PSL}(2, \mathbb{C})$ whose elements are all parabolic.

For a fixed oriented link diagram[2] $D$ of $L$, Wirtinger presentation gives an algorithmic expression of $\pi_1(L)$. For each arc $\alpha_k$ of $D$, we draw a small arrow labeled $a_k$ as in Figure 1, which represents a loop. (The details are in [Rolfsen 1976]. Here we are using the opposite orientation of $a_k$ to be consistent with the operation of the conjugation quandle.) This loop corresponds to one of the meridian curves of the boundary tori, so $\rho(a_k)$ is an element in $\mathcal{P}$. Hence we call $\{\rho(a_1), \dots, \rho(a_n)\}$ the *arc-coloring*[3] of $D$, where each $\rho(a_k)$ is assigned to the corresponding arc $\alpha_k$.

The Wirtinger presentation of the link group is given by

$$\pi_1(L) = \langle a_1, \dots, a_n \,;\, r_1, \dots, r_n \rangle,$$

where the relation $r_l$ is assigned to each crossing as in Figure 2. Note that $r_l$ coincides with (3), so we can write down the relation of the arc-colors as in Figure 3.

From now on, we always assume $\rho : \pi_1(L) \to \mathrm{PSL}(2, \mathbb{C})$ is a given boundary-parabolic representation. To avoid redundant notations, arc-coloring will be denoted by $\{a_1, \dots, a_n\}$ without indicating $\rho$ from now on. Choose an element $s_f \in \mathcal{P}$

[2] We always assume the diagram does not contain a trivial knot component which has only over-crossings or under-crossings or no crossing. (For example, any inseparable link diagram satisfies this condition.) If it happens, then we change the diagram of the trivial component slightly. For example, applying a Reidemeister second move to make different types of crossings or a Reidemeister first move to add a kink is good enough. This assumption is necessary to guarantee that the octahedral triangulation becomes a topological triangulation of $\mathbb{S}^3 \setminus (L \cup \{\text{two points}\})$

[3] Strictly speaking, an arc-coloring is a map from arcs of $D$ to $\mathcal{P}$, not a set. (A region-coloring, which will be defined below, is also a map from regions of $D$ to $\mathcal{P}$.) However, we abuse the set notation here for convenience.

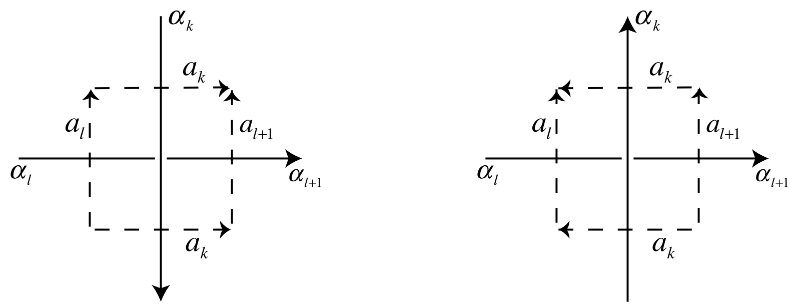



**Figure 2.** Relations at crossings, where $r_l : a_{l+1} = a_k^{-1} a_l a_k$ (left), or $r_l : a_l = a_k^{-1} a_{l+1} a_k$ (right).

corresponding to a region of the diagram $D$ and determine $s_1, s_2, \ldots, s_m \in \mathcal{P}$ corresponding to each regions using the relation in Figure 4.

The assignment of elements of $\mathcal{P}$ to all regions using the relation in Figure 4 is called the *region-coloring*. This assignment is well defined because the two curves in Figure 5, which we call the *cross-changing pair*, determine the same region-coloring, and any pair of curves with the same starting and ending points can be transformed into each other by a finite sequence of cross-changing pairs.

An arc-coloring together with a region-coloring is called a *shadow-coloring*. Lemma 2.4 shows an important property of shadow-colorings, which is crucial for showing the existence of solutions of certain equations.

**Definition 2.3.** The *Hopf map* $h : \mathcal{P} \longrightarrow \mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$ is defined by

$$\begin{pmatrix}\alpha & \beta\end{pmatrix} \mapsto \frac{\alpha}{\beta}.$$

Note that $h\begin{pmatrix}\alpha & \beta\end{pmatrix} = \alpha/\beta$ is the fixed point of the Möbius transformation

$$f(z) = \frac{(1+\alpha\beta)z - \alpha^2}{\beta^2 z + (1-\alpha\beta)}.$$

**Lemma 2.4.** *Let $L$ be a link and assume an arc-coloring is already given by the boundary-parabolic representation $\rho : \pi_1(L) \longrightarrow \mathrm{PSL}(2, \mathbb{C})$. Then there exists a*

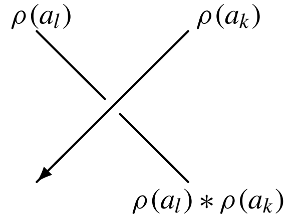



**Figure 3.** An arc-coloring.

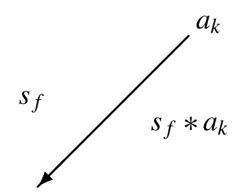



**Figure 4.** A region-coloring.

*region-coloring such that, for any edge of the link diagram with its arc-color* $a_k$ ($k = 1, \ldots, n$) *and its surrounding region-colors* $s_f$, $s_f * a_k$ (*see* Figure 4), *the following holds*:

$$h(a_k) \neq h(s_f) \neq h(s_f * a_k) \neq h(a_k). \tag{5}$$

*Proof.* Note that this was already proved inside the proof of Proposition 2 of [Inoue and Kabaya 2014]. However, finding the proof in the article is not easy, so we write it down below for the readers' convenience.

For the given arc-colors $a_1, \ldots, a_n$, we choose region-colors $s_1, \ldots, s_m$ so that

$$\{h(s_1), \ldots, h(s_m)\} \cap \{h(a_1), \ldots, h(a_n)\} = \varnothing. \tag{6}$$

This is always possible because each $h(s_k)$ is written as $h(s_k) = M_k(h(s_1))$ by a Möbius transformation $M_k$, which only depends on the arc-colors $a_1, \ldots, a_r$. If we choose $h(s_1) \in \mathbb{CP}^1$ away from the finite set

$$\bigcup_{1 \leq k \leq n} \left\{ M_k^{-1}(h(a_1)), \ldots, M_k^{-1}(h(a_r)) \right\},$$

we have $h(s_k) \notin \{h(a_1), \ldots, h(a_r)\}$ for all $k$. This choice of a region-coloring guarantees $h(a_k) \neq h(s_f)$ and $h(s_f * a_k) \neq h(a_k)$.

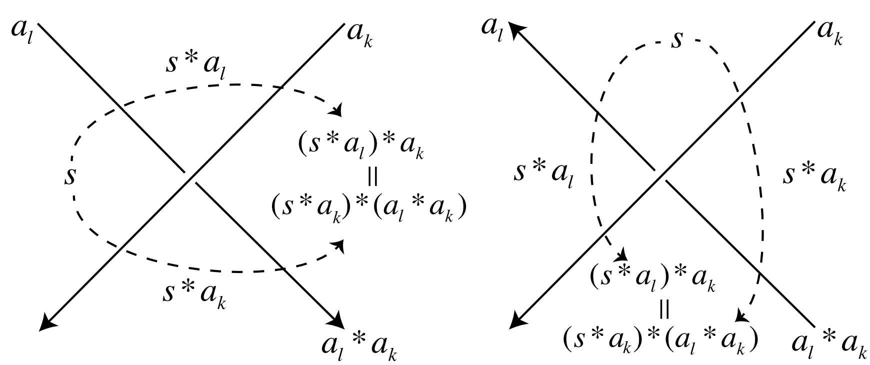



**Figure 5.** Well-definedness of region-coloring for a positive crossing (left) and a negative crossing (right).

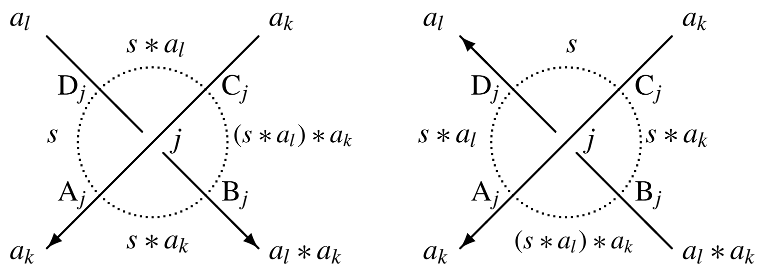



**Figure 6.** Positive (left) and negative (right) crossings of $j$ with shadow-coloring.

Now assume $h(s_f * a_k) = h(s_f)$ holds under the choice of the region-coloring above. Then we obtain

$$h(s_f * a_k) = \widehat{a_k}(h(s_f)) = h(s_f), \tag{7}$$

where $\widehat{a_k} : \mathbb{CP}^1 \to \mathbb{CP}^1$ is the Möbius transformation

$$\widehat{a_k}(z) = \frac{(1 + \alpha_k\beta_k)z - \alpha_k^2}{\beta_k^2 z + (1 - \alpha_k\beta_k)}$$

of $a_k = \begin{pmatrix}\alpha_k & \beta_k\end{pmatrix}$. Then (7) implies $h(s)$ is the fixed point of $\widehat{a_k}$, which means $h(a_k) = h(s)$, which contradicts (6). □

We remark that the condition (6) of a region-coloring is stronger than the condition in Lemma 2.4. For example, the region-colorings of the examples in Section 4 satisfy Lemma 2.4, but they do not satisfy (6). Even though we actually proved the stronger condition (6) in the proof, the region-colorings we consider are always assumed to satisfy Lemma 2.4 from now on. The arc-coloring induced by $\rho$ together with the region-coloring satisfying Lemma 2.4 is called the *shadow-coloring induced by* $\rho$. This shadow-coloring will determine the exact coordinates of points of the octahedral triangulation in the next section.

**2C.** ***Octahedral triangulations of link complements.*** In this section, we describe the ideal triangulation of $\mathbb{S}^3 \setminus (L \cup \{\text{two points}\})$ which appeared in [Cho et al. 2014]. Note that this triangulation naturally arises from the link diagram and has been widely used under various names. For example, the software SnapPea used this triangulation to obtain an ideal triangulation of the link complement $\mathbb{S}^3 \setminus L$ [Weeks 2005] (see also [Yokota 2011].) Another name of this construction is the *tunnel construction* in [Baseilhac and Benedetti 2007]. It seems the first written appearance of this construction was in [Thurston 1999].

To obtain the triangulation, we consider the crossing $j$ in Figure 6 and place an octahedron $A_jB_jC_jD_jE_jF_j$ on each crossing $j$ as in Figure 7 (left). Then we twist the

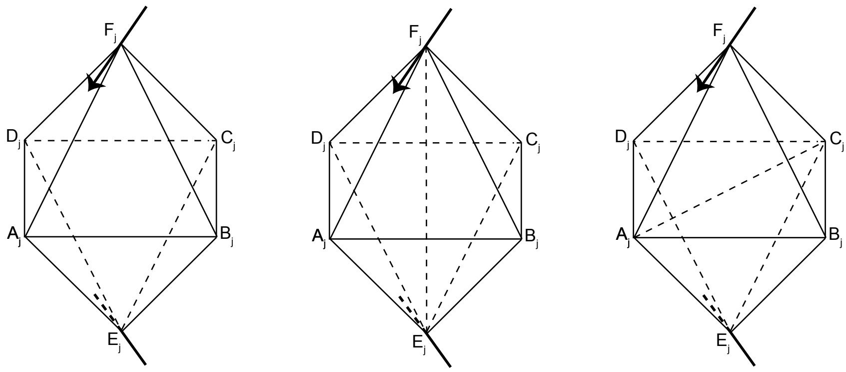



**Figure 7.** An octahedron on the crossing $j$.

octahedron by identifying edges $B_jF_j$ to $D_jF_j$ and $A_jE_j$ to $C_jE_j$, respectively. The edges $A_jB_j$, $B_jC_j$, $C_jD_j$ and $D_jA_j$ are called *horizontal edges* and we sometimes express these edges in the diagram as arcs around the crossing as in Figure 6.

Then we glue faces of the octahedra following the lines of the link diagram. Specifically, there are three gluing patterns as in Figure 8. In each of the cases (left, center and right), we identify the faces

$$\begin{aligned}
&\triangle A_jB_jE_j \cup \triangle C_jB_jE_j \quad \text{with} \quad \triangle C_{j+1}D_{j+1}F_{j+1} \cup \triangle C_{j+1}B_{j+1}F_{j+1},\\
&\triangle B_jC_jF_j \cup \triangle D_jC_jF_j \quad \text{with} \quad \triangle D_{j+1}C_{j+1}F_{j+1} \cup \triangle B_{j+1}C_{j+1}F_{j+1},\\
&\triangle A_jB_jE_j \cup \triangle C_jB_jE_j \quad \text{with} \quad \triangle C_{j+1}B_{j+1}E_{j+1} \cup \triangle A_{j+1}B_{j+1}E_{j+1},
\end{aligned}$$

respectively.

Note that this gluing process identifies vertices $\{A_j, C_j\}$ to one point, denoted by $-\infty$, and $\{B_j, D_j\}$ to another point, denoted by $\infty$, and finally $\{E_j, F_j\}$ to the other points, denoted by $P_t$ where $t = 1, \ldots, c$ and $c$ is the number of the components of the link $L$. The regular neighborhoods of $-\infty$ and $\infty$ are two 3-balls and that of $\bigcup_{t=1}^{c} P_t$ is a tubular neighborhood of the link $L$. Therefore, after removing all vertices of the gluing, we obtain an *octahedral decomposition* of $\mathbb{S}^3 \setminus (L \cup \{\pm\infty\})$. The *octahedral triangulation* is obtained by subdividing each octahedron of the decomposition into four tetrahedra in a certain way.

To apply the construction of the developing map of $\rho$ in Theorem 4.11 of [Zickert 2009], we subdivide each octahedron into four tetrahedra using the shadow-coloring of $\rho$ as follows.

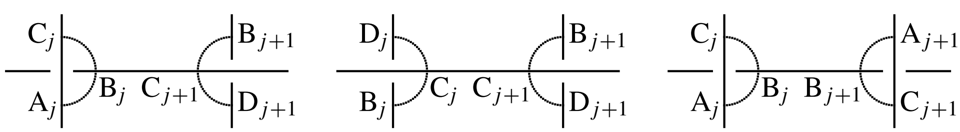



**Figure 8.** Three gluing patterns.

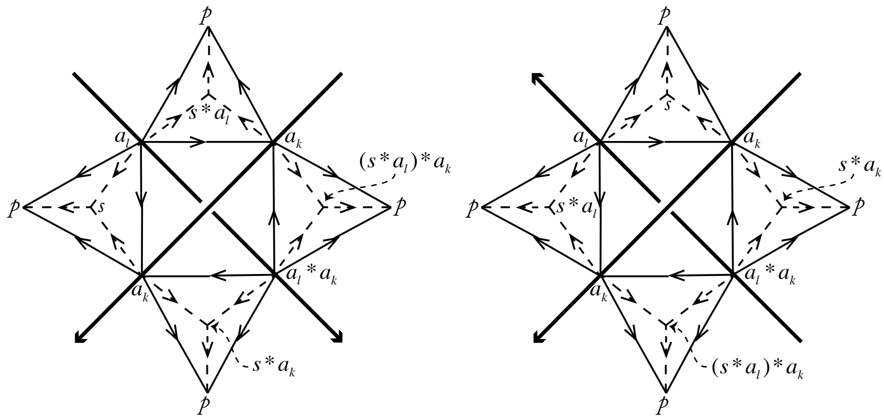



**Figure 9.** Coordinates of tetrahedra when $h(a_k) \neq h(a_l)$ with a positive crossing (left) and a negative cross (right).

**Definition 2.5.** Consider a crossing $j$ with the shadow-coloring in Figure 6. The crossing $j$ is called *nondegenerate* when $h(a_k) \neq h(a_l)$ and *degenerate* when $h(a_k) = h(a_l)$.

If a crossing $j$ is nondegenerate, then we subdivide the octahedron on the crossing $j$ into four tetrahedra by adding the edge $\mathrm{E}_j\mathrm{F}_j$ as in Figure 7 (center). Also, if a crossing $j$ is degenerate, then we subdivide it by adding edge $\mathrm{A}_j\mathrm{C}_j$ as in Figure 7 (right). This subdivision guarantees nondegeneracy of all tetrahedra, which will be proved at the end of this section. The resulting triangulation is called the *octahedral triangulation* of $\mathbb{S}^3 \setminus (L \cup \{\pm\infty\})$.

Consider the shadow-coloring of a link diagram $D$ induced by $\rho$, and let $\{a_1, a_2, \ldots, a_n\}$ be the arc-colors and $\{s_1, s_2, \ldots, s_m\}$ be the region-colors. The number of these colors is finite, so we can choose an element $p \in \mathcal{P}$ satisfying

$$h(p) \notin \{h(a_1), \ldots, h(a_n), h(s_1), \ldots, h(s_m)\}. \tag{8}$$

The geometric shape of the triangulation is determined by the shadow-coloring induced by $\rho$ in the following way. If the crossing $j$ in Figure 6 is nondegenerate and positive, then let the signed coordinates of the tetrahedra $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{B}_j$, and $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{B}_j$ be

$$\begin{gathered}(a_l, a_k, s * a_l, p), \\ -(a_l, a_k, s, p), \\ (a_l * a_k, a_k, s * a_k, p), \\ -(a_l * a_k, a_k, (s * a_l) * a_k, p),\end{gathered} \tag{9}$$

respectively. Here, the minus sign of the coordinate means the orientation of the tetrahedron does not coincide with the one induced by the vertex-ordering. Also, if

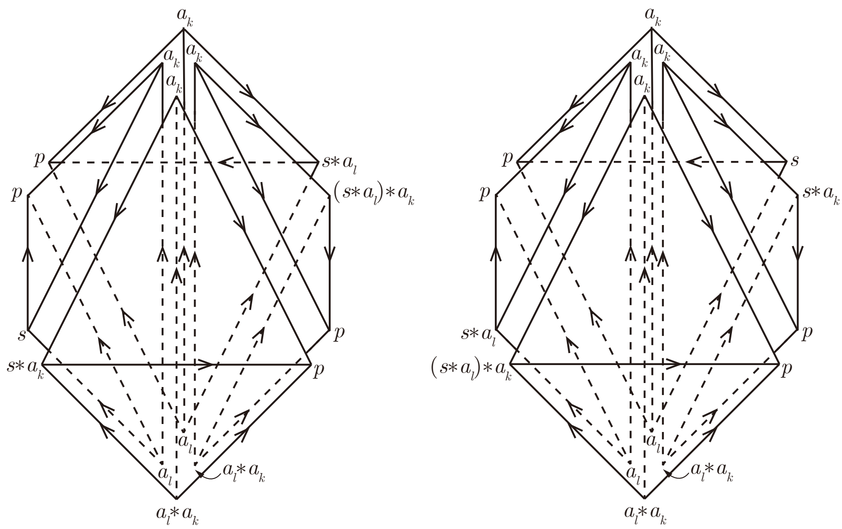



**Figure 10.** Figure 9 in octahedral position for a positive crossing (left) and a negative crossing (right).

the crossing $j$ is nondegenerate and negative, then let the signed coordinates of the tetrahedra $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{B}_j$, and $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{B}_j$ be

$$\begin{gathered}(a_l, a_k, s, p),\\ -(a_l, a_k, s * a_l, p),\\ (a_l * a_k, a_k, (s * a_l) * a_k, p),\\ -(a_l * a_k, a_k, s * a_k, p),\end{gathered} \tag{10}$$

respectively. Figures 9 and 10 show the signed coordinates of (9) and (10).

On the other hand, if the crossing $j$ in Figure 6 is degenerate and is positive, then let the signed coordinates of the tetrahedra $\mathrm{F}_j\mathrm{A}_j\mathrm{C}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{A}_j\mathrm{C}_j\mathrm{D}_j$, $\mathrm{E}_j\mathrm{A}_j\mathrm{C}_j\mathrm{B}_j$, and $\mathrm{F}_j\mathrm{A}_j\mathrm{C}_j\mathrm{B}_j$ be

$$\begin{gathered}-(a_k, s, s * a_l, p),\\ (a_l, s, s * a_l, p),\\ -(a_l * a_k, s * a_k, (s * a_l) * a_k, p),\\ (a_k, s * a_k, (s * a_l) * a_k, p),\end{gathered} \tag{11}$$

respectively. If $j$ is degenerate and negative, then let the signed coordinates be

$$\begin{gathered}-(a_k, s * a_l, s, p),\\ (a_l, s * a_l, s, p),\\ -(a_l * a_k, (s * a_l) * a_k, s * a_k, p),\\ (a_k, (s * a_l) * a_k, s * a_k, p),\end{gathered} \tag{12}$$

respectively.

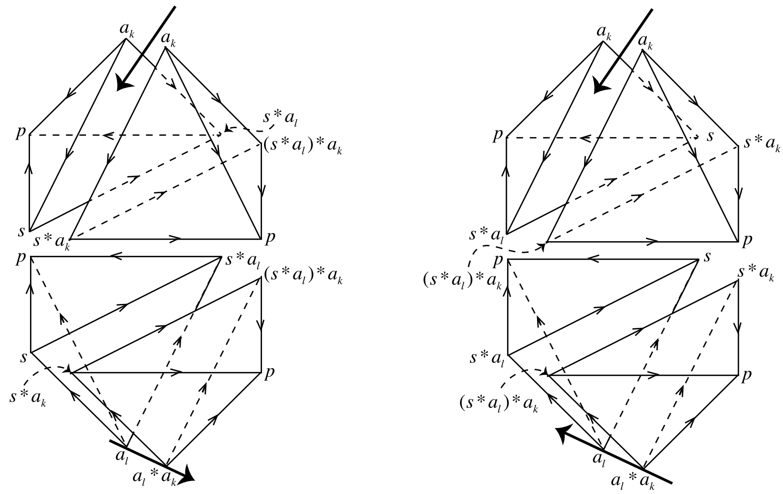



**Figure 11.** Coordinates of tetrahedra when $h(a_k) = h(a_l)$, for a positive crossing (left) and a negative crossing (right).

Figure 11 shows the signed coordinates of (11) and (12). Note that the orientations of (9)–(12) are different from [Inoue and Kabaya 2014] and match [Cho et al. 2014].

We remark that the signed coordinates (9)–(12) actually define an element in certain simplicial quandle homology in [Inoue and Kabaya 2014]. Although this homology is crucial for proving the main results of [Inoue and Kabaya 2014], we will use their results without the homology.

**Definition 2.6.** Let $v_0, v_1, v_2, v_3 \in \mathbb{CP}^1 = \mathbb{C} \cup \{\infty\} = \partial\mathbb{H}^3$. The hyperbolic ideal tetrahedron with signed coordinate $\sigma(v_0, v_1, v_2, v_3)$ with $\sigma \in \{\pm 1\}$ is called *degenerate* when some of the vertices $v_0, v_1, v_2, v_3$ coincide, and *nondegenerate* when all the vertices are different. The *cross-ratio* $[v_0, v_1, v_2, v_3]^\sigma$ of the nondegenerate signed coordinate $\sigma(v_0, v_1, v_2, v_3)$ is defined by

$$[v_0, v_1, v_2, v_3]^\sigma = \left(\frac{v_3 - v_0}{v_2 - v_0}\,\frac{v_2 - v_1}{v_3 - v_1}\right)^\sigma \in \mathbb{C} \setminus \{0, 1\}.$$

The tetrahedra in (9)–(12) have elements of the coordinates in $\mathcal{P}$. Therefore, we need to send them to points in the boundary of the hyperbolic 3-space $\partial\mathbb{H}^3$ so as to obtain hyperbolic ideal tetrahedra. The Hopf map $h$ (see Definition 2.3) plays this role.

**Lemma 2.7.** *The images of* (9)–(12) *under the Hopf map $h$ are nondegenerate tetrahedra. Specifically, if the crossing $j$ is nondegenerate and positive, then*

$$\begin{gathered}(h(a_l), h(a_k), h(s * a_l), h(p)),\\ -(h(a_l), h(a_k), h(s), h(p)),\\ (h(a_l * a_k), h(a_k), h(s * a_k), h(p)),\\ -(h(a_l * a_k), h(a_k), h((s * a_l) * a_k), h(p)),\end{gathered} \tag{13}$$

*are nondegenerate hyperbolic ideal tetrahedra and, if the crossing $j$ is nondegenerate and negative, then*

$$\begin{gathered}(h(a_l), h(a_k), h(s), h(p)),\\ -(h(a_l), h(a_k), h(s*a_l), h(p)),\\ (h(a_l*a_k), h(a_k), h((s*a_l)*a_k), h(p)),\\ -(h(a_l*a_k), h(a_k), h(s*a_k), h(p)),\end{gathered} \tag{14}$$

*are nondegenerate hyperbolic ideal tetrahedra also.*

*If the crossing $j$ is degenerate and positive, then*

$$\begin{gathered}(h(a_l), h(s), h(s*a_l), h(p)),\\ -(h(a_k), h(s), h(s*a_l), h(p)),\\ (h(a_k), h(s*a_k), h((s*a_l)*a_k), h(p)),\\ -(h(a_l*a_k), h(s*a_k), h((s*a_l)*a_k), h(p)),\end{gathered} \tag{15}$$

*are nondegenerate hyperbolic ideal tetrahedra and, if the crossing $j$ is degenerate and negative, then*

$$\begin{gathered}(h(a_l), h(s*a_l), h(s), h(p)),\\ -(h(a_k), h(s*a_l), h(s), h(p)),\\ (h(a_k), h((s*a_l)*a_k), h(s*a_k), h(p)),\\ -(h(a_l*a_k), h((s*a_l)*a_k), h(s*a_k), h(p)),\end{gathered} \tag{16}$$

*are nondegenerate hyperbolic ideal tetrahedra.*

*Proof.* Note that the region-coloring we are considering satisfies Lemma 2.4. To show the nondegeneracy of a tetrahedron, it is enough to show any two endpoints of an edge are different.

In the cases of (13)–(14), endpoints of any edge are adjacent, as a pair among $a_k, s, s*a_k$ in Figure 4 (to check the adjacency, refer to Figure 5), or one of them is $p$, except the edges $(a_l, a_k)$, $(a_l*a_k, a_k)$. Therefore, it is enough to show that $h(a_k) \neq h(a_l)$ implies $h(a_l*a_k) \neq h(a_k)$, which is trivial because $h(a_l*a_k) = h(a_k*a_k)$ implies $h(a_l) = h(a_k)$.

In the cases of (15)–(16), all endpoints of edges are adjacent or one of them is $p$, so we get the proof. □

Note that, when the crossing $j$ is degenerate, the first two tetrahedra in (15) share the same coordinate with different signs and the others do the same. Therefore, all tetrahedra cancel each other out geometrically and we can remove the octahedron of the crossing. (This is why the crossing is called degenerate.) Also, the same holds for (16). This idea will be used in Section 3.

The assignment of the coordinates to tetrahedra above is from [Inoue and Kabaya 2014]. Note that this assignment is based on the construction of the developing

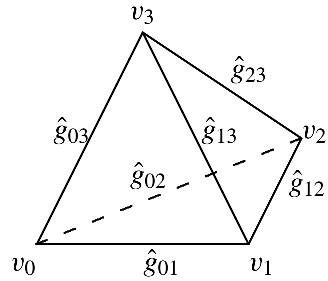



**Figure 12.** Edge parameters.

map of $\rho$ proposed in [Neumann and Yang 1999] and [Zickert 2009], so the shape of the triangulation determines the developing map of $\rho$.

**2D.** ***Complex volume of ρ.*** Consider an ideal tetrahedron with vertices $v_0$, $v_1$, $v_2$, and $v_3$, where $v_k \in \mathbb{CP}^1$. For each edge $v_k v_l$, we assign $g_{kl}$ and $\hat{g}_{kl} \in \mathbb{CP}^1$, and call them *long-edge parameter* and *edge parameter*, respectively. (See Figure 12.) Later, we will distinguish them by considering that $g_{kl}$ is assigned to the edge of a triangulation and $\hat{g}_{kl}$ to the edge of a tetrahedron.

**Definition 2.8.** For the edge parameter $\hat{g}_{kl}$ of an ideal tetrahedron, the *Ptolemy relation* is the following equation:

$$\hat{g}_{02}\hat{g}_{13} = \hat{g}_{01}\hat{g}_{23} + \hat{g}_{03}\hat{g}_{12}.$$

For example, if we define the edge parameter $\hat{g}_{kl} := v_l - v_k$, then direct calculation shows

$$(v_2 - v_0)(v_3 - v_1) = (v_1 - v_0)(v_3 - v_2) + (v_3 - v_0)(v_2 - v_1), \tag{17}$$

which is the Ptolemy relation. Furthermore, these edge parameters satisfy

$$[v_0, v_1, v_2, v_3] = \frac{\hat{g}_{03}\hat{g}_{12}}{\hat{g}_{02}\hat{g}_{13}}. \tag{18}$$

To apply the results of [Zickert 2009] and [Hikami and Inoue 2015], the edge parameters should satisfy the Ptolemy relation, (18) and one more condition that they should depend on the edge of the triangulation, not of the tetrahedron. In other words, if two edges are glued in the triangulation, the edge parameters should be the same. We call this latter condition the *coincidence condition*. When the edge-parameters satisfy the coincidence condition, we call them the *long-edge parameters* and denote this by $g_{kl}$. (We also need the extra condition that the orientations of the two glued edges induced by the vertex-orientations of each tetrahedron should coincide. However, the vertex-orientation in (13)–(16) always satisfies this.) Unfortunately, the edge-parameter $\hat{g}_{kl} = v_l - v_k$ defined above does not satisfy this condition, so we will redefine the edge-parameter and the long-edge parameter using [Inoue and Kabaya 2014] as follows.

At first, consider two elements $a = \begin{pmatrix}\alpha_1 & \alpha_2\end{pmatrix}$, $b = \begin{pmatrix}\beta_1 & \beta_2\end{pmatrix}$ in $\mathcal{P}$. We define the *determinant* $\det(a, b)$ by

$$\det(a, b) := \pm \det \begin{pmatrix}\alpha_1 & \alpha_2 \\ \beta_1 & \beta_2\end{pmatrix} = \pm(\alpha_1\beta_2 - \alpha_2\beta_1).$$

Note that the determinant is defined up to sign due to the choice of the representative $a = \begin{pmatrix}\alpha_1 & \alpha_2\end{pmatrix} = \begin{pmatrix}-\alpha_1 & -\alpha_2\end{pmatrix} \in \mathcal{P}$. To remove this ambiguity, we fix representatives[4] of arc-colors in $\mathbb{C}^2 \setminus \{0\}$ once and for all. Then we fix a representative of one region-color, which uniquely determines the representatives of all the other region-colors by the arc-coloring. (This is due to the fact that $s * (\pm a) = s * a$ for any $s, a \in \mathbb{C}^2 \setminus \{0\}$.)

After fixing all the representatives of the shadow-coloring, we obtain a well-defined determinant

$$\det(a, b) = \det \begin{pmatrix}\alpha_1 & \alpha_2 \\ \beta_1 & \beta_2\end{pmatrix} = \alpha_1\beta_2 - \alpha_2\beta_1. \tag{19}$$

**Lemma 2.9.** *For $a, b, c \in \mathbb{C}^2 \setminus \{0\}$, the determinant satisfies*

$$\det(a * c, b * c) = \det(a, b).$$

*Proof.* Let $a = \begin{pmatrix}\alpha_1 & \alpha_2\end{pmatrix}$, $b = \begin{pmatrix}\beta_1 & \beta_2\end{pmatrix}$, $c = \begin{pmatrix}\gamma_1 & \gamma_2\end{pmatrix}$, and

$$C = \begin{pmatrix}1 + \gamma_1\gamma_2 & \gamma_2^2 \\ -\gamma_1^2 & 1 - \gamma_1\gamma_2\end{pmatrix}.$$

Then

$$\det(a * c, b * c) = \det(aC, bC) = \det(a, b) \cdot \det C = \det(a, b). \qquad \square$$

Consider the shadow-coloring and the coordinates of tetrahedra in Figure 9 (or Figure 10) and Figure 11. We define the edge parameter $\hat{g}_{kl}$ using those coordinates. Specifically, when the signed coordinate of the tetrahedron is $\sigma(a_0, a_1, a_2, a_3)$ with $\sigma \in \{\pm 1\}$ and $a_k \in \mathbb{C}^2 \setminus \{0\}$, we define the edge parameter by

$$\hat{g}_{kl} = \det(a_k, a_l). \tag{20}$$

For example, the edge parameters of the tetrahedron $\mp(a_l, a_k, s, p)$ in the left-hand or the right-hand side of Figure 9 (or Figure 10) are defined by

$$\hat{g}_{01} = \det(a_l, a_k), \quad \hat{g}_{02} = \det(a_l, s), \quad \hat{g}_{03} = \det(a_l, p),$$
$$\hat{g}_{12} = \det(a_k, s), \quad \hat{g}_{13} = \det(a_k, p), \quad \hat{g}_{23} = \det(s, p).$$

[4] The difference in [Inoue and Kabaya 2014] is that they chose a sign of the determinant once and for all. Their choice is good enough to define the long-edge parameter $g_{jk}$, but not for the edge parameter $\hat{g}_{jk}$.

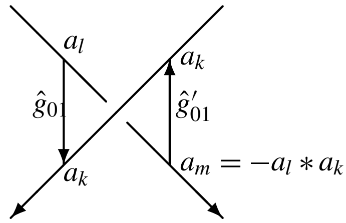



**Figure 13.** An example of the inconsistency of the edge parameter.

**Lemma 2.10.** *The edge parameter* $\hat{g}_{kl}$ *of the tetrahedron* $\sigma(a_0, a_1, a_2, a_3)$ *defined in* (20) *satisfies the Ptolemy identity and*

$$[h(a_0), h(a_1), h(a_2), h(a_3)] = \frac{\hat{g}_{03}\hat{g}_{12}}{\hat{g}_{02}\hat{g}_{13}}. \tag{21}$$

*Proof.* From (19), we obtain

$$h(x) - h(y) = \frac{x_1}{x_2} - \frac{y_1}{y_2} = \frac{\det(x, y)}{x_2 y_2}, \tag{22}$$

where $x = \begin{pmatrix} x_1 & x_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_1 & y_2 \end{pmatrix}$.

Let $a_k = \begin{pmatrix} \alpha_k & \beta_k \end{pmatrix}$ for $k = 0, \ldots, 3$, and let $v_k = h(a_k) = \alpha_k/\beta_k$. Then (17) and (22) imply

$$\frac{\det(a_0, a_2)}{\beta_0\beta_2}\frac{\det(a_1, a_3)}{\beta_1\beta_3} = \frac{\det(a_0, a_1)}{\beta_0\beta_1}\frac{\det(a_2, a_3)}{\beta_2\beta_3} + \frac{\det(a_0, a_3)}{\beta_0\beta_3}\frac{\det(a_1, a_2)}{\beta_1\beta_2},$$

which is equivalent to the Ptolemy identity $\hat{g}_{02}\hat{g}_{13} = \hat{g}_{01}\hat{g}_{23} + \hat{g}_{03}\hat{g}_{12}$.

Also, using (22), we obtain

$$[h(a_0), h(a_1), h(a_2), h(a_3)] = \frac{\dfrac{\det(a_0, a_3)}{\beta_0\beta_3}\dfrac{\det(a_1, a_2)}{\beta_1\beta_2}}{\dfrac{\det(a_1, a_3)}{\beta_1\beta_3}\dfrac{\det(a_0, a_2)}{\beta_0\beta_2}} = \frac{\hat{g}_{03}\hat{g}_{12}}{\hat{g}_{02}\hat{g}_{13}}. \qquad \square$$

Note that, by the same calculation as in the proof above, we obtain

$$[h(a_0), h(a_3), h(a_1), h(a_2)] = \frac{\hat{g}_{02}\hat{g}_{13}}{\hat{g}_{01}\hat{g}_{23}}, \quad [h(a_0), h(a_2), h(a_3), h(a_1)] = -\frac{\hat{g}_{01}\hat{g}_{23}}{\hat{g}_{03}\hat{g}_{12}}.$$

If we put $z^\sigma = [h(a_0), h(a_1), h(a_2), h(a_3)]$, using the Ptolemy identity, the above equations are expressed by

$$z^\sigma = \frac{\hat{g}_{03}\hat{g}_{12}}{\hat{g}_{02}\hat{g}_{13}}, \quad \frac{1}{1 - z^\sigma} = \frac{\hat{g}_{02}\hat{g}_{13}}{\hat{g}_{01}\hat{g}_{23}}, \quad 1 - \frac{1}{z^\sigma} = -\frac{\hat{g}_{01}\hat{g}_{23}}{\hat{g}_{03}\hat{g}_{12}}. \tag{23}$$

The edge parameter $\hat{g}_{jk}$ defined above satisfies all needed properties of the long-edge parameter $g_{jk}$ except the *coincidence* , which $\hat{g}_{jk}$ satisfies up to sign. To see this phenomenon, consider the two edges of Figure 9 (left) as in Figure 13, which are glued in the triangulation. Assume the chosen representative of $a_m$ in Figure 13 satisfies $a_m = -a_l * a_k \in \mathbb{C}^2 \setminus \{0\}$. (This actually happens often and is quite important. For example, the minus signs of (49) and (50) in Section 4 show this situation. This scenario will be discussed in depth in a later article.) Then the edge parameters satisfy

$$\hat{g}_{01} = \det(a_l, a_k) = \det(a_l * a_k, a_k) = -\det(a_m, a_k) = -\hat{g}'_{01}.$$

To obtain the long-edge parameter $g_{jk}$, we assign certain signs to the edge parameters

$$g_{jk} = \pm \hat{g}_{jk},$$

so that the consistency property holds. Due to Lemma 6 of [Inoue and Kabaya 2014], any choice of values of $g_{jk}$ determines the same complex volume. Actually, in Section 3, we do not need the exact values of $g_{jk}$, but we use the existence of them.

The relations of the edge parameters in (23) become

$$z^\sigma = \pm \frac{g_{03}g_{12}}{g_{02}g_{13}}, \quad \frac{1}{1-z^\sigma} = \pm \frac{g_{02}g_{13}}{g_{01}g_{23}}, \quad 1 - \frac{1}{z^\sigma} = \pm \frac{g_{01}g_{23}}{g_{03}g_{12}}. \tag{24}$$

Using (24), we define integers $p$ and $q$ by

$$\begin{cases} p\pi i = -\log z^\sigma + \log g_{03} + \log g_{12} - \log g_{02} - \log g_{13}, \\ q\pi i = \log(1 - z^\sigma) + \log g_{02} + \log g_{13} - \log g_{01} - \log g_{23}. \end{cases} \tag{25}$$

Now we consider the tetrahedron with the signed coordinate $\sigma(a_0, a_1, a_2, a_3)$ and the signed triples $\sigma[z^\sigma; p, q] \in \widehat{\mathcal{P}}(\mathbb{C})$. (The *extended pre-Bloch group* is denoted by $\widehat{\mathcal{P}}(\mathbb{C})$ here. For the definition, see Definition 1.6 of [Zickert 2009].) To consider all signed triples corresponding to all tetrahedra in the triangulation, we denote the triple by $\sigma_t[z_t^{\sigma_t}; p_t, q_t]$, where $t$ is the index of tetrahedra. We define a function $\widehat{L} : \widehat{\mathcal{P}}(\mathbb{C}) \to \mathbb{C}/\pi^2\mathbb{Z}$ by

$$[z; p, q] \mapsto \mathrm{Li}_2(z) + \tfrac{1}{2} \log z \log(1-z) + \frac{\pi i}{2}(q \log z + p \log(1-z)) - \frac{\pi^2}{6}, \tag{26}$$

where $\mathrm{Li}_2(z) = -\int_0^z \frac{1}{t} \log(1-t)dt$ is the dilogarithm function. (Well-definedness of $\widehat{L}$ was proved in [Neumann 2004].) Recall that, for a boundary-parabolic representation $\rho$, the hyperbolic volume $\mathrm{vol}(\rho)$ and the Chern–Simons invariant $\mathrm{cs}(\rho)$ were already defined in [Zickert 2009]. We call $\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)$ the *complex volume of* $\rho$. The following theorem is one of the main results of [Inoue and Kabaya 2014].

**Theorem 2.11** [Zickert 2009; Inoue and Kabaya 2014]. *For a given boundary-parabolic representation $\rho$ and the shadow-coloring induced by $\rho$, the complex* *volume of $\rho$ is calculated by*

$$\sum_t \sigma_t \,\widehat{L}[z_t^{\sigma_t};\, p_t, q_t] \equiv i(\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)) \pmod{\pi^2},$$

*where $t$ is over all tetrahedra of the triangulation defined in Section 2C.*

*Proof.* See Theorem 5 of [Inoue and Kabaya 2014]. □

Note that the removal of the tetrahedra in (15) and (16) does not have any effect on the complex volume. For example, if we let $[z;\, p, q]$ and $-[z';\, p', q']$ be the corresponding triples of the tetrahedron $(h(a_l), h(s), h(s * a_l), h(p))$ and $-(h(a_k), h(s), h(s * a_l), h(p))$ in (15), respectively, and put $\{g_{kl}\}$, $\{g'_{kl}\}$ the sets of long-edge parameters of the two tetrahedra, respectively, then, from $h(a_l) = h(a_k)$, we obtain $z = z'$. Furthermore, we can choose long-edge parameters so that $g_{kl} = g'_{kl}$ holds for all pairs of edges sharing the same coordinate, which induces $p = p'$, $q = q'$ and $\widehat{L}[z;\, p, q] - \widehat{L}[z';\, p', q'] = 0$.

## 3. Optimistic limit

In this section, we will use the result of Section 2 to redefine the optimistic limit of [Cho et al. 2014] and construct a solution of $\mathcal{H}$. At first, we consider a given boundary-parabolic representation $\rho$ and fix its shadow-coloring of a link diagram $D$. For the diagram, define the sides of the diagram to be the lines connecting two adjacent crossings. (The word *edge* is more common than *side* here. However, we want to keep the word *edge* for the edges of a triangulation.) For example, the diagram in Figure 14 has eight sides. We assign $z_1, \ldots, z_n$ to sides of $D$ as in Figure 14 and call them *side variables*.

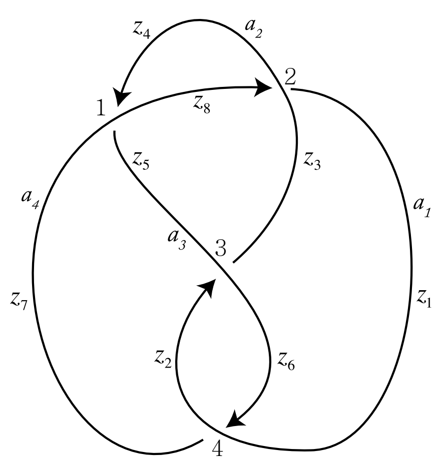



**Figure 14.** Sides of a link diagram.

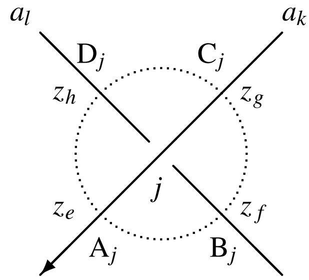



**Figure 15.** A crossing $j$ with arc-colors and side variables.

For the crossing $j$ in Figure 15, let $z_e, z_f, z_g, z_h$ be side variables and let $a_l, a_k$ be the arc-colors. If $h(a_k) \neq h(a_l)$, then we define the potential function $V_j$ of the crossing $j$ by

$$V_j(z_e, z_f, z_g, z_h) = \mathrm{Li}_2\left(\frac{z_f}{z_e}\right) - \mathrm{Li}_2\left(\frac{z_f}{z_g}\right) + \mathrm{Li}_2\left(\frac{z_h}{z_g}\right) - \mathrm{Li}_2\left(\frac{z_h}{z_e}\right). \tag{27}$$

On the other hand, if $h(a_l) = h(a_k)$ in Figure 15, then we introduce new variables $w_e^j, w_f^j, w_g^j$ of the crossing $j$ and define

$$\begin{aligned} &V_j(z_e, z_f, z_g, z_h, w_e^j, w_f^j, w_g^j) \\ &\quad = -\log w_e^j \log z_e + \log w_f^j \log z_f - \log w_g^j \log z_g + \log(w_e^j w_g^j / w_f^j) \log z_h. \end{aligned} \tag{28}$$

For notational convenience, we put $w_h^j := w_e^j w_g^j / w_f^j$. (In (28), we can choose any three variables among $w_e^j, w_f^j, w_g^j, w_h^j$ free variables.) We call the crossing $j$ in Figure 15 *degenerate* when $h(a_l) = h(a_k)$ holds. In particular, when the degenerate crossing forms a kink, as in Figure 16, we put

$$\begin{aligned} &V_j(z_e, z_f, z_g, w_e^j, w_f^j) \\ &\quad = -\log w_e^j \log z_e + \log w_f^j \log z_f - \log w_f^j \log z_f + \log(w_e^j w_f^j / w_f^j) \log z_g \\ &\quad = -\log w_e^j \log z_e + \log w_e^j \log z_g. \end{aligned}$$

Consider the crossing $j$ in Figure 15 and place the octahedron $\mathrm{A}_j\mathrm{B}_j\mathrm{C}_j\mathrm{D}_j\mathrm{E}_j\mathrm{F}_j$ as in Figure 7. When the crossing $j$ is nondegenerate, in other words $h(a_k) \neq h(a_l)$, we consider Figure 7 (center) and assign shape parameters $z_f/z_e$, $z_g/z_f$, $z_h/z_g$ and $z_e/z_h$ to the horizontal edges $\mathrm{A}_j\mathrm{B}_j$, $\mathrm{B}_j\mathrm{C}_j$, $\mathrm{C}_j\mathrm{D}_j$, $\mathrm{D}_j\mathrm{A}_j$, respectively. On the other hand, if the crossing $j$ is degenerate, in other words $h(a_k) = h(a_l)$, then we

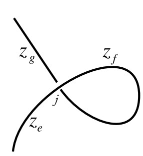



**Figure 16.** A kink.

consider Figure 7 (right) and assign shape parameters $w_e^j$, $w_f^j$, $w_g^j$ and $w_h^j$ to the edges $\mathrm{A}_j\mathrm{F}_j$, $\mathrm{B}_j\mathrm{E}_j$, $\mathrm{C}_j\mathrm{F}_j$ and $\mathrm{D}_j\mathrm{E}_j$, respectively.[5]

The *potential function* $V(z_1, \dots, z_n, w_k^j, \dots)$ of the link diagram $D$ is defined by

$$V(z_1, \dots, z_n, w_k^j, \dots) = \sum_j V_j,$$

where $j$ is over all crossings. For example, if $h(a_1) \neq h(a_2)$ in Figure 14, then $a_4 = a_1 * a_2$ implies[6] $h(a_4) \neq h(a_2)$, $a_2 = a_1 * a_3$ implies[7] $h(a_2) \neq h(a_3) \neq h(a_1)$, $a_2 = a_3 * a_4$ implies $h(a_4) \neq h(a_3)$, $a_4 = a_3 * a_1$ implies $h(a_4) \neq h(a_1)$, and the potential function becomes

$$\begin{aligned}(29)\qquad V(z_1, \dots, z_8) = &\left\{\mathrm{Li}_2\left(\frac{z_5}{z_7}\right) - \mathrm{Li}_2\left(\frac{z_5}{z_8}\right) + \mathrm{Li}_2\left(\frac{z_4}{z_8}\right) - \mathrm{Li}_2\left(\frac{z_4}{z_7}\right)\right\} \\ &+ \left\{\mathrm{Li}_2\left(\frac{z_1}{z_3}\right) - \mathrm{Li}_2\left(\frac{z_1}{z_4}\right) + \mathrm{Li}_2\left(\frac{z_8}{z_4}\right) - \mathrm{Li}_2\left(\frac{z_8}{z_3}\right)\right\} \\ &+ \left\{\mathrm{Li}_2\left(\frac{z_3}{z_6}\right) - \mathrm{Li}_2\left(\frac{z_3}{z_5}\right) + \mathrm{Li}_2\left(\frac{z_2}{z_5}\right) - \mathrm{Li}_2\left(\frac{z_2}{z_6}\right)\right\} \\ &+ \left\{\mathrm{Li}_2\left(\frac{z_6}{z_1}\right) - \mathrm{Li}_2\left(\frac{z_6}{z_2}\right) + \mathrm{Li}_2\left(\frac{z_7}{z_2}\right) - \mathrm{Li}_2\left(\frac{z_7}{z_1}\right)\right\}.\end{aligned}$$

Note that, if $h(a_l) \neq h(a_k)$ for any crossing $j$ in Figure 15, then the definition of the potential function above coincides with the definition in Section 2 of [Cho et al. 2014]. Therefore, the above definition is a slight modification of the previous one.

On the other hand, if $h(a_1) = h(a_2)$ in Figure 14, then $a_1 * a_2 = a_1$. This equation and the relations at crossings induce[8] $a_1 = a_2 = a_3 = a_4$, and the potential function becomes

$$\begin{aligned}V(z_1, \dots, z_8, w_8^1, w_4^1, w_7^1, w_4^2, w_8^2, w_3^2, w_6^3, w_3^3, w_5^3, w_2^4, w_7^4, w_1^4) = \\ -\log w_8^1 \log z_8 + \log w_4^1 \log z_4 - \log w_7^1 \log z_7 + \log w_5^1 \log z_5 \\ -\log w_4^2 \log z_4 + \log w_8^2 \log z_8 - \log w_3^2 \log z_3 + \log w_1^2 \log z_1 \\ -\log w_6^3 \log z_6 + \log w_3^3 \log z_3 - \log w_5^3 \log z_5 + \log w_2^3 \log z_2 \\ -\log w_2^4 \log z_2 + \log w_7^4 \log z_7 - \log w_1^4 \log z_1 + \log w_6^4 \log z_6,\end{aligned}$$

[5] Note that, when $h(a_k) = h(a_l)$, by adding one more edge $\mathrm{B}_j\mathrm{D}_j$ to Figure 7 (right), we obtain another subdivision of the octahedron with five tetrahedra. (This subdivision was already used in [Cho 2016b].) Focusing on the middle tetrahedron that contains all horizontal edges, we obtain $w_e^j w_g^j = w_f^j w_h^j$. Furthermore, the shape-parameters assigned to $\mathrm{D}_j\mathrm{F}_j$ and $\mathrm{B}_j\mathrm{F}_j$ are $(1 - 1/w_e^j)/(1 - w_g^j)$ and $(1 - 1/w_g^j)/(1 - w_e^j)$, respectively.

[6] If $h(a_4) = h(a_2)$, then $h(a_2 * a_2) = h(a_2) = h(a_4) = h(a_1 * a_2)$ induces $h(a_2) = h(a_1)$, which is a contradiction.

[7] If $h(a_2) = h(a_3)$, then $h(a_3 * a_3) = h(a_3) = h(a_2) = h(a_1 * a_3)$ induces $h(a_2) = h(a_3) = h(a_1)$, which is a contradiction. Likewise, if $h(a_1) = h(a_3)$, then $h(a_2) = h(a_1 * a_3) = h(a_1)$ is a contradiction.

[8] The relation $a_4 = a_1 * a_2$ induces $a_4 = a_1$, $a_4 = a_3 * a_1$ induces $a_4 = a_3$, and $a_2 = a_3 * a_4$ induces $a_2 = a_4$.

where $w_5^1 = w_8^1 w_7^1 / w_4^1,\ w_1^2 = w_4^2 w_3^2 / w_8^2,\ w_2^3 = w_6^3 w_5^3 / w_3^3$ and $w_6^4 = w_2^4 w_1^4 / w_7^4$.

For the potential function $V(z_1, \dots, z_n, w_k^j, \dots)$, let $\mathcal{H}$ be the set of equations

$$(30)\quad \mathcal{H} := \left\{ \exp\Bigl(z_k \frac{\partial V}{\partial z_k}\Bigr) = 1, \exp\Bigl(w_k^j \frac{\partial V}{\partial w_k^j}\Bigr) = 1 \;\Big|\; k = 1, \dots, n,\ j : \text{degenerate} \right\},$$

and $\mathcal{S} = \{(z_1, \dots, z_n, w_k^j, \dots)\}$ be the solution set of $\mathcal{H}$. Here, solutions are assumed to satisfy the properties that $z_k \neq 0$ for all $k = 1, \dots, n$ and $z_f/z_e \neq 1,\ z_g/z_f \neq 1,\ z_h/z_g \neq 1,\ z_e/z_h \neq 1,\ z_g/z_e \neq 1,\ z_h/z_f \neq 1$ in Figure 15 for any nondegenerate crossing, and $w_k^j \neq 0$ for any degenerate crossing $j$ and the index $k$. (All these assumptions are essential to avoid singularity of the equations in $\mathcal{H}$ and $\log 0$ in the formula $V_0$ defined in (1). Even though we allow $w_k^j = 1$ here, the value we are interested in always satisfies $w_k^j \neq 1$.)

**Proposition 3.1.** *For the arc-coloring of a link diagram $D$ induced by $\rho$ and the potential function $V(z_1, \dots, z_n, w_k^j, \dots)$, the set $\mathcal{H}$ induces the whole set of hyperbolicity equations of the octahedral triangulation defined in Section 2C.*

The *hyperbolicity equations* consist of Thurston's gluing equations of edges and the completeness condition.

*Proof of Proposition 3.1.* For the case where no crossing is degenerate, this proposition was already proved in Section 3 of [Cho et al. 2014]. To see the main idea, check Figures 10–13 and equations (3.1)–(3.3) of [Cho et al. 2014]. Equation (3.1) is a completeness condition along a meridian of a certain annulus, and (3.2)–(3.3) are gluing equations of certain edges. These three types of equations induce all the other gluing equations.

Therefore, we consider the case when the crossing $j$ in Figure 15 is degenerate. Then, the three equations

$$(31)\quad \exp\left(w_e^j \frac{\partial V}{\partial w_e^j}\right) = \frac{z_h}{z_e} = 1, \exp\left(w_f^j \frac{\partial V}{\partial w_f^j}\right) = \frac{z_f}{z_h} = 1, \exp\left(w_g^j \frac{\partial V}{\partial w_g^j}\right) = \frac{z_h}{z_g} = 1$$

induce $z_e = z_f = z_g = z_h$. This guarantees the gluing equations of horizontal edges trivially by the assigning rule of shape parameters. (Note that the shape parameters assigned to the horizontal edges of the octahedron at a degenerate crossing are always 1.)

There are four possible cases of gluing pattern as in Figure 17, and we assume the crossing $j$ is degenerate and $j+1$ is nondegenerate. (The case when both of $j$ and $j+1$ are degenerate can be proved similarly.)

The part of the potential function $V$ containing $z_k$ in Figure 17 (top left) is

$$V^{(a)} = \log w_k^j \log z_k + \mathrm{Li}_2\Bigl(\frac{z_e}{z_k}\Bigr) - \mathrm{Li}_2\Bigl(\frac{z_f}{z_k}\Bigr),$$

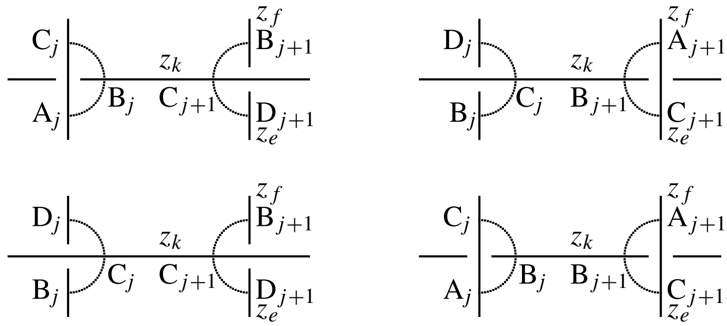



**Figure 17.** Four cases of a gluing pattern.

and

$$\exp\Big(z_k\frac{\partial V}{\partial z_k}\Big)=\exp\Big(z_k\frac{\partial V^{(a)}}{\partial z_k}\Big)=w_k^j\Big(1-\frac{z_e}{z_k}\Big)\Big(1-\frac{z_f}{z_k}\Big)^{-1}=1$$

is equivalent to the following completeness condition

$$\frac{1}{w_k^j}\Big(1-\frac{z_e}{z_k}\Big)^{-1}\Big(1-\frac{z_f}{z_k}\Big)=1$$

along a meridian $m$ in Figure 18 (top left). (Compare it with Figure 11 of [Cho et al. 2014].) Here, $a_j,\ b_j,\ c_j,\ b_{j+1},\ c_{j+1},\ d_{j+1}$ in Figure 18 (top left) are the points of the cusp diagram, which lie on the edges $\mathrm{A}_j\mathrm{E}_j,\ \mathrm{B}_j\mathrm{E}_j,\ \mathrm{C}_j\mathrm{E}_j,\ \mathrm{B}_{j+1}\mathrm{F}_{j+1},\ \mathrm{C}_{j+1}\mathrm{F}_{j+1},\ \mathrm{D}_{j+1}\mathrm{F}_{j+1}$ of Figure 7 (left), respectively.

The part of the potential function $V$ containing $z_k$ in Figure 17 (top right) is

$$V^{(b)}=-\log w_k^j\log z_k-\mathrm{Li}_2\Big(\frac{z_k}{z_e}\Big)+\mathrm{Li}_2\Big(\frac{z_k}{z_f}\Big),$$

and

$$\exp\Big(z_k\frac{\partial V}{\partial z_k}\Big)=\exp\Big(z_k\frac{\partial V^{(b)}}{\partial z_k}\Big)=\frac{1}{w_k^j}\Big(1-\frac{z_k}{z_e}\Big)\Big(1-\frac{z_k}{z_f}\Big)^{-1}=1$$

is equivalent to the completeness condition

$$\frac{1}{w_k^j}\Big(1-\frac{z_k}{z_f}\Big)^{-1}\Big(1-\frac{z_k}{z_e}\Big)=1$$

along a meridian $m$ in Figure 18 (top right). Here, $b_j,\ c_j,\ d_j,\ a_{j+1},\ b_{j+1},\ c_{j+1}$ in Figure 18 (top right) are the points of the cusp diagram, which lie on the edges $\mathrm{B}_j\mathrm{F}_j$, $\mathrm{C}_j\mathrm{F}_j,\ \mathrm{D}_j\mathrm{F}_j,\ \mathrm{A}_{j+1}\mathrm{E}_{j+1},\ \mathrm{B}_{j+1}\mathrm{E}_{j+1},\ \mathrm{C}_{j+1}\mathrm{E}_{j+1}$ of Figure 7 (left), respectively. (To simplify the cusp diagram in Figure 18 (top right), we subdivided the polygon $\mathrm{A}_j\mathrm{B}_j\mathrm{C}_j\mathrm{D}_j\mathrm{F}_j$ in Figure 7 (right) into three tetrahedra by adding the edge $\mathrm{B}_j\mathrm{D}_j$.)

The part of the potential function $V$ containing $z_k$ in Figure 17 (bottom left) is

$$V^{(c)}=-\log w_k^j\log z_k+\mathrm{Li}_2\Big(\frac{z_e}{z_k}\Big)-\mathrm{Li}_2\Big(\frac{z_f}{z_k}\Big),$$

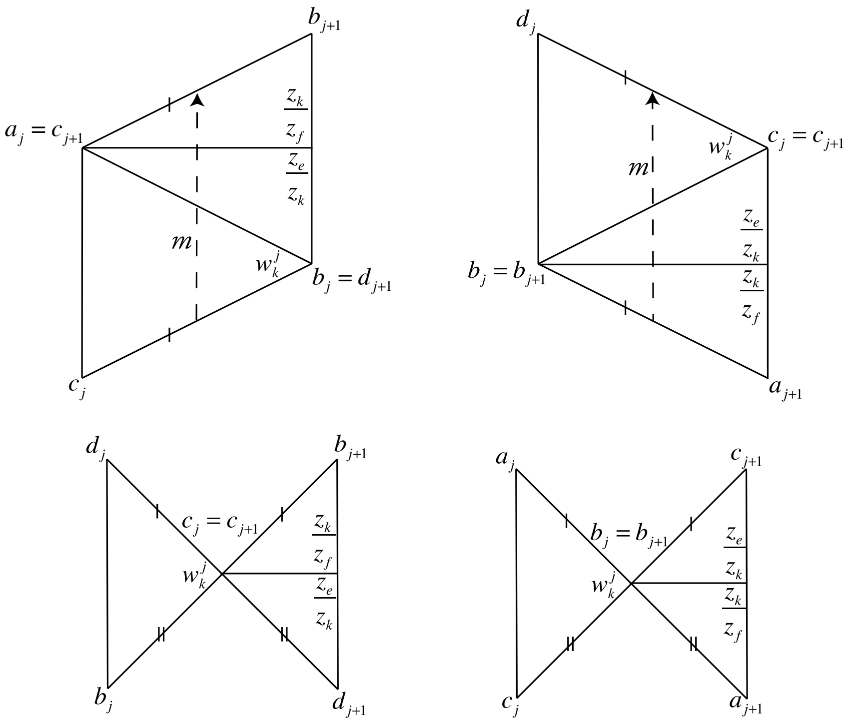



**Figure 18.** Four cusp diagrams from Figure 17.

and

$$\exp\Big(z_k \frac{\partial V}{\partial z_k}\Big) = \exp\Big(z_k \frac{\partial V^{(c)}}{\partial z_k}\Big) = \frac{1}{w_k^j}\Big(1 - \frac{z_e}{z_k}\Big)\Big(1 - \frac{z_f}{z_k}\Big)^{-1} = 1$$

is equivalent to the gluing equation

$$w_k^j\Big(1 - \frac{z_e}{z_k}\Big)^{-1}\Big(1 - \frac{z_f}{z_k}\Big) = 1$$

of $c_j = c_{j+1}$ in Figure 18 (bottom left). (Compare it with Figure 12 of [Cho et al. 2014].) Here, $b_j$, $c_j$, $d_j$, $b_{j+1}$, $c_{j+1}$, $d_{j+1}$ in Figure 18 (bottom left) are the points of the cusp diagram, which lie on the edges $\mathrm{B}_j\mathrm{F}_j$, $\mathrm{C}_j\mathrm{F}_j$, $\mathrm{D}_j\mathrm{F}_j$, $\mathrm{B}_{j+1}\mathrm{F}_{j+1}$, $\mathrm{C}_{j+1}\mathrm{F}_{j+1}$, $\mathrm{D}_{j+1}\mathrm{F}_{j+1}$ of Figure 7 (left), respectively, and the edges $d_j c_j$ and $b_j c_j$ are identified to $b_{j+1}c_{j+1}$ and $d_{j+1}c_{j+1}$, respectively. (To simplify the cusp diagram in Figure 18 (bottom left), we subdivided the polygon $\mathrm{A}_j\mathrm{B}_j\mathrm{C}_j\mathrm{D}_j\mathrm{F}_j$ in Figure 7 (right) into three tetrahedra by adding the edge $\mathrm{B}_j\mathrm{D}_j$.)

The part of the potential function $V$ containing $z_k$ in Figure 17 (bottom right) is

$$V^{(d)} = \log w_k^j \log z_k - \mathrm{Li}_2\Big(\frac{z_k}{z_e}\Big) + \mathrm{Li}_2\Big(\frac{z_k}{z_f}\Big),$$

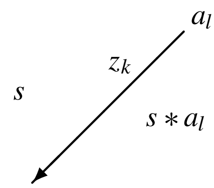



**Figure 19.** A region-coloring.

and

$$\exp\Big(z_k \frac{\partial V}{\partial z_k}\Big) = \exp\Big(z_k \frac{\partial V^{(d)}}{\partial z_k}\Big) = w_k^j \Big(1 - \frac{z_k}{z_e}\Big)\Big(1 - \frac{z_k}{z_f}\Big)^{-1} = 1$$

is equivalent to the gluing equation

$$w_k^j \Big(1 - \frac{z_k}{z_e}\Big)\Big(1 - \frac{z_k}{z_f}\Big)^{-1} = 1$$

of $b_j = b_{j+1}$ in Figure 18 (bottom right). (Compare it with Figure 13 of [Cho et al. 2014].) Here, $a_j,\ b_j,\ c_j,\ a_{j+1},\ b_{j+1},\ c_{j+1}$ in Figure 18 (bottom right) are the points of the cusp diagram, which lie on the edges $\mathrm{A}_j\mathrm{E}_j,\ \mathrm{B}_j\mathrm{E}_j,\ \mathrm{C}_j\mathrm{E}_j,\ \mathrm{A}_{j+1}\mathrm{E}_{j+1},\ \mathrm{B}_{j+1}\mathrm{E}_{j+1},\ \mathrm{C}_{j+1}\mathrm{E}_{j+1}$ of Figure 7 (left), respectively, and the edges $a_jb_j$ and $c_jb_j$ are identified to $c_{j+1}b_{j+1}$ and $a_{j+1}b_{j+1}$, respectively.

Note that the case when both of the crossings $j$ and $j+1$ in Figure 17 are degenerate can be proved in the same way.

On the other hand, it was already shown in [Cho et al. 2014] that all hyperbolicity equations are induced by these types of equations (see the discussion that follows Lemma 3.1 of [Cho et al. 2014]), so the proof is done. □

In [Cho et al. 2014], we could not prove the existence of a solution of $\mathcal{H}$, in other words $\mathcal{S} \neq \varnothing$, so we assumed it. However, the following theorem proves the existence by directly constructing one solution from the given boundary-parabolic representation $\rho$ together with the shadow-coloring.

**Theorem 3.2.** *Consider a shadow-coloring of a link diagram $D$ induced by $\rho$ and the potential function $V(z_1, \ldots, z_n, w_k^j, \ldots)$ from $D$. For each side of $D$ with the side variable $z_k$, arc-color $a_l$ and the region-color $s$, as in Figure 19, we define*

$$z_k^{(0)} := \frac{\det(a_l, p)}{\det(a_l, s)}. \tag{32}$$

*Also, if the positive crossing $j$ in Figure 20 (left) is degenerate, then we define*

$$(w_e^j)^{(0)} := \frac{\det(s, p)}{\det(s * a_k, p)}, \qquad (w_f^j)^{(0)} := \frac{\det((s * a_l) * a_k, p)}{\det(s * a_k, p)},$$
$$(w_g^j)^{(0)} := \frac{\det((s * a_l) * a_k, p)}{\det(s * a_l, p)}, \qquad (w_h^j)^{(0)} := \frac{\det(s, p)}{\det(s * a_l, p)}, \tag{33}$$

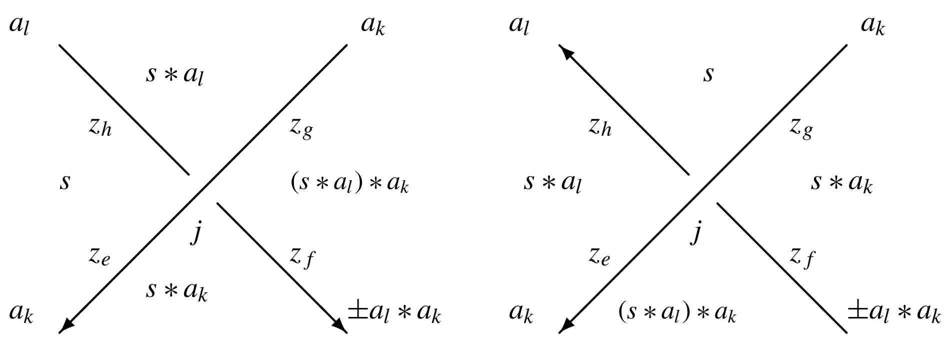



**Figure 20.** Crossings with shadow-colors and side-variables for a positive crossing (left) and a negative crossing (right).

*and, if the negative crossing $j$ in Figure 20 (right) is degenerate, then we define*

$$(w_e^j)^{(0)} := \frac{\det(s * a_l,\, p)}{\det((s * a_l) * a_k,\, p)}, \quad (w_f^j)^{(0)} := \frac{\det(s * a_k,\, p)}{\det((s * a_l) * a_k,\, p)},$$
$$(w_g^j)^{(0)} := \frac{\det(s * a_k,\, p)}{\det(s,\, p)}, \quad (w_h^j)^{(0)} := \frac{\det(s * a_l,\, p)}{\det(s,\, p)}.$$

*Then $z_k^{(0)} \neq 0, 1, \infty$, $(w_k^j)^{(0)} \neq 0, 1$ for all possible $j$, $k$, and*

$$(z_1^{(0)}, \ldots, z_n^{(0)}, (w_k^j)^{(0)}, \ldots) \in \mathcal{S}.$$

Note that the $\pm$ signs in the arc-colors of Figure 20 appear due to the representatives of the colors in $\mathbb{C}^2 \setminus \{0\}$. However, $\pm$ does not change the value of $z_k^{(0)}$ because

$$\frac{\det(\pm a_l,\, p)}{\det(\pm a_l,\, s)} = \frac{\det(a_l,\, p)}{\det(a_l,\, s)} = z_k^{(0)}.$$

Likewise, the value of $(w_k^j)^{(0)}$ does not depend on the choice of $\pm$ because the representatives of region-colors are uniquely determined from the fact $s*(\pm a) = s*a$ for any $s, a \in \mathbb{C}^2 \setminus \{0\}$.

*Proof of Theorem 3.2.* First, when the crossing $j$ in Figure 20 is degenerate, we will show

$$z_e^{(0)} = z_f^{(0)} = z_g^{(0)} = z_h^{(0)}, \tag{34}$$

which satisfies (31). Using $h(a_k) = h(a_l)$, we put $a_k = \begin{pmatrix}\alpha & \beta\end{pmatrix}$ and $a_l = \begin{pmatrix}c\,\alpha & c\,\beta\end{pmatrix} = c\,a_k$ for some constant $c \in \mathbb{C} \setminus \{0\}$. Then we obtain $a_l * a_k = a_l$ and, if $j$ is a positive crossing, then

$$z_e^{(0)} = \frac{c\ \det(a_k, p)}{c\ \det(a_k, s)} = \frac{\det(a_l, p)}{\det(a_l, s)} = z_h^{(0)},$$
$$z_f^{(0)} = \frac{\det(\pm a_l * a_k, p)}{\det(\pm a_l * a_k, s * a_k)} = \frac{\det(a_l * a_k, p)}{\det(a_l * a_k, s * a_k)} = \frac{\det(a_l, p)}{\det(a_l, s)} = z_h^{(0)},$$
$$z_g^{(0)} = \frac{c\ \det(a_k, p)}{c\ \det(a_k, s * a_l)} = \frac{\det(a_l, p)}{\det(a_l, s * a_l)} = z_h^{(0)}.$$

If $j$ is a negative crossing, then by exchanging the indices $e \leftrightarrow g$ in the above calculation, we obtain the same result.

Note that Lemma 2.4 and the definition of $p$ in Section 2C guarantee $z_k^{(0)} \neq 0, 1, \infty$ and $(w_k^j)^{(0)} \neq 0, 1$, so we will concentrate on proving

$$(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots) \in \mathcal{S}.$$

Consider the positive crossing $j$ in Figure 20 (top left) and assume it is nondegenerate. Also consider the tetrahedra in Figures 9 (left) and 10 (left), and assign variables $z_e, z_f, z_g, z_h$ to sides of the link diagram as in Figure 20 (top left). Then, using (21) and (32), the shape parameters assigned to the horizontal edges $\mathrm{A}_j\mathrm{B}_j$ and $\mathrm{D}_j\mathrm{A}_j$ are

$$1 \neq [h(s * a_k), h(p), h(\pm a_l * a_k), h(a_k)]$$
$$= \frac{\det(s, a_k)}{\det(s * a_k, \pm a_l * a_k)} \frac{\det(p, \pm a_l * a_k)}{\det(p, a_k)} = \frac{z_f^{(0)}}{z_e^{(0)}},$$
$$1 \neq [h(s), h(p), h(a_k), h(a_l)] = \frac{\det(s, a_l)}{\det(s, a_k)} \frac{\det(p, a_k)}{\det(p, a_l)} = \frac{z_e^{(0)}}{z_h^{(0)}},$$

respectively. Likewise, the shape parameters assigned to $\mathrm{B}_j\mathrm{C}_j$ and $\mathrm{C}_j\mathrm{D}_j$ are $z_g^{(0)}/z_f^{(0)}$ and $z_h^{(0)}/z_g^{(0)}$ respectively. Furthermore, for any $a, b \in \mathbb{C}^2 \setminus \{0\}$, we can easily show that $h(a * b - a) = h(b)$. If $z_g^{(0)}/z_e^{(0)} = \det(a_k, s)/\det(a_k, s * a_l) = 1$, then $h(a_k) = h(s * a_l - s) = h(a_l)$, which is contradiction. Therefore, we obtain $z_g^{(0)}/z_e^{(0)} \neq 1$, and $z_h^{(0)}/z_f^{(0)} \neq 1$ can be obtained similarly.

We can verify the same holds for nondegenerate negative crossings $j$ in the same way.

Now consider the case when the positive crossing $j$ in Figure 20 (top left) is degenerate. (See Figures 7 (right) and 11 (left).) Then, using (21) and (33), the shape parameters assigned to the edges $\mathrm{F}_j\mathrm{A}_j$, $\mathrm{E}_j\mathrm{B}_j$, $\mathrm{F}_j\mathrm{C}_j$ and $\mathrm{E}_j\mathrm{D}_j$ in Figure 7 (right) are

$$[h(a_k), h(s), h(p), h(s * a_l)][h(a_k), h(s * a_k), h((s * a_l) * a_k), h(p)]$$
$$= \frac{\det(s, p)}{\det(s * a_k, p)} = (w_e^j)^{(0)},$$

$$[h(\pm a_l * a_k), h(p), h((s * a_l) * a_k), h(s * a_k)] = \frac{\det(p, (s * a_l) * a_k)}{\det(p, s * a_k)} = (w_f^j)^{(0)},$$

$$[h(a_k), h((s*a_l)*a_k), h(p), h(s*a_k)][h(a_k), h(s*a_l), h(s), h(p)] = \frac{\det((s*a_l)*a_k, p)}{\det(s*a_l, p)} = (w_g^j)^{(0)},$$

$$[h(a_l), h(p), h(s), h(s*a_l)] = \frac{\det(p, s)}{\det(p, s*a_l)} = (w_h^j)^{(0)},$$

respectively. We can verify the same holds for degenerate negative crossings $j$ in the same way.

Therefore $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ satisfies the hyperbolicity equations of octahedral triangulation defined in Section 2C and, from Proposition 3.1, we get that $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ is a solution of $\mathcal{H}$. By the definition of $\mathcal{S}$, we obtain $(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots) \in \mathcal{S}$. □

To get the complex volume of $\rho$ from the potential function $V(z_1, \dots, z_n, (w_k^j), \dots)$, we modify it to

$$V_0(z_1, \dots, z_n, (w_k^j), \dots) := V(z_1, \dots, z_n, (w_k^j), \dots) - \sum_k \Big(z_k \frac{\partial V}{\partial z_k}\Big) \log z_k - \sum_{\substack{j:\text{degenerate} \\ k}} \Big(w_k^j \frac{\partial V}{\partial w_k^j}\Big) \log w_k^j. \tag{35}$$

This modification guarantees the invariance of the value under the choice of any log-branch. (See Lemma 2.1 of [Cho et al. 2014].) Note that $V_0(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots)$ means the evaluation of the function $V_0(z_1, \dots, z_n, (w_k^j), \dots)$ at

$$(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots).$$

**Theorem 3.3.** *Consider a hyperbolic link $L$, the shadow-coloring induced by $\rho$, the potential function $V(z_1, \dots, z_n, (w_k^j), \dots)$ and the solution*

$$(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots) \in \mathcal{S}$$

*defined in Theorem 3.2. Then,*

$$V_0(z_1^{(0)}, \dots, z_n^{(0)}, (w_k^j)^{(0)}, \dots) \equiv i(\text{vol}(\rho) + i\,\text{cs}(\rho)) \pmod{\pi^2}. \tag{36}$$

*Proof.* When the crossing $j$ is degenerate, direct calculation shows that the potential function $V_j$ of the crossing defined at (28) satisfies

$$(V_j)_0(z, z, z, z, w_1, w_2, w_3) = 0, \tag{37}$$

for any nonzero values of $z, w_1, w_2, w_3$. To simplify the potential function, we rearrange the side variables $z_1, \dots, z_n$ to $z_1, \dots, z_r,\ z_{r+1}, z_{r+1}^1, z_{r+1}^2, z_{r+1}^3, \dots, z_t, \dots, z_t^3$ so that all endpoints of sides with variables $z_1, \dots, z_r$ are nondegenerate crossings and the degenerate crossings induce $z_{r+1}^{(0)} = (z_{r+1}^1)^{(0)} = (z_{r+1}^2)^{(0)} = (z_{r+1}^3)^{(0)},\ \dots, z_t^{(0)} = \dots = (z_t^3)^{(0)}$. (Refer to (34).) Then we define the simplified potential function $\widehat{V}$ by

$$\widehat{V}(z_1,\dots,z_t) := \sum_{j:\text{nondegenerate}} V_j(z_1,\dots,z_r,z_{r+1},z_{r+1},z_{r+1},z_{r+1},\dots,z_t,z_t,z_t,z_t).$$

Note that $\widehat{V}$ is obtained from $V$ by removing the potential functions (28) of the degenerate crossings and substituting the side variables $z_e, z_f, z_g, z_h$ around the degenerate crossing with $z_e$. From (37), we have

$$\widehat{V}_0(z_1^{(0)},\dots,z_t^{(0)}) = V_0(z_1^{(0)},\dots,z_n^{(0)},(w_k^j)^{(0)},\dots),$$

which suggests $\widehat{V}$ is just a simplification of $V$ with the same value. Therefore, from now on, we will use only $\widehat{V}$ and substitute the side variables of the link diagram $z_{r+1}^1, z_{r+1}^2, z_{r+1}^3$ with $z_{r+1}$ and $z_t^1,\dots,z_t^3$ with $z_t$, etc, except at Lemma 3.4 below. Also, we remove octahedra (15) or (16) placed at all degenerate crossings (in other words, the octahedra in Figure 10) because they do not have any effect on the complex volume. (See the comment below the proof of Theorem 2.11.)

Now we will follow ideas of the proof of Theorem 1.2 in [Cho et al. 2014]. However, due to the degenerate crossings, we will improve the proof to cover more general cases. At first, we define $r_k$ by

$$r_k\pi i = z_k \frac{\partial \widehat{V}}{\partial z_k}\bigg|_{z_1=z_1^{(0)},\dots,z_t=z_t^{(0)}}, \tag{38}$$

for $k=1,\dots,t$, where $|_{z_1=z_1^{(0)},\dots,z_t=z_t^{(0)}}$ means the evaluation of the equation at $(z_1^{(0)},\dots,z_t^{(0)})$. Unlike [Cho et al. 2014], we cannot guarantee $r_k$ is an even integer yet, so we need the following lemma.

**Lemma 3.4.** *For the value $z_k^{(0)}$ defined in Theorem 3.2, $(z_1^{(0)},\dots,z_t^{(0)})$ is a solution of the set of equations*

$$\widehat{\mathcal{H}} = \left\{\exp\Big(z_k\frac{\partial \widehat{V}}{\partial z_k}\Big) = 1 \;\Big|\; k=1,\dots,t\right\}.$$

*Proof.* For a degenerate crossing $j$, from (28),

$$V_j(z_k,z_k,z_k,z_k,w_e^j,w_f^j,w_g^j) = (-\log w_e^j + \log w_f^j - \log w_g^j + \log w_h^j)\log z_k.$$

Therefore, using $w_f^j w_h^j/(w_e^j w_g^j) = 1$, we obtain

$$\exp\Big(z_k\frac{\partial V_j}{\partial z_k}(z_k,z_k,z_k,z_k,w_e^j,w_f^j,w_g^j)\Big) = 1.$$

This equation implies that, if we substitute the variables $z_{r+1}^1, z_{r+1}^2, z_{r+1}^3$ with $z_{r+1}$ and $z_t^1,\dots,z_t^3$ with $z_t$, etc., in the equation of $\mathcal{H}$, it becomes $\widehat{\mathcal{H}}$. Thus, Theorem 3.2 induces this lemma. □

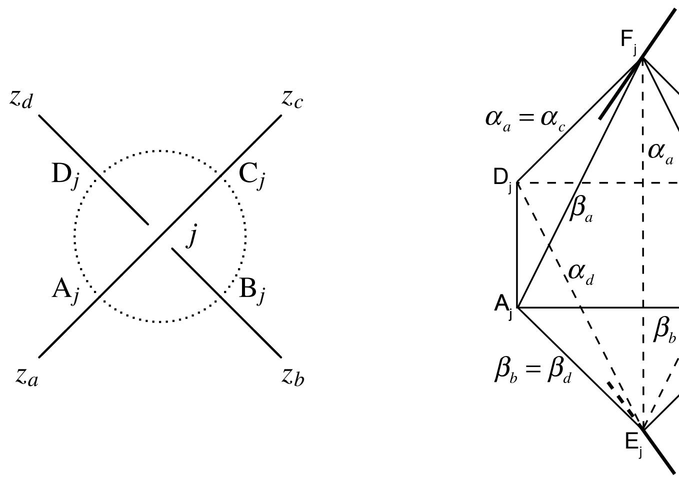



**Figure 21.** Long-edge parameters of nonhorizontal edges.

As a corollary of Lemma 3.4, now we know $r_k$ defined in (38) is an even integer.

To avoid redundant complicated indices, we use $z_k$ instead of $z_k^{(0)}$ in this proof from now on. Using the even integer $r_k$, we can denote $V_0(z_1, \dots, z_t)$ by

$$\widehat{V}_0(z_1, \dots, z_t) = \widehat{V}(z_1, \dots, z_t) - \sum_{k=1}^{t} r_k \pi i \log z_k. \tag{39}$$

Now we introduce notations $\alpha_m$, $\beta_m$, $\gamma_l$, $\delta_j$ for the long-edge parameters defined in (20). We assign $\alpha_m$ and $\beta_m$ to nonhorizontal edges as in Figure 21, where $m$ is over all sides of the link diagram. (Recall that the edges $\mathrm{A}_j\mathrm{B}_j$, $\mathrm{B}_j\mathrm{C}_j$, $\mathrm{C}_j\mathrm{D}_j$ and $\mathrm{D}_j\mathrm{A}_j$ in Figure 21 were named horizontal edges.) We also assign $\gamma_l$ to horizontal edges, where $l$ is over all regions, and $\delta_j$ to the edge $\mathrm{E}_j\mathrm{F}_j$ inside the octahedron. Although we have $\alpha_a = \alpha_c$ and $\beta_b = \beta_d$ because of the gluing, we use $\alpha_a$ for the tetrahedra $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{B}_j$ and $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{D}_j$, $\alpha_c$ for $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{B}_j$ and $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{D}_j$, $\beta_b$ for $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{B}_j$ and $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{B}_j$, and $\beta_d$ for $\mathrm{E}_j\mathrm{F}_j\mathrm{C}_j\mathrm{D}_j$ and $\mathrm{E}_j\mathrm{F}_j\mathrm{A}_j\mathrm{D}_j$. Note that the labeling is consistent even when some crossing is degenerate because, when the crossing $j$ in Figure 21 is degenerate, we obtain $z_a = z_b = z_c = z_d$ and, after removing the octahedron of the crossing, the long-edge parameters satisfy $\alpha_a = \alpha_b = \alpha_c = \alpha_d$ and $\beta_a = \beta_b = \beta_c = \beta_d$.

Now consider a side with variable $z_k$ and two possible cases in Figure 22. We consider the case when the crossing is nondegenerate, or equivalently, $z_a \neq z_k \neq z_b$. (If it is degenerate, we assume there is a degenerated octahedron[9] at the crossing.) For $m = a, b$, let $\sigma_k^m \in \{\pm 1\}$ be the sign of the tetrahedron[10] between the sides $z_k$ and $z_m$, and $u_k^m$ be the shape parameter of the tetrahedron assigned to the horizontal edge. We put $\tau_k^m = 1$ when $z_k$ is the numerator of $(u_k^m)^{\sigma_k^m}$ and $\tau_k^m = -1$ otherwise. We also

[9] An octahedron is called degenerate when two vertices at the top and the bottom coincide.

[10] The sign of a tetrahedron is the sign of the coordinate in (13) or (14).

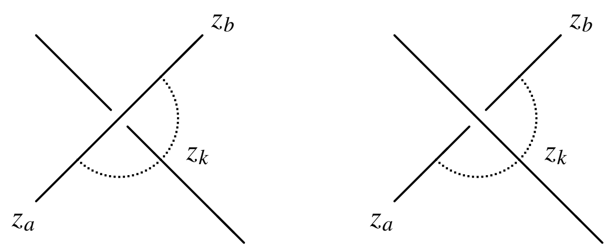



**Figure 22.** Two cases with respect to $z_k$.

define $p_k^m$ and $q_k^m$ by (25) so that $\sigma_k^m[(u_k^m)^{\sigma_k^m}; p_k^m, q_k^m]$ becomes the element of $\widehat{\mathcal{P}}(\mathbb{C})$ corresponding to the tetrahedron. Then $\frac{1}{2}\sum_{1\le k,m\le t}\sigma_k^m[(u_k^m)^{\sigma_k^m}; p_k^m, q_k^m]$ is the element[11] of $\widehat{\mathcal{B}}(\mathbb{C})$ corresponding to the octahedral triangulation in Section 2C, and

$$\frac{1}{2}\sum_{1\le k,m\le t}\sigma_k^m\widehat{L}[(u_k^m)^{\sigma_k^m}; p_k^m, q_k^m]\equiv i(\mathrm{vol}(\rho)+i\,\mathrm{cs}(\rho)) \pmod{\pi^2}, \tag{40}$$

from Theorem 2.11.

By definition, we know

$$u_k^a=\frac{z_k}{z_a},\quad u_k^b=\frac{z_b}{z_k}. \tag{41}$$

In the case of Figure 22 (left), we have

$$\sigma_k^a=1,\sigma_k^b=-1\quad\text{and}\quad\tau_k^a=\tau_k^b=1.$$

Using (25) and Figure 23 (left), we decide $p_k^m$ and $q_k^m$ as follows:

$$\begin{cases}\log(z_k/z_a)+p_k^a\pi i=(\log\alpha_k-\log\beta_k)-(\log\alpha_a-\log\beta_a),\\ \log(z_k/z_b)+p_k^b\pi i=(\log\alpha_k-\log\beta_k)-(\log\alpha_b-\log\beta_b),\end{cases} \tag{42}$$

$$\begin{cases}-\log(1-z_k/z_a)+q_k^a\pi i=\log\beta_k+\log\alpha_a-\log\gamma_1-\log\delta_1,\\ -\log(1-z_k/z_b)+q_k^b\pi i=\log\beta_k+\log\alpha_b-\log\gamma_2-\log\delta_1.\end{cases} \tag{43}$$

In the case of Figure 22 (right), we have

$$\sigma_k^a=-1,\sigma_k^b=1\quad\text{and}\quad\tau_k^a=\tau_k^b=-1.$$

Using (25) and Figure 23 (right), we decide $p_k^m$ and $q_k^m$ as follows:

$$\begin{cases}\log(z_a/z_k)+p_k^a\pi i=(\log\alpha_a-\log\beta_a)-(\log\alpha_k-\log\beta_k),\\ \log(z_b/z_k)+p_k^b\pi i=(\log\alpha_b-\log\beta_b)-(\log\alpha_k-\log\beta_k),\end{cases} \tag{44}$$

[11] The coefficient $\frac{1}{2}$ appears because the same tetrahedron is counted twice in the summation.

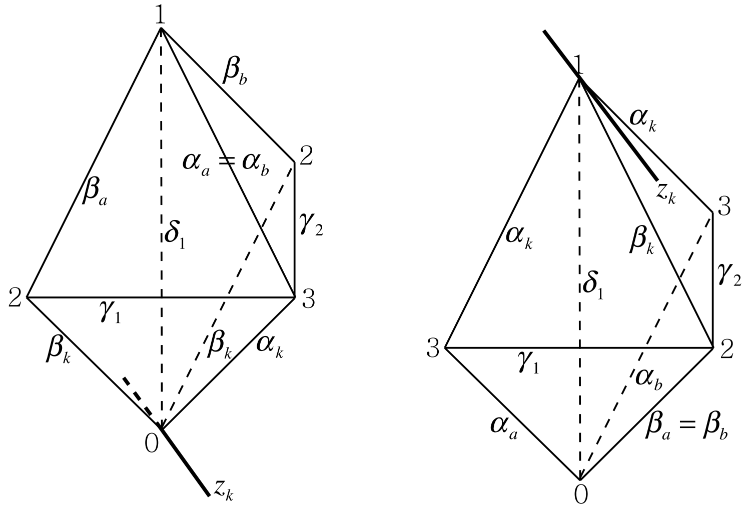



**Figure 23.** Tetrahedra of Figure 22.

$$\begin{cases} -\log(1 - z_a/z_k) + q_k^a \pi i = \log\beta_a + \log\alpha_k - \log\gamma_1 - \log\delta_1, \\ -\log(1 - z_b/z_k) + q_k^b \pi i = \log\beta_b + \log\alpha_k - \log\gamma_2 - \log\delta_1. \end{cases} \tag{45}$$

The equations (42) and (44) hold for all (nondegenerate and degenerate) crossings, so we get the following observation.

**Observation 3.5.** We have

$$\log\alpha_k - \log\beta_k \equiv \log z_k + A \pmod{\pi i},$$

for all $k = 1, \ldots, t$, where $A$ is a complex constant number independent of $k$.

Note that, by (27), the potential function $\widehat{V}$ is expressed by

$$\widehat{V}(z_1, \ldots, z_t) = \frac{1}{2} \sum_{1 \le k, m \le t} \sigma_k^m \mathrm{Li}_2((u_k^m)^{\sigma_k^m}) = \frac{1}{2} \sum_{k=1}^{t} \sum_{m=a,\ldots,d} \sigma_k^m \mathrm{Li}_2((u_k^m)^{\sigma_k^m}), \tag{46}$$

where the range of the index $m$ is determined by $k$ and we define the range of $m$ by $m = a, \ldots, d$[12] from now on. Recall that $r_k$ was defined in (38). Direct calculation shows

$$r_k \pi i = -\sum_{m=a,\ldots,d} \sigma_k^m \tau_k^m \log(1 - (u_k^m)^{\sigma_k^m}).$$

Combining (43) and (45), we obtain

$$\sum_{m=a,b} \sigma_k^m \tau_k^m \left\{ -\log(1 - (u_k^m)^{\sigma_k^m}) + q_k^m \pi i \right\} = -\log\gamma_1 + \log\gamma_2,$$

[12] The range $m = a, \ldots, d$ means that each side with one of the side variables $z_a, \ldots, z_d$ shares a nondegenerate crossing with a side with $z_k$.

for both cases in Figure 22. (Note that $\alpha_a = \alpha_b$ in (43) and $\beta_a = \beta_b$ in (45).) Therefore, we obtain

$$\sum_{m=a,\dots,d} \sigma_k^m \tau_k^m \left\{ -\log(1-(u_k^m)^{\sigma_k^m}) + q_k^m \pi i \right\} = 0,$$

and

$$r_k \pi i = -\sum_{m=a,\dots,d} \sigma_k^m \tau_k^m q_k^m \pi i. \tag{47}$$

**Lemma 3.6.** *For all possible $k$ and $m$, we have*

$$\frac{1}{2} \sum_{1\le k,m\le t} \sigma_k^m q_k^m \pi i \log(u_k^m)^{\sigma_k^m} \equiv -\sum_{k=1}^{t} r_k \pi i \log z_k \pmod{2\pi^2}. \tag{48}$$

*Proof.* Note that, by definition, $\sigma_k^m = \sigma_m^k$, $\tau_k^m = -\tau_m^k$ and

$$(u_k^m)^{\sigma_k^m} = \left(\frac{z_k}{z_m}\right)^{\tau_k^m} = (z_k)^{\tau_k^m} (z_m)^{\tau_m^k}.$$

Using the above and (47), we can directly calculate

$$\frac{1}{2} \sum_{k=1}^{t} \sum_{m=a,\dots,d} \sigma_k^m q_k^m \pi i \log(u_k^m)^{\sigma_k^m} \equiv \sum_{k=1}^{t} \Big( \sum_{m=a,\dots,d} \sigma_k^m \tau_k^m q_k^m \pi i \Big) \log z_k \pmod{2\pi^2}$$
$$= -\sum_{k=1}^{t} r_k \pi i \log z_k. \qquad \square$$

**Lemma 3.7.** *For all possible $k$ and $m$, we have*

$$\frac{1}{2} \sum_{1\le k,m\le t} \sigma_k^m \log\big(1-(u_k^m)^{\sigma_k^m}\big)\big(\log(u_k^m)^{\sigma_k^m} + p_k^m \pi i\big) \equiv -\sum_{k=1}^{t} r_k \pi i \log z_l \pmod{2\pi^2}.$$

*Proof.* From (42) and (44), we have

$$\log(u_k^m)^{\sigma_k^m} + p_k^m \pi i = \tau_k^m (\log\alpha_k - \log\beta_k) + \tau_m^k (\log\alpha_m - \log\beta_m).$$

Therefore,

$$\frac{1}{2} \sum_{1\le k,m\le t} \sigma_k^m \log\big(1-(u_k^m)^{\sigma_k^m}\big)\big(\log(u_k^m)^{\sigma_k^m} + p_k^m \pi i\big)$$
$$= \sum_{k=1}^{t} \Bigg( \sum_{m=a,\dots,d} \sigma_k^m \tau_k^m \log(1-(u_k^m)^{\sigma_k^m}) \Bigg) (\log\alpha_k - \log\beta_k)$$
$$= -\sum_{k=1}^{t} r_k \pi i (\log\alpha_k - \log\beta_k).$$

Note that

$$\sum_{k=1}^{t} r_k \pi i = \sum_{k=1}^{t} z_k \frac{\partial \widehat{V}}{\partial z_k} = 0$$

because $\widehat{V}$ is expressed by the summation of certain forms of $\mathrm{Li}_2(z_a/z_b)$ and

$$z_a \frac{\partial \mathrm{Li}_2(z_a/z_b)}{\partial z_a} + z_b \frac{\partial \mathrm{Li}_2(z_a/z_b)}{\partial z_b} = -\log\Big(1-\frac{z_a}{z_b}\Big) + \log\Big(1-\frac{z_a}{z_b}\Big) = 0.$$

By using Observation 3.5, the above, and the fact that $r_k$ is even, we have

$$\begin{aligned} -\sum_{k=1}^{t} r_k\pi i(\log\alpha_k - \log\beta_k) &\equiv -\sum_{k=1}^{t} r_k\pi i(\log z_k + A) \\ &= -\sum_{k=1}^{t} r_k\pi i \log z_k \pmod{2\pi^2}. \end{aligned}$$ □

Combining (40), (46), Lemma 3.6 and Lemma 3.7, we complete the proof of Theorem 3.3 as follows:

$$\begin{aligned} &i(\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)) \\ &\quad\equiv \frac{1}{2}\sum_{1\le k,m\le t} \sigma_k^m \widehat{L}[(u_k^m)^{\sigma_k^m};\, p_k^m, q_k^m] \\ &\quad= \frac{1}{2}\sum_{1\le k,m\le t} \sigma_k^m\Big(\mathrm{Li}_2\big((u_k^m)^{\sigma_k^m}\big) - \frac{\pi^2}{6}\Big) + \frac{1}{4}\sum_{1\le k,m\le t} \sigma_k^m q_k^m \pi i \log(u_k^m)^{\sigma_k^m} \\ &\qquad + \frac{1}{4}\sum_{1\le k,m\le t} \sigma_k^m \log\big(1-(u_k^m)^{\sigma_k^m}\big)\big(\log(u_k^m)^{\sigma_k^m} + p_k^m\pi i\big) \\ &\quad\equiv \widehat{V}(z_1,\dots,z_n) - \sum_{k=1}^{t} r_k\pi i \log z_k = \widehat{V}_0(z_1,\dots,z_t) \pmod{\pi^2}. \end{aligned}$$ □

## 4. Examples

**4A.** ***A figure-eight knot* $4_1$.** For the figure-eight knot diagram in Figure 24, let the elements of $\mathcal{P}$ corresponding to the arcs be

$$a_1 = \big(0 \ \ t\big), \quad a_2 = \big(1 \ \ 0\big), \quad a_3 = \big(-t \ \ 1+t\big), \quad a_4 = \big(-t \ \ t\big),$$

where $t$ is a solution of $t^2+t+1=0$. These elements satisfy

$$a_1 * a_2 = a_4, \quad a_3 * a_4 = a_2, \quad a_1 * a_3 = -a_2, \quad a_3 * a_1 = a_4, \tag{49}$$

where the identities are expressed in $\mathbb{C}^2\setminus\{0\}$, not in $\mathcal{P} = (\mathbb{C}^2\setminus\{0\})/\pm$. Let $\rho : \pi_1(4_1) \to \mathrm{PSL}(2,\mathbb{C})$ be the boundary-parabolic representation determined by $a_1,\dots,a_4$. We define the shadow-coloring of Figure 24 induced by $\rho$ by letting

$$\begin{aligned} &s_1 = \big(1 \ \ 1\big), && s_2 = \big(0 \ \ 1\big), && s_3 = \big(-t-1 \ \ t+2\big), \\ &s_4 = \big(-2t-1 \ \ 2t+3\big), && s_5 = \big(-2t-1 \ \ t+4\big), && s_6 = \big(1 \ \ t+2\big), \end{aligned}$$

and $p = \big(2 \ \ 1\big)$. Direct calculation shows this shadow-coloring satisfies (5) in Lemma 2.4. (However, this does not satisfy (6).)

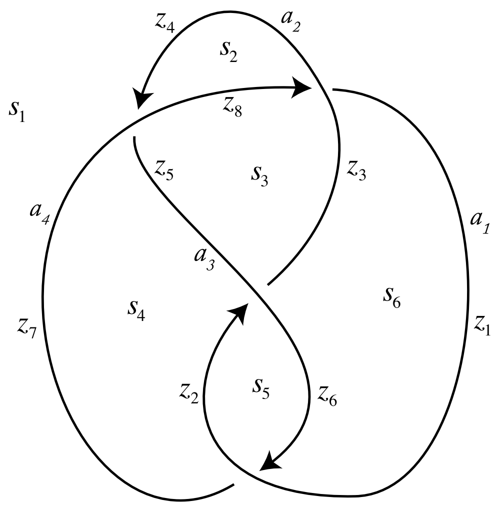



**Figure 24.** A figure-eight knot $4_1$ with parameters.

All values of $h(a_1), \ldots, h(a_4)$ are different, therefore the potential function $V(z_1, \ldots, z_8)$ of Figure 24 is (29). Applying Theorem 3.2, we obtain

$$z_1^{(0)} = \frac{\det(a_1, p)}{\det(a_1, s_6)} = 2, \qquad z_2^{(0)} = \frac{\det(a_1, p)}{\det(a_1, s_5)} = \frac{-2}{2t+1},$$
$$z_3^{(0)} = \frac{\det(a_2, p)}{\det(a_2, s_6)} = \frac{1}{t+2}, \qquad z_4^{(0)} = \frac{\det(a_2, p)}{\det(a_2, s_1)} = 1,$$
$$z_5^{(0)} = \frac{\det(a_3, p)}{\det(a_3, s_4)} = -3t-2, \qquad z_6^{(0)} = \frac{\det(a_3, p)}{\det(a_3, s_5)} = \frac{3t+2}{2t},$$
$$z_7^{(0)} = \frac{\det(a_4, p)}{\det(a_4, s_4)} = \frac{3}{2}, \qquad z_8^{(0)} = \frac{\det(a_4, p)}{\det(a_4, s_3)} = 3,$$

and $(z_1^{(0)}, \ldots, z_8^{(0)})$ becomes a solution of $\mathcal{H} = \{\exp(z_k \frac{\partial V}{\partial z_k}) = 1 \mid k = 1, \ldots, 8\}$. Applying Theorem 3.3, we obtain

$$V_0(z_1^{(0)}, \ldots, z_8^{(0)}) \equiv i(\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)) \pmod{\pi^2},$$

and numerical calculation verifies it by

$$V_0(z_1^{(0)}, \ldots, z_8^{(0)}) = \begin{cases} i(2.0299... + 0\,i) = i(\mathrm{vol}(4_1) + i\,\mathrm{cs}(4_1)) & \text{if } t = \frac{1}{2}(-1-\sqrt{3}\,i), \\ i(-2.0299... + 0\,i) = i(-\mathrm{vol}(4_1) + i\,\mathrm{cs}(4_1)) & \text{if } t = \frac{1}{2}(-1+\sqrt{3}\,i). \end{cases}$$

**4B. *Trefoil knot* $3_1$.** For the trefoil knot diagram in Figure 25, let the elements of $\mathcal{P}$ corresponding to the arcs be

$$a_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 & 1 \end{pmatrix}, \quad a_3 = a_4 = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

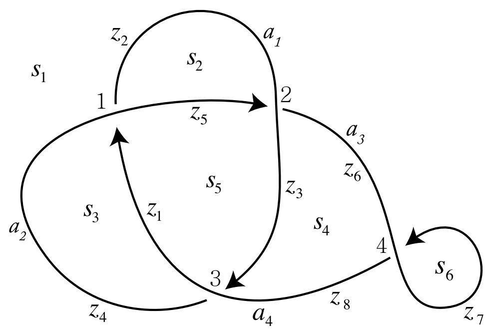



**Figure 25.** A trefoil knot $3_1$ with parameters.

(Note that crossing 4 is degenerate.) These elements satisfy

$$a_4 * a_2 = -a_1, \quad a_2 * a_1 = a_3, \quad a_1 * a_4 = a_2, \quad a_4 * a_3 = a_3, \tag{50}$$

where the identities are expressed in $\mathbb{C}^2 \setminus \{0\}$, not in $\mathcal{P} = (\mathbb{C}^2 \setminus \{0\})/\pm$. Let $\rho : \pi_1(3_1) \to \mathrm{PSL}(2, \mathbb{C})$ be the boundary-parabolic representation determined by $a_1, a_2, a_3, a_4$. We define the shadow-coloring of Figure 24 induced by $\rho$ by letting

$$s_1 = \begin{pmatrix}-1 & 2\end{pmatrix}, \quad s_2 = \begin{pmatrix}1 & 2\end{pmatrix}, \quad s_3 = \begin{pmatrix}-1 & 3\end{pmatrix},$$
$$s_4 = \begin{pmatrix}0 & 1\end{pmatrix}, \quad s_5 = \begin{pmatrix}1 & 1\end{pmatrix}, \quad s_6 = \begin{pmatrix}-2 & 3\end{pmatrix},$$

and $p = \begin{pmatrix}2 & 1\end{pmatrix}$. Direct calculation shows this shadow-coloring satisfies (5) in Lemma 2.4. (However, this does not satisfy (6).)

All values of $h(a_1), h(a_2), h(a_3) = h(a_4)$ are different, hence the potential function $V$ of Figure 25 is

$$\begin{aligned} V(z_1, \dots, z_8, w_6^4, w_7^4) = {} & \mathrm{Li}_2\left(\frac{z_2}{z_5}\right) - \mathrm{Li}_2\left(\frac{z_2}{z_4}\right) + \mathrm{Li}_2\left(\frac{z_1}{z_4}\right) - \mathrm{Li}_2\left(\frac{z_1}{z_5}\right) \\ & + \mathrm{Li}_2\left(\frac{z_6}{z_3}\right) - \mathrm{Li}_2\left(\frac{z_6}{z_2}\right) + \mathrm{Li}_2\left(\frac{z_5}{z_2}\right) - \mathrm{Li}_2\left(\frac{z_5}{z_3}\right) \\ & + \mathrm{Li}_2\left(\frac{z_4}{z_1}\right) - \mathrm{Li}_2\left(\frac{z_4}{z_8}\right) + \mathrm{Li}_2\left(\frac{z_3}{z_8}\right) - \mathrm{Li}_2\left(\frac{z_3}{z_1}\right) \\ & - \log w_6^4 \log z_6 + \log w_6^4 \log z_8, \end{aligned}$$

and the simplified potential function $\widehat{V}$ defined in the proof of Theorem 3.3 is

$$\begin{aligned} \widehat{V}(z_1, \dots, z_6) = {} & \mathrm{Li}_2\left(\frac{z_2}{z_5}\right) - \mathrm{Li}_2\left(\frac{z_2}{z_4}\right) + \mathrm{Li}_2\left(\frac{z_1}{z_4}\right) - \mathrm{Li}_2\left(\frac{z_1}{z_5}\right) \\ & + \mathrm{Li}_2\left(\frac{z_6}{z_3}\right) - \mathrm{Li}_2\left(\frac{z_6}{z_2}\right) + \mathrm{Li}_2\left(\frac{z_5}{z_2}\right) - \mathrm{Li}_2\left(\frac{z_5}{z_3}\right) \\ & + \mathrm{Li}_2\left(\frac{z_4}{z_1}\right) - \mathrm{Li}_2\left(\frac{z_4}{z_6}\right) + \mathrm{Li}_2\left(\frac{z_3}{z_6}\right) - \mathrm{Li}_2\left(\frac{z_3}{z_1}\right). \end{aligned}$$

Applying Theorem 3.2, we obtain

$$z_1^{(0)} = \frac{\det(a_4, p)}{\det(a_4, s_5)} = \frac{3}{2}, \qquad z_2^{(0)} = \frac{\det(a_1, p)}{\det(a_1, s_2)} = \frac{1}{2},$$
$$z_3^{(0)} = \frac{\det(a_1, p)}{\det(a_1, s_5)} = 1, \qquad z_4^{(0)} = \frac{\det(a_2, p)}{\det(a_2, s_3)} = -2,$$
$$z_5^{(0)} = \frac{\det(a_2, p)}{\det(a_2, s_5)} = 2, \qquad z_6^{(0)} = z_7^{(0)} = z_8^{(0)} = \frac{\det(a_3, p)}{\det(a_3, s_4)} = 3,$$
$$(w_6^4)^{(0)} = \frac{\det(s_1, p)}{\det(s_4, p)} = \frac{5}{2}, \quad (w_7^4)^{(0)} = \frac{\det(s_1, p)}{\det(s_6, p)} = \frac{5}{8}.$$

Note that $(z_1^{(0)}, \dots, z_8^{(0)}, (w_6^4)^{(0)}, (w_7^4)^{(0)})$ and $(z_1^{(0)}, \dots, z_6^{(0)})$ are solutions of

$$\mathcal{H} = \left\{\exp\left(z_k \frac{\partial V}{\partial z_k}\right) = 1, \exp\left(w_k^j \frac{\partial V}{\partial w_k^j}\right) = 1 \mid j = 4, k = 1, \dots, 8\right\}$$
$$\text{and } \widehat{\mathcal{H}} = \left\{\exp\left(z_k \frac{\partial \widehat{V}}{\partial z_k}\right) = 1 \mid k = 1, \dots, 6\right\},$$

respectively. Applying Theorem 3.3, we obtain

$$V_0(z_1^{(0)}, \dots, (w_7^4)^{(0)}) \equiv \widehat{V}_0(z_1^{(0)}, \dots, z_6^{(0)}) \equiv i(\mathrm{vol}(\rho) + i\,\mathrm{cs}(\rho)) \pmod{\pi^2},$$

and numerical calculation verifies it by

$$\widehat{V}_0(z_1^{(0)}, \dots, z_6^{(0)}) = i(0 + 1.6449...i),$$

where $\mathrm{vol}(3_1) = 0$ holds trivially and $1.6449... = \pi^2/6$ holds numerically.

## Acknowledgements

The author thanks Yuichi Kabaya and Jun Murakami for suggesting this research and having much discussion. Ayumu Inoue gave wonderful lectures on his work [Inoue and Kabaya 2014] at Seoul National University and it became the framework for Section 2 of this article. Many people, including Hyuk Kim, Seonhwa Kim, Roland van der Veen, Hitoshi Murakami, Satoshi Nawata, and Stephané Baseilhac heard my talks on the result and gave many suggestions. Special thanks are due to the reviewer who suggested the revised proof of Lemma 2.4.

The author is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1C1A1A02037540).

## References

[Baseilhac and Benedetti 2007] S. Baseilhac and R. Benedetti, "Quantum hyperbolic geometry", *Algebr. Geom. Topol.* **7** (2007), 845–917. MR Zbl

[Cho 2016a] J. Cho, "Optimistic limit of the colored Jones polynomial and the existence of a solution", *Proc. Amer. Math. Soc.* **144**:4 (2016), 1803–1814. MR Zbl

[Cho 2016b] J. Cho, "Optimistic limits of the colored Jones polynomials and the complex volumes of hyperbolic links", *J. Aust. Math. Soc.* **100**:3 (2016), 303–337. MR Zbl

[Cho and Murakami 2017] J. Cho and J. Murakami, "Reidemeister transformations of the potential function and the solution", *J. Knot Theory Ramifications* **26**:12 (2017), art. id. 1750079. MR Zbl

[Cho et al. 2014] J. Cho, H. Kim, and S. Kim, "Optimistic limits of Kashaev invariants and complex volumes of hyperbolic links", *J. Knot Theory Ramifications* **23**:9 (2014), art. id. 1450049. MR Zbl

[Elhamdadi and Nelson 2015] M. Elhamdadi and S. Nelson, *Quandles: an introduction to the algebra of knots*, Student Mathematical Library **74**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Garoufalidis et al. 2015] S. Garoufalidis, M. Goerner, and C. K. Zickert, "The Ptolemy field of 3-manifold representations", *Algebr. Geom. Topol.* **15**:1 (2015), 371–397. MR Zbl

[Hikami and Inoue 2015] K. Hikami and R. Inoue, "Braids, complex volume and cluster algebras", *Algebr. Geom. Topol.* **15**:4 (2015), 2175–2194. MR Zbl

[Inoue and Kabaya 2014] A. Inoue and Y. Kabaya, "Quandle homology and complex volume", *Geom. Dedicata* **171** (2014), 265–292. MR Zbl

[Kashaev 1995] R. M. Kashaev, "A link invariant from quantum dilogarithm", *Modern Phys. Lett. A* **10**:19 (1995), 1409–1418. MR Zbl

[Murakami 2000] H. Murakami, "The asymptotic behavior of the colored Jones function of a knot and its volume", preprint, 2000. arXiv

[Neumann 2004] W. D. Neumann, "Extended Bloch group and the Cheeger–Chern–Simons class", *Geom. Topol.* **8** (2004), 413–474. MR Zbl

[Neumann and Yang 1999] W. D. Neumann and J. Yang, "Bloch invariants of hyperbolic 3-manifolds", *Duke Math. J.* **96**:1 (1999), 29–59. MR Zbl

[Rolfsen 1976] D. Rolfsen, *Knots and links*, Mathematics Lecture Series **7**, Publish or Perish, Berkeley, CA, 1976. MR Zbl

[Sakuma and Yokota 2016] M. Sakuma and Y. Yokota, "An application of non-positively curved cubings of alternating links", preprint, 2016. arXiv

[Thurston 1999] D. Thurston, "Hyperbolic volume and the Jones polynomial", lecture notes, 1999, Available at http://pages.iu.edu/~dpthurst/speaking/Grenoble.pdf.

[Weeks 2005] J. Weeks, "Computation of hyperbolic structures in knot theory", pp. 461–480 in *Handbook of knot theory*, edited by W. Menasco and M. Thistlethwaite, Elsevier, Amsterdam, 2005. MR Zbl

[Yokota 2011] Y. Yokota, "On the complex volume of hyperbolic knots", *J. Knot Theory Ramifications* **20**:7 (2011), 955–976. MR Zbl

[Zickert 2009] C. K. Zickert, "The volume and Chern–Simons invariant of a representation", *Duke Math. J.* **150**:3 (2009), 489–532. MR Zbl

Received January 16, 2016. Revised January 30, 2018.

JINSEOK CHO
BUSAN NATIONAL UNIVERSITY OF EDUCATION
BUSAN
SOUTH KOREA
dol0425@bnue.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.367

# CLASSIFICATION OF POSITIVE SMOOTH SOLUTIONS TO THIRD-ORDER PDES INVOLVING FRACTIONAL LAPLACIANS

WEI DAI AND GUOLIN QIN

**In this paper, we are concerned with the third-order equations**

$$\begin{cases}(-\Delta)^{\frac{3}{2}}u=u^{\frac{d+3}{d-3}}, & x\in\mathbb{R}^d,\\ u\in C^3(\mathbb{R}^d), & u(x)>0,\ x\in\mathbb{R}^d,\end{cases}$$

**and**

$$\begin{cases}(-\Delta)^{\frac{3}{2}}u=\left(\frac{1}{|x|^6}*|u|^2\right)u, & x\in\mathbb{R}^d,\\ u\in C^3(\mathbb{R}^d), & u(x)>0,\ x\in\mathbb{R}^d,\ d\geq 7,\end{cases}$$

**with $\dot{H}^{\frac{3}{2}}$-critical nonlinearity. By showing the equivalence between the PDEs and the corresponding integral equations and using results from Chen et al. (2006) and Dai et al. (2018), we prove that positive classical solutions $u$ to the above equations are radially symmetric about some point $x_0\in\mathbb{R}^d$ and derive the explicit forms for $u$.**

## 1. Introduction

In this paper, we mainly consider the positive classical solutions to the following third-order conformal invariant equation with $\dot{H}^{\frac{3}{2}}$-critical nonlinearity:

$$\begin{cases}(-\Delta)^{\frac{3}{2}}u=u^{\frac{d+3}{d-3}}, & x\in\mathbb{R}^d,\\ u\in C^3(\mathbb{R}^d), & u(x)>0,\ x\in\mathbb{R}^d,\end{cases} \tag{1-1}$$

where $d\geq 4$ and the nonlocal fractional Laplacian $(-\Delta)^{\frac{1}{2}}$ can be defined by Fourier transform, that is,

$$\widehat{(-\Delta)^{\frac{1}{2}}f}(\xi):=(2\pi|\xi|)\hat{f}(\xi), \tag{1-2}$$

with $\hat{f}(\xi):=\int_{\mathbb{R}^d}f(x)e^{-2\pi i x\cdot\xi}\,dx$. If $f$ is in the Schwartz space $\mathcal{S}$ of rapidly decreasing $C^\infty$ functions in $\mathbb{R}^d$, then $(-\Delta)^{\frac{1}{2}}f$ can also be defined equivalently by

Dai was supported by the NNSF of China (No. 11501021).
*MSC2010:* primary 35R11; secondary 35B06, 35J91.
*Keywords:* fractional Laplacians, odd order, positive smooth solutions, radial symmetry, uniqueness, equivalence.

$$(-\Delta)^{\frac{\alpha}{2}} f(x) = C_{\alpha,d}\, \text{P.V.} \int_{\mathbb{R}^d} \frac{f(x)-f(y)}{|x-y|^{d+\alpha}}\, dy := C_{\alpha,d} \lim_{\epsilon\to 0} \int_{|y-x|\geq\epsilon} \frac{f(x)-f(y)}{|x-y|^{d+\alpha}}\, dy \tag{1-3}$$

with $\alpha = 1$, where the constant $C_{\alpha,d} = \bigl(\int_{\mathbb{R}^d} (1-\cos(2\pi\zeta_1))/|\zeta|^{d+\alpha} d\zeta\bigr)^{-1}$. For general $0<\alpha<2$, the definition (1-3) for $(-\Delta)^{\frac{\alpha}{2}} f$ can be extended and it is well defined for $f \in C^{1,1}_{\text{loc}} \cap \mathcal{L}_\alpha(\mathbb{R}^d)$ (see [Chen et al. 2015; 2017; Dai et al. 2017; Zhuo et al. 2014]) with

$$\mathcal{L}_\alpha(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; \int_{\mathbb{R}^d} \frac{|f(x)|}{1+|x|^{d+\alpha}}\, dx < \infty \right\}.$$

Throughout this paper, we define

$$(-\Delta)^{\frac{3}{2}} u := (-\Delta)^{\frac{1}{2}}(-\Delta u)$$

by definition (1-3) (with $f = -\Delta u$) provided that $\Delta u \in C^{1,1}_{\text{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ (i.e., (c) and (d) in Theorems 1.1 and 1.3), otherwise we will define $(-\Delta)^{\frac{3}{2}} u$ by Fourier transform (i.e., (a) and (b) in Theorems 1.1 and 1.3). See the extension method of defining $(-\Delta)^{\frac{\alpha}{2}}$ in [Caffarelli and Silvestre 2007]. The equation (1-1) is $\dot{H}^{\frac{3}{2}}$-critical in the sense that both it and the $\dot{H}^{\frac{3}{2}}$ norm are invariant under the same scaling

$$u_\rho(x) = \rho^{(d-3)/2} u(\rho x),$$

where the homogeneous Sobolev norm is defined as

$$\|u\|_{\dot{H}^{\frac{3}{2}}(\mathbb{R}^d)} := \|(-\Delta)^{\frac{3}{4}} u\|_{L^2(\mathbb{R}^d)} = \left( \int_{\mathbb{R}^d} |\xi|^3 |\hat{u}(\xi)|^2 d\xi \right)^{\frac{1}{2}}.$$

The quantitative and qualitative properties of solutions to fractional order or higher order conformally invariant equations of the form

$$(-\Delta)^{\frac{\alpha}{2}} u = u^{\frac{d+\alpha}{d-\alpha}} \tag{1-4}$$

have been extensively studied. In the special case $\alpha = 2$, (1-4) becomes the well-known Yamabe problem (for related results, please see Gidas, Ni and Nirenberg [Gidas et al. 1979] and Caffarelli, Gidas and Spruck [Caffarelli et al. 1989]); for $d = 2$, Chen and Li [2010] classified all the positive smooth solutions with finite total curvature of the equation

$$\begin{cases} -\Delta u = e^{2u}, & x \in \mathbb{R}^2, \\ \int_{\mathbb{R}^2} e^{2u}\, dx < \infty. \end{cases} \tag{1-5}$$

In general, when $\alpha = d$, under some assumptions, Chang and Yang [1997] classified the smooth solutions to

$$(-\Delta)^{\frac{d}{2}} u = (d-1)! e^{du}. \tag{1-6}$$

For $\alpha = 4$, Lin [1998] proved the classification results for all the positive smooth solutions of (1-4) ($d \geq 5$) and all the smooth solutions of

$$\begin{cases} \Delta^2 u = 6e^{4u}, & x \in \mathbb{R}^4, \\ \int_{\mathbb{R}^4} e^{4u}\,dx < \infty, & u(x) = o(|x|^2) \text{ as } |x| \to \infty. \end{cases} \tag{1-7}$$

Xu [2006] obtained similar results to Chang and Yang [1997] and Lin [1998] for (1-7) under the assumption $\Delta u(x) \to 0$ as $|x| \to \infty$. For $\alpha \in (0, d]$ an even integer, Wei and Xu [1999] classified the positive smooth solutions of (1-4), they also established the classification results for the smooth solutions of (1-6) with finite total curvature under the assumption $u(x) = o(|x|^2)$ as $|x| \to \infty$. Zhu [2004] classified all the smooth solutions with finite total curvature of the problem

$$\begin{cases} (-\Delta)^{\frac{3}{2}} u = 2e^{3u}, & x \in \mathbb{R}^3, \\ \int_{\mathbb{R}^3} e^{3u}\,dx < \infty, & u(x) = o(|x|^2) \text{ as } |x| \to \infty. \end{cases} \tag{1-8}$$

In [Chen et al. 2006], by developing the method of moving planes in integral forms, Chen, Li and Ou classified all the positive $L^{2d/(d-\alpha)}_{\text{loc}}$ solutions to the equivalent integral equation of PDE (1-4). As a consequence, they obtained the classification results for positive weak solutions to PDE (1-4), moreover, they also derived classification results for positive smooth solutions to (1-4) provided $\alpha \in (0, d)$ is an even integer. For more literature on the quantitative and qualitative properties of solutions to fractional order or higher order conformally invariant PDE and IE problems, please refer to [Chen and Li 2010; Chen et al. 2017; Dai et al. 2017; Xu 2005]. One should observe that, when $\alpha \in (0, d)$ is an odd integer, the classification for positive smooth solutions to (1-4) is still open.

By proving the equivalence between PDE (1-1) and the integral equation

$$u(x) = \int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} u^{\frac{d+3}{d-3}}(y)\,dy \tag{1-9}$$

and using the results for IE (1-9) from [Chen et al. 2006], we will study the classification of positive smooth solutions to the third-order equation (1-1) under assumptions which are similar to (or even weaker than) those in [Chen et al. 2017; Lin 1998; Xu 2006; Zhu 2004].

Our classification result for (1-1) is the following theorem.

**Theorem 1.1.** *Assume $d \geq 4$ and $u$ is a positive solution of* (1-1). *If $u$ satisfies one of the four assumptions*

(a) $\int_{\mathbb{R}^d} u^{\frac{2d}{d-3}}\,dx < \infty$ *and* $\Delta u(x) \to 0$ *as* $|x| \to \infty$,

(b) $\int_{\mathbb{R}^d} u^{\frac{2d}{d-3}}\,dx < \infty$ *and there exists some* $\tau < 3$ *such that* $u(x) = O(|x|^\tau)$ *as* $|x| \to \infty$,

(c) $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ *and* $\Delta u \leq 0$ *in* $\mathbb{R}^d$,

(d) $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \frac{u^{(d+3)/(d-3)}}{|x|^{d-3}}\, dx < \infty$ *and* $u(x) = o(|x|^2)$ *as* $|x| \to \infty$,

*then $u$ is radially symmetric and monotone decreasing about some point* $x_0 \in \mathbb{R}^d$; *in particular*, *the positive solution $u$ must assume the form*

$$u(x) = \left(\frac{1}{R_{3,d}\, I\big(\frac{d-3}{2}\big)}\right)^{\frac{d-3}{6}} \left(\frac{\lambda}{1+\lambda^2|x-x_0|^2}\right)^{\frac{d-3}{2}} \quad \text{for some } \lambda > 0,$$

*where* $R_{m,d} := \Gamma(\frac{d-m}{2})/(\pi^{\frac{d}{2}} 2^m \Gamma(\frac{m}{2}))$ *with* $0 < m < d$ *and*

$$I(s) := \frac{\pi^{\frac{d}{2}}\Gamma(\frac{1}{2}(d-2s))}{\Gamma(d-s)}$$

*for* $0 < s < \frac{d}{2}$.

**Remark 1.2.** In Theorem 1.1, we should observe that the integrable condition

$$\int_{\mathbb{R}^d} \frac{u^{\frac{d+3}{d-3}}}{|x|^{d-3}}\, dx < \infty$$

in (d) is much weaker than the condition $\int_{\mathbb{R}^d} u^{\frac{2d}{d-3}}\, dx < \infty$ in (a) and (b). In fact, one immediately has

$$\int_{|x|\geq 1} \frac{u^{\frac{d+3}{d-3}}(x)}{|x|^{d-3}}\, dx \leq \left(\int_{|x|\geq 1} u^{\frac{2d}{d-3}}\, dx\right)^{\frac{d+3}{2d}} \left(\int_{|x|\geq 1} \frac{1}{|x|^{2d}}\, dx\right)^{\frac{d-3}{2d}} < \infty,$$

provided that $\int_{\mathbb{R}^d} u^{\frac{2d}{d-3}}\, dx < \infty$. The assumption $\Delta u \in C^{1,1}_{\mathrm{loc}}$ in (c) and (d) in Theorem 1.1 can also be replaced by weaker assumptions $\Delta u \in C^{1,\epsilon}_{\mathrm{loc}}$ or $u \in C^{3,\epsilon}_{\mathrm{loc}}$ for arbitrarily small $\epsilon > 0$.

We also consider the classification of positive classical solutions to the following third-order $\dot{H}^{\frac{3}{2}}$-critical static Hartree equation with nonlocal nonlinearity:

$$\begin{cases} (-\Delta)^{\frac{3}{2}} u = \left(\frac{1}{|x|^6} * |u|^2\right) u, & x \in \mathbb{R}^d, \\ u \in C^3(\mathbb{R}^d), & u(x) > 0,\ x \in \mathbb{R}^d,\ d \geq 7. \end{cases} \tag{1-10}$$

The solution $u$ to problem (1-10) is also a stationary solution to the $\dot{H}^{\frac{3}{2}}$-critical focusing fractional order dynamic Schrödinger–Hartree equation

$$i\partial_t u + (-\Delta)^{\frac{3}{2}} u = \left(\frac{1}{|x|^6} * |u|^2\right) u, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d, \tag{1-11}$$

where $d \geq 7$. The Hartree equation has many interesting applications in the quantum theory of large systems of nonrelativistic bosonic atoms and molecules (see, e.g., [Fröhlich and Lenzmann 2004]). PDEs of the type (1-10) also arise in the Hartree–Fock theory of the nonlinear Schrödinger equations (see [Lieb and Simon 1977]).

There is lots of literature on the quantitative and qualitative properties of solutions to fractional order or higher order Hartree equations of the form

$$(-\Delta)^{\frac{\alpha}{2}} u = \left( \frac{1}{|x|^{2\alpha}} * |u|^2 \right) u \tag{1-12}$$

and various related Choquard equations, please see [Cao and Dai 2017; Dai et al. 2018; Liu 2009; Ma and Zhao 2010]. Cao and Dai [2017] classified all the positive $C^4$ solutions to the $\dot{H}^2$-critical biharmonic equation (1-12) with $\alpha = 4$; they also derived Liouville theorems in the subcritical cases. For general $0 < \alpha < \frac{d}{2}$, Dai et al. [2018] classified all the positive $L^{2d/(d-\alpha)}$ integrable solutions to the equivalent integral equation of PDE (1-12). As a consequence, they obtained the classification results for positive weak solutions to PDE (1-12).

By proving the equivalence between PDE (1-10) and the integral equation

$$u(x) = \int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} \left( \int_{\mathbb{R}^d} \frac{1}{|y-z|^6} |u(z)|^2 \, dz \right) u(y) \, dy \tag{1-13}$$

and using the results for IE (1-13) from [Dai et al. 2018], we establish the following classification theorem for positive smooth solutions of PDE (1-10) under similar assumptions as in Theorem 1.1.

**Theorem 1.3.** *Assume $u$ is a positive solution of* (1-10) *such that* $\int_{\mathbb{R}^d} u^{\frac{2d}{d-3}} \, dx < \infty$. *If $u$ satisfies one of the four assumptions*

(a) $\Delta u(x) \to 0$ *as* $|x| \to \infty$,

(b) *there exists some* $\tau < 3$ *such that* $u(x) = O(|x|^{\tau})$ *as* $|x| \to \infty$,

(c) $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ *and* $\Delta u \le 0$ *in* $\mathbb{R}^d$,

(d) $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ *and* $u(x) = o(|x|^2)$ *as* $|x| \to \infty$,

*then $u$ is radially symmetric and monotone decreasing about some point $x_0 \in \mathbb{R}^d$; in particular, the positive solution $u$ must assume the following form*:

$$u(x) = \sqrt{\frac{1}{R_{3,d} I(3) I\left(\frac{d-3}{2}\right)}} \left( \frac{\lambda}{1 + \lambda^2 |x - x_0|^2} \right)^{\frac{d-3}{2}} \quad \textit{for some } \lambda > 0.$$

The rest of our paper is organized as follows. In Section 2, we carry out our proof for Theorem 1.1. Section 3 is devoted to proving Theorem 1.3.

In the following, we will use $C$ to denote a general positive constant that may depend on $d$ and $u$, and whose value may differ from line to line.

## 2. Proof of Theorem 1.1

**Lemma 2.1** (Hardy–Littlewood–Sobolev inequality, [Lieb 1983]). *Letting $d \geq 1$, $0 < s < d$ and $1 < p < q < \infty$ be such that $\frac{d}{q} = \frac{d}{p} - s$, we have*

$$\left\| \int_{\mathbb{R}^d} \frac{f(y)}{|x-y|^{d-s}}\, dy \right\|_{L^q(\mathbb{R}^d)} \leq C_{d,s,p,q} \|f\|_{L^p(\mathbb{R}^d)}$$

*for all $f \in L^p(\mathbb{R}^d)$.*

Define

$$v(x) := -\int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} u^{\frac{d+3}{d-3}}(y)\, dy, \quad w(x) := u(x) + v(x), \tag{2-1}$$

where the Riesz potential's constants $R_{m,d} = \Gamma((d-m)/2)/(\pi^{\frac{d}{2}} 2^m \Gamma(m/2))$ with $0 < m < d$. Since $u$ is a solution to (1-1), we get immediately $(-\Delta)^{\frac{3}{2}} w \equiv 0$ and hence $\Delta^2 w \equiv 0$ in $\mathbb{R}^d$.

Under the following four entirely different assumptions (a), (b), (c) and (d) on $u$, we will prove that the solution $u$ to PDE (1-1) always satisfies the equivalent integral equation.

**(a)** Suppose $\Delta u \to 0$ as $|x| \to \infty$. By the Hardy–Littlewood–Sobolev inequality,

$$\|\Delta v\|_{L^{2d/(d+1)}(\mathbb{R}^d)} = C_d \left\| \int_{\mathbb{R}^d} \frac{1}{|x-y|^{d-1}} u^{\frac{d+3}{d-3}}(y)\, dy \right\|_{L^{2d/(d+1)}(\mathbb{R}^d)} \leq \widetilde{C}_d \|u\|_{L^{2d/(d-3)}(\mathbb{R}^d)}^{\frac{d+3}{d-3}}. \tag{2-2}$$

Now assume $z \in \mathbb{R}^d$ is arbitrary. We can infer from $\Delta v \in L^{2d/(d+1)}(\mathbb{R}^d)$ that there exists a sequence of radii $r_k \to \infty$ such that

$$r_k \cdot \int_{\partial B_{r_k}(z)} |\Delta v(x)|^{\frac{2d}{d+1}}\, d\sigma \to 0, \quad \text{as } k \to \infty. \tag{2-3}$$

Since $\Delta w$ is harmonic in $\mathbb{R}^d$, the mean value property yields that

$$\Delta w(z) = ⨍_{\partial B_{r_k}(z)} \Delta w(x)\, d\sigma, \tag{2-4}$$

where $⨍_{\partial B_{r_k}(z)} \Delta w(x)\, d\sigma$ is the integral average of $\Delta w$ over the sphere $|x-z| = r_k$. Therefore, by the Jensen inequality and (2-4), we get

$$\begin{aligned} |\Delta w(z)|^{\frac{2d}{d+1}} &\leq \left( ⨍_{\partial B_{r_k}(z)} \big(|\Delta u(x)| + |\Delta v(x)|\big)\, d\sigma \right)^{\frac{2d}{d+1}} \\ &\leq C_d \left\{ ⨍_{\partial B_{r_k}(z)} |\Delta u(x)|^{\frac{2d}{d+1}}\, d\sigma + ⨍_{\partial B_{r_k}(z)} |\Delta v(x)|^{\frac{2d}{d+1}}\, d\sigma \right\}. \end{aligned} \tag{2-5}$$

Letting $k \to \infty$ in (2-5), we can deduce from (2-3) and the assumption $\Delta u \to 0$ as $|x| \to \infty$ that

$$\Delta w(z) = 0. \tag{2-6}$$

Since $z \in \mathbb{R}^d$ is arbitrarily chosen, we actually have $\Delta w \equiv 0$ in $\mathbb{R}^d$.

Applying Hardy–Littlewood–Sobolev inequality again, we deduce that

$$\|v\|_{L^{2d/(d-3)}(\mathbb{R}^d)} \le C_d \|u^{\frac{d+3}{d-3}}\|_{L^{2d/(d+3)}(\mathbb{R}^d)} \le C_d \|u\|^{\frac{d+3}{d-3}}_{L^{2d/(d-3)}(\mathbb{R}^d)}. \tag{2-7}$$

Since $w \in L^{2d/(d-3)}(\mathbb{R}^d)$ is harmonic in $\mathbb{R}^d$, the Gagliardo–Nirenberg interpolation inequality implies that

$$\|\nabla w\|_{L^{2d/(d-1)}(\mathbb{R}^d)} \le C_d \|w\|^{\frac{1}{2}}_{L^{2d/(d-3)}(\mathbb{R}^d)} \|\Delta w\|^{\frac{1}{2}}_{L^{2d/(d+1)}(\mathbb{R}^d)} = 0, \tag{2-8}$$

thus we arrive at $w \equiv 0$ in $\mathbb{R}^d$. That is, $u$ also satisfies the integral equation

$$u(x) = \int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} u^{\frac{d+3}{d-3}}(y)\,dy. \tag{2-9}$$

**(b)** Suppose there exists some $\tau < 3$ such that $u(x) = O(|x|^\tau)$ as $|x| \to \infty$. Without loss of generality, we may assume $\tau > 2$. By the Hölder inequality, we have for $|x|$ sufficiently large,

$$\begin{aligned} |v(x)| &\le C_d \bigg[ \int_{|x-y|\ge 1} \frac{1}{|x-y|^{d-3}} u^{\frac{d+3}{d-3}}(y)\,dy \\ &\qquad\qquad + \int_{|x-y|\le 1} \frac{1}{|x-y|^{d-3}} u^{\frac{d+3}{d-3}}(y)\,dy \bigg] \\ &\le C_d + C_{d,\delta} \big( \sup_{\overline{B}_1(x)} u \big)^{1+\delta} \le C|x|^{(1+\delta)\tau}, \end{aligned}$$

where $\delta > 0$ is fixed sufficiently small such that $\tau < (1+\delta)\tau < 3$. It follows that $w(x) = O(|x|^{\tilde{\tau}})$ with $\tilde{\tau} := (1+\delta)\tau < 3$.

Since $\Delta w$ is harmonic in $\mathbb{R}^d$, from the mean value property, we get that, for any $x \in \mathbb{R}^d$ and $s > 0$,

$$\Delta w(x) = \frac{d}{\omega_{d-1} s^d} \int_{|y-x|\le s} \Delta w(y)\,dy = \frac{d}{\omega_{d-1} s^d} \int_{|y-x|=s} \frac{\partial w}{\partial s}(y)\,d\sigma, \tag{2-10}$$

where $\omega_{d-1}$ is the area of the unit sphere in $\mathbb{R}^d$. By integrating with respect to $s$ from 0 to $r$ in (2-10), we have

$$\frac{r^2}{2d} \Delta w(x) = \frac{1}{\omega_{d-1} r^{d-1}} \int_{|y-x|=r} w(y)\,d\sigma - w(x). \tag{2-11}$$

Therefore, we can deduce from $w(x) = O(|x|^{\tilde{\tau}})$ and (2-11) that, for any $x \in \mathbb{R}^d$ with $|x|$ sufficiently large and $r = |x|/2$,

$$|\Delta w(x)| \le \frac{2d}{r^2}\left\{\sup_{\frac{1}{2}|x|\le|y|\le\frac{3}{2}|x|}|w(y)| + |w(x)|\right\} \le C|x|^{\tilde{\tau}-2}, \tag{2-12}$$

that is, $\Delta w(x) = O(|x|^{\tilde{\tau}-2})$ as $|x| \to \infty$. Thus, by gradient estimates for harmonic functions, we have

$$\Delta w(x) \equiv C \quad \text{for all } x \in \mathbb{R}^d, \tag{2-13}$$

which implies that $w(x) - C/(2d)|x|^2$ is harmonic in $\mathbb{R}^d$. Since $w(x) - C/(2d)|x|^2 = O(|x|^{\tilde{\tau}})$, by gradient estimates for harmonic functions, $w$ must be a quadratic polynomial, that is,

$$w(x) = \sum_{i,j} a_{ij}x_i x_j + \sum_i b_i x_i + c. \tag{2-14}$$

Since $w \in L^{2d/(d-3)}(\mathbb{R}^d)$, all the coefficients $a_{ij}$, $b_i$ and $c$ in (2-14) must be zero, that is $w(x) \equiv 0$ in $\mathbb{R}^d$, thus $u$ also satisfies the equivalent integral equation (2-9).

**(c)** Suppose $\Delta u \in C^{1,1}_{\text{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ and $\Delta u \le 0$ in $\mathbb{R}^d$. We will prove the classical solution $u$ to PDE (1-1) also satisfies the equivalent integral equation (2-9) using the ideas from [Chen et al. 2015; Zhuo et al. 2014]. To this end, we will need the following two lemmas established in [Chen et al. 2017; Silvestre 2007; Zhuo et al. 2014].

**Lemma 2.2** (maximum principle, [Chen et al. 2017; Silvestre 2007]). *Let $\Omega$ be a bounded domain in $\mathbb{R}^d$ and $0 < \alpha < 2$. Assume that $u \in \mathcal{L}_\alpha \cap C^{1,1}_{\text{loc}}(\Omega)$ and is lower semicontinuous on $\overline{\Omega}$. If $(-\Delta)^{\frac{\alpha}{2}}u \ge 0$ in $\Omega$ and $u \ge 0$ in $\mathbb{R}^d \setminus \Omega$, then $u \ge 0$ in $\mathbb{R}^d$. Moreover, if $u = 0$ at some point in $\Omega$, then $u = 0$ almost everywhere in $\mathbb{R}^d$. These conclusions also hold for an unbounded domain $\Omega$ if we assume further that*

$$\liminf_{|x|\to\infty} u(x) \ge 0.$$

**Lemma 2.3** (Liouville theorem, [Zhuo et al. 2014]). *Assume $d \ge 2$ and $0 < \alpha < 2$. Let $u$ be a strong solution of*

$$\begin{cases} (-\Delta)^{\frac{\alpha}{2}}u = 0, & x \in \mathbb{R}^d, \\ u(x) \ge 0, & x \in \mathbb{R}^d, \end{cases}$$

*then $u \equiv C \ge 0$.*

**Remark 2.4.** Lemma 2.2 has been established first by Silvestre [2007] without the assumption $u \in C^{1,1}_{\text{loc}}(\Omega)$. In [Chen et al. 2017], Chen, Li and Li provided a much more elementary and simpler proof for Lemma 2.2 under the assumption $u \in C^{1,1}_{\text{loc}}(\Omega)$.

First, assume $u$ is a positive solution to (1-1) satisfying $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ and $\Delta u \le 0$ in $\mathbb{R}^d$; we will show that $-\Delta u$ also satisfies the integral equation

$$-\Delta u = \int_{\mathbb{R}^d} \frac{R_{1,d}}{|x-y|^{d-1}} u^{\frac{d+3}{d-3}}(y)\, dy + C_1, \tag{2-15}$$

where $C_1 \ge 0$ is a constant.

For arbitrary $R > 0$, let

$$\tilde{v}_R(x) = \int_{B_R(0)} G^1_R(x,y) u^{\frac{d+3}{d-3}}(y)\, dy, \tag{2-16}$$

where the Green's function for $(-\Delta)^{\frac{1}{2}}$ on $B_R(0)$ is given by

$$G^1_R(x,y) = \frac{C_d}{|x-y|^{d-1}} \int_0^{\frac{t_R}{s_R}} \frac{1}{b^{\frac{1}{2}}(1+b)^{\frac{d}{2}}}\, db, \quad \text{if } x, y \in B_R(0) \tag{2-17}$$

with $s_R = |x-y|^2/R^2$, $t_R = (1-|x|^2/R^2)(1-|y|^2/R^2)$, and $G^1_R(x,y) = 0$ if $x$ or $y \in \mathbb{R}^d \setminus B_R(0)$ (see [Kulczycki 1997]).

Then, we can derive

$$\begin{cases} (-\Delta)^{1/2} \tilde{v}_R(x) = u^{\frac{d+3}{d-3}}(x), & x \in B_R(0), \\ \tilde{v}_R(x) = 0, & x \in \mathbb{R}^d \setminus B_R(0). \end{cases} \tag{2-18}$$

Letting $\tilde{w}_R(x) = -\Delta u(x) - \tilde{v}_R(x)$, by (1-1) and (2-18), we have

$$\begin{cases} (-\Delta)^{1/2} \tilde{w}_R(x) = 0, & x \in B_R(0), \\ \tilde{w}_R(x) \ge 0, & x \in \mathbb{R}^d \setminus B_R(0). \end{cases} \tag{2-19}$$

By Lemma 2.2, we deduce that for any $R > 0$,

$$\tilde{w}_R(x) = -\Delta u(x) - \tilde{v}_R(x) \ge 0, \quad \text{for all } x \in \mathbb{R}^d. \tag{2-20}$$

Now, for each fixed $x \in \mathbb{R}^d$, letting $R \to \infty$ in (2-20), we have

$$-\Delta u(x) \ge \int_{\mathbb{R}^d} \frac{R_{1,d}}{|x-y|^{d-1}} u^{\frac{d+3}{d-3}}(y)\, dy =: \tilde{v}(x) > 0. \tag{2-21}$$

Taking $x = 0$ in (2-21), we get

$$\int_{\mathbb{R}^d} \frac{u^{\frac{d+3}{d-3}}(y)}{|y|^{d-1}}\, dy < \infty, \tag{2-22}$$

and it follows easily that $\int_{\mathbb{R}^d} |u(x)|/(1+|x|^d)\, dx < \infty$, and hence $u \in \mathcal{L}_\alpha$ for any $\alpha > 0$. One can easily observe that $\tilde{v}$ is a solution of

$$(-\Delta)^{\frac{1}{2}} \tilde{v}(x) = u^{\frac{d+3}{d-3}}(x), \quad x \in \mathbb{R}^d. \tag{2-23}$$

Define $\tilde{w}(x) = -\Delta u(x) - \tilde{v}(x)$, then it satisfies

$$\begin{cases} (-\Delta)^{\frac{1}{2}}\tilde{w}(x) = 0, & x \in \mathbb{R}^d, \\ \tilde{w}(x) \geq 0 & x \in \mathbb{R}^d. \end{cases} \tag{2-24}$$

From Lemma 2.3, we can deduce that

$$\tilde{w}(x) = -\Delta u(x) - \tilde{v}(x) \equiv C_1 \geq 0. \tag{2-25}$$

Therefore, we have proved (2-15), that is,

$$-\Delta u = \int_{\mathbb{R}^d} \frac{R_{1,d}}{|x-y|^{d-1}} u^{\frac{d+3}{d-3}}(y)\, dy + C_1 =: f(u) \geq C_1 \geq 0. \tag{2-26}$$

Next, we will prove $u$ also satisfies the equivalent integral equation (2-9). For arbitrary $R > 0$, let

$$v_R(x) = \int_{B_R(0)} G_R^2(x, y) f(u)(y)\, dy, \tag{2-27}$$

where the Green's function for $-\Delta$ on $B_R(0)$ is given by

$$G_R^2(x, y) = C_d \left[ \frac{1}{|x-y|^{d-2}} - \frac{1}{\big(|x| \cdot |Rx/|x|^2 - y/R|\big)^{d-2}} \right] \quad \text{if } x, y \in B_R(0),$$

and $G_R^2(x, y) = 0$ if $x$ or $y \in \mathbb{R}^d \setminus B_R(0)$. Then, we can get

$$\begin{cases} -\Delta v_R(x) = f(u)(x), & x \in B_R(0), \\ v_R(x) = 0, & x \in \mathbb{R}^d \setminus B_R(0). \end{cases} \tag{2-28}$$

Let $w_R(x) = u(x) - v_R(x)$, by (2-26) and (2-28), we have

$$\begin{cases} -\Delta w_R(x) = 0, & x \in B_R(0), \\ w_R(x) > 0, & x \in \mathbb{R}^d \setminus B_R(0). \end{cases} \tag{2-29}$$

By the maximum principle, we deduce that for any $R > 0$,

$$w_R(x) = u(x) - v_R(x) > 0, \qquad \text{for all } x \in \mathbb{R}^d. \tag{2-30}$$

Now, for each fixed $x \in \mathbb{R}^d$, letting $R \to \infty$ in (2-30), we have

$$u(x) \geq \int_{\mathbb{R}^d} \frac{R_{2,d}}{|x-y|^{d-2}} f(u)(y)\, dy =: V(x) > 0. \tag{2-31}$$

Taking $x = 0$ in (2-31), we get

$$\int_{\mathbb{R}^d} \frac{C_1}{|y|^{d-2}}\, dy \leq \int_{\mathbb{R}^d} \frac{f(u)(y)}{|y|^{d-2}}\, dy < \infty, \tag{2-32}$$

and it follows easily that $C_1 = 0$, and hence

$$-\Delta u = f(u) = \int_{\mathbb{R}^d} \frac{R_{1,d}}{|x-y|^{d-1}} u^{\frac{d+3}{d-3}}(y)\,dy.$$

One can easily observe that $V$ is a solution of

$$-\Delta V(x) = f(u)(x), \quad x \in \mathbb{R}^d. \tag{2-33}$$

Define $W(x) = u(x) - V(x)$, then it satisfies

$$\begin{cases} -\Delta W(x) = 0, & x \in \mathbb{R}^d, \\ W(x) \geq 0 & x \in \mathbb{R}^d. \end{cases} \tag{2-34}$$

From the Liouville theorem for harmonic functions, we can deduce that

$$W(x) = u(x) - V(x) \equiv C_2 \geq 0. \tag{2-35}$$

Therefore, we have proved that

$$u(x) = \int_{\mathbb{R}^d} \frac{R_{2,d}}{|x-y|^{d-2}} f(u)(y)\,dy + C_2 \geq C_2 \geq 0. \tag{2-36}$$

Now (2-22) implies that

$$\int_{\mathbb{R}^d} \frac{C_2^{\frac{d+3}{d-3}}(y)}{|y|^{d-1}}\,dy \leq \int_{\mathbb{R}^d} \frac{u^{\frac{d+3}{d-3}}(y)}{|y|^{d-1}}\,dy < \infty, \tag{2-37}$$

from which we can infer that $C_2 = 0$. Thus, by using the formula

$$\int_{\mathbb{R}^d} \frac{1}{|x-y|^{d-2}} \cdot \frac{1}{|y|^{d-1}}\,dy = \frac{R_{3,d}}{R_{1,d} R_{2,d}} \cdot \frac{1}{|x|^{d-3}} \tag{2-38}$$

(see [Stein 1970]) and direct calculations, we finally deduce from (2-36) that

$$\begin{aligned} u(x) &= \int_{\mathbb{R}^d} \frac{R_{2,d}}{|x-y|^{d-2}} \int_{\mathbb{R}^d} \frac{R_{1,d}}{|y-z|^{d-1}} u^{\frac{d+3}{d-3}}(z)\,dz\,dy \\ &= \int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-z|^{d-3}} u^{\frac{d+3}{d-3}}(z)\,dz, \end{aligned} \tag{2-39}$$

that is, $u$ also satisfies the equivalent integral equation (2-9).

**(d)** Suppose $\Delta u \in C^{1,1}_{\text{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \frac{u^{\frac{d+3}{d-3}}}{|x|^{d-3}}\,dx < \infty$$

and $u(x) = o(|x|^2)$ as $|x| \to \infty$. By the above proof under assumption (c), we only need to prove the super-harmonic property $-\Delta u \geq 0$ under assumption (d).

For that purpose, we will first estimate the upper bound for $-v(x)$. Since one can verify that

$$\Delta v(x)=\int_{\mathbb{R}^d}\frac{(d-3)R_{3,d}}{|x-y|^{d-1}}u^{\frac{d+3}{d-3}}(y)\,dy\geq 0, \tag{2-40}$$

we deduce that, for $|x|$ sufficiently large,

$$\begin{aligned}0\leq -v(x)&=\int_{\mathbb{R}^d}\frac{R_{3,d}}{|x-y|^{d-3}}u^{\frac{d+3}{d-3}}(y)\,dy\\&=\int_{|y-x|\geq\frac{|x|}{6}}\frac{R_{3,d}}{|x-y|^{d-3}}u^{\frac{d+3}{d-3}}(y)\,dy+\int_{|y-x|<\frac{|x|}{6}}\frac{R_{3,d}}{|x-y|^{d-3}}u^{\frac{d+3}{d-3}}(y)\,dy\\&\leq 7^{d-3}R_{3,d}\int_{|y-x|\geq\frac{|x|}{6}}\frac{u^{\frac{d+3}{d-3}}(y)}{|y|^{d-3}}\,dy+\frac{|x|^2}{36}\int_{|y-x|<\frac{|x|}{6}}\frac{R_{3,d}}{|x-y|^{d-1}}u^{\frac{d+3}{d-3}}(y)\,dy\\&\leq C_d\int_{\mathbb{R}^d}\frac{u^{\frac{d+3}{d-3}}(y)}{|y|^{d-3}}\,dy+\frac{|x|^2}{36(d-3)}\Delta v(x).\end{aligned}$$

As a consequence, we deduce from the assumption

$$\int_{\mathbb{R}^d}\frac{u^{\frac{d+3}{d-3}}}{|x|^{d-3}}\,dx<\infty$$

that, as $|x|\to\infty$,

$$0\leq -v(x)\leq O(1)+\frac{|x|^2}{36(d-3)}\Delta v(x). \tag{2-41}$$

Next, we can deduce from (2-11) that, for any $x\in\mathbb{R}^d$ with $|x|$ sufficiently large and $r=|x|/2$,

$$\begin{aligned}\Delta w(x)&\leq\frac{2d}{r^2}\left\{\sup_{\frac{1}{2}|x|\leq|y|\leq\frac{3}{2}|x|}w(y)-u(x)-v(x)\right\}\\&\leq\frac{2d}{r^2}\left\{\sup_{\frac{1}{2}|x|\leq|y|\leq\frac{3}{2}|x|}u(y)-v(x)\right\}.\end{aligned} \tag{2-42}$$

Therefore, we get from (2-40), (2-41), (2-42) and the assumption $u(x)=o(|x|^2)$ as $|x|\to\infty$ that, as $|x|\to\infty$,

$$\begin{aligned}\Delta w(x)=\Delta u(x)+\Delta v(x)&\leq\frac{8d}{|x|^2}\left\{o(|x|^2)+O(1)+\frac{|x|^2}{36(d-3)}\Delta v(x)\right\}\\&\leq o(1)+\frac{d}{4(d-3)}\Delta v(x).\end{aligned} \tag{2-43}$$

We can deduce from (2-43) that

$$\limsup_{|x|\to\infty} \Delta u(x) \le 0, \qquad \text{that is ,} \qquad \liminf_{|x|\to\infty}(-\Delta u(x)) \ge 0. \tag{2-44}$$

Therefore, from (1-1), (2-44) and the maximum principle (Lemma 2.2), we can infer

$$-\Delta u \ge 0 \quad \text{in } \mathbb{R}^d. \tag{2-45}$$

In conclusion, we have proved respectively under the four different assumptions (a), (b), (c) and (d) on $u$ that the classical solution $u$ to PDE (1-1) always satisfies the equivalent integral equation (2-9). Applying [Chen et al. 2006, Theorem 1.1] ($u \in L^{2d/(d-3)}_{\text{loc}}(\mathbb{R}^d)$ was assumed therein) to integral equation (2-9), we deduce immediately that $u$ is radially symmetric and monotone decreasing about some point $x_0 \in \mathbb{R}^d$ and thus assumes the form

$$u(x) = \left(\frac{1}{R_{3,d}\, I\left(\frac{d-3}{2}\right)}\right)^{\frac{d-3}{6}} \left(\frac{\lambda}{1+\lambda^2|x-x_0|^2}\right)^{\frac{d-3}{2}} \tag{2-46}$$

for some positive constant $\lambda$, where

$$I(s) := \frac{\pi^{\frac{d}{2}}\Gamma\left(\frac{d-2s}{2}\right)}{\Gamma(d-s)}$$

for $0 < s < \frac{d}{2}$. This concludes the proof of Theorem 1.1.

**Remark 2.5.** In the proof of Theorem 1.1 under assumption (d), one crucial step is to deduce $\Delta u \le 0$ from the assumptions

$$\int_{\mathbb{R}^d} \frac{u^{\frac{d+3}{d-3}}}{|x|^{d-3}}\, dx < \infty$$

and $u(x) = o(|x|^2)$ as $|x| \to \infty$, where the fractional Laplacian $(-\Delta)^{\frac{1}{2}}$ is given by definition (1-3). Suppose $(-\Delta)^{\frac{1}{2}}$ can be defined in terms of the Fourier transform, that is,

$$\widehat{(-\Delta)^{\frac{1}{2}} f}(\xi) := (2\pi|\xi|)\hat{f}(\xi)$$

with $\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\xi}\, dx$, then the super-harmonic property $\Delta u \le 0$ can be deduced directly from $\int_{\mathbb{R}^d} u^{(d+3)/(d-3)}/|x|^{d-1}\, dx < \infty$. Indeed, we only need to show that $\int_{\mathbb{R}^d}(-\Delta u)\phi dx \ge 0$ for any nontrivial $0 \le \phi \in C_0^\infty(\mathbb{R}^d)$. To this end, we define

$$\psi(x) := (-\Delta)^{-\frac{1}{2}}\phi(x) = \int_{\mathbb{R}^d} \frac{R_{1,d}}{|x-y|^{d-1}}\phi(y)\, dy \ge 0,$$

then $\psi \in C^\infty(\mathbb{R}^d)$ and satisfy $(2\pi|\xi|)\hat{\psi}(\xi) = \hat{\phi}(\xi)$ (see [Stein 1970]). Moreover, one can easily verify that $\psi(x) \sim 1/|x|^{d-1}$ for $|x|$ large enough, thus we have

$\int_{\mathbb{R}^d} u^{\frac{d+3}{d-3}} \psi \, dx < \infty$ provided $\int_{\mathbb{R}^d} u^{\frac{d+3}{d-3}} / |x|^{d-1} \, dx < \infty$. Therefore, we may multiply both sides of the PDE (1-1) by $\psi$ and integrate, then by Parseval's formula, we get

$$\infty > \int_{\mathbb{R}^d} u^{\frac{d+3}{d-3}} \psi \, dx = \int_{\mathbb{R}^d} (-\Delta)^{\frac{3}{2}} u \cdot \psi \, dx = \int_{\mathbb{R}^d} (2\pi |\xi|) \widehat{-\Delta u} \cdot \bar{\hat{\psi}} \, d\xi = \int_{\mathbb{R}^d} (-\Delta u) \cdot \phi \, dx \geq 0.$$

## 3. Proof of Theorem 1.3

We define

$$v(x) := -\int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} \left( \int_{\mathbb{R}^d} \frac{1}{|y-z|^6} |u(z)|^2 \, dz \right) u(y) \, dy, \tag{3-1}$$
$$w(x) := u(x) + v(x).$$

Since $u$ is a solution to (1-10), we get immediately $(-\Delta)^{\frac{3}{2}} w \equiv 0$ and hence $\Delta^2 w \equiv 0$ in $\mathbb{R}^d$.

Our goal is to show under the following four entirely different assumptions (a), (b), (c) and (d) that the solution $u$ to PDE (1-10) always satisfies the equivalent integral equation

$$u(x) = -v(x) = \int_{\mathbb{R}^d} \frac{R_{3,d}}{|x-y|^{d-3}} \left( \int_{\mathbb{R}^d} \frac{1}{|y-z|^6} |u(z)|^2 \, dz \right) u(y) \, dy. \tag{3-2}$$

**(a)** Suppose $\Delta u \to 0$ as $|x| \to \infty$. The key ingredients are showing $v \in L^{2d/(d-3)}(\mathbb{R}^d)$ and $\Delta v \in L^{2d/(d+1)}(\mathbb{R}^d)$.

Indeed, let $P(x) := 1/|x|^6 * |u|^2$, then by the Hardy–Littlewood–Sobolev inequality, one has

$$\|P\|_{L^{d/3}(\mathbb{R}^d)} \leq C \|u^2\|_{L^{d/(d-3)}(\mathbb{R}^d)} \leq C \|u\|^2_{L^{2d/(d-3)}(\mathbb{R}^d)}. \tag{3-3}$$

Therefore, by using Hardy–Littlewood–Sobolev inequality again, we get

$$\begin{aligned} \|v\|_{L^{2d/(d-3)}(\mathbb{R}^d)} &\leq C_d \|Pu\|_{L^{2d/(d+3)}(\mathbb{R}^d)} \leq C_d \|P\|_{L^{d/(3)}(\mathbb{R}^d)} \|u\|_{L^{2d/(d-3)}(\mathbb{R}^d)} \\ &\leq C_d \|u\|^3_{L^{\frac{2d}{d-3}}(\mathbb{R}^d)}, \end{aligned} \tag{3-4}$$

$$\begin{aligned} \|\Delta v\|_{L^{\frac{2d}{d+1}}(\mathbb{R}^d)} &= C_d \left\| \int_{\mathbb{R}^d} \frac{P(y)u(y)}{|x-y|^{d-1}} \, dy \right\|_{L^{\frac{2d}{d+1}}(\mathbb{R}^d)} \\ &\leq \widetilde{C}_d \|Pu\|_{L^{\frac{2d}{d+3}}(\mathbb{R}^d)} \leq \widetilde{C}_d \|u\|^3_{L^{\frac{2d}{d-3}}(\mathbb{R}^d)}. \end{aligned} \tag{3-5}$$

The rest of the proof is similar to the proof of Theorem 1.1 under assumption (a) in Section 2.

**(b)** Suppose there exists some $\tau < 3$ such that $u(x) = O(|x|^\tau)$ as $|x| \to \infty$. Without loss of generality, we may assume $\tau > 2$. The key ingredient is proving $w(x) = O(|x|^{\tilde{\tau}})$ for some $\tau < \tilde{\tau} < 3$.

In fact, using Hölder's inequality, one can verify that for $|x|$ large enough,

$$P(x) \le \int_{|x-y|\ge 1} \frac{1}{|x-y|^6}|u(y)|^2\,dy + \int_{|x-y|\le 1} \frac{1}{|x-y|^6}|u(y)|^2\,dy \le C_d + C_d\Big(\sup_{\overline{B}_1(x)} u\Big)^2 \le C|x|^{2\tau}. \tag{3-6}$$

Therefore, by $P \in L^{\frac{d}{3}}(\mathbb{R}^d)$ and the Hölder inequality, we have for $|x|$ sufficiently large,

$$|v(x)| \le C_d\bigg[\int_{|x-y|\ge 1} \frac{1}{|x-y|^{d-3}} P(y)u(y)\,dy + \int_{|x-y|\le 1} \frac{1}{|x-y|^{d-3}} P(y)u(y)\,dy\bigg] \le C_d + C_{d,\delta}\Big(\sup_{\overline{B}_1(x)} u\Big)\Big(\sup_{\overline{B}_1(x)} P\Big)^{\delta} \le C|x|^{(1+2\delta)\tau}, \tag{3-7}$$

where $\delta > 0$ is fixed sufficiently small such that $\tau < (1+2\delta)\tau < 3$. It follows that $w(x) = O(|x|^{\tilde{\tau}})$ with $\tilde{\tau} := (1+2\delta)\tau < 3$. The rest of the proof is similar to the proof of Theorem 1.1 under assumption (b) in Section 2.

**(c)** The proof is similar to the proof of Theorem 1.1 under assumption (c) in Section 2.

**(d)** Suppose $\Delta u \in C^{1,1}_{\mathrm{loc}} \cap \mathcal{L}_1(\mathbb{R}^d)$ and $u(x) = o(|x|^2)$ as $|x| \to \infty$. The key ingredient is proving $\int_{\mathbb{R}^d} P(x)u(x)/|x|^{d-3}\,dx < \infty$. Indeed, we have

$$\int_{\mathbb{R}^d} \frac{P(x)u(x)}{|x|^{d-3}}\,dx \le \int_{|x|\le 1} \frac{1}{|x|^{d-3}}\,dx \cdot \|Pu\|_{L^\infty(\overline{B}_1)} + \bigg(\int_{|x|>1} \frac{1}{|x|^{2d}}\,dx\bigg)^{\frac{d-3}{2d}} \|P\|_{L^{d/3}} \|u\|_{L^{2d/(d-3)}} < \infty.$$

The rest of the proof is similar to the proof of Theorem 1.1 under assumption (d) in Section 2.

In conclusion, we have proved respectively under the four different assumptions (a), (b), (c) and (d) on $u$ that the classical solution $u$ to PDE (1-10) always satisfies the equivalent integral equation (3-2). Applying [Dai et al. 2018, Theorem 1.4] ($u \in L^{\frac{2d}{d-3}}(\mathbb{R}^d)$ was assumed therein) to integral equation (3-2), we deduce immediately that $u$ is radially symmetric and monotone decreasing about some point $x_0 \in \mathbb{R}^d$ and thus assumes the form

$$u(x) = \sqrt{\frac{1}{R_{3,d}\,I(3)\,I\big(\frac{d-3}{2}\big)}}\bigg(\frac{\lambda}{1+\lambda^2|x-x_0|^2}\bigg)^{\frac{d-3}{2}} \tag{3-8}$$

for some positive constant $\lambda$. This concludes the proof of Theorem 1.3.

## Acknowledgements

The authors are grateful to the referees for their careful reading and valuable comments and suggestions that improved the presentation of the paper.

## References

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32**:7-9 (2007), 1245–1260. MR Zbl

[Caffarelli et al. 1989] L. A. Caffarelli, B. Gidas, and J. Spruck, "Asymptotic symmetry and local behavior of semilinear elliptic equations with critical Sobolev growth", *Comm. Pure Appl. Math.* **42**:3 (1989), 271–297. MR Zbl

[Cao and Dai 2017] D. Cao and W. Dai, "Classification of nonnegative solutions to a bi-harmonic equation with Hartree type nonlinearity", 2017. To appear in *Proc. Royal Soc. Edinburgh Sect. A*.

[Chang and Yang 1997] S.-Y. A. Chang and P. C. Yang, "On uniqueness of solutions of $n$th order differential equations in conformal geometry", *Math. Res. Lett.* **4**:1 (1997), 91–102. MR Zbl

[Chen and Li 2010] W. Chen and C. Li, *Methods on nonlinear elliptic equations*, AIMS Series on Differential Equations & Dynamical Systems **4**, Amer. Inst. Math. Sci., Springfield, MO, 2010. MR Zbl

[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 Zbl

[Chen et al. 2015] W. Chen, Y. Fang, and R. Yang, "Liouville theorems involving the fractional Laplacian on a half space", *Adv. Math.* **274** (2015), 167–198. MR Zbl

[Chen et al. 2017] W. Chen, C. Li, and Y. Li, "A direct method of moving planes for the fractional Laplacian", *Adv. Math.* **308** (2017), 404–437. MR Zbl

[Dai et al. 2017] W. Dai, Z. Liu, and G. Lu, "Liouville type theorems for PDE and IE systems involving fractional Laplacian on a half space", *Potential Anal.* **46**:3 (2017), 569–588. MR Zbl

[Dai et al. 2018] W. Dai, Y. Fang, J. Huang, Y. Qin, and B. Wang, "Regularity and classification of solutions to static Hartree equations involving fractional Laplacians", 2018, available at http://www.escience.cn/system/download/88591. To appear in *Discrete Contin. Dyn. Syst. Ser. A*.

[Fröhlich and Lenzmann 2004] J. Fröhlich and E. Lenzmann, "Mean-field limit of quantum Bose gases and nonlinear Hartree equation", exposé XIX in *Séminaire: Équations aux Dérivées Partielles, 2003–2004*, edited by J.-M. Bony et al., École Polytech., Palaiseau, 2004. MR Zbl arXiv

[Gidas et al. 1979] B. Gidas, W. M. Ni, and L. Nirenberg, "Symmetry and related properties via the maximum principle", *Comm. Math. Phys.* **68**:3 (1979), 209–243. MR Zbl

[Kulczycki 1997] T. Kulczycki, "Properties of Green function of symmetric stable processes", *Probab. Math. Statist.* **17**:2 (1997), 339–364. MR Zbl

[Lieb 1983] E. H. Lieb, "Sharp constants in the Hardy–Littlewood–Sobolev and related inequalities", *Ann. of Math.* (2) **118**:2 (1983), 349–374. MR Zbl

[Lieb and Simon 1977] E. H. Lieb and B. Simon, "The Hartree–Fock theory for Coulomb systems", *Comm. Math. Phys.* **53**:3 (1977), 185–194. MR

[Lin 1998] C.-S. Lin, "A classification of solutions of a conformally invariant fourth order equation in $\mathbb{R}^n$", *Comment. Math. Helv.* **73**:2 (1998), 206–231. MR Zbl

[Liu 2009] S. Liu, "Regularity, symmetry, and uniqueness of some integral type quasilinear equations", *Nonlinear Anal.* **71**:5-6 (2009), 1796–1806. MR Zbl

[Ma and Zhao 2010] L. Ma and L. Zhao, “Classification of positive solitary solutions of the nonlinear Choquard equation”, *Arch. Ration. Mech. Anal.* **195**:2 (2010), 455–467. MR Zbl

[Silvestre 2007] L. Silvestre, “Regularity of the obstacle problem for a fractional power of the Laplace operator”, *Comm. Pure Appl. Math.* **60**:1 (2007), 67–112. MR Zbl

[Stein 1970] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Math. Series **30**, Princeton Univ. Press, 1970. MR Zbl

[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 Zbl

[Xu 2005] X. Xu, “Exact solutions of nonlinear conformally invariant integral equations in $\mathbb{R}^3$”, *Adv. Math.* **194**:2 (2005), 485–503. MR Zbl

[Xu 2006] X. Xu, “Classification of solutions of certain fourth-order nonlinear elliptic equations in $\mathbb{R}^4$”, *Pacific J. Math.* **225**:2 (2006), 361–378. MR Zbl

[Zhu 2004] N. Zhu, “Classification of solutions of a conformally invariant third order equation in $\mathbb{R}^3$”, *Comm. Partial Differential Equations* **29**:11-12 (2004), 1755–1782. MR Zbl

[Zhuo et al. 2014] R. Zhuo, W. Chen, X. Cui, and Z. Yuan, “A Liouville theorem for the fractional Laplacian”, preprint, 2014. arXiv

Received July 21, 2017. Revised January 31, 2018.

WEI DAI
SCHOOL OF MATHEMATICS AND SYSTEMS SCIENCE
BEIHANG UNIVERSITY (BUAA)
BEIJING
CHINA
weidai@buaa.edu.cn

GUOLIN QIN
SCHOOL OF MATHEMATICS AND SYSTEMS SCIENCE
BEIHANG UNIVERSITY (BUAA)
BEIJING
CHINA
qinbuaa@foxmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.385

# THE PROJECTIVE LINEAR SUPERGROUP AND THE SUSY-PRESERVING AUTOMORPHISMS OF $\mathbb{P}^{1|1}$

RITA FIORESI AND STEPHEN D. KWOK

**The purpose of this paper is to describe the projective linear supergroup, its relation with the automorphisms of the projective superspace and to determine the supergroup of SUSY-preserving automorphisms of $\mathbb{P}^{1|1}$.**

## 1. Introduction

The works of Manin [1988; 1991] and more recently of Witten et al. [Witten 2012; Donagi and Witten 2015] have drawn attention to projective supergeometry and more specifically to SUSY curves and their moduli superspaces.

In this paper we study the automorphisms of the projective superspace $\mathbb{P}^{m|n}$ and its SUSY-preserving subsupergroup. We start by defining the projective linear supergroup $\mathrm{PGL}_{m|n}$, using the functor of points formalism, and then we show that this supergroup functor is indeed representable, that is, it is the functor of points of a superscheme. We achieve this by realizing $\mathrm{PGL}_{m|n}$ as a closed subsupergroup scheme of $\mathrm{GL}_{m^2+n^2|2mn}$, mimicking the ordinary procedure.

In relating this supergroup scheme to the automorphism supergroup of $\mathbb{P}^{m|n}$ we encounter a difficulty, not present in the ordinary setting, namely the fact that the Picard group of the projective superspace is not known in general and involves some difficulties. This is a consequence of the fact that the supergroup of automorphisms of the projective superspace is larger than $\mathrm{PGL}_{m|n}$ for $n > 1$. Nevertheless, going to the special case of $n = 1$, we are able to give the projective linear supergroup quite explicitly and to prove it coincides with the automorphisms of the projective superspace.

The question of singling out the SUSY-preserving automorphisms inside this supergroup was already settled over the complex field by Manin [1991] and Witten [2012]; we extend their considerations to an arbitrary algebraically closed field $k$, $\mathrm{char}(k) \neq 2$, and provide some extra details of their proofs.

The organization of this paper is as follows. In Section 2 we start by reviewing some generally known facts on the projective superspace and its functor of points to establish our notation. We then discuss line bundles and projective morphisms,

*MSC2010:* 14L30, 58A50.
*Keywords:* supergeometry, Lie theory, algebraic geometry.

proving, in Proposition 2.3, that the Picard group of $\mathbb{P}^{m|1}$ is $\mathbb{Z}$. To our knowledge this result is new and gives insight into projective supergeometry. In Section 3 we define the projective linear supergroup in terms of functor of points and we prove its representability by realizing it as a closed subsuperscheme of the general linear supergroup. Then, in Section 4 we prove that the projective linear supergroup is the supergroup of automorphisms of the projective superspace in the case of one odd dimension. Though the approach in both Sections 3 and 4 closely resembles the ordinary one, the results are novel in the supergeometric context. In Section 5, we use the machinery developed previously to prove that the subsupergroup of $\mathrm{Aut}(\mathbb{P}^{1|1})$ of SUSY-preserving automorphisms of $\mathbb{P}^{1|1}$ consists precisely of the irreducible component $(\mathrm{SpO}_{2|1})^0$ of the $2|1$-symplectic-orthogonal supergroup $\mathrm{SpO}_{2|1}$ containing the identity. This section is a generalization of the claims made in [Manin 1991] regarding complex supergeometry and provides proofs for such claims for a generic algebraically closed field.

## 2. The projective superspace $\mathbb{P}^{m|n}$

In this section we want to recall different, but equivalent definitions of projective superspace and we describe the line bundles on it. For all of our notation and main definitions of supergeometry, we refer the reader to [Manin 1988; Deligne and Morgan 1999; Carmeli et al. 2011].

Let $k$ be our ground ring.

We recall that, by definition, the functor of points of a superscheme $X=(|X|,\mathcal{O}_X)$ is the functor

$$X:(\text{sschemes})^o\to(\text{sets}),\quad X(S)=\mathrm{Hom}_{(\text{sschemes})}(S,X),\quad X(\phi)(f)=f\circ\phi,$$

where (sschemes) denotes the category of superschemes (it is customary to use the same letter for $X$ and its functor of points). Equivalently (see [Carmeli et al. 2011, Chapter 10]), we can view the functor of points of $X$ as $X:(\text{salg})\to(\text{sets})$:

$$X(R)=\mathrm{Hom}_{(\text{sschemes})}(\underline{\mathrm{Spec}}\,R,X),\quad X(\phi)(f)=f\circ\underline{\mathrm{Spec}}\,(\phi),$$

where (salg) denotes the category of superalgebras (over $k$), (we shall use the same letter for this functor also). In fact the functor of points of a superscheme is determined by its behavior on the affine superscheme subcategory, which in turn is equivalent to the category of superalgebras; see [Carmeli et al. 2011, Chapter 10, Theorem 10.2.5]. If $X=\underline{\mathrm{Spec}}\,\mathcal{O}(X)$, that is, $X$ is affine, we have that

$$X(R)=\mathrm{Hom}_{(\text{sschemes})}(\underline{\mathrm{Spec}}\,R,X)=\mathrm{Hom}_{(\text{salg})}(\mathcal{O}(X),R),$$

where $\mathcal{O}(X)$ denotes the superalgebra of global sections of the sheaf of superalgebras $\mathcal{O}_X$. We say that the $X(R)$ are the *$R$-points* of the superscheme $X$.

The algebraic superscheme $\mathbb{P}^{m|n}$ is defined as the patching of the $m+1$ affine superspaces $U_i = \underline{\mathrm{Spec}}\, \mathcal{O}(U_i)$, with $\mathcal{O}(U_i) = \underline{\mathrm{Spec}}\, k[x_0^i, \dots, \hat{x}_i^i, \dots, x_m^i, \xi_1^i, \dots, \xi_n^i]$ through the change of charts:

$$\begin{aligned} \phi_{ij} : \mathcal{O}(U_j)[(x_i^j)^{-1}] &\mapsto \mathcal{O}(U_i)[(x_j^i)^{-1}] \\ x_k^j &\mapsto x_k^i / x_j^i \\ x_i^j &\mapsto 1/x_j^i \\ \xi_k^j &\mapsto \xi_k^i / x_j^i, \end{aligned} \tag{1}$$

(where as usual $\hat{x}_i^i$ means that we are omitting the indeterminate $x_i^i$). Notice that $\mathcal{O}(U_j)[(x_i^j)^{-1}]$ is the superalgebra representing the open subscheme $U_j \cap U_i$ of $U_j$ (and similarly for $\mathcal{O}(U_i)[(x_j^i)^{-1}]$).

**Proposition 2.1.** *The $R$-points of $\mathbb{P}^{m|n}$, $R \in$ (salg) are given equivalently by*:

(i)
$$\mathbb{P}^{m|n}(R) = \{\alpha : R^{m+1|n} \to L,\ R\text{-linear, surjective}\}/\sim,$$
$$\mathbb{P}^{m|n}(\psi) : R^{m+1|n} \otimes_R T \to L \otimes_R T,$$

*where $L$ is locally free of rank $1|0$, $\psi : R \to T$ and $\alpha : R^{m+1|n} \to L \sim \alpha' : R^{m+1|n} \to L'$ if and only if $\ker(\alpha) = \ker(\alpha')$ (or equivalently, $\alpha \sim \alpha'$ if they differ by an automorphism of $L$ by multiplication of an element in $R^\times$).*

(ii)
$$\mathbb{P}^{m|n}(R) = \{\alpha : L \hookrightarrow R^{m+1|n}\ R\text{-linear, injective}\},$$
$$\mathbb{P}^{m|n}(\psi) : L \otimes_R T \to R^{m+1|n} \otimes_R T,$$

*where $L$ is locally free of rank $1|0$.*

*Let $\mathcal{O}_S^{m+1|n} = \mathcal{O}_S \otimes k^{m+1|n}$. The $S$-points of $\mathbb{P}^{m|n}$, $S \in$ (sschemes) are given equivalently by*:

(a)
$$\mathbb{P}^{m|n}(S) = \{\alpha : \mathcal{O}_S^{m+1|n} \to \mathcal{L},\ \text{surjective}\}/\sim,$$
$$\mathbb{P}^{m|n}(\psi) : (\psi^* \mathcal{O}_S)^{m+1|n} \to \psi^*(\mathcal{L}),$$

*where $\psi : T \to S$, $\mathcal{L}$ is a line bundle on $S$ (of rank $1|0$) and*

$$\alpha : \mathcal{O}_S^{m+1|n} \to \mathcal{L} \sim \alpha' : \mathcal{O}_S^{m+1|n} \to \mathcal{L}'$$

*if and only if $\ker(\alpha) = \ker(\alpha')$ (or equivalently, $\alpha \sim \alpha'$ if they differ by an automorphism of $\mathcal{L}$ by multiplication of an element in $\mathcal{O}_S^\times$).*

(b)
$$\mathbb{P}^{m|n}(S) = \{\alpha : \mathcal{L} \hookrightarrow \mathcal{O}_S^{m+1|n}\},$$
$$\mathbb{P}^{m|n}(\psi) : \psi^* \mathcal{L} \to (\psi^* \mathcal{O}_S)^{m+1|n}.$$

*Proof.* The proof relative to (i) and (a) works as in the ordinary setting and it is detailed in [Carmeli et al. 2011, Chapter 10]. The equivalence with (ii) and (b) is immediate. The equivalence between (i) and (ii) is essentially the same as in the ordinary setting (see [Eisenbud and Harris 2000, Chapter III, Section 2, Proposition III-40, Corollary III-42]). □

For every $A \in$ (salg), we denote by $(\text{salg})_A$ the category of superalgebras over $A$. We will need to consider also $\mathbb{P}_A^{m|n}$, that is, the projective superspace over a base $A \in$ (salg). This means that we are considering the superscheme obtained by patching the affine superspaces $U_i = A[x_j^i, \xi_k^i]$, $i, j = 0, \dots, m$, $j \neq i$, $k = 1, \dots, n$ as above. For example, in the second case in Proposition 2.1 each of the $T$-points, $T \in (\text{salg})_A$, is identified with a morphism $\alpha : L \to T^{m+1|n}$ of $A$-modules, where $L$ and $T^{m+1|n}$ are $T$-modules which become $A$-modules via the map $\phi : A \to T$:

$$\mathbb{P}_A^{m|n}(T) = \text{Hom}_{(\text{sschemes})_A}\big(\underline{\text{Spec}}\, T, \mathbb{P}_A^{m+1|n}\big) = \{\alpha : L \hookrightarrow T^{m+1|n}\}. \tag{2}$$

Notice that the functor of points of $\mathbb{P}_A^{m|n}$ is defined on the category of $A$-superalgebras or equivalently on the category of $A$-superschemes (that is, superschemes equipped with a morphism to the superscheme $\underline{\text{Spec}}\, A$ and morphisms compatible with it).

We leave to the reader the generalization of the other cases of Proposition 2.1 since it is straightforward.

We end this section with some observations on line bundles and morphisms on $\mathbb{P}_A^{m|n}$. We start with a result completely similar to the ordinary counterpart, left to the reader as a simple exercise; see also [Carmeli et al. 2011, Chapter 9].

**Proposition 2.2.** *We have a bijective correspondence between the following*:

(i) *The set of equivalence classes of $m+n+2$-tuples $(L, s_0, \dots, s_m, \sigma_1, \dots, \sigma_n)$, where $L$ is a line bundle on $\mathbb{P}_A^{m|n}$ globally generated by the global sections $s_0, \dots, s_m, \sigma_1, \dots, \sigma_n$ of $L$, under the relation*

$$(L, s_0, \dots, s_m, \sigma_1, \dots, \sigma_n) \sim (L, s_0', \dots, s_m', \sigma_1', \dots, \sigma_n')$$

*if and only if there exists some $c \in \mathcal{O}(\mathbb{P}_A^{m|n})_0^*$ such that $s_i' = cs_i$ and $\sigma_i' = c\sigma_i$ for all $i$.*

(ii) *The set of $A$-morphisms $\mathbb{P}_A^{m|n} \to \mathbb{P}_A^{m|n}$.*

In the ordinary setting we have that a line bundle on $\mathbb{P}_A^m$ is of the form $\mathcal{O}(n) \otimes \mathcal{L}$, where $\mathcal{L}$ is a line bundle on $\underline{\text{Spec}}\, A$. This nontrivial fact is still true in supergeometry for $\mathbb{P}_A^{m|1}$, and it will turn out to be crucial in our treatment.

**Proposition 2.3.** *Every line bundle on $\mathbb{P}_A^{m|1}$ is isomorphic to $\mathcal{O}(n) \otimes \mathcal{L}$, where $\mathcal{L}$ is a line bundle on* $\underline{\text{Spec}}$ *$A$.*

*Proof.* A line bundle on $\mathbb{P}_A^{m|1}$ is determined once we know its transition functions, say $g_{ij} \in \mathcal{O}_{\mathbb{P}_A^{m|1}}(U_i \cup U_j)_0^*$, which are even. We then need to prove that any such set of transition functions is equivalent, up to a coboundary, to a set of transition functions for a line bundle of the form $\mathcal{O}(n)\otimes\mathcal{L}$, for $\mathcal{L}$ a line bundle on $\underline{\mathrm{Spec}}\, A$. In other words we need to show

$$h_i|_{U_i\cap U_j}\, g_{ij}\, h_j^{-1}|_{U_i\cap U_j} = (x_j^i)^n, \qquad h_i \in \mathcal{O}_{\mathbb{P}_A^{m|1}}(U_i)_0^*.$$

Notice that

$$\mathcal{O}_{\mathbb{P}_A^{m|1}}(U_p)^* = (A[x_k^p, \xi^p])_0^* = (A[\xi^p][x_k^p])_0^*, \quad p = i, j.$$

Since $\phi_{ij}(\xi^j) = \xi^i/x_j^i$, $\phi_{ij}(x_i^j) = 1/x_j^i$ and $\phi_{ij}(x_k^j) = x_k^i/x_j^i$, where $\phi_{ij}$ is the change of chart as in (1), we can view the restrictions of the $h_p$'s $(p = i, j)$ to $U_i \cap U_j$, through this identification, as both belonging to $(A[\xi^i][x_j^i, (x_j^i)^{-1}])_0^*$. We now apply the classical result and obtain $h'_p \in (A[\xi^i][x_j^i, (x_j^i)^{-1}])_0^*$ such that

$$h'_i g_{ij} (h'_j)^{-1} = (x_j^i)^n.$$

The $h'_p$'s thus obtained are not yet the sections we want; since the odd dimension is one by hypothesis, the most general possible form for $h'_j$ is

$$h'_j = a_0 + \alpha_0\xi^i + \sum_K a_K x_K^i (x_j^i)^{-|K|} + \sum_L \alpha_L x_L^i (x_j^i)^{-|L|}\xi^i + \sum_k \beta_k (x_j^i)^{-k}\xi^i,$$

where $K$ and $L$ are multi-indices, $K = (k_1, \dots, k_r)$, $k_l \neq j$ $(r \in \mathbb{N})$ and $x_K^i := x_{k_1}^i \cdots x_{k_r}^i$ (similarly for $L$).

In order to eliminate the term $\alpha_0\xi^i$ which is not well defined on $U_j$, we define:

$$h_i := (a_0 + \alpha_0\xi^i)h'_i, \qquad h_j := (a_0^{-1} - a_0^{-2}\alpha_0\xi^i)h'_j,$$

and this gives the required sections. □

Notice that it was absolutely fundamental for our argument that there is only one odd dimension. This calculation will give us key information when we want to determine the automorphism supergroup of the projective linear supergroup.

## 3. The projective linear supergroup

In this section we want to define the supergroup functor of the projective linear supergroup and to show it is representable by producing an embedding of it as a closed subgroup into the general linear supergroup.

Let $\underline{\mathrm{M}}_{m|n}(R)$ denote the associative superalgebra of supermatrices of order $m|n$ by $m|n$ with entries in a commutative superalgebra $R$. More intrinsically, $\underline{\mathrm{M}}_{m|n}(R) = \underline{\mathrm{End}}_R(R^{m|n})$.

**Definition 3.1.** The *automorphism supergroup of supermatrices* is the supergroup functor $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n}) : (\mathrm{salg}) \to (\mathrm{grps})$,

$$[\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})](R) := \{f : \underline{\mathrm{M}}_{m|n}(R) \to \underline{\mathrm{M}}_{m|n}(R) \mid f \text{ is an } R\text{-superalgebra automorphism}\}.$$

In analogy with the ordinary setting we also will call this supergroup functor the *projective linear supergroup* and denote it with $\mathrm{PGL}_{m|n}$.

Since $\underline{\mathrm{M}}_{m|n}(R)$ is itself a free $R$-module of rank $M|N$, where $M = m^2 + n^2$ and $N = 2mn$, $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})$ is a subfunctor of $\mathrm{GL}_{M|N}$ in a natural way. We want to prove this is the functor of points of a closed subsuperscheme of $\mathrm{GL}_{M|N}$. Before proceeding we need a lemma characterizing the morphisms of the superalgebra of supermatrices.

**Lemma 3.2.** (i) *An $R$-linear parity-preserving map $\psi : \underline{\mathrm{M}}_{m|n}(R) \to \underline{\mathrm{M}}_{m|n}(R)$ is a morphism of the superalgebra of supermatrices $\underline{\mathrm{M}}_{m|n}(R)$ if and only if*

(a) $\psi(\mathrm{id}) = \mathrm{id}$;

(b) $\psi(e_{ij})\psi(e_{kl}) = \delta_{kj}\psi(e_{il})$,

*where $e_{ij}$ are the elementary matrices in $\underline{\mathrm{M}}_{m|n}(R)$.*

(ii) *If $R$ is a local superalgebra, all of the automorphisms of the superalgebra $\underline{\mathrm{M}}_{m|n}(R)$ are of the form*

$$\mathrm{M}_{m|n}(R) \to \mathrm{M}_{m|n}(R), (T, X) \mapsto TXT^{-1},$$

*for a suitable $T \in \mathrm{GL}_{m|n}(R)$.*

(iii) $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})$ *is a closed subsuperscheme of* $\mathrm{GL}_{M|N} = \underline{\mathrm{Spec}}\, k[x_{ij,kl}][d_1^{-1}, d_2^{-1}]$, $M = m^2 + n^2$ *and* $N = 2mn$, *defined by the equations*:

$$\sum_k x_{ij,kk} = \delta_{ij}, \qquad \sum_s x_{rs,ij}x_{st,kl} = \delta_{jk}x_{rt,il}, \tag{3}$$

*where $\mathrm{GL}_{M|N}(R)$ is identified with the parity-preserving automorphisms of the free $R$-module $\underline{\mathrm{M}}_{m|n}(R)$.*

*Proof.* (i) If $\psi$ is an $R$-superalgebra endomorphism of $\underline{\mathrm{M}}_{m|n}(R)$ then the two relations are obviously satisfied and vice versa.

(ii) Now assume $\psi$ is an automorphism of $\mathrm{M}_{m|n}(R)$, $R$ local, which satisfies the relations (a) and (b). We need to find $T \in \mathrm{GL}_{m|n}(R)$ such that $\psi(e_{ij}) = Te_{ij}T^{-1}$. This is an application of super Morita theory (see [Kwok 2013]), however we shall recall the main idea to make this proof self-contained. By (a) and (b) we have

$$\sum \psi(e_{ii}) = \mathrm{id}, \qquad \psi(e_{ii})^2 = \psi(e_{ii}), \qquad \psi(e_{ii})\psi(e_{jj}) = 0, \quad i \neq j,$$

hence we can write

$$R^{m|n} = \oplus\psi(e_{ii})R^{m|n}.$$

Since by (b), $\psi(e_{ji})\psi(e_{ii}) = \psi(e_{ji}) = \psi(e_{jj})\psi(e_{ji})$ we have $\psi(e_{ji}) : \psi(e_{ii})R^{m|n} \to \psi(e_{jj})R^{m|n}$ (recall that $R$ is local so projective implies free). Hence there exists a basis $\{t_i\}$ of the free module $R^{m|n}$ such that

$$\psi(e_{ii})R^{m|n} = \mathrm{span}_R\{t_i\}$$

and $\psi(e_{ji})t_i = t_j$. Let $T$ be the matrix whose columns are the $t_i$'s, $T = \sum t_i \otimes e_i^*$, $T^{-1} = \sum e_i \otimes t_i^*$. It is then immediate to verify $\psi(e_{ij}) = T e_{ij} T^{-1}$.

(iii). This is immediate from (i). □

Let us view the multiplicative algebraic supergroup $\mathbb{G}_m^{1|0}$ : (salg) → (grps) as the following subsupergroup of $\mathrm{GL}_{m|n}$:

$$\mathbb{G}_m^{1|0}(R) = \{aI \mid a \in R_0^*\} \subset \mathrm{GL}_{m|n}(R).$$

(Here $I$ denotes the identity matrix).

We do not specify the definition on the arrows whenever it is clear, as in this case.

**Definition 3.3.** We define the supergroup functor: $\widehat{\mathrm{PGL}}_{m|n}$ : (salg) → (grps),

$$\widehat{\mathrm{PGL}}_{m|n}(R) = \mathrm{GL}_{m|n}(R)/\mathbb{G}_m^{1|0}(R),$$

and we call its sheafification (as customary) $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$.

We wish to show that $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ is representable and coincides with the projective linear supergroup, that is, with the automorphism supergroup of supermatrices.

**Definition 3.4.** We say that a functor $F$ : (salg) → (grps) is *stalky* if for any superalgebra $R$, the natural map

$$\varinjlim_{f \notin \mathfrak{p}} F(R_f) \to F(R_{\mathfrak{p}})$$

is an isomorphism for any prime ideal $\mathfrak{p} \in R_0$.

The next two lemmas are standard and their proof is the same as in the ordinary case; see [Sun 2009].

**Lemma 3.5.** $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ *and* $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})$ *are stalky.*

**Lemma 3.6.** *Let $\mathcal{F}$, $\mathcal{G}$ be stalky Zariski sheaves* (salg) → (grps) *and $\alpha : \mathcal{F} \to \mathcal{G}$ be a morphism. If $\alpha_R : \mathcal{F}(R) \to \mathcal{G}(R)$ is an isomorphism for all local superrings $R$, then $\alpha$ is an isomorphism of sheaves.*

**Proposition 3.7.** *The supergroup functor* $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ *is representable and is realized as the closed subsupergroup* $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})$ *of* $\mathrm{GL}_{M|N}$ *for $M = m^2 + n^2$ and $N = 2mn$.*

*Proof.* We need to establish an isomorphism of sheaves between $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ and a closed subsupergroup of $\mathrm{GL}_{M|N}$. We will first give a morphism of sheaves and then show it is an isomorphism on local superalgebras; since $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ is a stalky sheaf, this will be enough. We start by giving a morphism of presheaves $\widehat{\mathrm{PGL}}_{m|n}$ and $\mathrm{GL}_{M|N}$; since $\mathrm{GL}_{M|N}$ is a sheaf then such a morphism will factor through the sheafification of $\widehat{\mathrm{PGL}}_{m|n}$ thus giving us a sheaf morphism.

Consider the action of $\mathrm{GL}_{M|N}$ on supermatrices $\underline{\mathrm{M}}_{m|n}$, where $M = m^2 + n^2$, $N = 2mn$:

$$\phi : \mathrm{GL}_{m|n}(R) \times \underline{\mathrm{M}}_{m|n}(R) \to \underline{\mathrm{M}}_{m|n}(R), \qquad (T, X) \mapsto TXT^{-1}.$$

This clearly factors through $\mathbb{G}_m^{1|0}(R)$ and hence gives a well defined action $\rho$ of $\widehat{\mathrm{PGL}}_{m|n}$ and then in turn of $\mathrm{GL}_{m|n}/\mathbb{G}^{1|0}$ (see comments at the beginning of the proof). Since $X \mapsto TXT^{-1}$ and $T \in (\mathrm{GL}_{m|n}/\mathbb{G}^{1|0})(R)$ is a parity-preserving $R$-superalgebra morphism, it is immediate to verify we have a morphism of sheaves,

$$\mathrm{GL}_{m|n}/\mathbb{G}^{1|0} \to \mathrm{Aut}\big(\underline{\mathrm{M}}_{m|n}\big).$$

By the first part of Lemma 3.2, we know that $\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})$ is represented by the closed subsuperscheme $H$ of $\mathrm{GL}_{M|N} = \underline{\mathrm{Spec}}\, k[x_{ij,kl}][d_1^{-1}, d_2^{-1}]$ defined by the equations

$$\sum_k x_{ij,kk} = \delta_{ij}, \qquad \sum_s x_{rs,ij} x_{st,kl} = \delta_{jk} x_{rt,il}. \tag{4}$$

(Here $d_i$ denotes as usual the determinants of the diagonal blocks of indeterminates). We want to show that the group homomorphism $(\mathrm{GL}_{m|n}/\mathbb{G}^{1|0})(R) \to [\mathrm{Aut}(\underline{\mathrm{M}}_{m|n})](R)$ is an isomorphism for $R$ local. The automorphism $\psi \in \mathrm{GL}_{M|N}(R)$ belongs to $H(R)$ if and only if its entries $\psi(e_{ij})_{kl}$ satisfy the above relations (4) (where in our convention $x_{ij,kl}$ corresponds to $\psi(e_{ij})_{kl}$). Hence by Lemma 3.2 we have the result for $R$ local. By Lemmas 3.5 and 3.6, it is true for any superalgebra $R$ and this concludes the proof. □

**Remark 3.8.** The projective linear supergroup may also be obtained through the Chevalley supergroup recipe as detailed in [Fioresi and Gavarini 2011; 2012; 2013]. It corresponds to the choice of the adjoint action of the Lie superalgebra $\mathfrak{sl}_{m|n}$. In fact one may readily check that the Lie superalgebra of $\mathrm{PGL}_{m|n}$ is indeed $\mathfrak{sl}_{m|n}$ and $(\mathrm{PGL}_{m|n})_0 = \mathrm{PGL}_m \times \mathrm{PGL}_n \times k^\times$.

## 4. The automorphisms of the projective superspace

We want to define the automorphism supergroup of the superscheme $\mathbb{P}^{m|n}$.

**Definition 4.1.** We define the supergroup functor of *automorphisms of the projective superspace*:

$$\mathrm{Aut}(\mathbb{P}^{m|n})(A) := \mathrm{Aut}_A(\mathbb{P}^{m|n} \times \underline{\mathrm{Spec}}\, A) = \mathrm{Aut}_A \mathbb{P}_A^{m|n}, \quad A \in (\mathrm{salg}).$$

$\mathrm{Aut}(\mathbb{P}^{m|n})$ is defined in an obvious way on the morphisms.

The equality in the definition is straightforward, noticing that we can identify the $T$-points of $\mathbb{P}^{m|n} \times \underline{\mathrm{Spec}}\, A$ and of $\mathbb{P}_A^{m|n}$. In fact, a $T$-point of $\mathbb{P}^{m|n} \times \underline{\mathrm{Spec}}\, A$ is a morphism $\phi : A \to T$ and a morphism $L \to T^{m|n}$ of $A$-modules via $\phi$. This is exactly an element of $\mathbb{P}_A^{m|n}(T)$ and vice versa.

An automorphism $\psi \in \mathrm{Aut}_A \mathbb{P}_A^{m|n}$ is a family of automorphisms $\psi_T$ for all $T \in (\mathrm{salg})_A$, which is functorial in $T$. The automorphism $\psi_T : \mathbb{P}_A^{m|n}(T) \to \mathbb{P}_A^{m|n}(T)$ must assign to a $T$-point of $\mathbb{P}_A^{m|n}(T)$, that is, a morphism $\alpha : L \to T^{m|n}$, another morphism $\alpha' : L' \to T^{m|n}$, where $L$ and $L'$ are projective rank $1|0$ $T$-modules, where the morphisms are interpreted as $A$-module morphisms. Similarly for the other characterizations of $T$-points as in Proposition 2.1.

We are now ready to relate the supergroup scheme $\mathrm{PGL}_{m|n}$ with the automorphisms of $\mathbb{P}^{m-1|n}$.

**Proposition 4.2.** *There is an embedding of supergroup functors*

$$\mathrm{PGL}_{m|n} \hookrightarrow \mathrm{Aut}(\mathbb{P}^{m-1|n}).$$

*Proof.* We first establish a morphism $\phi' : \mathrm{GL}_{m|n} \to \mathrm{Aut}(\mathbb{P}^{m-1|n})$. If $X \in \mathrm{GL}_{m|n}(A)$ and $\alpha \in \mathbb{P}_A^{m-1|n}(T) = \{T^{m|n} \to L\}/\sim,\ \psi : A \to T$ we define

$$\phi'(X) = \alpha \circ \mathrm{GL}_{m|n}(\psi)(X).$$

Clearly $\phi'$ factors through $\mathbb{G}_m(A)$. Since $\mathrm{Aut}(\mathbb{P}^{m-1|1})$ is a sheaf, we have defined a morphism

$$\phi : \mathrm{PGL}_{m|n} \to \mathrm{Aut}(\mathbb{P}^{m-1|n}).$$

The injectivity is clear. □

**Remark 4.3.** In general we cannot expect to get an isomorphism between $\mathrm{PGL}_{m|n}$ and $\mathrm{Aut}(\mathbb{P}^{m-1|n})$ for $n > 1$ and this is because of the peculiarity of the odd elements. Let us see this in a simple example, $\mathbb{P}^{1|2}$. Consider the morphism $\phi \in \mathbb{P}_A^{1|2}$ given on the affine pieces $U_0 = \underline{\mathrm{Spec}}\, A[u, \mu_1, \mu_2]$ and $U_1 = A[v, \nu_1, \nu_2]$ by

$$\phi|_{U_0}(u, \mu_1, \mu_2) = (u + \mu_1\mu_2, \mu_1, \mu_2), \qquad \phi|_{U_1}(v, \nu_1, \nu_2) = (v - \nu_1\nu_2, \nu_1, \nu_2).$$

As $\phi$ is invertible, $\phi \in \mathrm{Aut}(\mathbb{P}^{m|n})(A)$, but it is not obtained through an element of $\mathrm{PGL}_{2|2}(A)$. In fact the coefficient in $\phi|_{U_0}$ of $\mu_1\mu_2$ in an automorphism induced by a $\mathrm{PGL}_{2|2}(A)$ transformation must be nilpotent. Hence $\phi \notin \mathrm{PGL}_{2|2}(A)$.

We now want to show that we have an isomorphism between the projective linear supergroup and the automorphism of the super projective when $n = 1$. The argument we give follows along the lines of the calculation of $\mathrm{Aut}(\mathbb{P}^n)$ given in [Hartshorne 1977, Chapter 2, Section 7].

**Proposition 4.4.** *We have an isomorphism of supergroup functors*:

$$\mathrm{PGL}_{m+1|1} \cong \mathrm{Aut}(\mathbb{P}^{m|1}).$$

*In particular*, $\mathrm{Aut}(\mathbb{P}^{m|1})$ *is a supergroup scheme.*

*Proof.* Proposition 4.2 gives us an embedding of supergroup functors $\mathrm{PGL}_{m+1|1} \hookrightarrow \mathrm{Aut}(\mathbb{P}^{m|1})$. Now let $f \in \mathrm{Aut}(\mathbb{P}_A^{m|1})$ and let $g$ be its inverse. We want to show $f \in \mathrm{PGL}_{m+1|1}(A)$. The automorphism $f$ induces the two line bundle morphisms

$f^*\mathcal{O}_A(1) \to \mathcal{O}_A(1)$ and $g^*\mathcal{O}_A(1) \to \mathcal{O}_A(1)$, where $\mathcal{O}_A(1) := p_1^*(\mathcal{O}(1))$, with $p_1 : \mathbb{P}_A^{m|1} \to \mathbb{P}^{m|1}$ being the natural projection. By Proposition 2.3, we know that $f^*\mathcal{O}_A(1) = \mathcal{O}(k) \otimes \mathcal{L}_f$ and $g^*\mathcal{O}_A(1) = \mathcal{O}(l) \otimes \mathcal{L}_g$. Let us choose a suitable open cover of $A$ in which both $\mathcal{L}_f$ and $\mathcal{L}_g$ are trivial. By a common abuse of notation we shall still write $A$ to denote the ring of global sections of an element of the open cover, so we in fact are replacing $A$ with its localization. With such a choice we have $f^*\mathcal{O}_A(1) \cong \mathcal{O}_A(k)$ and $g^*\mathcal{O}_A(1) \cong \mathcal{O}_A(l)$. Since $f$ and $g$ are mutually inverse, we have

$$\mathcal{O}_A(1) = (f^* \circ g^*)(\mathcal{O}_A(1)) = f^*(g^*(\mathcal{O}_A(1))) = f^*(\mathcal{O}_A(l)) = \mathcal{O}_A(kl).$$

Hence $kl = 1$, whence $k = l = 1$, because for $k = l = -1$ we do not have global sections.

So $f^*(\mathcal{O}(1)) \cong \mathcal{O}(1)$, and choosing an isomorphism $F : f^*(\mathcal{O}(1)) \to \mathcal{O}(1)$ yields an isomorphism of the global sections $\Gamma(\mathbb{P}^m, f^*\mathcal{O}_A(1)) \cong \Gamma(\mathbb{P}^m, \mathcal{O}_A(1))$. By composing such an isomorphism with the natural isomorphism

$$f^* : \Gamma(\mathbb{P}^m, \mathcal{O}_A(1)) \to \Gamma(\mathbb{P}^m, f^*\mathcal{O}_A(1))$$

we obtain an $A$-linear automorphism,

$$T_F : \Gamma(\mathbb{P}^m, \mathcal{O}_A(1)) \to \Gamma(\mathbb{P}^m, \mathcal{O}_A(1)),$$

and identifying $\Gamma(\mathbb{P}^m, \mathcal{O}_A(1))$ with $A^{m+1|1}$ we see that $T_F \in \mathrm{GL}_{m+1|1}(A)$. However, $T_F$ depends on $F$. Suppose $G : f^*(\mathcal{O}(1)) \to \mathcal{O}(1)$ is another isomorphism, then $F^{-1} \circ G$ is an automorphism of $\mathcal{O}(1)$. Since $\underline{\mathrm{Hom}}(L, L) = L^* \otimes L = \mathcal{O}$ for any line bundle $L$, we see that an automorphism of $\mathcal{O}(1)$ is the same thing as an invertible even function on $\mathbb{P}_A^{m|1}$, and $F$ and $G$ differ by composing with multiplication by such a function.

Therefore $f$ determines $T_F$ only up to multiplication by an invertible even function, i.e., $f$ uniquely determines an element $T := [T_F]$ of $\mathrm{PGL}_{m+1|1}(A)$.

Now in suitable coordinates we have that $T$ induces (up to scalar multiplication) an automorphism of the $\mathbb{Z}$-graded superalgebra $A[z_0, \dots, z_m, \zeta]$. We leave to the reader the check that $\phi(T)$ is indeed $f$. □

## 5. The SUSY-preserving automorphisms of $\mathbb{P}_k^{1|1}$

In this section we want to consider those automorphisms of $\mathbb{P}_k^{1|1}$ which preserve its unique (up to isomorphism) SUSY structure. For all of the standard notation of supergeometry refer to [Carmeli et al. 2011].

Let $k$ be our ground field, $\mathrm{char}(k) \neq 2$, $k$ algebraically closed. All algebraic supergroups discussed below will be algebraic supergroups over $k$.

We recall that if $X$ is a smooth algebraic supervariety over $k$ of dimension $1|1$, we define a *SUSY structure* on $X$ as a $0|1$ distribution $\mathcal{D}$ on $X$ such that the Frobenius map

$$\mathcal{D} \otimes \mathcal{D} \to TX/\mathcal{D}, \qquad Y \otimes Z \mapsto [Y, Z] \bmod \mathcal{D}$$

is an isomorphism (see, for example, [Manin 1991] for the definition of a SUSY structure in the complex analytic case). If $X \to S$ is a smooth family of algebraic supervarieties of relative dimension $1|1$ over an algebraic $k$-supervariety $S$, then the notion of relative SUSY structure may be defined in the analogous way, as a relative distribution in the relative tangent sheaf $TX/S$. In this case we say that $X \to S$ is a *relative SUSY family*.

Our discussion is based on [Witten 2012].

We start by interpreting $\mathbb{P}_k^{1|1}$ as a homogeneous superspace. Let $\underline{k}^{2|1} = (k^2, \mathcal{O}_{\underline{k}^{2|1}})$ denote the affine superspace canonically associated to the $k$-super vector space $k^{2|1}$. Let us consider the action of the algebraic group $\underline{k}^\times$ on $\underline{k}^{2|1} \setminus \{0\}$, given in the functor of points notation by

$$t \cdot (z_0, z_1, \zeta) = (tz_0, tz_1, t\zeta).$$

Consider the projection (as topological map)

$$\pi : k^2 \setminus \{0\} \to k^2 \setminus \{0\}/k^\times \cong \mathbb{P}^1.$$

Define the sheaf on the topological space $\mathbb{P}_k^1$ consisting of the $\underline{k}^\times$-invariant sections

$$\mathcal{F}(U) := \mathcal{O}_{\underline{k}^{2|1}}(\pi^{-1}(U)))^{\underline{k}^\times}.$$

One can readily check that $(\mathbb{P}_k^1, \mathcal{F})$ is the superscheme $\mathbb{P}_k^{1|1}$ as defined in Section 2.

Let $z_0, z_1, \zeta$ be global coordinates on $\underline{k}^{2|1}$. We now consider the Euler vector field $E = z_0\partial_{z_0} + z_1\partial_{z_1} + \zeta\partial_\zeta$, which represents (in the chosen coordinates) the infinitesimal generator for the $\underline{k}^\times$ action on $\underline{k}^{2|1} \setminus \{0\}$. Since $E$ is everywhere nonsingular, it generates a trivial $1|0$ line bundle. As in the classical case, we have the Euler exact sequence of vector bundles on $\mathbb{P}_k^{1|1}$:

$$0 \to \mathcal{O}^{1|0} \xrightarrow{i} \mathcal{O}(1) \otimes \mathrm{Der}(S) \xrightarrow{j} T\mathbb{P}_k^{1|1} \to 0, \tag{5}$$

where $i$ is the inclusion of the trivial $1|0$ line bundle $\langle E \rangle$ with global basis the Euler vector field. Here $\mathrm{Der}(S)$ is the $k$-super vector space of $k$-linear derivations on $S := \underline{\mathrm{Sym}}((k^{2|1})^*)$; it has as basis the derivations $\partial_{z_i}, \partial_\zeta$. Thus $\mathcal{O}(1) \otimes \mathrm{Der}(S)$ is the sheaf whose sections on $U$ are the linear vector fields on $\pi^{-1}(U)$. Any local section of $\mathcal{O}(1) \otimes \mathrm{Der}(S)$ induces a corresponding local $k$-linear derivation on $\mathcal{O}_{\mathbb{P}_k^{1|1}}$ by restricting it to act on $\underline{k}^\times$-invariant functions; this defines $j$. Injectivity of $i$ and the inclusion $\mathrm{im}(i) \subseteq \ker(j)$ follow from the fact that $E$ is nonsingular and the infinitesimal generator for the $\underline{k}^\times$-action; a standard calculation in the usual affine cells shows that $\ker(j) \subseteq \mathrm{im}(i)$ and that $j$ is surjective. Note that the sequence continues to remain exact on $\mathbb{P}_A^{1|1}$ after base change to any affine $k$-supervariety

$\underline{\mathrm{Spec}}\,(A)$, with $T\mathbb{P}_k^{1|1}$ replaced by the relative tangent bundle $T\mathbb{P}_A^{1|1}/\mathrm{Spec}(A)$. We will denote the $A$-superalgebra $S\otimes_k A$ by $S_A$.

We now come to the SUSY structure.

**Definition 5.1.** Let $(X\to S, \mathcal{D})$ be a relative SUSY family. An $S$-automorphism $f: X\to X$ is *SUSY structure-preserving* (or simply *SUSY-preserving*) if and only if $(df_p)(\mathcal{D}_p)=\mathcal{D}_{f(p)}$ for any $p\in X$.

We will consider SUSY structures given by sections of $\mathcal{O}_A(1)\otimes\Omega_{S/A}$. Here $\Omega_{S/A}$ denotes the $A$-module of Kähler differentials on $S_A$, i.e., the $A$-dual to $\mathrm{Der}(S_A)$; it has as basis the differentials $dz_i, d\zeta$. When we speak of the kernel of a section $\omega$ of $\mathcal{O}_A(1)\otimes\Omega_{S/A}$, we mean the kernel of $\omega$ when $\omega$ is interpreted as a morphism of sheaves of $\mathcal{O}_{\mathbb{P}_A^{1|1}}$-modules from $\mathcal{O}_A(1)\otimes\mathrm{Der}(S_A)\to\mathcal{O}_A(2)$.

**Proposition 5.2.** *Let* $s:=z_1\,dz_0-z_0\,dz_1-\zeta\,d\zeta$. *Then the image of* $\ker(s)$ *under* $j$ *is a SUSY structure on* $\mathbb{P}_k^{1|1}$.

*Proof.* In the affine open subsupervariety $U_1:=\{z_1\neq 0\}\subset\mathbb{P}_k^{1|1}$, one calculates that the Euler vector field $E$ and the linear vector field $\widehat{Z}_1=\zeta\,\partial_{z_0}+z_1\partial_\zeta$ lie in $\ker(s)$ and are linearly independent. At any point $p\in\mathbb{P}_k^{1|1}$, $s$ induces a linear map of super vector spaces, $s_p:[\mathcal{O}(1)\otimes\mathrm{Der}(S)]_p\to[\mathcal{O}(2)]_p$, on the fibers. It is clear that $s$ is a basepoint-free section, hence $s_p$ is always surjective. By linear algebra, $\ker(s_p)$ is $1|1$ dimensional and hence $E_p$ and $\widehat{Z_{1,p}}$ span $\ker(s_p)$. By the super Nakayama's lemma, $E$ and $\widehat{Z_1}$ span $\ker(s)$ near $p$. Since $p$ was arbitrary, $E$ and $\widehat{Z_1}$ form a basis for $\ker(s)$ in $U_1$.

One sees that $Z_1:=j(\widehat{Z_1})=\partial_\eta+\eta\partial_w$, where $w=z_0/z_1$ and $\eta=\zeta/z_1$ are the usual affine coordinates in $U_1$. $Z_1^2=\partial_w$ and so $Z_1$ defines a SUSY structure in $U_1$. A similar calculation with the linear vector field $\widehat{Z}_0:=-\zeta\,\partial_{z_1}+z_0\partial_\zeta$ shows that $j(\ker(s))$ defines a SUSY structure on $U_0=\{z_0\neq 0\}$, hence the image of $\ker(s)$ under $j$ defines a SUSY structure on $\mathbb{P}_k^{1|1}$. □

We note that by the considerations of [Fioresi and Lledó 2015], this is the unique SUSY structure on $\mathbb{P}_k^{1|1}$, up to SUSY-isomorphism.

We now need the following proposition. The proof is completely similar to the one in [Fioresi and Lledó 2015, Proposition 5.2], however since the context here is more general, we include it for completeness.

**Lemma 5.3.** *Let $A$ be an affine $k$-superalgebra. Let $\omega, \omega'$ be two global sections of* $\mathcal{O}_A(1)\otimes\Omega_{S/A}$ *such that* $\mathcal{D}:=j(\ker(\omega))$ *and* $\mathcal{D}':=j(\ker(\omega'))$ *are* $0|1$ *distributions on* $\mathbb{P}_A^{1|1}$. *Suppose* $\mathcal{D}=\mathcal{D}'$. *Then* $\omega'=h\omega$ *for some even invertible function $h$ on* $\mathbb{P}_A^{1|1}$.

*Proof.* Let $p\in\mathbb{P}_A^{1|1}$ be a point. $\mathcal{D}$ is *locally a direct summand of* $T\mathbb{P}_A^{1|1}/\underline{\mathrm{Spec}}\,(A)$, so we have a local splitting $\mathcal{D}|_U\oplus\mathcal{E}=(T\mathbb{P}_A^{1|1}/\underline{\mathrm{Spec}}(A))|_U$ in some neighborhood $U\ni p$. Via the Euler exact sequence (base changed to $\mathrm{Spec}(A)$), we may lift $\mathcal{D}|_U$ (resp. $\mathcal{E}$) uniquely to a rank $1|1$ (resp. $2|0$) submodule $\widehat{\mathcal{D}}$ (resp. $\widehat{\mathcal{E}}$) of $[\mathcal{O}_A(1)\otimes\mathrm{Der}(S_A)]|_U$

containing the $1|0$ line bundle $\langle E\rangle$ spanned by the Euler vector field, such that $\widehat{\mathcal{D}}\cap\widehat{\mathcal{E}}=\langle E\rangle$. We may therefore find local sections $\widehat{Z}$ (resp. $\widehat{X}$) of $\widehat{\mathcal{D}}$ (resp. $\widehat{\mathcal{E}}$) such that $\widehat{Z}$, $E$ (resp. $\widehat{X}$, $E$) form a basis for $\widehat{\mathcal{D}}$ (resp. $\widehat{\mathcal{E}}$). Note that the condition $\widehat{\mathcal{D}}\cap\widehat{\mathcal{E}}=\langle E\rangle$ implies $\widehat{X}$, $\widehat{Z}$, $E$ form a basis of $[\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A)]|_U$.

Viewing $\omega|_U$ as an $\mathcal{O}_{\mathbb{P}^{1|1}_A}$-linear map from $[\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A)]|_U$ to $\mathcal{O}_A(2)|_U$, we have an induced linear map of super vector spaces,

$$\omega_p:(\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A))_p\to(\mathcal{O}_A(2))_p.$$

As $\ker(\omega_p)=\mathrm{span}\{\widehat{Z}_p,E_p\}$, we see by linear algebra that $\omega_p$ is a surjection, and that $\omega_p(\widehat{X}_p)$ is a basis for $(\mathcal{O}_A(2))_p$; the analogous conclusion holds for $\omega'_p$ and $\omega'_p(\widehat{X}_p)$. Hence by the super Nakayama's lemma, $\omega(\widehat{X})$ is a basis for $\mathcal{O}_A(2)|_U$, and the same is true of $\omega'(\widehat{X})$ (shrinking $U$ if necessary). Hence $\omega'(\widehat{X})/\omega(\widehat{X})$ is an invertible even function on $U$; let us denote it by $h$.

To show that $h$ is independent of the local complement $\mathcal{E}$ and the choice of basis element $\widehat{X}$, suppose $\mathcal{E}'$ is another local complement to $\mathcal{D}$ on $U$, and let $\widehat{X}'$, $E$ be a basis of the lift $\widehat{\mathcal{E}}'$ of $\mathcal{E}'$. Then we have $\widehat{X}'=a\widehat{X}+bE+\alpha\widehat{Z}$ for some $a,b,\alpha\in\mathcal{O}_{\mathbb{P}^{1|1}_A}(U)$, with $a,b$ even and $\alpha$ odd. As $\widehat{X}$, $E$, $\widehat{Z}$ and $\widehat{X}'$, $E$, $\widehat{Z}'$ are both local bases for $\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A)$, $a$ must be a unit.

Then we have

$$\omega'(\widehat{X}')/\omega(\widehat{X}')=\omega'(a\widehat{X}+bE+\alpha\widehat{Z})/\omega(a\widehat{X}+bE+\alpha\widehat{Z})=\omega'(\widehat{X})/\omega(\widehat{X}),$$

since $\omega$, $\omega'$ both annihilate $E$ and $\widehat{Z}$. This proves that the expression $\omega'(\widehat{X})/\omega(\widehat{X})$ is independent of all choices and hence $h$ is a well-defined function on all of $\mathbb{P}^{1|1}_A$. The equality $\omega'=h\omega$ clearly holds locally, and since $h$ is now known to be globally defined, it holds globally. □

**Proposition 5.4.** *Let $f$ be an automorphism of $\mathbb{P}^{1|1}_A$. Then $f$ preserves the SUSY structure defined by $s$ if and only if for some (hence every) lift $\tilde{f}$ of $f$ to* $\mathrm{GL}_{2|1}(A)$, *$\tilde{f}^*(s)=ts$ for some invertible function $t$.*

*Proof.* We begin by noting that $\mathrm{GL}_{2|1}(A)$ preserves $A_0^*$-invariant open subsets of $\mathbb{A}^{2|1}_A\setminus\{0\}$, hence it acts naturally by pullback of functions on $\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A)$, where we interpret the latter as the sheaf assigning to any open subset $U\subseteq\mathbb{P}^{1|1}_A$ the linear vector fields on $\pi^{-1}(U)\subseteq\mathbb{A}^{2|1}_A\setminus\{0\}$.

The subsupergroup of invertible scalar matrices $\{cI : c\in A_0^*\}$ is central in $\mathrm{GL}_{2|1}(A)$, hence this $\mathrm{GL}_{2|1}(A)$-action preserves the subalgebra of $A_0^*$-invariant functions on any $A_0^*$-invariant open subset of $\mathbb{A}^{2|1}_A\setminus\{0\}$. Hence we have an induced $\mathrm{GL}_{2|1}(A)$-action on the sheaf $\mathcal{O}_{\mathbb{P}^{1|1}_A}$. Clearly, invertible scalar matrices act trivially on $\mathcal{O}_{\mathbb{P}^{1|1}_A}$, thus the $\mathrm{GL}_{2|1}(A)$-action on $\mathcal{O}_{\mathbb{P}^{1|1}_A}$ factors through $\mathrm{PGL}_{2|1}(A)$.

We see from the above that the action of $\mathrm{GL}_{2|1}(A)$ on $\mathcal{O}_A(1)\otimes \mathrm{Der}(S_A)$ by pullback of functions induces naturally a $\mathrm{PGL}_{2|1}(A)$-action on $\mathcal{O}_{\mathbb{P}^{1|1}_A}$, hence on

$T\mathbb{P}^{1|1}_A/\underline{\mathrm{Spec}}\,(A)$, also given by pullback of functions. But this is precisely the $\mathrm{PGL}_{2|1}(A)$-action on $T\mathbb{P}^{1|1}_A/\underline{\mathrm{Spec}}\,(A)$ induced by the action of $\mathrm{PGL}_{2|1}(A)$ on $\mathbb{P}^{1|1}_A$ by automorphisms.

Since the sheaf morphism $j : \mathcal{O}_A(1) \otimes \mathrm{Der}(S_A) \to T\mathbb{P}^{1|1}_A/\underline{\mathrm{Spec}}\,(A)$ is just given by restricting a linear vector field to act on $A_0^*$-invariant functions, we see $j$ is equivariant with respect to the $\mathrm{GL}_{2|1}(A)$- and $\mathrm{PGL}_{2|1}(A)$-actions previously defined.

We also have a $\mathrm{GL}_{2|1}(A)$-action on $\mathcal{O}_A(1) \otimes \Omega_{S/A}$ by the natural action on both factors, and for $\omega \in \Gamma(\mathcal{O}_A(1) \otimes \Omega_{S/A}) = \Gamma(\mathcal{O}_A(1)) \otimes \Omega_{S/A}$, we write $g^*(\omega)$ for $g \cdot \omega$.

Since the action of $\mathrm{GL}_{2|1}(A)$ on $\mathcal{O}_A(1) \otimes \mathrm{Der}(S_A)$ is the same as the natural action on the individual factors, and the $\mathrm{GL}_{2|1}(A)$-action on $\Omega_{S/A}$ is dual to that on $\mathrm{Der}(S_A)$, it follows that the evaluation pairing

$$[\mathcal{O}_A(1) \otimes \mathrm{Der}(S_A)] \otimes [\mathcal{O}_A(1) \otimes \Omega_{S/A}] \to \mathcal{O}_A(2)$$

is $\mathrm{GL}_{2|1}(A)$-equivariant, where $\mathcal{O}_A(2)$ is endowed with the natural $\mathrm{GL}_{2|1}(A)$-action.

From the preceding discussion, we see that $f$ is SUSY-preserving if and only if $j\,[\ker(\omega)]_p = j\,[\ker(\tilde{f}^*(\omega)]_p$ for any point $p$.

We claim this is true if and only if $j\,[\ker(\omega)] = j\,[\ker(\tilde{f}^*(\omega))]$. One direction is clear. For the other, suppose $j\,[\ker(\omega)]_p = j\,[\ker(\tilde{f}^*(\omega))]_p$ for any point $p$. Then by the super Nakayama's lemma $j\,[\ker(\omega)] = j\,[\ker(\tilde{f}^*(\omega))]$ in a neighborhood of $p$, hence globally. The claim then follows from Lemma 5.3. □

In order to determine the supergroup of SUSY-preserving automorphisms of $\mathbb{P}^{1|1}_k$ we must discuss various other supergroups. We follow closely the discussion in [Manin 1991].

**Definition 5.5.** The $2|1$-dimensional *conformal symplectic-orthogonal supergroup* $\mathrm{C}_{2|1}$ is the subfunctor of $\mathrm{GL}_{2|1}$ that preserves, up to multiplication by an even invertible constant, the split nondegenerate supersymplectic form on $k^{2|1}$ given by $(v, w) = v^t H w$, where

$$H := \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{6}$$

and $^t$ denotes the super transpose of a matrix. More precisely, for every $k$-superalgebra $A$, $\mathrm{C}_{2|1}$ is the functor $(\mathrm{salg})_{\mathrm{k}} \to (\mathrm{grps})$ given by

$$\mathrm{C}_{2|1}(A) := \{B \in \mathrm{GL}_{2|1}(A) : B^t H B = Z(B) H\}, \tag{7}$$

where $Z : \mathrm{GL}_{2|1} \to \mathbb{G}_m^{1|0}$ is a fixed homomorphism.

The $2|1$-dimensional *projective conformal symplectic-orthogonal supergroup* $\mathrm{PC}_{2|1}$ is the image of $\mathrm{C}_{2|1}$ in $\mathrm{PGL}_{2|1}$, i.e, it is the sheafification of the group-valued functor $A \to \mathrm{C}_{2|1}(A)/\{aI : a \in A_0^*\}$.

**Proposition 5.6.** $\mathrm{C}_{2|1}$ *and* $\mathrm{PC}_{2|1}$ *are representable.*

*Proof.* Taking the Berezinian of both sides of (7), one sees that $Z(B) = \mathrm{Ber}(B)^2$. Thus, given

$$B = \begin{pmatrix} a & b & \alpha \\ c & d & \beta \\ \gamma & \delta & e \end{pmatrix} \in \mathrm{GL}_{2|1}(A),$$

a direct calculation shows that $B$ satisfies (7) if and only if the following equations hold: $e^2 + 2\alpha\beta = \mathrm{Ber}(B)^2$, $a\beta - c\alpha - e\gamma = 0$, $ad - bc - \gamma\delta = \mathrm{Ber}(B)^2$, $b\beta - d\alpha - e\delta = 0$. Thus these equations define $\mathrm{C}_{2|1}$ as a closed affine algebraic subsupergroup of $\mathrm{GL}_{2|1}$.

To prove that $\mathrm{PC}_{2|1}$ is representable, we use the trick of [Manin 1991]. Let $\mathrm{SC}_{2|1}$ denote the functor $(\mathrm{salg})_\mathrm{k} \to (\mathrm{grps})$ given by

$$\mathrm{SC}_{2|1}(A) := \{B \in \mathrm{C}_{2|1}(A) : \mathrm{Ber}(B) = 1\}.$$

Since its defining equations are those of $\mathrm{C}_{2|1}$ together with $\mathrm{Ber}(B) = 1$, $\mathrm{SC}_{2|1}$ is a closed affine algebraic subsupergroup of $\mathrm{GL}_{2|1}$. There is a short exact sequence of supergroups,

$$0 \to \mathrm{SC}_{2|1} \to \mathrm{C}_{2|1} \xrightarrow{\mathrm{Ber}} \mathbb{G}_m^{1|0} \to 0. \tag{8}$$

There is a splitting of this sequence, given on $A$-points by sending $a \in A_0^*$ to $aI$, and the image of $\mathbb{G}_m^{1|0}$ under the splitting is clearly normal in $\mathrm{C}_{2|1}$, hence $\mathrm{C}_{2|1}$ is the internal direct product of $\mathrm{SC}_{2|1}$ and the subsupergroup $\{aI : a \in A_0^*\}$. This direct product decomposition allows us to naturally identify the functor $\mathrm{PC}_{2|1}$ with the functor of points of $\mathrm{SC}_{2|1}$; in particular, we see $\mathrm{PC}_{2|1}$ is an affine algebraic supergroup, isomorphic to $\mathrm{SC}_{2|1}$. □

**Definition 5.7.** The $2|1$-dimensional *symplectic-orthogonal supergroup* $\mathrm{SpO}_{2|1}$ is the functor $(\mathrm{salg})_\mathrm{k} \to (\mathrm{grps})$,

$$\mathrm{SpO}_{2|1}(A) := \{B \in \mathrm{GL}_{2|1}(A) : B^t H B = H\}. \tag{9}$$

**Remark 5.8.** $\mathrm{SpO}_{2|1}$ is well known to be representable; the reader may readily write down defining equations for $\mathrm{SpO}_{2|1}$, completely analogous to those for $\mathrm{C}_{2|1}$, which show that $\mathrm{SpO}_{2|1}$ is a closed affine algebraic subsupergroup of $\mathrm{GL}_{2|1}$.

**Proposition 5.9.** $\mathrm{PC}_{2|1}$ *is isomorphic to the irreducible component* $(\mathrm{SpO}_{2|1})^0$ *of* $\mathrm{SpO}_{2|1}$ *containing the identity.*

*Proof.* Taking the Berezinian of both sides of (9) shows that $\mathrm{Ber}(B) = \pm 1$ for any $B \in \mathrm{SpO}_{2|1}(A)$. This yields a short exact sequence of supergroups

$$0 \to \mathrm{SC}_{2|1} \to \mathrm{SpO}_{2|1} \xrightarrow{\mathrm{Ber}} \{\pm 1\} \to 0, \tag{10}$$

which is split by the morphism $\pm 1 \mapsto \pm I$ and $\{\pm I\}$ is obviously normal in $\mathrm{SpO}_{2|1}$. Thus $\mathrm{SpO}_{2|1}$ is the internal direct product of $\{\pm I\}$ and $\mathrm{SC}_{2|1}$. Note that $\mathrm{SC}_{2|1}$ is irreducible (one sees from its defining equations that its reduced algebraic group is $\mathrm{SL}_2$, which is known to be irreducible). Let $(\mathrm{SpO}_{2|1})^0$ denote the irreducible component of $\mathrm{SpO}_{2|1}$ that contains the identity. We claim $\mathrm{SC}_{2|1} = (\mathrm{SpO}_{2|1})^0$. Since $I \in \mathrm{SC}_{2|1} \cap (\mathrm{SpO}_{2|1})^0$, it is clear $\mathrm{SC}_{2|1} \subseteq (\mathrm{SpO}_{2|1})^0$. Conversely, we see that $(\mathrm{SpO}_{2|1})^0 \subseteq \mathrm{SC}_{2|1}$: the restriction of the morphism Ber to the irreducible supervariety $(\mathrm{SpO}_{2|1})^0$ must be constant, hence equal to 1. Since we previously showed $\mathrm{PC}_{2|1}$ is isomorphic to $\mathrm{SC}_{2|1}$, the proposition is proven. □

**Theorem 5.10.** *The algebraic supergroup* $\mathrm{Aut}_{\mathrm{SUSY}}(\mathbb{P}_k^{1|1})$ *of SUSY-preserving automorphisms of* $\mathbb{P}_k^{1|1}$ *is isomorphic to* $(\mathrm{SpO}_{2|1})^0$.

*Proof.* As $\mathrm{Aut}_{\mathrm{SUSY}}(\mathbb{P}_k^{1|1})$ is a sheaf, the theorem reduces to the case of calculating $\mathrm{Aut}_{\mathrm{SUSY}}(\mathbb{P}_k^{1|1})(A)$ where $A$ is a $k$-superalgebra. For this, we note that $\mathbb{P}_A^{1|1}$ has the SUSY structure over $A$ induced by base change from $\mathbb{P}_k^{1|1}$, given by $s$.

Let $g \in \mathrm{PGL}_{2|1}(A)$ be an automorphism of $\mathbb{P}_A^{1|1}$, and $\hat{g}$ a lift of $g$ to $\mathrm{GL}_{2|1}(A)$. Recall that we have a natural action of the group of $A$-points of $\mathrm{GL}_{2|1}(A)$ on $\Gamma(\mathcal{O}_A(1) \otimes \Omega_{S/A})$. More concretely, in the given coordinates we have for any matrix $\hat{g} \in \mathrm{GL}_{2|1}(A)$,

$$\hat{g} \cdot \begin{pmatrix} z_0 \\ z_1 \\ \zeta \end{pmatrix} = \hat{g} \begin{pmatrix} z_0 \\ z_1 \\ \zeta \end{pmatrix}, \qquad \hat{g} \cdot \begin{pmatrix} dz_0 \\ dz_1 \\ d\zeta \end{pmatrix} = \hat{g} \begin{pmatrix} dz_0 \\ dz_1 \\ d\zeta \end{pmatrix},$$

where we write $z_i$ for $z_i \otimes 1$ and so on.

By Lemma 5.3, $g$ is SUSY-preserving if and only if $\hat{g}$ sends

$$s = z_1 dz_0 - z_0 dz_1 - \zeta d\zeta = \begin{pmatrix} z_0 & z_1 & \zeta \end{pmatrix} H \begin{pmatrix} dz_0 \\ dz_1 \\ d\zeta \end{pmatrix}, \qquad H = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

to a multiple of $s$ by an invertible even function. Hence

$$\begin{pmatrix} z_0 & z_1 & \zeta \end{pmatrix} \hat{g}^t H \hat{g} \begin{pmatrix} dz_0 \\ dz_1 \\ d\zeta \end{pmatrix} = \begin{pmatrix} z_0 & z_1 & \zeta \end{pmatrix} Z(\hat{g}) H \begin{pmatrix} dz_0 \\ dz_1 \\ d\zeta \end{pmatrix},$$

i.e., $\hat{g} \in \mathrm{C}_{2|1}(A)$. It follows from (8) that $g$ lies in $\mathrm{PC}_{2|1}(A)$, which is naturally identified with $(\mathrm{SpO}_{2|1})^0(A)$ by Proposition 5.9. □

***Acknowledgements.*** We are indebted to Prof. D. Gaitsgory for clarifying to us the structure of line bundles over $\mathbb{P}_A^n$ in the ordinary setting. We also thank Prof. L. Migliorini for helpful discussions. We are also grateful to the anonymous referee for suggestions and remarks on the paper.

## References

[Carmeli et al. 2011] C. Carmeli, L. Caston, and R. Fioresi, *Mathematical foundations of supersymmetry*, European Mathematical Society, Zürich, 2011. MR Zbl

[Deligne and Morgan 1999] P. Deligne and J. W. Morgan, "Notes on supersymmetry (following Joseph Bernstein)", pp. 41–97 in *Quantum fields and strings: a course for mathematicians, I* (Princeton, 1996/1997), edited by P. Deligne et al., American Mathematical Society, Providence, RI, 1999. MR Zbl

[Donagi and Witten 2015] R. Donagi and E. Witten, "Supermoduli space is not projected", pp. 19–71 in *String-math 2012*, edited by R. Donagi et al., Proc. Sympos. Pure Math. **90**, American Mathematical Society, Providence, RI, 2015. MR Zbl

[Eisenbud and Harris 2000] D. Eisenbud and J. Harris, *The geometry of schemes*, Graduate Texts in Mathematics **197**, Springer, 2000. MR Zbl

[Fioresi and Gavarini 2011] R. Fioresi and F. Gavarini, "On the construction of Chevalley supergroups", pp. 101–123 in *Supersymmetry in mathematics and physics* (Los Angeles, 2010), edited by S. Ferrara et al., Lecture Notes in Math. **2027**, Springer, 2011. MR Zbl

[Fioresi and Gavarini 2012] R. Fioresi and F. Gavarini, *Chevalley supergroups*, Mem. Amer. Math. Soc. **1014**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[Fioresi and Gavarini 2013] R. Fioresi and F. Gavarini, "Algebraic supergroups with Lie superalgebras of classical type", *J. Lie Theory* **23**:1 (2013), 143–158. MR Zbl

[Fioresi and Lledó 2015] R. Fioresi and M. A. Lledó, *The Minkowski and conformal superspaces*, World Sci., Hackensack, NJ, 2015. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977. MR Zbl

[Kwok 2013] S. D. Kwok, "Super Morita theory", submitted, 2013. arXiv

[Manin 1988] Y. I. Manin, *Gauge field theory and complex geometry*, Grundlehren der Math. Wissenschaften **289**, Springer, 1988. MR Zbl

[Manin 1991] Y. I. Manin, *Topics in noncommutative geometry*, Princeton Univ. Press, 1991. MR Zbl

[Sun 2009] Y. Sun, "$PGL_n$ and automorphisms of projective space", presentation notes, Harvard University, 2009, Available at https://tinyurl.com/yinotespdf.

[Witten 2012] E. Witten, "Notes on super Riemann surfaces and their moduli", preprint, 2012. arXiv

Received July 14, 2017.

RITA FIORESI
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI BOLOGNA
BOLOGNA
ITALY
rita.fioresi@unibo.it

STEPHEN D. KWOK
MATHEMATICS RESEARCH UNIT
UNIVERSITY OF LUXEMBOURG
LUXEMBOURG
stephen.kwok@uni.lu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.403

# THE GROMOV WIDTH OF COADJOINT ORBITS OF THE SYMPLECTIC GROUP

IVA HALACHEVA AND MILENA PABINIAK

**We prove that the Gromov width of a coadjoint orbit of the symplectic group through a regular point $\lambda$, lying on some rational line, is at least equal to:**

$$\min\{|\langle\alpha^\vee, \lambda\rangle| : \alpha^\vee \text{ a coroot}\}.$$

**Together with the results of Zoghi and Caviedes concerning the upper bounds, this establishes the actual Gromov width. This fits in the general conjecture that for any compact connected simple Lie group $G$, the Gromov width of its coadjoint orbit through $\lambda \in \mathrm{Lie}(G)^*$ is given by the above formula. The proof relies on tools coming from symplectic geometry, algebraic geometry and representation theory: we use a toric degeneration of a coadjoint orbit to a toric variety whose polytope is the string polytope arising from a string parametrization of elements of a crystal basis for a certain representation of the symplectic group.**

## 1. Introduction

The nonsqueezing theorem of Gromov motivated the question of finding the biggest ball that could be symplectically embedded into a given symplectic manifold $(M, \omega)$. Consider the ball of *capacity* $a$:

$$B_a^{2N} = \left\{(x_1, y_1, \dots, x_N, y_N) \in \mathbb{R}^{2N} \;\middle|\; \pi \sum_{i=1}^{N} (x_i^2 + y_i^2) < a\right\} \subset \mathbb{R}^{2N},$$

with the standard symplectic form $\omega_{\mathrm{std}} = \sum dx_j \wedge dy_j$. The *Gromov width* of a $2N$-dimensional symplectic manifold $(M, \omega)$ is the supremum of the set of $a$'s such that $B_a^{2N}$ can be symplectically embedded in $(M, \omega)$. It follows from Darboux's theorem that the Gromov width is positive unless $M$ is a point.

Coadjoint orbits form an important class of symplectic manifolds. Let $K$ be a compact Lie group. It acts on itself by conjugation

$$K \ni g : K \to K, \quad g(h) = ghg^{-1}.$$

*MSC2010:* 20G05, 53D99.

*Keywords:* Gromov width, coadjoint orbits, toric degenerations, Okounkov bodies, crystal bases, string polytopes.

Associating to $g \in K$ the derivative of the above map, taken at the identity, $dg_e \colon T_eK \to T_eK$, one obtains the adjoint action of $K$ on $\mathfrak{k} = \mathrm{Lie}(K) = T_eK$. This induces the action of $K$ on $\mathfrak{k}^* = \mathrm{Lie}(K)^*$, the dual of its Lie algebra, called the *coadjoint action*. Each orbit $\mathcal{O} \subset \mathrm{Lie}(K)^*$ of the coadjoint action is naturally equipped with the *Kostant–Kirillov–Souriau symplectic form*:

$$\omega_\xi(X^\#, Y^\#) = \langle \xi, [X, Y] \rangle, \quad \xi \in \mathcal{O} \subset \mathrm{Lie}(K)^*, \ X, Y \in \mathrm{Lie}(K),$$

where $X^\#, Y^\#$ are the vector fields on $\mathrm{Lie}(K)^*$ corresponding to $X, Y \in \mathrm{Lie}(K)$, induced by the coadjoint $K$ action. The coadjoint action of $K$ on $\mathcal{O}$ is Hamiltonian, and the momentum map is the inclusion $\mathcal{O} \hookrightarrow \mathrm{Lie}(K)^*$. Every coadjoint orbit intersects a chosen positive Weyl chamber in a single point. Therefore there is a bijection between the coadjoint orbits and points in the positive Weyl chamber. Points in the interior of the positive Weyl chamber are called *regular* points. The orbits corresponding to regular points are of maximal dimension. They are diffeomorphic to $K/T$, for $T$ a maximal torus of $K$, and are called *generic orbits*. For example, when $K = U(n, \mathbb{C})$, the group of (complex) unitary matrices, a coadjoint orbit can be identified with the set of Hermitian matrices with a fixed set of eigenvalues. The generic orbits are diffeomorphic to the manifold of full flags in $\mathbb{C}^n$.

In this note we concentrate on the (compact) symplectic group

$$K = \mathrm{Sp}(n) = U(n, \mathbb{H}).$$

The main result of this manuscript is the following theorem.

**Theorem 1.1.** *Let $M := \mathcal{O}_\lambda$ be the coadjoint orbit of $K = \mathrm{Sp}(n)$ through a regular point $\lambda$ lying on some rational line in $\mathfrak{k}^*$, equipped with the Kostant–Kirillov–Souriau symplectic form. The Gromov width of $M$ is at least the minimum,*

$$\min\{|\langle \alpha^\vee, \lambda \rangle| : \alpha^\vee \text{ a coroot}\}.$$

If $\lambda = \lambda_1\omega_1 + \cdots + \lambda_n\omega_n$ where $\omega_1, \ldots, \omega_n$ are the fundamental weights, and $\lambda_j > 0$, then the above minimum is equal to, as we explain in Section 3, $\min\{\lambda_1, \ldots, \lambda_n\}$.

This particular lower bound is important because it coincides with the known upper bound. Zoghi [2010] proved that for a compact connected simple Lie group $K$, the above formula gives an upper bound for the Gromov width of a regular indecomposable coadjoint $K$-orbit through $\lambda$ ([Zoghi 2010, Proposition 3.16]). This result was later extended to nonregular orbits by Caviedes.

**Theorem 1.2** [Caviedes 2016, Theorem 8.3; Zoghi 2010, Proposition 3.16, regular orbits]. *Let $K$ be a compact connected simple Lie group. The Gromov width* *of a coadjoint orbit $\mathcal{O}_\lambda$ through $\lambda$, equipped with the Kostant–Kirillov–Souriau symplectic form, is at most*

$$\min\{|\langle \alpha^\vee, \lambda\rangle| : \alpha^\vee \text{ a coroot and } \langle \alpha^\vee, \lambda\rangle \neq 0\}.$$

Putting these results together we obtain the following corollary.

**Corollary 1.3.** *The Gromov width of a coadjoint orbit $\mathcal{O}_\lambda$ of* $\mathrm{Sp}(n)$ *through a regular point $\lambda$ lying on some rational line in $\mathfrak{t}^*$, is exactly*

$$\min\{|\langle \alpha^\vee, \lambda\rangle| : \alpha^\vee \text{ a coroot}\}.$$

What adds importance to our result is the fact that it is a special case of a general conjecture about the Gromov width of coadjoint orbits of compact Lie groups. Namely, it has been conjectured, and by now proved in many cases, that for any compact connected simple Lie group $K$, the Gromov width of its coadjoint orbit through $\lambda \in \mathrm{Lie}(K)^*$ is given by the formula from Theorem 1.2, i.e., it is the minimum over the positive results of pairings of $\lambda$ with coroots in the system. Karshon and Tolman [2005], and independently Lu [2006a], showed that the Gromov width of complex Grassmannians (which are degenerate coadjoint orbits of $U(n,\mathbb{C})$) is given by the above formula. Combining the results of Zoghi [2010] and Caviedes [2016] about upper bounds, and the results of [Pabiniak 2014] about lower bounds, one proves that the Gromov width of (not necessarily regular) coadjoint orbits of $U(n,\mathbb{C})$, $\mathrm{SO}(2n,\mathbb{R})$ and $\mathrm{SO}(2n+1,\mathbb{R})$ is also given by that formula. (The result for $\mathrm{SO}(2n+1,\mathbb{R})$ works only for orbits satisfying one mild technical condition; see [Pabiniak 2014] for more details).

To prove the main result we use tools from symplectic geometry, algebraic geometry and representation theory. Here is a brief outline. Using the work of [Harada and Kaveh 2015] one can construct a toric degeneration from the given coadjoint orbit $\mathcal{O}_\lambda$ to a toric variety. By "pulling back" the toric action from the toric variety one equips (an open dense subset of) $\mathcal{O}_\lambda$ with a toric action and can use its flow to construct embeddings of balls. If $\lambda$ is a dominant weight, there exists a particularly nice toric degeneration to a toric variety whose associated Newton–Okounkov body is the string polytope parametrizing a crystal basis for (the dual of) the irreducible representation with highest weight $\lambda$ ([Kaveh 2015a]). Such string polytopes have been studied by Littelmann [1998], and using his work we prove Theorem 1.1 for orbits $\mathcal{O}_\lambda$ with $\lambda$ a dominant weight. We then further extend this result to any regular $\lambda$ lying on a rational line in $\mathfrak{t}^*$.

The techniques used in this paper could be applied to other compact connected simple Lie groups to obtain a lower bound for the Gromov width by studying the structure of (more general) string polytopes. We do not pursue this idea here for the following reason. As the formula for the conjectured Gromov width is given in purely Lie-theoretic language, we believe that there should be a way of proving the (lower bound part of the) conjecture for all groups at once, by a proof described in purely Lie-theoretic language.

In Section 2 we introduce the tools that are used in Section 3 to prove the main result.

## 2. Tools

**2A.** ***Using a toric action to construct symplectic embeddings of balls.*** Toric geometry proves to be very helpful in finding lower bounds for the Gromov width. When a manifold $(M, \omega)$ is equipped with a Hamiltonian (so also effective) action of a torus $T$, one can use the flow of the vector field generated by this action to construct explicit embeddings of balls and therefore to obtain a lower bound for the Gromov width (a construction by Karshon and Tolman [2005]). If additionally the action is *toric*, that is $\dim T = \frac{1}{2}\dim M$, then more constructions are available (see, for example, [Traynor 1995; Schlenk 2005; Latschev et al. 2013]).

Recall that a Hamiltonian action of a torus $T$ on a symplectic manifold $(M, \omega)$ gives rise to a momentum map $\mu\colon M \to \mathrm{Lie}(T)^* =: \Lambda_{\mathbb{R}}$, from $M$ to the dual of the Lie algebra of $T$, which we denote by $\Lambda_{\mathbb{R}}$. This map is unique up to a translation in $\Lambda_{\mathbb{R}}$. A manifold $M$ equipped with a Hamiltonian $T$ action is often called a *Hamiltonian $T$-space*. When $M$ is compact, the image $\mu(M)$ is a Delzant polytope. Identifying $\Lambda_{\mathbb{R}}$ with $\mathbb{R}^{\dim T}$, we can view $\mu(M)$ as a polytope in $\mathbb{R}^{\dim T}$. Such an identification is not unique: it depends on the choice of a splitting of $T$ into a product of circles, and on the choice of an identification of the Lie algebra of $S^1$ with the real line $\mathbb{R}$. Changing the splitting of $T$ results in applying a $\mathrm{GL}(\dim T, \mathbb{Z})$ transformation to $\mathbb{R}^{\dim T}$, while changing the identification $\mathrm{Lie}(S^1) \cong \mathbb{R}$ results in rescaling. In this work, $S^1 = \mathbb{R}/\mathbb{Z}$, that is, the exponential map $\exp\colon \mathbb{R} = \mathrm{Lie}(S^1) \to S^1$ is given by $t \mapsto e^{2\pi i t}$. With this convention, the momentum map for the standard $S^1$-action on $\mathbb{C}$ by rotation with speed 1 is given (up to the addition of a constant) by $z \mapsto -\pi |z|^2$.

Consider the standard $T^n = (S^1)^n$ action on $\mathbb{C}^n$ where each circle rotates a corresponding copy of $\mathbb{C}$ with speed 1, with a momentum map

$$(z_1, \dots, z_n) \mapsto -\pi(|z_1|^2, \dots, |z_n|^2).$$

The image of the $n$-dimensional ball of capacity $a$ (radius $\sqrt{a/\pi}$) centered at the origin is $(-1)$ times the standard simplex of size $a$;

$$\Delta^n(a) := \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n_{\geq 0} \;\Big|\; \sum_{k=1}^{n} x_k < \pi \cdot (\sqrt{a/\pi})^2 = a \right\}.$$

Moreover, simplices embedded in the momentum map image signify the existence of embeddings of balls, as the following result explains.

**Proposition 2.1** [Lu 2006b, Proposition 1.3; Pabiniak 2014, Proposition 2.5]. *For any connected, proper (not necessarily compact) Hamiltonian $T^n$-space $M^{2n}$ of dimension $2n$ let*

$$\mathcal{W}(\Phi(M)) = \sup\big\{a > 0 \mid \text{ there exists } \Psi \in \mathrm{GL}(n, \mathbb{Z}),\ x \in \mathbb{R}^n, \\ \text{such that } \Psi(\Delta^n(a)) + x \subset \Phi(M)\big\},$$

*where $\Phi$ is some choice of momentum map. Then the Gromov width of $M$ is at least $\mathcal{W}(\Phi(M))$.*

**2B. *Coadjoint orbits as flag varieties.*** Coadjoint orbits of compact Lie groups can be viewed as flag manifolds of complex reductive groups. This interpretation allows us to later construct toric degenerations of coadjoint orbits (Section 2C).

Let $G$ be a connected reductive group over $\mathbb{C}$ and $B$ a Borel subgroup. Denote by $\Lambda$ the weight lattice of $G$ and by $\Lambda^+$ the dominant weights. Let $K$ be the compact form of $G$ and $T$ its maximal torus. A generic coadjoint orbit of $K$, $K/T$, is diffeomorphic to the flag manifold $G/B$. To equip the manifold $G/B$ with a symplectic structure, fix $\lambda \in \Lambda^+$ and let $V_\lambda$ denote the finite dimensional irreducible representation of $G$ with highest weight $\lambda$. There exists a very ample $G$-equivariant line bundle $\mathcal{L}_\lambda$ on $G/B$ whose space of sections $H^0(G/B, \mathcal{L}_\lambda)$ is isomorphic to $V_\lambda^*$ (Borel–Weil theorem). Embed $G/B$ into $\mathbb{P}(H^0(G/B, \mathcal{L}_\lambda)^*)$ (the Kodaira embedding), and use this embedding to pull back to $G/B$ the Fubini–Study symplectic structure. If $\omega_\lambda$ denotes the symplectic structure on $G/B$ obtained this way, then $(G/B, \omega_\lambda)$ is symplectomorphic to the coadjoint orbit $\mathcal{O}_\lambda$ with the Kostant–Kirillov–Souriau symplectic structure defined in the introduction.

In this manuscript, $G = \mathrm{Sp}(2n, \mathbb{C})$ and $K = \mathrm{Sp}(n) = U(n, \mathbb{H})$.

**2C. *Obtaining a toric action via a toric degeneration.*** Coadjoint orbits of a compact Lie group $K$ are naturally equipped with a Hamiltonian action of a maximal torus of $K$. This action, however, is rarely toric. We note that for $U(n, \mathbb{C})$, $\mathrm{SO}(n, \mathbb{R})$ a toric action can be constructed by Thimm's trick [Pabiniak 2014].

To obtain a toric action on a dense open subset of a coadjoint orbit of $\mathrm{Sp}(n)$, we apply a method developed by Harada and Kaveh [2015] using toric degenerations. We briefly sketch the main ingredients of their construction and for details direct the reader to [Harada and Kaveh 2015].

Consider the situation where $X$ is a $d$-dimensional projective algebraic variety, $\mathcal{L}$ an ample line bundle over $X$, $L = H^0(X, \mathcal{L})$, and let $\mathbb{C}(X)$ denote the field of rational functions on $X$. Given a valuation $\nu\colon \mathbb{C}(X)\setminus\{0\} \to \mathbb{Z}^d$ with one-dimensional leaves, one builds an additive semigroup

$$S = S(X, L, \nu, h) = \bigcup_{k>0} \big\{(k, \nu(f/h^k)) \mid f \in L^{\otimes k} \setminus \{0\}\big\}.$$

and a convex body

$$\Delta(S) = \overline{\operatorname{conv}\Big(\bigcup_{k>0}\big\{x/k \mid (k, x) \in S\big\}\Big)},$$

in $\mathbb{R}^d$, called an *Okounkov* (*or Newton–Okounkov*) *body*. Here $h$ is a fixed section of $\mathcal{L}$ and $L^{\otimes k}$ denotes the image of the $k$-fold product $L \otimes \cdots \otimes L$ in $H^0(X, \mathcal{L}^{\otimes k})$.

**Theorem 2.2** [Anderson 2013, Proposition 5.1 and Corollary 5.3; Harada and Kaveh 2015, Corollary 3.14]. *With the notation as above, assume in addition that $S$ is finitely generated. Then there exists a finitely generated, $\mathbb{N}$-graded, flat $\mathbb{C}[t]$-subalgebra $\mathcal{R} \subset \mathbb{C}(X)[t]$ inducing a flat family $\pi : \mathfrak{X} = \operatorname{Proj} \mathcal{R} \to \mathbb{C}$ such that*:

- *For any $z \neq 0$ the fiber $X_z = \pi^{-1}(z)$ is isomorphic to $X = \operatorname{Proj} \mathbb{C}(X)$, i.e., $\pi^{-1}(\mathbb{C} \setminus \{0\})$ is isomorphic to $X \times (\mathbb{C} \setminus \{0\})$.*
- *The special fiber $X_0 = \pi^{-1}(0)$ is isomorphic to $\operatorname{Proj} \mathbb{C}[S]$ and is equipped with an action of $(\mathbb{C}^*)^d$, where $d = \dim_{\mathbb{C}} X$. The normalization of the variety $\operatorname{Proj} \mathbb{C}[S]$ is the toric variety associated to the rational polytope $\Delta(S)$.*

Fix a Hermitian structure on the very ample line bundle $\mathcal{L}$ and equip $X$ with the symplectic structure $\omega$ induced from the Fubini–Study form on $\mathbb{P}(H^0(X, \mathcal{L})^*)$ via the Kodaira embedding.

**Theorem 2.3** [Harada and Kaveh 2015, Theorem 3.25]. *With the notation as above, assume in addition that $(X, \omega)$ is smooth and that the semigroup $S$ is finitely generated. Then*:

(1) *There exists an integrable system $\mu = (F_1, \dots, F_d) \colon X \to \mathbb{R}^d$ on $(X, \omega)$ in the sense of* [*Harada and Kaveh 2015, Definition* 1], *and the image of $\mu$ coincides with the Newton–Okounkov body $\Delta = \Delta(S)$.*

(2) *The integrable system generates a torus action on the inverse image under $\mu$ of the interior of the moment polytope $\Delta$.*[1]

In this manuscript we use valuations (with one-dimensional leaves) coming from the following examples.

**Example 2.4** [Harada and Kaveh 2015, Example 3.3]. Fix a linear ordering on $\mathbb{Z}^d$. Let $p$ be a smooth point in $X$, and let $u_1, \dots, u_d$ be a regular system of parameters in a neighborhood of $p$. Using this system, we can construct the lowest and the highest term valuations on $\mathbb{C}(X)$: the *lowest* (*resp. highest*) *term valuation* $v_{\text{low}}$ (resp. $v_{\text{high}}$) assigns to each $f(u_1, \dots, u_d) = \sum_{j=(j_1,\dots,j_d)} c_j u_1^{j_1} \cdots u_d^{j_d} \in \mathbb{C}(X)$ a $d$-tuple of integers which is the smallest (resp. biggest) among $j = (j_1, \dots, j_d)$ with $c_j \neq 0$, in the fixed order. To a rational function $f/h \in \mathbb{C}(X)$ this valuation

[1] In fact the action is defined on the set $U$ introduced in [Harada and Kaveh 2015, Definition 1], which contains, but might be strictly bigger than, the inverse image under $\mu$ of the interior of the moment polytope $\Delta$.

assigns $v_{\text{low}}(f) - v_{\text{low}}(h)$ (resp. $v_{\text{high}}(f) - v_{\text{high}}(h)$). Both of these valuations have one-dimensional leaves.

**Example 2.5.** What will be very relevant for this manuscript is a special case of the previous example. In the situation we consider here, $X$ is the flag variety $G/B$ of the symplectic group $G = \mathrm{Sp}(2n, \mathbb{C})$, with $B$ a fixed Borel subgroup of $G$. Choose a reduced decomposition $\underline{w_0} = (\alpha_{i_1}, \dots, \alpha_{i_N})$ of the longest word in the Weyl group $w_0 = s_{\alpha_{i_1}} \cdots s_{\alpha_{i_N}}$, where $s_{\alpha_i}$ is the reflection through the hyperplane orthogonal to the simple root $\alpha_i$:

$$s_{\alpha_i}(\beta) = \beta - 2\frac{\langle \beta, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle}\alpha_i.$$

It defines a sequence of (Schubert) subvarieties, i.e., a Parshin point

$$\{o\} = X_{w_N} \subset \cdots \subset X_{w_0} = X,$$

where $X_{w_k}$ is the Schubert variety corresponding to the Weyl group element $w_k = s_{\alpha_{i_{k+1}}} \cdots s_{\alpha_{i_N}}$, and $\{o\}$ is the unique $B$-fixed point in $X$. This sequence of varieties, in turn, gives rise to a regular system of parameters $u_1, \dots, u_d$, in which $X_{w_k} = \{u_1 = \cdots = u_k = 0\}$ (see Section 2.2 of [Kaveh 2015a]). Following Kaveh [2015a], we denote the associated highest term valuation (as in Example 2.4) on $\mathbb{C}(X) \setminus \{0\}$ by $v_{\underline{w_0}}$.

**2D.** ***Crystal bases and Newton–Okounkov bodies.*** We now return to analyzing the flag manifold. With $G$, $B$, $\lambda \in \Lambda^+$, $V_\lambda$, and $\mathcal{L}_\lambda$ as in Section 2B, recall that $G$ acts on the space of sections $H^0(G/B, \mathcal{L}_\lambda)$ giving a representation isomorphic to the dual representation $V_\lambda^*$. There exists a particular toric degeneration of the flag variety $G/B$ for which the associated Okounkov body is the string polytope parametrizing the elements of a crystal basis of the representation $V_\lambda^*$. Before analyzing this toric degeneration, we recall some basic facts about crystal bases.

Let $I$ denote the Dynkin diagram, and $\{\alpha_i\}_{i \in I}$, $\{\alpha_i^\vee\}_{i \in I}$ denote the simple roots and coroots respectively. We will look at the perfect basis for $V_\lambda^*$ coming from the specialization of Lusztig's canonical basis to $q = 1$ for the quantum enveloping algebra, which Kaveh [2015a] refers to as a crystal basis for $V_\lambda^*$. Note that this differs from Kashiwara's notion of crystal basis being the specialization at $q = 0$.

A *perfect basis* for a finite-dimensional representation $V$ of $G$ is a weight basis $B_V$ of the vector space $V$ together with a pair of operators, called Kashiwara operators, $\tilde{E}_\alpha, \tilde{F}_\alpha : B_V \to B_V \cup \{0\}$ for each simple root $\alpha$, and maps $\tilde{\epsilon}_\alpha, \tilde{\phi}_\alpha : V \setminus \{0\} \to \mathbb{Z}$ satisfying certain compatibility conditions. For further information, we refer the reader to [Kaveh 2015a, Section 3.1].

One can associate to a perfect basis $B_V$ a directed labeled graph, called the *crystal graph of the representation* $V$, whose vertices are the elements of $B_V \cup \{0\}$, and whose directed edges are labeled by the simple roots following the rule: There is an edge from $b$ to $b'$ labeled $\alpha$ if and only if $\tilde{E}_\alpha(b) = b'$ (equivalently, $\tilde{F}_\alpha(b') = b$). Also there is an edge from $b$ to 0 if $\tilde{E}_\alpha(b) = 0$, and from 0 to $b$ if $\tilde{F}_\alpha(b) = 0$. The graphs obtained in this way are isomorphic for each perfect basis of the given $G$-representation $V$ [Berenstein and Kazhdan 2007, Theorem 5.55].

A perfect basis $B_\lambda$ for the representation $V_\lambda$ with highest weight vector $v_\lambda$ can be obtained by considering the nonzero elements $gv_\lambda$ where $g$ is an element in the specialization to $q = 1$ of the Lusztig canonical basis of the quantum enveloping algebra of $G$. The dual basis $B^*_\lambda$ is then a perfect basis for the dual representation $V^*_\lambda$, and will be referred to as the *dual crystal basis* (see [Berenstein and Kazhdan 2007, Lemma 5.50]). The crystal $B_\lambda$ can be thought of as a combinatorial realization of $V_\lambda$ and reflects its internal structure. For more information about crystals see [Berenstein and Kazhdan 2007; Hong and Kang 2002; Henriques and Kamnitzer 2006].

There exists a nice parametrization of the elements of a (dual) crystal basis, called the *string parametrization*, by integral points in $\mathbb{Z}^N$ where $N$ is the length of the longest word in the Weyl group $W$. This parametrization depends on a choice of a reduced decomposition $\underline{w_0} = (\alpha_{i_1}, \dots, \alpha_{i_N})$ of the longest word $w_0 = s_{\alpha_{i_1}} \cdots s_{\alpha_{i_N}}$ in $W$:

$$\iota_{\underline{w_0}} : \coprod_{\lambda \in \Lambda^+} B^*_\lambda \to \Lambda^+ \times \mathbb{Z}^N_{\geq 0}, \qquad \iota_{\underline{w_0}}(B^*_\lambda) \subset \{\lambda\} \times \mathbb{Z}^N_{\geq 0}.$$

The image of $\iota_{\underline{w_0}}$ is the intersection of a rational convex polyhedral cone $\mathcal{C}_{\underline{w_0}}$ in $\Lambda_\mathbb{R} \times \mathbb{R}^N$ with the lattice $\Lambda \times \mathbb{Z}^N$. The projection of $\mathcal{C}_{\underline{w_0}}$ to $\mathbb{R}^N$ is a rational polyhedral cone in $\mathbb{R}^N$, called the *string cone*, and will be denoted by $C_{\underline{w_0}}$. Littelmann [1998] analyzed the image of string parametrizations (see also [Alexeev and Brion 2004, Theorem 1.1; Kaveh 2015a, Theorem 3.4]).

**Theorem 2.6** [Littelmann 1998, Proposition 1.5]. *For any dominant weight $\lambda$, the string parametrization is one-to-one. Moreover, $S_\lambda := \iota_{\underline{w_0}}(B^*_\lambda)$ is the set of integral points of a convex rational polytope $\Delta_{\underline{w_0}}(\lambda) \subset \mathbb{R}^N$ obtained as the intersection of the string cone, $C_{\underline{w_0}}$, and the $N$ half-spaces*

$$x_k \leq \langle \lambda, \alpha^\vee_{i_k} \rangle - \sum_{l=k+1}^{N} x_l \langle \alpha_{i_l}, \alpha^\vee_{i_k} \rangle, \quad k = 1, \dots, N.$$

(Note that in [Kaveh 2015a] the symbol $\mathcal{C}_{\underline{w_0}}$ denotes a slightly different object: the projection of $\mathcal{C}_{\underline{w_0}}$ from [Kaveh 2015a] to $\mathbb{R}^N$ is “our” $C_{\underline{w_0}}$ already intersected with the above $N$ half-spaces).

**Definition 2.7.** The polytope $\Delta_{\underline{w_0}}(\lambda) \subset \mathbb{R}^N$ is called the *string polytope* associated to $\lambda$.

For integral $\lambda$, the vertices of the polytope $\Delta_{\underline{w_0}}(\lambda)$ are rational, so

$$\operatorname{Cone}(\Delta_{\underline{w_0}}(\lambda)) = \{(t, tx); t \in \mathbb{R}_{\geq 0}, x \in \Delta_{\underline{w_0}}(\lambda)\} \subset \mathbb{R} \times \mathbb{R}^N,$$

the cone over $\Delta_{\underline{w_0}}(\lambda)$, is a strongly convex rational polyhedral cone.

Kaveh [2015a] observed the following relation between the string polytopes and Newton–Okounkov bodies associated to certain valuations that we have described in Section 2C.

**Theorem 2.8** [Kaveh 2015a, Theorem 1]. *The string parametrization for a dual crystal basis of $V_\lambda^* = H^0(G/B, \mathcal{L}_\lambda)$ is the restriction of the valuation $v_{\underline{w_0}}$ and the string polytope $\Delta_{\underline{w_0}}(\lambda)$ coincides with the Newton–Okounkov body of the algebra of sections of $\mathcal{L}_\lambda$ and the valuation $v_{\underline{w_0}}$.*

**Corollary 2.9.** *The semigroup associated to the valuation $v_{\underline{w_0}}$ is finitely generated.*

This is a consequence of Theorem 2.8, the observation above that the cone $\text{Cone}(\Delta_{\underline{w_0}}(\lambda)) \subset \mathbb{R} \times \mathbb{R}^N$ over $\Delta_{\underline{w_0}}(\lambda)$ is a strongly convex rational polyhedral cone, and Gordon's Lemma.

## 3. Proof of the main result

We aim to prove that the Gromov width of a generic coadjoint orbit $\mathcal{O}_\lambda$ of $\text{Sp}(n)$, passing through a point $\lambda$ in the interior of a chosen positive Weyl chamber and on a rational line, equipped with the Kostant–Kirillov–Souriau symplectic form, is

$$\min\{|\langle \lambda, \alpha^\vee| : \alpha^\vee \text{ a coroot}\}.$$

Recall that all generic coadjoint orbits $\mathcal{O}_\lambda$ are diffeomorphic to the flag manifold $G/B$, for $G = \text{Sp}(2n, \mathbb{C})$. For $i = 1, \dots, 2n$, let $\epsilon_i : \mathfrak{sp}(2n, \mathbb{C}) \to \mathbb{C}$ denote the linear functional assigning to a matrix its $i$-th diagonal entry, $\epsilon_i(x) = x_{ii}$. With this notation we can express the simple roots as:

$$\alpha_n = \epsilon_1 - \epsilon_2, \quad \alpha_{n-1} = \epsilon_2 - \epsilon_3, \quad \dots, \quad \alpha_2 = \epsilon_{n-1} - \epsilon_n, \quad \alpha_1 = 2\epsilon_n. \tag{3-1}$$

Note that the above enumeration is nonstandard. We follow Littelmann's enumeration, as we are going to quote some results from [Littelmann 1998]. All the roots are given by $\pm 2\epsilon_i$ and $\pm(\epsilon_i \pm \epsilon_j)$, $i \neq j$. The fundamental weights are $\omega_i = \epsilon_1 + \epsilon_2 + \dots + \epsilon_i$, $i = 1, 2, \dots, n$, and each $\lambda \in \Lambda_{\mathbb{R}}^+$ can be expressed as

$$\begin{aligned} \lambda &= \lambda_1\omega_1 + \lambda_2\omega_2 + \dots + \lambda_n\omega_n \qquad (\lambda_i \geq 0) \\ &= (\lambda_1 + \lambda_2 + \dots + \lambda_n)\epsilon_1 + (\lambda_2 + \dots + \lambda_n)\epsilon_2 + \dots + \lambda_n\epsilon_n. \end{aligned}$$

Then

$$\min\{|\langle \lambda, \alpha^\vee \rangle| : \alpha^\vee \text{ a coroot}\} = \min\{\lambda_1, \dots, \lambda_n\}.$$

We first analyze the situation when $\lambda$ *is integral*. Then $\lambda$ is a dominant weight and thus there exists a very ample line bundle $\mathcal{L}_\lambda$ on $G/B$ whose space of sections $H^0(G/B, \mathcal{L}_\lambda)$ is isomorphic to $V_\lambda^*$. The very ample line bundle $\mathcal{L}_\lambda$ induces the Kodaira embedding $j_\lambda : G/B \hookrightarrow \mathbb{P}(H^0(G/B, \mathcal{L}_\lambda)^*)$ and one can use $j_\lambda$ to pull back the Fubini–Study symplectic structure from the projective space to $G/B$. The thus obtained symplectic manifold $(G/B, \omega_\lambda = j_\lambda^*(\omega_{FS}))$ is symplectomorphic to $\mathcal{O}_\lambda$ with the standard Kostant–Kirillov–Souriau symplectic structure.

As explained in Section 2 (page 409), a choice of a reduced decomposition $\underline{w_0} = (\alpha_{i_1}, \dots, \alpha_{i_N})$ of the longest word $w_0 = s_{\alpha_{i_1}} \cdots s_{\alpha_{i_N}}$ in the Weyl group gives rise to a highest term valuation $v_{\underline{w_0}}$ with one-dimensional leaves, and to a semigroup $S$ with the associated Newton–Okounkov body $\Delta(S)$. This semigroup is finitely generated (Corollary 2.9). Theorems 2.2, 2.3 and 2.8 imply the following:

**Corollary 3.1.** *For integral $\lambda$, there exists a toric action on an open dense subset of $\mathcal{O}_\lambda$. Its moment map image is the interior of the string polytope $\Delta_{\underline{w_0}}(\lambda) \subset \mathbb{R}^{n^2}$.*

We prove the main theorem by exhibiting an embedding of (a $\mathrm{GL}(n^2, \mathbb{Z})$ image of) a simplex $\Delta^{n^2}(\min\{\lambda_1, \dots, \lambda_n\})$, of size equal to $\min\{\lambda_1, \dots, \lambda_n\}$, in the string polytope $\Delta_{\underline{w_0}}(\lambda)$. The polytope $\Delta_{\underline{w_0}}(\lambda)$ for the longest word decomposition

$$w_0 = s_1(s_2 s_1 s_2) \cdots (s_{n-1} \cdots s_1 \cdots s_{n-1})(s_n s_{n-1} \cdots s_1 \cdots s_{n-1} s_n),$$

(where $s_j = s_{\alpha_j}$, with the numbering of the simple roots from (3-1)), was described by Littelmann ([1998, Section 6, Theorem 6.1 and Corollary 6]; note the misprint in Corollary 6: $\lambda_{m-j+1}$ should be $\lambda_j$ as can be deduced from [Littelmann 1998, Proposition 1.5]).

**Proposition 3.2** [Littelmann 1998]. *Fix a dominant weight,*

$$\lambda = \lambda_1\omega_1 + \cdots + \lambda_n\omega_n = (\lambda_1 + \cdots + \lambda_n)\epsilon_1 + \cdots + \lambda_n\epsilon_n.$$

*Then the associated string polytope $\Delta_{\underline{w_0}}(\lambda)$ is the convex polytope in $\mathbb{R}^{n^2}$ given by $n^2$-tuples $\{a_{i,j} \mid 1 \le i \le n,\ i \le j \le 2n-i\}$ which satisfy*

$$a_{i,i} \ge a_{i,i+1} \ge \cdots \ge a_{i,2n-i} \ge 0, \quad \text{for all } i = 1, \dots n,$$

*and*

$$\begin{aligned} \bar{a}_{i,j} &\le \lambda_j + s(\bar{a}_{i,j-1}) - 2s(a_{i-1,j}) + s(a_{i-1,j+1}), \\ a_{i,j} &\le \lambda_j + s(\bar{a}_{i,j-1}) - 2s(\bar{a}_{i,j}) + s(a_{i,j+1}), \\ a_{i,n} &\le \lambda_n + s(\bar{a}_{i,n-1}) - s(a_{i-1,n}), \end{aligned}$$

*for all $1 \le i, j \le n$, where we use the notation*

$$\bar{a}_{i,j} := a_{i,2n-j} \quad \text{for } 1 \le j \le n,$$

*and*

$$s(\bar{a}_{i,j}) := \bar{a}_{i,j} + \sum_{k=1}^{i-1}(a_{k,j} + \bar{a}_{k,j}), \qquad s(a_{i,j}) := \sum_{k=1}^{i}(a_{k,j} + \bar{a}_{k,j}),$$

*for $j < n$ (so $s(a_{i,n}) = 2\sum_{k=1}^{i} a_{k,n}$).*

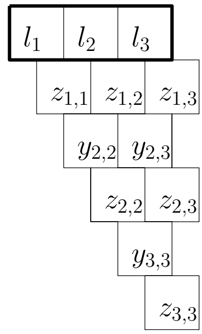



**Figure 1.** A graphical presentation of a Gelfand–Tsetlin pattern (for $n = 3$).

In the above formula we use the convention that $a_{i,j} = \bar{a}_{i,j} = 0$ if $j < i$. Note that if $i > 1$ then for $j < i$ the expression $s(\bar{a}_{i,j})$ is not 0 but equals $\sum_{k=1}^{i-1}(a_{k,j} + \bar{a}_{k,j})$.

Moreover, Littelmann [1998] defines a map from $\mathbb{R}^{n^2}$ to $\mathbb{R}^{n^2}$ which maps $\Delta_{\underline{w_0}}(\lambda)$ to the polytope $\mathrm{GT}(\lambda)$, obtained from a Gelfand–Tsetlin pattern,[2] which induces a bijection between the integral points of $\Delta_{\underline{w_0}}(\lambda)$ and $\mathrm{GT}(\lambda)$. We first recall from [Littelmann 1998] the definition of the polytope $\mathrm{GT}(\lambda)$. For simplicity of notation let

$$l_j := \lambda_j + \cdots + \lambda_n$$

so that $\lambda = l_1\epsilon_1 + \cdots + l_n\epsilon_n$. Let $\{y_{i,j}\}$, $2 \le i \le j \le n$, and $\{z_{i,j}\}$, $1 \le i \le j \le n$, denote coordinates in $\mathbb{R}^{n^2}$. A point

$$(y, z) := (z_{1,1}, \dots, z_{1,n}, y_{2,2}, \dots, y_{2,n}, z_{2,2}, \dots, z_{2,n}, \dots, y_{n,n}, z_{n,n})$$

in $\mathbb{R}^{n^2}_{\ge 0}$ is called a *Gelfand–Tsetlin pattern* for $\lambda = l_1\epsilon_1 + \cdots + l_n\epsilon_n$ if the entries satisfy the "betweenness " condition:

$$l_k \ge z_{1,k} \ge l_{k+1}, \quad z_{i-1,j-1} \ge y_{i,j} \ge z_{i-1,j}, \quad y_{i,j} \ge z_{i,j} \ge y_{i,j+1} \tag{3-2}$$

for $1 \le k \le n$, $1 \le i \le j \le n$, where $y_{1,j} = l_j$ for simplicity of notation. A convenient way to visualize these conditions is to organize the coordinates of $\mathbb{R}^{n^2}$ as in Figure 1 (for $n = 3$). The value of each coordinate must be between the values of its top right and top left neighbors. Littelmann's map from the string polytope $\Delta_{\underline{w_0}}(\lambda)$ to the Gelfand–Tsetlin polytope $\mathrm{GT}(\lambda)$ associates to each element $\underline{a} \in \mathbb{R}^{n^2}$ the pattern $P(\underline{a}) = (y_{i,j}, z_{i,j})$ of highest weight $\lambda = y_{1,1}\epsilon_1 + \cdots + y_{1,n}\epsilon_n$ defined by the equations

[2]Remark on notation: Performing Thimm's trick for the sequence of subgroups $\mathrm{Sp}(1) \subset \cdots \subset \mathrm{Sp}(n-1) \subset \mathrm{Sp}(n)$ produces a Hamiltonian action of a torus of dimension $\frac{1}{2}n(n-1)$ on $\mathcal{O}_\lambda$. The image of the momentum map for this torus (not toric) action is a polytope of dimension $\frac{1}{2}n(n-1)$ which is sometimes called a Gelfand–Tsetlin polytope. This polytope can be obtained from $\mathrm{GT}(\lambda)$ described here via a projection forgetting the $\{z_{i,j}\}$ coordinates.

in [Littelmann 1998] (note the misprint therein: $\alpha_{m-k+1}$ should be $\alpha_{m-j+1}$):

$$
\begin{aligned}
y_{i,1}\epsilon_1+\cdots+y_{i,n}\epsilon_n &= \lambda-\sum_{k=1}^{i-1}\Bigl(a_{k,n}\alpha_1+\sum_{j=k}^{n-1}(a_{k,j}+\bar{a}_{k,j})\alpha_{n-j+1}\Bigr)\\
z_{i,1}\epsilon_1+\cdots+z_{i,n}\epsilon_n &= \sum_{k=1}^{n} y_{i,k}\epsilon_k-\frac{a_{i,n}}{2}\alpha_1-\sum_{j=i}^{n-1}\bar{a}_{i,j}\alpha_{n-j+1},
\end{aligned}
\tag{3-3}
$$

where $\alpha_j$ are the simple roots as in (3-1):

$$\alpha_n=\epsilon_1-\epsilon_2,\quad \alpha_{n-1}=\epsilon_2-\epsilon_3,\quad \dots,\quad \alpha_2=\epsilon_{n-1}-\epsilon_n,\quad \alpha_1=2\epsilon_n.$$

In fact this map is a $\mathrm{GL}(n^2,\mathbb{Z})$-transformation followed by a translation, as we now show.

**Proposition 3.3.** *The map* (3-3) *which maps the polytope* $\Delta_{\underline{w_0}}(\lambda)$ *to the Gelfand–Tsetlin polytope* $\mathrm{GT}(\lambda)$ *is a* $\mathrm{GL}(n^2,\mathbb{Z})$*-transformation followed by a translation.*

We are grateful to the referee for suggesting we replace our original proof (by direct computation) with the following one.

*Proof.* Clearly (3-3) defines a composition of a linear map $\Phi\in\mathrm{GL}(n^2,\mathbb{R})$, defined by a matrix with integral entries (remember that $\alpha_1=2\epsilon_n$) and a translation. It suffices to show that $|\det\Phi|=1$ as this will imply that $\Phi^{-1}$ is also a matrix with integral entries, proving that $\Phi\in\mathrm{GL}(n^2,\mathbb{Z})$. The fact that (3-3) is a bijection between integral points of $\Delta_{\underline{w_0}}(k\lambda)=k\Delta_{\underline{w_0}}(\lambda)$ and integral points of $\mathrm{GT}(k\lambda)=k\,\mathrm{GT}(\lambda)$ for any $k\in\mathbb{N}$, together with the fact that the volume of any integral polytope $\Delta\in\mathbb{R}^{n^2}$, is the limit

$$\mathrm{vol}(\Delta)=\lim_{k\to\infty}\frac{\#(k\Delta\cap\mathbb{Z}^{n^2})}{k^{n^2}},$$

implies that $\mathrm{vol}(\Delta_{\underline{w_0}}(\lambda))=\mathrm{vol}\,\mathrm{GT}(\lambda)$. Therefore, we must have that $|\det\Phi|=1$. $\square$

**Example 3.4.** Let's take a closer look at the case $n=2$ and reprove the above proposition by direct computation. In this case, the simple roots are: $\alpha_1=2\epsilon_2$, $\alpha_2=\epsilon_1-\epsilon_2$. We fix a reduced word decomposition $w_0=s_1\,s_2\,s_1\,s_2$, and fix a weight

$$\lambda=\lambda_1 w_1+\lambda_2 w_2=(\lambda_1+\lambda_2)\epsilon_1+\lambda_2\epsilon_2.$$

The associated string polytope $\Delta=\Delta_{\underline{w_0}}(\lambda)$ is a subset of $\mathbb{R}^4$, for which we use coordinates $a_{22}, a_{11}, a_{12}, a_{13}$, and is defined by the inequalities

$$a_{22}\geq 0,\quad a_{11}\geq a_{12}\geq a_{13}\geq 0,$$

and

$$\begin{aligned} a_{13} = \bar{a}_{11} &\le \lambda_1, \\ a_{11} &\le \lambda_1 - 2s(\bar{a}_{11}) + s(a_{12}) = \lambda_1 - 2a_{13} + 2a_{12}, \\ a_{12} &\le \lambda_2 + s(\bar{a}_{11}) = \lambda_2 + a_{13}, \\ a_{22} &\le \lambda_2 + s(\bar{a}_{21}) - s(a_{12}) = \lambda_2 + a_{11} + a_{13} - 2a_{12}. \end{aligned}$$

We derive the second set of inequalities for the symplectic group (see also Corollary 6 of [Littelmann 1998]) from the description of the string polytope for a general $G$ given in [Littelmann 1998, definition on page 5, Proposition 1.5]. According to this description (using our fixed reduced word decomposition and numbering of simple roots):

$$\begin{aligned} a_{13} &\le \langle \lambda, \alpha_2^\vee \rangle = \langle \lambda, (\epsilon_1 - \epsilon_2)^\vee \rangle = (\lambda_1 + \lambda_2) - \lambda_2 = \lambda_1, \\ a_{12} &\le \langle \lambda - a_{13}\alpha_2, \alpha_1^\vee \rangle = \langle \lambda, 2\epsilon_2^\vee \rangle - a_{13}\langle \epsilon_1 - \epsilon_2, 2\epsilon_2^\vee \rangle = \lambda_2 + a_{13}, \\ a_{11} &\le \langle \lambda - a_{13}\alpha_2 - a_{12}\alpha_1, \alpha_2^\vee \rangle \\ &= \langle \lambda, (\epsilon_1 - \epsilon_2)^\vee \rangle - a_{13}\langle \epsilon_1 - \epsilon_2, (\epsilon_1 - \epsilon_2)^\vee \rangle - a_{12}\langle 2\epsilon_2, (\epsilon_1 - \epsilon_2)^\vee \rangle \\ &= \lambda_1 - 2a_{13} - a_{12}(-2), \\ a_{22} &\le \langle \lambda - a_{13}\alpha_2 - a_{12}\alpha_1 - a_{11}\alpha_2, \alpha_1^\vee \rangle \\ &= \lambda_2 + a_{13} - a_{12}\langle 2\epsilon_2, 2\epsilon_2^\vee \rangle - a_{11}\langle \epsilon_1 - \epsilon_2, 2\epsilon_2^\vee \rangle \\ &= \lambda_2 + a_{13} - 2a_{12} + a_{11}. \end{aligned}$$

We now analyze the map from the above string polytope to the Gelfand–Tsetlin polytope, given by equations (3-3). As

$$z_{11}\epsilon_1 + z_{12}\epsilon_2 = (\lambda_1 + \lambda_2)\epsilon_1 + \lambda_2\epsilon_2 - \frac{a_{12}}{2}(2\epsilon_2) - a_{13}(\epsilon_1 - \epsilon_2),$$

we get

$$\begin{aligned} z_{11} &= \lambda_1 + \lambda_2 - a_{13}, \\ z_{12} &= \lambda_2 - a_{12} + a_{13}. \end{aligned}$$

The value of $y_{22}$ is the coefficient of $\epsilon_2$ in $\lambda - a_{12}(2\epsilon_2) - (a_{11} + a_{13})(\epsilon_1 - \epsilon_2)$, and $z_{22}$ is the coefficient of $\epsilon_2$ in $y_{21}\epsilon_1 + y_{22}\epsilon_2 - \frac{1}{2}a_{22}(2\epsilon_2)$, thus

$$\begin{aligned} y_{22} &= \lambda_2 + a_{11} - 2a_{12} + a_{13}, \\ z_{22} &= y_{22} - a_{22}, \end{aligned}$$

i.e.,

$$\begin{bmatrix} z_{11} \\ z_{12} \\ y_{22} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & -2 & 1 \\ -1 & 1 & -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{22} \\ a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} + \begin{bmatrix} \lambda_1 + \lambda_2 \\ \lambda_2 \\ \lambda_2 \\ \lambda_2 \end{bmatrix}.$$

Therefore, the inequalities describing the string polytope translate to the following inequalities:

$$
\begin{aligned}
a_{22} \geq 0 &\iff y_{22} \geq z_{22},\\
a_{11} \geq a_{12} \iff y_{22} + 2a_{12} - a_{13} - \lambda_2 \geq a_{12} &\iff y_{22} \geq -a_{12} + a_{13} + \lambda_2 = z_{12},\\
a_{12} \geq a_{13} &\iff 0 \leq \lambda_2 - z_{12},\\
a_{13} \geq 0 &\iff \lambda_1 + \lambda_2 \geq z_{11},\\
a_{13} \leq \lambda_1 &\iff z_{11} \geq \lambda_2,\\
a_{12} - a_{13} \leq \lambda_2 \iff \lambda_2 - z_{12} \leq \lambda_2 &\iff 0 \leq z_{12},\\
a_{11} - 2a_{12} + 2a_{13} \leq \lambda_1 \iff y_{22} - z_{11} + \lambda_1 \leq \lambda_1 &\iff y_{22} \leq z_{11},\\
a_{22} - a_{11} + 2a_{12} - a_{13} \leq \lambda_2 \iff \lambda_2 - z_{22} \leq \lambda_2 &\iff 0 \leq z_{22}.
\end{aligned}
$$

The inequalities on the right are exactly the inequalities describing the Gelfand–Tsetlin polytope.

**Theorem 3.5.** *Let* $r = \min\{\lambda_1, \ldots, \lambda_n\}$ *and* $\Delta(r)$ *be an* $n^2$*-dimensional simplex of size* (*the lattice length of the edges*) $r$. *There exist* $\Psi \in \mathrm{GL}(n^2, \mathbb{Z})$ *and* $x \in \mathbb{R}^{n^2}$ *such that*

$$\Psi(\Delta(r)) + x \subset \mathrm{GT}(\lambda).$$

*Proof.* Recall from (3-2) the definition of $\mathrm{GT}(\lambda)$. Let $V_0 := V_0(\lambda)$ be a vertex of $\mathrm{GT}(\lambda)$ where all the coordinates $y_{i,j},\ z_{i,j}$ are equal to their upper bounds, i.e.,

$$z_{i,j} = y_{i,j} = z_{i-1,j-1} = y_{i-1,j-1} = \cdots = z_{1,j-i+1} = l_{j-i+1}.$$

We will analyze the edges starting from $V_0$. To obtain an edge starting from $V_0$, we pick one of the inequalities (3-2) defining $\mathrm{GT}(\lambda)$ which is an equality at $V_0$, and consider the set of points in $\mathrm{GT}(\lambda)$ satisfying all the same equations that $V_0$

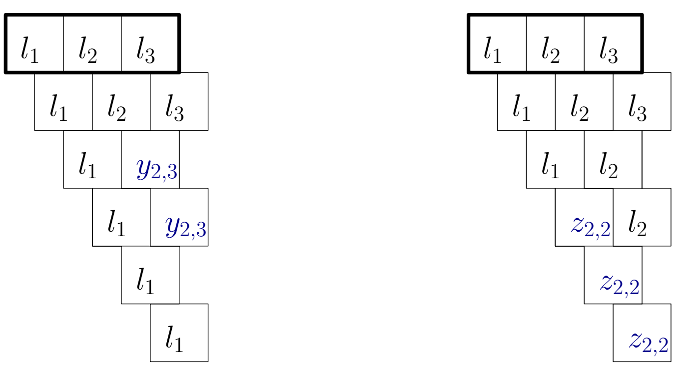



**Figure 2.** The edges $E_{2,3}$ and $F_{2,2}$, where $y_{2,3} \in [l_3, l_2]$ (left) and $z_{2,2} \in [l_2, l_1]$ (right).

satisfies, except possibly this chosen one. More precisely, each of the $\frac{1}{2}n(n-1)$ pairs $(i_0, j_0)$ with $2 \le i_0 \le j_0 \le n$ gives us an edge $E_{i_0,j_0}$ defined as the set of points $(y, z) \in \mathbb{R}^{n^2}$ satisfying

$$\begin{aligned} y_{i,j} &= z_{i,j} = l_{j-i+1} \text{ unless } j-i = j_0 - i_0 \text{ and } i \ge i_0, \\ y_{i_0,j_0} &= z_{i_0,j_0} = y_{i_0+1,j_0+1} = \cdots = z_{n-j_0+i_0,n} \in [l_{j_0-i_0+2}, l_{j_0-i_0+1}]. \end{aligned}$$

The lattice length of this edge is $l_{j_0-i_0+1} - l_{j_0-i_0+2} = \lambda_{j_0-i_0+1}$. An example of such an edge is presented in Figure 2, on the left.

Moreover, each of the $\frac{1}{2}n(n+1)$ pairs $(i_0, j_0)$ with $1 \le i_0 \le j_0 \le n$ gives us an edge $F_{i_0,j_0}$ defined as the set of points $(y, z) \in \mathbb{R}^{n^2}$ satisfying

$$\begin{aligned} y_{i,j} &= z_{i,j} = l_{j-i+1} \text{ unless } j-i = j_0 - i_0 \text{ and } i \ge i_0, \\ y_{i_0,j_0} &= l_{j_0-i_0+1}, \\ z_{i_0,j_0} &= y_{i_0+1,j_0+1} = z_{i_0+1,j_0+1} = \cdots = z_{n-j_0+i_0,n} \in [l_{j_0-i_0+2}, l_{j_0-i_0+1}]. \end{aligned}$$

The lattice length of this edge is also $l_{j_0-i_0+1} - l_{j_0-i_0+2} = \lambda_{j_0-i_0+1}$. An example of such an edge is presented in Figure 2, on the right.

The above collection gives $\frac{1}{2}n(n-1) + \frac{1}{2}n(n+1) = n^2$ edges. Observe that the directions of these $n^2$ edges from $V_0$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^{n^2} \subset \mathbb{R}^{n^2}$. Indeed, if we keep the ordering

$$z_{1,1},\ z_{1,2}, \ldots, z_{1,n},\ y_{2,2},\ y_{2,3}, \ldots, y_{2,n},\ z_{2,2}, \ldots, z_{2,n}, \ldots$$

of our usual coordinates on $\mathbb{R}^{n^2}$ and order the edge generators by

$$F_{1,1}, F_{1,2}, \ldots, F_{1,n}, E_{2,2}, E_{2,3}, \ldots, E_{2,n}, F_{2,2}, \ldots, F_{2,n}, \ldots,$$

then the matrix of edge generators expressed in our usual basis is an upper triangular matrix with $(-1)$'s on the diagonal. Therefore, there exist $\Psi \in \mathrm{GL}(n^2, \mathbb{Z})$ and $x \in \mathbb{R}^{n^2}$ such that

$$\Psi(\Delta(\min\{\lambda_j \mid j = 1, \ldots, n\})) + x \subset \mathrm{GT}(\lambda). \qquad \square$$

Combining the above claims, we prove our main result.

*Proof of Theorem 1.1.* Let

$$\lambda = \lambda_1\omega_1 + \cdots + \lambda_n\omega_n = (\lambda_1 + \cdots + \lambda_n)\epsilon_1 + \cdots + \lambda_n\epsilon_n$$

be a point in the interior of the chosen Weyl chamber $\Lambda^+_{\mathbb{R}}$ for the symplectic group $\mathrm{Sp}(n)$, which lies on some rational line. We want to show that the Gromov width of the coadjoint orbit $\mathcal{O}_\lambda$ through $\lambda$ is at least $\min\{\lambda_1, \ldots, \lambda_n\}$.

Recall that $\Lambda^+$ denotes the integral points of the positive Weyl chamber and let $\Lambda^+_{\mathbb{Q}}$ denote the rational ones. If $\lambda$ is integral then, by Corollary 3.1, an open dense subset of $\mathcal{O}_\lambda$ is equipped with a toric action. The momentum map image is the interior of a polytope equivalent under the action of $\mathrm{GL}(n^2, \mathbb{Z})$ and a translation to the Gelfand–Tsetlin polytope $\mathrm{GT}(\lambda)$ (see Propositions 3.2 and 3.3). Then Theorem 3.5 and Proposition 2.1 together with Theorem 1.2 prove that the Gromov width of $\mathcal{O}_\lambda$ is exactly $\min\{\lambda_1, \dots, \lambda_n\}$.

If $\lambda$ is not integral, let $a \in \mathbb{R}_+$ be such that $a\lambda$ is integral. Observe that the coadjoint orbits $\mathcal{O}_{a\lambda}$ and $\mathcal{O}_\lambda$ are diffeomorphic and differ only by a rescaling of their symplectic forms. Thus the Gromov width of $\mathcal{O}_{a\lambda}$, which is $\min\{a\lambda_1, \dots, a\lambda_n\}$, is $a$ times bigger than the Gromov width of $\mathcal{O}_\lambda$. This proves that the Gromov width of $\mathcal{O}_\lambda$ for $\lambda$ rational is exactly $\min\{\lambda_1, \dots, \lambda_n\}$. □

**3A.** ***Further comments.*** Note that the Gromov width of $\mathcal{O}_\lambda$ is lower semicontinuous as a function of $\lambda$, which one can prove by adjusting a "Moser type" argument from [Mandini and Pabiniak 2018]. However, to extend our result to orbits $\mathcal{O}_\lambda$ with arbitrary $\lambda$, what is in fact needed is upper semicontinuity. We are very grateful to the referee for this remark. It is not known in general if the Gromov width of $\mathcal{O}_\lambda$ is upper semicontinuous. It would be if, for example, all obstructions to embeddings of balls came from $J$-holomorphic curves. (The last condition is often called the "Biran Conjecture".) Note that an implication of the above conjecture of Biran is that the Gromov width of integral symplectic manifolds must be greater than or equal to 1. This statement was proved, under certain assumptions: using Seshadri constants by Lazarsfeld [2004a; 2004b] and by McDuff and Polterovich [1994], and also, using degenerations, by Kaveh [2015b].

## Acknowledgements

The authors are very grateful to Kiumars Kaveh for explaining his work to us. We also thank Yael Karshon, Joel Kamnitzer and Alexander Caviedes for useful discussions. We are very grateful to the anonymous referee for their corrections (a missing assumption that $\lambda$ is on a rational line) and pointing out misprints in the first version, as well as for their comments which improved the exposition of this paper.

The first author was supported by an NSERC Alexander Graham Bell CGS D and a Queen Elizabeth II graduate scholarship. The second author was supported by the Fundação para a Ciência e a Tecnologia (FCT), Portugal: fellowship SFRH/BPD/87791/2012 and projects PTDC/MAT/117762/2010, EXCL/MAT-GEO/0222/2012.

## References

[Alexeev and Brion 2004] V. Alexeev and M. Brion, "Toric degenerations of spherical varieties", *Selecta Math.* (*N.S.*) **10**:4 (2004), 453–478. MR Zbl

[Anderson 2013] D. Anderson, "Okounkov bodies and toric degenerations", *Math. Ann.* **356**:3 (2013), 1183–1202. MR Zbl

[Berenstein and Kazhdan 2007] A. Berenstein and D. Kazhdan, "Geometric and unipotent crystals, II: From unipotent bicrystals to crystal bases", pp. 13–88 in *Quantum groups*, edited by P. Etingof et al., Contemp. Math. **433**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Caviedes 2016] A. Caviedes Castro, "Upper bound for the Gromov width of coadjoint orbits of compact Lie groups", *J. Lie Theory* **26**:3 (2016), 821–860. MR Zbl

[Harada and Kaveh 2015] M. Harada and K. Kaveh, "Integrable systems, toric degenerations and Okounkov bodies", *Invent. Math.* **202**:3 (2015), 927–985. MR Zbl

[Henriques and Kamnitzer 2006] A. Henriques and J. Kamnitzer, "Crystals and coboundary categories", *Duke Math. J.* **132**:2 (2006), 191–216. MR Zbl

[Hong and Kang 2002] J. Hong and S.-J. Kang, *Introduction to quantum groups and crystal bases*, Graduate Studies in Mathematics **42**, Amer. Math. Soc., Providence, RI, 2002. MR Zbl

[Karshon and Tolman 2005] Y. Karshon and S. Tolman, "The Gromov width of complex Grassmannians", *Algebr. Geom. Topol.* **5** (2005), 911–922. MR Zbl

[Kaveh 2015a] K. Kaveh, "Crystal bases and Newton–Okounkov bodies", *Duke Math. J.* **164**:13 (2015), 2461–2506. MR Zbl

[Kaveh 2015b] K. Kaveh, "Toric degenerations and symplectic geometry of smooth projective varieties", preprint, 2015. arXiv

[Latschev et al. 2013] J. Latschev, D. McDuff, and F. Schlenk, "The Gromov width of 4-dimensional tori", *Geom. Topol.* **17**:5 (2013), 2813–2853. MR Zbl

[Lazarsfeld 2004a] R. Lazarsfeld, *Positivity in algebraic geometry, I*, Ergebnisse der Mathematik (3) **48**, Springer, Berlin, 2004. MR Zbl

[Lazarsfeld 2004b] R. Lazarsfeld, *Positivity in algebraic geometry, II*, Ergebnisse der Mathematik (3) **49**, Springer, Berlin, 2004. MR

[Littelmann 1998] P. Littelmann, "Cones, crystals, and patterns", *Transform. Groups* **3**:2 (1998), 145–179. MR Zbl

[Lu 2006a] G. Lu, "Gromov–Witten invariants and pseudo symplectic capacities", *Israel J. Math.* **156** (2006), 1–63. MR Zbl

[Lu 2006b] G. Lu, "Symplectic capacities of toric manifolds and related results", *Nagoya Math. J.* **181** (2006), 149–184. MR Zbl

[Mandini and Pabiniak 2018] A. Mandini and M. Pabiniak, "On the Gromov width of polygon spaces", *Transform. Groups* **23**:1 (2018), 149–183. MR

[McDuff and Polterovich 1994] D. McDuff and L. Polterovich, "Symplectic packings and algebraic geometry", *Invent. Math.* **115**:3 (1994), 405–434. MR Zbl

[Pabiniak 2014] M. Pabiniak, "Gromov width of non-regular coadjoint orbits of $U(n)$, $SO(2n)$ and $SO(2n+1)$", *Math. Res. Lett.* **21**:1 (2014), 187–205. MR Zbl

[Schlenk 2005] F. Schlenk, *Embedding problems in symplectic geometry*, De Gruyter Expositions in Mathematics **40**, Walter de Gruyter GmbH & Co., Berlin, 2005. MR Zbl

[Traynor 1995] L. Traynor, "Symplectic packing constructions", *J. Differential Geom.* **42**:2 (1995), 411–429. MR Zbl

[Zoghi 2010] M. Zoghi, *The Gromov width of coadjoint orbits of compact Lie groups*, Ph.D. thesis, University of Toronto, 2010, Available at https://search.proquest.com/docview/869989852. MR

Received June 25, 2016. Revised August 8, 2017.

IVA HALACHEVA
UNIVERSITY OF TORONTO
TORONTO, ON
CANADA
iva.halacheva@utoronto.ca

MILENA PABINIAK
MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
KÖLN
GERMANY
pabiniak@math.uni-koeln.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.421

# MINIMAL BRAID REPRESENTATIVES OF QUASIPOSITIVE LINKS

KYLE HAYDEN

**We show that every quasipositive link has a quasipositive minimal braid representative, partially resolving a question posed by Orevkov. These quasipositive minimal braids are used to show that the maximal self-linking number of a quasipositive link is bounded below by the negative of the minimal braid index, with equality if and only if the link is an unlink. This implies that the only amphichiral quasipositive links are the unlinks, answering a question of Rudolph's.**

## 1. Introduction

Quasipositive links in $S^3$ were introduced by Rudolph [1983] and defined in terms of special braid diagrams, the details of which we recall below. These links possess a variety of noteworthy features. Perhaps most strikingly, results from [Rudolph 1983; Boileau and Orevkov 2001] show that quasipositive links are precisely those links which arise as transverse intersections of the unit sphere $S^3 \subset \mathbb{C}^2$ with complex plane curves $\Sigma \subset \mathbb{C}^2$. The hierarchy of braid-positive, positive, strongly quasipositive, and quasipositive links interacts in compelling ways with conditions such as fiberedness [Etnyre and Van Horn-Morris 2011; Hedden 2010], sliceness [Rudolph 1993], homogeneity [Baader 2005], and symplectic or Lagrangian fillability [Boileau and Orevkov 2001; Hayden and Sabloff 2015]. Quasipositive links also have well-understood behavior with respect to invariants such as the four-ball genus, the maximal self-linking number, and the Ozsváth–Szabó concordance invariant $\tau$ [Hedden 2010]. For a different perspective, we can view quasipositive braids as a monoid in the mapping class group of a disk with marked points, where they lie inside the contact-geometrically important monoid of right-veering diffeomorphisms; see [Etnyre and Van Horn-Morris 2015] for more details.

The braid-theoretic description of quasipositivity is as follows: A braid is called *quasipositive* if it is the closure of a word

$$\prod_i \omega_i \sigma_{j_i} \omega_i^{-1},$$

*MSC2010:* primary 57M25; secondary 57R17.
*Keywords:* quasipositive links, braid index, self-linking number, amphichirality.

where $\omega_i$ is any word in the braid group and $\sigma_{j_i}$ is a positive standard generator. A link is then called *quasipositive* if it has a quasipositive braid representative. However, an arbitrary braid representative of a quasipositive link need not be a quasipositive braid. Along these lines, Orevkov [2000] posed the following question:

**Question 1.1** (Orevkov). Let $\mathcal{L}$ be a quasipositive link and $\beta$ a minimal braid index representative of $\mathcal{L}$. Is $\beta$ quasipositive?

Partial resolutions to this question have appeared in [Etnyre and Van Horn-Morris 2011; Feller and Krcatovich 2017]. The first of these showed that the answer to Question 1.1 is "yes" for fibered strongly quasipositive links. (In contrast, the answer to the analogue of Question 1.1 for positive braids is "no", as Stoimenow [2002] has provided examples of braid positive knots that have no positive minimal braid representatives. See also [Stoimenow 2006, §1].) The main purpose of this note is to provide another partial answer to Question 1.1.

**Theorem 1.2.** *Every quasipositive link has a quasipositive minimal braid index representative.*

This claim follows quickly from the proof of the generalized Jones conjecture in [LaFountain and Menasco 2014] — a substantial result in the theory of braid foliations. Our method of proof is similar to that of [Etnyre and Van Horn-Morris 2011; 2015].

A few simple consequences follow from Theorem 1.2. First, by considering the self-linking number of a quasipositive minimal braid index representative of a quasipositive link, we obtain a lower bound on the maximal self-linking number $\overline{\mathrm{sl}}$ in terms of the minimal braid index $b$:

**Theorem 1.3.** *If $\mathcal{L}$ is a quasipositive link, then*

$$\overline{\mathrm{sl}}(\mathcal{L}) \geq -b(\mathcal{L}),$$

*with equality if and only $\mathcal{L}$ is an unlink.*

The calculation underlying Theorem 1.3 also lets us resolve an earlier question of Rudolph's from [Morton 1988, Problem 9.2]:

**Question 1.4** (Rudolph). Are there any amphichiral quasipositive links other than the unlinks?

At the time this question was asked, it was already known that nontrivial strongly quasipositive knots were chiral; see [Rudolph 1999, Remark 4] for a discussion of precedent results. Additional evidence for a negative answer came in the form of strong constraints on invariants of amphichiral quasipositive links (including their being slice [Wu 2011]). We confirm that the answer to Rudolph's question is "no".

**Corollary 1.5.** *If a link $\mathcal{L}$ and its mirror $m(\mathcal{L})$ are both quasipositive, then $\mathcal{L}$ is an unlink. In particular, the unlinks are the only amphichiral quasipositive links.*

After recalling the necessary background in Section 2, we supply proofs for the above results in Section 3.

## 2. Background

The generalized Jones conjecture, first confirmed by Dynnikov and Prasolov [2013], relates the writhe $w$ and braid index $n$ of braids with a given link type.

**Theorem 2.1** [Dynnikov and Prasolov 2013, generalized Jones conjecture]. *Let $\beta$ and $\beta_0$ be closed braids with the same link type $\mathcal{L}$, where $n(\beta_0)$ is minimal for $\mathcal{L}$. Then there is an inequality*

$$|w(\beta) - w(\beta_0)| \leq n(\beta) - n(\beta_0).$$

Recall Bennequin's formula for the self-linking number of a braid $\beta$:

$$\mathrm{sl}(\beta) = w(\beta) - n(\beta).$$

It follows from the generalized Jones conjecture, Bennequin's formula, and the transverse Alexander theorem that a minimal braid index representative of $\mathcal{L}$ achieves the maximal self-linking number among all transverse representatives of $\mathcal{L}$, denoted $\overline{\mathrm{sl}}(\mathcal{L})$. For any braid $\beta$ representing a link type $\mathcal{L}$, we can plot the pair $(w(\beta), n(\beta))$ in a plane. The *cone* of $\beta$ is the collection of all pairs $(w, n)$ realized by braids which are stabilizations of $\beta$; see Figure 1 for an example. If $\beta_0$ is a minimal braid index representative of $\mathcal{L}$, we see that the right edge of its cone consists of all pairs $(w, n)$ corresponding to braids achieving the maximal self-linking number of $\mathcal{L}$.

The other tool central to the proof of Theorem 1.2 is due to Orevkov and concerns braid moves that preserve quasipositivity.

**Theorem 2.2** [Orevkov 2000]. *Suppose the braids $\beta$ and $\beta'$ are related by positive (de)stabilization. Then $\beta$ is quasipositive if and only if $\beta'$ is quasipositive.*

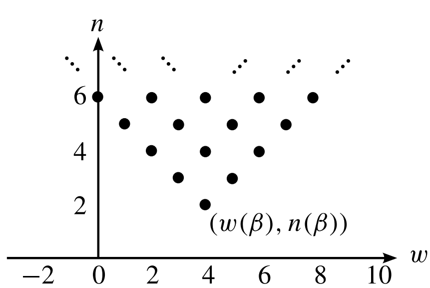



**Figure 1.** The cone of a braid $\beta$ with $(w(\beta), n(\beta)) = (4, 2)$.

**Remark 2.3.** In [Orevkov 2000], an $n$-stranded braid is viewed as an isotopy class of $n$-valued functions $f : [0, 1] \to \mathbb{C}$ where $f(0)$ and $f(1)$ equal $\{1, 2, \ldots, n\} \subset \mathbb{C}$. A braid is then quasipositive if one of its representatives can be expressed as a product of conjugates of the standard generators. For us, it is more convenient to study *closed* braids (up to isotopy through closed braids). Two closed braids are equivalent if and only if they can be expressed as closures of conjugate open braids. Since quasipositivity is a property of conjugacy classes of open braids, Theorem 2.2 holds equally well for closed braids.

## 3. Quasipositive minimal braids

We proceed to the proof of the of the main result, namely that every quasipositive link has a quasipositive minimal braid representative.

*Proof of Theorem 1.2.* Let $\mathcal{L}$ be a quasipositive link with a minimal braid index representative $\beta_0$ and a quasipositive braid representative $\beta_+$. Since the slice-Bennequin inequality is sharp for quasipositive links [Rudolph 1993; Hedden 2010], $\beta_+$ achieves the maximal self-linking number for $\mathcal{L}$. As noted above, it follows that $(w(\beta_+), n(\beta_+))$ lies along the right edge of the cone of $\beta_0$. The braids $\beta_0$ and $\beta_+$ have the same link type, so [LaFountain and Menasco 2014, Proposition 1.1] implies that there are braids $\beta_0'$ and $\beta_+'$ obtained from $\beta_0$ and $\beta_+$ by negative and positive stabilization, respectively, such that $\beta_0'$ and $\beta_+'$ cobound embedded annuli. Note that $\beta_0'$ and $\beta_+'$ lie along the left and right edges of the cone, respectively, as depicted on the left side of Figure 2. We also note that $\beta_+'$ is quasipositive since it is obtained from $\beta_+$ by positive stabilization.

Next, as in the proof of [LaFountain and Menasco 2014, Proposition 3.2], we can find braids $\beta_0''$ and $\beta_+''$ obtained from $\beta_0'$ and $\beta_+'$ by braid isotopy, destabilization, and exchange moves such that $w(\beta_+'') = w(\beta_0'')$ and $n(\beta_+'') = n(\beta_0'')$. We claim that $\beta_+''$ has minimal braid index (as does $\beta_0''$). Indeed, since $\beta_0'$ and $\beta_+'$ lie on the left and right edges of the cone of $\beta_0$, the destabilizations applied to them must be negative and positive, respectively. Given this and the fact that exchange moves preserve writhe and braid index, we see that $\beta_0''$ and $\beta_+''$ must also lie on the left and right edges of the cone of $\beta_0$, respectively. But since these braids occupy the same $(w, n)$-point, they must lie where the edges of the cone meet. As depicted on the right side of Figure 2, this implies that $\beta_0''$ and $\beta_+''$ have minimal braid index.

Finally, we show that the braid $\beta_+''$ is quasipositive. As noted above, any destabilizations of $\beta_+'$ must be positive, and these preserve quasipositivity by Theorem 2.2. An exchange move also preserves quasipositivity, since it can be expressed as a combination of one positive stabilization, one positive destabilization, and a number of conjugations; see [Birman and Wrinkle 2000, Figure 8]. □

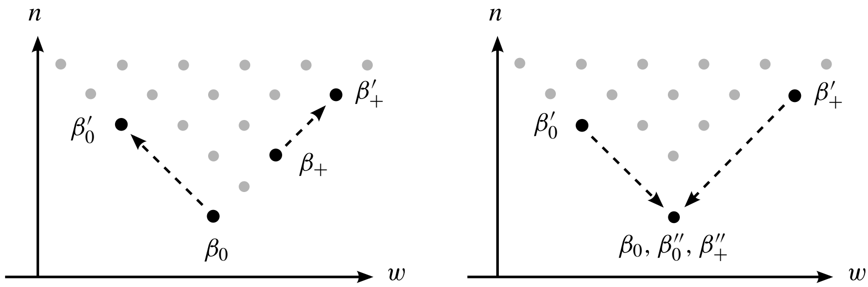



**Figure 2.** On the left, $\beta_0'$ and $\beta_+'$ are obtained from $\beta_0$ and $\beta_+$ by negative and positive stabilization, respectively. Then, on the right, $\beta_0''$ and $\beta_+''$ are obtained from $\beta_0'$ and $\beta_+'$ by negative and positive destabilization, respectively.

**Remark 3.1.** The question of whether or not *all* minimal braid index representatives of a quasipositive link are quasipositive remains open. The answer is seen to be "yes" for transversely simple link types: beginning with a quasipositive braid representative of a transversely simple link, the transverse Markov theorem implies that any minimal braid index representative can be related to it by positive (de)stabilization, which preserves quasipositivity. By the same reasoning, the answer to Question 1.1 is "yes" for any link type that has a unique transverse class achieving its maximal self-linking number (but is not necessarily transversely simple). This is the case for fibered strongly quasipositive links, as shown by Etnyre and Van Horn-Morris. But it fails to hold even for nonfibered strongly quasipositive links; as pointed out by Etnyre and Van Horn-Morris, there are infinite families of 3-braids found by Birman and Menasco [2006] which are (strongly) quasipositive and of minimal braid index but not transversely isotopic.

**Remark 3.2.** As pointed out by Eli Grigsby, the proof of Theorem 1.2 can be mirrored to show that any property of closed braids that is

(1) preserved under transverse isotopy, and
(2) satisfied by at least one braid representative of $\mathcal{L}$ with maximal self-linking number

is also satisfied by at least one minimal braid index representative of $\mathcal{L}$.

Now we obtain the lower bound in Theorem 1.3 by applying Bennequin's formula to a quasipositive minimal braid.

*Proof of Theorem 1.3.* Recall that a quasipositive braid always achieves the maximal self-linking number of its link type. Thus if $\beta$ is a quasipositive minimal braid index representative of $\mathcal{L}$, we have

$$\overline{\mathrm{sl}}(\mathcal{L}) = \mathrm{sl}(\beta) = w(\beta) - n(\beta) = w(\beta) - b(\mathcal{L}).$$

The desired inequality now follows from the fact that the writhe of a quasipositive braid is nonnegative, vanishing if and only if the braid is trivial. □

Finally, we prove the corollary that resolves Question 1.4.

*Proof of Corollary 1.5.* Observe that if $\beta$ is a minimal braid index representative of $\mathcal{L}$, then its mirror $m(\beta)$ is minimal for $m(\mathcal{L})$. Now suppose $\mathcal{L}$ and $m(\mathcal{L})$ are both quasipositive. The preceding proof implies that $w(\beta)$ and $w(m(\beta)) = -w(\beta)$ are both nonnegative, so $w(\beta)$ must be zero. Since we can choose the braid $\beta$ to be quasipositive, the vanishing of its writhe implies that the braid itself is trivial. □

## Acknowledgements

The author thanks John Baldwin, Peter Feller, and Eli Grigsby for several stimulating conversations and for introducing him to LaFountain and Menasco's proof of the generalized Jones conjecture.

## References

[Baader 2005] S. Baader, "Quasipositivity and homogeneity", *Math. Proc. Cambridge Philos. Soc.* **139**:2 (2005), 287–290. MR Zbl

[Birman and Menasco 2006] J. S. Birman and W. W. Menasco, "Stabilization in the braid groups, II: Transversal simplicity of knots", *Geom. Topol.* **10** (2006), 1425–1452. MR Zbl

[Birman and Wrinkle 2000] J. S. Birman and N. C. Wrinkle, "On transversally simple knots", *J. Differential Geom.* **55**:2 (2000), 325–354. MR Zbl

[Boileau and Orevkov 2001] M. Boileau and S. Orevkov, "Quasi-positivité d'une courbe analytique dans une boule pseudo-convexe", *C. R. Acad. Sci. Paris Sér. I Math.* **332**:9 (2001), 825–830. MR Zbl

[Dynnikov and Prasolov 2013] I. A. Dynnikov and M. V. Prasolov, "Bypasses for rectangular diagrams. A proof of the Jones conjecture and related questions", *Trans. Moscow Math. Soc.* (2013), 97–144. MR Zbl

[Etnyre and Van Horn-Morris 2011] J. B. Etnyre and J. Van Horn-Morris, "Fibered transverse knots and the Bennequin bound", *Int. Math. Res. Not.* **2011**:7 (2011), 1483–1509. MR Zbl

[Etnyre and Van Horn-Morris 2015] J. B. Etnyre and J. Van Horn-Morris, "Monoids in the mapping class group", pp. 319–365 in *Interactions between low-dimensional topology and mapping class groups*, edited by R. I. Bakyur et al., Geom. Topol. Monogr. **19**, 2015. MR Zbl

[Feller and Krcatovich 2017] P. Feller and D. Krcatovich, "On cobordisms between knots, braid index, and the Upsilon-invariant", *Math. Ann.* **369**:1-2 (2017), 301–329. MR Zbl

[Hayden and Sabloff 2015] K. Hayden and J. M. Sabloff, "Positive knots and Lagrangian fillability", *Proc. Amer. Math. Soc.* **143**:4 (2015), 1813–1821. MR Zbl

[Hedden 2010] M. Hedden, "Notions of positivity and the Ozsváth–Szabó concordance invariant", *J. Knot Theory Ramifications* **19**:5 (2010), 617–629. MR Zbl

[LaFountain and Menasco 2014] D. J. LaFountain and W. W. Menasco, "Embedded annuli and Jones' conjecture", *Algebr. Geom. Topol.* **14**:6 (2014), 3589–3601. MR Zbl

[Morton 1988] H. R. Morton, "Problems", pp. 557–574 in *Braids* (Santa Cruz, CA, 1986), edited by J. Birman and A. Libgober, Contemp. Math. **78**, Amer. Math. Soc., Providence, RI, 1988. MR Zbl

[Orevkov 2000] S. Y. Orevkov, “Markov moves for quasipositive braids”, *C. R. Acad. Sci. Paris Sér. I Math.* **331**:7 (2000), 557–562. MR Zbl

[Rudolph 1983] L. Rudolph, “Algebraic functions and closed braids”, *Topology* **22**:2 (1983), 191–202. MR Zbl

[Rudolph 1993] L. Rudolph, “Quasipositivity as an obstruction to sliceness”, *Bull. Amer. Math. Soc.* **29**:1 (1993), 51–59. MR Zbl

[Rudolph 1999] L. Rudolph, “Positive links are strongly quasipositive”, pp. 555–562 in *Proceedings of the Kirbyfest* (Berkeley, 1998), edited by J. Hass and M. Scharlemann, Geom. Topol. Monogr. **2**, 1999. MR Zbl

[Stoimenow 2002] A. Stoimenow, “On the crossing number of positive knots and braids and braid index criteria of Jones and Morton–Williams–Franks”, *Trans. Amer. Math. Soc.* **354**:10 (2002), 3927–3954. MR Zbl

[Stoimenow 2006] A. Stoimenow, “Properties of closed 3-braids and other link braid representations”, preprint, 2006. arXiv

[Wu 2011] H. Wu, “Generic deformations of the colored $\mathfrak{sl}(N)$-homology for links”, *Algebr. Geom. Topol.* **11**:4 (2011), 2037–2106. MR Zbl

Received April 21, 2017.

KYLE HAYDEN
DEPARTMENT OF MATHEMATICS
BOSTON COLLEGE
CHESTNUT HILL, MA
UNITED STATES
kyle.hayden@bc.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.429

# FOUR-DIMENSIONAL STATIC AND RELATED CRITICAL SPACES WITH HARMONIC CURVATURE

JONGSU KIM AND JINWOO SHIN

**We study any four-dimensional Riemannian manifold $(M, g)$ with harmonic curvature which admits a smooth nonzero solution $f$ to the equation**

$$\nabla df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right) + x\,\mathrm{Rc} + y(R)g,$$

**where Rc is the Ricci tensor of $g$, $x$ is a constant and $y(R)$ a function of the scalar curvature $R$. We show that a neighborhood of any point in some open dense subset of $M$ is locally isometric to one of the following five types: (i) $\mathbb{S}^2\left(\frac{R}{6}\right) \times \mathbb{S}^2\left(\frac{R}{3}\right)$ with $R > 0$, (ii) $\mathbb{H}^2\left(\frac{R}{6}\right) \times \mathbb{H}^2\left(\frac{R}{3}\right)$ with $R < 0$, where $\mathbb{S}^2(k)$ and $\mathbb{H}^2(k)$ are the two-dimensional Riemannian manifolds with constant sectional curvatures $k > 0$ and $k < 0$, respectively, (iii) the static spaces we describe in Example 3, (iv) conformally flat static spaces described by Kobayashi (1982), and (v) a Ricci flat metric.**

**We then get a number of corollaries, including the classification of the following four-dimensional spaces with harmonic curvature: static spaces, Miao–Tam critical metrics and $V$-static spaces.**

**For the proof we use some Codazzi-tensor properties of the Ricci tensor and analyze the equation displayed above depending on the various cases of multiplicity of the Ricci-eigenvalues.**

## 1. Introduction

In this article we consider an $n$-dimensional Riemannian manifold $(M, g)$ with constant scalar curvature $R$ which admits a smooth nonzero solution $f$ to the equation

$$\nabla df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right) + x \cdot \mathrm{Rc} + y(R)g, \tag{1-1}$$

Kim was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2017R1A2B4004460) .
*MSC2010:* 53C21, 53C25.
*Keywords:* static space, harmonic curvature, Codazzi tensor, critical point metric, Miao-Tam critical metric, $V$-static space.

where Rc is the Ricci curvature of $g$, $x$ is a constant and $y(R)$ a function of $R$. There are several well-known classes of spaces which admit such solutions. Below we describe them and briefly explain their geometric significance and recent developments.

A *static space* admits by definition a smooth nonzero solution $f$ to

$$\nabla df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right). \tag{1-2}$$

A Riemannian geometric interest of a static space comes from the fact that the scalar curvature functional $\mathfrak{S}$, defined on the space $\mathfrak{M}$ of smooth Riemannian metrics on a closed manifold, is locally surjective at $g \in \mathfrak{M}$ if there is no nonzero smooth function satisfying (1-2); see Chapter 4 of [Besse 1987].

This interpretation also holds in a local sense. Roughly speaking, if no nonzero smooth function on a compactly contained subdomain $\Omega$ of a smooth manifold satisfies (1-2) for a Riemannian metric $g$ on $\Omega$, then the scalar curvature functional defined on the space of Riemannian metrics on $\Omega$ is locally surjective at $g$ in a natural sense; see Theorem 1 of [Corvino 2000]. This local viewpoint has been developed to make remarkable progress in Riemannian and Lorentzian geometry [Chruściel et al. 2005; Corvino 2000; Corvino et al. 2013; Corvino and Schoen 2006; Qing and Yuan 2016].

Kobayashi [1982] studied a classification of conformally flat static spaces. In his study the list of *complete* ones is made. Moreover, all *local* ones are described for all varying parameter conditions and initial values of the static space equation. Indeed, they belong to the cases I–VI in Section 2 of [Kobayashi 1982] and the existence of solutions in each case is thoroughly discussed. Lafontaine [1983] independently proved a classification of closed conformally flat static spaces. Qing and Yuan [2013] classified complete Bach-flat static spaces which contain compact level hypersurfaces.

Next to static spaces we consider a Miao–Tam critical metric [2009; 2011], which is a compact Riemannian manifold $(M, g)$ that admits a smooth nonzero solution $f$, vanishing at the smooth boundary of $M$, to

$$\nabla df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right) - \frac{g}{n-1}. \tag{1-3}$$

In [Miao and Tam 2011], Miao–Tam critical metrics are classified when they are Einstein or conformally flat. In [Barros et al. 2015], Barros, Diógenes and Ribeiro proved that if $(M^4, g, f)$ is a Bach-flat simply connected, compact Miao–Tam critical metric with boundary isometric to a standard sphere $\mathbb{S}^3$, then $(M^4, g)$ is isometric to a geodesic ball in a simply connected space form $\mathbb{R}^4$, $\mathbb{H}^4$ or $\mathbb{S}^4$.

In [Corvino et al. 2013], Corvino, Eichmair and Miao defined a *V-static space* to be a Riemannian manifold $(M, g)$ which admits a nontrivial solution $(f, c)$, for a constant $c$, to the equation

$$\nabla\, df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right) - \frac{c}{n-1}g. \tag{1-4}$$

Note that $(M, g)$ is a $V$-static space if and only if it admits a solution $f$ to (1-2) or (1-3) on $M$, seen by scaling constants. Under a natural assumption, a $V$-static metric $g$ is a critical point of a geometric functional, as explained in Theorem 2.3 of [Corvino et al. 2013]. Like static spaces, *local* $V$-static spaces are still important; see, e.g., Theorems 1.1, 1.6 and 2.3 in [Corvino et al. 2013].

Lastly, one may consider Riemannian metrics $(M, g)$ which admit a nonconstant solution $f$ to

$$\nabla\, df = f\left(\mathrm{Rc} - \frac{R}{n-1}g\right) + \mathrm{Rc} - \frac{R}{n}g. \tag{1-5}$$

If $M$ is a closed manifold, then $g$ is a critical point of the total scalar curvature functional defined on the space of Riemannian metrics with unit volume and with constant scalar curvature on $M$. By an abuse of terminology we shall call a metric $g$ satisfying (1-5) *a critical point metric* even when $M$ is not closed. There are a number of works on this subject, including [Besse 1987, Section 4.F] and [Lafontaine 1983; Yun et al. 2014; Barros and Ribeiro 2014; Qing and Yuan 2013].

Finally we note that the existence of a nonzero or nonconstant solution to any of (1-2)–(1-5) guarantees the scalar curvature is constant. Indeed, it is shown for (1-2)–(1-4) in [Corvino 2000; Miao and Tam 2009; Corvino et al. 2013] and can be shown similarly for (1-5). But it does not hold true generally for (1-1).

In this paper we study spaces with harmonic curvature having a nonzero solution to (1-1). It is confined to four-dimensional spaces here, but our study may be extendible to higher dimensions. As motivated by Corvino's local deformation theory of scalar curvature, we study local (i.e., not necessarily complete) classification. We completely characterize nonconformally flat spaces, so that together with Kobayashi's work on conformally flat ones we get a full classification as follows.

**Theorem 1.1.** *Let* $(M, g)$ *be a four-dimensional* (*not necessarily complete*) *Riemannian manifold with harmonic curvature*, *satisfying* (1-1) *with nonconstant* $f$. *Then for each point* $p$ *in some open dense subset* $\widetilde{M}$ *of* $M$, *there exists a neighborhood* $V$ *of* $p$ *with one of the following properties*:

(i) $(V, g)$ *is isometric to a domain in* $\left(\mathbb{S}^2\left(\frac{R}{6}\right) \times \mathbb{S}^2\left(\frac{R}{3}\right), g_{R/6} + g_{R/3}\right)$ *with* $R > 0$, *where* $\mathbb{S}^2(k)$ *is the two-dimensional sphere with constant sectional curvature* $k > 0$ *and* $g_k$ *is the Riemannian metric of constant curvature* $k$, *and* $f = c_1 \cos\left(\sqrt{\frac{R}{6}}s\right) - x$

*for any constant* $c_1$, *where* $s$ *is the distance from a point on* $\mathbb{S}^2\big(\frac{R}{6}\big)$. *The constant* $R$ *equals the scalar curvature of* $g$. *It holds that* $\frac{1}{3}xR + y(R) = 0$.

(ii) $(V, g)$ *is isometric to a domain in* $\big(\mathbb{H}^2\big(\frac{R}{6}\big) \times \mathbb{H}^2(\frac{R}{3}), g_{R/6} + g_{R/3}\big)$ *with* $R < 0$, *where* $\mathbb{H}^2(k)$ *is the hyperbolic plane with constant sectional curvature* $k < 0$. *The metric* $g_{R/6}$ *can be written as* $g_{R/6} = ds^2 + p(s)^2\, dt^2$ *with* $p(s) = k_1 \sinh\big(\sqrt{-\frac{R}{6}}s\big) + k_2 \cosh\big(\sqrt{-\frac{R}{6}}s\big)$ *for constants* $k_1$, $k_2$, *and then* $f = c_2 p'(s) - x$ *for any constant* $c_2$. *It holds that* $\frac{1}{3}xR + y(R) = 0$.

(iii) $(V, g)$ *is isometric to a domain in one of the static spaces in Example 3 of Section 2A2*, *which is the Riemannian product* $(\mathbb{R}^1 \times W^3, dt^2 + ds^2 + h(s)^2\tilde{g})$ *of* $(\mathbb{R}^1, dt^2)$ *and some three-dimensional conformally flat static space* $(W^3, ds^2 + h(s)^2\tilde{g})$ *with zero scalar curvature*, *and* $f = c \cdot h'(s) - x$ *for any constant* $c$. *It holds that* $R = 0$ *and* $y(0) = 0$.

(iv) $(V, g)$ *is conformally flat. It is one of the metrics whose existence is described in Section* 2 *of [Kobayashi 1982]*; $g = ds^2 + h(s)^2 g_k$, *where* $h$ *is a solution of*

$$h'' + \tfrac{1}{12}Rh = ah^{-3} \quad \text{for a constant } a. \tag{1-6}$$

*For the constant* $k$, *the function* $h$ *satisfies*

$$(h')^2 + ah^{-2} + \tfrac{1}{12}Rh^2 = k, \tag{1-7}$$

*and* $f$ *is a nonconstant solution to the following ordinary differential equation for* $f$:

$$h'f' - fh'' = x\big(h'' + \tfrac{1}{3}Rh\big) + y(R)h. \tag{1-8}$$

*Conversely*, *any* $(V, g, f)$ *from* (i)–(iv) *has harmonic curvature and satisfies* (1-1).

Theorem 1.1 only considers the case when $f$ is a nonconstant solution, but the other case of $f$ being a nonzero constant solution is easier, which is described in Section 2A1.

Theorem 1.1 yields a number of classification theorems on four-dimensional spaces with harmonic curvature as follows. Theorem 8.2 classifies complete spaces satisfying (1-1). Then Theorems 9.1, 10.2 and 11.1 state the classification of *local* static spaces, $V$-static spaces and critical point metrics, respectively. Theorems 9.2 and 11.2 classify complete static spaces and critical point metrics, respectively. Theorem 10.3 gives a characterization of some four-dimensional Miao–Tam critical metrics with harmonic curvature, which is comparable to the aforementioned Bach-flat result [Barros et al. 2015].

To prove Theorem 1.1 we look into the eigenvalues of the Ricci tensor, which is a Codazzi tensor under the harmonic curvature condition. This Codazzi tensor encodes some geometric information, as investigated by Derdziński [1980]. In [Kim 2017], one of us has analyzed it in the Ricci soliton setting. We shall work in the same framework of arguments: we show that all Ricci-eigenvalues $\lambda_i,\ i=1,2,3,4$, locally depend on the function $f$ only, and then analyze case I when the three $\lambda_2, \lambda_3, \lambda_4$ are pairwise distinct and case II when exactly two of them are equal.

Our contribution in this paper is first to show the dependence of all Ricci-eigenvalues on $f$ in the setting of (1-1) by modifying the original soliton proof. Then in analyzing cases I and II, we manage to prove the desired key arguments of Propositions 4.2, 6.3 and 6.4 using involved formulas, which turns out to be fairly different from the soliton proof. Finally in the last five sections we discuss local-to-global results ranging from static spaces to critical point metrics.

This paper is organized as follows. In Section 2, we discuss examples and some properties from (1-1) and harmonic curvature. In Section 3, we prove that all Ricci-eigenvalues locally depend on only one variable. We study in Section 4 the case when the three eigenvalues $\lambda_2, \lambda_3, \lambda_4$ are pairwise distinct. In Sections 5 and 6 we analyze the case when exactly two of the three are equal. In Section 7 we characterize the case when all the three are equal, and then prove the local classification theorem as Theorem 1.1. We discuss the classification of complete spaces in Section 8. In Sections 9, 10 and 11 we treat static spaces, Miao–Tam critical and $V$-static spaces and critical point metrics respectively.

## 2. Examples and properties from (1-1) and harmonic curvature

We are going to describe some examples of spaces which satisfy (1-1) in Section 2A and state basic properties of spaces with harmonic curvature satisfying (1-1) in Section 2B.

### 2A. *Examples of spaces satisfying* (1-1).

**2A1.** *Spaces with a nonzero constant solution to* (1-1). When $(M, g)$ has a constant solution $f=-x$ to (1-1), then $y(R)+xR/(n-1)=0$. Conversely, any metric with its scalar curvature satisfying $y(R)+xR/(n-1)=0$ admits the constant solution $f=-x$ to (1-1) because

$$\nabla df = f\left(\mathrm{Rc}-\frac{R}{n-1}g\right)+x\mathrm{Rc}+y(R)g=(f+x)\left(\mathrm{Rc}-\frac{R}{n-1}g\right).$$

This proves the following lemma.

**Lemma 2.1.** *An $n$-dimensional Riemannian manifold $(M, g)$ of constant scalar curvature $R$ admits the constant solution $f=-x$ if and only if it satisfies $y(R)+xR/(n-1)=0$.*

If $(M, g)$ has a constant solution $f=c_0$, which does not equal $-x$, then $g$ is an Einstein metric. Conversely, if $g$ is Einstein, i.e., $\mathrm{Rc}=(R/n)g$ with $R\neq 0$, then any constant $c_0$ satisfying $c_0R=(n-1)xR+y(R)n(n-1)$ is a solution to (1-1); but if $g$ is Ricci-flat, then $f=c_0$ is a solution exactly when $y(0)=0$.

**2A2.** *Some examples of spaces which satisfy* (1-1) *with nonconstant $f$.*

**Example 1** (Einstein spaces satisfying (1-1) with nonconstant $f$). Let $(M, g, f)$ be a four-dimensional space satisfying (1-1), where $g$ is an Einstein metric. We shall show that $g$ has constant sectional curvature. We may use the argument in Section 1 of [Cheeger and Colding 1996]. In fact, the relation (1.6) of that paper corresponds to the equation

$$\nabla\, df = \left[-\tfrac{1}{12}Rf + x\tfrac{1}{4}R + y(R)\right]g \tag{2-1}$$

in our Einstein case. One can readily see that their argument to get their (1.19) still works; in some neighborhood of any point in $M$ we can write $g = ds^2 + (f'(s))^2\tilde{g}$, where $s$ is a function such that $\nabla s = \nabla f/|\nabla f|$ and $\tilde{g}$ is considered as a Riemannian metric on a level surface of $f$.

As $g$ is Einstein, so is $\tilde{g}$ from Lemma 4 in [Derdziński 1980]. As $\tilde{g}$ is three-dimensional, it has constant sectional curvature, say $k$. Moreover, $f$ satisfies $f'' = -\frac{1}{12}Rf + \frac{1}{4}xR + y(R)$, by feeding $(\partial/\partial s, \partial/\partial s)$ to (2-1).

Since $g$ is Einstein, we can readily see that our warped product metric $g$ has constant sectional curvature. In particular, a four-dimensional complete positive Einstein space satisfying (1-1) with nonconstant $f$ is a round sphere; see [Obata 1962; Yano and Nagano 1959].

**Example 2.** Assume $\frac{1}{3}xR + y(R) = 0$. Then (1-1) reduces to

$$\nabla df = (f + x)\left(\mathrm{Rc} - \frac{R}{n-1}g\right).$$

This is the static space equation for $g$ and $F = f + x$. We recall one example from [Lafontaine 1983]. On the round sphere $\mathbb{S}^2(1)$ of sectional curvature 1, we consider the local coordinates $(s, t) \in (0, \pi) \times \mathbb{S}^1$ so that the round metric is written $ds^2 + \sin^2(s)\, dt^2$. Let $f(s) = c_1 \cos s - x$ for any constant $c_1$. Then the product metric of $\mathbb{S}^2(1) \times \mathbb{S}^2(2)$ with $f$ satisfies (1-1). This example is neither Einstein nor conformally flat.

**Example 3.** Here we shall describe some four-dimensional nonconformally flat static space $g_W + dt^2$. We first recall some spaces among Kobayashi's warped product static spaces [1982] on $I \times N(k)$ with the metric $g = ds^2 + r(s)^2\bar{g}$, where $I$ is an interval and $(\bar{g}, N(k))$ is an $(n-1)$-dimensional Riemannian manifold of constant sectional curvature $k$. Moreover, $f = cr'$ for a nonzero constant $c$.

In order for $g$ to be a static space, the next equation needs to be satisfied; for a constant $\alpha$

$$r'' + \frac{R}{n(n-1)}r = \alpha r^{1-n}, \tag{2-2}$$

along with an integrability condition: for a constant $k$,

$$(r')^2 + \frac{2\alpha}{n-2}r^{2-n} + \frac{R}{n(n-1)}r^2 = k. \tag{2-3}$$

Existence of solutions depends on the values of $\alpha$, $R$, $k$. Here we consider only when $R = 0$. Then there are three cases:

(i) $R = 0$, $\alpha > 0$.

(ii) $R = 0$, $\alpha < 0$.

(iii) $R = 0$, $\alpha = 0$.

The above (i), (ii) and (iii) correspond respectively to the cases IV.1, III.1 and II in Section 2 of [Kobayashi 1982]. The solutions for these cases are discussed in Proposition 2.5, Example 5 and Proposition 2.4 in that paper. In particular, if $R = 0$, $\alpha > 0$ (then $k > 0$) and $n = 3$, we get the warped product metric on $\mathbb{R}^1 \times \mathbb{S}^2(1)$ which contains the spatial slice of a Schwarzschild space-time. Next, if $R = 0$, $\alpha < 0$, then there is an incomplete metric on $I \times N(k)$. If $R = 0$, $\alpha = 0$, then $g$ is readily seen to be a flat metric.

Let $(W^3, g_W, f)$ be one of the three-dimensional static spaces $(g, f)$ in the above paragraph. We now consider the four-dimensional product metric $g_W + dt^2$ on $W^3 \times \mathbb{R}^1$. One can check that $(W^3 \times \mathbb{R}^1, g_W + dt^2, f \circ \mathrm{pr}_1)$ is a static space, where $\mathrm{pr}_1$ is the projection of $W^3 \times \mathbb{R}^1$ onto the first factor. When $R = 0$ and $\alpha \neq 0$ for $g_W$, the metric $g_W + dt^2$ is not conformally flat and has three distinct Ricci-eigenvalues.

**2B.** ***Spaces with harmonic curvature.*** A Riemannian metric is said to have harmonic curvature [Besse 1987, Chapter 16] if the divergence of the curvature tensor is zero. The Ricci tensor Rc of a Riemannian metric, when evaluated on two vectors $(X, Y)$, shall be denoted by $R(X, Y)$ rather than $\mathrm{Rc}(X, Y)$, and its components in vector frames shall be written as $R_{ij}$.

By the differential Bianchi identity, the Ricci tensor of a Riemannian metric with harmonic curvature is a Codazzi tensor, written in local coordinates as $\nabla_k R_{ij} = \nabla_i R_{kj}$. A Riemannian metric with harmonic curvature has constant scalar curvature. We begin with a basic formula.

**Lemma 2.2.** *For a four-dimensional manifold* $(M^4, g, f)$ *with harmonic curvature satisfying* (1-1), *it holds that*

$$-R(X, Y, Z, \nabla f) = -R(X, Z)g(\nabla f, Y) + R(Y, Z)g(\nabla f, X) - \tfrac{1}{3}R\{g(\nabla f, X)g(Y, Z) - g(\nabla f, Y)g(X, Z)\}.$$

*Proof.* By the Ricci identity, $\nabla_i\nabla_j\nabla_k f - \nabla_j\nabla_i\nabla_k f = -\sum_l R_{ijkl}\nabla_l f$. The equation (1-1) gives

$$\begin{aligned}\sum_l - R_{ijkl}\nabla_l f &= \nabla_i\bigl\{f\bigl(R_{jk} - \tfrac{1}{3}Rg_{jk}\bigr) + xR_{jk} + y(R)g_{jk}\bigr\} \\ &\qquad - \nabla_j\bigl\{f\bigl(R_{ik} - \tfrac{1}{3}Rg_{ik}\bigr) + xR_{ik} + y(R)g_{ik}\bigr\} \\ &= \nabla_i f\bigl(R_{jk} - \tfrac{1}{3}Rg_{jk}\bigr) - \nabla_j f\bigl(R_{ik} - \tfrac{1}{3}Rg_{ik}\bigr),\end{aligned}$$

which yields the lemma. □

A Riemannian manifold with harmonic curvature is real analytic in harmonic coordinates [DeTurck and Goldschmidt 1989]. Equation (1-1) then implies that $f$ is real analytic in harmonic coordinates.

One may mimic arguments in [Cao and Chen 2013] and get the next lemma. We shall often denote the metric $g(X, Y)$ by $\langle X, Y\rangle$.

**Lemma 2.3.** *Let $(M^n, g, f)$ have harmonic curvature, satisfying* (1-1) *with non-constant $f$. Let $c$ be a regular value of $f$ and $\Sigma_c = \{x \mid f(x) = c\}$ be the level surface of $f$. Then the following hold*:

(i) $E_1 := \nabla f/|\nabla f|$ *is an eigenvector field of* Rc, *where* $\nabla f \neq 0$.

(ii) $|\nabla f|$ *is constant on any connected component of* $\Sigma_c$.

(iii) *There is a function $s$ locally defined with $s(x) = \int df/|\nabla f|$, so that $ds = df/|\nabla f|$ and $E_1 = \nabla s$.*

(iv) $R(E_1, E_1)$ *is constant on any connected component of* $\Sigma_c$.

(v) *Near a point in $\Sigma_c$, the metric $g$ can be written as*

$$g = ds^2 + \sum_{i,j>1} g_{ij}(s, x_2, \dots, x_n)\, dx_i \otimes dx_j,$$

*where $x_2, \dots, x_n$ is a local coordinate system on $\Sigma_c$.*

(vi) $\nabla_{E_1} E_1 = 0$.

*Proof.* In Lemma 2.2, put $Y = Z = \nabla f$ and $X \perp \nabla f$ to get

$$0 = -R(X, \nabla f, \nabla f, \nabla f) = -R(X, \nabla f)g(\nabla f, \nabla f).$$

So, $R(X, \nabla f) = 0$. Hence $E_1 = \nabla f/|\nabla f|$ is an eigenvector of Rc. By (1-1), $\frac{1}{2}\nabla_X|\nabla f|^2 = \langle\nabla_X\nabla f, \nabla f\rangle = fR(\nabla f, X) = 0$ for $X \perp \nabla f$. This proves (ii). Next

$$d\left(\frac{df}{|\nabla f|}\right) = -\frac{1}{2|\nabla f|^{\frac{3}{2}}} d|\nabla f|^2 \wedge df = 0$$

as $\nabla_X(|\nabla f|^2) = 0$ for $X \perp \nabla f$. So, (iii) is proved. As $\nabla f$ and the level surfaces of $f$ are perpendicular, one gets (v). One uses (v) to compute Christoffel symbols and gets (vi).

Now we shall prove (iv). Locally, $f$ is a function of the local variable $s$ only. We can write

$$E_1(f) = df(E_1) = \frac{df}{ds}\, ds(E_1) = \frac{df}{ds} g(\nabla s, \nabla s) = \frac{df}{ds},$$

which again depends on $s$ only. Similarly we get $E_1E_1(f) = d^2f/ds^2$. By (1-1), we have

$$\begin{aligned} E_1E_1f &= E_1E_1f - (\nabla_{E_1}E_1)f \\ &= \nabla\, df(E_1, E_1) = (f + x)R(E_1, E_1) - \frac{1}{n-1}Rf + y(R). \end{aligned}$$

Since $f+x$ is not zero on an open subset,

$$R(E_1, E_1) = \frac{1}{(f+x)}\left\{E_1E_1f + \frac{1}{n-1}Rf - y(R)\right\}$$

depends on $s$ only. So $R(E_1, E_1)$ is constant on any connected component of $\Sigma_c$. This proves (iv). □

As $(M, g)$ has harmonic curvature, the Ricci tensor Rc is a Codazzi tensor. Following [Derdziński 1980], for $x \in M$, let $E_{\mathrm{Rc}}(x)$ be the number of distinct eigenvalues of $\mathrm{Rc}_x$, and set $M_{\mathrm{Rc}} = \{x \in M \mid E_{\mathrm{Rc}} \text{ is constant in a neighborhood of } x\}$. The open subset $M_{\mathrm{Rc}}$ is dense in $M$. To see this, one may argue as follows. For each point $x \in M$, consider any open ball $B$ centered at $x$. As the range of the map $E_{\mathrm{Rc}}$ is finite, there is a point $q \in B$ where $E_{\mathrm{Rc}}(q)$ equals the maximum of $E_{\mathrm{Rc}}$ on $B$. By definition $E_{\mathrm{Rc}} \geq E_{\mathrm{Rc}}(q)$ near $q$. So, $E_{\mathrm{Rc}} \equiv E_{\mathrm{Rc}}(q)$ near $q$. Then $q \in M_{\mathrm{Rc}}$. This implies that $M_{\mathrm{Rc}}$ is dense.

Now we have:

**Lemma 2.4.** *For a Riemannian metric $g$ of dimension $n \geq 4$ with harmonic curvature, consider orthonormal vector fields $E_i$, $i = 1, \ldots, n$, such that $R(E_i, \cdot) = \lambda_i g(E_i, \cdot)$. Then the following hold in each connected component of $M_{\mathrm{Rc}}$:*

(i) $(\lambda_j - \lambda_k)\langle \nabla_{E_i} E_j, E_k\rangle + E_i\{R(E_j, E_k)\} = (\lambda_i - \lambda_k)\langle \nabla_{E_j} E_i, E_k\rangle + E_j\{R(E_k, E_i)\}$, *for any $i, j, k = 1, \ldots, n$.*

(ii) *If $k \neq i$ and $k \neq j$, then $(\lambda_j - \lambda_k)\langle \nabla_{E_i} E_j, E_k\rangle = (\lambda_i - \lambda_k)\langle \nabla_{E_j} E_i, E_k\rangle$.*

(iii) *Given distinct Ricci-eigenvalues $\lambda$, $\mu$ and local vector fields $v$, $u$ such that $R(v, \cdot) = \lambda g(v, \cdot)$ and $R(u, \cdot) = \mu g(u, \cdot)$ with $|u| = 1$, it holds that $v(\mu) = (\mu - \lambda)\langle \nabla_u u, v\rangle$.*

(iv) *For each eigenvalue $\lambda$, the $\lambda$-eigenspace distribution is integrable and its leaves are totally umbilic submanifolds of $M$.*

*Proof.* The statement (i) was proved in [Kim 2017]. Parts (ii) and (iii) follow from (i). Parts (iii) and (iv) are from Section 2 of [Derdziński 1980]. □

Given $(M^n, g, f)$ with harmonic curvature satisfying (1-1), $f$ is real analytic in harmonic coordinates, so $\{\nabla f \neq 0\}$ is open and dense in $M$. Lemma 2.3 gives that for any point $p$ in the open dense subset $M_r \cap \{\nabla f \neq 0\}$ of $M^n$, there is a neighborhood $U$ of $p$ where there exist orthonormal Ricci-eigenvector fields $E_i$, $i = 1, \ldots, n$, such that

(i) $E_1 = \nabla f / |\nabla f|$,

(ii) $E_i$ is tangent to smooth level hypersurfaces of $f$ for $i > 1$.

These local orthonormal Ricci-eigenvector fields $\{E_i\}$ shall be called an *adapted frame field* of $(M, g, f)$.

## 3. Constancy of $\lambda_i$ on level hypersurfaces of $f$

For an adapted frame field of $(M^n, g, f)$ with harmonic curvature satisfying (1-1), we set $\zeta_i := -\langle \nabla_{E_i} E_i, E_1\rangle = \langle E_i, \nabla_{E_i} E_1\rangle$ for $i > 1$. Then by (1-1),

$$\nabla_{E_i} E_1 = \nabla_{E_i}\left(\frac{\nabla f}{|\nabla f|}\right) = \frac{\nabla_{E_i}\nabla f}{|\nabla f|} = \frac{fR(E_i,\cdot) - fR/(n-1)g(E_i,\cdot) + xR(E_i,\cdot) + y(R)g(E_i,\cdot)}{|\nabla f|}.$$

So we may write

$$\nabla_{E_i} E_1 = \zeta_i E_i, \quad \text{where } \zeta_i = \frac{(f+x)R(E_i,E_i) - fR/(n-1) + y(R)}{|\nabla f|}. \tag{3-1}$$

Due to Lemma 2.3, in a neighborhood of a point $p \in M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$, $f$ may be considered as a function of $s$ only, and we write the derivative in $s$ by a prime: $f' = df/ds$.

**Lemma 3.1.** *Let $(M, g, f)$ be a four-dimensional space with harmonic curvature, satisfying (1-1) with nonconstant $f$. Then the Ricci-eigenvalue $\lambda_i$ associated to an adapted frame field $E_i$ is constant on any connected component of a regular level hypersurface $\Sigma_c$ of $f$, and so depend on the local variable $s$ only. Moreover, $\zeta_i$, $i = 2, 3, 4$, in (3-1) also depend on $s$ only. In particular, we have $E_i(\lambda_j) = E_i(\zeta_k) = 0$ for $i, k > 1$ and any $j$.*

*Proof.* We denote $\nabla_{E_i} f$ by $f_i$ and $\nabla_{E_j}\nabla_{E_i} f$ by $f_{ij}$. We have

$$\sum_{j=1}^{4} \tfrac{1}{2}\nabla_{E_j}\nabla_{E_j}(|\nabla f|^2) = \sum_{i,j} \tfrac{1}{2}\nabla_{E_j}\nabla_{E_j}(f_i f_i) = \sum_{i,j} \nabla_{E_j}(f_i f_{ij}).$$

We use $f_{ij} = f\big(R_{ij} - \frac{1}{3}Rg_{ij}\big) + xR_{ij} + y(R)g_{ij}$ from (1-1) to compute:

$$\begin{aligned}\sum_{i,j}\nabla_{E_j}(f_i f_{ij}) &= \sum_{i,j}\nabla_{E_j}\big\{ff_i\big(R_{ij}-\tfrac{1}{3}Rg_{ij}\big)+xf_iR_{ij}+y(R)f_ig_{ij}\big\}\\ &= \sum_{i,j} f_jf_i\big(R_{ij}-\tfrac{1}{3}Rg_{ij}\big)+ff_{ij}\big(R_{ij}-\tfrac{1}{3}Rg_{ij}\big)+xf_{ij}R_{ij}+y(R)f_{ij}g_{ij}\\ &= \big(R_{11}-\tfrac{1}{3}R\big)|\nabla f|^2+\sum_{i,j}(f+x)^2R_{ij}R_{ij}-\tfrac{2}{9}R^2f^2-\tfrac{2}{3}xR^2f\\ &\qquad\qquad +\big(2x-\tfrac{2}{3}f\big)y(R)R+4y(R)^2,\end{aligned}$$

where in obtaining the second equality we use the Bianchi identity $\nabla_k R_{jk} = \frac{1}{2}\nabla_k R$ and the fact that $R$ is constant.

Meanwhile,

$$\sum_{j=1}^{4} \nabla_{E_j}\nabla_{E_j}(|\nabla f|^2) = \sum_{j=1}^{4} E_j E_j(|\nabla f|^2) - (\nabla_{E_j} E_j)(|\nabla f|^2)$$
$$= (|\nabla f|^2)'' + \sum_{j=2}^{4} \zeta_j (|\nabla f|^2)'.$$

Since $R$ and $\lambda_1 = R_{11}$ depend on $s$ only by Lemma 2.3, the function $\sum_{j=2}^{4} \zeta_j$ depends only on $s$ by (3-1). We compare the above two expressions of

$$\sum_{j=1}^{4} \nabla_{E_j}\nabla_{E_j}(|\nabla f|^2)$$

to see that

$$\sum_{i,j} (f+x)^2 R_{ij} R_{ij}$$

depends only on $s$. As $f$ is nonconstant real analytic, $\sum_{i,j} R_{ij} R_{ij}$ depends only on $s$.

We compute

$$\begin{aligned}
\sum_{i,j,k} &\nabla_k(f_i f_{ij} R_{jk}) \\
&= \sum_{i,j,k} \nabla_k\bigl[f_i R_{jk}\bigl\{f(R_{ij} - \tfrac{1}{3} R g_{ij}) + x R_{ij} + y(R) g_{ij}\bigr\}\bigr] \\
&= \sum_{i,j,k} \nabla_k\bigl[f_i\bigl\{(f+x) R_{ij} R_{jk} - \bigl(\tfrac{1}{3} f R - y(R)\bigr) g_{ij} R_{jk}\bigr\}\bigr] \\
&= \sum_{i,j,k} f_{ik}\bigl\{(f+x) R_{ij} R_{jk} - \bigl(\tfrac{1}{3} f R - y(R)\bigr) g_{ij} R_{jk}\bigr\} \\
&\qquad + \sum_{i,j,k} f_i\bigl\{f_k R_{ij} R_{jk} + (f+x) R_{jk} \nabla_k R_{ij} - \tfrac{1}{3} f_k R g_{ij} R_{jk}\bigr\} \\
&= \sum_{i,j,k} \bigl\{(f+x) R_{ik} - \bigl(\tfrac{1}{3} f R - y(R)\bigr) g_{ik}\bigr\}\bigl\{(f+x) R_{ij} R_{jk} - \bigl(\tfrac{1}{3} f R - y(R)\bigr) g_{ij} R_{jk}\bigr\} \\
&\qquad + \sum_{i,j,k} f_i f_k R_{ij} R_{jk} + (f+x) f_i R_{jk} \nabla_k R_{ij} - \tfrac{1}{3} f_i f_k R g_{ij} R_{jk} \\
&= \sum_{i,j,k} (f+x)^2 R_{ik} R_{ij} R_{jk} + (f+x) f_i R_{jk} \nabla_k R_{ij} + L(s),
\end{aligned}$$

where $L(s)$ is a function of $s$ only, and the Bianchi identity $\nabla_k R_{jk} = \frac{1}{2} \nabla_k R = 0$ is used in obtaining the third equality.

Using $\nabla_k R_{ij} = \nabla_i R_{jk}$, we get

$$\sum_{i,j,k} \nabla_k(f_i f_{ij} R_{jk}) = \sum_{i,j,k} (f+x)^2 R_{ik} R_{ij} R_{jk} + \tfrac{1}{2}(f+x) f_i \nabla_i(R_{jk} R_{jk}) + L(s). \tag{3-2}$$

All terms except $(f+x)^2 R_{ij}R_{jk}R_{ik}$ in the right-hand side of (3-2) depend on $s$ only. From the constancy of $R$ and (3-1) we also get

$$\begin{aligned}
(3\text{-}3)\quad &\sum_{i,j,k} 2\nabla_k(f_i f_{ij} R_{jk}) \\
&= \sum_{i,j,k} \nabla_k(2 f_i f_{ij})\cdot R_{jk} = \sum_{i,j,k} \nabla_k\nabla_j(f_i f_i)\cdot R_{jk} \\
&= \sum_{i,j,k} E_k E_j(f_i f_i)\cdot R_{jk} - (\nabla_{E_k}E_j)(f_i f_i)\cdot R_{jk} \\
&= \sum_{j,i} E_j E_j(f_i f_i)\cdot R_{jj} - (\nabla_{E_j}E_j)(f_i f_i)\cdot R_{jj} \\
&= \sum_{i} E_1 E_1(f_i f_i)\cdot R_{11} + \sum_{j=2}^{4} \zeta_j E_1(|\nabla f|^2)\cdot R_{jj} \\
&= (|\nabla f|^2)''\cdot R_{11} + \sum_{j=2}^{4} \frac{(f+x)R_{jj}R_{jj} - \frac{1}{3}R f R_{jj} + y(R)R_{jj}}{|\nabla f|} E_1(|\nabla f|^2),
\end{aligned}$$

which depends only on $s$.

So, we compare (3-2) with (3-3) to see that $R_{ij}R_{jk}R_{ik}$ depends only on $s$. Now $\lambda_1$ and $\sum_{i=1}^{4}(\lambda_i)^k$, $k=1,2,3$, depend only on $s$. This implies that each $\lambda_i$, $i=1,2,3,4$, depends only on $s$. By (3-1), $\zeta_i$, $i=2,3,4$, depends on $s$ only. □

## 4. Four-dimensional space with distinct $\lambda_2, \lambda_3, \lambda_4$

Let $(M, g, f)$ be a four-dimensional Riemannian manifold with harmonic curvature satisfying (1-1). For an adapted frame field $\{E_j\}$ with its eigenvalue $\lambda_j$ in an open subset of $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$, we may only consider three cases depending on the distinctiveness of $\lambda_2, \lambda_3, \lambda_4$; the first case is when $\lambda_i$, $i=2,3,4$, are all equal (on an open subset), and the second is when exactly two of the three are equal. And the last is when the three $\lambda_i$, $i=2,3,4$, are mutually distinct. In this section we shall study the last case. Note that by (3-1) two eigenvalues $\lambda_i$ and $\lambda_j$ are distinct if and only if $\zeta_i$ and $\zeta_j$ are. We set $\Gamma_{ij}^k := \langle \nabla_{E_i}E_j, E_k\rangle$.

**Lemma 4.1.** *Let $(M, g, f)$ be a four-dimensional Riemannian manifold with harmonic curvature satisfying* (1-1) *with nonconstant $f$. Suppose that for an adapted frame field $E_j$, $j=1,2,3,4$, in an open subset $W$ of $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$, the eigenvalues $\lambda_2, \lambda_3, \lambda_4$ are distinct from each other. Then the following hold in $W$*:

$$\begin{aligned}
R_{1ij1} &= 0 \quad \textit{for distinct } i, j > 1, \\
R_{1ii1} &= -\zeta_i' - \zeta_i^2, \\
R_{1ii1} &= -R_{ii} + \tfrac{1}{3}R,
\end{aligned}$$

*where*

$$R_{11} = -\zeta_2' - \zeta_2^2 - \zeta_3' - \zeta_3^2 - \zeta_4' - \zeta_4^2,$$
$$R_{22} = -\zeta_2' - \zeta_2^2 - \zeta_2\zeta_3 - \zeta_2\zeta_4 - 2\Gamma_{34}^2\Gamma_{43}^2,$$
$$R_{33} = -\zeta_3' - \zeta_3^2 - \zeta_3\zeta_2 - \zeta_3\zeta_4 + 2\frac{\zeta_2 - \zeta_4}{\zeta_3 - \zeta_4}\Gamma_{34}^2\Gamma_{43}^2,$$
$$R_{44} = -\zeta_4' - \zeta_4^2 - \zeta_4\zeta_2 - \zeta_4\zeta_3 + 2\frac{\zeta_2 - \zeta_3}{\zeta_4 - \zeta_3}\Gamma_{34}^2\Gamma_{43}^2,$$

*Proof.* Now $\nabla_{E_1}E_1 = 0$ from Lemma 2.3(vi) and $\nabla_{E_i}E_1 = \zeta_i E_i$ from (3-1). Let $i, j > 1$ be distinct. From Lemma 2.4(iii) and Lemma 3.1, $\langle \nabla_{E_i}E_i, E_j\rangle = 0$. Since $\langle \nabla_{E_i}E_i, E_1\rangle = -\langle E_i, \nabla_{E_i}E_1\rangle = -\zeta_i$, we get $\nabla_{E_i}E_i = -\zeta_i E_1$. Now,

$$\langle \nabla_{E_i}E_j, E_i\rangle = -\langle \nabla_{E_i}E_i, E_j\rangle = 0,$$
$$\langle \nabla_{E_i}E_j, E_j\rangle = 0,$$
$$\langle \nabla_{E_i}E_j, E_1\rangle = -\langle \nabla_{E_i}E_1, E_j\rangle = 0.$$

So, $\nabla_{E_i}E_j = \Gamma_{ij}^k E_k$, where $k$ is the number such that $\{2, 3, 4\} = \{i, j, k\}$. Clearly $\Gamma_{ij}^k = -\Gamma_{ik}^j$. From Lemma 2.4(ii), $(\lambda_i - \lambda_j)\langle \nabla_{E_1}E_i, E_j\rangle = (\lambda_1 - \lambda_j)\langle \nabla_{E_i}E_1, E_j\rangle$. As $\langle \nabla_{E_i}E_1, E_j\rangle = 0$, we have $\langle \nabla_{E_1}E_i, E_j\rangle = 0$. This gives $\nabla_{E_1}E_i = 0$. Summarizing, we have the following for $i, j > 1,\ i \neq j$:

$$\nabla_{E_1}E_1 = 0, \quad \nabla_{E_i}E_1 = \zeta_i E_i, \quad \nabla_{E_i}E_i = -\zeta_i E_1, \quad \nabla_{E_1}E_i = 0,$$
$$\nabla_{E_i}E_j = \Gamma_{ij}^k E_k, \quad \text{where } k \text{ is the number such that } \{2, 3, 4\} = \{i, j, k\}.$$

One uses Lemma 3.1 in computing curvature components. For $i > 1$, we get $R_{1ii1} = -\zeta_i' - \zeta_i^2$, and for distinct $i, j, k > 1$, we get

$$R_{jiij} = -\zeta_j\zeta_i - \Gamma_{ji}^k\Gamma_{ik}^j - \Gamma_{ji}^k\Gamma_{ki}^j + \Gamma_{ij}^k\Gamma_{ki}^j,$$
$$R_{kijk} = E_k(\Gamma_{ij}^k),$$
$$R_{1ij1} = 0.$$

From Lemma 2.4, for distinct $i, j, k > 1$, we have

$$(\zeta_j - \zeta_k)\Gamma_{ij}^k = (\zeta_i - \zeta_k)\Gamma_{ji}^k, \tag{4-1}$$

which helps to express $R_{ii}$. Lemma 2.2 gives

$$-R(E_1, E_i, E_i, \nabla f) = \big(R_{ii} - \tfrac{1}{3}R\big)g(\nabla f, E_1)$$

for $i > 1$. From this we get

$$R_{1ii1} = -R_{ii} + \tfrac{1}{3}R. \tag{4-2}$$

□

From the proof of the above lemma, we may write

$$[E_2, E_3] = \alpha E_4, \quad [E_3, E_4] = \beta E_2, \quad [E_4, E_2] = \gamma E_3. \tag{4-3}$$

From the Jacobi identity $[[E_1, E_2], E_3] + [[E_2, E_3], E_1] + [[E_3, E_1], E_2] = 0$, we have

$$E_1(\alpha) = \alpha(\zeta_4 - \zeta_2 - \zeta_3). \tag{4-4}$$

Moreover, (4-1) gives

$$\beta = \frac{(\zeta_3 - \zeta_4)^2}{(\zeta_2 - \zeta_3)^2}\alpha, \quad \gamma = \frac{(\zeta_2 - \zeta_4)^2}{(\zeta_2 - \zeta_3)^2}\alpha. \tag{4-5}$$

We set $a := \zeta_2$, $b := \zeta_3$ and $c := \zeta_4$. Lemma 4.1 states two formulas for $R_{1ii1}$: $R_{1ii1} = -\zeta_i' - \zeta_i^2$ and $R_{1ii1} = -R_{ii} + \frac{1}{3}R$ for $i > 1$. So we have $R_{22} - R_{33} = a' + a^2 - b' - b^2$. The Ricci curvature formulas in Lemma 4.1 also give

$$R_{22} - R_{33} = -a' - a^2 + b' + b^2 - ac - 2\Gamma_{34}^2\Gamma_{43}^2 + bc - 2\frac{a-c}{b-c}\Gamma_{34}^2\Gamma_{43}^2.$$

Adding the last two equalities, we obtain

$$2(R_{22} - R_{33}) = (b-a)c - 2\Gamma_{34}^2\Gamma_{43}^2 - 2\frac{a-c}{b-c}\Gamma_{34}^2\Gamma_{43}^2.$$

From (1-1), $\zeta_i f' = f\big(R_{ii} - \frac{1}{3}R\big) + xR_{ii} + y(R)$ for $i > 1$. Then we get

$$(a-b)\frac{f'}{f} = \left(1 + \frac{x}{f}\right)(R_{22} - R_{33}) = \frac{1}{2}\left(1 + \frac{x}{f}\right)\left[(b-a)c - 2\left\{1 + \frac{a-c}{b-c}\right\}\Gamma_{34}^2\Gamma_{43}^2\right].$$

So,

$$-\frac{f'}{f} = \frac{1}{2}\left(1 + \frac{x}{f}\right)\left[c + 2\frac{a+b-2c}{(a-b)(b-c)}\Gamma_{34}^2\Gamma_{43}^2\right]. \tag{4-6}$$

Similarly,

$$(a-c)\frac{f'}{f} = \frac{1}{2}\left(1 + \frac{x}{f}\right)\left[(c-a)b - 2\left\{1 + \frac{a-b}{c-b}\right\}\Gamma_{34}^2\Gamma_{43}^2\right].$$

So,

$$-\frac{f'}{f} = \frac{1}{2}\left(1 + \frac{x}{f}\right)\left[b + 2\frac{a+c-2b}{(a-c)(c-b)}\Gamma_{34}^2\Gamma_{43}^2\right]. \tag{4-7}$$

From (4-6) and (4-7), we get

$$4\Gamma_{34}^2\Gamma_{43}^2 = \frac{(a-b)(a-c)(b-c)^2}{(a^2 + b^2 + c^2 - ab - bc - ac)}, \tag{4-8}$$

$$-\frac{f'}{f} = \frac{1}{2}\left(1 + \frac{x}{f}\right)\frac{a^2b + a^2c + ab^2 + ac^2 + b^2c + c^2b - 6abc}{2(a^2 + b^2 + c^2 - ab - bc - ac)}. \tag{4-9}$$

The formula (4-2) gives $R_{1212} - R_{1313} = R_{22} - R_{33}$, which reduces to

$$2(a'-b') = -2(a^2-b^2)+bc-ac+\frac{(a-b)(b-c)(c-a)(a+b-2c)}{2(a^2+b^2+c^2-ab-bc-ac)} = -2(a^2-b^2)+\frac{a-b}{2P}A, \tag{4-10}$$

where we set $P := a^2 + b^2 + c^2 - ab - bc - ac$, and $A := 6abc - a^2b - ab^2 - a^2c - ac^2 - b^2c - bc^2$. By symmetry, we get

$$\zeta_i' - \zeta_j' = -(\zeta_i^2 - \zeta_j^2) + \frac{\zeta_i - \zeta_j}{4P}A \quad \text{for } i, j \in \{2, 3, 4\}. \tag{4-11}$$

The formula (4-11) looks different from the corresponding one in the soliton case in [Kim 2017]: $\zeta_i' - \zeta_j' = -(\zeta_i^2 - \zeta_j^2)$. But surprisingly the next proposition still works in resolving (1-1); refer to Proposition 3.4 in [Kim 2017]. Here the formula (4-9) is crucial.

**Proposition 4.2.** *Let $(M, g, f)$ be a four-dimensional Riemannian manifold with harmonic curvature, satisfying* (1-1) *with nonconstant $f$. For any adapted frame field $E_j$, $j = 1, 2, 3, 4$, in an open dense subset $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$ of $M$, the three eigenfunctions $\lambda_2, \lambda_3, \lambda_4$ cannot be pairwise distinct, i.e., at least two of the three coincide.*

*Proof.* Suppose that $\lambda_2, \lambda_3, \lambda_4$ are pairwise distinct. We shall prove then that $f$ should be a constant, a contradiction to the hypothesis.

From (4-8) and (4-1),

$$(\alpha - \gamma + \beta)^2 = 4(\Gamma_{34}^2)^2 = 4\Gamma_{34}^2\Gamma_{43}^2\frac{a-b}{a-c} = \frac{(a-b)^2(b-c)^2}{(a^2+b^2+c^2-ab-bc-ac)}.$$

From (4-5),

$$(\alpha - \gamma + \beta)^2 = \alpha^2\left\{1 - \frac{(a-c)^2}{(a-b)^2} + \frac{(b-c)^2}{(a-b)^2}\right\}^2 = \frac{4\alpha^2(b-c)^2}{(a-b)^2}.$$

So, $\alpha^2 = (a-b)^4/(4P)$. Since $a, b, c$ are all functions of $s$ only, so is $\alpha$. We compute from (4-11)

$$\begin{aligned}(a-b)(a'-b') + (a-c)(a'-c') + (b-c)(b'-c') &= -(a-b)(a^2-b^2) - (a-c)(a^2-c^2) - (b-c)(b^2-c^2) \\ &\qquad + \frac{A}{4P}\{(a-b)^2 + (a-c)^2 + (b-c)^2\} \\ &= -2(a^3+b^3+c^3) + a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2 + \tfrac{1}{2}A \\ &= -2(a^3+b^3+c^3-3abc) - \tfrac{1}{2}A.\end{aligned} \tag{4-12}$$

Differentiating $\alpha^2 = (a-b)^4/(4P)$ in $s$ and using (4-11) and (4-12),

$$\begin{aligned} 2\alpha\alpha' &= \frac{(a-b)^3(a'-b')}{P} - \frac{(a-b)^4(2aa'+2bb'+2cc'-ab'-ba'-ac'-ca'-cb'-bc')}{4P^2} \\ &= \frac{-(a-b)^3(a^2-b^2)}{P} + \frac{(a-b)^4}{4P^2}A \\ &\qquad - \frac{(a-b)^4\{(a-b)(a'-b')+(a-c)(a'-c')+(b-c)(b'-c')\}}{4P^2} \\ &= -\frac{(a-b)^4(a+b)}{P} + \frac{(a-b)^4}{4P^2}A + \frac{(a-b)^4\{2(a^3+b^3+c^3-3abc)\}}{4P^2} + \frac{(a-b)^4\{\frac{1}{2}A\}}{4P^2} \\ &= -\frac{(a-b)^4}{P}\frac{(a+b-c)}{2} + \frac{3(a-b)^4}{8P^2}A. \end{aligned}$$

Meanwhile, from (4-4) and $\alpha^2 = (a-b)^4/(4P)$,

$$2\alpha\alpha' = 2\alpha^2(c-a-b) = -\frac{(a-b)^4}{2P}(a+b-c).$$

Equating these two expressions for $2\alpha\alpha'$, we get $A = 0$. From (4-9), $f' = 0$. □

## 5. Four-dimensional space with $\lambda_2 \neq \lambda_3 = \lambda_4$

In this section we study when exactly two of $\lambda_2, \lambda_3, \lambda_4$ are equal. We may well assume that $\lambda_2 \neq \lambda_3 = \lambda_4$. By (3-1) we then have $\zeta_2 \neq \zeta_3 = \zeta_4$. We use (3-1), Lemma 2.4 and Lemma 3.1 to compute $\nabla_{E_i} E_j$ and get the next lemma.

**Lemma 5.1.** *Let $(M, g, f)$ be a four-dimensional Riemannian manifold with harmonic curvature satisfying* (1-1) *with nonconstant $f$. Suppose that $\lambda_2 \neq \lambda_3 = \lambda_4$ for an adapted frame field $E_j$, $j = 1, 2, 3, 4$, on an open subset $U$ of $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$. Then we have*

$$[E_1, E_2] = -\zeta_2 E_2,$$
$$\langle \nabla_{E_i} E_j, E_2\rangle = 0 \quad \textit{and} \quad \langle \nabla_{E_i} E_j, E_1\rangle = -\delta_{ij}\zeta_3 \quad \textit{for } i, j \in \{3, 4\}.$$

*In particular, the distribution spanned by $E_1$ and $E_2$ is integrable. So is that spanned by $E_3$ and $E_4$.*

*Proof.* From Lemma 2.4 (ii) and (3-1),

$$(\lambda_2 - \lambda_i)\langle \nabla_{E_1} E_2, E_i\rangle = (\lambda_1 - \lambda_i)\langle \nabla_{E_2} E_1, E_i\rangle = (\lambda_1 - \lambda_i)\langle \zeta_2 E_2, E_i\rangle = 0$$

for $i = 3, 4$. This gives $\nabla_{E_1} E_2 = 0$, and so $[E_1, E_2] = -\zeta_2 E_2$.

From Lemma 2.4 (ii), $(\lambda_2 - \lambda_4)\langle \nabla_{E_3} E_2, E_4\rangle = (\lambda_3 - \lambda_4)\langle \nabla_{E_2} E_3, E_4\rangle = 0$. So, $\langle \nabla_{E_3} E_2, E_4\rangle = -\langle E_2, \nabla_{E_3} E_4\rangle = 0$. This and (3-1) yield $\nabla_{E_3} E_4 = \beta_3 E_3$ for some function $\beta_3$. Similarly, $\nabla_{E_4} E_3 = -\beta_4 E_4$ for some function $\beta_4$. Then $[E_3, E_4] =$

$\beta_3 E_3 + \beta_4 E_4$. For $i = 3, 4$, Lemma 2.4(iii) and Lemma 3.1 give $\langle \nabla_{E_i} E_i, E_2 \rangle = 0$ and (3-1) gives $\langle \nabla_{E_i} E_j, E_1 \rangle = -\delta_{ij} \zeta_3$ for $i, j \in \{3, 4\}$. □

We shall express the metric $g$ in a simple form as in the next lemma.

**Lemma 5.2.** *Under the same hypothesis as Lemma 5.1, for each point $p_0$ in $U$, there exists a neighborhood $V$ of $p_0$ in $U$ with coordinates $(s, t, x_3, x_4)$ such that $\nabla s = \nabla f / |\nabla f|$ and $g$ can be written on $V$ as*

$$g = ds^2 + p(s)^2\, dt^2 + h(s)^2 \tilde{g}, \tag{5-1}$$

*where $p := p(s)$ and $h := h(s)$ are smooth functions of $s$ and $\tilde{g}$ is (a pull-back of) a Riemannian metric of constant curvature, say $k$, on a two-dimensional domain with $x_3, x_4$ coordinates.*

*Proof.* Once Lemma 5.1 is in hand, this lemma may follow from the proof of Lemma 4.3 in [Kim 2017]. We produce a simplified proof for the sake of completeness.

We let $D^1$ be the two-dimensional distribution spanned by $E_1 = \nabla s$ and $E_2$, and let $D^2$ be the one spanned by $E_3$ and $E_4$. Then $D^1$ and $D^2$ are both integrable by Lemma 5.1. We may consider the coordinates $(x_1, x_2, x_3, x_4)$ from Lemma 4.2 of [Kim 2017], so that $D^1$ is tangent to the two-dimensional level sets $\{(x_1, x_2, x_3, x_4) \,|\, x_3, x_4 \text{ constants}\}$ and $D^2$ is tangent to the level sets $\{(x_1, x_2, x_3, x_4) \,|\, x_1, x_2 \text{ constants}\}$. We may write $g$ as

$$g = g_{11}\, dx_1^2 + g_{12}\, dx_1 \odot dx_2 + g_{22}\, dx_2^2 + g_{33}\, dx_3^2 + g_{34}\, dx_3 \odot dx_4 + g_{44}\, dx_4^2,$$

where $\odot$ is the symmetric tensor product and $g_{ij}$ are functions of $(x_1, x_2, x_3, x_4)$.

Defining a 1-form $\omega_2(\,\cdot\,) := g(E_2, \cdot\,)$, we can see that

$$ds^2 + \omega_2^2 = g_{11}\, dx_1^2 + g_{12}\, dx_1 \odot dx_2 + g_{22}\, dx_2^2.$$

Setting a function

$$p(s) := e^{\int_{s_0}^{s} \zeta_2(u)\, du}$$

for a constant $s_0$, we can check that $d(\omega_2 / p) = 0$ from Lemma 5.1. So, $\omega_2 / p = dt$ for some local function $t$ modulo a constant. The metric $g$ can be now written as

$$g = ds^2 + p(s)^2\, dt^2 + g_{33}\, dx_3^2 + g_{34}\, dx_3 \odot dx_4 + g_{44}\, dx_4^2. \tag{5-2}$$

Writing $\partial_i := \partial / \partial x_i$ in new coordinates $(x_1 := s, x_2 := t, x_3, x_4)$, from Lemma 5.1, we compute $0 = \langle \nabla_{\partial_i} \partial_j, \partial_2 \rangle = -\frac{1}{2} \partial_2 g_{ij}$ for $i, j \in \{3, 4\}$.

We consider the second fundamental form of a leaf for $D^2$ with respect to $E_1$: $H^{E_1}(u, v) = -\langle \nabla_u v, E_1 \rangle$. For $i, j \in \{3, 4\}$, from Lemma 5.1

$$\zeta_3 g_{ij} = H^{E_1}(\partial_i, \partial_j) = -\left\langle \nabla_{\partial_i} \partial_j, \frac{\partial}{\partial s} \right\rangle = \frac{1}{2} \frac{\partial}{\partial s} g_{ij}.$$

If $g_{34} > 0$ or $g_{34} < 0$ in a neighborhood of $p_0$, we can integrate the above and get

$$\ln |g_{ij}| = \int_{c_0}^{s} 2\zeta_3(u)\, du + C_{ij}(x_3, x_4)$$

for $i, j \in \{3, 4\}$ and a constant $c_0$. Setting

$$h(s) := e^{\int_{c_0}^{s} \zeta_3(u)\, du},$$

we have $|g_{ij}| = (h(s))^2 e^{C_{ij}(x_3,x_4)}$. Then we may write

$$G := g_{33}\, dx_3^2 + g_{34}\, dx_3 \odot dx_4 + g_{44}\, dx_4^2 = (h(s))^2 \tilde{g},$$

where $\tilde{g}$ is a Riemannian metric in a domain of the $(x_3, x_4)$-plane.

If $g_{34}(p_0) = 0$, by changing coordinates as $x_3 = z_3$ and $x_4 = z_3 + z_4$, we get

$$\begin{aligned} G &= g_{33}\, dz_3^2 + g_{34}\, dz_3 \odot (dz_3 + dz_4) + g_{44}(dz_3 + dz_4)^2 \\ &= a_{33}\, dz_3^2 + a_{34}\, dz_3 \odot dz_4 + a_{44}\, dz_4^2, \end{aligned}$$

where $a_{ij} = g(\partial/\partial z_i, \partial/\partial z_j)$. As $g_{44}(p_0) > 0$, we have $a_{34}(p_0) \neq 0$. So, $a_{34} \neq 0$ in a neighborhood of $p_0$. In $z_i$-coordinates we can still have $\partial_2 a_{ij} = 0$ and $\zeta_3 a_{ij} = \frac{1}{2}(\partial/\partial s)a_{ij}$. Arguing as the above paragraph, we can write $G$ in the form $G = (h(s))^2 \tilde{g}$, where

$$h(s) := e^{\int_{c_1}^{s} \zeta_3(u)\, du}$$

for a constant $c_1$ and $\tilde{g}$ is a Riemannian metric in a domain of the $(z_3, z_4)$-plane which is also a domain of the $(x_3, x_4)$-plane.

In any case $g$ can be written as $g = ds^2 + p(s)^2\, dt^2 + h(s)^2 \tilde{g}$, where $\tilde{g}$ can be viewed as a Riemannian metric in a domain of the $(x_3, x_4)$-plane.

The argument used in the proof of Lemma 4 in [Derdziński 1980] can prove that $\tilde{g}$ has constant curvature, say $k$. □

## 6. Analysis of the metric when $\lambda_2 \neq \lambda_3 = \lambda_4$

We continue to suppose that $\lambda_2 \neq \lambda_3 = \lambda_4$ for an adapted frame field $E_j$, $j = 1, 2, 3, 4$.

The metric $\tilde{g}$ in (5-1) can be written locally: $\tilde{g} = dr^2 + u(r)^2 d\theta^2$ on a domain in $\mathbb{R}^2$ with polar coordinates $(r, \theta)$, where $u''(r) = -ku$. We set an orthonormal basis

$$e_3 = \frac{\partial}{\partial r} \quad \text{and} \quad e_4 = \frac{1}{u(r)} \frac{\partial}{\partial \theta}.$$

**Lemma 6.1.** *For the local metric $g = ds^2 + p(s)^2\, dt^2 + h(s)^2 \tilde{g}$ with harmonic curvature satisfying* (1-1) *with nonconstant $f$, obtained in Lemma 5.2, if we set*

$$E_1 = \frac{\partial}{\partial s}, \quad E_2 = \frac{1}{p(s)} \frac{\partial}{\partial t}, \quad E_3 = \frac{1}{h(s)} e_3, \quad E_4 = \frac{1}{h(s)} e_4,$$

*where* $e_3$ *and* $e_4$ *are as in the above paragraph, then we have the following. Here* $R_{ij} = R(E_i, E_j)$ *and* $R_{ijkl} = R(E_i, E_j, E_k, E_l)$:

$$\nabla_{E_1} E_1 = 0,$$

$$\textit{for } i = 2, 3, 4, \quad \nabla_{E_1} E_i = 0, \quad \nabla_{E_i} E_1 = \zeta_i E_i, \quad \textit{where } \zeta_2 = \frac{p'}{p},\ \zeta_3 = \zeta_4 = \frac{h'}{h},$$

$$\nabla_{E_2} E_2 = -\zeta_2 E_1, \quad \nabla_{E_2} E_3 = 0, \quad \nabla_{E_2} E_4 = 0, \quad \nabla_{E_3} E_2 = 0,$$

$$\nabla_{E_3} E_3 = -\zeta_3 E_1, \quad \nabla_{E_3} E_4 = 0, \quad \nabla_{E_4} E_2 = 0, \quad \nabla_{E_4} E_3 = -\beta_4 E_4,$$

$$\nabla_{E_4} E_4 = -\zeta_4 E_1 + \beta_4 E_3 \quad \textit{for some function } \beta_4,$$

*and*

$$R_{1221} = -\frac{p''}{p} = -\zeta_2' - \zeta_2^2,$$

$$R_{1ii1} = -\zeta_i' - \zeta_i^2 = -\frac{h''}{h} \quad \textit{for } i = 3, 4,$$

$$R_{11} = -\zeta_2' - \zeta_2^2 - 2\zeta_3' - 2\zeta_3^2 = -\frac{p''}{p} - 2\frac{h''}{h},$$

$$R_{22} = -\zeta_2' - \zeta_2^2 - 2\zeta_2\zeta_3 = -\frac{p''}{p} - 2\frac{p'}{p}\frac{h'}{h},$$

$$R_{33} = R_{44} = -\zeta_3' - \zeta_3^2 - \zeta_3\zeta_2 - (\zeta_3)^2 + \frac{k}{h^2} = -\frac{h''}{h} - \frac{p'}{p}\frac{h'}{h} - \frac{(h')^2}{h^2} + \frac{k}{h^2},$$

$$R_{ij} = 0 \quad \textit{for } i \neq j.$$

*Proof.* Now $\nabla_{E_1} E_1 = 0$ from Lemma 2.3(vi) and $\nabla_{E_i} E_1 = \zeta_i E_i,\ i > 1$, from (3-1). From the proof of Lemma 5.1, we already have $\nabla_{E_1} E_2 = 0,\ \nabla_{E_3} E_4 = \beta_3 E_3$ and $\nabla_{E_4} E_3 = -\beta_4 E_4$.

As $\langle \nabla_{E_1} E_3, E_2 \rangle = -\langle E_3, \nabla_{E_1} E_2 \rangle = 0$, one can readily get $\nabla_{E_1} E_3 = \rho E_4$ for some function $\rho$ and $\nabla_{E_1} E_4 = -\rho E_3$. We get $\rho = 0$ by computing directly (in coordinates)

$$\nabla_{E_1} E_3 = \nabla_{\partial/\partial s} \frac{1}{h(s)} \frac{\partial}{\partial r} = 0.$$

From Lemma 3.1 and Lemma 2.4(iii), we have

$$(\lambda_2 - \lambda_i)\langle \nabla_{E_2} E_2, E_i \rangle = E_i(\lambda_2) = 0 \quad \text{for } i = 3, 4,$$
$$\langle \nabla_{E_2} E_2, E_1 \rangle = -\langle E_2, \nabla_{E_2} E_1 \rangle = -\zeta_2(s).$$

So, $\nabla_{E_2} E_2 = -\zeta_2(s) E_1$. By a similar argument, $\nabla_{E_3} E_3 = -\zeta_3 E_1 - \beta_3 E_4$ and $\nabla_{E_4} E_4 = -\zeta_4 E_1 + \beta_4 E_3$. Direct computation of the coordinates gives $\beta_3 = 0$.

Then $\nabla_{E_2} E_3 = q E_4$ for some function $q$ and $\nabla_{E_2} E_4 = -q E_3$. One computes directly that $q = 0$. We similarly get $\nabla_{E_3} E_2 = 0$ and $\nabla_{E_4} E_2 = 0$.

We compute directly that $\nabla_{E_2} E_1 = (p'/p) E_2$ and $\nabla_{E_3} E_1 = (h'/h) E_3$ so that (3-1) gives $\zeta_2 = p'/p$ and $\zeta_3 = \zeta_4 = h'/h$. We now get $\nabla_{E_3} E_4 = 0$ and $\nabla_{E_4} E_3 = -\beta_4 E_4$, where $\beta_4 = u'(r)/(h(s)u(r))$.

With these computations in hand, it is straightforward to compute the curvature components. □

We set $a := \zeta_2$ and $b := \zeta_3$.

**Lemma 6.2.** *For the local metric $g = ds^2 + p(s)^2\, dt^2 + h(s)^2 \tilde{g}$ with harmonic curvature satisfying* (1-1) *with nonconstant $f$, obtained in Lemma 5.2, it holds that*

$$\big(ab + \tfrac{1}{12} R\big) b = 0. \tag{6-1}$$

*Proof.* Equation (4-2) gives

$$2a' + 2a^2 + 2ab + \tfrac{1}{3} R = 0, \tag{6-2}$$

$$2b' + 3b^2 + ab - \frac{k}{h^2} + \tfrac{1}{3} R = 0. \tag{6-3}$$

From $\nabla df(E_i, E_i) = f\big(\mathrm{Rc} - \frac{1}{3} Rg\big)(E_i, E_i) + x R(E_i, E_i) + y(R)$, we get

$$-(\nabla_{E_i} E_i) f = f\big(R_{ii} - \tfrac{1}{3} R\big) + x R_{ii} + y(R) = -f R_{1ii1} + x R_{ii} + y(R)$$

for $i = 2, 3$. From Lemma 6.1 we have

$$f' a = f(a' + a^2) - x(a' + a^2 + 2ab) + y(R), \tag{6-4}$$

$$f' b = f(b' + b^2) - x\left(b' + 2b^2 + ab - \frac{k}{h^2}\right) + y(R). \tag{6-5}$$

From the harmonic curvature condition we have

$$\begin{aligned} 0 &= \nabla_{E_1} R_{22} - \nabla_{E_2} R_{12} = \nabla_{E_1}(R_{22}) + R(\nabla_{E_2} E_1, E_2) + R(\nabla_{E_2} E_2, E_1) \\ &= (R_{22})' + a(R_{22} - R_{11}) \\ &= (-a' - a^2 - 2ab)' + a(-2ab + 2b' + 2b^2) \\ &= -a'' - 2aa' - 2a'b - 2a^2 b + 2ab^2. \end{aligned} \tag{6-6}$$

We differentiate (6-2) to get $a'' + 2aa' + a'b + ab' = 0$. Together with (6-6) we obtain

$$ab' - a'b - 2a^2 b + 2ab^2 = 0. \tag{6-7}$$

Putting (6-2) and (6-3) into (6-7) we get

$$\begin{aligned} 0 &= -a\left(3b^2 + ab - \frac{k}{h^2} + \tfrac{1}{3} R\right) + 2\big(a^2 + ab + \tfrac{1}{6} R\big) b - 4a^2 b + 4ab^2 \\ &= a\frac{k}{h^2} + \tfrac{1}{3} R(b - a) + 3ab(b - a). \end{aligned}$$

Then, as $a \neq b$,

(6-8)
$$\frac{a}{a-b}\frac{k}{h^2} = \tfrac{1}{3}R + 3ab.$$

From (6-4) and (6-5) we get

$$\frac{f'}{f}(a-b) = (a'+a^2-b'-b^2) - \frac{x}{f}\left(a'+a^2+2ab-b'-2b^2-ab+\frac{k}{h^2}\right).$$

With (6-3) and (6-2), the above gives

$$2\frac{f'}{f}(a-b) = \left(1+\frac{x}{f}\right)\left(b^2-ab-\frac{k}{h^2}\right).$$

Then by (6-8),

$$2\frac{f'}{f}a = \left(1+\frac{x}{f}\right)\left(-ab-\frac{ka}{h^2(a-b)}\right) = \left(1+\frac{x}{f}\right)\left(-4ab-\tfrac{1}{3}R\right).$$

Meanwhile, (6-4) and (6-2) give $f'a = -f\big(ab+\frac{1}{6}R\big) - x\big(ab-\frac{1}{6}R\big) + y(R)$, so

$$-2\big(ab+\tfrac{1}{6}R\big) - \frac{2x}{f}\big(ab-\tfrac{1}{6}R\big) + \frac{2y(R)}{f} = 2\frac{f'}{f}a = \left(1+\frac{x}{f}\right)\left(-4ab-\tfrac{1}{3}R\right).$$

So we obtain

(6-9)
$$x\big(ab+\tfrac{1}{3}R\big) + y(R) = -fab.$$

Differentiating (6-9) and dividing by $f$,

$$\frac{f'}{f}ab = -\frac{x}{f}(a'b+ab') - (a'b+ab').$$

From (6-4) we get

$$\frac{f'}{f}ab = (a'+a^2)b - \frac{x}{f}(a'+a^2+2ab)b + \frac{yb}{f}.$$

Equating the above and arranging terms, we get

$$\frac{x}{f}(-ab'+a^2b+2ab^2) = 2a'b+ab'+a^2b+\frac{yb}{f}.$$

Using (6-9) we get

(6-10)
$$\frac{x}{f}\big(-ab'+a^2b+3ab^2+\tfrac{1}{3}Rb\big) = 2a'b+ab'+a^2b-ab^2.$$

Using (6-7) and (6-2), the left-hand side of (6-10) equals $(x/f)\big(6ab^2+\frac{1}{2}Rb\big)$, while the right-hand side equals $-\big(6ab^2+\frac{1}{2}Rb\big)$.

So we get $(1+x/f)\big(6ab+\frac{1}{2}R\big)b = 0$. Then $(ab+\frac{1}{12}R)b = 0$. □

**Proposition 6.3.** *For the local metric* $g = ds^2 + p(s)^2\,dt^2 + h(s)^2\tilde{g}$ *with harmonic curvature satisfying* (1-1) *with nonconstant* $f$, *obtained in Lemma 5.2, suppose that* $ab = -\frac{1}{12}R$.

*Then* $R = 0$, $y(0) = 0$ *and* $p$ *is a constant. The metric* $g$ *is locally isometric to a domain in the nonconformally flat static space* $(W^3 \times \mathbb{R}^1, g_W + dt^2)$ *of Example 3 in Section 2A2. Moreover,* $f = ch'(s) - x$.

*Proof.* As $ab = -\frac{1}{12}R$, (6-9) gives $\frac{1}{4}Rx + y(R) = \frac{1}{12}Rf$.

If $R \neq 0$, then $f$ is a constant, a contradiction to the hypothesis. Therefore $R = 0$. Then $y(0) = 0$ from the preceding equation. From (6-2), $a' + a^2 = 0$ and we have two cases: (i) $a = 1/(s+c)$ for a constant $c$ or (ii) $a = 0$.

Case (i): $a = 1/(s+c)$. From (6-4), $f'a = 0$, so $f$ is a constant, a contradiction to the hypothesis.

Case (ii): $a = 0$, i.e., $p$ is a constant. From (6-5) and (6-3), we get $f'(h'/h) = (f+x)(h''/h)$. If $h'$ vanishes, we get $\lambda_2 = \lambda_3$, a contradiction. So we may assume that $h$ is not constant. Then $ch' = f + x$ for a constant $c \neq 0$. Evaluating (1-1) at $(E_1, E_1)$,

$$f'' = (f+x)R(E_1, E_1) - \tfrac{1}{3}Rf + y(R). \tag{6-11}$$

Here we get $f'' = -2(f+x)(h''/h)$, so $h''' = -2h'(h''/h)$. Hence, for a constant $\alpha$,

$$h^2h'' = \alpha. \tag{6-12}$$

From (6-3),

$$0 = 2b' + 3b^2 - \frac{k}{h^2} = 2\left(\frac{h''}{h}\right) + \left(\frac{h'}{h}\right)^2 - \frac{k}{h^2} = \frac{2\alpha}{h^3} + \left(\frac{h'}{h}\right)^2 - \frac{k}{h^2}.$$

So we have

$$(h')^2 + \frac{2\alpha}{h} - k = 0. \tag{6-13}$$

We have exactly (2-2) and (2-3) in the case $R = 0$ and $n = 3$. At this point we may write

$$g = ds^2 + dt^2 + h(s)^2\tilde{g} = \left(k - \frac{2\alpha}{h}\right)^{-1} dh^2 + dt^2 + h(s)^2\tilde{g}.$$

When $\alpha = 0$, we have $(h')^2 = k \geq 0$. As $h$ is not constant, $k > 0$. When $h' = \pm\sqrt{k} \neq 0$, we have $h = \pm\sqrt{k}s + c_0$ for a constant $c_0$. One can see that $g$ is a flat metric, a contradiction to $\lambda_2 \neq \lambda_3$.

When $\alpha > 0$, then $k > 0$ from (6-13). We set $r := h/\sqrt{k}$, and then

$$g = \left(1 - \frac{2\alpha}{k\sqrt{k}r}\right)^{-1} dr^2 + dt^2 + r^2\tilde{g}_1,$$

where $\tilde{g}_1$ is the metric of constant curvature 1 on $S^2$. When $\alpha < 0$, the three-dimensional metric $(1 - 2\alpha/(k\sqrt{k}r))^{-1}dr^2 + r^2\tilde{g}_1$ corresponds to case III.1 of Kobayashi's conditions [1982, p. 670]. It is incomplete as explained in his Proposition 2.4.

In these two cases of $\alpha > 0$ and $\alpha < 0$, we get the same Riemannian metrics as those of static spaces $(W^3 \times \mathbb{R}^1, g_W + dt^2)$ explained in Example 3, and $f = ch' - x$.

Conversely, these metrics have harmonic curvature and satisfy (1-1) with the above $f$. Indeed, nontrivial components of (1-1) are (6-4), (6-5) and (6-11), whereas the harmonic curvature condition essentially consists of (6-6) and the equation $\nabla_{E_1} R_{33} - \nabla_{E_3} R_{13} = 0$; all these can be verified from $a = R = y(0) = 0$ and $h$, $f$ which satisfy (6-12), (6-13) and $f = ch' - x$. □

**Proposition 6.4.** *For the local metric $g = ds^2 + p(s)^2\,dt^2 + h(s)^2\tilde{g}$ with harmonic curvature satisfying* (1-1) *with nonconstant $f$, obtained in Lemma 5.2, suppose that $b = 0$ and that $ab = 0 \neq -\frac{1}{12}R$. Then the following hold*:

(i) $\frac{1}{3}xR + y(R) = 0$.

(ii) *If $R > 0$, then $g$ is locally isometric to the Riemannian product $\big(\mathbb{S}^2\big(\frac{R}{6}\big) \times \mathbb{S}^2\big(\frac{R}{3}\big),\ g_{R/6} + g_{R/3}\big)$, where $g_\delta$ is the two-dimensional Riemannian metric of constant curvature $\delta$, and $f = c_1\cos\big(\sqrt{\frac{R}{6}}s\big) - x$ for any constant $c_1$, where $s$ is the distance from a point on $\mathbb{S}^2\big(\frac{R}{6}\big)$.*

(iii) *If $R < 0$, then $g$ is locally isometric to $\big(\mathbb{H}^2\big(\frac{R}{6}\big) \times \mathbb{H}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$. The metric $g_{R/6}$ can be written as $g_{R/6} = ds^2 + p(s)^2\,dt^2$ with $p(s) = k_1\sinh\big(\sqrt{-\frac{R}{6}}s\big) + k_2\cosh\big(\sqrt{-\frac{R}{6}}s\big)$ for constants $k_1$, $k_2$, and then $f = c_2p' - x$ for any constant $c_2$.*

*Proof.* As $b = 0$, (6-9) gives (i). Next, (6-3) gives $k/h^2 = \frac{1}{3}R$ and (6-2) gives $a' + a^2 + \frac{1}{6}R = p''/p + \frac{1}{6}R = 0$. Along with (6-4) these give

$$f'a = -\tfrac{1}{6}R(f + x). \tag{6-14}$$

Assume $R > 0$. Set $r_0 = \sqrt{\frac{R}{6}}$. For some constants $C_1 \neq 0$ and $s_0$, we have $p = C_1\sin(r_0(s + s_0))$ so that $a = r_0\cot(r_0(s + s_0))$. Then (6-14) and (i) give $f = c_1\cos(r_0(s + s_0)) - x$. Then $g = ds^2 + \sin^2(r_0(s + s_0))\,dt^2 + \tilde{g}_{R/3}$ by absorbing a constant into $dt^2$ and using $k/h^2 = \frac{1}{3}R$.

Replacing $s + s_0$ by a new $s$, we have $g = ds^2 + \sin^2(r_0s)\,dt^2 + \tilde{g}_{R/3}$. Here $s$ becomes the distance from a point on $\mathbb{S}^2\big(\frac{R}{6}\big)$. And $f = c_1\cos(r_0s) - x$.

Assume $R < 0$. From $p''/p + \frac{1}{6}R = 0$ we get $p(s) = k_1\sinh(r_1s) + k_2\cosh(r_1s)$ for constants $k_1$, $k_2$, where $r_1 = \sqrt{-\frac{R}{6}}$, and $f = c_2p' - x$ for any constant $c_2$.

Conversely, the above product metrics clearly have harmonic curvature. One can check they satisfy (1-1). Indeed, as in the proof of Proposition 6.3 one may check (6-4), (6-5) and (6-11). □

## 7. Local four-dimensional space with harmonic curvature

We first treat the remaining case of $\lambda_2 = \lambda_3 = \lambda_4$ and then give the proof of Theorem 1.1.

**Proposition 7.1.** *Let $(M, g, f)$ be a four-dimensional Riemannian manifold with harmonic curvature satisfying* (1-1) *with nonconstant $f$. Suppose that $\lambda_2 = \lambda_3 = \lambda_4 \neq \lambda_1$ for an adapted frame field in an open subset $U$ of $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$.*

*Then for each point $p_0$ in $U$, there exists a neighborhood $V$ of $p_0$ in $U$ where $g$ is a warped product,*

$$g = ds^2 + h(s)^2 \tilde{g}, \tag{7-1}$$

*where $h$ is a positive function and the Riemannian metric $\tilde{g}$ has constant curvature, say $k$. In particular, $g$ is conformally flat.*

*As a Riemannian manifold, $(M, g)$ is locally one of Kobayashi's warped product spaces, as described in Sections* 2 *and* 3 *of [Kobayashi 1982], so that*

$$h'' + \tfrac{1}{12} R h = a h^{-3} \tag{7-2}$$

*for a constant $a$, so that by integration we have for some constant $k$*

$$(h')^2 + a h^{-2} + \tfrac{1}{12} R h^2 = k. \tag{7-3}$$

*Moreover, $f$ is a nonconstant solution to*

$$h' f' - f h'' = x\big(h'' + \tfrac{1}{3} R h\big) + y(R) h. \tag{7-4}$$

*Conversely, any $(h, f)$ satisfying* (7-2), (7-3) *and* (7-4) *gives rise to $(g, f)$ which has harmonic curvature and satisfies* (1-1).

*Proof.* To prove that $g$ is in the form of (7-1), we may use Lemma 2.3(v) and Lemma 2.4(iii)–(iv). For a detailed proof we refer to that of Proposition 7.1 of [Kim 2017] since the argument is almost the same as in the gradient Ricci soliton case. To prove that $\tilde{g}$ has constant curvature, we use Lemma 4 in [Derdziński 1980]. It then follows that the metric $g$ in (7-1) is conformally flat.

In the setting of Lemma 2.3, $f$ is a function of $s$ only. For $g = ds^2 + h(s)^2 \tilde{g}$, in a local adapted frame field, we have

$$\begin{gathered} R_{11} = -3\frac{h''}{h}, \quad R_{ii} = -\frac{h''}{h} - 2\frac{(h')^2}{h^2} + 2\frac{k}{h^2}, \\ R_{ij} = 0 \quad \text{for } i \neq j, \\ R = -6\frac{h''}{h} - 6\frac{(h')^2}{h^2} + 6\frac{k}{h^2}. \end{gathered} \tag{7-5}$$

Feeding $(E_i, E_i)$, $i = 1, 2$ to (1-1) we obtain

$$f'' = -3f\frac{h''}{h} - f\tfrac{1}{3}R - 3x\frac{h''}{h} + y(R), \tag{7-6}$$

$$h' f' - f h'' = x\big(h'' + \tfrac{1}{3} R h\big) + y(R) h. \tag{7-7}$$

Differentiating (7-7) and using (7-6), we get

$$(f+x)\left\{h''' + 3\frac{h''h'}{h} + \tfrac{1}{3}Rh'\right\} = 0.$$

As $f \neq -x$, we get

$$h''' + 3\frac{h''h'}{h} + \tfrac{1}{3}Rh' = 0.$$

Multiplying this by $h^3$, we get $\left(h^3h'' + \frac{1}{12}Rh^4\right)' = 0$. Then we have (7-2) and then (7-3). Kobayashi solved these completely according to each parameter and initial condition.

One can check that any $h$ and $f$ satisfying (7-7), (7-2) and (7-3) satisfy (7-5) and (7-6). □

We are ready to prove Theorem 1.1.

*Proof of Theorem 1.1.* Recall that we have already discussed the case $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ in Example 1 of Section 2A2. The conformally flat spaces in Example 1 belong to the type (iv) of Theorem 1.1; in particular $a = 0$ in (1-6) and (1-7).

As the metrics $g$ and $f$ are real analytic, the Ricci-eigenvalues $\lambda_i$ are real analytic on $M_{\mathrm{Rc}} \cap \{\nabla f \neq 0\}$. And $\zeta_i$'s are real analytic from (3-1). So we can combine Proposition 4.2, Lemma 6.2, Propositions 6.3, 6.4, 7.1 and Example 1 of Section 2A2, to obtain a classification of four-dimensional *local* spaces with harmonic curvature satisfying (1-1) as Theorem 1.1. □

**Remark 7.2.** In the statement of Theorem 1.1, among the types (i)–(iv), there is possibly only one type of neighborhood $V$ on a *connected* space $(M, g, f)$; this holds by a continuity argument of Riemannian metrics. Then one can prove that $\widetilde{M} = M$ if $M$ is of type (i), (ii) or (iii).

## 8. Complete four-dimensional space with harmonic curvature

It is not hard to describe complete spaces corresponding to parts (i), (ii), (iii) of Theorem 1.1.

For the complete conformally flat case corresponding to (iv) of Theorem 1.1, we may use Theorem 3.1 of Kobayashi's classification [1982]. Then $(M, g)$ can be either $\mathbb{S}^4$, $\mathbb{H}^4$, a flat space or one of the spaces in Examples 1–5 in [Kobayashi 1982]. Now our task is to determine $f$, which is described by (1-8).

We first recall the spaces in Examples 3–5 in [Kobayashi 1982]. Any space in Examples 3 and 4 in that paper is a quotient of a warped product $\mathbb{R} \times_h N(1)$ where $h$ is a smooth periodic function on $\mathbb{R}$; recall that $N(k)$ is a Riemannian manifold of constant sectional curvature $k$. Any space in Example 5 in that paper is a quotient of a warped product $\mathbb{R} \times_h N(k)$ where $h$ is smooth on $\mathbb{R}$. Here $h \geq \rho_1 > 0$.

We verify the following lemma.

**Lemma 8.1.** *For any one of the spaces in Examples* 3, 4 *and* 5 *in [Kobayashi 1982], the following hold*:

(i) *The solution* $f$ *to* (1-1) *can be defined and is smooth on* $\mathbb{R}$.

(ii) *If* $h$ *is periodic and* $\frac{1}{3}xR + y(R) = 0$, *then* $f$ *is periodic.*

*Proof.* As stated in Proposition 7.1, any $(h, f)$ satisfying (7-2), (7-3) and (7-4) gives rise to $(g, f)$ which satisfies (1-1). So, $(h, f)$ satisfies (7-6).

Choose some point $s_0$ with $h''(s_0) \neq 0$. For any constant $c$, we consider the initial-value problem

$$f'' = -f\left(\tfrac{1}{12}R + 3ah^{-4}\right) + 3x\left(\tfrac{1}{12}R - ah^{-4}\right) + y(R), \tag{8-1}$$

with initial conditions $f'(s_0) = c$ and

$$f(s_0) = \frac{ch'(s_0) - \left\{x(h''(s_0) + \frac{1}{3}Rh(s_0)) + y(R)h(s_0)\right\}}{h''(s_0)}$$

so that (1-8) holds at $s_0$. Note that (8-1) is equivalent to (7-6) since $h$ satisfies (1-6).

As $h$ exists smoothly on $\mathbb{R}$ as a solution of (1-6), by global Lipschitz continuity of the right-hand side of (8-1), the solution $f$ exists globally on $\mathbb{R}$.

From (1-6) we obtain

$$h''' = -\left(\tfrac{1}{12}R + 3ah^{-4}\right)h'. \tag{8-2}$$

Then by (8-1) and (8-2) it satisfies

$$h'f'' - fh''' = x\left(h''' + \tfrac{1}{3}Rh'\right) + y(R)h',$$

which is the derivative of (1-8). So, (1-8) holds on $\mathbb{R}$. As $h$ and $f$ satisfy (1-8), the induced $(g, f)$ satisfies (1-1) on $\mathbb{R}$.

If $\frac{1}{3}xR + y(R) = 0$, then from (1-8) we get $f(s) = -x + Ch'(s)$ for a constant $C$, which is periodic as $h$. □

About Lemma 8.1(ii), we note that if $\frac{1}{3}xR + y(R) \neq 0$ and $h$ is periodic, then the periodicity of $f$ should be checked by computation.

We are ready to state the following result.

**Theorem 8.2.** *Let* $(M, g)$ *be a four-dimensional complete Riemannian manifold with harmonic curvature satisfying* (1-1) *with nonconstant* $f$. *Then it is one of the following*:

(8.2-i) $(M, g)$ *is isometric to a quotient of* $\left(\mathbb{S}^2\left(\frac{R}{6}\right) \times \mathbb{S}^2\left(\frac{R}{3}\right), g_{R/6} + g_{R/3}\right)$ *with* $R > 0$, *where* $f = c_1 \cos\left(\sqrt{\frac{R}{6}}s\right) - x$ *for any constant* $c_1$, *where* $s$ *is the distance from a point on* $\mathbb{S}^2\left(\frac{R}{6}\right)$. *It holds that* $\frac{1}{3}xR + y(R) = 0$.

(8.2-ii) *$(M, g)$ is isometric to a quotient of* $\big(\mathbb{H}^2\big(\frac{R}{6}\big) \times \mathbb{H}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$ *with* $R < 0$. *The metric $g_{R/6}$ can be written as $g_{R/6} = ds^2 + p(s)^2\,dt^2$ with* $p(s) = k_1 \sinh\big(\sqrt{-\frac{R}{6}}s\big) + k_2 \cosh\big(\sqrt{-\frac{R}{6}}s\big)$ *for constants $k_1$, $k_2$, and then $f = c_2 p' - x$ for any constant $c_2$. It holds that* $\frac{1}{3}xR + y(R) = 0$.

(8.2-iii) *$(M, g)$ is isometric to a quotient of one of the static spaces in Example 3 of Section 2A2, which is the Riemannian product* $(\mathbb{R}^1 \times W^3, dt^2 + ds^2 + h(s)^2\tilde{g})$ *of* $\mathbb{R}^1$ *and some three-dimensional conformally flat static space* $(W^3 = \mathbb{R}^1 \times \mathbb{S}^2(1), ds^2 + h(s)^2\tilde{g})$ *with zero scalar curvature, which contains the spatial slice of the Schwarzschild space-time*

*And $f = c \cdot h'(s) - x$ for a constant $c$. It holds that* $R = y(0) = 0$.

(8.2-iv) *$(M, g)$ is conformally flat. It is either $\mathbb{S}^4$, $\mathbb{H}^4$, a flat space or one of the spaces in Examples* 1–5 *in [Kobayashi 1982]. Below we describe $f$ in each subcase*:

(8.2-iv-1) $\mathbb{S}^4(k^2)$ *with the metric* $g = ds^2 + (\sin(ks)^2/k^2)g_1$ *for any constant $c$*,

$$f(s) = c \cdot \cos(ks) + 3x + \frac{y(12k^2)}{k^2}.$$

(8.2-iv-2) $\mathbb{H}^4(-k^2)$ *with* $g = ds^2 + (\sinh(ks)^2/k^2)g_1$ *for any constant $c$*,

$$f(s) = c \cdot \cosh(ks) + 3x - \frac{y(-12k^2)}{k^2}.$$

(8.2-iv-3) *A flat space,* $f = a + \sum_i +b_i x_i + \frac{1}{2}y(0)x_i^2$ *in local Euclidean coordinates $x_i$ for constants $a$ and $b_i$.*

(8.2-iv-4) *Examples* 1 *and* 2 *in [Kobayashi 1982]*: *the Riemannian product* $(\mathbb{R}\times N(k), ds^2 + g_k)$ *or its quotient, $k \neq 0$, where $N(k)$ is three-dimensional complete space of constant sectional curvature $k$*,

$$f = \begin{cases} c_1 \sin\sqrt{\frac{R}{3}}s + c_2 \cos\sqrt{\frac{R}{3}}s - x & \text{when } R > 0,\\ c_1 \sinh\sqrt{-\frac{R}{3}}s + c_2 \cosh\sqrt{-\frac{R}{3}}s - x & \text{when } R < 0.\end{cases}$$

*It holds that* $\frac{1}{3}xR + y(R) = 0$ *and* $R = 6k$.

(8.2-iv-5) *Examples* 3 *and* 4 *in [Kobayashi 1982]*: *a warped product* $\mathbb{R} \times_h N(1)$ *or its quotient, where $h$ is a periodic function on* $\mathbb{R}$, *$f$ is on* $\mathbb{R}$, *satisfying* (1-8).

(8.2-iv-6) *Example* 5 *in [Kobayashi 1982]*: *a warped product* $\mathbb{R} \times_h N(k)$ *where $h$ is defined on* $\mathbb{R}$, *$f$ is on* $\mathbb{R}$, *satisfying* (1-8).

*Proof.* To obtain (8.2-i), (8.2-ii) and (8.2-iii), we use the continuity argument of Riemannian metrics from Theorem 1.1. To describe $f$ in the subcases of (8.2-iv), we use (1-8) and (7-6). □

## 9. Four-dimensional static spaces with harmonic curvature

In this section we study static spaces, i.e., those satisfying (1-2). As explained in the Introduction, studying local static spaces is interesting due to Corvino's local deformation theory of scalar curvature. Qing and Yuan's work [2016] on local scalar curvature rigidity arouses another motivation. Here we state a local classification which can be read off from Theorem 1.1:

**Theorem 9.1.** *Let $(M, g, f)$ be a four-dimensional (not necessarily complete) static space with harmonic curvature and nonconstant $f$. Then for each point $p$ in some open dense subset $\widetilde{M}$ of $M$, there exists a neighborhood $V$ of $p$ with one of the following properties*:

(9.1-i) *$(V, g)$ is isometric to a domain in $\big(\mathbb{S}^2\big(\frac{R}{6}\big) \times \mathbb{S}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$ with $R > 0$. And $f = c_1 \cos\big(\sqrt{\frac{R}{6}}(s + s_0)\big)$, where $s$ is the distance from a point on $\mathbb{S}^2\big(\frac{R}{6}\big)$ and $c_1, s_0$ are constants.*

(9.1-ii) *$(V, g)$ is isometric to a domain in $\big(\mathbb{H}^2\big(\frac{R}{6}\big) \times \mathbb{H}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$ with $R < 0$. The metric $g_{R/6}$ can be written as $g_{R/6} = ds^2 + p(s)^2\, dt^2$ with $p(s) = k_1 \sinh\big(\sqrt{-\frac{R}{6}}s\big) + k_2 \cosh\big(\sqrt{-\frac{R}{6}}s\big)$ for constants $k_1, k_2$, and then $f = c_2 p'$ for any constant $c_2$.*

(9.1-iii) *$(V, g)$ is isometric to a domain in one of the static spaces in Example 3 of Section 2A2, which is the Riemannian product $\mathbb{R}^1 \times W^3$ of $\mathbb{R}^1$ and some three-dimensional conformally flat static space $(W^3, ds^2 + h(s)^2\tilde{g})$ with zero scalar curvature, and $f = ch'$.*

(9.1-iv) *$(V, g)$ is conformally flat. So, it is one of the warped product metrics of the form $ds^2 + h(s)^2 g_k$ whose existence is described in Section* 2 *of [Kobayashi 1982]. The function $h$ satisfies* (1-6) *and* (1-7), *and we have $f(s) = Ch'(s)$.*

For complete conformally flat case corresponding to (9.1-iv) in Theorem 9.1, if we use Theorem 3.1 of Kobayashi's classification, we get either $\mathbb{S}^4$, $\mathbb{H}^4$, a flat space or one of the spaces in Examples 1–5 in [Kobayashi 1982]. We may thus obtain classification of complete four-dimensional static spaces with harmonic curvature:

**Theorem 9.2.** *Let $(M, g, f)$ be a complete four-dimensional static space with harmonic curvature. Then it is one of the following*:

(9.2-i) *$(M, g)$ is isometric to a quotient of $\big(\mathbb{S}^2\big(\frac{R}{6}\big) \times \mathbb{S}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$ with $R > 0$. And $f = c_1 \cos\big(\sqrt{\frac{R}{6}}s\big)$, where $s$ is the distance function from a point on $\mathbb{S}^2\big(\frac{R}{6}\big)$.*

(9.2-ii) *$(M, g)$ is isometric to a quotient of $\big(\mathbb{H}^2\big(\frac{R}{6}\big) \times \mathbb{H}^2\big(\frac{R}{3}\big), g_{R/6} + g_{R/3}\big)$ with $R < 0$. The metric $g_{R/6}$ can be written as $g_{R/6} = ds^2 + p(s)^2\, dt^2$ with $p(s) = k_1 \sinh\big(\sqrt{-\frac{R}{6}}s\big) + k_2 \cosh\big(\sqrt{-\frac{R}{6}}s\big)$ for constants $k_1, k_2$, and then $f = c_2 p'$ for any constant $c_2$.*

(9.2-iii) $(M, g)$ *is isometric to a quotient of the Riemannian product* $(\mathbb{R}^1 \times W^3, dt^2 + \tilde{g})$, *where* $(W^3, \tilde{g})$ *denotes the warped product manifold on the smooth product* $\mathbb{R}^1 \times \mathbb{S}^2(1)$ *which contains the spatial slice of the Schwarzschild space-time*; *see Example 3 of Section 2A2.*

(9.2-iv) $(M, g, f)$ *is* $\mathbb{S}^4$, $\mathbb{H}^4$, *a flat space or one of the spaces in Examples* 1–5 *in [Kim 2017].*

(9.2-v) $g$ *is a complete Ricci-flat metric with* $f$ *a constant function.*

*Proof.* It follows from Theorem 8.2. When $f$ is a nonzero constant, $g$ is clearly Ricci-flat. So we get (v). □

Fischer and Marsden [1974] made the conjecture that any closed static space is Einstein. But it was disproved by conformally flat examples in [Lafontaine 1983; Kobayashi 1982]. Now we ask:

**Question 1.** Does there exist a closed static space which does not have harmonic curvature?

The space in (9.2-iii) of Theorem 9.2 has three distinct Ricci-eigenvalues. We only know examples of static spaces with at most three distinct Ricci-eigenvalues. So we ask the following:

**Question 2.** Does there exist a static space with more than three distinct Ricci-eigenvalues? Is there a limit on the number of distinct Ricci-eigenvalues for a static space?

## 10. Miao–Tam critical metrics and $V$-critical spaces

In this section we treat Miao–Tam critical metrics. These metrics originate from [Miao and Tam 2009], where they studied the critical points of the volume functional on the space $\mathcal{M}_\gamma^K$ of metrics with constant scalar curvature $K$ on a compact manifold $M$ with a prescribed metric $\gamma$ at the boundary of $M$. Miao–Tam critical metrics are precisely described [Miao and Tam 2011] in case they are Einstein or conformally flat.

Here we first describe four-dimensional metrics with harmonic curvature which have a nonzero solution $f$ to (1-3). We do not assume the condition $f_{|\Sigma} = 0$ but still can show that any such metric must be conformally flat;

**Theorem 10.1.** *Let* $(M, g)$ *be a four-dimensional* (*not necessarily complete*) *Riemannian manifold with harmonic curvature*, *satisfying* (1-3) *with nonconstant* $f$. *Then* $(M, g)$ *is conformally flat. It is one of the warped product metrics of the form* $ds^2 + h(s)^2 g_k$ *whose existence is described in Section* 2 *of [Kobayashi 1982]. The function* $h$ *satisfies* (1-6) *and* (1-7), *and* $f$ *satisfies* $h' f' - f h'' = -h/(n-1)$.

*Proof.* The proof is immediate from Theorem 1.1; the cases (i)–(ii) of Theorem 1.1 require $\frac{1}{3}xR + y(R) = 0$ and (iii) requires $y(0) = 0$, which contradict the conditions $x = 0$ and $y(R) = -\frac{1}{3}$ that (1-3) has. The description of Theorem 1.1(iv) holds for $g$ and $f$ of Theorem 10.1, and in particular $g$ is conformally flat. □

Theorem 10.1 shows an advantage of our local approach over [Barros et al. 2015] in analyzing (1-3). In fact, the integration argument of Lemma 5 of that paper only works for compact manifolds, but our analysis can resolve local solutions.

From Theorems 9.1 and 10.1 we can classify local four-dimensional $V$-static spaces with harmonic curvature:

**Theorem 10.2.** *Let* $(M, g, f)$ *be a four-dimensional* (*not necessarily complete*) $V$*-static space with harmonic curvature and nonconstant* $f$*. Then for each point* $p$ *in some open dense subset* $\widetilde{M}$ *of* $M$*, there exists a neighborhood* $V$ *of* $p$ *with one of the following properties*:

(10.2-i) $(V, g)$ *is isometric to a domain in* $\left(\mathbb{S}^2\left(\frac{R}{6}\right) \times \mathbb{S}^2\left(\frac{R}{3}\right), g_{R/6} + g_{R/3}\right)$ *with* $R > 0$*. And* $f = c_1 \cos\left(\sqrt{\frac{R}{6}}(s + s_0)\right)$*, where* $s$ *is the distance function from a point on* $\mathbb{S}^2\left(\frac{R}{6}\right)$ *and* $c_1, s_0$ *are constants.*

(10.2-ii) $(V, g)$ *is isometric to a domain in* $\left(\mathbb{H}^2\left(\frac{R}{6}\right) \times \mathbb{H}^2\left(\frac{R}{3}\right), g_{R/6} + g_{R/3}\right)$ *with* $R < 0$*. The metric* $g_{R/6}$ *can be written as* $g_{R/6} = ds^2 + p(s)^2\, dt^2$ *with* $p(s) = k_1 \sinh\left(\sqrt{-\frac{R}{6}}s\right) + k_2 \cosh\left(\sqrt{-\frac{R}{6}}s\right)$ *for constants* $k_1, k_2$*, and then* $f = c_2 p'$ *for any constant* $c_2$*.*

(10.2-iii) $(V, g)$ *is isometric to a domain in one of the static spaces in Example 3 of Section 2A2 which is the Riemannian product* $\mathbb{R}^1 \times W^3$ *of* $\mathbb{R}^1$ *and some three-dimensional conformally flat static space* $(W^3, ds^2 + h(s)^2\tilde{g})$ *with zero scalar curvature. And* $f = ch'$ *for any constant* $c$*.*

(10.2-iv) $(V, g)$ *is conformally flat. It is one of the warped product metrics of the form* $ds^2 + h(s)^2 g_k$ *whose existence is described in Section* 2 *of [Kobayashi 1982]. The function* $h$ *satisfies* (1-6) *and* (1-7), *and we have* $f(s) = ch'(s)$ *for any constant* $c$*.*

(10.2-v) $(V, g)$ *is conformally flat. It is one of the warped product metrics of the form* $ds^2 + h(s)^2 g_k$ *whose existence is described in Section* 2 *of [Kobayashi 1982]. The function* $h$ *satisfies* (1-6) *and* (1-7) *and* $f$ *is any constant multiple of a solution* $f_0$ *satisfying* $h' f_0' - f_0 h'' = -h/(n-1)$*.*

Note that the last equation in (10.2-v) comes from (1-4), which allows any constant multiple of one solution.

As a corollary of Theorem 10.1, we could state an extension of Theorem 1.2 in [Miao and Tam 2011] to the case of harmonic curvature. Instead we choose to state the following version, which is a twin to Corollary 1 of [Barros et al. 2015].

**Theorem 10.3.** *If* $(M^4, g, f)$ *is a simply connected, compact Miao–Tam critical metric of harmonic curvature with boundary isometric to a standard sphere* $S^3$, *then* $(M^4, g)$ *is isometric to a geodesic ball in a simply connected space form* $\mathbb{R}^4$, $\mathbb{H}^4$ *or* $\mathbb{S}^4$.

One can also make classification statements of complete spaces with harmonic curvature satisfying (1-3) or (1-4). We omit them.

Theorem 10.1 gives a speculation that it might hold in general dimension. So, we ask the following:

**Question 3.** Let $(M, g)$ be an $n$-dimensional Miao–Tam critical metric with harmonic curvature. Is it conformally flat?

It is also interesting to find examples of nonconformally flat Miao–Tam critical metrics in any dimension.

## 11. On critical point metrics

In this section we study a critical point metric, i.e., a Riemannian metric $g$ on a manifold $M$ which admits a nonzero solution $f$ to (1-5). According to [Yun et al. 2014], these critical point metrics with harmonic curvature on closed manifolds in any dimension are Einstein.

On a closed manifold, by taking the trace of this equation, $R$ must be positive and $f$ satisfies $\int_M f\, dv = 0$. Here $M$ is not necessarily closed and $g$ may have nonpositive scalar curvature. From Theorem 1.1, we can easily obtain the next theorem.

**Theorem 11.1.** *Let* $(M, g)$ *be a four-dimensional* (*not necessarily complete*) *Riemannian manifold with harmonic curvature, satisfying* (1-5) *with nonconstant* $f$. *Then one of the following holds*:

(11.1-i) $(M, g)$ *is locally isometric to a domain in one of the static spaces of Example 3 in Section 2A2, which is the Riemannian product* $(\mathbb{R}^1 \times W^3, dt^2 + ds^2 + h(s)^2 \tilde{g})$ *of* $\mathbb{R}^1$ *and a three-dimensional conformally flat static space* $(W^3, ds^2 + h(s)^2 \tilde{g})$ *with zero scalar curvature. And* $f = c \cdot h'(s) - 1$.

(11.1-ii) $(M, g)$ *is conformally flat and is locally one of the metrics whose existence is described in Section* 2 *of [Kobayashi 1982]*: $g = ds^2 + h(s)^2 g_k$ *where* $h$ *and* $f$ *satisfy* (1-6), (1-7) *and* (1-8).

*Proof.* We have $\frac{1}{3} x R + y(R) = 0$ and $R \neq 0$ in the cases (i), (ii) of Theorem 1.1. This is not compatible with (1-5). □

Complete spaces with harmonic curvature which admit a solution $f$ to (1-5) are described in the next theorem. We obtain nonconformally flat examples with zero scalar curvature in (11.2-i), which is in contrast to the above result of [Yun et al.

2014] for closed manifolds. The case (11.2-v) is also noteworthy; it is conformally flat with positive scalar curvature and the metric $g$ can exist on a compact quotient but the function $f$ can survive on the universal cover $\mathbb{R} \times_h N(1)$.

**Theorem 11.2.** *Let $(M, g)$ be a four-dimensional complete Riemannian manifold with harmonic curvature, satisfying* (1-5) *with nonconstant $f$. Then $(M, g)$ is one of the following*:

(11.2-i) *$(M, g)$ is isometric to a quotient of one of the static spaces of Example 3 in Section 2A2, which is the Riemannian product $(\mathbb{R}^1 \times W^3, dt^2 + ds^2 + h(s)^2\tilde{g})$ of $\mathbb{R}^1$ and a three-dimensional conformally flat static space $(W^3, ds^2 + h(s)^2\tilde{g})$ with zero scalar curvature which contains the spatial slice of the Schwarzschild space-time. And $f = c \cdot h'(s) - 1$ for a constant $c$.*

(11.2-ii) $\mathbb{S}^4(k^2)$ *with the metric* $g = ds^2 + (\sin^2(ks)/k^2)g_1$, *with* $f(s) = c \cdot \cos(ks)$.

(11.2-iii) $\mathbb{H}^4(-k^2)$ *with* $g = ds^2 + (\sinh(ks)^2/k^2)g_1$, *with* $f(s) = c \cdot \cosh(ks)$.

(11.2-iv) *A flat space, $f = a + \sum_i b_i x_i$ in a local Euclidean coordinate $x_i$ and constants $a, b_i$.*

(11.2-v) *Example* 3 *in [Kobayashi 1982]: a warped product $\mathbb{R} \times_h N(1)$ where $h$ is a periodic function on $\mathbb{R}$, $f$ is smooth on $\mathbb{R}$ but is not periodic. Here $R > 0$.*

(11.2-vi) *Example* 5 *in [Kobayashi 1982]: a warped product $\mathbb{R} \times_h N(k)$ where $h$ is defined on $\mathbb{R}$, $f$ is smooth on $\mathbb{R}$. Here $R \leq 0$.*

*Proof.* We may check the list in Theorem 8.2. The spaces of (8.2-i) and (8.2-ii) in Theorem 8.2 are excluded as in the proof of Theorem 11.1. The space for (8.2-iv-4) of Theorem 8.2, where $R \neq 0$, does not satisfy the equation $h'f' - fh'' = x\big(h'' + \frac{1}{3}Rh\big) + y(R)h$; when $x = 1$, $y(R) = -\frac{1}{4}R$ and $h = 1$, it reduces to $0 = \frac{1}{12}R$.

On the space of (8.2-iv-5) in Theorem 8.2, $f$ is defined and smooth on $\mathbb{R}$ by Lemma 8.1 (i). As $\frac{1}{3}xR + y(R) \neq 0$, Lemma 8.1(ii) does not apply. According to Section E.2 of [Lafontaine 1983], $f$ cannot be periodic. This yields (11.2-v). □

## References

[Barros and Ribeiro 2014] A. Barros and E. Ribeiro, Jr., "Critical point equation on four-dimensional compact manifolds", *Math. Nachr.* **287**:14-15 (2014), 1618–1623. MR Zbl

[Barros et al. 2015] A. Barros, R. Diógenes, and E. Ribeiro, Jr., "Bach-flat critical metrics of the volume functional on 4-dimensional manifolds with boundary", *J. Geom. Anal.* **25**:4 (2015), 2698–2715. MR Zbl

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **10**, Springer, 1987. MR Zbl

[Cao and Chen 2013] H.-D. Cao and Q. Chen, "On Bach-flat gradient shrinking Ricci solitons", *Duke Math. J.* **162**:6 (2013), 1149–1169. MR Zbl

[Cheeger and Colding 1996] J. Cheeger and T. H. Colding, "Lower bounds on Ricci curvature and the almost rigidity of warped products", *Ann. of Math.* (2) **144**:1 (1996), 189–237. MR Zbl

[Chruściel et al. 2005] P. T. Chruściel, J. Isenberg, and D. Pollack, "Initial data engineering", *Comm. Math. Phys.* **257**:1 (2005), 29–42. MR Zbl

[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 Zbl

[Corvino and Schoen 2006] J. Corvino and R. M. Schoen, "On the asymptotics for the vacuum Einstein constraint equations", *J. Differential Geom.* **73**:2 (2006), 185–217. MR Zbl

[Corvino et al. 2013] J. Corvino, M. Eichmair, and P. Miao, "Deformation of scalar curvature and volume", *Math. Ann.* **357**:2 (2013), 551–584. MR Zbl

[Derdziński 1980] A. Derdziński, "Classification of certain compact Riemannian manifolds with harmonic curvature and nonparallel Ricci tensor", *Math. Z.* **172**:3 (1980), 273–280. MR Zbl

[DeTurck and Goldschmidt 1989] D. DeTurck and H. Goldschmidt, "Regularity theorems in Riemannian geometry, II: Harmonic curvature and the Weyl tensor", *Forum Math.* **1**:4 (1989), 377–394. MR Zbl

[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 Zbl

[Kim 2017] J. Kim, "On a classification of 4-d gradient Ricci solitons with harmonic Weyl curvature", *J. Geom. Anal.* **27**:2 (2017), 986–1012. MR Zbl

[Kobayashi 1982] O. Kobayashi, "A differential equation arising from scalar curvature function", *J. Math. Soc. Japan* **34**:4 (1982), 665–675. MR Zbl

[Lafontaine 1983] J. Lafontaine, "Sur la géométrie d'une généralisation de l'équation différentielle d'Obata", *J. Math. Pures Appl.* (9) **62**:1 (1983), 63–72. MR Zbl

[Miao and Tam 2009] P. Miao and L.-F. Tam, "On the volume functional of compact manifolds with boundary with constant scalar curvature", *Calc. Var. Partial Differential Equations* **36**:2 (2009), 141–171. MR Zbl

[Miao and Tam 2011] P. Miao and L.-F. Tam, "Einstein and conformally flat critical metrics of the volume functional", *Trans. Amer. Math. Soc.* **363**:6 (2011), 2907–2937. MR Zbl

[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 Zbl

[Qing and Yuan 2013] J. Qing and W. Yuan, "A note on static spaces and related problems", *J. Geom. Phys.* **74** (2013), 18–27. MR Zbl

[Qing and Yuan 2016] J. Qing and W. Yuan, "On scalar curvature rigidity of vacuum static spaces", *Math. Ann.* **365**:3-4 (2016), 1257–1277. MR Zbl

[Yano and Nagano 1959] K. Yano and T. Nagano, "Einstein spaces admitting a one-parameter group of conformal transformations", *Ann. of Math.* (2) **69** (1959), 451–461. MR Zbl

[Yun et al. 2014] G. Yun, J. Chang, and S. Hwang, "Total scalar curvature and harmonic curvature", *Taiwanese J. Math.* **18**:5 (2014), 1439–1458. MR Zbl

Received December 18, 2016. Revised November 2, 2017.

JONGSU KIM
DEPARTMENT OF MATHEMATICS
SOGANG UNIVERSITY
SEOUL
SOUTH KOREA
jskim@sogang.ac.kr

Jinwoo Shin
Department of Mathematics
Sogang University
Seoul
South Korea
shinjin@sogang.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.463

# BOUNDARY SCHWARZ LEMMA FOR NONEQUIDIMENSIONAL HOLOMORPHIC MAPPINGS AND ITS APPLICATION

YANG LIU, ZHIHUA CHEN AND YIFEI PAN

**In this paper, we give a boundary Schwarz lemma for holomorphic mappings between nonequidimensional unit balls. As an application, a new boundary rigidity result is presented.**

## 1. Introduction

Let $B^n$ be the unit ball in $\mathbb{C}^n$ for $n \geq 1$. Denote by $\mathrm{Hol}(B^n, B^N)$ the set of all holomorphic mapping from the unit ball $B^n \subset \mathbb{C}^n$ into $B^N \subset \mathbb{C}^N$. For a bounded domain $V \subset \mathbb{C}^n$, let $C^{1+\alpha}(V)$ be the set of all functions $f$ on $V$ whose first order partial derivatives exist and are Hölder continuous. For $z_0 \in \partial B^n$, the tangent space $T_{z_0}(\partial B^n)$ and holomorphic tangent space $T^{1,0}_{z_0}(\partial B^n)$ at $z_0$ are defined by

$$T_{z_0}(\partial B^n) = \{\beta \in \mathbb{C}^n \mid \mathrm{Re}(\overline{z_0}^T \beta) = 0\}, \quad T^{(1,0)}_{z_0}(\partial B^n) = \{\beta \in \mathbb{C}^n \mid \overline{z_0}^T \beta = 0\},$$

respectively. In this paper, we give a general boundary Schwarz lemma for holomorphic mappings between unit balls in any dimensions as follows.

**Theorem 1.1.** *Let $f \in \mathrm{Hol}(B^n, B^N)$ for any $n, N \geq 1$, and denote by $J_f(z)$ the Jacobian matrix of $f$ at $z$. If $f$ is $C^{1+\alpha}$ at $z_0 \in \partial B^n$ and $f(z_0) = w_0 \in \partial B^N$, then we have*:

(I) *$J_f(z_0)\beta \in T_{w_0}(\partial B^N)$ for any $\beta \in T_{z_0}(\partial B^n)$, and $J_f(z_0)\beta \in T^{(1,0)}_{w_0}(\partial B^N)$ for any $\beta \in T^{(1,0)}_{z_0}(\partial B^n)$.*

(II) *There exists $\lambda \in \mathbb{R}$ such that*

$$\overline{J_f(z_0)}^T w_0 = \lambda z_0$$

*with $\lambda \geq |1 - \bar{a}^T w_0|^2/(1 - \|a\|^2) > 0$, where $a = f(0)$.*

**Remark 1.2.** For the case of biholomorphic mapping, item (I) holds; see Chapter 3 of [Krantz 1992]. Here we conclude the same result for holomorphic mappings between unit balls of different dimensions. For $n = N = 1$, the theorem says

*MSC2010:* primary 32H02; secondary 30C80.
*Keywords:* Boundary Schwarz lemma, boundary rigidity, holomorphic mapping, unit ball.

$f'(z_0) > 0$, so the image $f(\partial B^1)$ at $w_0$ is always smooth. For $n > 1$, if $f(\partial B^n)$ is a smooth manifold, then conclusion (I) is almost trivial. However, we would like to point out that $f(\partial B^n)$ may be not a smooth manifold.

In the special case when $n = N$, Theorem 1.1 reduces to (1) and (2) in Theorem 3.1 of [Liu et al. 2015]. For $n = N = 1$, part (II) of the theorem gives the classical boundary Schwarz lemma in [Garnett 1981].

As an application of Theorem 1.1, we will present a new boundary rigidity result. First, recall the following famous rigidity result for holomorphic self-mappings on $B^n$.

**Theorem 1.3** [Burns and Krantz 1994]. *Let $f \in \mathrm{Hol}(B^n, B^n)$ with $n \geq 1$ such that*

$$f(z) = z + O(|z - \mathbf{1}|^4)$$

*as $z \to \mathbf{1}$, where $\mathbf{1} = (1, 0, \ldots, 0)^T \in \partial B^n$. Then $f(z) \equiv z$.*

Notice that the order of the estimation $O(|z - \mathbf{1}|^4)$ is sharp in Theorem 1.3, as shown by the example [Burns and Krantz 1994]

$$f(z) = z - \tfrac{1}{10}(z - 1)^3, \quad z \in D,$$

where $D$ is the unit disk.

On the other hand, Huang [1995] shows that if $f \in \mathrm{Hol}(B^n, B^n)$ satisfies $f(z) = z + O(|z - \mathbf{1}|^3)$ as $z \to \mathbf{1}$, and $f(z_0) = z_0$ with $z_0 \in B^n$, then $f(z) = z$ on the unit ball. This result gives a condition under which the order of the estimation $O(|z - \mathbf{1}|^4)$ in [Burns and Krantz 1994] can be lower with a fixed point.

A problem of the boundary rigidity for nonequidimensional mappings was given by Krantz [2011]. Using Theorem 1.1, we give a positive answer to this problem, and provide a new boundary rigidity result for holomorphic mappings between nonequidimensional unit balls. In fact, we find conditions under which the order of the estimation can be lower and is also sharp without internal fixed point. Our result is given as follows.

**Theorem 1.4.** *Let $f \in \mathrm{Hol}(B^n, B^N)$ for $N \geq n \geq 1$, such that*

$$f(z) = (z^T, 0)^T + O(|z - \mathbf{1}|^3) \tag{1-1}$$

*as $z \to \mathbf{1}$. If $f$ is $C^2$ at $\mathbf{1}$ and $f_1(z) = z_1$, where $f_1$ is the first component of $f$, then $f(z) \equiv (z^T, 0)^T$.*

**Example.** Let $f(z_1, z_2) = (z_1, z_2 z_1^k, 0)^T \in \mathrm{Hol}(B^2, B^3)$ for integer $k \geq 1$. Since $f(z) - (z_1, z_2, 0)^T = (0, z_2(z_1^k - 1), 0)^T$, and

$$\frac{|f(z) - (z_1, z_2, 0)|}{|z - \mathbf{1}|^2} = \frac{|z_2(z_1^k - 1)|}{|z_1 - 1|^2 + |z_2|^2} \leq \frac{1}{2}\frac{|z_1^k - 1|^2 + |z_2|^2}{|z_1 - 1|^2 + |z_2|^2} \leq \tfrac{1}{2}(k^2 + 1),$$

it satisfies $f(z) = (z_1, z_2, 0)^T + O(|z - \mathbf{1}|^2)$. However, it is obvious that $f(z) \neq (z_1, z_2, 0)^T$, which indicates that the order of $O(|z - \mathbf{1}|^3)$ is sharp.

## 2. Proof of Theorem 1.1

To prove the main result, we first give some notation and lemmas. For any $z = (z_1, \ldots, z_n)^T$, $w = (w_1, \ldots, w_n)^T \in \mathbb{C}^n$, the inner product and the corresponding norm are given by $\langle z, w\rangle = \sum_{j=1}^n z_j \overline{w_j}$ and $\|z\| = \langle z, z\rangle^{\frac{1}{2}}$ respectively. $\partial B^n$ denotes the boundary of $B^n$.

**Lemma 2.1** [Rudin 1980]. *Let $f \in \mathrm{Hol}(B^n, B^N)$ with $n, N \geq 1$. If $f(0) = 0$, then $\|f(z)\| \leq \|z\|$, $z \in B^n$.*

**Lemma 2.2** [Dai et al. 2010; Liu et al. 2016]. *For given $p \in B^n \cup \partial B^n$ and $q \in \mathbb{C}^n$ with $q \neq 0$, let $L(\xi) = p + \xi q$ for $\xi \in \mathbb{C}$. Then*

$$L(D_{p,q}) \subset B^n, \quad L(\partial D_{p,q}) \subset \partial B^n,$$

*where $D_{p,q} = \{\xi \in \mathbb{C} \mid |\xi - c_{p,q}| < r_{p,q}\}$, with*

$$c_{p,q} = -\frac{\langle p, q\rangle}{\|q\|^2}, \quad r_{p,q} = \sqrt{\frac{1 - \|p\|^2}{\|q\|^2} + \left|\frac{\langle p, q\rangle}{\|q\|^2}\right|^2}.$$

*Proof.* Assume $\|L(D_{p,q})\|^2 < 1$, which means

$$\|p\|^2 + 2\,\mathrm{Re}(\bar{p}^T \xi q) + \|\xi q\|^2 < 1,$$

and

$$\frac{\|p\|^2}{\|q\|^2} + 2\frac{\mathrm{Re}(\bar{p}^T q\xi)}{\|q\|^2} + |\xi|^2 < \frac{1}{\|q\|^2},$$

i.e.,

$$\left|\xi + \frac{\langle p, q\rangle}{\|q\|^2}\right|^2 < \frac{1 - \|p\|^2}{\|q\|^2} + \left|\frac{\langle p, q\rangle}{\|q\|^2}\right|^2. \qquad \square$$

*Proof of Theorem 1.1.* We prove the theorem in five steps.

<u>Step 1</u>. Denote by $e_i^n$ the $i$-th column of the $n \times n$ identity matrix. Assume $z_0 = e_1^n = \mathbf{1} \in \partial B^n$, and $f$ is $C^{1+\alpha}$ in a neighborhood $V$ of $z_0$. Moreover, assume $f(0) = 0$ and $f(z_0) = w_0 = e_1^N$.

We first show that for any $q \in H = \{z \in \mathbb{C}^n \mid \mathrm{Re}\, z_1 < 0\}$, there exists a $r_q > 0$ such that

$$\mathbf{1} + tq \in B^n, \quad 0 < t < r_q. \tag{2-1}$$

Assume $q = (q_1, \ldots, q_n)^T \in H$ and $\mathrm{Re}\, q_1 < 0$. Then for $t \in \mathbb{R}$,

$$\mathbf{1} + tq \in B^n \iff \|\mathbf{1} + tq\|^2 < 1 \iff |1 + t\,\mathrm{Re}\, q_1|^2 + |t\,\mathrm{Im}\, q_1|^2 + \sum_{j=2}^n |q_j|^2 t^2 < 1,$$

which is equivalent to

$$0 < t < \frac{-2\operatorname{Re} q_1}{\sum_{j=1}^n |q_j|^2}.$$

Letting $r_q = -2\operatorname{Re} q_1/(\sum_{j=1}^n |q_j|^2)$ implies the claim.

Let $p = z_0,\ q = (-1+ik)z_0$ for any given $k \in \mathbb{R}$. Then from (2-1), when $t \to 0^+$, $p + tq \in B^n \cap V$. For such $t$, taking the Taylor expansion of $f((1-t+ikt)z_0)$ at $t = 0$, we have

$$f((1-t+ikt)z_0) = w_0 + J_f(z_0)(-1+ik)z_0 t + O(t^{1+\alpha}).$$

By Lemma 2.1,

$$\|f((1-t+ikt)z_0)\|^2 = \|w_0 + J_f(z_0)(-1+ik)z_0 t + O(t^{1+\alpha})\|^2 \le \|(1-t+ikt)z_0\|^2,$$

i.e.,

$$1 + 2\operatorname{Re}\big(\overline{w_0}^T J_f(z_0)(-1+ik)z_0 t\big) + O(t^{1+\alpha}) \le 1 - 2t + O(t^2).$$

Substituting $w_0 = e_1^N,\ z_0 = e_1^n$ and let $t \to 0^+$, we have

$$\operatorname{Re}\big(\overline{e_1^N}^T J_f(z_0)(-1+ik)e_1^n\big) \le -1,$$

i.e.,

$$\operatorname{Re}\bigg(\frac{\partial f_1(z_0)}{\partial z_1}(-1+ik)\bigg) \le -1.$$

Let $\partial f_1(z_0)/\partial z_1 = \operatorname{Re}(\partial f_1(z_0)/\partial z_1) + i\operatorname{Im}(\partial f_1(z_0)/\partial z_1)$. Then from the above inequality, one gets

$$-\operatorname{Re}\frac{\partial f_1(z_0)}{\partial z_1} - k\operatorname{Im}\frac{\partial f_1(z_0)}{\partial z_1} \le -1,$$

i.e.,

$$-k\operatorname{Im}\frac{\partial f_1(z_0)}{\partial z_1} \le \operatorname{Re}\frac{\partial f_1(z_0)}{\partial z_1} - 1. \tag{2-2}$$

Since (2-2) is valid for any $k \in \mathbb{R}$, we have

$$\operatorname{Im}\frac{\partial f_1(z_0)}{\partial z_1} = 0,$$

which implies

$$0 \le \operatorname{Re}\frac{\partial f_1(z_0)}{\partial z_1} - 1,$$

and

$$\frac{\partial f_1(z_0)}{\partial z_1} = \operatorname{Re}\frac{\partial f_1(z_0)}{\partial z_1} \ge 1. \tag{2-3}$$

Step 2. Let $p = z_0,\ q = -z_0 + ike_j^n$ for $2 \le j \le n$ and $k \in \mathbb{R}$. Then as $t \to 0^+$, $p + tq \in B^n \cap V$. Similarly, taking the Taylor expansion of $f\big((1-t)z_0 + ikte_j^n\big)$ at $t = 0$, we have

$$f\big((1-t)z_0 + ikte_j^n\big) = w_0 + J_f(z_0)(-z_0 + ike_j^n)t + O(t^{1+\alpha}).$$

By Lemma 2.1,

$$\begin{aligned}\big\| f\big((1-t)z_0 + ikte_j^n\big)\big\|^2 &= \|w_0 + J_f(z_0)(-z_0 + ike_j^n)t + O(t^{1+\alpha})\|^2 \\ &\le \|(1-t)z_0 + ikte_j^n\|^2,\end{aligned}$$

i.e.,

$$1 + 2\,\mathrm{Re}\big(\overline{w_0}^T J_f(z_0)(-z_0 + ike_j^n)t\big) + O(t^{1+\alpha}) \le 1 - 2t + O(t^2).$$

Substituting $w_0 = e_1^N$, $z_0 = e_1^n$ and letting $t \to 0^+$, we have

$$\mathrm{Re}\big(\overline{e_1^N}^T J_f(z_0)(-e_1^n + ike_j^n)\big) \le -1,$$

i.e.,

$$\mathrm{Re}\bigg(-\frac{\partial f_1(z_0)}{\partial z_1} + ik\frac{\partial f_1(z_0)}{\partial z_j}\bigg) \le -1.$$

From the above inequality as well as inequality (2-3), one has

$$-k\,\mathrm{Im}\,\frac{\partial f_1(z_0)}{\partial z_j} \le \frac{\partial f_1(z_0)}{\partial z_1} - 1.$$

With an argument similar to Step 1, we have

$$\mathrm{Im}\,\frac{\partial f_1(z_0)}{\partial z_j} = 0, \quad 2 \le j \le n.$$

Meanwhile, if we assume $p = z_0,\ q = -z_0 + ke_j^n$ for $2 \le j \le n$ and any $k \in \mathbb{R}$. It is easy to find

$$\mathrm{Re}\,\frac{\partial f_1(z_0)}{\partial z_j} = 0, \quad 2 \le j \le n.$$

Therefore,

$$\frac{\partial f_1(z_0)}{\partial z_j} = 0, \quad 2 \le j \le n, \tag{2-4}$$

as well. As a result of (2-3) and (2-4), we have

$$\overline{J_f(z_0)}^T w_0 = \lambda_f z_0 \tag{2-5}$$

for $w_0 = e_1^N,\ z_0 = e_1^n$ and $\lambda_f = \partial f_1(z_0)/\partial z_1 \ge 1$.

Step 3. Now let $z_0$ be any given point at $\partial B^n$. Then there exists a unitary matrix $U_{z_0}$ such that $U_{z_0}(z_0) = e_1^n$. Assume $f(0) = 0$, $f(z_0) = w_0$ and $w_0$ is not necessarily $e_1^N$ at $\partial B^N$. Similarly, there is a unitary matrix $U_{w_0}$ such that $U_{w_0}(w_0) = e_1^N$. Let

$$g(z) = U_{w_0} \circ f \circ \overline{U_{z_0}}^T;$$

then $g(0) = 0$, $g(e_1^n) = e_1^N$. Moreover,

$$J_g(z) = U_{w_0} J_f(\overline{U_{z_0}}^T z)\overline{U_{z_0}}^T. \tag{2-6}$$

From Steps 1 and 2, we have

$$\overline{J_g(e_1^n)}^T e_1^N = \lambda_g e_1^n$$

for $z_0 = e_1^n$ and $\lambda_g = \partial g_1(e_1^n)/\partial z_1 \geq 1$, which implies

$$\overline{U_{w_0} J_f(\overline{U_{z_0}}^T e_1^n)\overline{U_{z_0}}^T}^T e_1^N = \lambda_g e_1^n,$$

i.e.,

$$U_{z_0} \overline{J_f(z_0)}^T \overline{U_{w_0}}^T e_1^N = \lambda_g e_1^n.$$

After multiplying by $\overline{U_{z_0}}^T$ on both sides of the above equation, we obtain

$$\overline{U_{z_0}}^T U_{z_0} \overline{J_f(z_0)}^T \overline{U_{w_0}}^T e_1^N = \lambda_g \overline{U_{z_0}}^T e_1^n,$$

i.e.,

$$\overline{J_f(z_0)}^T w_0 = \lambda_g z_0, \tag{2-7}$$

where $\lambda_g = \partial g_1(e_1^n)/\partial z_1 \geq 1$.

Step 4. Let $f(z_0) = w_0$ with $z_0 \in \partial B^n$, $w_0 \in \partial B^N$. If $f(0) = a \neq 0$, then we use the automorphism of $B^N$ to get the result. Assume $\phi_a(w)$ is an automorphism of $B^N$ such that $\phi_a(a) = 0$. Then $\phi_a(w_0) \in \partial B^N$ as well. With a similar analysis to Step 3, there exists a $U_{\phi_a(w_0)}$ such that $U_{\phi_a}(\phi_a(w_0)) = w_0$. Let

$$h = U_{\phi_a} \circ \phi_a \circ f,$$

then $h(0) = 0$, $h(z_0) = w_0$. As a result of Step 3, there is a real number $\gamma \geq 1$ such that

$$\overline{J_h(z_0)}^T w_0 = \gamma z_0.$$

Using the expression for $h$, we obtain

$$\overline{J_h(z_0)}^T w_0 = \overline{U_{\phi_a} J_{\phi_a}(w_0) J_f(z_0)}^T w_0 = \overline{J_f(z_0)}^T \overline{J_{\phi_a}(w_0)}^T \overline{U_{\phi_a}}^T w_0 = \gamma z_0. \tag{2-8}$$

Since $U_{\phi_a}(\phi_a(w_0)) = w_0$, we have $\overline{U_{\phi_a}}^T w_0 = \phi_a(w_0)$. From the expression for the automorphism $\phi_a$ given by [Rudin 1980], we have the following equality:

$$\overline{J_{\phi_a}(w_0)}^T \overline{U_{\phi_a}}^T w_0 = \overline{J_{\phi_a}(w_0)}^T \phi_a(w_0) = \frac{1 - \|a\|^2}{|1 - \bar{a}^T w_0|^2} w_0.$$

Therefore, combining with (2-8) we get

$$\overline{J_f(z_0)}^T \frac{1-\|a\|^2}{|1-\bar{a}^T w_0|^2} w_0 = \gamma z_0.$$

Consequently,

$$\overline{J_f(z_0)}^T w_0 = \lambda z_0, \tag{2-9}$$

where

$$\lambda = \frac{|1-\bar{a}^T w_0|^2}{1-\|a\|^2}\gamma \geq \frac{|1-\bar{a}^T w_0|^2}{1-\|a\|^2} > 0 \quad \text{and} \quad a = f(0).$$

The proof of (II) is completed.

Step 5. For any $\beta \in T_{z_0}(\partial B^n)$, we have

$$\operatorname{Re}(\overline{z_0}^T \beta) = 0. \tag{2-10}$$

To prove $J_f(z_0)\beta \in T_{w_0}(\partial B^N)$, it is sufficient to verify

$$\operatorname{Re}\big(\overline{w_0}^T J_f(z_0)\beta\big) = 0. \tag{2-11}$$

From (2-9), $\overline{J_f(z_0)}^T w_0 = \lambda z_0$, which means

$$\overline{w_0}^T J_f(z_0) = \overline{\overline{J_f(z_0)}^T w_0}^T = \lambda \overline{z_0}^T. \tag{2-12}$$

Then

$$\operatorname{Re}\big(\overline{w_0}^T J_f(z_0)\beta\big) = \operatorname{Re}(\lambda \overline{z_0}^T \beta) = \lambda \operatorname{Re}(\overline{z_0}^T \beta) = 0,$$

where the last equality comes from (2-10). Therefore, (2-11) is proved and hence

$$J_f(z_0)\beta \in T_{w_0}(\partial B^N).$$

On the other hand, for any $\beta \in T_{z_0}^{(1,0)}(\partial B^n)$, we have

$$\overline{z_0}^T \beta = 0. \tag{2-13}$$

To prove $J_f^{(1,0)}(z_0)\beta \in T_{w_0}^{(1,0)}(\partial B^N)$, it is sufficient to get

$$\overline{w_0}^T J_f(z_0)\beta = 0.$$

From (2-12) and (2-13),

$$\overline{w_0}^T J_f(z_0)\beta = \lambda \overline{z_0}^T \beta = \lambda \overline{z_0}^T \beta = 0,$$

Therefore, $J_f(z_0)\beta \in T_{w_0}^{(1,0)}(\partial B^N)$. The proof of (I) is completed. □

## 3. Proof of Theorem 1.4

For any fixed point $b \in B^n$, let $\mathcal{L}_b$ be the complex (straight) line joining $b$ and $\mathbf{1}$:

$$\mathcal{L}_b = \{z \in \mathbb{C}^n \mid z = \mathbf{1} + \xi(\mathbf{1} - b),\ \forall \xi \in \mathbb{C}\},$$

and let $d_b$ be the complex disc given by $\mathcal{L}_b \cap B^n$. In particular,

$$d_0 = \{z \in B^n \mid z_2 = \cdots = z_n = 0\}.$$

From Lemma 2.2, it is found that $d_b = L(D_{\mathbf{1},\mathbf{1}-b})$.

**Lemma 3.1.** *Let $f = (f_1, \dots, f_N)^T \in \mathrm{Hol}(B^n, B^N)$ with $N \geq n \geq 1$, and $f_1(z) = z_1$, $z \in B^n$. Then*

$$f(z_1, 0, \dots, 0) = (z_1, 0, \dots, 0)^T, \quad z \in d_0.$$

*Proof.* Restricting $f(z) = (z_1, f_2, \dots, f_N)^T$ on $d_0$, then $f|_{d_0}$ can be regarded as a holomorphic mapping from $D$ into $B^N$, which implies $|z_1|^2 + \sum_{j=2}^N |f_j(z)|^2 < 1$, $z \in d_0$ and then $\sum_{j=2}^N |f_j(z)|^2 < 1 - |z_1|^2$, $z \in d_0$. By $z_1 \to 1$, the maximum principle of subharmonic function guarantees $f_j|_{d_0} \equiv 0$ for any $2 \leq j \leq N$. Therefore, $f|_{d_0} = (z_1, 0, \dots, 0)^T$. □

*Proof of Theorem 1.4.* Step 1. Given $f = (f_1, \dots, f_N)^T \in \mathrm{Hol}(B^n, B^N)$ such that (1-1) holds and $f_1(z) \equiv z_1$ on $B^n$. From Lemma 3.1, one gets $f|_{d_0} = (z_1, 0, \dots, 0)^T$. We aim to prove $f_j(z) = z_j$ for $2 \leq j \leq n$ and $f_j(z) = 0$ for $n+1 \leq j \leq N$ on the unit ball.

Represent $f_j$ by

$$f_j(z) = \sum_{k=2}^{n} \phi_{jk}(z) z_k, \qquad z \in B^n, \quad 2 \leq j \leq N, \tag{3-1}$$

where $\phi_{jk}(z)$ are all holomorphic functions on the unit ball. In fact, taking the Taylor expansion for $f_j(z)$ at 0 for $2 \leq j \leq N$, one gets

$$f_j(z) = f_j(0) + \sum_{k=1}^{\infty} \sum_{|v|=k} C_v z^v, \quad z \in B^n.$$

Let $\phi_{j1}(z_1) = \sum_{i=1}^{\infty} C_i z_1^i$. Then there are holomorphic functions $\phi_{jk}(z)$ satisfying

$$f_j(z) = f_j(0) + \sum_{k=1}^{\infty} \sum_{|v|=k} C_v z^v = f_j(0) + \phi_{j1}(z_1) + \sum_{k=2}^{n} \phi_{jk}(z) z_k, \quad z \in B^n.$$

We notice that, for $2 \leq k \leq n$, the $\phi_{jk}(z)$ are not necessarily unique in this expression for $f_j(z)$. Since $f_j(z_1, 0, \dots, 0) = 0$ for any $(z_1, 0, \dots, 0)^T \in B^n \cup \{\mathbf{1}\}$, we have $f_j(0) = 0$ and $\phi_{j1}(z_1) \equiv 0$, $z \in B^n \cup \{\mathbf{1}\}$, so that (3-1) holds.

In particular, if

$$\phi_{jk}(z) \equiv \delta_{jk}, \qquad 2 \le j \le N, \quad 2 \le k \le n, \tag{3-2}$$

then the theorem is proved. If not, due to $f(z) \in B^N$,

$$|z_1|^2 + \sum_{j=2}^{N} \left| \sum_{k=2}^{n} \phi_{jk}(z) z_k \right|^2 < 1, \quad z \in B^n. \tag{3-3}$$

Given a $b \in B^n$ with $\tilde{b} = (b_2, \dots, b_n)^T \neq 0$, there at least exists one $b_j \neq 0$ for $2 \le j \le n$; without loss of generality, let $b_2 \neq 0$. We consider $d_b = L(D_{\mathbf{1},\mathbf{1}-b})$ from Lemma 2.2, where the expression for $D_{\mathbf{1},\mathbf{1}-b}$ can be given by

$$D_{\mathbf{1},\mathbf{1}-b} = \left\{ \xi \in \mathbb{C} \;\middle|\; \left| \xi + \frac{1-\bar{b}_1}{\|\mathbf{1}-b\|^2} \right| < \frac{|1-b_1|}{\|\mathbf{1}-b\|^2} \right\}. \tag{3-4}$$

Notice that $\xi = 0 \in \partial D_{\mathbf{1},\mathbf{1}-b}$ and $z = \mathbf{1} \in \partial d_b$. Furthermore, for any $z \in d_b$, $z = L(\xi) = \mathbf{1} + \xi(\mathbf{1}-b) \in d_b$, $\xi \in D_{\mathbf{1},\mathbf{1}-b}$, i.e.,

$$(z_1, z_2, \dots, z_n)^T = (1 + \xi(1-b_1), -\xi b_2, \dots, -\xi b_n)^T, \quad \xi \in D_{\mathbf{1},\mathbf{1}-b},$$

which gives that for $z \in d_b \cup \partial d_b$, the following inequality holds:

$$\frac{1-|z_1|^2}{|z_2|^2} \ge \sum_{j=2}^{n} \frac{|z_j|^2}{|z_2|^2} = \sum_{j=2}^{n} \left| \frac{b_j}{b_2} \right|^2. \tag{3-5}$$

The equality is available only for $z \in \partial d_b$ and $z \neq \mathbf{1}$, i.e., $z_2 \neq 0$ ($\xi \neq 0$).

Step 2. Since (1-1) holds as $z \to \mathbf{1}$, it follows that

$$f(z) - (z_1, \dots, z_n, 0, \dots, 0)^T = O(|z-\mathbf{1}|^3).$$

Restricting $z \in d_b$, we obtain

$$\begin{aligned} &f(z) - (z_1, \dots, z_n, 0, \dots, 0)^T|_{z \in d_b} \\ &= \Bigg(0, \sum_{k=2}^{n} \phi_{2k}(z) z_k - z_2, \dots, \sum_{k=2}^{n} \phi_{nk}(z) z_k - z_n, \\ &\qquad \sum_{k=2}^{n} \phi_{(n+1)k}(z) z_k, \dots, \sum_{k=2}^{n} \phi_{Nk}(z) z_k \Bigg)^T \\ &= \Bigg(0, \left(\sum_{k=2}^{n} \phi_{2k}(z) \frac{b_k}{b_2} - \frac{b_2}{b_2}\right) z_2, \dots, \left(\sum_{k=2}^{n} \phi_{nk}(z) \frac{b_k}{b_2} - \frac{b_n}{b_2}\right) z_2, \\ &\qquad \left(\sum_{k=2}^{n} \phi_{(n+1)k}(z) \frac{b_k}{b_2}\right) z_2, \dots, \left(\sum_{k=2}^{n} \phi_{Nk}(z) \frac{b_k}{b_2}\right) z_2 \Bigg)^T, \end{aligned} \tag{3-6a}$$

and

$$O(|z-\mathbf{1}|^3)|_{z\in d_b} = O\left(\left(\left|\frac{1-b_1}{b_2}\right|^2 + \sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2\right)^{\frac{3}{2}}|z_2|^3\right) = O(|z_2|^3). \tag{3-6b}$$

Setting

$$\begin{aligned}\Gamma(z) &= (\Gamma_2(z),\dots,\Gamma_N(z))^T\\ &\triangleq \left(\sum_{k=2}^{n}\phi_{2k}(z)\frac{b_k}{b_2},\dots,\sum_{k=2}^{n}\phi_{nk}(z)\frac{b_k}{b_2},\sum_{k=2}^{n}\phi_{(n+1)k}(z)\frac{b_k}{b_2},\dots,\sum_{k=2}^{n}\phi_{Nk}(z)\frac{b_k}{b_2}\right)^T,\end{aligned}$$

we have from (3-6a) and (3-6b),

$$\Gamma(z) - \left(\frac{b_2}{b_2},\dots,\frac{b_n}{b_2},0\dots,0\right)^T = O(|z_2|^2), \quad z\in d_b. \tag{3-7}$$

Letting $z\to\mathbf{1}\in\partial d_b$, gives $z_2\to 0$ and hence (3-7) yields the following equalities:

$$\begin{aligned}\sum_{k=2}^{n}\phi_{jk}(\mathbf{1})\frac{b_k}{b_2} - \frac{b_j}{b_2} &= 0, \quad 2\le j\le n,\\ \sum_{k=2}^{n}\phi_{jk}(\mathbf{1})\frac{b_k}{b_2} &= 0, \quad n+1\le j\le N.\end{aligned} \tag{3-8}$$

We consider the first order derivative of (3-7) at $\mathbf{1}$ and obtain

$$\sum_{k=2}^{n}\phi'_{jk}(\mathbf{1})\frac{b_k}{b_2} = 0, \quad 2\le j\le N. \tag{3-9}$$

We now set

$$A_0 = (\phi_{ij}(\mathbf{1}))_{(N-1)\times(n-1)}, \quad A_1 = (\phi'_{ij}(\mathbf{1}))_{(N-1)\times(n-1)},$$

so (3-8) and (3-9) are equivalent to

$$A_0\tilde{b} = (\tilde{b},0,\dots,0)^T, \quad A_1\tilde{b} = 0, \tag{3-10}$$

where $\tilde{b} = (b_2,\dots,b_n)^T$. Since (3-10) is valid for any $\tilde{b}\neq 0$, we have $A_0 = (I_{n-1},0)^T$ and $A_1 = 0$, which implies that

$$\phi_{ij}(\mathbf{1}) = \delta_{ij}, \quad \phi'_{ij}(\mathbf{1}) = 0, \qquad 2\le i\le N, \quad 2\le j\le n, \tag{3-11}$$

Step 3. Restricting $f$ on $d_b$, from (3-3), we have

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)z_k\right|^2 = \sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)\frac{b_k}{b_2}\right|^2|z_2|^2 < 1-|z_1|^2, \quad z\in d_b.$$

Then

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)\frac{b_k}{b_2}\right|^2<\frac{1-|z_1|^2}{|z_2|^2},\quad z\in d_b.$$

From (3-5),

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)\frac{b_k}{b_2}\right|^2\le\frac{1-|z_1|^2}{|z_2|^2}=\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2,\qquad z\in\partial d_b,\quad z\neq\mathbf{1}.\tag{3-12}$$

For $z=\mathbf{1}$, i.e., $z_2=0$, it follows from (3-11) that

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(\mathbf{1})\frac{b_k}{b_2}\right|^2=\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2.\tag{3-13}$$

Combining (3-12) and (3-13), we have

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)\frac{b_k}{b_2}\right|^2\le\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2,\quad z\in\partial d_b.\tag{3-14}$$

Since $d_b=L(D_{\mathbf{1},\mathbf{1}-b})$, (3-14) is equivalent to

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(L(\xi))\frac{b_k}{b_2}\right|^2\le\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2,\quad \xi\in\partial D_{\mathbf{1},\mathbf{1}-b}.\tag{3-15}$$

Considering the maximum principle for the subharmonic function

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(L(\xi))\frac{b_k}{b_2}\right|^2$$

on $D_{\mathbf{1},\mathbf{1}-b}$, we have

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(L(\xi))\frac{b_k}{b_2}\right|^2\le\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2,\quad \xi\in D_{\mathbf{1},\mathbf{1}-b},$$

which means that

$$\sum_{j=2}^{N}\left|\sum_{k=2}^{n}\phi_{jk}(z)\frac{b_k}{b_2}\right|^2\le\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2,\quad z\in d_b.\tag{3-16}$$

Step 4. Consider the mapping $\Gamma(z)$ on $d_b$, which is a holomorphic mapping from $d_b$ to the closure of the ball in $\mathbb{C}^{n-1}$ with the center 0 and radius $\left(\sum_{j=2}^{n}|b_j/b_2|^2\right)^{\frac{1}{2}}$ from (3-16). From the expression of $D_{\mathbf{1},\mathbf{1}-b}$ given by (3-4), let

$$\eta_1(\xi)=\frac{\xi+(1-\bar{b}_1)/\|\mathbf{1}-b\|^2}{|1-b_1|/\|\mathbf{1}-b\|^2}:\overline{D}_{\mathbf{1},\mathbf{1}-b}\to\overline{D},$$

and

$$\eta_2(\xi) = \frac{|1-b_1|}{1-\bar{b}_1}\xi : \bar{D} \to \bar{D},$$

where $\bar{D}_{\mathbf{1},\mathbf{1}-b}$ and $\bar{D}$ denote the closures of $D_{\mathbf{1},\mathbf{1}-b}$ and $D$, respectively. Constructing a mapping

$$\Psi(\xi) = \left(\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2\right)^{-\frac{1}{2}} \cdot \Gamma \circ \eta_1^{-1} \circ \eta_2^{-1} : D \to \bar{B}^{N-1},$$

we have from (3-11) that

$$\Psi(1) = \left(\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2\right)^{-\frac{1}{2}} \cdot \left(\frac{b_2}{b_2}, \dots, \frac{b_n}{b_2}, 0, \dots, 0\right)^T \in \partial B^{N-1}.$$

Moreover, the mapping $f$ is holomorphic on $B^n$ and satisfies (1-1) as $z \to \mathbf{1}$; from the construction, $\Psi$ is holomorphic on $D$ and $C^2$ at 1. In addition $\Psi(1) = w_0 \in \partial B^{N-1}$. According to Theorem 1.1, there exists a $\lambda > 0$ such that

$$\overline{J_\Psi(1)}^T w_0 = \lambda \cdot 1 > 0$$

unless $\Psi$ is a constant mapping. In other words, the above inequality means that

$$\left(\sum_{j=2}^{n}\left|\frac{b_j}{b_2}\right|^2\right)^{-1} \cdot \frac{|1-b_1|}{\|\mathbf{1}-b\|^2} \cdot \frac{\overline{1-\bar{b}_1}}{|1-b_1|} \cdot \overline{\Gamma'(\mathbf{1})} \cdot \left(\frac{b_2}{b_2}, \dots, \frac{b_n}{b_2}\right)^T > 0.$$

However, from (3-11), it is found that $\Gamma'(\mathbf{1}) = 0$, which is a contradiction and forces $\Psi$ to be a constant mapping such that $\Gamma$ satisfies (3-11), i.e.,

$$\phi_{ij}(z) = \phi_{ij}(\mathbf{1}) \equiv \delta_{ij}, \qquad 2 \le i \le N, \quad 2 \le j \le n.$$

Consequently, from the expression for $f_j(z)$ in (3-1), one gets $f_j(z) = z_j$ for $2 \le j \le n$ and $f_j(z) = 0$ for $n+1 \le j \le N$. Therefore, we have $f(z) \equiv (z^T, 0)^T$ on the unit ball. □

## Acknowledgments

The work was finished while Liu visited Pan at the Department of Mathematical Sciences, Indiana University-Purdue University Fort Wayne. The work was supported by the China Scholarship Council (CSC). It was also supported by the National Natural Science Foundation of China under Grants 11671361, 11571256, and the China Postdoctoral Science Foundation under Grants 2016T90406, 2015M580378.

## References

[Burns and Krantz 1994] D. M. Burns and S. G. Krantz, "Rigidity of holomorphic mappings and a new Schwarz lemma at the boundary", *J. Amer. Math. Soc.* **7**:3 (1994), 661–676. MR Zbl

[Dai et al. 2010] S. Dai, H. Chen, and Y. Pan, "The high order Schwarz–Pick lemma on complex Hilbert balls", *Sci. China Math.* **53**:10 (2010), 2649–2656. MR Zbl

[Garnett 1981] J. B. Garnett, *Bounded analytic functions*, Pure and Applied Mathematics **96**, Academic Press, New York, 1981. MR Zbl

[Huang 1995] X. J. Huang, "A boundary rigidity problem for holomorphic mappings on some weakly pseudoconvex domains", *Canad. J. Math.* **47**:2 (1995), 405–420. MR Zbl

[Krantz 1992] S. G. Krantz, *Function theory of several complex variables*, 2nd ed., Wadsworth & Brooks/Cole, Pacific Grove, CA, 1992. MR Zbl

[Krantz 2011] S. G. Krantz, "The Schwarz lemma at the boundary", *Complex Var. Elliptic Equ.* **56**:5 (2011), 455–468. MR Zbl

[Liu et al. 2015] T. Liu, J. Wang, and X. Tang, "Schwarz lemma at the boundary of the unit ball in $\mathbb{C}^n$ and its applications", *J. Geom. Anal.* **25**:3 (2015), 1890–1914. MR Zbl

[Liu et al. 2016] Y. Liu, S. Dai, and Y. Pan, "Boundary Schwarz lemma for pluriharmonic mappings between unit balls", *J. Math. Anal. Appl.* **433**:1 (2016), 487–495. MR Zbl

[Rudin 1980] W. Rudin, *Function theory in the unit ball of* $\mathbf{C}^n$, Grundlehren der Math. Wissenschaften **241**, Springer, 1980. MR Zbl

Received October 16, 2016. Revised May 27, 2017.

YANG LIU
DEPARTMENT OF MATHEMATICS
ZHEJIANG NORMAL UNIVERSITY
JINHUA, 321004
CHINA
liuyang@zjnu.edu.cn

and

DEPARTMENT OF MATHEMATICAL SCIENCES
INDIANA UNIVERSITY-PURDUE UNIVERSITY FORT WAYNE
FORT WAYNE, IN 46805-1499
UNITED STATES

ZHIHUA CHEN
DEPARTMENT OF MATHEMATICS
TONGJI UNIVERSITY
SHANGHAI, 200092
CHINA
zzzhhc@tongji.edu.cn

YIFEI PAN

DEPARTMENT OF MATHEMATICAL SCIENCES
INDIANA UNIVERSITY-PURDUE UNIVERSITY FORT WAYNE
FORT WAYNE, IN 46805-1499
UNITED STATES

pan@ipfw.edu

and

SCHOOL OF MATHEMATICS AND INFORMATICS
JIANGXI NORMAL UNIVERSITY
NANCHANG 330022
CHINA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.477

# THETA CORRESPONDENCE AND THE PRASAD CONJECTURE FOR SL(2)

HENGFEI LU

**We use relations between the base change representations and theta lifts, to give a new proof to the local period problems of SL(2) over a nonarchimedean quadratic field extension $E/F$. Then we verify the Prasad conjecture for SL(2). With a similar strategy, we obtain a certain result for the Prasad conjecture for Sp(4).**

## 1. Introduction

Assume that $F$ is a nonarchimedean local field with characteristic 0. Let $G$ be a connected reductive group defined over $F$ and $H$ be a closed subgroup of $G$. Given a smooth irreducible representation $\pi$ of $G(F)$, one may consider the complex vector space $\mathrm{Hom}_{H(F)}(\pi, \mathbb{C})$. If it is nonzero, then we say that $\pi$ is $H(F)$-distinguished, or has a nonzero $H(F)$-period.

Period problems, which are closely related to harmonic analysis, have been extensively studied for classical groups. The most general situations have been studied in [Sakellaridis and Venkatesh 2017] when $G$ is split. Given a spherical variety $X = H\setminus G$, Sakellaridis and Venkatesh [2017] introduce a certain complex reductive group $\hat{G}_X$ associated with the variety $X$, to deal with the spectral decomposition of $L^2(H\setminus G)$ under the assumption that $G$ is split. In a similar way, Prasad [2015, §9] introduces a certain quasisplit reductive group $G^{\mathrm{op}}$ to deal with the period problem when the subgroup $H$ is the Galois fixed points of $G$, i.e., $H = G^{\mathrm{Gal}(E/F)}$, where $E$ is a quadratic field extension of $F$. In this paper, we will mainly focus on the cases $G = R_{E/F}\mathrm{SL}_2$ and $H = \mathrm{SL}_2$, where $R_{E/F}$ denotes the Weil restriction of scalars, i.e., the Prasad conjecture [2015, Conjecture 2] for $\mathrm{SL}_2$.

Let $W_F$ and $W_E$ be the Weil groups of $F$ and $E$, and let $WD_F$ and $WD_E$ be the Weil–Deligne groups. Let $\psi$ be any additive character of $F$ and $\psi_E = \psi \circ \mathrm{tr}_{E/F}$. Assume that $\tau$ is an irreducible smooth representation of $\mathrm{SL}_2(F)$, with a Langlands parameter $\phi_\tau : WD_F \to \mathrm{PGL}_2(\mathbb{C})$ and a character $\lambda$ of the component group $S_{\phi_\tau} = C(\phi_\tau)/C^\circ(\phi_\tau)$, where $C(\phi_\tau)$ is the centralizer of $\phi_\tau$ in $\mathrm{PGL}_2(\mathbb{C})$ and $C^\circ(\phi_\tau)$ is the connected component of $C(\phi)$. Then $\phi_\tau|_{WD_E}$ gives a Langlands

*MSC2010:* 11F27, 11F70, 22E50.
*Keywords:* theta lifts, periods, base change, Prasad's conjecture.

parameter of $\mathrm{SL}_2(E)$. The map $\phi_\tau \to \phi_\tau|_{WD_E}$ is called the base change map. Prasad's conjecture for SL(2) predicts the following result, which was shown in [Anandavardhanan and Prasad 2003].

**Theorem 1.1.** *Let $E$ be a quadratic field extension of a nonarchimedean local field $F$ with associated Galois group $\mathrm{Gal}(E/F) = \{1, \sigma\}$ and associated quadratic character $\omega_{E/F}$ of $F^\times$. Assume that $\tau$ is an irreducible smooth admissible representation of $\mathrm{SL}_2(E)$ with central character $\omega_\tau$ satisfying $\omega_\tau(-1) = 1$. Then the following are equivalent*:

(i) *$\tau$ is $\mathrm{SL}_2(F)$-distinguished.*

(ii) *$\phi_\tau = \phi_{\tau'}|_{WD_E}$ for some irreducible representation $\tau'$ of $\mathrm{SL}_2(F)$ and $\tau$ has a Whittaker model with respect to a nontrivial additive character of $E$ which is trivial on $F$.*

Anandavardhanan and Prasad [2003] deal with the cases for the principal series and square-integrable representations separately, using the restriction of $\mathrm{GL}_2(F)$-distinguished representations of $\mathrm{GL}_2(E)$. There is a key lemma [Anandavardhanan and Prasad 2003, Lemma 3.1] that if $\tau$ is $\mathrm{SL}_2(F)$-distinguished, then $\tau$ has a Whittaker model with respect to a nontrivial additive character of $E$ which is trivial on $F$. Moreover, the multiplicity $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C})$ is invariant under the $\mathrm{GL}_2(F)$-conjugation action on $\tau$. In [Anandavardhanan and Prasad 2016], they use a similar idea to deal with the case for $\mathrm{SL}_n$, involving the restriction of $\mathrm{GL}_n(F)$-distinguished representations of $\mathrm{GL}_n(E)$. In this paper, we will use the local theta correspondence to give a new proof for a tempered representation of $\mathrm{SL}_2(E)$. Then we use Mackey theory and the double coset decomposition to deal with the principal series, instead of involving representations of GL(2). In order to verify Prasad's conjecture [2015, Conjecture 2] for SL(2), we will list all possible explicit parameter lifts

$$\tilde{\phi} : WD_F \to \mathrm{PGL}_2(\mathbb{C})$$

such that $\tilde{\phi}|_{WD_E} = \phi_\tau$, which are different from Prasad's descriptions in [2015, §18]. Our methods can also be used for the Sp(4)-distinction problems over a quadratic field extension; see Theorem 4.2.

**Theorem 1.2.** *Assume that $\tau$ is an irreducible $\mathrm{SL}_2(F)$-distinguished representation of $\mathrm{SL}_2(E)$, with an enhanced $L$-parameter $(\phi_\tau, \lambda)$, where $\lambda$ is a character of the component group $S_{\phi_\tau}$, then*

$$\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = |F(\phi_\tau)|,$$

*where $F(\phi_\tau) = \{\tilde{\phi} : WD_F \to \mathrm{PGL}_2(\mathbb{C}) : \tilde{\phi}|_{WD_E} = \phi_\tau$ and $\lambda|_{S_{\tilde{\phi}}} \supset \mathbf{1}\}$ and $|F(\phi_\tau)|$ denotes its cardinality.*

**Remark 1.3.** The statement in Theorem 1.2 is slightly different from the original Prasad conjecture for SL(2). We have used the fact that the degree of the base change map

$$\Phi : \mathrm{Hom}(WD_F, \mathrm{PGL}_2(\mathbb{C})) \to \mathrm{Hom}(WD_E, \mathrm{PGL}_2(\mathbb{C}))$$

at each parameter $\tilde{\phi}$ is equal to the size of the cokernel

$$\mathrm{coker}\{S_{\tilde{\phi}} \to S_{\phi_\tau}^{\mathrm{Gal}(E/F)}\}$$

for $\tilde{\phi} \in F(\phi_\tau)$ when $G = \mathrm{SL}(2)$, which is easy to check; see [Prasad 2015, §18].

**Remark 1.4.** Raphael Beuzart-Plessis [2017, Theorem 1] uses the relative trace formula to give an identity for the multiplicity $\dim_{\mathbb{C}} \mathrm{Hom}_{H'(F)}(\pi', \chi_{H'})$, where $H'$ is an inner form of $H$ defined over $F$, $\chi_{H'}$ is a quadratic character of $H'(F)$ and $\pi'$ is a stable square-integrable representation of $(R_{E/F}H')(F) = H'(E)$. For example, $H' = \mathrm{SL}_1(D)$ and $H'(E) = \mathrm{SL}_2(E)$, where $D$ is a quaternion division algebra defined over $F$. We plan to use the local theta correspondence to deal with the distinction problems for the pair $(\mathrm{SL}_2(E), \mathrm{SL}_1(D))$ in a subsequent paper. More precisely, we will figure out the multiplicity $\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{SL}_1(D)}(\tau, \mathbb{C})$ for a smooth irreducible representation $\tau$ of $\mathrm{SL}_2(E)$.

**Remark 1.5.** Anandavardhanan and Prasad [2006; 2013] discuss the global period problems for $\mathrm{SL}_2$ over a quadratic number field extension $\mathbb{E}/\mathbb{F}$. More generally, there are several results for the global period problems of $\mathrm{SL}_1(D)$ in [Anandavardhanan and Prasad 2013, §9], where $\mathrm{SL}_1(D)$ is an inner form of $\mathrm{SL}_2$ defined over a number field $\mathbb{F}$. We hope that we can also use the global theta correspondence to revisit these questions in future.

Now we briefly describe the contents and the organization of this paper. In §2, we set up the notation about the local theta lifts. In §3, we give the proof of Theorem 1.1, and then we verify Prasad's conjecture for SL(2), i.e., Theorem 1.2 in §4. Finally, we give a partial result for the Prasad conjecture for $\mathrm{Sp}_4$, i.e., Theorem 4.2.

## 2. The local theta correspondences

In this section, we will briefly recall some results about the local theta correspondence, following [Kudla 1996].

Let $F$ be a local field of characteristic zero. Consider the dual pair $\mathrm{O}(V) \times \mathrm{Sp}(W)$. For simplicity, we may assume that $\dim V$ is even. Fix a nontrivial additive character $\psi$ of $F$. Let $\omega_\psi$ be the Weil representation for $\mathrm{O}(V) \times \mathrm{Sp}(W)$, which can be described as follows. Fix a Witt decomposition $W = X \oplus Y$ and let $P(Y) = \mathrm{GL}(Y)N(Y)$ be the parabolic subgroup stabilizing the maximal isotropic subspace $Y$. Then

$$N(Y) = \{b \in \mathrm{Hom}(X, Y) \mid b^t = b\},$$

where $b^t \in \mathrm{Hom}(Y^*, X^*) \cong \mathrm{Hom}(X, Y)$. The Weil representation $\omega_\psi$ can be realized on the Schwartz space $S(X \otimes V)$ and the action of $P(Y) \times \mathrm{O}(V)$ is given by the usual formula

$$\begin{cases} \omega_\psi(h)\phi(x) = \phi(h^{-1}x), & \text{for } h \in \mathrm{O}(V), \\ \omega_\psi(a)\phi(x) = \chi_V(\det_Y(a))|\det_Y a|^{\frac{1}{2}\dim V}\phi(a^{-1}\cdot x), & \text{for } a \in \mathrm{GL}(Y), \\ \omega_\psi(b)\phi(x) = \psi(\langle bx, x\rangle)\phi(x), & \text{for } b \in N(Y), \end{cases}$$

where $\chi_V$ is the quadratic character associated to the disc $V \in F^\times/F^{\times 2}$ and $\langle -, - \rangle$ is the natural symplectic form on $W \otimes V$. To describe the full action of $\mathrm{Sp}(W)$, one needs to specify the action of a Weyl group element, which acts by a Fourier transform.

If $\pi$ is an irreducible representation of $\mathrm{O}(V)$ (resp. $\mathrm{Sp}(W)$), the maximal $\pi$-isotypic quotient has the form

$$\pi \boxtimes \Theta_\psi(\pi)$$

for some smooth representation of $\mathrm{Sp}(W)$ (resp. $\mathrm{O}(V)$). We call $\Theta_\psi(\pi)$ the big theta lift of $\pi$. It is known that $\Theta_\psi(\pi)$ is of finite length and hence is admissible. Let $\theta_\psi(\pi)$ be the maximal semisimple quotient of $\Theta_\psi(\pi)$, which is called the small theta lift of $\pi$. Then there is a conjecture of Howe which states that

- $\theta_\psi(\pi)$ is irreducible whenever $\Theta_\psi(\pi)$ is nonzero.
- the map $\pi \mapsto \theta_\psi(\pi)$ is injective on its domain.

This has been proved by Waldspurger [1990] when the residual characteristic $p$ of $F$ is not 2. Recently, it has been proved completely in [Gan and Takeda 2016a; 2016b].

**Theorem 2.1.** *The Howe conjecture holds.*

***First occurrence indices for pairs of orthogonal Witt towers.*** Let $W_n$ be the $2n$-dimensional symplectic vector space with associated symplectic group $\mathrm{Sp}(W_n)$ and consider the two towers of orthogonal groups attached to the quadratic spaces with nontrivial discriminant. Let $V_E$ and $\epsilon V_E$ be 2-dimensional quadratic spaces with discriminant $E$ and Hasse invariants $+1$ and $-1$, respectively, and let $\mathbb{H}$ be the 2-dimensional hyperbolic quadratic space over $F$,

$$V_r^+ = V_E \oplus \mathbb{H}^{r-1} \quad \text{and} \quad V_r^- = \epsilon V_E \oplus \mathbb{H}^{r-1},$$

and denote the orthogonal groups by $\mathrm{O}(V_r^+)$ and $\mathrm{O}(V_r^-)$, respectively. For an irreducible representation $\pi$ of $\mathrm{Sp}(W_n)$, one may consider the theta lifts $\theta_r^+(\pi)$ and $\theta_r^-(\pi)$ to $\mathrm{O}(V_r^+)$ and $\mathrm{O}(V_r^-)$, respectively, with respect to a fixed nontrivial additive character $\psi$. Set

$$\begin{cases} r^+(\pi) = \inf\{2r : \theta_r^+(\pi) \neq 0\}; \\ r^-(\pi) = \inf\{2r : \theta_r^-(\pi) \neq 0\}. \end{cases}$$

Then Kudla and Rallis [2005] and B. Sun and C. Zhu [2015] showed the following:

**Theorem 2.2** (conservation relation). *For any irreducible representation $\pi$ of* $\mathrm{Sp}(W_n)$, *we have*

$$r^+(\pi)+r^-(\pi)=4n+4=4+2\dim W_n.$$

On the other hand, one may consider the mirror situation, where one fixes an irreducible representation of $\mathrm{O}(V_r^+)$ or $\mathrm{O}(V_r^-)$, and consider its theta lifts $\theta_n(\pi)$ to the tower of the symplectic group $\mathrm{Sp}(W_n)$. Then with $n(\pi)$ defined in the analogous fashion, due to [Sun and Zhu 2015, Theorem 1.10], we have

$$n(\pi)+n(\pi\otimes\det)=\dim V_r^{\pm}.$$

***See-saw identities.*** Let $(V, q)$ be a quadratic vector space over $E$ of even dimension. Let $V'=\mathrm{Res}_{E/F} V$ be the same space $V$ but now thought of as a vector space over $F$ with a quadratic form

$$q'(v)=\tfrac{1}{2}\operatorname{tr}_{E/F} q(v).$$

If $W_0$ is a symplectic vector space over $F$, then $W_0\otimes_F E$ is a symplectic vector space over $E$. Then we have the following isomorphism of symplectic spaces:

$$\mathrm{Res}_{E/F}[(W_0\otimes_F E)\otimes_E V]\cong W_0\otimes V'=\boldsymbol{W}.$$

There is a pair

$$(\mathrm{Sp}(W_0),\mathrm{O}(V'))\quad\text{and}\quad(\mathrm{Sp}(W_0\otimes E),\mathrm{O}(V))$$

of dual reductive pairs in the symplectic group $\mathrm{Sp}(\boldsymbol{W})$. A pair $(G_1, H_1)$ and $(G_2, H_2)$ of dual reductive pairs in a symplectic group is called a see-saw pair if $H_1\subset G_2$ and $H_2\subset G_1$.

**Lemma 2.3** [Kudla 1984]. *For a see-saw pair of dual reductive pairs* $(G_1, H_1)$ *and* $(G_2, H_2)$, *let* $\pi_1$ *be an irreducible representation of* $H_1$ *and* $\pi_2$ *of* $H_2$, *then we have the isomorphism*

$$\mathrm{Hom}_{H_1}(\Theta_\psi(\pi_2),\pi_1)\cong\mathrm{Hom}_{H_2}(\Theta_\psi(\pi_1),\pi_2).$$

***Quadratic spaces.*** Let $K/E$ be a quadratic field extension and $V=V_K$ be a 2-dimensional quadratic space over $E$ with the norm map $N_{K/E}$. Set $\varpi$ to be the uniformizer of $\mathcal{O}_F$ and $\mathrm{Gal}(K/E)=\langle s\rangle$. Let $u$ be a unit in $\mathcal{O}_F^\times\setminus\mathcal{O}_F^{\times 2}$. Assume that the Hilbert symbol $(\varpi, u)_F$ is $-1$.

**Example 2.4.** Assume that $p$ is odd. Let $L=F(\sqrt{-\varpi})$ be a quadratic field extension over $F$ with associated quadratic character $\omega_{L/F}=\omega_{F(\sqrt{-\varpi})/F}$ by local class field theory. Let $K$ be a quadratic field extension over $E$, then $V_K$ is a 2-dimensional quadratic space over $E$ with norm map $N_{K/E}$. We may regard $V_K$ as a 4-dimensional quadratic space $V'$ over $F$ with quadratic form $q'(k)=\frac{1}{2}\operatorname{tr}_{E/F} N_{K/E}(k)$ for $k\in K$.

(i) If $E = F(\sqrt{\varpi})$ is ramified, then:

- If $K = E(\sqrt{u})$, then the discriminant $\mathrm{disc}(V') = 1 \in F^\times/F^{\times 2}$ and the Hasse invariant $\epsilon(V') = -1$.
- If $K = E(\sqrt[4]{\varpi})$, then $V' = V_L \oplus \mathbb{H}$ and $\mathrm{disc}(V') = -\varpi \in F^\times/F^{\times 2}$.
- If $K = E(\sqrt[4]{\varpi} \cdot \sqrt{u})$, then $\mathrm{disc}(V') = L$.

(ii) If $E = F(\sqrt{u})$ is unramified, then:

- If $K = E(\sqrt{\varpi})$, then $\mathrm{disc}(V') = 1$ and
$$\epsilon(V') = -(-1, \varpi)_F = \begin{cases} +1 & \text{if } -1 \in uF^{\times 2}; \\ -1 & \text{if } -1 \in F^{\times 2}. \end{cases}$$
- If $K = E(\sqrt{u'})$ and $u' \notin F^\times$, then $\mathrm{disc}(V') = N_{E/F}(u') \in F^\times/F^{\times 2}$.

If $-1 \in (F^\times)^2$ is a square in $F^\times$ and the discriminant of $V' = \mathrm{Res}_{E/F} V_K$ is the same as the discriminant of the 2-dimensional vector space $E$ over $F$, i.e., $\mathrm{disc}(V') = E$, then $\chi_{V'}$ is $\omega_{E/F}$ and its special orthogonal group, denoted by $\mathrm{SO}(V') = \mathrm{SO}(3, 1)$, is isomorphic to

$$\begin{aligned} \mathrm{SO}(3,1) &= \frac{\{(g, \lambda) \in \mathrm{GL}_2(E) \times F^\times : \lambda^2 N_{E/F}(\det g) = 1\}}{\{(t, N_{E/F}(t)^{-1}) : t \in E^\times\}} \\ &\cong \frac{\{g \in \mathrm{GL}_2(E) : \det(g) \in F^\times\}}{F^\times}. \end{aligned}$$

Set $K^1 = \{k \in K^\times : k \cdot k^s = 1\}$, then there is a natural embedding

$$\mathrm{O}(V_K) = K^1 \rtimes \mu_2 \subset \mathrm{SO}(3,1) \quad \text{where } K^1 = \mathrm{SO}(V_K) \subset \mathrm{GL}_2(E).$$

In general, the discriminant $\mathrm{disc}(V')$ may not be equal to $E$. There is a group embedding $K^1 \hookrightarrow \mathrm{GL}_2(L')$ where $L' = F(\delta)$ and $\delta^2 = N_{E/F}(u')$ if $K = E(\sqrt{u'})$.

**Remark 2.5.** If $V' = \mathrm{Res}_{E/F} V_K$ has discriminant $1 \in F^\times/F^{\times 2}$ and Hasse invariant $+1$, then $V'$ is called a split 4-dimensional quadratic space over $F$. Set $\mathrm{SO}_{2,2}(F) = \mathrm{SO}(V')$ to be the special orthogonal group.

***Degenerate principal series representations.*** Let $V_K$ be a 2-dimensional quadratic space over $E$ with the norm map $N_{K/E}$. Assume that $V' = \mathrm{Res}_{E/F} V_K$ is a split 4-dimensional quadratic space over $F$. There is a natural embedding $\mathrm{O}(V_K) \hookrightarrow \mathrm{O}_{2,2}(F)$. Let $P$ be a Siegel parabolic subgroup of $\mathrm{O}_{2,2}(F)$. Assume that $\mathcal{I}(s)$ is the degenerate principal series of $\mathrm{O}_{2,2}(F)$. Let us consider the double coset decomposition $P \backslash \mathrm{O}_{2,2}(F)/\mathrm{O}(V_K)$.

- If $K$ is a field, then there are four open orbits in $P \backslash \mathrm{O}_{2,2}(F)/\mathrm{O}(V_K)$.
- If $K = E \oplus E$, then there are one closed orbit and three open orbits in $P \backslash \mathrm{O}_{2,2}(F)/\mathrm{O}_{1,1}(E)$.

Assume that there is a stratification $P\backslash \mathrm{O}_{2,2}(F)/\mathrm{O}(V_K)=\sqcup_{i=0}^{r} X_i$ such that $\bigsqcup_{i=0}^{k} X_i$ is open for each $k$ lying in $\{0, 1, 2, \ldots, r\}$. Then there is an $\mathrm{O}(V_K)$-equivariant filtration $\{I_i\}_{i=0,1,2,\ldots,r}$ of $\mathcal{I}(s)|_{\mathrm{O}(V_K)}$ such that

$$0 = I_{-1} \subset I_0 \subset I_1 \subset \cdots \subset I_r = \mathcal{I}(s)|_{\mathrm{O}(V_K)}$$

and the smooth functions in the quotient $I_i/I_{i-1}$ are supported on a single orbit $X_i$ in $P \backslash \mathrm{O}_{2,2}(F)/\mathrm{O}(V_K)$.

**Definition 2.6.** Given an irreducible representation $\pi$ of $\mathrm{O}(V_K)$, if

$$\mathrm{Hom}_{\mathrm{O}(V_K)}(I_{i+1}/I_i, \pi) \neq 0$$

implies that $I_{i+1}/I_i$ is supported on the open orbits in $P \backslash \mathrm{O}_{2,2}(F)/\mathrm{O}(V_K)$, then we say that the representation $\pi$ does not occur on the boundary of $\mathcal{I}(s)$.

It is well known that only the open orbits can support supercuspidal representations. Due to the Casselman criterion for a tempered representation, only the open orbits can support the tempered representations in our case if $s = \frac{1}{2}$; see [Lu 2017, Lemma 4.2.9].

## 3. Proof of Theorem 1.1

Before we prove Theorem 1.1, let us recall some facts.

**Lemma 3.1.** *If the discriminant of $V' = \mathrm{Res}_{E/F} V_K$ is $E$, then the theta lift of the trivial representation from $\mathrm{SL}_2(F)$ to $\mathrm{SO}(3, 1) = \mathrm{SO}(V')$ is a character, i.e.,*

$$\Theta_\psi(\mathbf{1}) = \mathbf{1} \boxtimes \omega_{E/F}.$$

*Proof.* Due to [Lu 2017, Theorem 2.4.11], the big theta lift of the Steinberg representation St from $\mathrm{GL}_2^+(F)$ to GSO(3, 1) is $\Theta_\psi(\mathrm{St}) = \mathrm{St}_E \boxtimes \omega_{E/F}$. By a similar argument, one can get $\Theta_\psi(\mathbf{1}) = \mathbf{1} \boxtimes \omega_{E/F}$. Notice that

$$\Theta_\psi(\mathbf{1}|_{\mathrm{SL}_2}) = \Theta_\psi(\mathbf{1})|_{\mathrm{SO}(3,1)},$$

then we are done. □

**Remark 3.2.** In fact, the theta lift $\theta'_\psi(\mathbf{1})$ from $\mathrm{SL}_2(F)$ to O(3, 1) remains irreducible when restricted to SO(3, 1), see [Prasad 1993, §5].

Now we begin the proof of Theorem 1.1, which we will complete in Section 4.

***Proof of Theorem 1.1.*** According to the representation $\tau$, we separate the proof into four cases:

- $\tau$ is a supercuspidal representation; see (A).
- $\tau$ is an irreducible principal series representation; see (B).

- $\tau$ is a Steinberg representation $\mathrm{St}_E$; see (C).
- $\tau$ is a constituent of a reducible principle series $I(\chi)$ with $\chi^2 = 1$; see (D).

These exhaust all irreducible smooth representations of $\mathrm{SL}_2(E)$.

(A) If $\tau$ is supercuspidal, then there exists a character $\mu : K^\times \to \mathbb{C}^\times$ such that $\phi_\tau = i \circ (\mathrm{Ind}_{W_K}^{W_E} \mu)$, where

- $W_K$ is the Weil group of $K$, where $K$ is a quadratic field extension over $E$;
- $\mu$ does not factor through the norm map $N_{K/E}$, so the irreducible Langlands parameter
$$\mathrm{Ind}_{W_K}^{W_E} \mu : W_E \to \mathrm{GL}_2(\mathbb{C})$$
corresponds to a dihedral supercuspidal representation of $\mathrm{GL}_2(E)$ with respect to $K$;
- $i : \mathrm{GL}_2(\mathbb{C}) \to \mathrm{PGL}_2(\mathbb{C})$ is the projection map, which coincides with the adjoint map
$$\mathrm{Ad} : \mathrm{GL}(2) \to \mathrm{SO}(3).$$

In fact, the Langlands parameter $\phi$ of the representation $\Sigma$ of $\mathrm{O}(V_K)$, where $\tau = \theta_\psi(\Sigma)$, is given by

$$\phi(g) = \begin{cases} \begin{pmatrix} \chi_K(g) & \\ & \chi_K^{-1}(g) \end{pmatrix} & \text{if } g \in W_K, \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \text{if } g = s, \end{cases}$$

where $s \in W_E \setminus W_K$ and the character $\chi_K : W_K \to \mathbb{C}^\times$ is the pull back of a nontrivial character $\mu_1$ of $K^1$ under the map $K^\times \to K^1$ via $k \mapsto k^s k^{-1}$, i.e., $\chi_K(k) = \mu_1(k^s k^{-1})$, see [Kudla 1996, §6.4]. Furthermore, there is an isomorphism between two Langlands parameters of O(2),

$$\phi \otimes \omega_{K/E} \cong \mathrm{Ind}_{W_K}^{W_E} \frac{\mu^s}{\mu}.$$

In other words, one has $\chi_K = \mu^s \mu^{-1}$ and $\mu_1 = \mu|_{K^1}$ is the restricted character.

Moreover, if $\mu_1^2 \neq \mathbf{1}$, then

$$\tau = \theta_\psi(\mathrm{Ind}_{\mathrm{SO}(V_K)}^{\mathrm{O}(V_K)}(\mu_1)).$$

If $\mu_1^2 = \mathbf{1}$, then there are two extensions of $\mu_1$ from $\mathrm{SO}(V_K)$ to $\mathrm{O}(V_K)$, denoted by $\mu_1^\pm$. For convenience, if $\mu_1^2 \neq \mathbf{1}$, we denote the irreducible representation

$$\mathrm{Ind}_{\mathrm{SO}(V_K)}^{\mathrm{O}(V_K)}(\mu_1)$$

by $\mu_1^+$ as well. Assume that $\tau = \Theta_\psi(\mu_1^+)$ is supercuspidal.

If the discriminant disc $V' = L \in F^\times/(F^\times)^2$ is nontrivial, by the see-saw diagram

$$\begin{array}{cccc} \tau & \mathrm{SL}_2(E) & \mathrm{O}(V') & \Theta_\psi(\mathbf{1}) \\ & & & \\ \mathbf{1} & \mathrm{SL}_2(F) & \mathrm{O}(V_K) & {\mu_1}^+ \end{array}$$

one has an isomorphism

$$\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) \cong \mathrm{Hom}_{\mathrm{O}(V_K)}(\mathbf{1} \boxtimes \omega_{L/F}, \mu_1^+)$$

which is nonzero if and only if $\mu_1 = \mathbf{1}$. But $\mathrm{Hom}_{K^1}(\mathbf{1}, \mu_1) = 0$, and therefore $\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 0$.

If the discriminant of $V'$ is $1 \in F^\times/(F^\times)^2$ and its Hasse invariant is $-1$, then the theta lift $\theta_\psi(\mathbf{1})$ from $\mathrm{SL}_2(F)$ to $\mathrm{O}(V')$ is zero by the conservation relation, so that

$$\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = \mathrm{Hom}_{\mathrm{O}(V_K)}(\Theta_\psi(\mathbf{1}), \theta_\psi(\tau)) = 0.$$

If $V' \cong \mathbb{H}^2$ is a split 4-dimensional quadratic space over $F$, we denote by $\mathcal{I}(s)$ the degenerate principal series of $\mathrm{O}_{2,2}(F)$ and we assume that $F^\times/(F^\times)^2 \supset \{1, u, \varpi, u\varpi\}$ and $E = F(\sqrt{u})$ with associated Galois group $\mathrm{Gal}(E/F) = \langle\sigma\rangle$. Then

$$\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = \mathrm{Hom}_{\mathrm{O}(V_K)}\big(\mathcal{I}\big(\tfrac{1}{2}\big), \mu_1^+\big) \cong \bigoplus_{j=1}^{4} \mathrm{Hom}_{\mathrm{O}(V_j)}(\mu_1^+, \mathbb{C}), \tag{3-1}$$

where $K = F(\sqrt{\varpi}, \sqrt{u})$ is a biquadratic field over $F$, and

- $V_1 = V_{E'}$ (where $E' = F(\sqrt{\varpi})$ is a quadratic field extension over $F$) is a 2-dimensional quadratic space over $F$ with quadratic form $q(e') = N_{E'/F}(e')$, Hasse invariant $+1$ and quadratic character $\chi_{V_1} = \omega_{E'/F} = \omega_{F(\sqrt{\omega})/F}$;
- $V_2 = \epsilon' V_1$ ($\epsilon' \in F^\times \setminus N_{E'/F}(E')^\times$) is the 2-dimensional quadratic space $F(\sqrt{\varpi})$ with quadratic form $\epsilon' N_{E'/F}$, Hasse invariant $-1$ and quadratic character $\chi_{V_2} = \chi_{V_1}$;
- $V_3 = V_{E''}$ is a 2-dimensional quadratic space over $F$ with quadratic character $\omega_{F(\sqrt{\varpi u})/F}$ and Hasse invariant $+1$, where $E'' = F(\sqrt{\varpi u})$ is a quadratic field extension over $F$; and
- $V_4 = \epsilon'' V_3$ with Hasse invariant $-1$, where $\epsilon'' \in F^\times \setminus N_{E''/F}(E'')^\times$.

In the latter case, (3-1) can be rewritten as the identity

$$\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = \sum_{j=1}^{4} \dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{O}(V_j)}(\mu_1^+, \mathbb{C}), \tag{3-2}$$

which is nonzero if and only if one of the following holds:

- $\mu(x - y\sqrt{\varpi}) = \mu(x + y\sqrt{\varpi})$ for $x, y \in F$.
- $\mu(x - y\sqrt{u\varpi}) = \mu(x + y\sqrt{u\varpi})$ for $x, y \in F$.

**Remark 3.3.** Because $\mu^s \neq \mu$, these two conditions cannot hold at the same time unless $p = 2$.

We would like to highlight a fact about the group embeddings $\mathrm{O}(V_j) \hookrightarrow K^1 \rtimes \langle s \rangle$ for $j \in \{1, 2\}$. There is a natural group embedding $\mathrm{SO}(V_1) \rtimes \langle s \rangle \to K^1 \rtimes \langle s \rangle$. Via the isomorphism between two quadratic $E$-vector spaces $(V_{E'} \otimes_F E, \epsilon' N_{E'/F}) \cong (V_K, N_{K/E})$, one has an identity

$$\dim \mathrm{Hom}_{\mathrm{O}(\epsilon' V_{E'})}(\mu_1^+, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_{E'})}((\mu_1^+)^{g_{\epsilon'}}, \mathbb{C}),$$

where $(\mu_1^+)^{g_{\epsilon'}}$ is a representation of $\mathrm{O}(V_K)$ given by

$$(\mu_1^+)^{g_{\epsilon'}}(x) = \mu_1^+(g_{\epsilon'}^{-1} x g_{\epsilon'}), \quad x \in \mathrm{O}(V_K), \quad g_{\epsilon'} \in \mathrm{GSO}(V_K) = K^\times \quad \text{with } N_{K/E}(g_{\epsilon'}) = \epsilon'.$$

Further, if the Whittaker datum is fixed, then the enhanced $L$-parameter of $(\mu_1^+)^{g_{\epsilon'}}$ is known if the enhanced $L$-parameter of $\mu_1^+$ is given; see [Atobe and Gan 2017, §3.6].

***The case $p \neq 2$.*** (i) If $\mu_1^2 \neq \mathbf{1}$, then $\mathrm{Ind}_{\mathrm{SO}(V_K)}^{\mathrm{O}(V_K)}(\mu_1)$ is irreducible and

$$\dim \mathrm{Hom}_{\mathrm{O}(V_2)}(\mathrm{Ind}_{\mathrm{SO}(V_K)}^{\mathrm{O}(V_K)}(\mu_1), \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_1)}(\mathrm{Ind}_{\mathrm{SO}(V_K)}^{\mathrm{O}(V_K)}(\mu_1), \mathbb{C}).$$

(ii) If $\mu_1^2 = \mathbf{1}$, then $\mu^2 = \chi_E \circ N_{K/E}$ and $\mu^s = -\mu$, so

$$\dim \mathrm{Hom}_{\mathrm{O}(V_2)}(\mu_1^+, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_1)}(\mu_1^-, \mathbb{C}).$$

Hence, if $p \neq 2$, (3-2) implies the following:

- If $\mu_1^2 \neq \mathbf{1}$ and $\mu|_{E'}$ factors through the norm map $N_{E'/F}$ for $E' \neq E$, then
$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 2.$$
- If $\mu_1^2 = \mathbf{1}$ and $\mu|_{E'}$ factors through the norm map $N_{E'/F}$ for $E' \neq E$, then
$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 1.$$

If $\mu_1^2 = \mathbf{1}$ and $\tau = \theta_\psi(\mu_1^+)$ is $\mathrm{SL}_2(F)$-distinguished, then

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\mu_1^-), \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_K)}\big(\mathcal{I}(\tfrac{1}{2}), \mu_1^-\big)$$

which is equal to

$$\sum_{j=1}^{4} \dim \mathrm{Hom}_{\mathrm{O}(V_j)}(\mu_1^-, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\mu_1^+), \mathbb{C}).$$

Hence

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\mu_1^-), \mathbb{C}) = 1$$

if and only if $\mu|_{E'}$ factors through the norm map $N_{E'/F}$ for $E' \neq E$.

***The case $p = 2$.*** (i) Suppose that there are two distinct quadratic fields $E'$ and $E''$ over $F$ such that $\mu|_{E'} = \chi'_F \circ N_{E'/F}$ and $\mu|_{E''} = \chi''_F \circ N_{E''/F}$. Furthermore, $\chi'_F/\chi''_F$ is a quadratic character of $F^\times$ that is not trivial restricted on the Weil group $W_K$ of $K$, i.e., $\chi'_F/\chi''_F$ is different from three quadratic characters $\omega_{E/F}$, $\omega_{E'/F}$ and $\omega_{E''/F}$, which may happen only when $p = 2$. In this case, $\mu^s(t) = \mu(t) \cdot \chi'_F/\chi''_F(t)$ for $t \in W_K$,

$$\dim \mathrm{Hom}_{\mathrm{O}(V_1)}(\mu_1^+, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_2)}(\mu_1^+, \mathbb{C}),$$

and $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 4$ by the identity (3-2).

(ii) Given a cuspidal representation $\pi$ of $\mathrm{GL}_2(E)$ with $\pi|_{\mathrm{SL}_2(E)} \supset \tau$, if $\pi$ is not dihedral with respect to any quadratic extension $K$ over $E$, then $\pi|_{\mathrm{SL}_2(E)} = \tau$ is irreducible.

We consider a 4-dimensional quadratic space $X$ over $F$ with discriminant $E$, then the orthogonal group $\mathrm{O}(X) = \mathrm{O}(3, 1)$ can be naturally embedded into the orthogonal group $\mathrm{O}(X \otimes_F E) = \mathrm{O}(2, 2)(E)$. Let $\pi \boxtimes \pi$ be the irreducible representation of the similitude special orthogonal group $\mathrm{GSO}(2, 2)(E)$. By the property of the big theta lift $\Theta(\pi)$ from $\mathrm{GL}_2(E)$ to $\mathrm{GSO}(2, 2)(E)$,

$$(\pi \boxtimes \pi)|_{\mathrm{SO}(2,2)(E)} = \Theta(\pi)|_{\mathrm{SO}(2,2)(E)} = \Theta(\pi|_{\mathrm{SL}_2(E)}) = \Theta(\tau)$$

is irreducible since $\tau$ is supercuspidal. Let $\mathfrak{I}(s)$ be the degenerate principal series of $\mathrm{Sp}_4(F)$. Assume that $(\pi \boxtimes \pi)^+$ is the unique extension from $\mathrm{GSO}(2, 2)(E)$ to $\mathrm{GO}(2, 2)(E)$ which participates with the theta correspondence with $\mathrm{GL}_2(E)$. Then $(\pi \boxtimes \pi)^+|_{\mathrm{O}(2,2)(E)}$ is irreducible. Considering the see-saw diagram

$$\begin{array}{cccc} \mathfrak{I}\big(\tfrac{1}{2}\big) & \mathrm{Sp}_4(F) & \mathrm{O}(2, 2)(E) & (\pi \boxtimes \pi)^+ \\ & & \!\!\!\!\!\!\times\!\!\!\!\!\! & \\ \pi|_{\mathrm{SL}_2(E)} & \mathrm{SL}_2(E) & \mathrm{O}(3, 1)(F) & \mathbb{C} \end{array}$$

due to the structure of $\mathfrak{I}\big(\frac{1}{2}\big)$ in [Gan and Ichino 2014, Proposition 7.2], one can get an equality

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(E)}\big(\mathfrak{I}\big(\tfrac{1}{2}\big), \pi\big) = \dim \mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi \boxtimes \pi)^+, \mathbb{C}).$$

The supercuspidal representation $\pi|_{\mathrm{SL}_2(E)}$ does not occur on the boundary of $\mathfrak{I}\big(\frac{1}{2}\big)$, therefore

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(E)}\big(\mathfrak{I}\big(\tfrac{1}{2}\big), \pi\big) = \dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\pi^\vee, \mathbb{C}).$$

By the conservation relation, the fact that the first occurrence index of the determinant map det of $\mathrm{O}(3, 1)(F)$ is 4 implies that $\Theta_\psi(\det)$ from $\mathrm{O}(3, 1)(F)$ to

$\mathrm{Sp}(W_2)=\mathrm{Sp}_4(F)$ is zero and

$$\begin{aligned}\mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi\boxtimes\pi)^-,\mathbb{C})&\cong\mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi\boxtimes\pi)^+,\det)\\&=\mathrm{Hom}_{\mathrm{SL}_2(E)}(\Theta_\psi(\det),\pi|_{\mathrm{SL}_2(E)})=0.\end{aligned}$$

Hence

$$\begin{aligned}\text{(3-3)}\quad &\dim\mathrm{Hom}_{\mathrm{SL}_2(F)}(\pi^\vee,\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi\boxtimes\pi)^+,\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi\boxtimes\pi)^+,\mathbb{C})+\dim\mathrm{Hom}_{\mathrm{O}(3,1)(F)}((\pi\boxtimes\pi)^-,\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{O}(3,1)(F)}(\mathrm{Ind}_{\mathrm{SO}(2,2)(E)}^{\mathrm{O}(2,2)(E)}(\pi\boxtimes\pi)|_{\mathrm{SO}(2,2)(E)},\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{SO}(3,1)(F)}((\pi\boxtimes\pi),\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{GSO}(3,1)(F)}(\pi\boxtimes\pi,\mathbb{C})\\&=\dim\mathrm{Hom}_{\mathrm{GL}_2(E)}(\pi^\sigma,\pi^\vee).\end{aligned}$$

Therefore, if $\pi$ is not dihedral with respect to any quadratic field extension $K$ over $E$ then $\tau=\pi|_{\mathrm{SL}_2(E)}$ is irreducible, and so the following are equivalent:

- $\pi^\sigma\cong\pi^\vee$, i.e., $\phi_\pi$ is conjugate-self-dual in the sense of [Gan et al. 2012, §3].
- $\dim\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau,\mathbb{C})=1$.

**Remark 3.4.** This method can be used to deal with the case when $\tau$ is the Steinberg representation $\mathrm{St}_E$ of $\mathrm{SL}_2(E)$, which will imply $\dim\mathrm{Hom}_{\mathrm{SL}_2(F)}(\mathrm{St}_E,\mathbb{C})=1$ directly. It will appear in the proof of Theorem 4.2 as well.

(B) Let $\chi$ be a unitary character of $E^\times$. If $\tau=I(z,\chi)=\mathrm{Ind}_{B(E)}^{\mathrm{SL}_2(E)}\chi|-|_E^z$ (normalized induction) is an irreducible principal series, by the double coset decomposition for $B(E)\setminus\mathrm{SL}_2(E)/\mathrm{SL}_2(F)$

$$\mathrm{SL}_2(E)=B(E)\mathrm{SL}_2(F)\sqcup B(E)\eta_1\mathrm{SL}_2(F)\sqcup B(E)\eta_2\mathrm{SL}_2(F),$$

where

$$\eta_1=\begin{pmatrix}1&\\\sqrt{d}&1\end{pmatrix}\quad\text{and}\quad\eta_2=\begin{pmatrix}1&\\\epsilon\sqrt{d}&1\end{pmatrix},$$

$\epsilon\in F^\times\setminus N_{E/F}(E^\times)$, then there is a short exact sequence

$$\begin{aligned}\text{(3-4)}\quad \mathrm{Hom}_{F^\times}(|-|_E^z\chi,\mathbb{C})&\hookrightarrow\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau,\mathbb{C})\\&\to\textstyle\prod_{j=1}^2\mathrm{Hom}_{E^1}(\tau^{\eta_j},\mathbb{C})\to\mathrm{Ext}^1_{F^\times}(|-|_E^z\chi,\mathbb{C}),\end{aligned}$$

where $\tau^{\eta_j}\left(\begin{smallmatrix}a&*\\&\bar{a}\end{smallmatrix}\right)=\chi(a)$ for $a\in E^1=\ker\{N_{E/F}:E^\times\to F^\times\}$. Then $\mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau,\mathbb{C})$ is not equal to 0 if and only if one of the following conditions holds:

- $\chi|_{F^\times}=\mathbf{1}$ and $z=0$;
- $\chi=\chi_F\circ N_{E/F}$.

In order to verify the Prasad conjecture, we need to figure out the exact dimension $\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C})$.

(i) If $\chi$ is trivial and $z=0$, then $\tau = I(\mathbf{1})$ is irreducible and $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 2$.

(ii) If $\chi = \chi_F \circ N_{E/F}$ with $\chi^2 = \mathbf{1} \neq \chi$ and $z = 0$, then $I(\chi)$ is reducible, which belongs to the tempered cases and we will discuss later; see (D).

(iii) If $\chi = \chi_F \circ N_{E/F}$ with $\chi^2 \neq \mathbf{1}$, then $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 2$.

(iv) If $\chi$ does not factor through $N_{E/F}$ but $\chi|_{F^\times} = \mathbf{1}$ and $s = 0$, then

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 1.$$

(C) If $\tau = \mathrm{St}_E$ is a Steinberg representation of $\mathrm{SL}_2(E)$, then the exact sequence (3-4) implies that

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(I(|-|_E), \mathbb{C}) = 2,$$

so that $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\mathrm{St}_E, \mathbb{C}) = 2 - 1 = 1$.

(D) Assume that $\tau$ is tempered. If $\tau \subset I(\omega_{K/E})$ is an irreducible constituent of a reducible principal series, set $\chi = \omega_{K/E},\ \chi^+(\omega) = 1,\ \omega = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$, then from [Kudla 1996, page 86], we can see that

$$I(\omega_{K/E}) = \theta_\psi(\chi^+) \oplus \theta_\psi(\chi^-) \quad \text{where } \chi^- = \chi^+ \otimes \det$$

and $\tau = \theta_\psi(\chi^+) = \Theta_\psi(\chi^+)$, where $\theta_\psi(\chi^+)$ is the theta lift of $\chi^+$ from $\mathrm{O}_{1,1}(E)$ to $\mathrm{SL}_2(E)$. By the see-saw diagram

$$\begin{array}{cccc} \tau & \mathrm{SL}_2(E) & \mathrm{O}_{2,2}(F) & \mathcal{I}(\frac{1}{2}) \\ & & \times & \\ \mathbb{C} & \mathrm{SL}_2(F) & \mathrm{O}_{1,1}(E) & \chi^+ \end{array}$$

where $\mathcal{I}(s)$ is the principal series of $\mathrm{O}_{2,2}(F)$, we have an identity,

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}_{1,1}(E)}\big(\mathcal{I}(\tfrac{1}{2}), \chi^+\big),$$

which is equal to

$$\dim \mathrm{Hom}_{\mathrm{O}_{1,1}(F)}(\chi^+, \mathbb{C}) + \dim \mathrm{Hom}_{\mathrm{O}(V_E)}(\chi^+, \mathbb{C}) + \dim \mathrm{Hom}_{\mathrm{O}(\epsilon V_E)}(\chi^+, \mathbb{C}).$$

If $\chi|_{F^\times} = 1$, then $\dim \mathrm{Hom}_{\mathrm{O}_{1,1}(F)}(\chi^+, \mathbb{C}) = 1$ and $\dim \mathrm{Hom}_{\mathrm{O}_{1,1}(F)}(\chi^-, \mathbb{C}) = 0$. If $\chi = \chi_F \circ N_{E/F}$, then $\dim \mathrm{Hom}_{\mathrm{O}(V_E)}(\chi^+, \mathbb{C}) = 1$. Hence we have the conclusion:

- If $\chi = \omega_{K/E} = \chi_F \circ N_{E/F}$ with $\chi_F^2 = 1$, then

$$\dim \mathrm{Hom}_{\mathrm{O}(\epsilon V_E)}(\chi^+, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_E)}(\chi^+, \mathbb{C}) = 1$$

and

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 3.$$

- If $\chi = \chi_F \circ N_{E/F}$ with $\chi_F^2 = \omega_{E/F}$, then

$$\dim \mathrm{Hom}_{\mathrm{O}(\epsilon V_E)}(\chi^+, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{O}(V_E)}(\chi^-, \mathbb{C})$$

and

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\chi^+), \mathbb{C}) = \dim \mathrm{Hom}_{E^1}(\chi, \mathbb{C}) = 1.$$

- If $\chi$ does not factor through the norm map $N_{E/F}$, but $\chi|_{F^\times} = 1$, then

$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 1.$$

In this case, if $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\chi^+), \mathbb{C}) \neq 0$, then $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\theta_\psi(\chi^-), \mathbb{C})$ is equal to the sum

$$\dim \mathrm{Hom}_{\mathrm{O}_{1,1}(F)}(\chi^+, \det) + \dim \mathrm{Hom}_{\mathrm{O}(V_E)}(\chi^+, \det) + \dim \mathrm{Hom}_{\mathrm{O}(\epsilon V_E)}(\chi^+, \det),$$

which is nonzero if and only if $\chi = \chi_F \circ N_{E/F}$ with $\chi_F^2 = \omega_{E/F}$.

After the discussions for the parameter side in Section 4, we finish the proof of Theorem 1.1.

## 4. The Prasad conjecture for SL(2)

Let us recall a well known result for $\mathrm{SL}_2$.

**Proposition 4.1** [Shelstad 1979]. *Let $\phi : WD_F \to \mathrm{GL}_2(\mathbb{C})$ be an irreducible representation and $\tau = i(\phi) = \mathrm{Ad}(\phi) : WD_F \to \mathrm{PGL}_2(\mathbb{C})$ be the associated discrete series $L$-parameter for $\mathrm{SL}_2$, then there is a short exact sequence of component groups,*

$$1 \to S_\phi \to S_\tau \to I(\phi) \to 1,$$

*where $I(\phi) = \{\chi : F^\times \to \mathbb{C}^\times \mid \chi^2 = 1 \text{ and } \phi \otimes \chi = \phi\}$.*

Assume that $\tau$ is $\mathrm{SL}_2(F)$-distinguished and $\ell \in W_F \setminus W_E$, $\omega_{E/F}(\ell) = -1$. We start to verify the Prasad conjecture for $\mathrm{SL}_2$. The main work here is to choose a proper element $A \in \mathrm{PGL}_2(\mathbb{C})$ such that $\tilde{\phi}(\ell) = A$ and $\tilde{\phi}|_{WD_E} = \phi_\tau$ for a certain Langlands parameter $\tilde{\phi} \in \mathrm{Hom}(WD_F, \mathrm{PGL}_2(\mathbb{C}))$ under the assumption that $\tau$ is $\mathrm{SL}_2(F)$-distinguished. In accordance with the discussions in Section 3, we separate the possible cases for $\tau$ into four parts.

Recall that $F^\times/F^{\times 2} \supset \{1, u, \varpi, u\varpi\}$, $E = F(\sqrt{u})$, $E'' = F(\sqrt{u\varpi})$ and $E' = F(\sqrt{\varpi})$. Let $K = F(\sqrt{u}, \sqrt{\varpi})$ be a biquadratic field extension over $F$ with Galois group $\mathrm{Gal}(K/F) = \langle 1, s, \sigma, s\sigma \rangle$ and Weil group $W_K$. Suppose that $\mathrm{Gal}(K/E) = \langle s \rangle$, $\mathrm{Gal}(K/E'') = \langle s\sigma \rangle$ and $\mathrm{Gal}(K/E') = \langle \sigma \rangle$.

(A) Assume that $\tau \subset \pi|_{\mathrm{SL}_2(E)}$ is a supercuspidal representation of $\mathrm{SL}_2(E)$. If the Langlands parameter of $\tau$,

$$\phi_\tau = i(\mathrm{Ind}_{W_K}^{W_E} \mu) = \omega_{K/E} \oplus \mathrm{Ind}_{W_K}^{W_E}\left(\frac{\mu^s}{\mu}\right)$$

with $\mu|_{E''} = \chi_F \circ N_{E''/F}$, then $\mu(t)\mu^{s\sigma}(t) = \chi_F(t)$ for $t \in W_K$. So

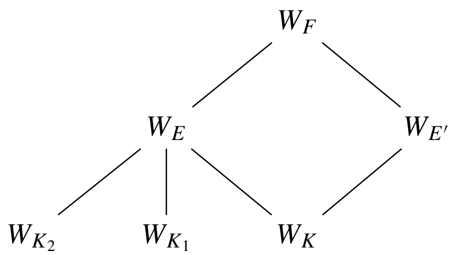
$$\left(\frac{\mu^s}{\mu}\right)^\sigma(t) = \frac{\mu^{s\sigma}(t)}{\mu^\sigma(t)} = \frac{\chi_F(t)}{\mu(t)\mu^\sigma(t)} = \frac{\chi_F(sts^{-1})}{\mu(t)\mu^\sigma(t)} = \frac{\mu^s(t)}{\mu(t)} \quad \text{for } t \in W_K,$$

i.e., $\mu^s/\mu = \chi_{E'} \circ N_{K/E'}$ for a character $\chi_{E'}$ of $E'^\times$.

***The case $p \neq 2$.*** • If $\mu_1^2 = \mathbf{1}$, then the Langlands parameter satisfies

$$\phi_\tau = \omega_{K/E} \oplus \omega_{K_2/E} \oplus \omega_{K_1/E},$$

where each $K_j \neq K$ is a quadratic field extension over $E$:




Set

$$\tilde{\phi} = \omega_{E'/F} \oplus \mathrm{Ind}_{W_{E'}}^{W_F} \chi_{E'}, \tag{4-1}$$

where $E' \neq E$ are two distinct quadratic field extensions over $F$, then $\tilde{\phi}|_{W_E} = \phi_\tau$.

• If $\mu_1^2 \neq \mathbf{1}$, then the Langlands parameter

$$\phi_\tau = \omega_{K/E} \oplus \mathrm{Ind}_{W_K}^{W_E} \frac{\mu^s}{\mu}$$

has a lift $\tilde{\phi}$ defined in (4-1). Moreover, there is one more lift,

$$\tilde{\phi}' = \omega_{E'/F} \oplus \mathrm{Ind}_{W_{E'}}^{W_F} \chi_{E'}^{-1} \quad \text{with} \quad \chi_{E'} \circ N_{K/E'} = \frac{\mu^s}{\mu}$$

since $\mathrm{Ind}_{W_K}^{W_E}(\mu/\mu^s) = \mathrm{Ind}_{W_K}^{W_E}(\mu^s/\mu)$ is irreducible. In the $L$-packet $\Pi_{\phi_\tau}$ containing $\phi_\tau$, set $\phi = \mathrm{Ind}_{W_K}^{W_E} \mu$ and $\phi_\tau = \mathrm{Ad}(\phi)$.

If the component group $S_{\phi_\tau}$ has order 4, then we denote the four characters of $S_{\phi_\tau}$ by $\{\lambda^{++}, \lambda^{--}, \lambda^{-+}, \lambda^{+-}\}$ which corresponds to the $L$-packet

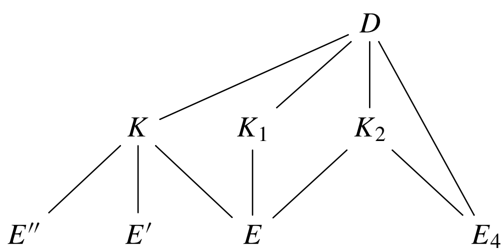
$$\Pi_{\phi_\tau} = \{\tau^{++}, \tau^{--}, \tau^{-+}, \tau^{+-}\}.$$

If the order of $S_{\phi_\tau}$ is 2, then we denote its two characters as $\{\lambda^+, \lambda^-\}$, which corresponds to $\Pi_{\phi_\tau} = \{\tau^+, \tau^-\}$.

- If $\mu_1^2 = \mathbf{1}$, then $|I(\phi)| = 4$, two representations in $\Pi_{\phi_\tau}$ are $\mathrm{SL}_2(F)$-distinguished and of dimension 1, say $\tau^{++}$ and $\tau^{--}$. Since the component group $S_{\tilde{\phi}} = \mu_2 \hookrightarrow S_{\phi_\tau}$ is the diagonal embedding, $\tau^{+-}$ and $\tau^{-+}$ are not $\mathrm{SL}_2(F)$-distinguished, which is compatible with the fact that neither the restricted representation $\lambda^{+-}|_{S_{\tilde{\phi}}}$ nor $\lambda^{-+}|_{S_{\tilde{\phi}}}$ contains the trivial character of $S_{\tilde{\phi}}$, where $\lambda^{+-}$ and $\lambda^{-+}$ correspond to the representations $\tau^{+-}$ and $\tau^{-+}$, respectively.
- If $\mu_1^2 \neq \mathbf{1}$, then $|I(\phi)| = 2$ and only one of them is $\mathrm{SL}_2(F)$-distinguished, say $\tau^+ = \theta_{\psi, V_K, W}((\mu^s/\mu)^+)$. If $\tau^- = \theta_{\psi, \epsilon V_K, W}((\mu^s/\mu)^+)$ corresponds to the nontrivial character of $S_{\phi_\tau}$, denoted by $\lambda^-$, where $\epsilon V_K$ is the 2-dimensional quadratic space $K$ over $E$ with a quadratic form $\epsilon N_{K/E}$, $\epsilon \in E^\times \setminus N_{K/E}(K^\times)$ and the Hasse invariant of $\mathrm{Res}_{E/F}(\epsilon V_K)$ is $-1$, then
$$\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau^-, \mathbb{C}) = 0.$$
Note that $S_{\tilde{\phi}} = \mu_2 \cong S_{\phi_\tau}$, then $\lambda^-|_{S_{\tilde{\phi}}}$ is nontrivial.

***The case p* = 2.** There are some special cases if $p = 2$.

- If $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 4$, then $\mu_1^2 = \mathbf{1}$ and there is a quadratic field extension $D$ over $K$ such that $\chi_K = \omega_{D/K}$ and $D$ is the composite field $KE_4$, where $E_4$ is the quadratic field extension of $F$ corresponding to the quadratic character $\chi'_F/\chi''_F$ where $\mu|_{E'} = \chi'_F \circ N_{E'/F}$, $\mu|_{E''} = \chi''_F \circ N_{E''/F}$ and $E'$ and $E''$ are two distinct quadratic field extensions over $F$, which are different from $E$:




Set $\{1, u, \varpi, d, du, \varpi u, \varpi d, \varpi du\} \subset F^\times/F^{\times 2}$, $E_4 = F(\sqrt{d})$, $K = F(\sqrt{u}, \sqrt{\varpi})$, $K_2 = F(\sqrt{u}, \sqrt{d})$, and $K_1 = F(\sqrt{u}, \sqrt{d\varpi})$. There are four distinct Langlands parameter lifts of $\phi_\tau$:

$$\tilde{\phi}_1 = \omega_{E_4/F} \oplus \omega_{F(\sqrt{\varpi u})/F} \oplus \omega_{F(\sqrt{d\varpi u})/F},$$
$$\tilde{\phi}_2 = \omega_{E_4/F} \oplus \omega_{F(\sqrt{\varpi})/F} \oplus \omega_{F(\sqrt{d\varpi})/F},$$
$$\tilde{\phi}_3 = \omega_{F(\sqrt{du})/F} \oplus \omega_{F(\sqrt{\varpi u})/F} \oplus \omega_{F(\sqrt{d\varpi})/F},$$
$$\tilde{\phi}_4 = \omega_{F(\sqrt{du})/F} \oplus \omega_{F(\sqrt{\varpi})/F} \oplus \omega_{F(\sqrt{\varpi ud})/F},$$

where $\omega_{F(\sqrt{\varpi})/F}$ is the quadratic character associated to the quadratic field extension $F(\sqrt{\varpi})/F$, and similarly for the other quadratic characters $\omega_{F(\sqrt{du})/F}$ and so on.

Since $S_{\tilde{\phi}_i} = S_{\phi_\tau} \cong \mu_2 \times \mu_2$, only $\tau^{++}$ can survive, i.e., the rest of the elements in the $L$-packet $\Pi_{\phi_\tau}$ cannot be $\mathrm{SL}_2(F)$-distinguished.

- If $\dim \mathrm{Hom}_{\mathrm{SL}_2(F)}(\tau, \mathbb{C}) = 1$ and $\pi$ is not dihedral, i.e., $\tau = \pi|_{\mathrm{SL}_2(E)}$ is irreducible, then $\phi_\tau = \phi_\tau^\sigma$. There exists one element $A \in \mathrm{PGL}_2(\mathbb{C})$ such that

$$\phi_\tau(\ell \cdot t \cdot \ell^{-1}) = A \cdot \phi_\tau(t) \cdot A^{-1}$$

for $t \in WD_E$. Set $\tilde{\phi}(\ell) = A$ and $\tilde{\phi}(t) = \phi_\tau(t)$ for $t \in WD_E$. Since $\phi_\tau$ is irreducible, $A$ is unique. Hence $\phi_\tau$ admits a unique lift $\tilde{\phi} : W_F \to \mathrm{PGL}_2(\mathbb{C})$ such that $\tilde{\phi}|_{W_E} = \phi_\tau$.

(B) If $\phi_\tau(t) = \begin{pmatrix} \chi(t)|t|^z & \\ & 1 \end{pmatrix} \in \mathrm{PGL}_2(\mathbb{C})$, then

- if $z = 0$ and $\chi$ is trivial, $\tilde{\phi}(\ell)$ can be chosen as $\begin{pmatrix} \omega_{E/F}(\ell) & \\ & 1 \end{pmatrix} = \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$;

- if $z = 0$, $\chi$ does not factor through the norm $N_{E/F}$ but $\chi|_{F^\times} = 1$, set $\chi = \nu^\sigma/\nu$ for a quadratic character $\nu$ of $E^\times$, then there is only one lift,

$$\tilde{\phi} = i(\mathrm{Ind}_{W_E}^{W_F} \nu);$$

- if $\chi = \chi_F \circ N_{E/F}$, $\chi^2 \neq 1$, then there are two lifts

$$\tilde{\phi}(\ell) = \begin{pmatrix} \chi_F(\ell) & \\ & 1 \end{pmatrix} \text{ or } \begin{pmatrix} -\chi_F(\ell) & \\ & 1 \end{pmatrix}.$$

(C) If $\phi_\tau = \mathrm{Ad}(\mathbf{1} \otimes S_2)$ corresponds to the Steinberg representation $\mathrm{St}_E$ of $\mathrm{SL}_2(E)$, then there is only one lift $\tilde{\phi} = \mathrm{Ad}(\mathbf{1} \otimes S_2) : WD_F \to \mathrm{PGL}_2(\mathbb{C})$.

(D) If $\phi_\tau(t) = \begin{pmatrix} \omega_{K/E}(t) & \\ & 1 \end{pmatrix} \in \mathrm{PGL}_2(\mathbb{C})$, then there are several subcases.

- If $\omega_{K/E} = \chi_F \circ N_{E/F}$ with $\chi_F^2 = \mathbf{1}$, then

$$\tilde{\phi}(\ell) = \begin{pmatrix} \chi_F(\ell) & \\ & 1 \end{pmatrix} \text{ or } \begin{pmatrix} -\chi_F(\ell) & \\ & 1 \end{pmatrix}.$$

  Moreover, $\omega_{K/E}|_{F^\times} = \chi_F^2 = \mathbf{1}$, and $\omega_{K/E} = \nu^\sigma/\nu$ so for a quadratic character $\nu$ of $E^\times$, we may set

$$\tilde{\phi}_3 = i(\mathrm{Ind}_{W_E}^{W_F} \nu) = \omega_{E/F} \oplus \mathrm{Ind}_{W_E}^{W_F}\left(\frac{\nu^\sigma}{\nu}\right).$$

- If $\omega_{K/E} = \chi_F \circ N_{E/F}$ with $\chi_F^2 = \omega_{E/F}$, then there is only one extension

$$\tilde{\phi}(\ell) = \begin{pmatrix} \chi_F(\ell) & \\ & 1 \end{pmatrix}.$$

- If $\omega_{K/E}$ does not factor through the norm map $N_{E/F}$ but $\omega_{K/E}|_{F^\times} = 1$, then

$$\tilde{\phi} = i(\mathrm{Ind}_{W_E}^{W_F} \nu) \quad \text{where} \quad \omega_{K/E} = \nu^\sigma/\nu.$$

Hence, we finish the proof of Theorem 1.1 and Theorem 1.2. □

***Further discussion.*** Inspired by the case that $\tau = \pi|_{\mathrm{SL}_2(E)}$ is an irreducible representation of $\mathrm{SL}_2(E)$, where $\pi$ is a representation of $\mathrm{GL}_2(E)$, we have a certain result of the Prasad conjecture for $G = \mathrm{Sp}_4$.

**Theorem 4.2.** *Let $E$ be a quadratic field extension over a nonarchimedean local field $F$ with characteristic zero. Assume that $\tau$ is an irreducible representation of $\mathrm{Sp}_4(E)$. Let $\pi$ be an irreducible representation of $\mathrm{GSp}_4(E)$ and $\pi|_{\mathrm{Sp}_4(E)} \supset \tau$, then*

(i) *if $\pi$ is tempered and nongeneric, then $\mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C}) = 0$;*

(ii) *if $\pi$ is a generic square-integrable representation of $\mathrm{GSp}_4(E)$ and $\pi|_{\mathrm{Sp}_4(E)}$ is irreducible, then the $L$-packet $\Pi_{\phi_\tau}$ is a singleton and*

$$\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C}) = |F(\phi_\tau)|,$$

*where $F(\phi_\tau) = \{\tilde{\phi} : WD_F \to \mathrm{SO}_5(\mathbb{C}) \mid \tilde{\phi}|_{WD_E} = \phi_\tau\}$ and $|F(\phi_\tau)|$ denotes its cardinality.*

*Proof.* (i) If $\pi$ is tempered and nongeneric, then $\pi = \Theta(\Sigma)$ where $\Sigma$ is an irreducible representation of $\mathrm{GSO}(V_{D_E})$, where $V_{D_E}$ is the nonsplit 4-dimensional quadratic space over $E$ with trivial discriminant and Hasse invariant $-1$. Since $\mathrm{Res}_{E/F} V_{D_E}$ is an 8-dimensional quadratic space over $F$ with trivial discriminant and Hasse invariant $-1$, the conservation relation implies that the theta lift of the trivial representation from $\mathrm{Sp}_4(F)$ to $\mathrm{O}(\mathrm{Res}_{E/F} V_{D_E})$ is zero. Due to the see-saw diagram

$$\begin{array}{cccc} \tau & \mathrm{Sp}_4(E) & \mathrm{O}(\mathrm{Res}_{E/F}(V_{D_E})) & 0 \\ & & \times & \\ \mathbb{C} & \mathrm{Sp}_4(F) & \mathrm{O}(V_{D_E}) & \theta(\tau) \end{array}$$

one has the desired equality, $\mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C}) = 0$.

(ii) By the assumption, $\tau = \pi|_{\mathrm{Sp}_4(E)}$ is a square-integrable representation. Fix $\ell \in W_F \setminus W_E$.

- If the theta lift $\Theta^{2,2}(\pi)$ from $\mathrm{GSp}_4(E)$ to $\mathrm{GSO}(2,2)(E)$ is zero, then one can use a similar method appearing in the proof of [Lu 2017, Theorem 4.2.18(iii)] to obtain the equality

$$\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\pi, \mathbb{C}) = \dim \mathrm{Hom}_{\mathrm{SO}(3,3)(F)}(\Theta^{3,3}(\pi), \mathbb{C}),$$

which is equal to the number

$$\left|\{\chi : F^\times \to \mathbb{C}^\times \mid \mathrm{Hom}_{\mathrm{GSO}(3,3)(F)}(\Theta^{3,3}(\pi), \chi \circ \lambda) \neq 0\}\right|,$$

where $\Theta^{3,3}(\pi)$ is the theta lift of $\pi$ from $\mathrm{GSp}_4(E)$ to $\mathrm{GSO}(3,3)(E)$ and $\lambda$ is the similitude character of the group $\mathrm{GSO}(3,3)(F)$. Therefore, the dimension

$\dim \mathrm{Hom}_{\mathrm{Sp}_4(E)}(\tau, \mathbb{C}) = 1$ if and only if the Langlands parameter $\phi_\pi$ of $\pi$ is conjugate-self-dual, i.e., $\phi_\pi^\vee = \phi_\pi^\sigma$.

On the parameter side, $\phi_\tau : WD_E \to \mathrm{PGSp}_4(\mathbb{C}) = \mathrm{SO}_5(\mathbb{C})$ is irreducible and $\phi_\tau \cong \phi_\tau^\vee \cong \phi_\tau^\sigma$. There exists a unique element $A \in \mathrm{SO}_5(\mathbb{C})$ such that

$$\phi_\tau(\ell \cdot t \cdot \ell^{-1}) = A \cdot \phi_\tau(t) \cdot A^{-1}$$

for $t \in WD_E$. Set $\tilde{\phi}(\ell) = A$ and $\tilde{\phi}(t) = \phi_\tau(t)$ for $t \in WD_E$. Then $\tilde{\phi}$ is what we want.

- If $\Theta^{2,2}(\pi) \neq 0$, then $\phi_\pi = \phi_1 \oplus \phi_2$ where $\phi_i : WD_E \to \mathrm{GL}_2(\mathbb{C})$ is irreducible and $\phi_1 \neq \phi_2$. Moreover, $\phi_\tau = \mathbf{1} \oplus (\phi_1^\vee \otimes \phi_2)$; see [Gan and Takeda 2010, page 3008]. Let $\Sigma$ be the irreducible representation of $\mathrm{GSO}(2,2)(E)$ satisfying $\theta_\psi(\Sigma) = \pi$, then $\Sigma|_{\mathrm{SO}(2,2)(E)}$ is irreducible since $\pi|_{\mathrm{Sp}_4(E)}$ is irreducible. Using a similar method appearing in [Lu 2017, Theorem 4.2.18(ii)], one can get that the dimension

$$\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C})$$

has an upper bound

$$\dim \mathrm{Hom}_{\mathrm{SO}(3,3)(F)}(\Theta^{3,3}(\pi), \mathbb{C}) + \dim \mathrm{Hom}_{\mathrm{SO}(4,0)(F)}(\Sigma, \mathbb{C}) \tag{4-2}$$

and a lower bound

$$\sum_X \dim \mathrm{Hom}_{\mathrm{SO}(X,F)}(\Sigma, \mathbb{C}), \tag{4-3}$$

where $X$ runs over all elements in the kernel $\ker\{H^1(F, \mathrm{O}(4)) \to H^1(E, \mathrm{O}(4))\}$. We will show that both the lower bound (4-3) and the upper bound (4-2) are equal to 2 if $\pi|_{\mathrm{Sp}_4(E)}$ is an irreducible $\mathrm{Sp}_4(F)$-distinguished representation. Then

$$\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C}) = 2.$$

There are two subcases.

(a) If $\phi_1^\vee = \phi_1^\sigma$, then $\phi_1^\vee \neq \phi_2^\sigma$, otherwise $\phi_1 = \phi_2$, which contradicts $\phi_1 \neq \phi_2$. Since $\phi_1$ is irreducible, the Langlands parameter $\phi_1$ is either conjugate-orthogonal or conjugate-symplectic, but cannot be both. Note that there is an equality

$$\dim \mathrm{Hom}_{\mathrm{SO}(3,3)(F)}(\Theta^{3,3}(\pi), \mathbb{C}) = \big|\{\chi : F^\times \to \mathbb{C}^\times \mid \mathrm{Hom}_{\mathrm{GSO}(3,3)(F)}(\Theta^{3,3}(\pi), \chi \circ \lambda) \neq 0\}\big|.$$

We have a similar result for $\dim \mathrm{Hom}_{\mathrm{SO}(4,0)(F)}(\Sigma, \mathbb{C})$ and $\dim \mathrm{Hom}_{\mathrm{SO}(2,2)(F)}(\Sigma, \mathbb{C})$. If $\phi_2^\vee = \phi_2^\sigma$ is conjugate-self-dual with the same sign as $\phi_1$, then

$$\dim \mathrm{Hom}_{\mathrm{Sp}_4(E)}(\tau, \mathbb{C}) = 2.$$

Otherwise, $\tau$ is not $\mathrm{Sp}_4(F)$-distinguished.

On the parameter side, $1/\det\phi_1 = (\det\phi_1)^\sigma$. Without loss of generality, suppose that $\phi_1$ is conjugate-orthogonal, i.e., $\det\phi_1 = \nu^\sigma/\nu = \det\phi_2$, then $\nu\otimes\phi_j$ is $\mathrm{Gal}(E/F)$-invariant. For each $j$, there exists a parameter $\tilde{\phi}_j : WD_F \to \mathrm{GL}_2(\mathbb{C})$ such that $\tilde{\phi}_j|_{WD_E} = \phi_j \otimes \nu$. Set $\rho_1 = \tilde{\phi}_1 \oplus \tilde{\phi}_2$ and $\rho_2 = \tilde{\phi}_1 \oplus \tilde{\phi}_2\omega_{E/F}$. Let $i : \mathrm{GSp}_4(\mathbb{C}) \to \mathrm{SO}_5(\mathbb{C})$ be the natural projection map. Then the parameters $i(\rho_1)$ and $i(\rho_2)$ are what we want.

(b) If $\phi_1^\vee = \phi_2^\sigma$, then $\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C}) = 2$ since the upper bound (4-2) is 2 and the lower bound (4-3) is at least 2. On the parameter side, $\phi_\tau = \mathbf{1} \oplus (\phi_2^\sigma \otimes \phi_2)$ is $\mathrm{Gal}(E/F)$-invariant. There exist two natural parameters $\tilde{\phi}_j : WD_F \to \mathrm{GL}_5(\mathbb{C})$ such that $\tilde{\phi}_j|_{WD_E} = \phi_\tau$, which are $\omega_{E/F} \oplus \mathrm{As}^+(\phi_2)$ and $\omega_{E/F} \oplus \mathrm{As}^-(\phi_2)$, where $\mathrm{As}^\pm(\phi_2)$ are the Asai lifts of $\phi_2$; see [Gan et al. 2012, §7]. Then the images of $\tilde{\phi}_j$ lie in $\mathrm{SO}_5(\mathbb{C})$. Therefore, we have finished the proof. □

**Remark 4.3.** If $\tau = \pi|_{\mathrm{Sp}_4(E)}$ is irreducible, one can also use the method appearing in [Anandavardhanan and Prasad 2003] directly to get that the dimension $\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C})$ equals the sum

$$\sum_{\chi : F^\times/(F^\times)^2 \to \mathbb{C}^\times} \dim \mathrm{Hom}_{\mathrm{GSp}_4(F)}(\pi, \chi). \tag{4-4}$$

Combining this with the results in [Lu 2017, Theorem 4.2.18], we can obtain $\dim \mathrm{Hom}_{\mathrm{Sp}_4(F)}(\tau, \mathbb{C})$ if $\pi$ is tempered.

**Remark 4.4.** Let $U_2(D)$ be the unique inner form of $\mathrm{Sp}_4(F)$ defined over $F$. Suppose that $\pi$ is a generic representation of $\mathrm{GSp}_4(E)$. Thanks to [Beuzart-Plessis 2017, Theorem 1], if $\pi|_{\mathrm{Sp}_4(E)} = \tau$ is an irreducible square-integrable representation of $\mathrm{Sp}_4(E)$ and $\Theta^{2,2}(\pi)$ is 0, then

$$\dim \mathrm{Hom}_{U_2(D)}(\tau, \mathbb{C}) = 1.$$

## Acknowledgements

The author thanks Wee Teck Gan for his guidance and numerous discussions when he was doing his Ph.D. study at National University of Singapore. He would like to thank Dipendra Prasad for useful discussions as well. He also thanks the anonymous referees for the careful reading and helpful comments, especially for pointing out the inaccurate statement in Theorem 1.1 in the earlier version.

## References

[Anandavardhanan and Prasad 2003] U. K. Anandavardhanan and D. Prasad, "Distinguished representations for SL(2)", *Math. Res. Lett.* **10**:5-6 (2003), 867–878. MR Zbl

[Anandavardhanan and Prasad 2006] U. K. Anandavardhanan and D. Prasad, "On the SL(2) period integral", *Amer. J. Math.* **128**:6 (2006), 1429–1453. MR Zbl

[Anandavardhanan and Prasad 2013] U. K. Anandavardhanan and D. Prasad, "A local-global question in automorphic forms", *Compos. Math.* **149**:6 (2013), 959–995. MR Zbl

[Anandavardhanan and Prasad 2016] U. K. Anandavardhanan and D. Prasad, "Distinguished representations for SL($n$)", 2016. To appear in *Math. Res. Lett.* arXiv

[Atobe and Gan 2017] H. Atobe and W. T. Gan, "On the local Langlands correspondence and Arthur conjecture for even orthogonal groups", *Represent. Theory* **21** (2017), 354–415. MR Zbl

[Beuzart-Plessis 2017] R. Beuzart-Plessis, "On distinguished square-integrable representations for Galois pairs and a conjecture of Prasad", preprint, 2017. arXiv

[Gan and Ichino 2014] W. T. Gan and A. Ichino, "Formal degrees and local theta correspondence", *Invent. Math.* **195**:3 (2014), 509–672. MR Zbl

[Gan and Takeda 2010] W. T. Gan and S. Takeda, "The local Langlands conjecture for Sp(4)", *Int. Math. Res. Not.* **2010**:15 (2010), 2987–3038. MR Zbl

[Gan and Takeda 2016a] W. T. Gan and S. Takeda, "On the Howe duality conjecture in classical theta correspondence", pp. 105–117 in *Advances in the theory of automorphic forms and their $L$-functions* (Vienna, 2013), edited by D. Jiang et al., Contemp. Math. **664**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Gan and Takeda 2016b] W. T. Gan and S. Takeda, "A proof of the Howe duality conjecture", *J. Amer. Math. Soc.* **29**:2 (2016), 473–493. MR Zbl

[Gan et al. 2012] W. T. Gan, B. H. Gross, and D. Prasad, "Symplectic local root numbers, central critical $L$-values, and restriction problems in the representation theory of classical groups", pp. 1–109 in *Sur les conjectures de Gross et Prasad, I*, Astérisque **346**, Société Mathématique de France, Paris, 2012. MR Zbl arXiv

[Kudla 1984] S. S. Kudla, "Seesaw dual reductive pairs", pp. 244–268 in *Automorphic forms of several variables* (Katata, 1983), edited by Y. Morita, Progr. Math. **46**, Birkhäuser, Boston, 1984. MR Zbl

[Kudla 1996] S. S. Kudla, "Notes on the local theta correspondence", lecture notes, 1996, Available at http://www.math.toronto.edu/skudla/castle.pdf.

[Kudla and Rallis 2005] S. S. Kudla and S. Rallis, "On first occurrence in the local theta correspondence", pp. 273–308 in *Automorphic representations, $L$-functions and applications: progress and prospects* (Columbus, OH, 2003), edited by J. W. Cogdell et al., Ohio State Univ. Math. Res. Inst. Publ. **11**, de Gruyter, Berlin, 2005. MR Zbl

[Lu 2017] H. Lu, GSp(4)*-period problems over a quadratic field extension*, Ph.D. thesis, National University of Singapore, 2017, Available at http://scholarbank.nus.edu.sg/handle/10635/135863.

[Prasad 1993] D. Prasad, "On the local Howe duality correspondence", *Int. Math. Res. Not.* **1993**:11 (1993), 279–287. MR Zbl

[Prasad 2015] D. Prasad, "A 'relative' local Langlands correspondence", preprint, 2015. arXiv

[Sakellaridis and Venkatesh 2017] Y. Sakellaridis and A. Venkatesh, *Periods and harmonic analysis on spherical varieties*, Astérisque **396**, Société Mathématique de France, Paris, 2017. Zbl

[Shelstad 1979] D. Shelstad, "Notes on $L$-indistinguishability (based on a lecture of R. P. Langlands)", pp. 193–203 in *Automorphic forms, representations and $L$-functions, II* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR Zbl

[Sun and Zhu 2015] B. Sun and C.-B. Zhu, "Conservation relations for local theta correspondence", *J. Amer. Math. Soc.* **28**:4 (2015), 939–983. MR Zbl

[Waldspurger 1990] J.-L. Waldspurger, "Démonstration d'une conjecture de dualité de Howe dans le cas $p$-adique, $p \neq 2$", pp. 267–324 in *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, I* (Ramat Aviv, 1989), edited by S. Gelbert et al., Israel Math. Conf. Proc. **2**, Weizmann, Jerusalem, 1990. MR Zbl

Received September 12, 2017. Revised February 6, 2018.

HENGFEI LU
SCHOOL OF MATHEMATICS
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
SCHOOL OF MATHEMATICS
COLABA
MUMBAI
INDIA
hengfei@math.tifr.res.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 295, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.295.499

# CONVEXITY OF LEVEL SETS AND A TWO-POINT FUNCTION

BEN WEINKOVE

**We establish a maximum principle for a two-point function in order to analyze the convexity of level sets of harmonic functions. We show that this can be used to prove a strict convexity result involving the smallest principal curvature of the level sets.**

## 1. Introduction

The study of the convexity of level sets of solutions to elliptic PDEs has a long history, starting with the well-known result that the level curves of the Green's function of a convex domain $\Omega$ in $\mathbb{R}^2$ are convex [Ahlfors 1973]. Gabriel [1957] proved the analogous result in three dimensions and this was extended by Lewis [1977] and later Caffarelli and Spruck [1982] to higher dimensions and more general elliptic PDEs. These results show that for a large class of PDEs, there is a principle that convexity properties of the boundary of the domain $\Omega$ imply convexity of the level sets of the solution $u$.

There are several approaches to these kinds of convexity results; see for example [Kawohl 1985, Section III.11]. One is the "macroscopic" approach, which uses a globally defined function of two points $x$, $y$ (which could be far apart) such as $u\big(\frac{1}{2}(x+y)\big) - \min(u(x), u(y))$. Another is the "microscopic" approach, which computes with functions of the principal curvatures of the level sets at a single point. This is often used together with a constant rank theorem. There is now a vast literature on these and closely related results, see for example [Alvarez et al. 1997; Bian and Guan 2009; Bianchini et al. 2009; Borell 1982; Brascamp and Lieb 1976; Caffarelli and Friedman 1985; Caffarelli et al. 2007; Diaz and Kawohl 1993; Hamel et al. 2016; Korevaar 1983; 1990; Korevaar and Lewis 1987; Rosay and Rudin 1989; Shiffman 1956; Singer et al. 1985; Székelyhidi and Weinkove 2016; Wang 2014].

It is natural to ask whether these ideas can be extended to cases where the boundary of the domain is *not* convex. Are the level sets of the solution at least as

The author thanks G. Székelyhidi for some helpful discussions and the referee for useful comments. Supported in part by National Science Foundation grant DMS-1406164.
*MSC2010:* 31B05, 35J05.
*Keywords:* convexity, two point function, level sets, principal curvature, maximum principle, harmonic functions.

convex as the boundary in some appropriate sense? In this short note we introduce a global "macroscopic" function of two points which gives a kind of measure of convexity and makes sense for nonconvex domains. Our function

$$(Du(y) - Du(x)) \cdot (y - x) \tag{1-1}$$

is evaluated at two points $x$, $y$, which are constrained to lie on the same level set of $u$. Under suitable conditions, a level set of $u$ is convex if and only if this quantity has the correct sign on that level set. We prove a maximum principle for this function using the method of Rosay and Rudin [1989], who considered a different two-point function

$$\tfrac{1}{2}(u(x) + u(y)) - u\Big(\frac{x+y}{2}\Big). \tag{1-2}$$

In addition, we show that our "macroscopic" approach can be used to prove a "microscopic" result. Namely, we localize our function and show that it gives another proof of a result of Chang, Ma, and Yang [Chang et al. 2010] on the principal curvatures of the level sets of a harmonic function $u$. In this paper, we consider only the case of harmonic functions. However, we expect that our techniques extend to some more general types of PDEs.

We now describe our results more precisely. Let $\Omega_0$ and $\Omega_1$ be bounded domains in $\mathbb{R}^n$ with $\overline{\Omega}_1 \subset \Omega_0$. Define $\Omega = \Omega_0 \setminus \Omega_1$. Assume that $u \in C^1(\overline{\Omega})$ satisfies

$$\Delta u = 0 \text{ in } \Omega = \Omega_0 \setminus \overline{\Omega}_1, \quad u = 0 \text{ on } \partial\Omega_0, \quad u = 1 \text{ on } \partial\Omega_1, \tag{1-3}$$

and

$$Du \text{ is nowhere vanishing in } \Omega. \tag{1-4}$$

It is well known that (1-4) is satisfied if $\Omega_0$ and $\Omega_1$ are both starshaped with respect to some point $p \in \Omega_1$. A special case of interest is when both $\Omega_0$ and $\Omega_1$ are convex, but this is not required for our main result.

To introduce our two-point function, first fix a smooth function $\psi : [0, \infty) \to \mathbb{R}$ satisfying

$$\psi'(t) - 2|\psi''(t)|t \geq 0. \tag{1-5}$$

For example, we could take $\psi(t) = at$ for $a \geq 0$. Then define

$$Q(x, y) = (Du(y) - Du(x)) \cdot (y - x) + \psi(|y - x|^2) \tag{1-6}$$

restricted to $(x, y)$ in

$$\Sigma = \{(x, y) \in \overline{\Omega} \times \overline{\Omega} \mid u(x) = u(y)\}.$$

Comparing with the Rosay–Rudin function (1-2), note that the function $Q(x, y)$ does not require $\frac{1}{2}(x + y) \in \overline{\Omega}$ and makes sense whether or not $\partial\Omega_0$ or $\partial\Omega_1$ are convex. Taking $\psi = 0$, the level set $\{u = c\}$ is convex if and only if the quantity $Q$ is nonpositive on $\{u = c\}$. If $\psi(t) = at$ for $a > 0$ then $Q \leq 0$ implies strict convexity of the level set. More generally $Q$ gives quantitative information about the convexity of the level sets $\{u = c\}$, relative to the gradient $Du$.

We also remark that the function (1-6) looks formally similar to the two-point function of Andrews and Clutterbuck [2011], a crucial tool in their proof of the fundamental gap conjecture. However, here $x$ and $y$ are constrained to lie on the same level set of $u$ and so the methods of this paper are quite different.

Our main result is the following:

**Theorem 1.1.** *$Q$ does not attain a strict maximum at a point in the interior of $\Sigma$.*

Roughly speaking, this result says that the level sets $\{u = c\}$ for $0 \leq c \leq 1$ are "the least convex" when $c = 0$ or $c = 1$. As mentioned above, the result holds even in the case that $\partial\Omega_0$ and $\partial\Omega_1$ are nonconvex.

The proof of Theorem 1.1 follows quite closely the paper of Rosay and Rudin [1989]. Indeed a key tool of [Rosay and Rudin 1989] is Lemma 2.1 below, which gives a map from points $x$ to points $y$ with the property that $x$, $y$ lie on the same level set.

Next we localize our function (1-6) to prove a strict convexity result on the level sets of $u$. If we assume now that $\partial\Omega_0$ and $\partial\Omega_1$ are strictly convex, we can apply the technique of Theorem 1.1 to obtain an alternative proof of the following result of Chang, Ma, and Yang [Chang et al. 2010].

**Theorem 1.2.** *Assume in addition that $\partial\Omega_0$ and $\partial\Omega_1$ are strictly convex and $C^2$. Then the quantity $|Du|\kappa_1$ attains its minimum on the boundary of $\Omega$, where $\kappa_1$ is the smallest principal curvature of the level sets of $u$.*

Note that many other strict convexity results of this kind are proved in [Chang et al. 2010; Jost et al. 2012; Longinetti 1983; Ma et al. 2010; 2011; Ortel and Schneider 1983; Zhang and Zhang 2013].

## 2. Proof of Theorem 1.1

First we assume that $n$ is even. We suppose for a contradiction that $Q$ attains a maximum at an interior point, and assume that $\sup_\Sigma Q > \sup_{\partial\Sigma} Q$. Then we may choose $\delta > 0$ sufficiently small so that

$$Q_\delta(x, y) = Q(x, y) + \delta|x|^2$$

still attains a maximum at an interior point.

We use a lemma from [Rosay and Rudin 1989]. Suppose $(x_0, y_0)$ is an interior point with $u(x_0) = u(y_0)$. We may assume that $Du(x_0)$ and $Du(y_0)$ are nonzero vectors. Let $L$ be an element of $\mathrm{O}(n)$ with the property that

$$L(Du(x_0)) = c\,Du(y_0) \quad \text{for } c = |Du(x_0)|/|Du(y_0)|. \tag{2-1}$$

Note that there is some freedom in the definition of $L$. We will make a specific choice later. Rosay and Rudin [1989, Lemma 1.3] show the following — it is a special case of the lemma:

**Lemma 2.1.** *There exists a real analytic function $\alpha(w) = O(|w|^3)$ such that for all $w \in \mathbb{R}^n$ sufficiently close to the origin,*

$$u(x_0 + w) = u(y_0 + cLw + f(w)\xi + \alpha(w)\xi), \quad \text{where } \xi = \frac{Du(y_0)}{|Du(y_0)|}, \tag{2-2}$$

*where $f$ is a harmonic function defined in a neighborhood of the origin in $\mathbb{R}^n$, given by*

$$f(w) = \frac{1}{|Du(y_0)|}(u(x_0 + w) - u(y_0 + cLw)). \tag{2-3}$$

*Proof of Lemma 2.1.* We include the brief argument here for the sake of completeness. Define a real analytic map $G$ which takes $(w, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ sufficiently close to the origin to

$$G(w, \alpha) = u(y_0 + cLw + f(w)\xi + \alpha\xi) - u(x_0 + w),$$

for $c$, $L$, $\xi$, and $f$ defined by (2-1), (2-2), and (2-3). Note that $G(0, 0) = 0$ and, by the definition of $\xi$,

$$\frac{\partial G}{\partial \alpha}(0, 0) = D_i u(y_0)\xi_i = |Du(y_0)| > 0,$$

where here and henceforth we are using the convention of summing repeated indices.

Hence by the implicit function theorem there exists a real analytic map $\alpha = \alpha(w)$ defined in a neighborhood $U$ of the origin in $\mathbb{R}^n$ to $\mathbb{R}$ with $\alpha(0) = 0$ such that $G(w, \alpha(w)) = 0$ for all $w \in U$. It only remains to show that $\alpha(w) = O(|w|^3)$.

Write $y = y_0 + cLw + f(w)\xi + \alpha(w)\xi$, $x = x_0 + w$, and $L = (L_{ij})$ so that $L_{ij} D_j u(x_0) = c D_i u(y_0)$ and $cL_{ij} D_i u(y_0) = D_j u(x_0)$. Then at $w \in U$,

$$\begin{aligned} 0 &= \frac{\partial G}{\partial w_j} \\ &= D_i u(y)\left(cL_{ij} + \frac{(D_j u(x) - cD_k u(y_0 + cLw)L_{kj})}{|Du(y_0)|}\xi_i + \frac{\partial \alpha}{\partial w_j}\xi_i\right) - D_j u(x), \end{aligned} \tag{2-4}$$

and evaluating at $w = 0$ gives $0 = |Du(y_0)|\,\partial\alpha/\partial w_j(0)$ and hence $\partial\alpha/\partial w_j(0) = 0$ for all $j$.

Differentiating (2-4) and evaluating at $w=0$, we obtain for all $j$, $\ell$,

$$\begin{aligned}0 &= \frac{\partial^2 G}{\partial w_\ell \partial w_j}\\ &= D_k D_i u(y_0) c^2 L_{ij} L_{k\ell} - D_\ell D_j u(x_0)\\ &\qquad + D_i u(y_0)\biggl(\frac{(D_\ell D_j u(x_0) - c^2 D_m D_k u(y_0) L_{kj} L_{m\ell})}{|Du(y_0)|}\xi_i + \frac{\partial^2 \alpha}{\partial w_\ell \partial w_j}(0)\xi_i\biggr)\\ &= |Du(y_0)|\frac{\partial^2 \alpha}{\partial w_\ell \partial w_j}(0).\end{aligned}$$

Hence $\alpha(w) = O(|w|^3)$, as required. □

Now assume that $Q_\delta$ achieves a maximum at the interior point $(x_0, y_0)$. Write $x = x_0 + w = (x_1, \dots, x_n)$ and $y = y_0 + cLw + f(w)\xi + \alpha(w)\xi = (y_1, \dots, y_n)$ and

$$F(w) = Q_\delta(x, y) = Q(x_0 + w, y_0 + cLw + f(w)\xi + \alpha(w)\xi) + \delta|x_0 + w|^2.$$

To prove the lemma it suffices to show that $\Delta_w F(0) > 0$, where we write $\Delta_w = \sum_j \partial^2/\partial w_j^2$. Observe that

$$\Delta_w x(0) = 0 = \Delta_w y(0).$$

Hence, evaluating at 0, we get

$$\begin{aligned}\Delta_w F &= \sum_j \biggl(\frac{\partial^2}{\partial w_j^2}(D_i u(y) - D_i u(x))\biggr)(y_i - x_i)\\ &\qquad + 2\frac{\partial}{\partial w_j}(D_i u(y) - D_i u(x))\frac{\partial}{\partial w_j}(y_i - x_i) + \sum_j \frac{\partial^2}{\partial w_j^2}\psi(|y-x|^2) + 2n\delta.\end{aligned}$$

First we compute

$$\begin{aligned}\sum_j \frac{\partial^2}{\partial w_j^2}\psi(|y-x|^2) &= 2\psi' \sum_{i,j}(cL_{ij} - \delta_{ij})^2 + 4\psi'' \sum_j \biggl(\sum_i (y_i - x_i)(cL_{ij} - \delta_{ij})\biggr)^2\\ &\geq 2\psi' \sum_{i,j}(cL_{ij} - \delta_{ij})^2 - 4|\psi''||y-x|^2 \sum_{i,j}(cL_{ij} - \delta_{ij})^2 \geq 0\end{aligned}$$

using the Cauchy–Schwarz inequality and the condition (1-5).

Next, at $w=0$,

$$\frac{\partial}{\partial w_j} D_i u(y) = D_k D_i u(y) \frac{\partial y_k}{\partial w_j} = c D_k D_i u(y) L_{kj},$$

$$\sum_j \frac{\partial^2}{\partial w_j^2} D_i u(y) = D_\ell D_k D_i u(y) \frac{\partial y_k}{\partial w_j}\frac{\partial y_\ell}{\partial w_j} = c^2 D_\ell D_k D_i u(y) L_{kj} L_{\ell j} = 0,$$

$$\frac{\partial}{\partial w_j} D_i u(x) = D_j D_i u(x), \qquad \sum_j \frac{\partial^2}{\partial w_j^2} D_i u(x) = D_j D_j D_i u(x) = 0,$$

where for the second line we used the fact that $\Delta_w y(0)=0$ and $L_{kj}L_{\ell j}D_\ell D_k u = \Delta u = 0$. Hence, combining the above,

$$\begin{aligned}\Delta_w F &> 2(cD_kD_iu(y)L_{kj} - D_jD_iu(x))(cL_{ij}-\delta_{ij})\\ &= 2c^2\Delta u(y) - 2cL_{ki}D_kD_iu(y) - 2cL_{ij}D_jD_iu(x) + 2\Delta u(x)\\ &= -2cL_{ki}D_kD_iu(y) - 2cL_{ij}D_jD_iu(x).\end{aligned}$$

Now we use the fact that $n$ is even, and we make an appropriate choice of $L$ following [Rosay and Rudin 1989, Lemma 4.1(a)]. Namely, after making an orthonormal change of coordinates, we may assume, without loss of generality that $Du(x_0)/|Du(x_0)|$ is $e_1$, and

$$Du(y_0)/|Du(y_0)| = \cos\theta\, e_1 + \sin\theta\, e_2,$$

for some $\theta \in [0, 2\pi)$. Here we are writing $e_1 = (1, 0, \ldots 0)$ and $e_2 = (0, 1, 0, \ldots)$, etc., for the standard unit basis vectors in $\mathbb{R}^n$. Then define the isometry $L$ by

$$L(e_i) = \begin{cases} \cos\theta\, e_i + \sin\theta\, e_{i+1} & \text{for } i = 1, 3, \ldots, n-1,\\ -\sin\theta\, e_{i-1} + \cos\theta\, e_i & \text{for } i = 2, 4, \ldots, n.\end{cases}$$

In terms of entries of the matrix $(L_{ij})$, this means that $L_{kk} = \cos\theta$ for $k = 1, \ldots, n$ and for $\alpha = 1, 2, \ldots, \frac{1}{2}n$, we have

$$L_{2\alpha-1,2\alpha} = -\sin\theta, \quad L_{2\alpha,2\alpha-1} = \sin\theta,$$

with all other entries zero. Then

$$\begin{aligned}\text{(2-5)}\quad \sum_{i,k} L_{ki}D_kD_iu(y) &= \sum_{k=1}^{n} L_{kk}D_kD_ku(y) + \sum_{\alpha=1}^{n/2}(L_{2\alpha-1,2\alpha} + L_{2\alpha,2\alpha-1})D_{2\alpha-1}D_{2\alpha}u(y)\\ &= (\cos\theta)\Delta u(y) = 0.\end{aligned}$$

Similarly $\sum_{i,k} L_{ki}D_kD_iu(x) = 0$. This completes the proof of Theorem 1.1 in the case of $n$ even.

For $n$ odd, we argue in the same way as in [Rosay and Rudin 1989]. Let $L$ be an isometry of the even-dimensional $\mathbb{R}^{n+1}$, defined in the same way as above, but now

$$L(Du(x_0), 0) = (c(Du)(y_0), 0).$$

In Lemma 2.1, replace $w \in \mathbb{R}^n$ by $w \in \mathbb{R}^{n+1}$. Define $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ to be the projection $(w_1, \dots, w_{n+1}) \mapsto (w_1, \dots, w_n)$ and replace (2-2) and (2-3) by

$$u(x_0 + \pi(w)) = u\big(y_0 + c\pi(Lw) + f(w)\xi + \alpha(w)\xi\big), \tag{2-6}$$

where $\xi = Du(y_0)/|Du(y_0)|$ and $f$ is given by

$$f(w) = \frac{1}{|Du(y_0)|}\big(u(x_0 + \pi(w)) - u(y_0 + c\pi(Lw))\big). \tag{2-7}$$

As in [Rosay and Rudin 1989], note that if $g : \mathbb{R}^n \to \mathbb{R}$ is harmonic in $\mathbb{R}^n$ then $w \mapsto g(\pi(Lw))$ is harmonic in $\mathbb{R}^{n+1}$. In particular, $f$ is harmonic in a neighborhood of the origin in $\mathbb{R}^{n+1}$. The function $G$ above becomes $G(w, \alpha) = u(y_0 + c\pi(Lw) + f(w)\xi + \alpha\xi) - u(x_0 + \pi(w))$ with $w \in \mathbb{R}^{n+1}$, and we make similar changes to $F$. It is straightforward to check that the rest of the proof goes through. □

**Remark 2.2.** The proof of Theorem 1.1 also shows that when $\psi = 0$ the quantity $Q(x, y)$ does not attain a strict interior *minimum*.

## 3. Global to infinitesimal

Here we give a proof of Theorem 1.2 using the quantity $Q$. We first claim that, for $x \in \Omega$ and $a > 0$,

$$(Du(y) - Du(x)) \cdot (y - x) + a|y - x|^2 \le O(|y - x|^3) \quad \text{for } y \sim x,\ u(x) = u(y)$$

if and only if

$$(\kappa_1 |Du|)(x) \ge a.$$

Indeed, to see this, first choose coordinates such that at $x$ we have $Du = (0, \dots, 0, D_n u)$ and $(D_i D_j u)_{1 \le i, j \le n-1}$ is diagonal with

$$D_1 D_1 u \ge \cdots \ge D_{n-1} D_{n-1} u.$$

For the "if" direction of the claim, choose $y(t) = x + te_1 + O(t^2)$ such that $u(x) = u(y(t))$, for $t$ small. By Taylor's theorem,

$$\big(Du(y(t)) - Du(x)\big) \cdot (y(t) - x) + a|y(t) - x|^2 = t^2 D_1 D_1 u(x) + at^2 + O(t^3),$$

giving $D_1 D_1 u(x) \le -a$, which is the same as $|Du|\kappa_1 \ge a$. Indeed, from a well-known and elementary calculation (see for example [Chang et al. 2010, § 2]),

$$\kappa_1 = \frac{-D_1 D_1 u}{|Du|}$$

at $x$. Hence $|Du|\kappa_1 \geq a$. The "only if" direction of the claim follows similarly.

We will make use of this correspondence in what follows.

*Proof of Theorem 1.2.* By assumption, $\kappa_1|Du| \geq a > 0$ on $\partial\Omega$. It follows from Theorem 1.1 and the discussion above that the level sets of $u$ are all strictly convex. Assume for a contradiction that $\kappa_1|Du|$ achieves a strict (positive) minimum at a point $x_0$ in the interior of $\Omega$, say

$$(\kappa_1|Du|)(x_0) = a - \eta > 0 \text{ for some } \eta > 0. \tag{3-1}$$

We may assume without loss of generality that $\eta < \frac{1}{6}a$. Indeed, if not then if $x_0$ lies on the level set $\{u = c\}$ for some $c \in (0, 1)$ we can replace $\Omega$ by a convex ring $\{c_0 < u < c_1\}$ for $c_0, c_1$ with $0 \leq c_0 < c < c_1 \leq 1$. We still denote by $a$ the minimum value of $\kappa_1|Du|$ on the boundary of this new $\Omega$. For appropriately chosen $c_0, c_1$ we have (3-1) and $\eta < \frac{1}{6}a$. This changes the boundary conditions on $\partial\Omega_0$ and $\partial\Omega_1$ to $u = c_0$ and $u = c_1$, but this will not affect any of the arguments.

Pick $\varepsilon > 0$ sufficiently small, so that the distance from $x_0$ to the boundary of $\Omega$ is much larger than $\varepsilon$, and in addition, so that $\varepsilon^{1/3} \ll \eta$.

Consider the quantity

$$Q(x, y) = (Du(y) - Du(x)) \cdot (y - x) + a|y - x|^2 - \frac{a}{6\varepsilon^2}|y - x|^4,$$

and restrict to the set

$$\Sigma^\varepsilon = \{(x, y) \in \bar{\Omega} \times \bar{\Omega} \mid u(x) = u(y), |y - x| \leq \varepsilon\}.$$

Suppose that $Q$ attains a maximum on $\Sigma^\varepsilon$ at a point $(x, y)$. First assume that $(x, y)$ lies in the boundary of $\Sigma^\varepsilon$. There are two possible cases:

(1) If $x, y \in \Sigma^\varepsilon$ with $x$ and $y$ in $\partial\Omega$ (note that since $u(x) = u(y)$, if one of $x, y$ is a boundary point then so is the other), then since $\kappa_1|Du| \geq a$ on $\partial\Omega$ we have

$$(Du(y) - Du(x)) \cdot (y - x) + a|y - x|^2 \leq O(\varepsilon^3).$$

Hence in this case $Q(x, y) \leq O(\varepsilon^3)$.

(2) If $|y - x| = \varepsilon$ then since $\kappa_1|Du| \geq a - \eta$ everywhere,

$$Q(x, y) \leq -(a - \eta)\varepsilon^2 + O(\varepsilon^3) + a\varepsilon^2 - \tfrac{1}{6}a\varepsilon^2 = \big(\eta - \tfrac{1}{6}a\big)\varepsilon^2 + O(\varepsilon^3) < 0,$$

by the assumption $\eta < \frac{1}{6}a$.

We claim that neither case can occur. Indeed, consider $y = x_0 + tv + O(t^2)$ for $t$ small, where $v$ is vector in the direction of the smallest curvature of the level set of

$u$ and $x_0$ satisfies (3-1). Then since $(|Du|\kappa_1)(x_0) = a - \eta$,

$$Q(x, y) = -(a-\eta)|y-x_0|^2 + O(|y-x_0|^3) + a|y-x_0|^2 - \frac{a}{6\varepsilon^2}|y-x_0|^4$$
$$= \eta|y-x_0|^2 - \frac{a}{6\varepsilon^2}|y-x_0|^4 + O(|y-x_0|^3).$$

If $|y-x_0| \sim \varepsilon^{4/3}$ say then $Q(x_0, y) \sim \eta\varepsilon^{8/3} + O(\varepsilon^3) \gg \varepsilon^3$ since we assume $\eta \gg \varepsilon^{1/3}$. Since $Q$ here is larger than in (1) or (2), this rules out (1) or (2) as being possible cases for the maximum of $Q$.

This implies that $Q$ must attain an interior maximum, contradicting the argument of Theorem 1.1. Here we use the fact that if $\psi(t) = at - a/(6\varepsilon^2)\, t^2$ then for $t$ with $0 \le t \le \varepsilon^2$,

$$\psi'(t) - 2|\psi''(t)|t = a(1 - t/\varepsilon^2) \ge 0. \qquad \square$$

**Remark 3.1.** In [Chang et al. 2010] and also [Ma et al. 2011] it was shown that when $n = 3$ the smallest principal curvature $\kappa_1$ also satisfies a minimum principle. It would be interesting to know whether a modification of the quantity (1-6) can give another proof of this.

## References

[Ahlfors 1973] L. V. Ahlfors, *Conformal invariants: topics in geometric function theory*, McGraw-Hill, New York, 1973. MR Zbl

[Alvarez et al. 1997] O. Alvarez, J.-M. Lasry, and P.-L. Lions, “Convex viscosity solutions and state constraints”, *J. Math. Pures Appl.* (9) **76**:3 (1997), 265–288. MR Zbl

[Andrews and Clutterbuck 2011] B. Andrews and J. Clutterbuck, “Proof of the fundamental gap conjecture”, *J. Amer. Math. Soc.* **24**:3 (2011), 899–916. MR Zbl

[Bian and Guan 2009] B. Bian and P. Guan, “A microscopic convexity principle for nonlinear partial differential equations”, *Invent. Math.* **177**:2 (2009), 307–335. MR Zbl

[Bianchini et al. 2009] C. Bianchini, M. Longinetti, and P. Salani, “Quasiconcave solutions to elliptic problems in convex rings”, *Indiana Univ. Math. J.* **58**:4 (2009), 1565–1589. MR Zbl

[Borell 1982] C. Borell, “Brownian motion in a convex ring and quasiconcavity”, *Comm. Math. Phys.* **86**:1 (1982), 143–147. MR Zbl

[Brascamp and Lieb 1976] H. J. Brascamp and E. H. Lieb, “On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation”, *J. Functional Analysis* **22**:4 (1976), 366–389. MR Zbl

[Caffarelli and Friedman 1985] L. A. Caffarelli and A. Friedman, “Convexity of solutions of semilinear elliptic equations”, *Duke Math. J.* **52**:2 (1985), 431–456. MR Zbl

[Caffarelli and Spruck 1982] L. A. Caffarelli and J. Spruck, “Convexity properties of solutions to some classical variational problems”, *Comm. Partial Differential Equations* **7**:11 (1982), 1337–1379. MR Zbl

[Caffarelli et al. 2007] L. Caffarelli, P. Guan, and X.-N. Ma, “A constant rank theorem for solutions of fully nonlinear elliptic equations”, *Comm. Pure Appl. Math.* **60**:12 (2007), 1769–1791. MR Zbl

[Chang et al. 2010] S.-Y. A. Chang, X.-N. Ma, and P. Yang, "Principal curvature estimates for the convex level sets of semilinear elliptic equations", *Discrete Contin. Dyn. Syst.* **28**:3 (2010), 1151–1164. MR Zbl

[Diaz and Kawohl 1993] J. I. Diaz and B. Kawohl, "On convexity and starshapedness of level sets for some nonlinear elliptic and parabolic problems on convex rings", *J. Math. Anal. Appl.* **177**:1 (1993), 263–286. MR Zbl

[Gabriel 1957] R. M. Gabriel, "A result concerning convex level surfaces of 3-dimensional harmonic functions", *J. London Math. Soc.* **32** (1957), 286–294. MR Zbl

[Hamel et al. 2016] F. Hamel, N. Nadirashvili, and Y. Sire, "Convexity of level sets for elliptic problems in convex domains or convex rings: two counterexamples", *Amer. J. Math.* **138**:2 (2016), 499–527. MR Zbl

[Jost et al. 2012] J. Jost, X.-N. Ma, and Q. Ou, "Curvature estimates in dimensions 2 and 3 for the level sets of $p$-harmonic functions in convex rings", *Trans. Amer. Math. Soc.* **364**:9 (2012), 4605–4627. MR Zbl

[Kawohl 1985] B. Kawohl, *Rearrangements and convexity of level sets in PDE*, Lecture Notes in Mathematics **1150**, Springer, 1985. MR Zbl

[Korevaar 1983] N. Korevaar, "Capillary surface convexity above convex domains", *Indiana Univ. Math. J.* **32**:1 (1983), 73–81. MR Zbl

[Korevaar 1990] N. J. Korevaar, "Convexity of level sets for solutions to elliptic ring problems", *Comm. Partial Differential Equations* **15**:4 (1990), 541–556. MR Zbl

[Korevaar and Lewis 1987] N. J. Korevaar and J. L. Lewis, "Convex solutions of certain elliptic equations have constant rank Hessians", *Arch. Rational Mech. Anal.* **97**:1 (1987), 19–32. MR Zbl

[Lewis 1977] J. L. Lewis, "Capacitary functions in convex rings", *Arch. Rational Mech. Anal.* **66**:3 (1977), 201–224. MR Zbl

[Longinetti 1983] M. Longinetti, "Convexity of the level lines of harmonic functions", *Boll. Un. Mat. Ital. A* (6) **2**:1 (1983), 71–75. MR Zbl

[Ma et al. 2010] X.-N. Ma, Q. Ou, and W. Zhang, "Gaussian curvature estimates for the convex level sets of $p$-harmonic functions", *Comm. Pure Appl. Math.* **63**:7 (2010), 935–971. MR Zbl

[Ma et al. 2011] X.-N. Ma, J. Ye, and Y.-H. Ye, "Principal curvature estimates for the level sets of harmonic functions and minimal graphs in $\mathbb{R}^3$", *Commun. Pure Appl. Anal.* **10**:1 (2011), 225–243. MR Zbl

[Ortel and Schneider 1983] M. Ortel and W. Schneider, "Curvature of level curves of harmonic functions", *Canad. Math. Bull.* **26**:4 (1983), 399–405. MR Zbl

[Rosay and Rudin 1989] J.-P. Rosay and W. Rudin, "A maximum principle for sums of subharmonic functions, and the convexity of level sets", *Michigan Math. J.* **36**:1 (1989), 95–111. MR Zbl

[Shiffman 1956] M. Shiffman, "On surfaces of stationary area bounded by two circles, or convex curves, in parallel planes", *Ann. of Math.* (2) **63** (1956), 77–90. MR Zbl

[Singer et al. 1985] I. M. Singer, B. Wong, S.-T. Yau, and S. S.-T. Yau, "An estimate of the gap of the first two eigenvalues in the Schrödinger operator", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **12**:2 (1985), 319–333. MR Zbl

[Székelyhidi and Weinkove 2016] G. Székelyhidi and B. Weinkove, "On a constant rank theorem for nonlinear elliptic PDEs", *Discrete Contin. Dyn. Syst.* **36**:11 (2016), 6523–6532. MR Zbl

[Wang 2014] X.-J. Wang, "Counterexample to the convexity of level sets of solutions to the mean curvature equation", *J. Eur. Math. Soc.* **16**:6 (2014), 1173–1182. MR Zbl

[Zhang and Zhang 2013] T. Zhang and W. Zhang, "On convexity of level sets of $p$-harmonic functions", *J. Differential Equations* **255**:7 (2013), 2065–2081. MR Zbl

Received February 21, 2017. Revised September 11, 2017.

BEN WEINKOVE
DEPARTMENT OF MATHEMATICS
NORTHWESTERN UNIVERSITY
EVANSTON, IL
UNITED STATES
weinkove@math.northwestern.edu

## Guidelines for Authors

Authors may submit articles at msp.org/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 295 No. 2 August 2018





0030-8730(201808)295:2;1-O