Pacific Journal of Mathematics

Volume 231 No. 1 May 2007

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

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

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

Regular subscription rate for 2007: \$425.00 a year (10 issues). Special rate: \$212.50 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

NONEXISTENCE RESULTS AND CONVEX HULL PROPERTY FOR MAXIMAL SURFACES IN MINKOWSKI THREE-SPACE

ROSA MARIA BARREIRO CHAVES AND LEONOR FERRER

We study properly immersed maximal surfaces with nonempty boundary and singularities in three-dimensional Minkowski space. We use the maximum principle and scaling arguments to obtain nonexistence results for these surfaces when the boundary is planar. We also give sufficient conditions for such surfaces to satisfy the convex hull property.

1. Introduction

In recent years, maximal hypersurfaces in a Lorentzian manifold — that is, spacelike submanifolds of codimension one with zero mean curvature — have been the object of considerable interest. Such hypersurfaces, and in general those having constant mean curvature, have a special significance in classical relativity [Marsden and Tipler 1980].

When the ambient space is the flat Minkowski space $\mathbb{L}^{n+1}$, Calabi [1970] (for $n \leq 3$) and Cheng and Yau [1976] (for arbitrary dimension) proved that a complete maximal hypersurface is necessarily a spacelike hyperplane. This result remains valid if we replace the completeness hypothesis by properness; see [Fernández and López 2004b]. Therefore, it does not make sense to consider global problems on regular maximal hypersurfaces in $\mathbb{L}^{n+1}$. Interesting problems are then those that deal with hypersurfaces with nonempty boundary or having certain type of singularities. In this line, Bartnik and Simon [1982/83] obtained results on the existence and regularity of spacelike solutions to the boundary value problem for the mean curvature operator in $\mathbb{L}^{n+1}$, and Kobayashi [1984] investigated surfaces with cone-like singularities. Estudillo and Romero [1992] defined a class of maximal surfaces with singularities of other types and studied criteria for such a surface to be a plane. On the other hand, Klyachin and Mikyukov [1993] have tackled the problem of existence of solutions to the maximal hypersurface equation in $\mathbb{L}^{n+1}$ with prescribed boundary conditions and a finite number of singularities. Fernández, López and Souam [Fernández et al. 2005] proved that a complete embedded maximal surface

---

MSC2000: primary 53C50; secondary 53C42, 53C80.
Keywords: maximal surfaces.
Research partially supported by MCYT-FEDER grant number MTM2004-00160.

with a finite set of singularities is an entire graph over any spacelike plane and that this family of maximal graphs has a structure of moduli space. We also mention the work of Umehara and Yamada [2006] where topological obstructions to the existence of this type of surfaces are given.

Maximal surfaces in $\mathbb{L}^3$ and minimal surfaces in Euclidean space are closely related. Both are solutions of variational problems, namely they are local maxima (minima) for the area functional. Both admit a Weierstrass representation (see [Kobayashi 1983] for maximal surfaces). The maximal surface equation and the minimal surface equation are both quasilinear elliptic equations and therefore enjoy a maximum principle. But contrary to the minimal case, solutions to the maximal surface equation can have isolated singularities, that is to say, points where the solution is not differentiable. Such points correspond to possible degeneracy of the ellipticity of the maximal surface equation. Geometrically at these singular points the Gauss curvature blows up, the Gauss map has no well-defined limit and the surface is asymptotic to the light cone.

In the minimal case, the maximum principle has been used by Schoen [1983], Hoffman and Meeks [1990], Meeks and Rosenberg [1993], López and Martín [2001], and others to derive remarkable results. In this paper we apply the maximum principle and scaling arguments to properly immersed maximal surfaces with nonempty boundary and isolated singularities in $\mathbb{L}^3$. We get two types of results: nonexistence results for properly immersed maximal surfaces with singularities and planar boundary contained in a timelike or lightlike plane, and results generalizing the convex hull property for such surfaces. Recall that a surface satisfies the convex hull property if it lies in the convex hull of its boundary. Although compact maximal surfaces in $\mathbb{L}^3$ satisfy this property, since they have nonpositive euclidean Gauss curvature (see [Osserman 1971/72]), this is not true if compactness is not assumed. We give sufficient conditions for a properly immersed maximal surface (not necessarily compact and with singularities) to satisfy the convex hull property.

Organization of paper. Section 2 contains the necessary notations and definitions, a description of the behavior of maximal surfaces around an isolated singularity, and a discussion of the maximal surfaces we use as barriers: Lorentzian catenoids, maximal surfaces of Riemann and Scherk type, and spacelike planes. We finish the section giving a first generalization of the convex hull property to compact maximal surfaces with singularities.

In Section 3 we obtain nonexistence results for properly immersed maximal surfaces with singularities and boundary contained in a timelike plane. Letting

$$\mathsf{C}^+ = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 - x_3^2 \le 0, x_3 \ge 0\}$$

be the positive solid half-cone, we show:

Theorem A. *There exists no connected properly immersed maximal surface M such that $M \subset \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 \geq 0,\ -ax_2 + x_3 \geq 0\}$ and $\partial M \subset \mathsf{C}^+ \cap \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 = 0\}$, for $a \in]-1, 1[$.*

This theorem holds even if we allow certain singularities (see Theorem 3.5).

Section 4 is devoted to the study of properly immersed maximal surfaces whose boundary is contained in a spacelike plane. Consider any region V of the form

$$V = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 \geq 0,\ -ax_2 + x_3 \leq 0,\ x_1 + bx_2 + c \geq 0\},$$

with $a \in]0, 1[$ and $b, c \in]-\infty, \infty[$.

Theorem B. *Let M be a connected properly immersed maximal surface contained in V and such that ∂M lies in a spacelike plane. Then M is a planar region.*

This result, too, holds even if we allow certain singularities (see Theorem 4.2 and Corollary 4.3). In the proof we construct a barrier surface *ad hoc* using the aforementioned Bartnik and Simon existence result. Theorem B is still valid if we replace V by C^+ (see Proposition 4.4).

Finally, in Section 5 we exploit the results of the preceding sections to give nonexistence results for properly immersed maximal surfaces with the boundary on a lightlike plane. We also prove:

Theorem C. *Any connected properly immersed maximal surface with singularities contained either in V or C^+ lies in the convex hull of its boundary and some of its singularities.*

Propositions 5.3 and 5.4 provide a precise formulation of this result.

2. Preliminaries

We denote by $\mathbb{L}^3$ the three dimensional Lorentz–Minkowski space $(\mathbb{R}^3, \langle\ ,\ \rangle)$, where the inner product corresponds to the form $dx_1^2 + dx_2^2 - dx_3^2$. A nonzero vector $v \in \mathbb{R}^3$ is called *spacelike, timelike or lightlike* if $\langle v, v\rangle$ is positive, negative or zero, respectively. The vector $(0, 0, 0)$ is considered spacelike. We say that a plane in $\mathbb{L}^3$ is *spacelike, timelike or lightlike* if the induced metric is Riemannian, nondegenerate and indefinite or degenerate, respectively. We also say that an affine plane in $\mathbb{L}^3$ is *spacelike, timelike or lightlike* if it is parallel to a spacelike, timelike or lightlike vector plane.

The *light cone* at $y = (y_1, y_2, y_3) \in \mathbb{L}^3$ is defined as

$$C(y) = \{x \in \mathbb{L}^3 \mid \langle x - y, x - y\rangle = 0\}.$$

We also set $C^+(y) = C(y) \cap \{x_3 \geq y_3\}$ and $C^-(y) = C(y) \cap \{x_3 \leq y_3\}$. Observe that lightlike vectors in $\mathbb{L}^3$ lie in the light cone $C((0, 0, 0))$.

Further, set $\mathbb{H}^2 = \mathbb{H}^2_+ \cup \mathbb{H}^2_-$, where $\mathbb{H}^2_+ = \{x \in \mathbb{L}^3 \mid \langle x, x\rangle = -1\} \cap \{x_3 \geq 0\}$ and $\mathbb{H}^2_- = \{x \in \mathbb{L}^3 \mid \langle x, x\rangle = -1\} \cap \{x_3 \leq 0\}$.

Consider the stereographic projection $\sigma : \overline{\mathbb{C}} - \{|z| = 1\} \to \mathbb{H}^2$ for $\mathbb{H}^2$, given by

$$\sigma(z) = \left(\frac{2\operatorname{Im} z}{|z|^2-1}, \frac{2\operatorname{Re} z}{|z|^2-1}, \frac{|z|^2+1}{|z|^2-1} \right), \tag{2-1}$$

where $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and $\sigma(\infty) = (0, 0, 1)$.

An immersion $X : M \to \mathbb{L}^3$ is *spacelike* if the tangent plane at any point is spacelike. In this case M must be orientable, that is to say, the Gauss map N is globally well defined and $N(M)$ lies in one of the components of $\mathbb{H}^2$.

A *maximal immersion* is a spacelike immersion $X : M \to \mathbb{L}^3$ whose mean curvature vanishes. In this case $X(M)$ is said to be a *maximal surface* in $\mathbb{L}^3$. Using isothermal parameters compatible with a fixed orientation $N : M \to \mathbb{H}^2$, M acquires a natural conformal structure, and the map $g = \sigma^{-1} \circ N$ is meromorphic. Moreover, there exists a holomorphic 1-form Φ_3 on M such that the 1-forms

$$\Phi_1 = \frac{i}{2}\left(\frac{1}{g} - g\right)\Phi_3,\ \Phi_2 = -\frac{1}{2}\left(\frac{1}{g} + g\right)\Phi_3 \tag{2-2}$$

are holomorphic, and together with Φ_3, have no real periods on M and no common zeros. Up to a translation, the immersion is given by

$$X = \operatorname{Re} \int (\Phi_1, \Phi_2, \Phi_3). \tag{2-3}$$

The induced Riemannian metric ds^2 on M is given by $ds^2 = \lambda(du^2 + dv^2)$, where $z = u + iv$ is a conformal parameter and

$$\lambda = \tfrac{1}{2}\left(|\Phi_1|^2 + |\Phi_2|^2 - |\Phi_3|^2\right) = \left(\frac{|\Phi_3|}{2} \left(\frac{1}{|g|} - |g| \right) \right)^2 .$$

Since M is spacelike, we have $|g| \neq 1$ on M and we can assume $|g| < 1$.

Conversely, let M, g and Φ_3 be a Riemann surface, a meromorphic map on M and a holomorphic 1-form on M. If $|g(p)| \neq 1$ for all $p \in M$, and if the 1-forms Φ_1, Φ_2, Φ_3 defined as above are holomorphic, have no real periods and no common zeros, then the conformal immersion X defined in (2-3) is maximal and its Gauss map is $\sigma \circ g$. We call (M, g, Φ_3) the Weierstrass representation of X. For more details see [Kobayashi 1983].

A maximal surface in $\mathbb{L}^3$ can be represented locally as a graph $x_3 = u(x_1, x_2)$ of a smooth function u such that $u_{x_1}^2 + u_{x_2}^2 < 1$ and

$$(1 - u_{x_1}^2)u_{x_2x_2} + 2u_{x_1}u_{x_2}u_{x_1x_2} + (1 - u_{x_2}^2)u_{x_1x_1} = 0. \tag{2-4}$$

The maximum principle for elliptic quasilinear equations then gives rise to:

Maximum principle for maximal surfaces. *Let S_1 and S_2 be two maximal surfaces in $\mathbb{L}^3$ which intersect tangentially at a point $p \in S_1 \cap S_2$. Suppose that u_i, for $i = 1, 2$ denotes the function defining S_i around p and that $u_1 \geq u_2$ (we say S_1 is above S_2 or S_2 is below S_1). Then $S_1 = S_2$ locally around p.*

Maximal surfaces with singularities. If in a maximal immersion $X : M \to \mathbb{L}^3$ we allow points $q \in M$ where the induced metric is not Riemannian we say that X (respectively, $X(M)$) has singularities and q (respectively, $X(q)$) is called a *singular point*. The different kinds of isolated singularities of maximal surfaces and the behavior of maximal surfaces around these points are well known; see [Kobayashi 1984; Ecker 1986; Miklyukov 1992; Fernández et al. 2005]. We recall the necessary material.

Let D be an open disc and $X : D \to \mathbb{L}^3$ a maximal immersion with a singular point in $q \in D$. There are two possibilities: either N extends continuously to q (q is a *spacelike* singular point) or not (q is a *lightlike* singular point).

In the second case $D - \{q\}$ with the induced metric is conformally equivalent to $\{z \in \mathbb{C}, 0 < r < |z| < 1\}$ and X extends to a conformal map $X : A_r \to \mathbb{L}^3$ with $X(\mathbb{S}^1) = X(q) = p$, where $A_r = \{z \in \mathbb{C}, r < |z| \leq 1\}$ and $\mathbb{S}^1 = \{z \in \mathbb{C}, |z| = 1\}$. Denote by $J(z) = 1/\bar{z}$ the inversion about $\mathbb{S}^1$. Then Schwarz reflection allows us to assert that X extends analytically to $B_r = \{z \in \mathbb{C}, r < |z| < 1/r\}$ and satisfies $X \circ J = -X + 2p$. Therefore if (g, Φ_3) are the Weierstrass data of the extended immersion we have $J^*(\Phi_k) = -\overline{\Phi}_k$ for $k = 1, 2, 3$, where $J^*(\Phi_k)$ denotes the pullback of Φ_k under J: if $\Phi_k = f_k\, dz$ then $J^*(\Phi_k) = -\bar{z}^{-2}(f_k \circ J)\, d\bar{z}$. Thus $g \circ J = 1/\bar{g}$ and consequently $|g| = 1$ on $\mathbb{S}^1$. Let Π be a spacelike plane containing $p = X(\mathbb{S}^1)$ and label $\pi : \mathbb{L}^3 \to \Pi$ as the Lorentzian orthogonal projection. If n (always even) is the number of zeros of Φ_3 on $\mathbb{S}^1$ and m denotes the degree of the map $g : \mathbb{S}^1 \to \mathbb{S}^1$, we have:

Lemma 2.1 [Fernández et al. 2005]. *There exists a small closed disc U in Π centered at p such that $(\pi \circ X)^{-1}(p) \cap V = \mathbb{S}^1$ and $(\pi \circ X) : V - \mathbb{S}^1 \to U - \{p\}$ is a covering of $m + \frac{1}{2}n$ sheets, where V is the annular connected component of $(\pi \circ X)^{-1}(U)$ containing $\mathbb{S}^1$.*

As a consequence, X is an embedding around q if and only if $m = 1$ and $n = 0$. In this case the point $p = X(\mathbb{S}^1)$ is said to be a *conelike singularity* of the maximal surface $X(D)$. Moreover, for r_0 close enough to 1, $X(A_{r_0})$ is the graph of a function u over Π. Locally, conelike singularities are points where the function defining the graph is not differentiable and correspond to possible degeneracy of the equation (2-4). Moreover, the graph of u is either above Π and asymptotic to $C^+(p)$ or below Π and asymptotic to $C^-(p)$, and the point p is called a *downward* or *upward pointing conelike singularity*, respectively.

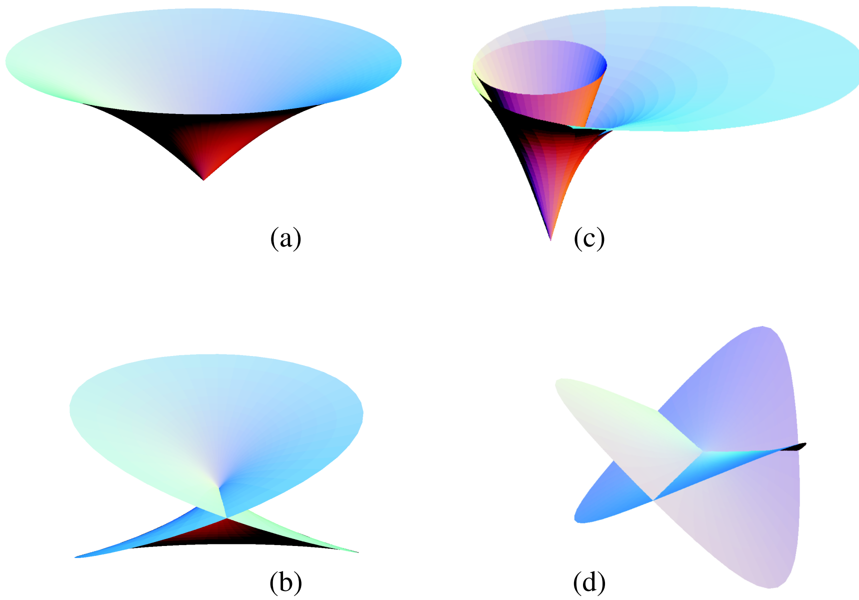


Figure 1. Different types of isolated singularities. (a) A downward pointing conelike singularity ($m = 1, n = 0$). (b) a downward pointing lightlike singularity with $m = 2$, $n = 0$. (c) a lightlike singularity with $m = 1$, $n = 2$. (d) a spacelike singularity with $n = 2$.

Lemma 2.2. *Let D be an open disc and $X : D \to \mathbb{L}^3$ a maximal immersion with a lightlike singular point in $q \in D$. Set $p = X(q)$. The neighborhoods U and V of Lemma 2.1 can be chosen so that*:

(i) *If p is a lightlike singularity with $n = 0$, then $X(V)$ is either over Π and asymptotic to $C^+(p)$ or below Π and asymptotic to $C^-(p)$* (see Figure 1a,b).

(ii) *If, on the contrary, p is a lightlike singularity with $n > 0$, there exist points of $X(V)$ in both sides of the plane Π. In particular there exist a pair of curves α, β in V starting at q such that $X(\alpha) - \{p\}$ is over Π and asymptotic to $C^+(p)$ and $X(\beta) - \{p\}$ is below Π and asymptotic to $C^-(p)$* (see Figure 1c).

Proof. Up to a Lorentzian isometry we can assume $\Pi = \{x_3 = 0\}$ and $p = (0, 0, 0)$. Let $X : A_r \to \mathbb{R}^3$ be a conformal reparametrization of the maximal immersion with $X(\mathbb{S}^1) = p$ and consider U, V as in Lemma 2.1. A thoughtful reading of the proof of Lemma 2.1 in [Fernández et al. 2005] will convince the reader that the same arguments prove (i).

For the proof of (ii) we use again the ideas of the same lemma. The Weierstrass data can be written in a neighborhood of $\mathbb{S}^1$ as

$$g(z) = z^m, \quad \Phi_3(z) = i\,\frac{\prod_{j=1}^n (z - a_j)}{z^{n/2+1}} f(z)\,dz, \tag{2-5}$$

where $a_1, \dots, a_n$ are the zeros of Φ_3 on $\mathbb{S}^1$ (with multiplicity) and f is a nonvanishing holomorphic function. Recall that the multiplicity of the zero of Φ_3 at a_i coincides with the number of nodal curves of the harmonic function x_3 meeting at a_i minus one. By the maximum principle there are no domains bounded by nodal curves and x_3 changes sign when crossing a nodal curve. Since $n \geq 2$ there are points of $X(V)$ in both sides of Π and there exist at least a pair of domains $\Gamma, \Gamma' \subset V$ bounded by a pair of nodal curves of x_3, a piece of $\partial V - \mathbb{S}^1$ and a point or a piece of $\mathbb{S}^1$, such that $x_3(X(\Gamma)) > 0$ and $x_3(X(\Gamma')) < 0$.

To conclude we prove that the image of all the curves $\rho_\theta(s) = se^{i\theta}$, for $\theta \in K = [0, 2\pi] - \{\arg(a_1), \dots, \arg(a_n)\}$, is asymptotic to the cone $C(p)$. Taking into account (2-5) we can write

$$X(\rho_\theta(s)) = \operatorname{Re} \int_1^s \frac{i \prod_{j=1}^n (te^{i\theta} - a_j)}{t^{n/2+1}(e^{i\theta})^{n/2}} f(te^{i\theta}) \times \left(\frac{i}{2}\Big(\frac{e^{-im\theta}}{t^m} - t^m e^{im\theta}\Big), -\frac{1}{2}\Big(\frac{e^{-im\theta}}{t^m} + t^m e^{im\theta}\Big), 1 \right) dt.$$

Since $J^*(\Phi_3) = -\overline{\Phi_3}$, we deduce that

$$\operatorname{Im}\left(\frac{i \prod_{j=1}^n (e^{i\theta} - a_j)}{(e^{i\theta})^{\frac{n}{2}}} f(e^{i\theta}) \right) = 0.$$

Using this it is straightforward to see that

$$\lim_{s\to 1} \left\| \frac{X(\rho_\theta(s))}{x_3(X(\rho_\theta(s)))} - \big(\sin(m\theta), -\cos(m\theta), 1\big) \right\|_1 = 0,$$

where $\|\cdot\|_1$ is the $\mathscr{C}^1$ norm in $\mathscr{C}^1(K, \mathbb{R}^3)$. Therefore, we can consider a pair of curves $\alpha \in \Gamma$ and $\beta \in \Gamma'$ satisfying the requirements of statement (ii). $\square$

Definition 2.3. A point p as in Lemma 2.2(i) is called a *downward* or *upward pointing lightlike singularity*, as the case may be. We also call it a *general conelike singularity*.

If D is an open disc and $X : D \to \mathbb{L}^3$ is a maximal immersion with a spacelike singular point in $q \in D$, the local behavior at the singularity is similar to the case of minimal surfaces in $\mathbb{R}^3$ (see [Dierkes et al. 1992; Estudillo and Romero 1992; Fernández et al. 2005]): X is not a topological embedding, $D - \{q\}$ with the induced metric is conformally equivalent to $\{z \in \mathbb{C} \mid 0 < |z| < 1\}$, the Weierstrass data (g, Φ_3)

extend analytically to q, $|g(q)| < 1$ and $\Phi = (\Phi_1, \Phi_2, \Phi_3)$ has a zero at q. Up to a Lorentzian isometry we can assume that the tangent plane of $X(D)$ at $p = X(q)$ is $\Pi = \{x_3 = 0\}$ and $p = (0, 0, 0)$. The Weierstrass data of the immersion can be written as

$$g(z) = z^m f(z), \qquad \Phi_3(z) = z^{m+n}\, dz,$$

where $m > 0$, n is the order of the zero of Φ at q and f is a holomorphic function with $f(0) \neq 0$. Up to a rotation around the axis x_3, we can assume $\mathrm{Im}(f(0)) = 0$. From here it is easy to derive that the asymptotic behavior of the immersion around the singularity is in polar coordinates

$$X(se^{i\theta}) = \Bigg(\frac{-s^{n+1}}{2f(0)(n+1)} \sin((n+1)\theta) + O(s^{n+2}), \\ \frac{-s^{n+1}}{2f(0)(n+1)} \cos((n+1)\theta) + O(s^{n+2}), \ \frac{s^{m+n+1}}{m+n+1} \cos((m+n+1)\theta) \Bigg),$$

where by $O(s^{n+2})$ we denote a function such that $s^{-n-2}O(s^{n+2})$ is bounded as $s \to 0$. Therefore, it is clear that X has a branch point at q of order n in the sense of [Gulliver et al. 1973].

Lemma 2.4 [Gulliver et al. 1973, Lemma 2.12]. *Let $X : D \to \mathbb{L}^3$ be a maximal immersion with a spacelike singular point at $q \in D$. Set $p = X(q)$ and let be S an embedded surface in $\mathbb{L}^3$ with $p \in S$. Suppose that for a neighborhood V of q, $X(V)$ lies on one side of S. Then the tangent plane to S at p coincides with the tangent plane to $X(D)$ at p.*

Remark 2.5. In the case of spacelike singularities, we always assume that the immersion $X : D \to \mathbb{L}^3$ is not a branched covering of an embedded surface; that is to say, q is not a false branch point. See [Gulliver et al. 1973].

Finally, we mention a property of maximal surfaces with singularities (see for example [Fernández and López 2004a]).

Lemma 2.6. *Let $X : M \to \mathbb{L}^3$ be a maximal immersion with isolated singularities. Then for all $q \in M$ there exists a neighborhood V, such that $X(V) - \{X(q)\}$ is contained in the exterior of $C(X(q))$.*

Remark 2.7. Let S be an embedded surface and $p \in S$. If the tangent plane of S at p is spacelike then S can be written in a neighborhood of p as the graph of a function h on a domain Ω of the plane $\{x_3 = 0\}$. Let M be another surface (possibly with singularities) and denote by π the orthogonal projection on $\{x_3 = 0\}$. In this context, we say that M lies *above* S in a neighborhood of p if $x_3(p') \geq h(x_1, x_2)$ for all $(x_1, x_2) \in \Omega$ and $p' \in M \cap \pi^{-1}(x_1, x_2)$. Naturally, M lies *below* S if $x_3(p') \leq h(x_1, x_2)$ instead.

Maximal surfaces with boundary. Let S' be a maximal surface, possibly with isolated singularities. Consider $S \subset S'$ such that the topological boundary of S in S' is nonempty and piecewise $\mathscr{C}^1$. Then S is called a *maximal surface with boundary*; the topological boundary of S in S' is called the *boundary* of S, written ∂S. The *interior* of S is $\operatorname{Int} S = S - \partial S$. Our definition allows singularities on the boundary of S.

Since the components of a maximal immersion are harmonic functions, the intersection of such an S with any plane Π having ∂S entirely to one side is a union of piecewise analytic curves, and each connected component of $S-(S\cap\Pi)$ is itself a maximal surface with boundary.

We say that S is a *properly immersed maximal surface with boundary* if, in the preceding situation, S' is a maximal surface properly immersed in $\mathbb{L}^3$.

Theorem [Fernández and López 2004b]. *Let M be a properly immersed maximal surface with boundary such that, except for a compact set, it is contained in the region $\{x \in \mathbb{L}^3 \mid \langle x, x\rangle \geq \varepsilon\}$, for $\varepsilon > 0$. Then M is relative parabolic, it is to say, bounded harmonic functions on M are determined uniquely by their values at the boundary and the interior isolated singularities.*

(Note that in [Fernández and López 2004b] the definition of a maximal surface with boundary is more general than in this paper.)

Corollary 2.8. *Let M be a connected properly immersed maximal surface with boundary such that $M \subset \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 0 \leq x_3 \leq k\}$ and the boundary and the singularities are contained in $\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 = k\}$, for $k > 0$. Then M is a planar region.*

Barrier surfaces. For any $v \in \mathbb{R}^3 - \{(0,0,0)\}$ and $y \in \mathbb{R}^3$, define

$$\begin{aligned} H(y, v) &= \{x \in \mathbb{R}^3 \mid \langle v, x-y\rangle_e = 0\},\\ H^+(y, v) &= \{x \in \mathbb{R}^3 \mid \langle v, x-y\rangle_e \geq 0\},\\ H^-(y, v) &= \{x \in \mathbb{R}^3 \mid \langle v, x-y\rangle_e \leq 0\}, \end{aligned}$$

where $\langle\, ,\, \rangle_e$ is the Euclidean metric on $\mathbb{R}^3$. Next, for $\theta \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ and $t \in \mathbb{R}$, set

$$\begin{aligned} \Pi_{\theta t} &= H\big((0,0,t), (0, -\tan\theta, 1)\big),\\ \Pi^+_{\theta t} &= H^+\big((0,0,t), (0, -\tan\theta, 1)\big),\\ \Pi^-_{\theta t} &= H^-\big((0,0,t), (0, -\tan\theta, 1)\big). \end{aligned}$$

In the case of $t = 0$ we write simply Π_θ instead of $\Pi_{\theta 0}$, and so on.

We also consider, for $\alpha \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[$,

$$\begin{aligned}\Sigma_\alpha &= H\big((0,0,0), (0,1,-\tan\alpha)\big),\\ \Sigma_\alpha^+ &= H^+\big((0,0,0), (0,1,-\tan\alpha)\big),\\ \Sigma_\alpha^- &= H^-\big((0,0,0), (0,1,-\tan\alpha)\big).\end{aligned}$$

Observe that $\Pi_{\pi/4,t}$ and $\Pi_{-\pi/4,t}$ are lightlike planes, while the $\Pi_{\theta t}$ are spacelike planes for $\theta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[$. Also, for any $\theta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[$, there is an orthochronous hyperbolic rotation f_s of $\mathbb{L}^3$ of the form

$$f_s\begin{pmatrix} x_1\\ x_2\\ x_3\end{pmatrix} = \begin{pmatrix} 1 & 0 & 0\\ 0 & \cosh s & \sinh s\\ 0 & \sinh s & \cosh s\end{pmatrix}\begin{pmatrix} x_1\\ x_2\\ x_3\end{pmatrix},$$

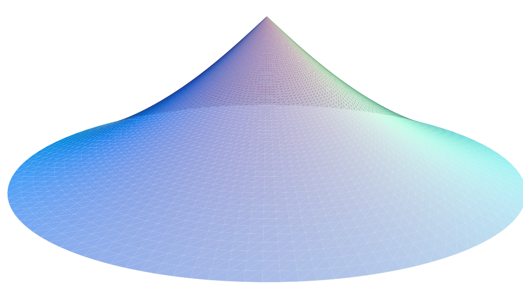
which preserves (individually) the light half-cones $C^+((0,0,0))$ and $C^-((0,0,0))$ and satisfies $f_s(\Pi_\theta) = \Pi_0$. Analogously, for $\theta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[$, there is an orthochronous isometry $\widetilde{f_s}$ of $\mathbb{L}^3$ composed of an orthochronous hyperbolic rotation f_s and a vertical translation, and such that $\widetilde{f_s}(\Pi_{\theta t}^+) = \Pi_0^+$ and $\widetilde{f_s}(\Pi_{\theta t}^-) = \Pi_0^-$, so also $\widetilde{f_s}(\Pi_{\theta t}) = \Pi_0$. As for the Σ_α, they are timelike planes and there is an orthochronous hyperbolic rotation f_s of $\mathbb{L}^3$ that preserves the light half-cones and satisfies $f_s(\Sigma_\alpha) = \Sigma_0$. For more details about these isometries of $\mathbb{L}^3$, see [Fernández and López 2004a].

Now we present the maximal surfaces that we use as barriers.

Lorentzian catenoids. The (vertical) *Lorentzian catenoid* $\mathscr{C}_a$ is the maximal surface given on $\overline{\mathbb{D}} - \{0\} = \{z \in \mathbb{C} \mid 0 < |z| \leq 1\}$ by the Weierstrass data $g = z$ and $\Phi_3 = a\,dz/z$ (see figure for the case $a = 1$). We can express $\mathscr{C}_a$ as the graph of the radially symmetric function

$$u(r) = -\int_0^r \frac{a}{\sqrt{t^2 + a^2}}\,dt, \qquad r > 0.$$

Let $\mathscr{C} = \{\mathscr{C}_a, a \in \left]0, \infty\right[\}$ be the family of such catenoids. Lorentzian catenoids have been used as barriers for applications of the maximum principle in [Bartnik and Simon 1982/83] and [Ecker 1986].



Maximal surfaces of Riemann type. R. López, F. J. López and R. Souam studied in [López et al. 2000] the set of maximal surfaces in $\mathbb{L}^3$ that are foliated by pieces of circles. From among them, we take the one-parameter family of Riemann-type maximal surfaces. This is a family of singly periodic maximal surfaces that plays the same role that Riemann's minimal examples play in Euclidean space, and whose fundamental piece is a graph over a spacelike plane, having one planar end and two conelike singularities:

We recall the Weierstrass representation of one-half of a fundamental piece of such surfaces. For $r \in]1, \infty[$, consider the four-punctured torus

$$\mathcal{N} = \{(z, w) \in \mathbb{C}^* \times \mathbb{C} \mid w^2 = z(z^2 + 2rz + 1)\}$$

and define in the z-plane

$$s_0 = \{z \in \mathbb{C} \mid |z| = 1\}, \quad s_1 = [r_1, 0[\times \{0\}, \quad s_2 =]-\infty, r_2] \times \{0\},$$

where $r_1 = -r + \sqrt{r^2 - 1}$ and $r_2 = -r - \sqrt{r^2 - 1}$. Observe that $r_2 < -1 < r_1 < 0$. Define $N \subset \mathcal{N}$ as the connected component of $z^{-1}\big(\mathbb{C} - \bigcup_{i=0}^{2} s_i\big)$ containing the point

$$\left(\tfrac{1}{2}, \sqrt{\tfrac{5}{8} + \tfrac{1}{2}r}\right).$$

Finally set $M = \overline{N}$, the closure of N in $\mathcal{N}$.

For brevity, when $z(z^2 + 2rz + 1) \in \mathbb{R}^+$, we set

$$z_+ = (z, +\sqrt{z(z^2 + 2rz + 1)}), \quad z_- = (z, -\sqrt{z(z^2 + 2rz + 1)}).$$

On M we consider the Weierstrass data $g = z$ and $\Phi_3 = dz/w$ and the 1-forms Φ_j, $j = 1, 2$ given by (2-2). The lift γ of s_0 to M generates $\mathcal{H}_1(M, \mathbb{Z})$. It is not difficult to see that Φ_1 is exact and that Φ_2, Φ_3 have no real periods on γ, so we can consider the maximal immersion $X = (X_1, X_2, X_3) = \operatorname{Re} \int_{z_0}^{z} (\Phi_1, \Phi_2, \Phi_3)$.

Denote by γ_1 the lift to M of s_1. It is not hard to prove that $X(\gamma_1)$ is a line parallel to $\{x_2 = x_3 = 0\}$. The set of singularities of the immersion is the trace of γ, and the image of these points under the immersion X is a single point, which we label P^r. We can choose z_0 so that $X(\gamma_1) = \{x_2 = x_3 = 0\}$ and $P^r = (0, P_2^r, P_3^r)$.

Let $\Theta(r) \in [-\pi, \pi[$ be the angle between $(0, 1, 0)$ and the vector P^r, given by

$$\cos\Theta(r) = \frac{P_2^r}{\sqrt{(P_2^r)^2 + (P_3^r)^2}}, \quad \sin\Theta(r) = \frac{P_3^r}{\sqrt{(P_2^r)^2 + (P_3^r)^2}}.$$

To use the surfaces of this family as barriers, we study the function $\Theta(r)$. We have $\Theta(r) = \arctan(h(r)/d(r))$, where

$$h(r) = X_3(-1_-) - X_3(r_1) = X_3(1_+) - X_3(0), \quad d(r) = X_2(-1_-) - X_2(r_1).$$

Hence

$$\begin{aligned} h(r) &= \mathrm{Re} \int_{r_1}^{-1_-} \Phi_3 = \int_{-1}^{r_1} \frac{dt}{\sqrt{t(t^2+2rt+1)}}, \\ h(r) &= \mathrm{Re} \int_0^{1_+} \Phi_3 = \int_0^1 \frac{dt}{\sqrt{t(t^2+2rt+1)}}, \\ d(r) &= \mathrm{Re} \int_{r_1}^{-1_-} \Phi_2 = -\frac{1}{2} \int_{-1}^{r_1} \frac{(1+t^2)\,dt}{t\sqrt{t(t^2+2rt+1)}}. \end{aligned} \tag{2-6}$$

Since h and d are positive functions, $\Theta(r)$ lies in $]0, \frac{\pi}{2}[$. Moreover,

$$d(r) = rh(r) + I(r), \tag{2-7}$$

where

$$I(r) = \frac{1}{2} \int_{-1}^{r_1} \frac{\sqrt{t(t^2+2rt+1)}\,dt}{t^2}. \tag{2-8}$$

From (2-7) and (2-8) we see that $\lim_{r\to 1} \Theta(r) = \frac{\pi}{4}$ and $\lim_{r\to+\infty} \Theta(r) = 0$. From (2-6) we observe that

$$h'(r) = \int_0^1 \frac{-t^2 dt}{(t(t^2+2rt+1))^{3/2}}. \tag{2-9}$$

On the other hand, from (2-7) and (2-8) the derivative of d respect to r is

$$d'(r) = rh'(r) + \tfrac{3}{2} h(r). \tag{2-10}$$

According to (2-7) and (2-10) we have

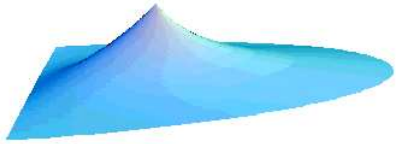
$$\Theta'(r) = \frac{h'(r)\,d(r) - h(r)\,d'(r)}{h(r)^2 + d(r)^2} = \frac{I(r)h'(r) - \frac{3}{2}h(r)^2}{h(r)^2 + d(r)^2}.$$

Taking into account (2-8) and (2-9) we get $\Theta'(r) < 0$, so Θ is a one-to-one function $\Theta :]1, \infty[\to]0, \frac{\pi}{4}[$.

For $\delta \in]0, \frac{\pi}{4}[$ we shall denote by R_δ the maximal surface with boundary defined in $\mathbb{L}^3$ by the above immersion for $r = \Theta^{-1}(\delta)$ (see figure at the top of next page). We also set

$$\mathcal{R} = \{R_\delta \mid \delta \in]0, \tfrac{\pi}{4}[\}.$$

Finally, we need to prove that $R_\delta \subset \Pi_\delta^- \cap \{x_3 \geq 0\}$. It is not difficult to see that the point $\{0\}$ is a planar end of the surface asymptotic to the plane $\{x_3 = 0\}$. Therefore,

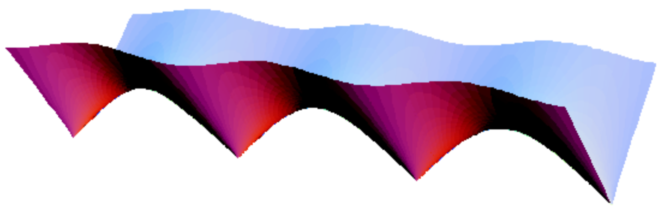


R_δ for $\delta = 0.595881$.

X_3 is bounded on M. From Corollary 2.8 we deduce that $R_\delta \subset \{x_3 \geq 0\}$. Moreover, from the above facts there exists $t > 0$ such that $R_\delta \subset \Pi^-_{\delta t}$. The maximum principle allows us to assert that $R_\delta \subset \Pi^-_\delta$.

Maximal surfaces of Scherk type. This family of singly periodic maximal surfaces of Scherk type was studied in depth in [Fernández and López 2004a], although an example had already appeared in [Kobayashi 1984]. For $b \in \,]0, 1[$, consider the maximal surface given on $\overline{\mathbb{D}} - \{b, -b\}$ by the Weierstrass data $g(z) = iz$ and

$$\Phi_3(z) = \frac{z\,dz}{(z^2 - b^2)(b^2z^2 - 1)}.$$



The surface is a graph over a spacelike plane, it is invariant under translation by $\big(0, \pi/(2b(b^2+1)), 0\big)$, and each fundamental piece of it has a conelike singularity. Up to translation we can assume that one of these singularities is at $(0, 0, 0)$, and then all the conelike singularities lie on the line $\{x_1 = x_3 = 0\}$. The ends are asymptotic to the totally geodesic horizontal half-cylinder ∂W_δ, where

$$W_\delta = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid -\tan\delta\, x_1 + x_3 \geq 0, \tan\delta\, x_1 + x_3 \geq 0\},$$

for $\delta = \arctan\big(2b/(1+b^2)\big) \in \,]0, \frac{\pi}{4}[$; for this reason we denote this Scherk-type surface by S_δ. By Corollary 2.8, S_δ lies entirely in W_δ.

The convex hull property. We now prove that a compact maximal surface, even one with isolated singularities, satisfies the convex hull property, that is, it lies in the convex hull of its boundary plus singularities. We will need the following version of the maximum principle for maximal surfaces with singularities. We would like to point out that the proof is inspired in the work [Gulliver et al. 1973].

Proposition 2.9. *Let $X : D \to \mathbb{L}^3$ be a maximal immersion with an isolated singular point in $q \in D$. Set $p = X(q)$ and let S be an embedded maximal surface* (*without singularities*) *in $\mathbb{L}^3$ with $p \in S$.*

(i) *If $X(D)$ is above S, then p is a downward pointing lightlike singularity.*

(ii) *If $X(D)$ is below S, then p is an upward pointing lightlike singularity.*

Proof. We prove (i); the proof of part (ii) is similar. Suppose first that p is a lightlike singularity but not downward pointing. Denote by Π the tangent plane to S at p that is a spacelike plane. From Lemma 2.2 we obtain a curve in $X(D)$ asymptotic to $C^-(p)$. Since S is asymptotic to Π in a neighborhood of p we deduce that there are points of $X(D)$ below S and this contradicts our assumptions.

Now assume p is a spacelike singularity. By Lemma 2.4, the tangent plane to $X(D)$ at p coincides with the tangent plane to S at p. Denote this plane by Π and by π the Lorentzian orthogonal projection on Π. Up to a Lorentzian isometry we can suppose that $p = (0, 0, 0)$ and $\Pi = \Pi_0$. Consider a disk Δ in Π centered at $(0, 0)$ such that S is the graph of a function h on Δ and $\Delta \subset \pi(X(D))$. Set $M = X(V)$, where V is the connected component of $(\pi \circ X)^{-1}(\Delta)$ containing q. If $\partial(M) \cap h(\Delta) \neq \varnothing$, we have an interior regular point in $X(D) \cap S$. By applying the maximum principle we obtain $M = h(\Delta)$ and then $h(\Delta)$ must contain a spacelike singularity. Taking into account Remark 2.5 we get a contradiction.

Suppose then that ∂M lies strictly above $h(\Delta)$. Then there exists $\theta \in \left]-\frac{\pi}{4}, 0\right[$ sufficiently small and f an hyperbolic rotation in $\mathbb{L}^3$ such that $f(\Pi) = \Pi_\theta$ and $f(\partial M)$ remains strictly above $h(\Delta)$. Note that the tangent plane to the maximal surface $f(M)$ at $(0, 0, 0)$ is Π_θ and thus we can assert that there are points of $f(M)$ below $h(\Delta)$. Translating in the positive x_3-direction, we find a last contact point with $h(\Delta)$ which must be an interior regular point. As in the previous case, by using the maximum principle we obtain that $h(\Delta)$ coincides with the translate of $f(M)$ by some vector $(0, 0, t_0)$, for $t_0 > 0$. From Lemma 2.6 we see that $\pi^{-1}(0, 0, 0)$ intersects this translate at $(0, 0, t_0)$. But this implies $(0, 0, t_0) = (0, 0, 0)$, again a contradiction. □

Proposition 2.10. *Let M be a compact maximal surface with isolated singularities. Then M lies in the convex hull of ∂M and its general conelike singularities.*

Proof. Let A be the set of general conelike singularities. If M is contained in a plane the result is obvious. Assume M is not flat and consider $v \in \mathbb{S}^2$ and $y \in \mathbb{R}^3$ such that $(\partial(M) \cup A) \subset H^+(y, v)$. We prove that $M \subset H^+(y, v)$.

We proceed by contradiction. Suppose that $M \cap \big(H^-(y, v) - H(y, v)\big) \neq \varnothing$. Let M' be a connected component of $M \cap H^-(y, v)$; then M' does not contain general conelike singularities.

First, assume v is a timelike vector, that is, $H(y, v)$ is a spacelike plane. There exists an interior point $p \in M'$ such that M' is contained in the slab determined by the parallel planes $H(p, v)$ and $H(y, v)$. Therefore we can use Proposition 2.9 to infer that p is a regular point of M'. Using the maximum principle we find $M' = H(p, v)$, a contradiction.

Analogously, if v is either spacelike or lightlike, we can deduce the existence of an interior point $p \in M'$ such that M' is contained in the slab determined by the parallel planes $H(p, v)$ and $H(y, v)$. If p were a spacelike singularity, Lemma 2.4 implies $H(p, v)$ is the tangent plane to M' at p, contradicting $|g(q)| < 1$. Assume p is a lightlike singularity. Up to a Lorentzian isometry we can assume that $p = (0, 0, 0)$ and

- $H(p, v) = \Sigma_\theta$ and $M' \subset \Sigma_\alpha^-$ if v is spacelike,
- $H(p, v) = \Pi_{\pi/4}$ and $M' \subset \Pi_{\pi/4}^+$ if v is lightlike.

By Lemma 2.6, M' is in the exterior of $C((0, 0, 0))$. Consider π, the Lorentzian orthogonal projection onto Π_0. It is easy to prove that the preceding conditions imply that $\pi(M') \subset (\Pi_0 - \{(0, y, 0) \mid y \in \mathbb{R}\})$ in a neighborhood of $(0, 0, 0)$. This contradicts Lemma 2.1. Therefore, since p is not a singular point we infer that $H(y, v)$ is the tangent plane to M' at p, in contradiction with the fact that M is spacelike. □

Remark 2.11. Proposition 2.10 holds even if M cannot be extended to an open maximal surface M'.

3. Maximal surfaces whose boundary is contained in a timelike plane

Recall that we defined

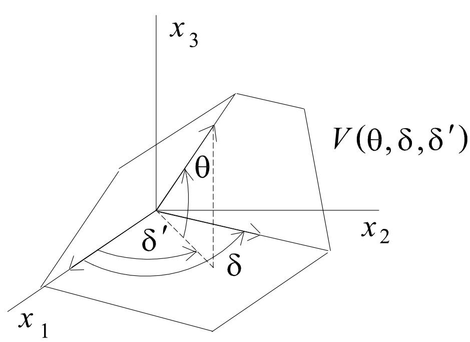
$$\mathsf{C}^+ = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 - x_3^2 \leq 0,\ x_3 \geq 0\}.$$

For $\theta \in]0, \frac{\pi}{4}[$ and $\delta, \delta' \in]0, \pi[$, we also define the convex region

$$V(\theta, \delta, \delta') = \Pi_0^+ \cap \Pi_\theta^- \cap H^+\big((0, 0, 0),\ (1, -\cot\delta, \cot\theta(\cot\delta - \cot\delta'))\big).$$

This is the convex hull of the half-lines with origin in $(0, 0, 0)$ and directions $(1, 0, 0)$, $(\cot\delta, 1, 0)$ and $(\cot\delta', 1, \tan\theta)$ (see figure).

Let τ_t denote the translation along the vector $(0, 0, t)$, where $t \in \mathbb{R}$.



Proposition 3.1. *Let $\alpha \in]-\frac{\pi}{4}, \frac{\pi}{4}[$ be arbitrary. Suppose M is a connected, properly immersed maximal surface contained in $\Pi^+_{\pi/4} \cap \Sigma^+_\alpha$, such that $\partial M \subset \Sigma_\alpha$. If there is a point $p_0 \in \partial M$ minimizing x_3 on ∂M, there must be some downward pointing lightlike singularity in the interior of M.*

Proof. After applying an orthochronous hyperbolic rotation f_s of $\mathbb{L}^3$, we can assume $\alpha = 0$.

Let M be as in the hypothesis of the proposition and define $\hat{t} = x_3(p_0)$. For any $\theta \in [0, \frac{\pi}{4}[$, consider the set

$$\mathcal{I}_\theta = \{t \in [0, \hat{t}] \mid M \subset \Pi^+_{\theta t}\}.$$

Since $\mathcal{I}_\theta$ contains 0, it is nonempty. Suppose there is no singularity in $\operatorname{Int} M$ as in the conclusion; we shall prove that $\mathcal{I}_\theta = [0, \hat{t}]$, and from there we will derive a contradiction.

We may assume $\hat{t} > 0$, the case $\hat{t} = 0$ being obvious. Clearly $\mathcal{I}_\theta$ is closed; we show that it is open. If $t \in \mathcal{I}_\theta$ then $[0, t] \subset \mathcal{I}_\theta$. We claim that if $t \in \mathcal{I}_\theta \cap [0, \hat{t}[$, there exists $\varepsilon > 0$ such that $[t, t+\varepsilon[\subset \mathcal{I}_\theta$. If not, we have two possibilities: either

(a) there is an interior point p of M in the plane $\Pi_{\theta t}$, or

(b) M is asymptotic to $\Pi_{\theta t}$ at infinity.

In case (a), p is not a singularity, for if it were, it would be downward pointing by Proposition 2.9(i), contrary to assumption. But if p is not a singularity, the interior maximum principle implies that M and $\Pi_{\theta t}$ coincide, in contradiction with the inequality $x_3(p_0) = \hat{t} > t$.

In case (b), we can assume $M \cap \Pi_{\theta t} = \varnothing$; otherwise there exists an interior point of M in $\Pi_{\theta t}$ and we may apply the previous argument. Consider an orthochronous isometry f of $\mathbb{L}^3$ such that $f(\Pi_{\theta t}) = \Pi_0$ and $f(\Pi^+_{\theta t}) = \Pi^+_0$. Set $\widetilde{M} = f(M)$. Then we have a properly immersed maximal surface $\widetilde{M} \subset \Pi^+_0$ asymptotic to Π_0 and disjoint from it.

Since the immersion is proper and $(0, 0, 0) \notin \widetilde{M}$ we can find $\epsilon > 0$ sufficiently small so that the ball $B(\epsilon)$ of radius ϵ around $(0, 0, 0)$ is disjoint from $\widetilde{M}$. Hence, there exists $\epsilon' \in]0, \epsilon[$ and $a_0 > 0$ small enough such that $\tau_{\epsilon'}(C_{a_0}) \subset B(\epsilon) \cup \Pi^-_0$. Now define

$$A = \{a \in]0, a_0] \mid \tau_{\epsilon'}(C_a) \cap \widetilde{M} = \varnothing\}.$$

Clearly, $a_0 \in A$ and we can consider the infimum a' of A. We claim that $a' = 0$. Assume on the contrary that $a' > 0$. Since $\tau_{\epsilon'}(C_a)$ and $\widetilde{M}$ do not have a contact at infinity, there exists an interior point p of $\widetilde{M}$ in $\tau_{\epsilon'}(C_{a'})$. Taking into account Proposition 2.9(i) and the assumed absence of singularities, we see that p is a regular point of $\widetilde{M}$. Applying the maximum principle, we obtain $\widetilde{M} = \tau_{\epsilon'}(C_{a'})$, in contradiction with $\partial \widetilde{M} \subset \Sigma_0 \cap \Pi^+_0$.

We have shown that if the conclusion of the theorem fails, then $\mathcal{I}_\theta$ is open and $\mathcal{I}_\theta = [0, \hat{t}]$. Hence $M \subset \Pi^+_{\theta\hat{t}}$ for all $\theta \in [0, \frac{\pi}{4}[$ and thus $M \subset \Pi^+_{\pi/4,\hat{t}}$. Now consider p_0: if it is *not* a singular point in ∂M, the fact that $p_0 \in \partial(M) \cap \Pi_{\pi/4,\hat{t}}$ and $M \subset \Pi^+_{\pi/4,\hat{t}} \cap \Sigma^+_0$ implies that the tangent plane p_0 is lightlike of timelike, which is a contradiction. If instead p_0 *is* a singular point, Lemma 2.6 implies that around p_0 the surface is in the exterior of $C(p_0)$, which again contradicts $M \subset \Pi^+_{\pi/4,\hat{t}}$. □

Corollary 3.2. *There exists no connected properly immersed maximal surface M without downward pointing lightlike singularities in the interior and such that $M \subset \mathsf{C}^+ \cap \Sigma^+_\alpha$ and $\partial M \subset \Sigma_\alpha$, for any $\alpha \in]-\frac{\pi}{4}, \frac{\pi}{4}[$.*

Proof. This follows immediately from Proposition 3.1. □

Corollary 3.3. *There exists no connected properly immersed maximal surface M without downward pointing lightlike singularities in the interior and such that $M \subset \mathsf{C}^+$ and ∂M lies in the intersection of C^+ with a timelike plane P.*

Proof. Assume that there exists such a maximal surface and consider a connected component M' of $M - (M \cap P)$. Up to an elliptic rotation and a translation we can assume the timelike plane is the plane Σ_α, for $\alpha \in]-\frac{\pi}{4}, \frac{\pi}{4}[$, $M' \subset \mathsf{C}^+ \cap \Sigma^+_\alpha$ and $\partial M' \subset \Sigma_\alpha$. An immediate application of Corollary 3.2 to M' leads to a contradiction. □

Theorem 3.4. *There exists no connected properly immersed maximal surface M without downward pointing lightlike singularities in the interior and such that $M \subset W_\delta \cap \Sigma^+_\alpha$ and $\partial M \subset \Sigma_\alpha \cap \mathsf{C}^+$, for $\delta \in]0, \frac{\pi}{4}[$, $\alpha \in]-\frac{\pi}{4}, \frac{\pi}{4}[$.*

Proof. Consider the isometry f_ε of $\mathbb{L}^3$ with $\tanh \varepsilon = \tan \frac{\pi}{8}$. It is not difficult to see that $f_\varepsilon(M) \subset \Pi^+_{\pi/8} \cap W_{\delta'} \cap \Sigma^+_{\alpha'}$ and $\partial f_\varepsilon(M) \subset \Sigma_{\alpha'} \cap \mathsf{C}^+$, where

$$\tan \alpha' = \frac{\tan \alpha + \tan \frac{\pi}{8}}{\tan \frac{\pi}{8} \tan \alpha + 1},$$

$$\tan \delta' = \min\{\tan \tfrac{\pi}{8}, \cosh \varepsilon \tan \delta (\tan \tfrac{\pi}{8} \tan \alpha + 1)\}.$$

For simplicity of notation we consider $M \subset \Pi^+_{\pi/8} \cap W_\delta \cap \Sigma^+_\alpha$ and $\partial M \subset \Sigma_\alpha \cap \mathsf{C}^+$.

We claim that $(0,0,0) \notin \partial M$ and so $\partial M \subset \Pi^+_0 - \Pi_0$. If not, we deduce from Lemma 2.6 that around $(0,0,0)$ the maximal surface M is in the exterior of $C((0,0,0))$, but this contradicts $\partial M \subset \mathsf{C}^+$.

Now consider the Scherk-type maximal surface $S_{\delta/2}$ (page 13) asymptotic to the boundary of the region $W_{\delta/2}$. We will prove that $S_{\delta/2} \cap M = \varnothing$. It is clear that there exist $t_0 \in]-\infty, 0]$ and $t_1 \in]0, \infty[$ such that $\tau_{t_0}(S_{\delta/2}) \cap M = \varnothing$ and $\tau_{t_1}(S_{\delta/2}) \cap M \neq \varnothing$. Therefore, we can define

$$\hat{t} = \text{infimum}\big\{t \in]t_0, \infty[\;\big|\; \tau_t(S_{\delta/2}) \cap M \neq \varnothing\big\}.$$

Suppose $\hat{t} \leq 0$. Observe that, since $M \subset \Pi^+_{\pi/8} \cap W_\delta \cap \Sigma^+_\alpha$, $\partial M \subset \mathrm{C}^+ - \{(0,0,0)\}$ and $S_{\delta/2} \cap \mathrm{C}^+ = \{(0,0,0)\}$, then $\tau_{\hat{t}}(S_{\delta/2})$ and M can have a contact point neither at infinity nor at the boundary. Hence there exists an interior point of M in $\tau_{\hat{t}}(S_{\delta/2})$. Taking into account our assumptions on the singularities and Proposition 2.9(i) we deduce that this point is not a singularity. Then, by applying the maximum principle we get that M and $\tau_{\hat{t}}(S_{\delta/2})$ coincide which contradicts the hypothesis on ∂M. Thus $\hat{t} > 0$ and $S_{\delta/2} \cap M = \varnothing$.

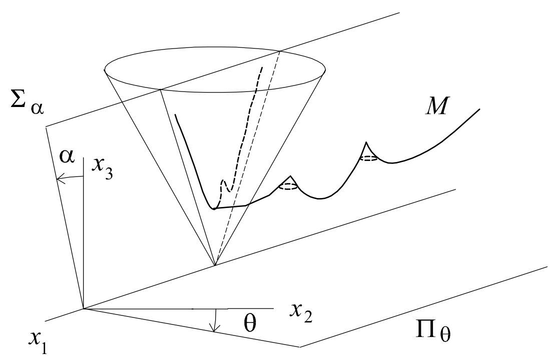
Now consider $S^\lambda_{\delta/2}$, the homothetic shrinking of $S_{\delta/2}$ by λ, $\lambda \geq 1$. We shall prove that $S^\lambda_{\delta/2} \cap M = \varnothing$ for all $\lambda \geq 1$. Suppose on the contrary that there exists $\lambda' \geq 1$ such that $S^{\lambda'}_{\delta/2} \cap M \neq \varnothing$. We set

$$\hat{\lambda} = \text{infimum}\big\{\lambda \in]1,\infty[\;\big|\; S^\lambda_{\delta/2} \cap M \neq \varnothing\big\}.$$

Clearly $S^{\hat{\lambda}}_{\delta/2}$ and M do not touch either at infinity or at the boundary. Therefore there must exist an interior point of M in $S^{\hat{\lambda}}_{\delta/2}$. Again using Proposition 2.9(i) and our hypothesis on the singularities we deduce that this point is not a singularity and so by applying the maximum principle we obtain that $S^{\hat{\lambda}}_{\delta/2}$ and M coincide. But this contradicts our assumptions on ∂M.

Thus $S^\lambda_{\delta/2} \cap M = \varnothing$ for all $\lambda \geq 1$. Taking into account that $S_{\delta/2}$ is asymptotic to $C^+((0,0,0))$ near the conelike singularity $(0,0,0)$, we deduce that $M \subset \mathrm{C}^+$ and the Corollary 3.2 finishes the proof. □

Theorem 3.5. *There exist no connected properly immersed maximal surface M without downward pointing lightlike singularities in the interior and such that $M \subset \Pi^+_\theta \cap \Sigma^+_\alpha$ and $\partial M \subset \Sigma_\alpha \cap \mathrm{C}^+$, for $\theta, \alpha \in]-\frac{\pi}{4}, \frac{\pi}{4}[$.*



Proof. Suppose there exists such an M. We observe that if $\theta \leq 0$ or $\alpha < 0$ we can consider an orthochronous hyperbolic rotation f_s such that $f_s(M) \subset \Pi^+_{\theta'} \cap \Sigma^+_{\alpha'}$ and $\partial f_s(M) \subset \Sigma_{\alpha'} \cap \mathrm{C}^+$ for some $\theta' \in]0, \frac{\pi}{4}[$, $\alpha' \in [0, \frac{\pi}{4}[$. As in the previous theorem, for the sake of simplicity of notation we assume $M \subset \Pi^+_\theta \cap \Sigma^+_\alpha$ and $\partial M \subset \Sigma_\alpha \cap \mathrm{C}^+$, for $\theta \in]0, \frac{\pi}{4}[$, $\alpha \in [0, \frac{\pi}{4}[$.

Since $\partial M \subset \Sigma_\alpha \cap \mathsf{C}^+$ we have that there exists $p_0 \in \partial M$ such that $x_3(p_0) \leq x_3(p)$ for all $p \in \partial M$. As in the preceding theorem it is easy to see that $p_0 \neq (0,0,0)$ and so $\lambda = x_3(p_0) > 0$. Then, reasoning as in Proposition 3.1 we can conclude $M \subset \Pi^+_{\theta\lambda} \cap \Sigma^+_\alpha$.

Denote by $\widetilde{R}_\delta$ the Riemann type maximal example that results after applying an elliptic rotation of $\frac{\pi}{2}$ along the axis x_3 on R_δ for any $\delta \in]0, \frac{\pi}{4}[$. We assert that $M \cap \widetilde{R}_\delta = \varnothing$. Observe that we can consider $t_0 \leq 0$ and $t_1 \in \mathbb{R}$ such that $\tau_{t_0}(\widetilde{R}_\delta) \cap M = \varnothing$ and $\tau_{t_1}(\widetilde{R}_\delta) \cap M \neq \varnothing$. Now define

$$\hat{t} = \text{infimum}\big\{t \in]t_0, \infty[\,\big|\, \tau_t(\widetilde{R}_\delta) \cap M \neq \varnothing\big\}.$$

Suppose $\hat{t} \leq 0$. Note that $\tau_{\hat{t}}(\widetilde{R}_\delta)$ and M can have a contact point neither at infinity nor at the boundary. Hence there exists an interior point of M in $\tau_{\hat{t}}(\widetilde{R}_\delta)$. Making use of Proposition 2.9(i) and taking into account our assumptions on singularities, we deduce that the point is not a singularity. Therefore, by applying the maximum principle we get that $\tau_{\hat{t}}(\widetilde{R}_\delta)$ and M coincide. But this contradicts our hypothesis on ∂M. Thus $\hat{t} > 0$ and $\widetilde{R}_\delta \cap M = \varnothing$.

Consider now $\widetilde{R}^\lambda_\delta$ the homothetic shrinking of $\widetilde{R}_\delta$ by λ, $\lambda > 0$. Next we prove that $\widetilde{R}^\lambda_\delta \cap M = \varnothing$ for all $\lambda \geq 1$. Assume that there exists $\lambda' > 1$ such that $\widetilde{R}^{\lambda'}_\delta \cap M \neq \varnothing$. We define

$$\hat{\lambda} = \text{infimum}\big\{\lambda \in]1, \lambda'[\,\big|\, \widetilde{R}^\lambda_\delta \cap M \neq \varnothing\big\}.$$

Observe that $\widetilde{R}^{\lambda'}_\delta$ and M do not touch either at infinity or at the boundary for all $\lambda \geq 1$. Therefore there is an interior point of M in $\widetilde{R}^{\lambda'}_\delta$. Using again our assumptions on singularities and Proposition 2.9(i) we deduce that the point is not a singularity. Then by applying the maximum principle we obtain that $\widetilde{R}^{\lambda'}_\delta$ and M coincide, which contradicts our hypothesis on ∂M. The same argument proves that $\widetilde{R}^\lambda_\delta \cap M = \varnothing$ for all $\lambda \leq 1$.

Analogously, considering $\widehat{R}_\delta$ the Riemann type maximal example that results after applying a rotation of $-\frac{\pi}{2}$ along the axis x_3 on R_δ for any $\delta \in]0, \frac{\pi}{4}[$, we obtain $\widehat{R}^\lambda_\delta \cap M = \varnothing$ for all $\lambda \in \mathbb{R}$.

Furthermore, it is not difficult to prove that

$$(\Pi^+_\theta \cap \Sigma^+_\alpha) - \left(\bigcup_{\lambda\in\mathbb{R}} \widetilde{R}^\lambda_\delta \cup \bigcup_{\lambda\in\mathbb{R}} \widehat{R}^\lambda_\delta\right) \subset \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid -\tan\delta\, x_1 + x_2 + x_3 \geq 0\} \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid \tan\delta x_1 + x_2 + x_3 \geq 0\}.$$

Taking this into account, we can assert

$$M \subset (\Pi^+_\theta \cap \Sigma^+_\alpha) \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid -\tan\delta\, x_1 + x_2 + x_3 \geq 0,\ \tan\delta\, x_1 + x_2 + x_3 \geq 0\}.$$

A direct computation shows that

$$\Pi^+_\theta \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid -\tan\delta\, x_1 + x_2 + x_3 \geq 0,\ \tan\delta\, x_1 + x_2 + x_3 \geq 0\} \subset W_{\delta'},$$

where $\delta' \in]0, \frac{\pi}{4}[$ is given by $\tan\delta' = \tan\delta \tan\theta/(1+\tan\theta)$. Then Theorem 3.4 concludes the proof. $\square$

To finish this section, we analyze the case of maximal surfaces whose boundary is contained in a timelike plane but not necessarily in C^+.

Proposition 3.6. *There exists no connected properly immersed maximal surface M with at least one connected component of ∂M contained in the intersection $\Sigma_\alpha^+ \cap \Sigma_{-\alpha}^+ \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid x_1 = 0\}$, where $\alpha \in]0, \frac{\pi}{4}[$.*

Proof. Let B a connected component of ∂M satisfying

$$B \subset \Sigma_\alpha^+ \cap \Sigma_{-\alpha}^+ \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid x_1 = 0\},$$

where $\alpha \in]0, \frac{\pi}{4}[$. The function x_2 has at least one minimum on B, and this minimum cannot be a singularity. Then, the tangent vector to the boundary at this point is vertical and therefore the tangent plane of the maximal surface at this point is timelike, which is contrary to our assumptions. $\square$

Corollary 3.7. *There exists no connected properly immersed maximal surface M contained in $V(\theta,\delta,\delta')$ with ∂M contained in a timelike plane.*

Proof. From the hypothesis it is not difficult to see that there exists an isometry of $\mathbb{L}^3$ that sends the timelike plane to the plane $\{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid x_1 = 0\}$ and in particular the image of ∂M lies in $\Sigma_\alpha^+ \cap \Sigma_{-\alpha}^+ \cap \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid x_1 = 0\}$ for some $\alpha \in]0, \frac{\pi}{4}[$. The result is then a consequence of Proposition 3.6. $\square$

4. Maximal surfaces whose boundary is contained in a spacelike plane

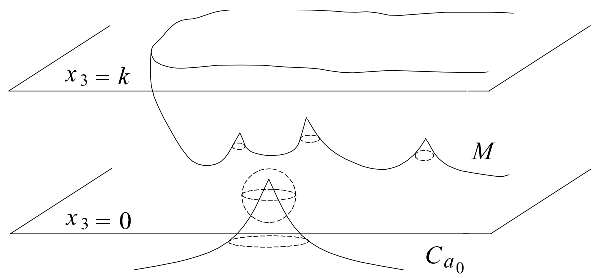
We now obtain, using the maximum principle, other results about maximal surfaces whose boundary is contained in a spacelike plane but that cannot be inferred from the theorem of Fernández and López quoted on page 9. We start with a result similar to Corollary 2.8.

Proposition 4.1. *Let M be a connected properly immersed maximal surface without downward pointing lightlike singularities in the interior such that*

$$M \subset \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid 0 \le x_3 \le k\} \quad \textit{and} \quad \partial M \subset \{(x_1,x_2,x_3) \in \mathbb{R}^3 \mid x_3 = k\},$$

for $k > 0$. Then M is a planar region.

Proof. We proceed by contradiction. Assume that there exists $t \ge 0$ such that $M \subset \Pi_{0t}^+$ but $M \not\subset \Pi_{0t'}^+$ for any $t < t'$. We claim that $M \cap \Pi_{0t} = \varnothing$. If not, there exists an interior point p of M in Π_{0t}. Then, from Proposition 2.9(i) and our assumptions on singularities, we deduce that p is a regular point. It then follows from the maximum principle that $M = \Pi_{0t}$, contradicting the hypothesis on ∂M.



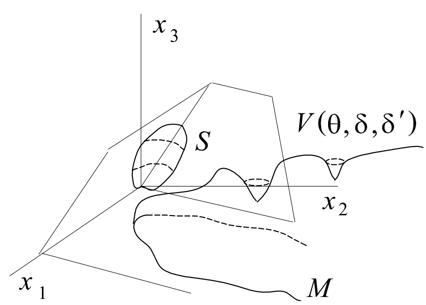
Since M is properly immersed and $M \cap \Pi_{0t} = \varnothing$ we can find $\varepsilon > 0$ such that the ball of radius ε about $(0, 0, t)$ does not meet M. Hence, there are constants $\varepsilon' \in\ ]0, \varepsilon[$ and $a_0 > 0$ sufficiently small such that $\tau_{t+\varepsilon'}(C_{a_0})$ is contained in Π^-_{0t} and also in the ball of radius ε around $(0, 0, t)$. Now define

$$A = \{a \in\ ]0, a_0] \mid \tau_{t+\varepsilon'}(C_a) \cap M = \varnothing\}.$$

Clearly, $a_0 \in A$ and we can consider the infimum a' of A. We claim $a' = 0$. Assume on the contrary that $a' > 0$. Then as $\tau_{t+\varepsilon'}(C_a)$ and M do not have a contact either at infinity or at the boundary, we infer that there is an interior point of M in $\tau_{t+\varepsilon'}(C_{a'})$. Taking into account Proposition 2.9(i) and our assumptions on the singularities we infer that the interior point is not a singularity and then, by applying the interior maximum principle we obtain $M = \tau_{t+\varepsilon'}(C_{a'})$ which contradicts the hypothesis on ∂M.

Therefore, $a' = 0$ and so $M \subset \Pi^+_{0t+\varepsilon'}$ contradicting our assumption at the beginning of the proof. □

Theorem 4.2. *Let M be a connected properly immersed maximal surface without upward pointing lightlike singularities in the interior such that $M \subset V(\theta, \delta, \delta')$ and $\partial M \subset \Pi_0$. Then M is a planar region.*



Proof. Up a translation we can assume that

$$M \cap (\Pi_\theta \cup H\big((0, 0, 0), (1, -\cot\delta, \cot\theta(\cot\delta - \cot\delta'))\big) = \varnothing.$$

Now we observe that

$$\tau_{-1}(\mathbb{H}^2_+) \cap \Pi_\theta \cap V(\theta, \delta, \delta') = \alpha_1,$$
$$\tau_{-1}(\mathbb{H}^2_+) \cap H((0,0,0),(1, -\cot\delta, \cot\theta(\cot\delta - \cot\delta'))) \cap V(\theta, \delta, \delta') = \alpha_2,$$

where α_1 and α_2 are regular curves. The union of these curves is a continuous curve of $\tau_{-1}(\mathbb{H}^2_+)$ and the tangent vectors to these curves at the point $(0,0,0)$ are contained in the plane Π_0 and are linearly independent.

Since $\tau_{-1}(\mathbb{H}^2_+)$ is a spacelike surface, it is well-known (see Theorem 4.1 in [Bartnik and Simon 1982/83]) that there exists S a maximal surface (it is even a graph on the x_3-plane) spanned by the curve $\alpha_1 \cup \alpha_2$. Note that the tangent plane of S at $(0,0,0)$ is the plane Π_0. On the other hand, using Proposition 2.10 and Remark 2.11, we see that S is contained in the convex hull of its boundary and thus $S \subset V(\theta, \delta, \delta')$.

Now, we denote by S^λ the homothetic shrinking of S by λ, $\lambda > 0$. As M is properly immersed it is possible to find $\lambda_0 > 0$ such that $S^{\lambda_0} \cap M = \varnothing$. Next we prove that $S^\lambda \cap M = \varnothing$ for all $\lambda > 0$. Assume that there exists $\lambda' > 0$ such that $S^{\lambda'} \cap M \neq \varnothing$. We denote by

$$\hat{\lambda} = \text{infimum}\big\{\lambda \in \,]\lambda_0, \lambda'] \;\big|\; S^\lambda \cap M \neq \varnothing\big\}.$$

Observe that S^λ and M do not contact at the boundary for all $\lambda > 0$. Therefore there is an interior point of M in $S^{\hat{\lambda}}$. It follows from Proposition 2.9(ii) and the conditions on the singularities that the contact point is not a singularity. Then, by applying the maximum principle we obtain $S^{\hat{\lambda}} = M$ which contradicts the assumptions on ∂M.

Hence, taking into account that the tangent plane of S at $(0,0,0)$ is Π_0 we obtain

$$V(\theta, \delta, \delta') - \bigcup_{\lambda \in \mathbb{R}} S^\lambda \subset \Pi_0,$$

from which we deduce that $M \subset \Pi_0$. □

Corollary 4.3. *Let M be a connected properly immersed maximal surface without general conelike singularities in the interior such that $M \subset V(\theta, \delta, \delta')$ and ∂M is contained in a spacelike plane. Then M is a planar region.*

Proof. Let Π be the spacelike plane such that $\partial M \subset \Pi$ and M' a connected component of $M - (M \cap \Pi)$. Denote by Π^+ the half-space determined by Π such that $M' \subset \Pi^+$. Then, it is not difficult to see that there exists an isometry of $\mathbb{L}^3$, f, that verifies $f(\Pi) = \Pi_0$ and $f(V(\theta, \delta, \delta') \cap \Pi^+) \subset V(\hat{\theta}, \hat{\delta}, \hat{\delta}')$, for some $\hat{\theta}$, $\hat{\delta}$ and $\hat{\delta}'$. Therefore, the corollary follows from Theorem 4.2. □

By contrast:

Proposition 4.4. *Let M be a connected properly immersed maximal surface without general conelike singularities in the interior such that $M \subset \mathsf{C}^+$ and ∂M is contained in a spacelike plane. Then M is a planar region.*

Proof. Our hypotheses imply that ∂M is compact. We consider the intersection of M with all the timelike planes $H(y, v)$ such that $\partial M \subset H^+(y, v)$. By applying Corollary 3.3 to the connected components of M contained in $H^-(y, v)$ we obtain that $M \subset H^+(y, v)$ for all the timelike planes described above. Then M is also compact and Proposition 2.10 proves that M is a planar region. □

5. Maximal surfaces whose boundary is contained in a lightlike plane and the convex hull property

As a consequence of the previous sections we deduce the following results for maximal surfaces whose boundary is contained in a lightlike plane.

Proposition 5.1. *There exists no connected properly immersed maximal surface M without general conelike singularities in the interior and such that $M \subset V(\theta, \delta, \delta')$ and ∂M is contained in a lightlike plane.*

Proof. Suppose there exists such an M. Let Π be the lightlike plane such that $\partial M \subset \Pi$. Then, we can consider the pencil of planes through the line $L = \Pi \cap \Pi_0$, that is, the set of planes sharing the line L. Since M cannot be flat, it is possible to find a spacelike or timelike plane in the pencil that intersects M transversally. But Corollaries 4.3 and 3.7 lead to a contradiction in each case. □

Proposition 5.2. *There exists no connected properly immersed maximal surface M without general conelike singularities in the interior and such that $M \subset \mathsf{C}^+$ and ∂M is contained in a lightlike plane.*

Proof. This can be demonstrated like Proposition 5.1, but using Proposition 4.4 and Corollary 3.3. □

As we saw in Section 2, a compact maximal surface lies in the convex hull of its boundary and the set of its general conelike singularities. This is not true for noncompact maximal surfaces in general. However, Theorem 4.2 and Proposition 4.4 can be seen as a convex hull type property. We have proved that if certain conditions are satisfied then the surfaces lie in the convex hull of their boundary. In the remainder of the section, we use the results obtained in the previous sections to give a generalization of these results. More precisely:

Proposition 5.3. *Any connected properly immersed maximal surface contained in $V(\theta, \delta, \delta')$ lies in the convex hull of its boundary and its general conelike singularities.*

Proof. Let M be a minimal surface satisfying the hypotheses of the proposition and denote by A the set of general conelike singularities of M. If M is contained in a plane the result is obvious. Assume then that M is not flat and consider $v \in \mathbb{S}^2$ and $y \in \mathbb{R}^3$ such that $(\partial(M) \cup A) \subset H^+(y, v)$. We have to prove that $M \subset H^+(y, v)$ too.

We proceed by contradiction, and suppose that $M \cap \big(H^-(y, v) - H(y, v)\big) \neq \varnothing$. Let M' be a connected component of $M \cap H^-(y, v)$.

If v is spacelike, so $H(y, v)$ is a timelike plane, Corollary 3.7 leads to a contradiction.

If v is timelike, so $H(y, v)$ is spacelike, the assumption that M' is not flat contradicts Corollary 4.3.

Finally, if v is lightlike, so is the plane $H(y, v)$, and Proposition 5.1 gives a contradiction. □

Proposition 5.4. *Any connected properly immersed maximal surface contained in* C^+ *lies in the convex hull of its boundary and its general conelike singularities.*

Proof. The proof is obtained as for the preceding proposition, using Corollary 3.3 and Propositions 5.2 and 4.4. □

Acknowledgments

We are indebted to F. Martín, I. Fernández and especially to F. J. López for helpful conversations during the preparation of this work. This paper was carried out during Ferrer's visit to IME-USP, Brazil. She is grateful to the people at the Institute for their hospitality. We also thank the referee for valuable suggestions.

References

[Bartnik and Simon 1982/83] R. Bartnik and L. Simon, "Spacelike hypersurfaces with prescribed boundary values and mean curvature", *Comm. Math. Phys.* **87**:1 (1982/83), 131–152. MR 84j:58126 Zbl 0512.53055

[Calabi 1970] E. Calabi, "Examples of Bernstein problems for some nonlinear equations", pp. 223–230 in *Global analysis* (Berkeley, 1968), vol. 2, edited by S.-S. Chern and S. Smale, Proc. Sympos. Pure Math. **15**, Amer. Math. Soc., Providence, RI, 1970. MR 41 #8806 Zbl 0211.12801

[Cheng and Yau 1976] S. Y. Cheng and S. T. Yau, "Maximal space-like hypersurfaces in the Lorentz-Minkowski spaces", *Ann. of Math.* (2) **104**:3 (1976), 407–419. MR 55 #4063 Zbl 0352.53021

[Dierkes et al. 1992] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab, *Minimal surfaces, I: Boundary value problems*, Grundlehren der Math. Wissenschaften **295**, Springer, Berlin, 1992. MR 94c:49001a Zbl 0777.53012

[Ecker 1986] K. Ecker, "Area maximizing hypersurfaces in Minkowski space having an isolated singularity", *Manuscripta Math.* **56**:4 (1986), 375–397. MR 88b:53075 Zbl 0594.58023

[Estudillo and Romero 1992] F. J. M. Estudillo and A. Romero, "Generalized maximal surfaces in Lorentz-Minkowski space L^3", *Math. Proc. Cambridge Philos. Soc.* **111**:3 (1992), 515–524. MR 93b:53010 Zbl 0824.53061

[Fernández and López 2004a] I. Fernández and F. J. López, "Periodic maximal surfaces in the Lorentz–Minkowski space $\mathbb{L}^3$", preprint, 2004. To appear in *Math. Z.* math.DG/0412461

[Fernández and López 2004b] I. Fernández and F. J. López, "Relative parabolicity of zero mean curvature surfaces in $\mathbb{R}^3$ and $\mathbb{R}^3_1$", preprint, 2004. math.DG/0410435

[Fernández et al. 2005] I. Fernández, F. J. López, and R. Souam, "The space of complete embedded maximal surfaces with isolated singularities in the 3-dimensional Lorentz-Minkowski space", *Math. Ann.* **332**:3 (2005), 605–643. MR 2006h:58012 Zbl 1081.53052

[Gulliver et al. 1973] R. D. Gulliver, II, R. Osserman, and H. L. Royden, "A theory of branched immersions of surfaces", *Amer. J. Math.* **95** (1973), 750–812. MR 50 #14595 Zbl 0295.53002

[Hoffman and Meeks 1990] D. Hoffman and W. H. Meeks, III, "The strong halfspace theorem for minimal surfaces", *Invent. Math.* **101**:2 (1990), 373–377. MR 92e:53010 Zbl 0722.53054

[Klyachin and Miklyukov 1993] A. A. Klyachin and V. M. Miklyukov, "Existence of solutions with singularities for the maximal surface equation in Minkowski space", *Mat. Sb.* **184**:9 (1993), 103–124. In Russian; translated in *Russian Acad. Sci. Sb. Math.* **80** (1995), 87–104. MR 95j:53014 Zbl 0824.53062

[Kobayashi 1983] O. Kobayashi, "Maximal surfaces in the 3-dimensional Minkowski space L^3", *Tokyo J. Math.* **6**:2 (1983), 297–309. MR 85d:53003 Zbl 0535.53052

[Kobayashi 1984] O. Kobayashi, "Maximal surfaces with conelike singularities", *J. Math. Soc. Japan* **36**:4 (1984), 609–617. MR 86d:53008 Zbl 0548.53006

[López and Martín 2001] F. J. López and F. Martín, "Minimal surfaces in a wedge of a slab", *Comm. Anal. Geom.* **9**:4 (2001), 683–723. MR 2002k:53013 Zbl 1026.53002

[López et al. 2000] F. J. López, R. López, and R. Souam, "Maximal surfaces of Riemann type in Lorentz-Minkowski space $\mathbb{L}^3$", *Michigan Math. J.* **47**:3 (2000), 469–497. MR 2002c:53009 Zbl 1029.53014

[Marsden and Tipler 1980] J. E. Marsden and F. J. Tipler, "Maximal hypersurfaces and foliations of constant mean curvature in general relativity", *Phys. Rep.* **66**:3 (1980), 109–139. MR 82d:83048

[Meeks and Rosenberg 1993] W. H. Meeks, III and H. Rosenberg, "The geometry and conformal structure of properly embedded minimal surfaces of finite topology in $\mathbf{R}^3$", *Invent. Math.* **114**:3 (1993), 625–639. MR 94i:53003 Zbl 0803.53007

[Miklyukov 1992] V. M. Miklyukov, "Maximal tubes and bands in Minkowski space", *Mat. Sb.* **183**:12 (1992), 45–76. In Russian; translated in *Russian Acad. Sci. Sb. Math.* **77**:2 (1994), 427–453. MR 94a:53019 Zbl 0801.53042

[Osserman 1971/72] R. Osserman, "The convex hull property of immersed manifolds", *J. Differential Geometry* **6** (1971/72), 267–270. MR 45 #7647 Zbl 0226.53009

[Schoen 1983] R. M. Schoen, "Uniqueness, symmetry, and embeddedness of minimal surfaces", *J. Differential Geom.* **18**:4 (1983), 791–809. MR 85f:53011 Zbl 0575.53037

[Umehara and Yamada 2006] M. Umehara and K. Yamada, "Maximal surfaces with singularities in Minkowski space", *Hokkaido Math. J.* **35**:1 (2006), 13–40. MR 2007a:53018 Zbl 05059482

Received December 15, 2005.

Rosa Maria Barreiro Chaves
Departamento de Matemática
Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão, 1010
05508-090 São Paulo, SP
Brazil

rosab@ime.usp.br
http://www.ime.usp.br

Leonor Ferrer
Departamento de Geometría y Topología
Universidad de Granada
18071, Granada
Spain

lferrer@ugr.es
http://www.ugr.es/~lferrer

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

ENERGY AND TOPOLOGY OF SINGULAR UNIT VECTOR FIELDS ON S^3

PABLO M. CHACÓN AND GIOVANNI DA SILVA NUNES

The energy of unit vector fields on odd-dimensional spheres is a functional that has a minimum in dimension 3 and an infimum in higher dimensions. Vector fields with isolated singularities arise naturally in the study of this functional. We consider the class of fields in $\mathbf{S}^3$ having two antipodal singularities. We prove a lower bound, attained for the radial vector field, for the energy of this class of fields in terms of the indices of the singularities. A similar inequality is not to be expected in other dimensions.

1. Introduction

Let (M^{n+1}, g) be a compact oriented Riemannian manifold without boundary. We denote by ∇ the Levi-Civita connection and by ν the volume form. Given a unit vector field $\vec{v}$ on M, the *energy* of $\vec{v}$ is (see [Wiegmink 1995])

$$\mathscr{E}(\vec{v}) = \frac{1}{2}\int_M \|\nabla\vec{v}\|^2\nu + \tfrac{1}{2}(n+1)\operatorname{vol} M. \tag{1}$$

This integral (times a constant) is also known as the *total bending* of $\vec{v}$ or the *vertical energy* of $\vec{v}$ as a section of the tangent bundle. The integrand $\frac{1}{2}\|\nabla\vec{v}\|^2$ is called the *energy density* of $\vec{v}$.

The absolute minimum of the energy functional is $\frac{1}{2}(n+1)\operatorname{vol} M$, attained for vector fields whose integral curves are geodesic and such that the orthogonal distribution $\vec{v}^\perp$ is integrable and totally geodesic. Such vector fields, called *parallel*, are rare because if M admits a unit parallel vector field then M is locally a Riemannian product.

The simplest spaces to be studied are spheres.

Theorem 1 [Brito 2000]. *Hopf vector fields are the only unit vector fields on* $\mathbf{S}^3$ *to minimize* $\mathscr{E}$.

MSC2000: primary 53C20; secondary 57R25, 58C25.
Keywords: energy of vector field, index of vector field.
Chacón is partially supported by MEC/FEDER Grant No. MTM2004-04934-C04-02 (Spain). This work was initiated during a postdoctoral stay of his funded by DGU No. HBE2002-008 (Spain).

By a Hopf vector field we mean a unit vector field tangent to the classical Hopf fibration. In dimensions 5 and higher, Hopf vector fields are unstable critical points of $\mathscr{E}$; see [Wood 2000; Gil-Medrano and Llinares-Fuster 2001].

The *radial* vector field V_R plays a fundamental role in the study of the energy. It is defined, relative to fixed antipodal points $N, S \in \mathbf{S}^{n+1}$, as the unit vector field tangent to the geodesics from N to S. It has $\{N, S\}$ as its only singularities.

Theorem 2 [Brito and Walczak 2000]. *The energy of any unit vector field $\vec{v}$ with isolated singularities on $\mathbf{S}^{n+1}$, $n \geq 2$, satisfies the inequality*

$$\mathscr{E}(\vec{v}) \geq \frac{n^2+n-1}{2(n-1)} \operatorname{vol} \mathbf{S}^{n+1},$$

and if $n \geq 3$, equality holds if and only if $\vec{v} = V_R$.

The existence of a global energy-minimizing unit vector field was left open at that time. This question has since been solved:

Theorem 3 [Borrelli et al. 2003]. *The infimum of $\mathscr{E}$ among all smooth, globally defined unit vector fields on sphere $\mathbf{S}^{2k+1}$, for $k \geq 2$, is*

$$\frac{4k^2+2k-1}{2(2k-1)} \operatorname{vol} \mathbf{S}^{2k+1}.$$

In fact, the proof of Theorem 3 in [Borrelli et al. 2003] shows a family of unit vector fields whose energy converges to $\mathscr{E}(V_R)$.

More results on energy can be found, for example, in [Berndt et al. 2003; Boeckx et al. 2002; Brito and Salvai 2004; Chacón and Naveira 2004; Gil-Medrano et al. 2004; Gil-Medrano and Hurtado 2004; González-Dávila and Vanhecke 2002; Wiegmink 1996].

From Theorem 1 and 3 it follows that the energy on $\mathbf{S}^{2k+1}$ has a minimum only in dimension 3. It is easy to see that on $\mathbf{S}^3$ Hopf vector fields and radial vector fields have the same energy. Thus vector fields with singularities appear naturally in the study of the energy of global unit vector fields.

In this paper we relate the energy and topology of vector fields with singularities. More precisely, we prove:

Theorem 4. *Let $\vec{v}$ be a unit vector field on $\mathbf{S}^3$ with exactly two antipodal singularities of index $\pm k$ with $k \in \mathbb{N}$. Then*

$$\mathscr{E}(\vec{v}) \geq \left(|2k-1| + \tfrac{3}{2}\right) \operatorname{vol}(\mathbf{S}^3).$$

The radial vector field V_R satisfies the equality in Theorem 4. The indices of V_R at N and S are respectively $+1$ and -1. For vector fields with indices ± 2, we do not know if there is a vector field for which equality obtains in Theorem 4.

A similar result is not expected for higher-dimensional spheres. The energy density of $\vec{v}$ has degree 2. The index of the singularity of a vector field on $\mathbf{S}^{n+1}$ is defined through a map from $\mathbf{S}^n$ to $\mathbf{S}^n$. The underlying technique in proving Theorem 4 is to relate the energy density (of degree 2) to an integral of a n-form of $\mathbf{S}^n$ that gives the index. For $n = 2$, all these objects are comparable but not for higher dimensions; the n-form of the index will not provide the energy density.

2. Notation and definitions

On $\mathbf{S}^3$, take two antipodal points $\{N, S\}$, determining the northern and southern hemispheres. Consider a vector field $\vec{v}$ on $\mathbf{S}^3$ such that $\vec{v}(N) = \vec{v}(S) = 0$ and $\|\vec{v}(p)\| = 1$ for $p \notin \{N, S\}$. The energy of $\vec{v}$ is defined in (1).

We fix an orientation for $\mathbf{S}^3$ and choose an oriented orthonormal local frame $\{e_1, e_2, e_3\}$ on $\mathbf{S}^3 \setminus \{N, S\}$ such that $e_3 = \vec{v}$. The corresponding dual 1-forms are denoted by $\{\theta_1, \theta_2, \theta_3\}$. The connection forms of the Levi-Civita connection ∇ of $\mathbf{S}^3$ will be denoted by ω_{ij}:

$$\omega_{ij}(X) = g(\nabla_X e_i, e_j) \quad \text{for } i, j = 1, 2, 3,$$

where X is a vector in the corresponding tangent space.

On $\mathbf{S}^3 \setminus \{N, S\}$ the field $\vec{v}$ determines a subbundle $\vec{v}^\perp$ of $T\mathbf{S}^3$ formed by the vectors orthogonal to $\vec{v}$. The second fundamental form of $\vec{v}^\perp$, possibly nonsymmetric, is given by the coefficients $h_{ij} = \omega_{j3}(e_i)$ for $i, j = 1, 2$. The energy density of $\vec{v}$ is

$$\tfrac{1}{2}\|\nabla\vec{v}\|^2 = \frac{1}{2}\left(\sum_{i,j=1}^{2} h_{ij}^2 + \sum_{i=1}^{2} a_i^2\right),$$

where the a_i are the components of the acceleration of $\vec{v}$:

$$\nabla_{e_3}\vec{v} = \nabla_{\vec{v}}\vec{v} = a_1 e_1 + a_2 e_2.$$

3. Proof of Theorem 4

Since $\mathbf{S}^3$ has Euler characteristic zero, the sum of the indices of $\vec{v}$ at the singularities must be zero. If the singularities of $\vec{v}$ have index zero, Theorem 4 follows from Theorem 2. We may therefore assume that $\vec{v}$ has nontrivial singularities.

We may also assume that N is the singularity with index $+k$ and S is the singularity with index $-k$. For simplicity, we also assume that $N = (0, 0, 0, 1) \in \mathbb{R}^4$ and $S = (0, 0, 0, -1) \in \mathbb{R}^4$.

Consider the 2-form on $\mathbf{S}^3 \setminus \{N, S\}$ given by $\omega_{13} \wedge \omega_{23}$. It is not hard to see that

$$\begin{aligned}(2)\quad \omega_{13} \wedge \omega_{23} = {} & (h_{11}h_{22} - h_{12}h_{21})\theta_1 \wedge \theta_2 \\ & + (h_{12}a_1 - h_{11}a_2)\theta_1 \wedge \theta_3 + (h_{22}a_1 - h_{21}a_2)\theta_2 \wedge \theta_3.\end{aligned}$$

This form is independent of the oriented frame chosen (satisfying $e_3 = \vec{v}$). Let

$$S_\theta = \left\{(x, y, z, t) \in \mathbf{S}^3 \mid t = \sin\theta\right\} \equiv \mathbf{S}^2(\cos\theta)$$

be the "parallel of latitude" in $\mathbf{S}^3$ determined by $\theta \in (-\pi/2, \pi/2)$. The field $\vec{v}$ has no singularities on any S_θ. Let $\alpha \in [0, \pi/2]$ be the angle between $\vec{v}$ and TS_θ. Note that $\cos\alpha$ and $\sin\alpha$ are well-defined functions on S_θ. Take an oriented orthonormal local frame $\{e_1, e_2, \vec{v}\}$ such that $e_1 \in TS_\theta$. We choose the orientation on the parallel S_θ given by the normal vector field V_R (pointing downward). Depending on the index of the singularity, along a fix parallel S_θ near the singularity, the field $\vec{v}$ can cross the plane TS_θ several times, or equivalently V_R and $\vec{v}$ can be on the same or on different sides of $T\mathbf{S}^3$ in relation with TS_θ. We define $\epsilon_1 = -1$ if TS_θ leaves V_R and $\vec{v}$ on different sides and $\epsilon_1 = 1$ otherwise. Next set

$$\tilde{e} = \sin\alpha\, e_2 + \epsilon_2 \cos\alpha\, \vec{v}, \tag{3}$$

where $\epsilon_2 = 1$ if TS_θ separates e_2 and $\vec{v}$ on different sides of $T\mathbf{S}^3$ and $\epsilon_2 = -1$ otherwise. Then $\{e_1, \epsilon_1\tilde{e}\}$ is an oriented local frame of S_θ. When $e_2 \in TS_\theta$, we have $\cos\alpha = 0$ and the value of ϵ_2 is irrelevant. For the special case $\vec{v} \in TS_\theta$, we have $\sin\alpha = 0$, and the oriented frame of S_θ is $\{e_1, -\vec{v}\}$ when e_2 agrees with V_R or $\{e_1, \vec{v}\}$ when $e_2 = -V_R$. Observe that although the positions of $\vec{v}$, V_R, TS_θ and e_2 can change, the frame $\{e_1, \epsilon_1\tilde{e}\}$ is a C^∞ local frame of S_θ (not global but differentiable).

If $i : S_\theta \to \mathbf{S}^3$ is the inclusion map, we have the 2-form $i^*(\omega_{13} \wedge \omega_{23})$ on S_θ. Using (2), we can evaluate

$$\begin{aligned} i^*(\omega_{13} \wedge \omega_{23})(e_1, \epsilon_1\tilde{e}) &= \epsilon_1 \omega_{13} \wedge \omega_{23}(e_1, \sin(\alpha)e_2 + \epsilon_2\cos(\alpha)\vec{v}) \\ &= \epsilon_1 \sin\alpha(h_{11}h_{22} - h_{12}h_{21}) + \epsilon_1\epsilon_2\cos\alpha(h_{12}a_1 - h_{11}a_2). \end{aligned}$$

This value depends only on $\vec{v}$.

We next compute

$$\begin{aligned} &\sum_{i,j=1}^{2} h_{ij}^2 + \sum_{i=1}^{2} a_i^2 \\ &\quad = (\sin\alpha)^2 h_{11}^2 + h_{22}^2 + (\sin\alpha)^2 h_{21}^2 + h_{12}^2 \\ &\qquad\qquad + (\epsilon_2\cos\alpha)^2 h_{21}^2 + a_1^2 + (\epsilon_2\cos\alpha)^2 h_{11}^2 + a_2^2 \\ &\quad \geq 2|\sin\alpha\, h_{11}h_{22}| + 2|\sin\alpha\, h_{21}h_{12}| + 2|\epsilon_2\cos\alpha\, h_{21}a_1| + 2|\epsilon_2\cos\alpha\, h_{11}a_2| \\ &\quad \geq 2|\sin\alpha\,(h_{11}h_{22} - h_{12}h_{21})| + 2|\epsilon_2\cos\alpha\,(h_{21}a_1 - h_{11}a_2)| \\ &\quad \geq 2\left|\sin\alpha\,(h_{11}h_{22} - h_{12}h_{21}) + \epsilon_2\cos\alpha\,(h_{12}a_1 - h_{11}a_2)\right| \\ &\quad = 2\left|i^*(\omega_{13} \wedge \omega_{23})(e_1, \tilde{e})\right| = 2\left|i^*(\omega_{13} \wedge \omega_{23})(e_1, \epsilon_1\tilde{e})\right| \end{aligned}$$

(recall that $x^2 + y^2 \geq 2|xy|$ for $x, y \in \mathbb{R}$). In this way, we can bound the integral in (1) from below by

$$\begin{aligned}\int_{\mathbf{S}^3} \|\nabla\vec{v}\|^2\nu &= \int_{\mathbf{S}^3}\Big(\sum_{i,j=1}^{2} h_{ij}^2 + \sum_{i=1}^{2} a_i^2\Big)\nu \geq \int_{-\pi/2}^{\pi/2}\int_{S_\theta} 2\left|i^*(\omega_{13}\wedge\omega_{23})(e_1, \epsilon_1\tilde{e})\right|\nu \\ &= \int_0^{\pi/2}\int_{S_\theta} 2\left|i^*(\omega_{13}\wedge\omega_{23})(e_1, \epsilon_1\tilde{e})\right|\nu_\theta\, d\theta \\ &\qquad + \int_{-\pi/2}^{0}\int_{S_\theta} 2\left|i^*(\omega_{13}\wedge\omega_{23})(e_1, \epsilon_1\tilde{e})\right|\nu_\theta\, d\theta,\end{aligned}$$

where ν_θ is the volume form of S_θ. Therefore,

$$\begin{aligned}\int_{\mathbf{S}^3} \|\nabla\vec{v}\|^2\nu \ \geq\ &\left|2\int_0^{\pi/2}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1, \epsilon_1\tilde{e})\nu_\theta\, d\theta\right| \\ &+\left|2\int_{-\pi/2}^{0}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1, \epsilon_1\tilde{e})\nu_\theta\, d\theta\right|.\end{aligned} \tag{4}$$

The integral over $(0, \pi/2)$ in (4) is related to the index of $\vec{v}$ at N and the integral over $(-\pi/2, 0)$ is related to the index of $\vec{v}$ at S. We consider the first integral.

By the structure equations of $\mathbf{S}^3$, the form $\omega_{13}\wedge\omega_{23}+\theta_1\wedge\theta_2$ is closed and hence

$$d(\omega_{13}\wedge\omega_{23}) = -d(\theta_1\wedge\theta_2). \tag{5}$$

To integrate the first term of (4), we consider the region of $\mathbf{S}^3$ defined by

$$A_{\phi\psi} = \{(x, y, z, t) \in \mathbf{S}^3 \mid \sin\phi \leq t \leq \sin\psi\},$$

where $0 \leq \phi < \psi \leq \pi/2$.

By Stokes' Theorem and the appropriate orientation of $A_{\phi\psi}$ we know that

$$\int_{A_{\phi\psi}} d(\omega_{13}\wedge\omega_{23}) = \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) - \int_{S_\phi} i^*(\omega_{13}\wedge\omega_{23}). \tag{6}$$

On the other hand, by (5) and Stokes' Theorem we have

$$\int_{A_{\phi\psi}} d(\omega_{13}\wedge\omega_{23}) = -\int_{A_{\phi\psi}} d(\theta_1\wedge\theta_2) = \int_{S_\phi} i^*(\theta_1\wedge\theta_2) - \int_{S_\psi} i^*(\theta_1\wedge\theta_2). \tag{7}$$

We can calculate the value of $i^*(\theta_1\wedge\theta_2)$:

$$i^*(\theta_1\wedge\theta_2)(e_1, \epsilon_1\tilde{e}) = \epsilon_1(\theta_1\wedge\theta_2)(e_1, \sin(\alpha)e_2 + \epsilon_2\cos(\alpha)\vec{v}) = \epsilon_1\sin\alpha. \tag{8}$$

So, for any ϕ, $0 \leq \phi < \pi/2$, by (6), (7) and (8) we obtain

$$
\begin{aligned}
(9)\qquad \int_{S_\phi} i^*(\omega_{13}\wedge\omega_{23}) &= \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) + \int_{S_\psi} i^*(\theta_1\wedge\theta_2) - \int_{S_\phi} i^*(\theta_1\wedge\theta_2)\\
&= \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) + \int_{S_\psi} \epsilon_1 \sin\alpha\,\nu_\psi - \int_{S_\phi} \epsilon_1 \sin\alpha\,\nu_\phi\\
&\geq \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) - \operatorname{vol} S_\psi - \operatorname{vol} S_\phi.
\end{aligned}
$$

From the definition of the index and the fact that the degree of a map $f : S_\psi \to \mathbf{S}^2$ is equal to $(\operatorname{vol}\mathbf{S}^2)^{-1}\int_{S_\psi} f^*\eta$, where η is a volume form of $\mathbf{S}^2$, we obtain an integral expression for the index of $\vec{v}$ at the singularity N:

$$
(10)\qquad \lim_{\psi\to\frac{\pi}{2}} \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) = 4\pi k.
$$

From (9) and (10) we obtain $\int_{S_\phi} i^*(\omega_{13}\wedge\omega_{23}) \geq 4\pi k - 4\pi(\cos\phi)^2$. Now, integrating on the northern hemisphere we have

$$
\begin{aligned}
(11)\qquad & 2\int_0^{\pi/2}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1,\epsilon_1\tilde{e})\nu_\theta\, d\theta\\
&\qquad \geq 2\int_0^{\pi/2}\big(4\pi k - 4\pi\cos^2\theta\big)\, d\theta = 4\pi^2 k - 2\pi^2 > 0.
\end{aligned}
$$

Taking absolute values in (11) we obtain

$$
(12)\qquad \left| 2\int_0^{\pi/2}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1,\epsilon_1\tilde{e})\nu_\theta\, d\theta \right| \geq |4\pi^2 k - 2\pi^2|.
$$

For the second integral of (4) we work in a similar way. Now the index of $\vec{v}$ at S is $-k < 0$ and the aim is to obtain an inequality equivalent to (11) in such a way that we know the sign of both sides of the inequality.

For $-\pi/2 \leq \phi < \psi \leq 0$, the inequality $\int_{S_\phi} i^*(\omega_{13}\wedge\omega_{23}) \geq \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) - \operatorname{vol} S_\psi - \operatorname{vol} S_\phi$ in (9) still holds. Now, the index of $\vec{v}$ at S is given by

$$
(13)\qquad \lim_{\phi\to-\pi/2} \int_{S_\phi} i^*(\omega_{13}\wedge\omega_{23}) = -4\pi k.
$$

From this and the inequality in (9) we obtain $-4\pi k \geq \int_{S_\psi} i^*(\omega_{13}\wedge\omega_{23}) - 4\pi(\cos\psi)^2$. So, on the southern hemisphere we have

$$
\begin{aligned}
& 2\int_{-\pi/2}^{0}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1,\epsilon_1\tilde{e})\nu_\theta\, d\theta\\
&\qquad \leq 2\int_{-\pi/2}^{0}\big(-4\pi k + 4\pi\cos^2\theta\big)\, d\theta = -4\pi^2 k + 2\pi^2 < 0.
\end{aligned}
$$

For the absolute value we get the bound

$$\left|2\int_{-\pi/2}^{0}\int_{S_\theta} i^*(\omega_{13}\wedge\omega_{23})(e_1,\epsilon_1\tilde{e})\nu_\theta\,d\theta\right| \geq |-4\pi^2k+2\pi^2| = |4\pi^2k-2\pi^2|. \tag{14}$$

Finally, the energy of $\vec{v}$ follows from (4), (12) and (14) that is bounded by:

$$\begin{aligned}\mathscr{E}(\vec{v}) &\geq \tfrac{1}{2}\left|4\pi^2k-2\pi^2\right| + \tfrac{1}{2}\left|4\pi^2k-2\pi^2\right| + \tfrac{3}{2}\operatorname{vol}\mathbf{S}^3\\ &\geq \tfrac{1}{2}\left|4\pi^2k+4\pi^2k-4\pi^2\right| + \tfrac{3}{2}\operatorname{vol}\mathbf{S}^3 = \left(|2k-1|+\tfrac{3}{2}\right)\operatorname{vol}\mathbf{S}^3.\end{aligned}$$

This completes the proof of Theorem 4. □

Remarks. (1) When $\vec{v} = V_R$, equality is attained in Theorem 4. This can be shown by direct computation, but also by reproducing the proof with the following considerations: the acceleration of V_R is zero, $V_R^\perp$ is integrable and umbilical, and the angle α is equal to $\pi/2$.

(2) In a sense, the assumption of Theorem 4 about the quantity and the antipodal position of the singularities is a necessary restriction. It is possible to construct a sequence of unit vector fields, each with an arbitrary number of singularities in free positions, whose energy converges to the energy of the radial vector field V_R.

Acknowledgements

We thank Professor Fabiano Brito from the University of São Paulo (Brazil) for encouraging us to tackle this line of investigations.

References

[Berndt et al. 2003] J. Berndt, L. Vanhecke, and L. Verhóczki, "Harmonic and minimal unit vector fields on Riemannian symmetric spaces", *Illinois J. Math.* **47** (2003), 1273–1286. MR 2005i:53064 Zbl 1045.53036

[Boeckx et al. 2002] E. Boeckx, J. C. González-Dávila, and L. Vanhecke, "Stability of the geodesic flow for the energy", *Comment. Math. Univ. Carolin.* **43**:2 (2002), 201–213. MR 2003g:53147 Zbl 02185229

[Borrelli et al. 2003] V. Borrelli, F. Brito, and O. Gil-Medrano, "The infimum of the energy of unit vector fields on odd-dimensional spheres", *Ann. Global Anal. Geom.* **23**:2 (2003), 129–140. MR 2003m:53046 Zbl 1031.53090

[Brito 2000] F. G. B. Brito, "Total bending of flows with mean curvature correction", *Differential Geom. Appl.* **12**:2 (2000), 157–163. MR 2001g:53065 Zbl 0995.53023

[Brito and Salvai 2004] F. Brito and M. Salvai, "Solenoidal unit vector fields with minimum energy", *Osaka J. Math.* **41**:3 (2004), 533–544. MR 2005h:53044 Zbl 1076.53078

[Brito and Walczak 2000] F. G. B. Brito and P. Walczak, "On the energy of unit vector fields with isolated singularities", *Ann. Polon. Math.* **73**:3 (2000), 269–274. MR 2001k:53049 Zbl 0997.53025

[Chacón and Naveira 2004] P. M. Chacón and A. M. Naveira, "Corrected energy of distributions on Riemannian manifolds", *Osaka J. Math.* **41**:1 (2004), 97–105. MR 2004k:53043 Zbl 1055.53021

[Gil-Medrano and Hurtado 2004] O. Gil-Medrano and A. Hurtado, "Spacelike energy of timelike unit vector fields on a Lorentzian manifold", *J. Geom. Phys.* **51**:1 (2004), 82–100. MR 2005f:53115 Zbl 1076.53086

[Gil-Medrano and Llinares-Fuster 2001] O. Gil-Medrano and E. Llinares-Fuster, "Second variation of volume and energy of vector fields. Stability of Hopf vector fields", *Math. Ann.* **320**:3 (2001), 531–545. MR 2002m:53100 Zbl 0989.53020

[Gil-Medrano et al. 2004] O. Gil-Medrano, J. C. González-Dávila, and L. Vanhecke, "Harmonicity and minimality of oriented distributions", *Israel J. Math.* **143** (2004), 253–279. MR 2005g:58004 Zbl 1073.58007

[González-Dávila and Vanhecke 2002] J. C. González-Dávila and L. Vanhecke, "Energy and volume of unit vector fields on three-dimensional Riemannian manifolds", *Differential Geom. Appl.* **16**:3 (2002), 225–244. MR 2003b:53032 Zbl 1035.53089

[Wiegmink 1995] G. Wiegmink, "Total bending of vector fields on Riemannian manifolds", *Math. Ann.* **303**:2 (1995), 325–344. MR 97a:53050 Zbl 0834.53034

[Wiegmink 1996] G. Wiegmink, "Total bending of vector fields on the sphere S^3", *Differential Geom. Appl.* **6**:3 (1996), 219–236. MR 97e:53067 Zbl 0866.53025

[Wood 2000] C. M. Wood, "The energy of Hopf vector fields", *Manuscripta Math.* **101**:1 (2000), 71–88. MR 2001c:53090 Zbl 0946.53032

Received November 7, 2005. Revised February 2, 2007.

PABLO M. CHACÓN
DEPARTAMENTO DE MATEMÁTICAS
FACULTAD DE CIENCIAS
UNIVERSIDAD DE SALAMANCA
PLAZA DE LA MERCED 1-4
37008 SALAMANCA
SPAIN

pmchacon@usal.es

GIOVANNI DA SILVA NUNES
UNIVERSIDADE FEDERAL DE PELOTAS
INSTITUTO DE FÍSICA E MATEMÁTICA
CAMPUS UNIVERSITÁRIO - CAPÃO DO LEÃO
CAIXA POSTAL 354
96010-900 PELOTAS, RS
BRAZIL

gsnunes@portoweb.com.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

SINGULAR ANGLES OF WEAK LIMITING METRICS UNDER CERTAIN INTEGRAL CURVATURE BOUNDS

QING CHEN, XIUXIONG CHEN AND WEIYONG HE

We prove that a nonvanishing weak limit of Riemannian metrics in surfaces with an integral curvature bound admits only weak cusp singularities. The result is useful toward generalizing classical uniformization theory to surfaces with boundary.

1. Introduction

Classical uniformization theorem says that in a compact Riemann surface without boundary there is a constant curvature metric in any conformal class of metrics. There have been many attempts to generalize this theory to surfaces with boundary. The main focus, started by the independent work of Troyanov [1991] and McOwen [1988], has been to study the existence or nonexistence of constant curvature metrics in surfaces with conical singularities. Much work has followed since; see, for example, [Chen and Li 1991; Chang and Yang 1988; Luo and Tian 1992].

The disadvantage of this classical approach is that the Gaussian curvature is a second-order differential operator of the metric, while the condition of prescribing conical singularities is equivalent to prescribing both the metric and its derivatives near the set of singular points at infinitesimal level. Thus, the constant curvature equation with prescribed conical singularities is an overdetermined elliptic equation. In general, one should not expect to get a clear-cut statement about the existence (or nonexistence) of solutions.

We now describe another approach. Given a compact Riemann surface Ω and a Hermitian metric g_0 on Ω, any metric g on Ω is said to be conformal to g_0 if there exists a smooth function $e^{2\varphi}$ such that $g = e^{2\varphi} g_0$ on Ω. Define the variational space $\mathcal{G}(\Omega)$ to be the set of all metrics that are conformal to g_0 and agree with g_0 on $\partial\Omega$ up to first derivatives. For each $g \in \mathcal{G}(\Omega)$, define the energy functional

$$E(g) = \int_\Omega K^2 \, dg,$$

MSC2000: primary 53C20; secondary 53C23.
Keywords: Riemannian metric, weak limit, singular angle.
Qing Chen was supported by the NCFT Program of the Ministry of Education of China. Xiuxiong Chen was partially supported by an NSF grant.

where K is the Gaussian curvature of g. The problem is to minimize this functional with the area constraint

$$A(g) = \int_\Omega dg = \text{const}.$$

A critical point of this functional is called an extremal Hermitian metric; see [Calabi 1982]. The Euler–Lagrange equation is

$$\triangle_g K + K^2 = C,$$

where we have, in a local system,

$$-\triangle\varphi = Ke^{2\varphi} \quad \text{if } g = e^{2\varphi}\,|dz|^2.$$

By a theorem of Calabi, any metric solving this Euler–Lagrange equation on a surface without boundary must have constant scalar curvature. This result conforms with the classical uniformization theorem.

One of us (X. Chen) has been using the approach just described, namely the study of the variational problem of minimizing $E(g)$ in $\mathscr{G}(\Omega)$ with fixed area, with the goal of generalizing uniformization results to surfaces with boundary. Some of the ideas go back to E. Calabi (private communications).

To attack this variational problem, one first studies the weak compactness of any subset of metrics in $\mathscr{G}(\Omega)$ with finite energy and area. Generally speaking, for a sequence of metrics $\{g_n, n \in \mathbb{N}\}$ with finite energy and area, there exists a weak limit metric $\underline{g}$ such that g_n converges to $\underline{g}$ weakly in any compact subset away from a finite set of singular points. An important feature of the limit is the *bubbling phenomenon*, first observed by Sacks and Uhlenbeck in 1979, when they studied the existence theorem for harmonic maps between two spheres [Uhlenbeck 1982]. The key observation was that the noncompactness is associated with the concentration of the energy density at isolated *bubble points*. Around each such point, one can define a *weak singular angle* in the approximation sense (see Definition 2 below). If the weak singular angle is 0, it is called a weak cusp singular point. An intriguing question is that whether all points is weak cusp singular points if the weak limit metric $\underline{g} \neq 0$. In this paper, we will prove (see Section 2 for terminology and notation):

Theorem 1.1. *Let $\{g_n\}$ be a sequence of conformal metrics in domain Ω with finite area and finite energy. There exist a subsequence of $\{g_n\}$, a limit metric g_0 and a finite set of bubble points $\{p_1, p_2, \dots, p_m\}$ such that $g_n \rightharpoonup g_0$ in $H^{2,2}_{\text{loc}}(\Omega \setminus \{p_1, p_2, \dots, p_m\})$. If $g_0 \neq 0 (\varphi_0 \neq -\infty)$, then g_0 has a weak cusp singularity at each bubble point p_i. There is no ghost vertex in the bubble-tree decomposition.*

This motivates the following classification result:

Theorem 1.2 [Chen 1999]. *Let Ω be a surface without boundary and let g be an extremal Hermitian metric in $\Omega \setminus \{p_1, \dots, p_n\}$ with finite energy and area, and having only weak cusp singularities at all singular points.*

(1) *If $\chi(\Omega) \leqslant 0$, then K is a negative constant.*

(2) *If $n \geqslant 3$ and $\chi(\Omega) = 2$ (that is, Ω is a sphere), then K is a negative constant.*

(3) *If $n = 2$ and $\chi(\Omega) = 2$, there exists no extremal Hermitian metric.*

(4) *If $n = 1$ and $\chi(\Omega) = 2$, there exists a unique extremal Hermitian metric determined by total area, and the metric must be rotationally symmetric.*

These results are critical for the generalization of uniformization theory to a surface with boundary.

The main result was stated in [Chen 1998a; 1999] without proof and was used crucially in [Chen 2001], in deriving the long-time existence of the Calabi flow in Riemannian surfaces. At present, there is strong interest in the Calabi flow for general Kähler manifolds. Clarification of this important technical step will likely be indicative of what happens in more general settings.

2. The problem from an analytic viewpoint

In a coordinate chart (D, z), any metric g can be written as

$$g = e^{2\varphi}(dx^2 + dy^2),$$

and the curvature function is

$$K = -\frac{\triangle\varphi}{e^{2\varphi}}.$$

A metric is said to have finite area C_1 and finite energy C_2 if

$$\int_D e^{2\varphi}\,dx\,dy \leqslant C_1 \quad \text{and} \quad \int_D \frac{(\triangle\varphi)^2}{e^{2\varphi}}\,dx\,dy \leqslant C_2. \tag{2-1}$$

A sequence of metrics $\{g_n\}$, where $g_n = e^{2\varphi_n}(dx^2 + dy^2)$, is said to have finite area C_1 and finite energy C_2 if each φ_n satisfies (2-1). From now on we will always use either $\{\varphi_n\}$ or $\{g_n\}$ to denote a sequence of metrics with finite area and finite energy.

For any subdomain Ω in D, relabel the energy and area for a conformal parameter function as

$$A_c(\varphi, \Omega) = \int_\Omega e^{2\varphi}\,dx\,dy, \quad K_c(\varphi, \Omega) = \int_\Omega \frac{(\triangle\varphi)^2}{e^{2\varphi}}\,dx\,dy. \tag{2-2}$$

A "zero metric" should have zero area and zero energy. Since the zero metric has a conformal parameter function identically equal to $-\infty$, we define $A_c(-\infty, \Omega) = K_c(-\infty, \Omega) = 0$. For notational convenience, define $\hat{H}^{2,2}_{\mathrm{loc}}(\Omega) = H^{2,2}(\Omega) \cup \{-\infty\}$.

A sequence of functions $\{\varphi_n\} \in H^{2,2}(\Omega)$ is said to converge weakly to a function φ_0 in $\hat{H}^{2,2}_{\text{loc}}(\Omega)$ if one of two mutually exclusive alternatives holds:

1 (Vanishing case): If $\varphi_0 \equiv -\infty$, then $\varphi_n \to -\infty$ uniformly in any relatively compact subdomain of Ω.

2 (Nonvanishing case): If $\varphi_0 \in H^{2,2}(\Omega)$, then $\varphi_n \rightharpoonup \varphi_0$ weakly in $H^{2,2}_{\text{loc}}(\Omega)$.

Consequently, a sequence of Riemannian metrics $\{g_n\} \in H^{2,2}(\Omega)$ converges weakly to a limit metric g_0 in $H^{2,2}_{\text{loc}}(\Omega)$ if and only if either

1 $g_n \to 0$ everywhere (and $g_0 \equiv 0$), or

2 $\varphi_n \rightharpoonup \varphi_0$ in $H^{2,2}_{\text{loc}}(\Omega)$, where $g_n = e^{2\varphi_n} g_{\text{bk}}$ and $g_0 = e^{2\varphi_0} g_{\text{bk}}$, for g_{bk} a smooth background metric in Ω.

Definition 2.1. *A point p is said to be a bubble point of $\{\varphi_n\}$ if, for any $r > 0$,*

$$\liminf_{n\to\infty} \int_{D_r(p)} \frac{\triangle(\varphi_n)^2}{e^{2\varphi_n}}\, dx\, dy \geqslant \alpha > 0 \quad \textit{and} \quad \liminf_{n\to\infty} \int_{D_r(p)} e^{2\varphi_n}\, dx\, dy \geqslant \beta > 0,$$

where $D_r(p)$ is a coordinate disk centered at p with radius r.

The largest possible numbers α and β are the concentration weights of the energy function and area function at point p.

For convenience we restate here the three main theorems from [Chen 1998b]. Their proofs can be found there.

Theorem 2.2. *Let $\{\varphi_n,\ n \in \mathbb{N}\}$ be a sequence of metrics in $H^{2,2}(D)$ with finite area C_1 and energy C_2. There exists a subsequence $\{\varphi_{n_j},\ j \in \mathbb{N}\}$ of $\{\varphi_n\}$, a finite number of bubble points $\{p_1, p_2, \dots, p_m\}$ with respect to $\{\varphi_{n_j},\ j \in \mathbb{N}\}$ (where $0 \leqslant m \leqslant \sqrt{C_1 C_2/(4\pi)}$), and a metric $\varphi_0 \in \hat{H}^{2,2}_{\text{loc}}(D \setminus \{p_1, p_2, \dots, p_m\})$ such that*

$$\varphi_{n_j} \rightharpoonup \varphi_0 \quad \textit{in } \hat{H}^{2,2}_{\text{loc}}(D \setminus \{p_1, p_2, \dots, p_m\}).$$

Theorem 2.3 (Bubbles on bubbles). *Let $\{\varphi_n\}$ be a sequence of metrics in D with finite area C_1 and finite energy C_2. Suppose that $p = 0$ is the only bubble point in D, that it has area concentration A_p and energy concentration K_p, and that there exists a metric $\varphi_0 \in \hat{H}_{2,2}(D \setminus \{p\})$ such that $\varphi_n \rightharpoonup \varphi_0$ in $\hat{H}_{2,2}(D \setminus \{p\})$. A sequence of numbers $\{\delta_n \searrow 0\}$ can be chosen to renormalize the sequence of metrics as $\phi_n(x, y) = \varphi_n(\delta_n x, \delta_n y) + \log \delta_n$, for $n \in \mathbb{N}$.*

There exists a subsequence $\{\varphi_{n_j},\ j \in \mathbb{N}\}$ of $\{\varphi_n\}$, a finite number of bubble points $\{q_1, q_2, \dots, q_m\}$ with respect to the subsequence of metrics $\{\phi_{n_j}\}$ (where $0 \leqslant m \leqslant \sqrt{A_p K_p/(4\pi^2)}$), and a metric $\phi_0 \in \hat{H}^{2,2}(S^2 \setminus \{\infty, q_1, q_2, \dots, q_m\})$ such that

$$\phi_{n_j} \rightharpoonup \phi_0 \in \hat{H}^{2,2}_{\text{loc}}(S^2 \setminus \{\infty, q_1, q_2, \dots, q_m\}).$$

If $\phi_0 \equiv -\infty$ (vanishing case), then $m \geqslant 2$ and p is a bubble point of $\{\phi_{n_j},\ j \in \mathbb{N}\}$.

Theorem 2.4 (Bubble tree). *The limit of any locally weakly convergent sequence of metrics $\{g_k,\ k \in \mathbb{N}\}$ in $\mathcal{G}(\Omega)$ encompasses the following data*:

(1) *A finite rooted tree T, possibly reduced to just the base vertex f.*

(2) *The base vertex $f \in T$ is a limit metric in Ω with a finite number of bubble points $\{p_i\}$ deleted*; *the edge emanating from the base vertex is $\{p_i\}$.*

(3) *Any other vertex f_s is a limit metric defined on $S^2 \setminus \{\infty, p_{s_i}\}$*; *the edges emanating from this vertices are $\{p_{s_i}\}$.*

(4) *For each pair of vertex f_{s_1} and f_{s_2} bounding a common edge in T, they are tenuously connected at the pair of respective singular points. The number of vertices whose valence $\neq 2$, is bounded from above by $\sqrt{C_1 C_2}$. The depth of the tree is also finite in a reasonable sense. Each vertex $f_I = f_{i_1 i_2 \cdots i_k}$ has the property that*, *if it vanishes in any point in its domain*, *it vanishes everywhere*; *in this case we call it a ghost vertex. The number of ghost vertices is finite.*

Definition 2.5. *Around each bubble point*, *define the weak singular angle*

$$\alpha = \lim_{r \to 0} \int_{\partial D_r(p)} k_g ds_g = \lim_{r \to 0} \lim_{n \to \infty} \int_{\partial D_r(p)} k_{g_n} ds_{g_n}$$

if the limit is well defined. Here k_g and k_{g_n} are geodesic curvatures of g and g_n.

The weak angle does exist, as we shall see in the next section, and has some interesting properties. From now on, we will always assume that $\{g_n\}$ converges, by substituting a convergent subsequence if necessary.

If the weak singular angle is zero, the bubble point p is called a *weak cusp singularity*. Our purpose is to prove:

Main Theorem 2.6. *Let $\{g_n\}$ be a sequence of conformal metrics in Ω with finite area and finite energy. There exists a subsequence of $\{g_n\}$, a limit metric g_0 and a finite set of bubble points $\{p_1, p_2, \ldots, p_m\}$ such that $g_n \rightharpoonup g_0$ in $H^{2,2}_{\text{loc}}(\Omega \setminus \{p_1, p_2, \ldots, p_m\})$* (following the notations of Theorem 1.1). *If $g_0 \neq 0$, then g_0 has a weak cusp singularity at each bubble point p_i, and there is no ghost vertex in the bubble-tree decomposition.*

3. Asymptotic geometry of singular points

In this section, we introduce the blowing-up process and the definition of a neck. We also prove several lemmas that are essential to our main theorem.

Definition 3.1. *For any metric φ with finite area C_1 and energy C_2 in $D \setminus \{0\}$*, *set*

$$\bar{\varphi}(r) = \frac{1}{2\pi} \int_0^{2\pi} \varphi(r \cos\theta, r \sin\theta)\, d\theta.$$

Lemma 3.2 [Chen 1998b]. *Suppose φ is a metric with finite area C_1 and energy C_2 in $D\setminus\{0\}$.*

(1) $\lim_{r\to 0}(\varphi(r\cos\theta, r\sin\theta)+\log r) = -\infty$.

(2) $\lim_{r\to 0}\bar{\varphi}_r(r)\, r$ *exists and is finite.*

(3) *There exists constants $\lambda \in (0, 1)$ and C_3, C_4 such that*

$$\frac{1}{\lambda}(\bar{\varphi}(r)+\log r)+C_3 \le \varphi(r\cos\theta, r\sin\theta)+\log r \le \lambda(\bar{\varphi}(r)+\log r)+C_4.$$

Similarly, one can prove:

Corollary 3.3. *Suppose ϕ is a metric in $R^2\setminus\bar{D}_{r_0}(0)$ with finite area C_1 and energy C_2, where $\bar{D}_{r_0}(0)$ is a closed disk with radius r_0.*

(1) $\lim_{r\to\infty}(\phi(r\cos\theta, r\sin\theta)+\log r) = -\infty$.

(2) $\lim_{r\to\infty}\bar{\phi}_r(r) r$ *exists and is finite.*

(3) *There exists constants $\mu \in (0, 1)$ and C_5, C_6 such that*

$$\frac{1}{\mu}(\bar{\varphi}(r)+\log r)+C_5 \le \varphi(r\cos\theta, r\sin\theta)+\log r \le \mu(\bar{\varphi}(r)+\log r)+C_6.$$

Lemma 3.4 [Chen 1998b]. *Let $\{\varphi_n\}$ be a sequence of metrics with finite area C_1 and energy C_2. There exists a constant ϵ_0 such that if*

$$\max_{r\leqslant\rho}|\partial D_r|_{g_n} = \max_{r\leqslant\rho}\int_0^{2\pi} e^{\varphi_n(r\cos\theta, r\sin\theta)}\, r\, d\theta \leqslant \epsilon_0 \quad \textit{for all } n\in\mathbb{N},$$

then $\{\varphi_n\}$ has no bubble points in D.

We can now explain the blowing-up process. To simplify the problem, let $\{g_n = e^{\varphi_n}(dx^2+dy^2)\}$ be a sequence of metrics in D with finite area and finite energy, converging to a limit metric $g = e^{\varphi_0}(dx^2+dy^2)\neq 0$, and having $\{0\}$ as its only bubble point. Following Lemma 3.2, we have

$$\lim_{r\to 0}\max_{0\leqslant\theta\leqslant 2\pi}(\varphi_0(r\cos\theta, r\sin\theta)+\log r) = -\infty.$$

Then there exists $r_1 > 0$ such that

$$\max_{0\leqslant\theta\leqslant 2\pi}(\varphi_0(r\cos\theta, r\sin\theta)+\log r)\ll 0 \quad \text{for all } r\le r_1.$$

If n is large enough, the convergence implies that

$$\max_{0\leqslant\theta\leqslant 2\pi}(\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1)\ll 0 \quad \text{for all } n > N;$$

equivalently, if the length of the circle $|z| = r_1$ is very small, then

$$|\partial D_{r_1}|_{g_n} = \int_0^{2\pi} e^{\varphi_n(r_1 \cos\theta, r_1 \sin\theta)} r_1 \, d\theta \le \varepsilon \quad \text{for all } n > N.$$

According to Lemma 3.4, if ε is small enough, for each $n > N$ we can choose δ_n such that

$$|\partial D_r|_{g_n} = \int_0^{2\pi} e^{\varphi_n(r \cos\theta, r \sin\theta)} r \, d\theta < \varepsilon \quad \text{for all } r \text{ with } r_1 \ge r > \delta_n,$$

and

$$|\partial D_{\delta_n}|_{g_n} = \int_0^{2\pi} e^{\varphi_n(\delta_n \cos\theta, \delta_n \sin\theta)} \delta_n \, d\theta = \varepsilon.$$

Renormalize this sequence of metrics as

$$\phi_n(z) = \varphi(\delta_n z) + \log \delta_n \quad \text{for all } |z| < \frac{1}{\delta_n}$$

and

$$\tilde{g}_n(z) = e^{2\phi_n(z)} |dz|^2 = g_n(\delta_n z).$$

The theorems in Section 2 assert that there is a subsequence $\{\tilde{g}_{n_j}\}$ of metrics $\{\tilde{g}_n\}$, a limit metric $\tilde{g}$ and finitely many bubble points $\{q_1, q_2, \ldots, q_m\}$ such that either

$$\tilde{g} \equiv 0 (\phi_{n_j} \to -\infty) \quad \text{or} \quad \tilde{g}_{n_j} \rightharpoonup \tilde{g}$$

in $\hat{H}^{2,2}_{\text{loc}}(S^2 \setminus \{\infty, q_1, q_2, \ldots, q_m\})$.

Choose r_2 big enough that $\{q_1, q_2, \ldots, q_m\} \subset D_{r_2}$. The annulus bounded by the two circles $|z| = r_1$ and $|z| = r_2 \delta_n$ is called the neck of the blowing-up process and denoted by $\text{Neck}(r_1, r_2)$. The blowing up procedure or the renormalization procedure depends only on the filter size $\varepsilon > 0$ (once a coordinate system is fixed). Since $g_n \rightharpoonup g_0$ in $\hat{H}^{2,2}_{\text{loc}}(D \setminus \{p\})$, the surface $(g_0, D \setminus \{p\})$ and $(\tilde{g}_0, S^2 \setminus \{\infty, q_1, q_2, \ldots, q_m\})$ are called tenuously connected at p and at $z = \infty$.

Lemma 3.5. *If the limit metric g is nonzero, the weak angle α exists and is finite, and $\alpha \geqslant 0$.*

Proof. In polar coordinates (r, θ), we have

$$\alpha = \lim_{r \to 0} \int_{\partial D_r(p)} k_g \, ds_g = \frac{1}{2\pi} \lim_{r \to 0} \int_0^{2\pi} (\partial_r \varphi(r \cos\theta, r \sin\theta) r + 1) \, d\theta = \lim_{r \to 0} \bar{\varphi}'(r) + 1.$$

According to Lemma 3.2, the weak angle exists and is finite. Now we only have to prove $\alpha \ge 0$. Assume $\alpha < 0$. By the preceding equality we know that $\bar{\varphi}_r(r) r + 1 < 0$ for r small enough, say for $r < r_0$; that is,

$$\bar{\varphi}_r(r) + \frac{1}{r} < 0 \quad \text{when } r < r_0.$$

Then, for $0 < r_1 < r_2 < r_0$,

$$0 > \int_{r_1}^{r_2} \left(\bar{\varphi}_r(r) + \frac{1}{r}\right) = (\bar{\varphi}(r) + \log r)|_{r1}^{r_2} = (\bar{\varphi}(r_2) + \log r_2) - (\bar{\varphi}(r_1) + \log r_1).$$

Letting $r_1 \to 0$, since $\bar{\varphi}(r_1) + \log r_1 \to -\infty$, it follows that

$$0 > \int_{r_1}^{r_2} \left(\bar{\varphi}_r(r) + \frac{1}{r}\right) = (\bar{\varphi}(r_2) + \log r_2) - (\bar{\varphi}(r_1) + \log r_1) \to \infty,$$

which is a contradiction. □

Definition 3.6. *Let $\{\varphi_n\}$ be a sequence of metrics in D with finite area C_1 and finite energy C_2. Suppose that $p = 0$ is the only bubble point in D in the blowing-up process just described in Lemma 3.4. We define the out angle β around p in* Neck(r_1, r_2) *as*

$$\beta = \lim_{r_2\to\infty} \lim_{n\to\infty} \int_{\partial D_{r_2\delta_n}(p)} k_{g_n} ds_n$$

if the limit exists, where k_{g_n} is the geodesic curvature.

Lemma 3.7. *In the notations in Lemma 3.4, if the (renormalized) limit metric ϕ_0 is not $-\infty$, the out angle β exists and is finite. Furthermore, $\beta \leqslant 0$.*

Proof. By the definition of β, we have

$$\beta = \lim_{r_2\to\infty} \lim_{n\to\infty} \int_{\partial D_{r_2\delta_n}(p)} k_{g_n} ds_n = \lim_{r_2\to\infty} \lim_{n\to\infty} \int_{\partial D_{r_2}(p)} k_{\tilde{g}_n} d\tilde{s}_n.$$

If the limit metric $\tilde{g}_0$ does not vanish, that is, if $\phi_0 \neq -\infty$, we have

$$\beta = -\lim_{r_2\to\infty} \int_{\partial D_{r_2}} k_{\tilde{g}_0} d\tilde{s}_0.$$

By Lemma 3.2, β is well defined and finite. Just as the argument in the proof of Lemma 3.7, we can then prove that $\beta \leqslant 0$. □

Remark 3.8. The out angle β (more precisely, its negative) can be regarded as the weak angle at $\{\infty\}$ in the blowing-up process.

When the limit metric vanishes, the weak angle can still be defined:

Lemma 3.9. *If a sequence of metrics $\{g_n = e^{\varphi_n}(dx^2 + dy^2)\}$ on D has vanishing limit metric and $\{0\}$ as its only bubble point, the weak angle at $\{0\}$ exists and finite.*

Proof. It suffices to show the existence of the limit

$$\alpha = \lim_{r\to 0} \lim_{n\to\infty} \int_{\partial D_r(p)} k_{g_n} ds_{g_n} = \lim_{r\to 0} \lim_{n\to\infty} \int_0^{2\pi} (\partial_r \varphi_n r + 1)\, d\theta.$$

But for any $0 < r_1 < r_2$,

$$\begin{aligned}\left|\int_0^{2\pi}(\partial_r\varphi_n(r_2\cos\theta, r_2\sin\theta)r_2+1)\,d\theta-\int_0^{2\pi}(\partial_r\varphi_n(r_1\cos\theta, r_1\sin\theta)r_1+1)\,d\theta\right|\\ =\int_{r_1}^{r_2}\int_0^{2\pi}\partial_r(\partial_r\varphi_n r)\,d\theta\,dr=\int_{r_1}^{r_2}\int_0^{2\pi}\triangle\varphi_n\,d\theta\,dr\\ \leqslant\int_{r_1}^{r_2}\int_0^{2\pi}\frac{(\triangle\varphi_n)^2}{e^{2\varphi_n}}\,d\theta\,dr\int_{r_1}^{r_2}\int_0^{2\pi}e^{2\varphi_n}\,d\theta\,dr\\ =K_n(r_1,r_2)A_n(r_1,r_2),\end{aligned}$$

which tends to 0 as $n\to\infty$. Hence the limit exists and α is defined. □

4. Proof of the main theorem

In the neck of the blowing-up process (page 41), the sequence of metrics has an average property around concentric circles:

Lemma 4.1 (Max-min inequality). *Let $\{g_n = e^{2\varphi_n}(dx^2+dy^2)\}$ be a sequence of metrics with finite area and finite energy, converging weakly to a limit metric $g = e^{2\varphi}(dx^2+dy^2)$ in $H^{2,2}_{\mathrm{loc}}(D\setminus\{0\})$ and having $\{0\}$ as its only bubble point. In* $\mathrm{Neck}(r_1, r_2)$, *there exists a constant $c\in(0,1)$, independent of r, such that*

$$c\leqslant\left|\frac{\max_\theta(\varphi_n(r\cos\theta, r\sin\theta)+\log r)}{\min_\theta(\varphi_n(r\cos\theta, r\sin\theta)+\log r)}\right|\leqslant 1.$$

Proof. Renormalize the sequence of metrics around $p=\{0\}$ as

$$\phi_n(z)=\varphi(\delta_n z)+\log\delta_n,$$
$$\tilde g_n(z)=e^{2\phi_n(z)}|dz|^2=g_n(\delta_n z).$$

Lemma 3.2 yields, for the limit metrics φ and ϕ,

$$\lambda^{-1}(\bar\varphi(r)+\log r)+C_3\le\varphi(r\cos\theta, r\sin\theta)+\log r\le\lambda(\bar\varphi(r)+\log r)+C_4,$$
$$\mu^{-1}(\bar\phi(r)+\log r)+C_5\le\phi(r\cos\theta, r\sin\theta)+\log r\le\mu(\bar\phi(r)+\log r)+C_6,$$

with $\lambda,\mu\in(0,1)$. For r_1 small enough and n, r_2 big enough, we then have

$$\begin{aligned}\lambda^{-1}(\bar\varphi_n(r_1)+\log r_1)+C_3&\le\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1\\&\le\lambda(\bar\varphi_n(r_1)+\log r_1)+C_4,\\ \mu^{-1}(\bar\phi_n(r_2)+\log r_2)+C_5&\le\phi_n(r_2\cos\theta, r_2\sin\theta)+\log r_2\\&\le\mu(\bar\phi_n(r_2)+\log r_2)+C_6,\end{aligned}\tag{4-1}$$

where C_3, C_4, C_5, C_6 are independent of r. This last pair of inequalities leads to

$$\mu^{-1}(\bar{\varphi}_n(r_2\delta_n)+\log(r_2\delta_n)+C_5 \le \varphi_n(r_2\delta_n\cos\theta, r_2\delta_n\sin\theta)+\log(r_2\delta_n)$$
$$\le \mu(\bar{\varphi}_n(r_2\delta_n)+\log(r_2\delta_n))+C_6.$$

Combining this with (4-1), one obtains, still for r_1 small enough and n, r_2 big enough,

$$\frac{\lambda^2}{2} \leqslant \left|\frac{\max_\theta(\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1)}{\min_\theta(\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1)}\right| \leqslant 1$$

and

$$\frac{\mu^2}{2} \leqslant \left|\frac{\max_\theta(\varphi_n(r_2\delta_n\cos\theta, r_2\delta_n\sin\theta)+\log(r_2\delta_n))}{\min_\theta(\varphi_n(r_2\delta_n\cos\theta, r_2\delta_n\sin\theta)+\log(r_2\delta_n))}\right| \leqslant 1.$$

Thus the max-min inequality holds at the two boundary circles.

Set $u = -\log r = -\log\sqrt{x^2+y^2}$ and $\theta = \tan^{-1}(y/x)$. The domain $D\setminus\{0\}$ becomes an infinite annulus $\{(u,\theta): 0 < u \leqslant \infty,\ -\pi \leqslant \theta \leqslant \pi\}$ via this transformation. Let

$$\xi(u,\theta) = \varphi(e^{-u}\cos\theta, e^{-u}\sin\theta) - u\ ,\ \xi_n(u,\theta) = \varphi_n(e^{-u}\cos\theta, e^{-u}\sin\theta) - u$$

and define $\triangle_{u,\theta} = \partial_u^2 + \partial_\theta^2$. For any small $r_1 = e^{-u_0} > 0$, define $\tilde{\varphi}(v,\theta) = \xi(v+u_0,\theta)$ and $\tilde{\varphi}_n(v,\theta) = \xi_n(v+u_0,\theta)$. Then

$$-\triangle_{v,\theta}\tilde{\varphi} = K(v+u_0,\theta)e^{2\tilde{\varphi}} \qquad \text{for all } (v,\theta)\in\tilde{D},$$
$$-\triangle_{v,\theta}\tilde{\varphi}_n = K_n(v+u_0,\theta)e^{2\tilde{\varphi}_n} \quad \text{for all } (v,\theta)\in\tilde{D},$$

where $\tilde{D} = [-1,1]\times S^1$. There exists a constant C such that $\tilde{\varphi}, \tilde{\varphi}_n \leqslant C$ for n big enough; thus the right-hand sides of both equalities are uniformly bounded in $L^2(\tilde{D})$.

Define ω, ω_n by

$$\begin{cases} -\triangle\omega = K(v+u_0,\theta)e^{2\tilde{\varphi}}, \\ \omega|_{\partial\tilde{D}} = 0, \end{cases} \qquad \begin{cases} -\triangle\omega_n = K_n(v+u_0,\theta)e^{2\tilde{\varphi}_n}, \\ \omega_n|_{\partial\tilde{D}} = 0. \end{cases}$$

Then $\|\omega\|_{L^\infty}, \|\omega_n\|_{L^\infty}$ are uniformly bounded from above; the bound is actually independent of u_0, since L^2 norm of $\tilde{\varphi}(v,\theta)$ and $\tilde{\varphi}_n(v,\theta)$ in $\tilde{D}$ converge uniformly to 0 as $n\to\infty$ and $u\to\infty$.

The function $h_n = \tilde{\varphi}_n - \omega_n$ is harmonic. Consider the two concentric circles $v_1 = 0$ ($|z| = r_1$), $v_2 = \log r_1 - \log r_2\delta_n$ ($|z| = r_2\delta_n$). For any circle $v_3 = \log r_1 - \log r_3$ ($|z| = r_3$) between the two, set

$$M_n(v) = \max_v h_n(v,\theta), \quad m_n(v) = \min_v h_n(v,\theta).$$

Apply Hadamard's Three-Circle Theorem to obtain

$$M_n(v_3) \leqslant \frac{M_n(v_1)\big(\log(v_2+u_0)-\log(v_3+u_0)\big) + M_n(v_2)\big(\log(v_3+u_0)-\log(v_1+u_0)\big)}{\log(v_2+u_0)-\log(v_3+u_0)}$$

and

$$m_n(v_3) \geqslant \frac{m_n(v_1)\big(\log(v_2+u_0)-\log(v_3+u_0)\big) + m_n(v_2)\big(\log(v_3+u_0)-\log(v_1+u_0)\big)}{\log(v_2+u_0)-\log(v_3+u_0)}.$$

Since the max-min inequality holds at the two boundary circles, we obtain

$$M_n(v_1) = \max\nolimits_\theta h_n(0,\theta) = \max\nolimits_\theta(\tilde{\varphi}_n(0,\theta)-\omega_n(0,\theta)) \leqslant \max\nolimits_\theta \tilde{\varphi}_n(0,\theta)+C,$$
$$m_n(v_1) = \min\nolimits_\theta h_n(0,\theta) = \min\nolimits_\theta(\tilde{\varphi}_n(0,\theta)-\omega_n(0,\theta)) \geqslant \min\nolimits_\theta \tilde{\varphi}_n(0,\theta)-C,$$

where C is the uniform bound of $\|\omega_n\|_{L^\infty}$. By the definition of $\tilde{\varphi}_n(0,\theta)$, we have

$$\begin{aligned}\max\nolimits_\theta \tilde{\varphi}_n(0,\theta) &= \max\nolimits_\theta(\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1)\\ &\leqslant \tfrac{1}{2}\lambda^2 \min\nolimits_\theta(\varphi_n(r_1\cos\theta, r_1\sin\theta)+\log r_1) = \tfrac{1}{2}\lambda^2 \min\nolimits_\theta \tilde{\varphi}_n(0,\theta).\end{aligned}$$

This implies

$$M_n(v_1) \leqslant \tfrac{1}{2}\lambda^2 m_n(v_1)+2C.$$

Similarly, we obtain

$$M_n(v_2) \leqslant \tfrac{1}{2}\mu^2 m_n(v_2)+2C.$$

For any $v_3 \in (v_1, v_2)$, the inequality at the top of this page implies

$$M_n(v_3) \leqslant \tfrac{1}{2}\min(\lambda^2,\mu^2) m_n(v_3)+4C;$$

thus

$$\max_\theta \tilde{\varphi}_n(v_3,\theta) \leqslant \tfrac{1}{2}\min(\lambda^2,\mu^2)\min\nolimits_\theta \tilde{\varphi}_n(v_3,\theta)+6C,$$

and so

$$\max_\theta(\varphi_n(r_3,\theta)+\log r_3) \leqslant \tfrac{1}{2}\min(\lambda^2,\mu^2)\min\nolimits_\theta(\varphi_n(r_3,\theta)+\log r_3)+6C.$$

Because $\varphi+\log r \to -\infty$ when $r\to 0$, and C is independent of r, we see that if r_1 small enough and n big enough, then

$$\min\nolimits_\theta(\varphi_n(r_3,\theta)+\log r_3) \leqslant \max_\theta(\varphi_n(r_3,\theta)+\log r_3) \leqslant c\min\nolimits_\theta(\varphi_n(r_3,\theta)+\log r_3)$$

for any $r_3 \in (r_2\delta_n, r_1)$, where $c\in(0,1)$ is a constant. This immediately implies that the max-min inequality holds in $\mathrm{Neck}(r_1, r_2)$. □

Proof of Main Theorem 2.6. Case 1: The renormalized limit metric ϕ does not vanish. According to the definition of the weak angle and the out angle, for r_1 small enough and r_2, n large enough we have

$$\left| \int_{|z|=r_1} k_{g_n} ds_n - \alpha \right| < \varepsilon, \quad \left| \int_{|z|=r_2\delta_n} k_{g_n} ds_n - \beta \right| < \varepsilon.$$

By Lemmas 3.7 and 3.9, we have $\alpha \geqslant 0$ and $\beta \leqslant 0$. If $\alpha > 0$, there exists a circle $|z| = r_{3,n}$ such that

$$\int_{|z|=r_{3,n}} k_{g_n} ds_n = \frac{\alpha}{2},$$

and we can assume that $r_{3,n}$ is maximal satisfying this condition. Define

$$\bar{\varphi}(r) = \frac{1}{2\pi} \int_0^{2\pi} \varphi(r\cos\theta, r\sin\theta)\, d\theta,$$
$$\bar{\varphi}_n(r) = \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(r\cos\theta, r\sin\theta)\, d\theta.$$

Then

$$\begin{aligned} \text{(4-2)} \qquad \iint_{r_{3,n} \leqslant |z| \leqslant r_1} \frac{\alpha}{2} e^{2\bar{\varphi}_n} r\, dr\, d\theta &\leqslant \frac{1}{2} \int_0^{2\pi} \int_{r_{3,n}}^{r_1} e^{2\bar{\varphi}_n} r \int_{|z|=r} k_{g_n} ds_n\, dr\, d\theta \\ &= \pi \int_{r3,n}^{r_1} e^{2\bar{\varphi}_n} r(\partial_r \bar{\varphi}_n r + 1) dr \\ &= \pi \int_{r3,n}^{r_1} e^{2(\bar{\varphi}_n + \log r)} \Big(\partial_r \bar{\varphi}_n + \frac{1}{r} \Big) dr \\ &= \frac{\pi}{2} e^{2(\bar{\varphi}_n + \log r)} \Big|_{r_{3,n}}^{r_1}. \end{aligned}$$

Using the max-min inequality (Lemma 4.1), one easily gets

$$\text{(4-3)} \qquad e^{2(\bar{\varphi}_n + \log r)}|_{r_1} \leqslant \left(\frac{1}{2\pi} \int_0^{2\pi} e^{(\varphi_n + \log r)|_{r_1}}\, d\theta \right)^{1/c},$$

where $c \in (0, 1)$ is the constant in the max-min inequality. Since

$$\int_0^{2\pi} e^{(\varphi_n + \log r)|_{r_1}}\, d\theta < \epsilon$$

(from the blowing up process), inequalities (4-2) and (4-3) then imply

$$\iint_{r_{3,n} \leqslant |z| \leqslant r_1} \frac{\alpha}{2} e^{2\bar{\varphi}_n} r\, dr\, d\theta \to 0.$$

If $\alpha > 0$, then

$$\iint_{r_{3,n}\leqslant|z|\leqslant r_1} e^{2\bar{\varphi}_n} r\,dr\,d\theta \to 0$$

as $r_1 \to 0$, $r_2 \to \infty$ and $n \to \infty$. Using the max-min inequality again, we obtain

$$\iint_{r_{3,n}\leqslant|z|\leqslant r_1} e^{2\varphi_n} r\,dr\,d\theta \to 0.$$

On the other hand, for the annulus bounded by two concentric circle $|z| = r_1$ and $|z| = r_{3,n}$, we apply the Gauss–Bonnet Theorem to get

$$\int_{|z|=r_1} k_{g_n} ds_n - \int_{|z|=r_{3,n}} k_{g_n} ds_n + \int_{r_{3,n}\leqslant|z|\leqslant r_1} K_n\,dg_n = 0, \tag{4-4}$$

while

$$\left(\int_{r_{3,n}\leqslant|z|\leqslant r_1} K_n\,dg_n\right)^2 \leqslant \int_{r_{3,n}\leqslant|z|\leqslant r_1} K_n^2 dg_n \int_{r_{3,n}\leqslant|z|\leqslant r_1} dg_n$$

and

$$\int_{r_{3,n}\leqslant|z|\leqslant r_1} dg_n = \iint_{r_{3,n}\leqslant|z|\leqslant r_1} e^{2\varphi_n} r\,dr\,d\theta \to 0.$$

This means that $\int_{r_{3,n}\leqslant|z|\leqslant r_1} K_n\,dg_n \to 0$. Taking the limit in (4-4), we get $\alpha - \frac{1}{2}\alpha = 0$, a contradiction. Thus we have proved that $\alpha = 0$ in this case.

Case 2: The renormalized limit metric ϕ_0 vanishes. If the out angle β is nonpositive, we apply the Gauss–Bonnet theorem directly:

$$\int_{|z|=r_1} k_{g_n} ds_n - \int_{|z|=r_2\delta_n} k_{g_n} ds_n + \int_{r_2\delta_n\leqslant|z|\leqslant r_1} K_n\,dg_n = 0.$$

Since the limit metric vanishes,

$$\int_{r_2\delta_n\leqslant|z|\leqslant r_1} K_n\,dg_n \to 0.$$

Taking the limit in the last equality, we have $\alpha - \beta = 0$, so $\alpha = \beta = 0$.

If $\beta > 0$, we consider the bubble tree decomposition. The renormalized limit metric ϕ has two bubble points at least, by Theorem 1.2. Assume the bubble points are $\{q_1, q_2, \ldots, q_m\}$, with $m \geqslant 2$, and the weak angle at q_i is α_i. Let $h_n = \tilde{g}_n$ be the renormalized metrics, and take disks $D_{r^{(i)}}(q_i)$ around q_i with radius $r^{(i)}$ such that

$$\left|\alpha_i - \int_{\partial D_{r^{(i)}}(q_i)} k_{h_n} ds_n\right| < \varepsilon$$

for n big enough, and a fixed small ε. Set

$$\Omega = D_{r_2}(p) \setminus \bigcup_{1\leqslant i\leqslant m} D_{r^{(i)}}(q_i)$$

(before renormalization, the disk is $D_{r_2\delta_n}(p)$). Now apply to the Gauss–Bonnet Theorem to Ω:

$$\int_{|z|=r_2} k_{h_n} ds_n - \sum_{i=1}^{m} \int_{|z|=r^{(i)}} k_{h_n} ds_n + \int_\Omega K_n dh_n = -(m-1). \tag{4-5}$$

Because $\{h_n\} \to 0$, $A_n(\Omega) = \int_\Omega dh_n \to 0$ when $n \to \infty$, we have

$$\left(\int_\Omega K_n dh_n\right)^2 \leqslant \int_\Omega K_n^2 dh_n \int_\Omega dh_n \to 0.$$

Taking limits in (4-5) we get $\sum_{i=1}^{m} \alpha_i = m - 1 + \beta$. Thus there must exist some i with $1 \leqslant i \leqslant m$ such that $\alpha_i > 0$. We can assume $\alpha_1 > 0$, and consider a pair of angles $\{\alpha_1, \beta_1\}$ (where β_1 is the out angle at the point q_1) and the blowing-up process around $\{q_1\}$. If $\beta_1 \leqslant 0$, then by the argument above we have $\alpha_1 = 0$, a contradiction. If $\beta_1 > 0$, we argue as in the case $\beta > 0$: we get a new pair of angles $\{\alpha_{11}, \beta_{11}\}$ at the bubble point q_{11} in the blowing-up process around q_1, such that $\alpha_{11} > 0$. If $\beta_{11} \leqslant 0$, there is a contradiction. If $\beta_{11} > 0$, then we apply the blowing-up once more. Since $\beta > 0$, the limit metric of renormalization is zero and the limit metric is a ghost vertex in the bubble tree. Thus, if $\alpha > 0$, we get an infinite series of ghost vertices in the bubble tree. But we know that there are only finitely ghost vertices in the bubble tree, so we reach a contradiction. This proves that α is zero in this case too, as desired. □

References

[Calabi 1982] E. Calabi, "Extremal Kähler metrics", pp. 259–290 in *Seminar on Differential Geometry*, Ann. of Math. Stud. **102**, Princeton Univ. Press, Princeton, NJ, 1982. MR 83i:53088 Zbl 0487.53057

[Chang and Yang 1988] S.-Y. A. Chang and P. C. Yang, "Conformal deformation of metrics on S^2", *J. Differential Geom.* **27**:2 (1988), 259–296. MR 89b:53079 Zbl 0649.53022

[Chen 1998a] X. Chen, "Extremal Hermitian metrics on Riemannian surfaces", *Internat. Math. Res. Notices* 15 (1998), 781–797. MR 99i:58035 Zbl 0955.30032

[Chen 1998b] X. Chen, "Weak limits of Riemannian metrics in surfaces with integral curvature bound", *Calc. Var. Partial Differential Eq.* **6**:3 (1998), 189–226. MR 99k:53087 Zbl 0894.53040

[Chen 1999] X. Chen, "Extremal Hermitian metrics on Riemann surfaces", *Calc. Var. Partial Differential Eq.* **8**:3 (1999), 191–232. MR 2000f:58029 Zbl 0947.53038

[Chen 2001] X. X. Chen, "Calabi flow in Riemann surfaces revisited: a new point of view", *Internat. Math. Res. Notices* **2001**:6 (2001), 275–297. MR 2002c:53111 Zbl 1078.53065

[Chen and Li 1991] W. X. Chen and C. Li, "Classification of solutions of some nonlinear elliptic equations", *Duke Math. J.* **63**:3 (1991), 615–622. MR 93e:35009 Zbl 0768.35025

[Luo and Tian 1992] F. Luo and G. Tian, "Liouville equation and spherical convex polytopes", *Proc. Amer. Math. Soc.* **116**:4 (1992), 1119–1129. MR 93b:53034 Zbl 0806.53012

[McOwen 1988] R. C. McOwen, “Point singularities and conformal metrics on Riemann surfaces”, *Proc. Amer. Math. Soc.* **103**:1 (1988), 222–224. MR 89m:30089 Zbl 0657.30033

[Troyanov 1991] M. Troyanov, “Prescribing curvature on compact surfaces with conical singularities”, *Trans. Amer. Math. Soc.* **324**:2 (1991), 793–821. MR 91h:53059 Zbl 0724.53023

[Uhlenbeck 1982] K. K. Uhlenbeck, “Connections with L^p bounds on curvature”, *Comm. Math. Phys.* **83**:1 (1982), 31–42. MR 83e:53035 Zbl 0499.58019

Received December 20, 2005.

QING CHEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI, ANHUI, 230026
CHINA

qchen@ustc.edu.cn

XIUXIONG CHEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
480 LINCOLN DRIVE
MADISON WI 53706-1388
UNITED STATES

xxchen@math.wisc.edu

WEIYONG HE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI, ANHUI, 230026
CHINA

wyhe@mail.ustc.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

ON AHLFORS' SCHWARZIAN DERIVATIVE AND KNOTS

MARTIN CHUAQUI

We extend Ahlfors' definition of the Schwarzian derivative for curves in euclidean space to include curves on arbitrary manifolds, and give applications to the classical spaces of constant curvature. We also derive in terms of the Schwarzian a sharp criterion for a closed curve in $\mathbb{R}^3$ to be unknotted.

1. Introduction

This paper is a continuation of [Chuaqui and Gevirtz 2004], in which we developed sharp bounds on the real part of Ahlfors' [1988] Schwarzian derivative for curves C in $\mathbb{R}^n$ which imply that C is simple. We begin with a geometrically simpler definition of the Schwarzian for such curves, the real part $S_1 f$ of which coincides with that of Ahlfors. This approach has the advantage of suggesting a Schwarzian for curves in arbitrary manifolds, the results we obtain strongly suggesting that its real part, at least, is appropriately defined. After our discussion of the Schwarzian for curves in the general manifold context we focus on the particular cases of hyperbolic n-space $\mathbb{H}^n$ and the n-sphere $\mathbb{S}^n$ and derive the relationship between $S_1 f$ as calculated with respect to the metrics on $\mathbb{H}^n$ and $\mathbb{S}^n$ on the one hand, and with respect to the euclidean metric on the underlying ball and $\mathbb{R}^n \cup \{\infty\}$, on the other. Using these calculations together with results of [Chuaqui and Gevirtz 2004] we obtain a very short proof of a theorem of C. Epstein [1985] to the effect a curve in $\mathbb{H}^n$ is necessarily simple if the absolute value of its geodesic curvature is everywhere bounded by 1; we also prove the theorem's spherical counterpart. Lastly, we derive a sharp bound on $S_1 f$ which implies that the corresponding curve is unknotted.

2. Preliminaries

Let $f : (a, b) \to \mathbb{R}^n$ be a C^3 curve with $f' \neq 0$, and let $X \cdot Y$ stand for the euclidean inner product of vectors X, Y in $\mathbb{R}^n$. Set $|X|^2 = X \cdot X$. As pointed out in [Chuaqui

---

MSC2000: primary 53A04, 53A30; secondary 53A55.
Keywords: Schwarzian derivative, simple curve, conformal metric, knot.
The author was partially supported by Fondecyt Grant # 1030589 and by MECESUP PROJECT PUC 0103.

and Gevirtz 2004], the real part of Ahlfors' Schwarzian, defined by

$$S_1 f = \frac{f' \cdot f'''}{|f'|^2} - 3\frac{(f' \cdot f'')^2}{|f'|^4} + \frac{3}{2}\frac{|f''|^2}{|f'|^2},$$

can be written in terms of the velocity $v = |f'|$ and the curvature k of the trace of f as

$$S_1 f = \left(\frac{v'}{v}\right)' - \frac{1}{2}\left(\frac{v'}{v}\right)^2 + \frac{1}{2}v^2k^2, \tag{1}$$

and this expression is invariant under Möbius transformations of $\mathbb{R}^n \cup \{\infty\}$. Our main result in that paper was:

Theorem 1. *Let $p = p(x)$ be a continuous real-valued function on an open interval I such that any nontrivial solution of $u'' + pu = 0$ has at most one zero on I. Let $f : I \to \mathbb{R}^n$ be a C^3 curve with $f' \neq 0$. If $S_1 f \leq 2p$, then f is one-to-one on I and admits a spherically continuous extension to the closed interval, which is also one-to-one unless the trace of f is a circle, in which case $S_1 f \equiv 2p$.*

Although the formal expression on the right side of (1) is meaningful in the context of manifolds, its appropriateness is made apparent by the following considerations. Let T denote the tangent vector along the trace of f, and let ∇ stand for usual covariant differential operator on M. Then $\nabla_T T$ corresponds to f''. We regard the 2-dimensional subspace spanned by T and $\nabla_T T$ as the complex plane $\mathbb{C}$ (the orientation being irrelevant), so that $T = a = a(t)$ and $\nabla_T T = b(t)$ are complex-valued functions of the parametrizing variable $t \in I$. Following the classical definition of the Schwarzian, given by

$$\left(\frac{f''}{f'}\right)' - \frac{1}{2}\left(\frac{f''}{f'}\right)^2,$$

we are led to consider the complex function

$$\left(\frac{b}{a}\right)' - \frac{1}{2}\left(\frac{b}{a}\right)^2 \tag{2}$$

as the manifold analogue of the Schwarzian. A straightforward calculation shows that the real part of the expression in (2) coincides with (1).

Let $\mathbb{H}^n$ denote the hyperbolic n-space with constant sectional curvature -1, for which we use the standard model $\mathbb{B}^n = \{x \in \mathbb{R}^n : |x| < 1\}$ with metric tensor $g_h = 4(1-|x|^2)^{-2}g$, where g is the euclidean metric. Let $\mathbb{S}^n$ stand for the n-dimensional sphere, as modeled by $\mathbb{R}^n \cup \{\infty\}$ with the metric $g_e = 4(1 + |x|^2)^{-2}g$; here the sectional curvature is 1. Both are special cases of a domain $\Omega \subset \mathbb{R}^n$ endowed with a conformal metric tensor, that is, a metric tensor of the form $\bar{g} = e^{2\varphi(x)}g$. In this generality one can relate the Schwarzian corresponding to the resulting manifold M with the standard euclidean Schwarzian defined on Ω itself. To do so one needs to determine how the velocity and curvature of a curve change under conformal changes of metric. Any object (velocity, curvature, covariant derivative, etc.) associated with the manifold M will be distinguished from the corresponding object in the underlying Ω by a bar. Thus, let v, k denote the velocity and curvature on Ω so that $\bar{v}$, $\bar{k}$ are their counterparts on M. Obviously, $\bar{v} = e^{\varphi} v$, from which routine calculations yield

$$\begin{aligned}(3)\quad &\left(\frac{\bar{v}'}{\bar{v}}\right)' - \frac{1}{2}\left(\frac{\bar{v}'}{\bar{v}}\right)^2 \\ &= \left(\frac{v'}{v}\right)' - \frac{1}{2}\left(\frac{v'}{v}\right)^2 + v^2 \operatorname{Hess}(\varphi)(t,t) + v^2 k(\operatorname{grad}\varphi \cdot n) - \tfrac{1}{2} v^2 (\operatorname{grad}\varphi \cdot t)^2,\end{aligned}$$

where t and n are the euclidean unitary tangent and normal vectors to the curve, $\operatorname{Hess}(\varphi)$ is the (euclidean) Hessian bilinear form and grad is the standard gradient.

In order to derive the relationship between k and $\bar{k}$ one needs to know how the covariant derivative changes under conformal changes of metric. The classical formula is

$$(4)\qquad \overline{\nabla}_X Y = \nabla_X Y + (\operatorname{grad}\varphi \cdot X)Y + (\operatorname{grad}\varphi \cdot Y)X - (X \cdot Y)\operatorname{grad}\varphi.$$

The curvature $\bar{k}$ is determined by the equation

$$\overline{\nabla}_{\bar{t}}\bar{t} = \bar{k}\bar{n},$$

where $\bar{t} = e^{-\varphi} t$ and $\bar{n} = e^{-\varphi} n$. Using (4) one obtains

$$\overline{\nabla}_{\bar{t}}\bar{t} = e^{-2\varphi}\left(kn + (\operatorname{grad}\varphi \cdot t)t - \operatorname{grad}\varphi\right).$$

After taking the euclidean norm of both sides we get

$$\bar{k}^2 = e^{-2\varphi}\left(k^2 - (\operatorname{grad}\varphi \cdot t)^2 - 2k(\operatorname{grad}\varphi \cdot n) + |\operatorname{grad}\varphi|^2\right),$$

and using (3) we have

$$(5)\qquad \bar{S}_1 f = S_1 f + v^2 \operatorname{Hess}(\varphi)(t,t) - v^2(\operatorname{grad}\varphi \cdot t)^2 + \frac{v^2}{2}|\operatorname{grad}\varphi|^2.$$

The terms on the right-hand side depending on φ are best expressed in terms of the Schwarzian tensor $B(\varphi)$ of the metric $\bar{g}$ with respect to g, as defined in [Osgood and Stowe 1992] by

$$B(\varphi) = \operatorname{Hess}(\varphi) - d\varphi \otimes d\varphi - \frac{1}{n}(\Delta\varphi - |\operatorname{grad}\varphi|^2)g.$$

Then (5) can be rewritten as

$$
\begin{aligned}
\bar{S}_1 f &= S_1 f + v^2 B(\varphi)(t,t) + \frac{v^2}{n}\Delta\varphi + \frac{n-2}{2n}\, v^2\,|\mathrm{grad}\,\varphi|^2 \\
&= S_1 f + v^2 B(\varphi)(t,t) - \frac{v^2}{2}\frac{\mathrm{scal}\,\bar{g}}{n(n-1)}\, e^{2\varphi},
\end{aligned} \tag{6}
$$

where $\mathrm{scal}\,\bar{g}$ is the scalar curvature of the metric $\bar{g}$, that is, the sum of the sectional curvatures of any complete set orthogonal 2-planes of the tangent space at a given point. The Schwarzian tensor appears in the work of Osgood and Stowe as a suitable generalization of the classical Schwarzian derivative when studying conformal local diffeomorphisms between Riemannian manifolds, or more generally, when studying metrics on a given manifold that are conformally related. They show that conformal changes of metric with vanishing Schwarzian tensor, called Möbius changes of metric, are rare on arbitrary manifolds. On euclidean space, nevertheless, Möbius changes can be described completely and include, in particular, the hyperbolic and the spherical metric. In other words, $B(\varphi) = 0$ when e^{φ} is either $2(1-|x|^2)^{-1}$ or $2(1+|x|^2)^{-1}$.

Since $\mathrm{scal}\,\bar{g} = -n(n-1)$ when $\bar{g} = g_h$ we obtain from (6)

$$
S_1^h f = S_1 f + \frac{v^2}{2}\, e^{2\varphi}. \tag{7}
$$

For the spherical metric we have $\mathrm{scal}(\bar{g}) = n(n-1)$, hence (6) gives

$$
S_1^s f = S_1 f - \frac{v^2}{2}\, e^{2\varphi}. \tag{8}
$$

We use (7) to give a very short proof of this result:

Theorem 2 [Epstein 1985]. *Let $\gamma \subset \mathbb{H}^n$ be a curve with geodesic curvature bounded in absolute value by* 1. *Then γ is simple.*

Proof. Let $f : (-l, l) \to \gamma$ be a hyperbolic arclength parametrization. Note that the value $l = \infty$ is possible. Then $v_h \equiv 1$, so that $S_1^h f = k_h^2/2$. But since $v = e^{-\varphi} = (1-|x|^2)/2$ it follows from (7) that

$$
S_1 f = \frac{k_h^2 - 1}{2} \le 0.
$$

By appealing to Theorem 1 with the choice $p(x) \equiv 0$, we conclude that γ is simple. □

In the same vein, we can use (8) to derive the corresponding simplicity criterion for curves on $\mathbb{S}^n$.

Theorem 3. *Let $\gamma \subset \mathbb{S}^n$ be a curve of length $l \le 2\pi$ and geodesic curvature k_s satisfying*

$$k_s^2 \le \frac{4\pi^2 - l^2}{l^2}.$$

Then γ is simple except when it is a circle of constant curvature $\sqrt{4\pi^2 - l^2}/l$.

Proof. We proceed as before and consider $f : [0, l] \to \gamma$, a spherical arclength parametrization. Then $S_1^s f = k_s^2/2$ and $v_s = (1 + |x|^2)/2$, so (8) gives

$$S_1 f = \frac{1 + k_s^2}{2} \le \frac{2\pi^2}{l^2}.$$

This time we apply Theorem 1 with $p(x) \equiv \pi^2/l^2$ to conclude that $f((0, l))$ is simple. The extended curve $f([0, l])$ remains simple unless it is a circle, of constant curvature $\sqrt{4\pi^2 - l^2}/l$. □

3. Knots

Theorem 4. *Let $f : [-1, 1) \to \mathbb{R}^3$ parametrize a simple closed curve in $\mathbb{R}^3$. If the periodic continuation of f is C^3 and $S_1 f(t) \le 2\pi^2$ for all $t \in (-1, 1)$, then $f([-1, 1))$ is unknotted.*

Proof. The idea is to show that, if knotted, the curve $\Gamma = f([-1, 1])$ can be laid out to form a planar, closed, nonsimple curve for which the real part of the Schwarzian has not increased. The process used to do this is based on the following ideas, developed by Brickell and Hsiung [1974] in the course of their proof of the Fary–Milnor theorem.

For $p \in \mathbb{R}^3$ we define the *shell* C_p of Γ with vertex p to be the developable surface made up of all segments $[p, q]$ with $q \ne p$ on Γ. The *indicatrix* of C_p, denoted by I_p, is the curve on $\mathbb{S}^2 = \{u \in \mathbb{R}^3 : |u| = 1\}$ traced by the vectors $(q - p)/|q - p|$; its length $l(I_p)$ is called the total angle of I_p. A key fact established by Brickell and Hsiung is that Γ is unknotted if $l(I_p) < 3\pi$ for all $p \in \Gamma$. The proof of this uses Crofton's formula

$$\int n(G)dG = 4l(I_p)$$

giving the length of I_p in terms of the number $n(G)$ of intersection points of I_p with great circles $G \subset \mathbb{S}^2$. The integral is performed over $\mathbb{S}^2$, after identifying a point on the sphere with the normal direction of a plane containing a great circle. The authors show that $n(G) \ge 1$ for all G and that $\{G : n(G) = 2\}$ has measure zero [Brickell and Hsiung 1974, Lemma 8, p. 188]. Since the measure of the entire set of great circles is 4π, if $l(I_p) < 3\pi$ then $\{G : n(G) = 1\}$ must have positive measure. Hence there exists at least one great circle G with $n(G) = 1$, which means there exists one plane through the point p intersecting Γ at exactly one other point $q \neq p$. Such a plane is called *transversal* to Γ. The curve Γ has the *transversal property* if for any $p \in \Gamma$ there exists a plane through p transversal to Γ.

Theorem [Brickell and Hsiung 1974, Lemma 6, p. 191]. *Let C be a closed smooth curve embedded in hyperbolic or euclidean space of dimension three. If C has the transversal property then C is a trivial knot.*

We conclude from this discussion that if Γ is a knot there is a point $p \in \Gamma$ for which $l(I_p) \geq 3\pi$. The two cases $l(I_p) > 3\pi$ and $l(I_p) = 3\pi$ require a slightly different analysis. Suppose first that $l(I_p) > 3\pi$. As we move p to a point p' slightly away from Γ, the number $l(I_{p'})$ varies continuously, except for a jump increment by π. Hence there exists $p' \notin \Gamma$ for which $l(I_{p'}) > 4\pi$. On the other hand, since $l(I_r)$ is a continuous function of $r \in \mathbb{R}^3 \setminus \Gamma$ and since $l(I_r) \to 0$ as $|r| \to \infty$, we can find $p_0 \notin \Gamma$ such that $l(I_{p_0}) = 4\pi$. We now lay out the shell C_{p_0} isometrically onto the plane so that Γ traces out a closed curve γ that is not simple. To do this, let $\Gamma = \Gamma(s)$ be an arclength parametrization, $0 \leq s \leq L$, and set $r(s) = |\Gamma(s) - p_0|$. We lay out Γ onto the plane curve γ given by $z = z(s) = r(s)e^{i\theta(s)}$, where the function θ is chosen so that $|z'(s)| = 1$, or equivalently

$$|r'(s) + ir(s)\theta'(s)| = 1.$$

The function

$$\theta(s) = \int_0^s \frac{\sqrt{1-(r'(t))^2}}{r(t)}\,dt$$

has this property. The point p_0 corresponds to $z = 0 \notin \gamma$, and the polar angle $\theta = \theta(s)$ increases at the same rate as the spatial angle of the rays $[p_0, \Gamma(s)]$ at the vertex p_0. Because $l(I_{p_0}) = 4\pi$ it follows that γ is a closed curve with winding number 2 with respect to the origin.

If, on the other hand, $l(I_p) = 3\pi$, we let $p_0 = p$ and lay out Γ as before. We may assume that $p_0 = \Gamma(0)$. Since the point p_0 belongs to Γ, the curve γ obtained is closed because $r(s) \to 0$ as $s \to 0^+$ and as $s \to L^-$. Also, because Γ possesses a tangent at p_0, it is easy to see that the integrand in the equation for $\theta(s)$ above behaves like $h(s)/\sqrt{s(L-s)}$, where h is continuous on $[0, L]$. In other words, $\gamma(s) = z(s)$ is a planar curve passing through $z = 0$ with the property that $\theta(s) = \arg\{z(s)\}$ is increasing and has total variation of 3π. A variant of the argument principle allowing for zeros *on* the curve (see [Nehari 1952, p. 131], for instance) implies that γ cannot be simple: the point $0 \in \gamma$ contributes π to the total variation of argument and therefore γ must in addition wind around the origin once.

In either case, let $g : [-1, 1) \to \mathbb{R}^2$ be the induced parametrization of γ defined on the original interval of definition of f. We claim that $S_1 g \leq S_1 f$. First, $v_g =$

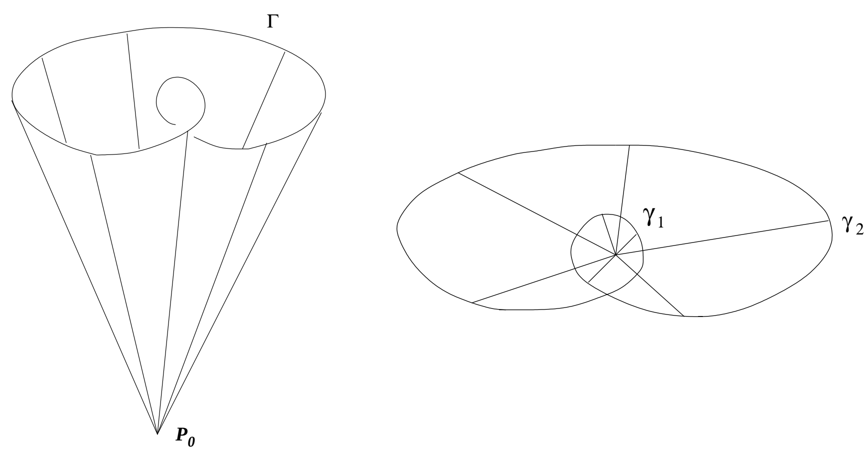
$|g'| = |f'| = v_f$ because the laying-out process preserves arclength. Secondly, the term involving the curvature does not increase because the curvature of γ is equal to the curvature of Γ relative to the surface C_{p_0}, that is, equal to the length of the projection of the curvature vector of Γ in $\mathbb{R}^3$ onto the tangent plane to the shell. We see from (1) that $S_1 g \le S_1 f$.

Since γ is not simple, it can be subdivided into closed curves γ_1, γ_2 that are differentiable except at the point where γ self-intersects. Because g is periodic, one can find intervals $[a, b]$ and $[c, d]$ of total length 2 such that

(i) $g_1 = g|_{[a,b]} : [a, b] \to \gamma_1$ and $g_2 = g|_{[c,d]} : [c, d] \to \gamma_2$, and

(ii) the parametrizations g_1, g_2 are C^3 on the open subintervals.

This sketch represents the case when $p_0 \notin \Gamma$ together with the corresponding nonsimple curve g:



We will show that both γ_1 and γ_2 are circles and that and each subinterval $[a, b]$, $[c, d]$ has length 1. In effect, it follows from Theorem 1 that the optimal C constant for a univalence criterion $S_1 h \le C$ on an open interval of length d is $C = 2\pi^2/d$, and that the extended curve can be closed only if it is a circle and $S_1 h \equiv 2\pi^2/d$. Because $S_1 g_1$, $S_1 g_2$ are bounded above by $2\pi^2$ on the open intervals and the curves γ_1 and γ_2 are closed, we conclude that the length of each subinterval $[a, b]$, $[c, d]$ cannot be less than 1. Because the total length is 2, each subinterval must have length 1, and since γ_1 and γ_2 are closed, they must be circles with $S_1 g_1 = S_1 g_2 \equiv 2\pi^2$. Hence $S_1 g \equiv 2\pi^2$, which can only happen if $S_1 f \equiv 2\pi^2$ and the curvature of γ remains the same as that of Γ. Hence Γ is an asymptotic curve, that is, the normal curvature vanishes at each point of Γ. Because the segments $[p_0, q]$ on the shell C_{p_0} are lines of curvature with corresponding principal curvature equal to zero, it follows that either Γ lies entirely on one such segment or else the shell is planar.

In the first case, Γ could not be closed, and in the second, it could not be knotted. This contradiction proves the theorem. $\square$

4. Example

In this final section we will show with that the assumption in Theorem 4 that the periodic continuation of f be smooth is essential. We will construct a closed curve $f : [-1, 1] \to \mathbb{R}^3$ with $S_1 f \leq 2\pi^2$ on $(-1, 1)$, whose image is a knot that is not of class C^3 at $f(1) = f(-1)$. The function f will be a Möbius transformation of the following curve g.

Let $g : (-1, 1) \to \mathbb{C}$. We write

$$S_1 g = \left(\frac{v'}{v}\right)' - \frac{1}{2}\left(\frac{v'}{v}\right)^2 + \frac{1}{2}k^2v^2 = 2q + \frac{1}{2}k^2v^2. \tag{9}$$

We will make $S_1 g \leq 2\pi^2$ everywhere on the open interval, but with different weights for the terms $2q = (v'/v)' - (1/2)(v'/v)^2$ and $k^2v^2/2$. Intuitively, the term q determines how fast one traverses the curve, while the second term determines the shape.

Let $\delta > 0$ be small. On $[-\frac{1}{2} + \delta, \frac{1}{2} - \delta]$ the curve g will have $q \equiv 0$, $v \equiv 1$ and $k \equiv 2\pi$. In other words, on this interval g describes almost a complete circle. We define g on $(\frac{1}{2} - \delta, 1) = (\frac{1}{2} - \delta, \frac{1}{2} + \delta] \cup (\frac{1}{2} + \delta, 1)$, and on $(-1, -\frac{1}{2} + \delta)$ in a symmetric way. On $(\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ we increase the value of q smoothly; this produces an increment in v, which forces us to decrease the value of k. We will do this in a way that

$$\int_{\frac{1}{2}-\delta}^{\frac{1}{2}+\delta} kv\,dx = \int k\,ds = 2\pi\delta. \tag{10}$$

Because of the symmetry on $[-\frac{1}{2} - \delta, -\frac{1}{2} + \delta)$ we will have

$$\int_{-\frac{1}{2}-\delta}^{\frac{1}{2}+\delta} kv\,dx = \left(\int_{-\frac{1}{2}-\delta}^{-\frac{1}{2}+\delta} + \int_{-\frac{1}{2}+\delta}^{\frac{1}{2}-\delta} + \int_{\frac{1}{2}-\delta}^{\frac{1}{2}+\delta}\right) kv\,dx = 2\pi.$$

On the remaining interval $(\frac{1}{2} + \delta, 1)$ we will decrease k sharply to 0, shifting all the weight to $q \equiv \pi^2$. Therefore, g will map this interval to a straight line. We will show that this can done in a way that the value of v'/v at $x = \frac{1}{2} + \delta$ is large enough to allow the parametrization of a straight line with $S_1 g = 2\pi^2$ on an interval of length $\frac{1}{2} - \delta$ to reach the point at infinity.

The details are as follows:

(I) The interval $\left(\frac{1}{2}-\delta,\, \frac{1}{2}+\delta\right]$: We see from (9) that $kv=\sqrt{4\pi^2-4q}=2\pi\sqrt{1-h}$, where $h=q/\pi^2$. From (10) we seek $h=h(x)\in[0,1]$ such that

$$\int_{\frac{1}{2}-\delta}^{\frac{1}{2}+\delta}\sqrt{1-h}\,dx=\delta.$$

If we shift the interval in question to $(0, 2\delta]$, we can choose h, for example, so that

$$\sqrt{1-h(x)}=1-\frac{x}{2\delta},$$

that is,

$$h(x)=\frac{x}{\delta}-\left(\frac{x}{2\delta}\right)^2.$$

(This choice requires only to be smoothed out at the endpoints of the interval.) With this,

$$\int_0^{2\delta}\sqrt{1-h}\,dx=2\delta-\frac{1}{2\delta}\frac{(2\delta)^2}{2}=\delta.$$

Observe that

$$\int_0^{2\delta}h\,dx=\int_0^{2\delta}\left(\frac{x}{\delta}-\left(\frac{x}{2\delta}\right)^2\right)dx=\frac{(2\delta)^2}{2\delta}-\frac{(2\delta)^3}{3(2\delta)^2}=\frac{4\delta}{3}, \tag{11}$$

a fact that will be important ahead.

(II) The term v'/v: Let $y=v'/v$. Then

$$y'=2q+\tfrac{1}{2}y^2=2\pi^2h+\tfrac{1}{2}y^2. \tag{12}$$

For convenience, once more we replace the interval $(\frac{1}{2}-\delta,\, \frac{1}{2}+\delta]$ by $(0, 2\delta]$. The initial condition for (12) is $y(0)=0$. We want to know whether $y(2\delta)$ (which corresponds to the original value of v'/v at $\frac{1}{2}+\delta$) is sufficiently large so that the parametrization of a straight line with velocity $v=e^{\int y\,dx}$ reaches the point at infinity before time $\frac{1}{2}-\delta$.

The parametrization of a straight line with Schwarzian identically equal to $2\pi^2$ reaches the point at infinity in time exactly $\frac{1}{2}$ if its initial velocity has $v'=0$. To verify this we consider the differential equation

$$w'=2\pi^2+\tfrac{1}{2}w^2,\quad w(0)=0,$$

which has the solution $w(x)=2\pi\tan(\pi x)$. The corresponding parametrization of the straight is then given by $\frac{1}{\pi}\tan(\pi x)$, which indeed becomes infinite at $x=\frac{1}{2}$. Now we need to verify that the solution y of (12) has

$$y(2\delta)>w(\delta). \tag{13}$$

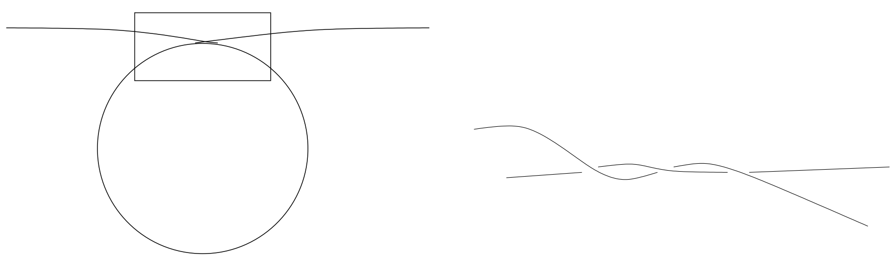
By integrating (12) we see from (11) that

$$y(2\delta) > 2\pi^2 \int_0^{2\delta} h\,dx = \frac{8\pi^2\delta}{3},$$

while

$$w(\delta) = 2\pi^2\delta + O(\delta^3),$$

so that (13) will hold if δ is small enough. Thus g reaches the point at infinity symmetrically at $1-\epsilon$ and $-1+\epsilon$, for some $\epsilon = O(\delta)$. In order to rectify the fact that g is defined only on $(-1+\epsilon, 1-\epsilon)$, we consider the scaled parametrization $g((1-\epsilon)x)$ defined on $(-1, 1)$, the Schwarzian of which is equal to $(1-\epsilon)^2 S_1 g < 2\pi^2$. We keep the notation g for the scaled curve; its trace together with the knot to be produced are shown in the following figure.



In the final step we produce a knot on g with a very small cost in $S_1 g$. The knot can be accomplished by replacing a small portion of one of the arcs at the point of self-intersection of g by a very thin tubular neighborhood, along which the new arc of g will go around once. Although this procedure introduces torsion, $S_1 g$ does not depend on it. It is easy to see that both the modified curvature and velocity remain arbitrarily close to their original values as long as the tubular neighborhood is thin enough. To finish the construction, we consider some Möbius transformation T for which $f = T(g)$ lies in the finite plane.

Acknowledgements

We thank Julian Gevirtz for many conversations regarding this paper and the proof of Theorem 4 in particular, as well as for bringing to our attention the work of Brickell and Hsiung.

References

[Ahlfors 1988] L. V. Ahlfors, "Cross-ratios and Schwarzian derivatives in $\mathbf{R}^n$", pp. 1–15 in *Complex analysis: Articles dedicated to Albert Pfluger*, edited by J. Hersch and A. Huber, Birkhäuser, Basel, 1988. MR 90a:30055 Zbl 0675.30021

[Brickell and Hsiung 1974] F. Brickell and C. C. Hsiung, "The total absolute curvature of closed curves in Riemannian manifolds", *J. Differential Geometry* **9** (1974), 177–193. MR 49 #3795 Zbl 0277.53004

[Chuaqui and Gevirtz 2004] M. Chuaqui and J. Gevirtz, "Simple curves in $\mathbb{R}^n$ and Ahlfors' Schwarzian derivative", *Proc. Amer. Math. Soc.* **132**:1 (2004), 223–230. MR 2005f:53001 Zbl 1045.53003

[Epstein 1985] C. L. Epstein, "Envelopes of horospheres and Weingarten surfaces in hyperbolic 3-space", preprint, Princeton Univ., 1985.

[Nehari 1952] Z. Nehari, *Conformal mapping*, McGraw-Hill, New York, 1952. Reprinted Dover, New York, 1975. MR 51 #13206 Zbl 0048.31503

[Osgood and Stowe 1992] B. Osgood and D. Stowe, "The Schwarzian derivative and conformal mapping of Riemannian manifolds", *Duke Math. J.* **67**:1 (1992), 57–99. MR 93j:53062 Zbl 0766.53034

Received December 26, 2005. Revised September 5, 2006.

MARTIN CHUAQUI
FACULTAD DE MATEMÁTICAS
P. UNIVERSIDAD CATÓLICA DE CHILE
CASILLA 306
SANTIAGO 22
CHILE

mchuaqui@mat.puc.cl
www.mat.puc.cl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

ON ISOPERIMETRIC SURFACES IN GENERAL RELATIVITY

JUSTIN CORVINO, AYDIN GEREK,
MICHAEL GREENBERG AND BRIAN KRUMMEL

We obtain the isoperimetric profile for the standard initial slices in the Reissner–Nordstrom and Schwarzschild anti-de Sitter spacetimes, following recent work of Bray and Morgan on isoperimetric comparison. We then discuss these results in the context of Bray's isoperimetric approach to the Penrose inequality.

1. Introduction

One of the major recent developments in mathematical relativity is the resolution of the Riemannian case of the Penrose conjecture, by Huisken and Ilmanen [2001] and by Bray [2001]. Bray had obtained earlier partial results in his thesis [1997] by using isoperimetric surface techniques. As a key step, Bray established that the isoperimetric profile of the time-symmetric Schwarzschild initial data (of positive mass) is given by the radially symmetric spheres, the method of proof of which has been codified in [Bray and Morgan 2002]. The main idea is that one can deduce the isoperimetric profile of a given metric if one can construct an appropriate map to a model space (for instance, Euclidean space or hyperbolic space) in which the profile is known. We obtain below as a direct corollary the isoperimetric profile for the Reissner–Nordstrom initial data. We then carry out an extension of the method to derive the isoperimetric profile for the Schwarzschild anti-de Sitter (AdS) data; unlike the previous two families, which are asymptotically flat, Schwarzschild AdS is asymptotically hyperbolic. In all these cases, the spaces are rotationally symmetric, and the rotationally symmetric spheres give the isoperimetric profile. For contrast, in the negative mass Schwarzschild, the analogous family of spheres is unstable, as we discuss below.

We will review Bray's isoperimetric surface approach to the Penrose inequality and discuss its extension to certain asymptotically flat solutions of the Einstein–Maxwell constraint equations. We also include computations relevant to a form

MSC2000: 53C21, 83C99.
Keywords: differential geometry, isoperimetric problem, general relativity.
This research was partly supported by an NSF REU grant DMS-0243763. Gerek was funded by a Lafayette College EXCEL grant.

of the Penrose inequality for a class of asymptotically hyperbolic spaces. For background and references on the Penrose inequality, see [Bray 2002; Bray and Chruściel 2004]; for recently announced work by Huisken which explores the relation between isoperimetric inequalities and the mass of asymptotically flat metrics, see [Huisken 2005; 2006].

2. Preliminaries

We recall the isoperimetric problem and introduce the families of metrics (Schwarzschild, Reissner–Nordstrom, and Schwarzschild AdS) whose isoperimetric profiles we will discuss.

Isoperimetric problems. The isoperimetric problem is the classical problem of how to enclose a given volume V with a surface of least area. In Euclidean and hyperbolic space, homogeneity allows one to conclude that if a volume V can be enclosed with a surface of area A, a volume $V_0 < V$ can be enclosed with an area $A_0 < A$. In general spaces, one can pose an isoperimetric problem to find the minimum area that encloses a volume of at least V. It is a classical result that in Euclidean and hyperbolic spaces, the most efficient way to enclose a volume V is by using a sphere [Chavel 1993; Howards et al. 1999]. We will in fact consider the problem of minimizing volume *against* a (two-sided) hypersurface Σ_0; that is, we consider the problem of finding least-area enclosures in the homology class of Σ_0 of net volume (at least) V with Σ_0.

The metrics of interest. We will focus on three families of spherically symmetric metrics which appear as constant time slices in well-known solutions of the Einstein equations of general relativity. Let $\mathbb{S}^2$ be the two-dimensional sphere and let $d\Omega^2$ be the standard round metric on the unit two-sphere. Each of the following metrics is defined on the smooth manifold $(r_0, +\infty) \times \mathbb{S}^2$, where $m > 0$, and $r_0 > 0$ is specified below:

(1) Schwarzschild metric:

$$\left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2\, d\Omega^2, \quad r_0 = 2m.$$

(2) Schwarzschild AdS metric:

$$\left(1 + r^2 - \frac{2m}{r}\right)^{-1} dr^2 + r^2\, d\Omega^2, \quad \text{where } r_0 \text{ satisfies } 1 + r_0^2 - \frac{2m}{r_0} = 0.$$

(3) Reissner–Nordstrom metric:

$$\left(1 - \frac{2m}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2,$$

where $m^2 > Q^2$ and r_0 is the larger solution of

$$1 - \frac{2m}{r_0} + \frac{Q^2}{r_0^2} = 0.$$

The parameter m measures the deviation (in the third example, the top-order deviation) of the metrics from the model Euclidean or hyperbolic metrics. It is called the *mass*, and indeed it has an interpretation in terms of the energy of isolated gravitational systems [Bray 2002; Bray and Chruściel 2004]. See the Appendix for useful formulas (Christoffel symbols, curvatures) for these metrics. Each metric extends to r_0, where there is a minimal sphere (*horizon*) S_{r_0}. This minimal sphere is in fact totally geodesic, and the metrics can be smoothly reflected across the horizon (using inversion in the horizon sphere with respect to the metric distance along radial geodesics) to produce complete metrics with two ends. These metrics are conformally flat; for example, the extended Schwarzschild metric in appropriate coordinates is precisely $\big(1 + m/(2r)\big)^4 \delta$, where δ is the Euclidean metric. In these coordinates the horizon is located at $r = m/2$, and the inversion $r \mapsto m^2/(4r)$ is an isometry.

The Schwarzschild metric with $m > 0$ can be isometrically embedded into the Euclidean space $\mathbb{R}^4$ as the set $\{(x, y, z, w) : r = w^2/(8m) + 2m\}$, where $r^2 = x^2 + y^2 + z^2$. To see this, we look for an embedding which in terms of spherical coordinates on Schwarzschild is of the form $(r, \omega) \mapsto (r\omega, \xi(r)) \in \mathbb{R}^4$. Using the above form of the Schwarzschild metric, we see that the map is an isometry if and only if $(\xi'(r))^2 + 1 = (1 - 2m/r)^{-1}$, which can be rewritten (choosing $\xi'(r) > 0$) as

$$\xi'(r) = \sqrt{\frac{2m}{r - 2m}}.$$

We note that $\xi(r) = \sqrt{8m(r - 2m)}$ does indeed satisfy this equation. Interestingly enough, this derivation breaks down for $m < 0$; however (as was pointed out to us by Greg Galloway and Hubert Bray), the same idea can be pushed through in the negative mass case to obtain an isometric embedding of the negative mass Schwarzschild into *Minkowski* space.

The Einstein constraint equations. The three families of metrics above give particular solutions to the Einstein constraint equations, as we now recall. The Einstein equations for the corresponding four-dimensional Lorentzian spacetimes $(\mathcal{S}, \bar{g})$ in which these three-dimensional Riemannian spaces embed as totally geodesic spacelike slices are $\mathrm{Ric}(\bar{g}) = 0$, $\mathrm{Ric}(\bar{g}) = -3\bar{g}$, and $\mathrm{Ric}(\bar{g}) - \frac{1}{2}R(\bar{g})\bar{g} = 8\pi T$, respectively, where T is the stress-energy tensor of a Maxwell field [Wald 1984]. Consider in general any spacetime $(\mathcal{S}, \bar{g})$ satisfying one of these Einstein equations; then the Gauss and Codazzi equations (together with the Einstein equation) imply constraint equations on the geometry (intrinsic and extrinsic) of spacelike slices. If g is the induced metric and II the second fundamental form (with trace H) of a spacelike slice, then using the Einstein equation along with the Gauss equation, we obtain the Hamiltonian constraint, which in the first two cases yields $R(g) - \|\mathrm{II}\|^2 + H^2 = 0$ and $R(g) - \|\mathrm{II}\|^2 + H^2 = -6$, respectively. In the totally geodesic case ($\mathrm{II} = 0$), these constraints reduce to the condition of constant scalar curvature $R(g) = 0$ or $R(g) = -6$, respectively; in the case of a maximal slice ($H = 0$), the constraints imply the inequalities $R(g) \geq 0$ and $R(g) \geq -6$. Similarly, the (totally geodesic) Einstein–Maxwell constraint equations for a metric g and an electromagnetic field E are given by the Hamiltonian constraint $R(g) = 2|E|^2$ coupled with the Maxwell field equation $div_g E = 0$. If we let $\mathbf{e}_r$ be the unit outward radial vector, and couple the field $E = (Q/r^2)\mathbf{e}_r$ to the Reissner–Nordstrom metric, we produce a solution to the Einstein–Maxwell constraints.

On the isoperimetric inequality and the mass. In Euclidean space, the isoperimetric inequality for a closed surface Σ of area A enclosing a volume V can be written $V \leq A^{3/2}/(6\sqrt{\pi})$, with equality precisely when Σ is a round sphere. We compare this to Schwarzschild, where (using Corollary 3.5) it is easy to compute the volume $V(\sigma)$ enclosed by the isoperimetric sphere of area σ. In fact, if we use the conformally flat coordinates for Schwarzschild, in which the metric is $(1 + m/(2r))^4\delta$, we have

$$A(S_r) = 4\pi r^2\Big(1 + \frac{m}{2r}\Big)^4.$$

Thus

$$\frac{A(S_r)^{3/2}}{6\sqrt{\pi}} = \frac{4\pi}{3}r^3\Big(1 + \frac{3m}{r} + mO\Big(\frac{1}{r^2}\Big)\Big).$$

The net volume enclosed by S_r has the expansion

$$4\pi\int_{m/2}^{r}\Big(1 + \frac{m}{2t}\Big)^6 t^2\,dt = \frac{4\pi}{3}r^3\Big(1 + \frac{9m}{2r} + mO\Big(\frac{1}{r^2}\Big)\Big).$$

From this it is easy to see that the volume enclosed by the isoperimetric sphere of area σ has the expansion

$$V(\sigma) = \frac{\sigma^{3/2}}{6\pi^{1/2}}\bigg(1 + \frac{(3\sqrt{\pi})m}{\sqrt{\sigma}} + mO\Big(\frac{1}{\sigma}\Big)\bigg).$$

This is yet another quantitative way in which the mass m measures the deviation of the geometry from that of Euclidean, which is explored in the recent work of Huisken [2005; 2006].

3. Isoperimetric profiles by comparison

We review the isoperimetric comparison theorem of Bray and Morgan and apply it to the Schwarzschild and Reissner–Nordstrom spaces. Let $I \subset \mathbb{R}$ be an interval. Suppose we have a rotationally symmetric model space $M_0 = I \times \mathbb{S}^2$ with the twisted product metric $dr^2 + \varphi_0^2(r)\, d\Omega^2$ for which we know the isoperimetric surfaces are the radially symmetric spheres $S_c = \{r = c\}$. We consider another rotationally symmetric space $M = I \times \mathbb{S}^2$ with the metric $dr^2 + \varphi^2(r)\, d\Omega^2$. Bray and Morgan showed that under certain geometric conditions, the isoperimetric surfaces in M are also the radially symmetric spheres. We now recall their argument, which as in [Bray and Morgan 2002] can be more generally applied to twisted products $I \times N$ with a closed manifold fiber N.

Let $F : M \to M_0$ map radially symmetric spheres in M to radially symmetric spheres in M_0, so that $F(r, \omega) = (\psi(r), \omega)$. We assume that ψ is increasing, so that F is orientation-preserving. We define the area stretch AS_Σ for a surface $\Sigma \subset M$ by the equation $F^*(dA_{F(\Sigma)}) = \mathrm{AS}_\Sigma\, dA_\Sigma$, where dA_Σ and $dA_{F(\Sigma)}$ are the area forms of $\Sigma \subset M$ and $F(\Sigma) \subset M_0$, respectively. The volume stretch VS is defined similarly by $F^*(dV_{M_0}) = \mathrm{VS}\, dV_M$, where dV_M and dV_{M_0} are the volume forms of M and M_0, respectively. By symmetry, VS depends only on r. Finally, let $A(\Sigma)$ be the area of the surface $\Sigma \subset M$, and let $A_0(\Sigma_0)$ be the area of the surface $\Sigma_0 \subset M_0$.

Let $a = A(S_{r_1})/A_0(F(S_{r_1}))$. Suppose the map F can be constructed so that the area stretch under F satisfies $\mathrm{AS}_\Sigma \le 1/a$, so the volume stretch satisfies $\mathrm{VS}(r) \le b$ for $r < r_1$, and $\mathrm{VS}(r) \ge b$ for $r > r_1$. Now suppose there were a surface $\Sigma \subset M$ bounding nonnegative net volume against S_{r_1} (that is, Σ bounds no less volume against S_{r_0} than S_{r_1} does), so that Σ has the same or less surface area as S_{r_1}. We will show that in fact $A(\Sigma) = A(S_{r_1})$, which will then imply that S_{r_1} is an isoperimetric surface. Since the volume stretch for $r > r_1$ is no less than the volume stretch for $r < r_1$, the net volume bounded by $F(\Sigma)$ contained in $\{r > \psi(r_1)\}$ is no less than the net volume bounded by $F(\Sigma)$ contained in $\{r < \psi(r_1)\}$. Thus the net volume bounded by $F(\Sigma)$ is greater than or equal to the volume bounded by $F(S_{r_1})$. Since the area stretch $\mathrm{AS}_\Sigma \le 1/a = A_0(F(S_{r_1}))/A(S_{r_1})$, and $A(\Sigma) \le A(S_{r_1})$, we obtain

$$\begin{aligned} A_0(F(\Sigma)) &= \int_{F(\Sigma)} dA_{F(\Sigma)} = \int_\Sigma F^* dA_{F(\Sigma)} = \int_\Sigma \mathrm{AS}_\Sigma\, dA_\Sigma \\ &\le \frac{1}{a} A(\Sigma) = A_0(F(S_{r_1}))\frac{A(\Sigma)}{A(S_{r_1})} \le A_0(F(S_{r_1})). \end{aligned}$$

Since $F(S_{r_1}) = S_{\psi(r_1)}$ is an isoperimetric surface in M_0, $F(\Sigma)$ and $F(S_{r_1})$ must thus bound the same amount of volume and have the same surface areas, $A_0(F(\Sigma)) = A_0(F(S_{r_1}))$. Thus the inequalities above must be equalities, and we see that indeed

$A(\Sigma) = A(S_{r_1})$. Therefore we have shown by comparison that S_{r_1} is an isoperimetric surface in M; if we have uniqueness for the isoperimetric surfaces, we can go further to assert $\Sigma = S_{r_1}$.

To put this observation to work, one identifies concrete geometric conditions that allow such a map F to be constructed. Indeed the main theorem in [Bray and Morgan 2002] is stated in geometric terms from which the following is readily established as a corollary. We note that the comparison space M_0 for this corollary is Euclidean space, so the comparison metric is $dr^2 + r^2\, d\Omega^2$.

Theorem 3.1 [Bray and Morgan 2002]. *Consider a rotationally symmetric three-manifold $M = I \times \mathbb{S}^2$ with the metric $dr^2 + \varphi^2(r)\, d\Omega^2$. Suppose* (1) *$\varphi'$ is nondecreasing for all r, and* (2) *$0 \le \varphi' \le 1$ for all $r \ge r_0$. Then for all $r \ge r_0$, the radially symmetric spheres S_r minimize surface area among smooth surfaces enclosing the same volume with S_{r_0}, where volume inside $\{r < r_0\}$ is counted as negative. Furthermore, these spheres are unique minimizers if $\varphi'(r) < 1$.*

Condition (1) *holds if and only if M has nonpositive radial Ricci curvature. For any r, condition* (2) *holds if and only if S_r has nonnegative* (*inward*) *mean curvature and M has nonnegative tangential sectional curvature, or equivalently, S_r has nonnegative Hawking mass.*

We take the mean curvature to be the trace of the second fundamental form (the sum of the principal curvatures), not the average of the principal curvatures as in [Bray and Morgan 2002]. We recall the Hawking mass of a surface Σ is

$$m_H(\Sigma) = \sqrt{\frac{A(\Sigma)}{16\pi}}\left(1 - \frac{1}{16\pi}\int_\Sigma H^2\, dA\right).$$

We will see the Hawking mass play a role in the Penrose inequality below; in fact, the underlying motivation for the Huisken–Ilmanen inverse mean curvature flow is the monotonicity of the Hawking mass under the flow [Geroch 1973].

We are interested in spaces (M, g) with $M = I \times \mathbb{S}^2$ and with $g = f(r)\, dr^2 + r^2\, d\Omega^2$, where f is a positive function. The metrics which we study here, in both the forms given above and for the metrics suitably extended by reflection, all have this form. In order to apply Theorem 3.1 to such spaces, we note:

Lemma 3.2. *The metric $g = f(r)\, dr^2 + r^2 d\Omega^2$ can be written as a twisted product metric $g = dt^2 + \varphi^2(t)\, d\Omega^2$, where $\varphi(t) > 0$.*

Proof. The result is equivalent to $dt = \sqrt{f(r)}\, dr$, for $r = \varphi(t)$. We integrate to find $t = t(r)$; by the equation t is increasing, and we write the inverse as $r = \varphi(t)$. □

Theorem 3.3. *Consider the space $M = I \times \mathbb{S}^2$ with metric $g = f(r)\, dr^2 + r^2 d\Omega^2$. Suppose $f'(r) \le 0$ for all r and $f(r) \ge 1$ for $r \ge r_0$. Then every sphere of revolution S_r for $r \ge r_0$ minimizes perimeter among smooth surfaces enclosing fixed volume with S_{r_0}, uniquely if $f(r) > 1$ for $r \ge r_0$.*

Proof. It suffices to show that M satisfies the conditions for Theorem 3.1, in particular that M has nonpositive radial Ricci curvature, S_r has nonnegative mean curvature (with respect to the inward unit normal), and M has nonnegative tangential sectional curvature. For indices, let $(1, 2, 3)$ represent (r, ϕ, θ). We find $R_{12} = R_{13} = 0$ and $R_{11} = f'/(rf)$ (see the Appendix), so the radial Ricci curvature is nonpositive if and only if $f' \leq 0$. We know that $H_{S_r} = 2/(r\sqrt{f}) > 0$ as required. We compute the sectional curvature K of the plane containing ∂_ϕ and ∂_θ as

$$K = \frac{g_{33}R^3_{232}}{g_{22}g_{33} - (g_{23})^2} = \frac{\left(1 - \dfrac{1}{f(r)}\right) r^2 \sin^2\phi}{r^2 r^2 \sin^2\phi - 0^2} = r^{-2}\left(1 - \frac{1}{f(r)}\right)$$

Thus $K \geq 0$ if and only if $f \geq 1$. The spheres S_r are uniquely minimizing provided $f > 1$. □

Remark 3.4. It is often convenient to consider the function $1/f$ instead of f. If $h = 1/f$, $f' = -h'/h^2$, so $f' \leq 0$ if and only if $h' \geq 0$. To check if $f \geq 1$, we check if $h \leq 1$ and similarly for strict inequality, in which case the tangential sectional curvature is strictly positive.

The Schwarzschild profile. We let g be the Schwarzschild metric with $m > 0$, which we recall has the form $(1 - 2m/r)^{-1}\, dr^2 + r^2\, d\Omega^2$ on $(2m, +\infty) \times \mathbb{S}^2$. We recall the following result from [Bray 1997], proved as in [Bray and Morgan 2002].

Corollary 3.5 [Bray 1997]. *In the Schwarzschild metric with positive mass $m > 0$, every sphere of revolution S_r for $r \geq 2m$ uniquely minimizes perimeter among smooth surfaces enclosing fixed volume with the horizon S_{2m}.*

Proof. Let $h(r) = 1 - 2m/r$. We note $h(r) = 1 - 2m/r < 1$ for positive mass. Also, $h'(r) = 2m/r^2 > 0$, so by Theorem 3.3, every sphere of revolution S_r for $r \geq 2m$ uniquely minimizes perimeter among smooth surfaces enclosing fixed volume with the horizon S_{2m}. □

Remark 3.6. Of course if we consider the full Schwarzschild space with reflection symmetry, then uniqueness is with respect to one chosen end. Similar considerations apply to Reissner–Nordstrom and Schwarzschild AdS below.

The Reissner–Nordstrom profile. Let g be the Reissner–Nordstrom metric, which on $(r_0, \infty) \times \mathbb{S}^2$ takes the form $g = h(r)^{-1}\, dr^2 + r^2\, d\Omega^2$, with

$$h(r) = 1 - \frac{2m}{r} + \frac{Q^2}{r^2}.$$

We shall assume $m^2 > Q^2$, so that h has two positive roots, and we take r_0 to be the larger of the two. Then $r_0 > m > Q^2/m$.

Corollary 3.7. *In Reissner–Nordstrom with $m^2 > Q^2$, every sphere of revolution S_r for $r \geq r_0$ uniquely minimizes perimeter among smooth surfaces enclosing fixed volume with S_{r_0}.*

Proof. We have $h(r) < 1$ for $r > Q^2/(2m)$. We also have $h'(r) = 2m/r^2 - 2Q^2/r^3$, so that $h'(r) \geq 0$ for $r \geq Q^2/m$. Both conditions of Theorem 3.3 hold for $r \geq r_0$, so every sphere of revolution S_r for $r \geq r_0$ uniquely minimizes perimeter among smooth surfaces enclosing fixed volume with S_{r_0}. □

4. Isoperimetric profile for Schwarzschild AdS

We now let the comparison space M_0 be hyperbolic three-space with hyperbolic metric $(1+r^2)^{-1}\,dr^2 + r^2\,d\Omega^2$. Consider $M = (r_0, \infty) \times \mathbb{S}^2$ with the Schwarzschild AdS metric $g = \big(1 + r^2 - 2m/r\big)^{-1}\,dr^2 + r^2\,d\Omega^2$. We will construct a comparison map $F : M \to M_0$ given by $F(r, \omega) = (\psi(r), \omega)$ to show that the radially symmetric spheres are the isoperimetric surfaces in Schwarzschild AdS space.

We will be concerned with two particular types of area stretches. The first one encodes the area stretch for a radially symmetric sphere, $F^*(dA_{F(S_r)}) = \mathrm{AS}_1\,dA_{S_r}$:

$$\mathrm{AS}_1(r) = \frac{\displaystyle\int_{\mathbb{S}^2} \psi^2(r)\,dA_{\mathbb{S}^2}}{\displaystyle\int_{\mathbb{S}^2} r^2\,dA_{\mathbb{S}^2}} = \frac{\psi^2(r)}{r^2}.$$

For example, in the previous section, we had $\mathrm{AS}_1 = 1/a$.

The second stretch factor encodes the area stretch for an annular surface $\Sigma = J \times \mathbb{S}^1$ ($J \subset (r_0, +\infty)$), obtained by flowing some great circle $\mathbb{S}^1$ (with element of arclength ds) along the radial direction field ∂_r, $F^*(dA_{F(\Sigma)}) = \mathrm{AS}_2\,dA_\Sigma$:

$$\mathrm{AS}_2(r) = \frac{\dfrac{d}{dr}\displaystyle\int_{\psi(r_0)}^{\psi(r)}\int_{\mathbb{S}^1} \rho(1+\rho^2)^{-1/2}\,ds\,d\rho}{\dfrac{d}{dr}\displaystyle\int_{r_0}^{r}\int_{\mathbb{S}^1} \rho(1+\rho^2 - 2m/\rho)^{-1/2}\,ds\,d\rho} = \frac{\psi(r)(1+\psi^2(r))^{-1/2}\psi'(r)}{r(1+r^2-2m/r)^{-1/2}}.$$

The volume stretch VS, where $F^*(dV_{M_0}) = \mathrm{VS}\,dV_M$, is given by

$$\mathrm{VS}(r) = \frac{\dfrac{d}{dr}\displaystyle\int_{\psi(r_0)}^{\psi(r)}\int_{\mathbb{S}^2} \rho(1+\rho^2)^{-1/2}\,dA_{\mathbb{S}^2}\,d\rho}{\dfrac{d}{dr}\displaystyle\int_{r_0}^{r}\int_{\mathbb{S}^2} \rho^2(1+\rho^2-2m/\rho)^{-1/2}\,dA_{\mathbb{S}^2}\,d\rho} = \frac{\psi^2(r)(1+\psi^2(r))^{-1/2}\psi'(r)}{r^2(1+r^2-2m/r)^{-1/2}}.$$

Note that $\mathrm{VS} = \sqrt{\mathrm{AS}_1}\,\mathrm{AS}_2$.

Lemma 4.1. *The area stretch AS_Σ for any surface does not exceed the maximum of AS_1 and AS_2.*

Proof. By dimension considerations, if Σ is any smooth surface, $T_p\Sigma$ contains at least one tangent direction to the radial sphere through p. We let E_1 be such a unit vector; let E_2 be an orthogonal unit vector tangent to the radial sphere, and let E_3 be the unit outward radial vector. There exist α and β with $\alpha^2+\beta^2=1$ so that $\alpha E_2+\beta E_3 \in T_p\Sigma$. We have by orthogonality

$$\begin{aligned}\mathrm{AS}_\Sigma &= dA_{F(\Sigma)}(F_*(E_1), \alpha F_*(E_2)+\beta F_*(E_3)) = \left|F_*(E_1)\right| \left|\alpha F_*(E_2)+\beta F_*(E_3)\right| \\ &= \frac{\psi(r)}{r}\sqrt{\alpha^2\frac{\psi(r)^2}{r^2}+\beta^2\frac{(\psi'(r))^2((1+\psi(r))^{-1}}{(1+r^2-2m/r)^{-1}}}.\end{aligned}$$

Thus $\mathrm{AS}_\Sigma^2=\alpha^2\mathrm{AS}_1^2+\beta^2\mathrm{AS}_2^2$, from which the claim follows. □

As above, we will produce a map $F: M \to M_0$ with the following properties: at $r=r_1$, the area stretch $\mathrm{AS}_1(r_1)=1/a$ and the volume stretch $\mathrm{VS}(r_1)=b$, for some $a, b>0$; for all Σ, $\mathrm{AS}_\Sigma \leq 1/a$; $\mathrm{VS}(r) \geq b$ for $r>r_1$, and $\mathrm{VS}(r) \leq b$ for $r<r_1$. By the lemma, it suffices to show that $\mathrm{AS}_1, \mathrm{AS}_2 \leq 1/a$ everywhere. (The construction in [Bray and Morgan 2002] uses only the parameter a, in which case $\mathrm{VS}(r_1)=1/a$; this suffices for the asymptotically flat cases above, but we require slightly more flexibility in constructing the map F in the asymptotically hyperbolic case, and so we introduce the parameter b.) As above, it follows that for any competitor surface Σ bounding at least as much volume as S_{r_1} with equal or less surface area, the image $F(\Sigma)$ will bound no less volume with no more surface area than $F(S_{r_1})$. (In hyperbolic space M_0, we can shrink $F(\Sigma)$ to produce a surface Σ' bounding the same volume as $F(S_{r_1})$ with less or equal surface area.) But the radially symmetric spheres are isoperimetric surfaces in hyperbolic space M_0, so all previously mentioned area and volume inequalities must be equalities; hence radially symmetric spheres are isoperimetric surfaces in M. Furthermore, if the maximal area stretch is strictly tangential ($\mathrm{AS}_2<1/a$), radially symmetric spheres are the unique isoperimetric surfaces in M.

Theorem 4.2. *In Schwarzschild AdS, every sphere of revolution S_r uniquely minimizes perimeter among smooth surfaces enclosing fixed volume with S_{r_0}.*

Proof. First we consider $r_1>2m$. Let $a=1-2m/r_1<1$, and define F using $\psi(r)=a^{-1/2}r$ for all $r \geq r_0$. Then $\mathrm{AS}_1=1/a$ everywhere. Also

$$\mathrm{AS}_2(r)=\frac{\sqrt{1+r^2-2m/r}}{a\sqrt{1+a^{-1}r^2}}.$$

Hence $\mathrm{AS}_2(r)<1/a$ is equivalent to $1-2mr^{-3}<1/a$. Since $1-2mr^{-3}<1<1/a$ for all $r>0$, the maximal area stretch equals $\mathrm{AS}_1=1/a$, is strictly tangential, and occurs on S_{r_1}.

At at $r = r_1$, we have $\mathrm{AS}_2(r_1) = a^{-1/2}$, while $\mathrm{AS}_2(r) \to a^{-1/2}$ as $r \to \infty$. We have

$$\frac{d}{dr}\left(\frac{1+r^2-2m/r}{1+a^{-1}r^2}\right) = \frac{(2r+2m/r^2)(1+a^{-1}r^2)-(1+r^2-2m/r)(2a^{-1}r)}{(1+a^{-1}r^2)^2}$$
$$= \frac{2(a-1)r^3+6mr^2+2am}{ar^2(1+a^{-1}r^2)^2}.$$

The cubic numerator has a positive local minimum at $r = 0$ and one other critical point at some $r > 0$, so in particular it has only one root (which is positive). Thus AS_2 has a unique maximum on the set $r \geq r_0$. Since AS_2 decreases to $a^{-1/2}$ as $r \to \infty$, the maximum occurs on (r_1, ∞), and on this interval

$$\mathrm{AS}_2(r) > \mathrm{AS}_2(r_1) = a^{-1/2}.$$

Hence $\mathrm{VS}(r) = \sqrt{\mathrm{AS}_1(r)}\,\mathrm{AS}_2(r) \leq a^{-1/2}\,a^{-1/2} = 1/a$ for $r \leq r_1$, and $\mathrm{VS}(r) = \sqrt{\mathrm{AS}_1(r)}\,\mathrm{AS}_2(r) \geq a^{-1/2}\,a^{-1/2} = 1/a$ for $r \geq r_1$. Since areas and volumes stretch in the required manner, S_r are the unique isoperimetric surfaces for $r > 2m$.

Now suppose $r_1 \leq 2m$. Choose $a \in (0, 1)$ and let $\psi(r) = a^{-1/2}r$ for all $r \geq r_0$. Then $\mathrm{AS}_1 = 1/a$ and $\mathrm{AS}_2 < 1/a$ everywhere as before. Note that at $r = r_1$, $\mathrm{AS}_2(r_1) < a^{-1/2}$ since $1 - 2m/r_1 \leq 0$. As before, AS_2 has a unique maximum for $r \geq r_0$ and AS_2 decreases to $a^{-1/2}$ as $r \to \infty$. Hence the maximum occurs for some $r_{max} > r_1$, and AS_2 is increasing on (r_0, r_{max}). Thus the volume stretch $\mathrm{VS} = a^{-1/2}\mathrm{AS}_2$ is also increasing on (r_0, r_{max}), and so $\mathrm{VS}(r) \leq b := \mathrm{VS}(r_1)$ for $r < r_1$, and $\mathrm{VS}(r) \geq b$ for $r \in [r_1, r_{max}]$. Furthermore, $b = \mathrm{VS}(r_1) = \sqrt{\mathrm{AS}_1(r_1)}\mathrm{AS}_2(r_1) < 1/a$, so for $r > r_{max}$, $\mathrm{VS}(r) = \sqrt{\mathrm{AS}_1(r)}\mathrm{AS}_2(r) \geq 1/a > b$. Since areas and volumes stretch in the required manner, S_r are the unique isoperimetric surfaces for $r \leq 2m$. □

5. Remarks on the negative mass Schwarzschild

If we let the mass m be negative in the formula $(1-2m/r)^{-1}\,dr^2 + r^2\,d\Omega^2$ for the Schwarzschild metric, we obtain an inextendible metric with no minimal sphere. The coordinates are only singular at the origin; in fact the metric is incomplete, as radial geodesics have finite length as $r \to 0^+$, but the Ricci tensor blows up on approach to the origin. The Bray–Morgan construction for the positive-mass Schwarzschild does not extend to the negative mass case; in fact we will show below that radial spheres are unstable.

Instability of the radial spheres. We consider now the variations of area and volume enclosed by the coordinate spheres, and compute the second variation of area with respect to volume-preserving perturbations. The variation formulas are standard [Chavel 1993; Taylor 1996]. We note that H below is the trace of the second fundamental form computed with respect to the *inward* unit normal $-\nu$, which accounts for a sign difference from some versions of the variation formulas.

We consider a smooth family of surfaces Σ_t obtained from $\Sigma = \Sigma_0$ using the variation field given by $V(x,t) = \eta(x,t)\nu(x,t)$. Then we have the first variation $A'(t) = \int_{\Sigma_t} H\eta\, dA$, and the second variation

$$A''(0) = \int_\Sigma \Big(\eta\big(-\Delta_\Sigma \eta - \eta\|\mathrm{II}\|^2 - \eta\,\mathrm{Ric}(\nu,\nu)\big) + H\frac{\partial\eta}{\partial t} + H^2\eta^2\Big)\, dA.$$

The first variation of volume $V(t)$ inside Σ_t is given by $V'(t) = \int_{\Sigma_t} \eta\, dA$, so the second variation is $V''(t) = \int_{\Sigma_t} (H\eta^2 + \partial\eta/\partial t)\, dA$.

The radial spheres $\Sigma = S_r$ have constant mean curvature, and hence they are critical points for the area functional with respect to volume-preserving perturbations. Indeed, from the variation of volume formula, we have $0 = V'(0) = \int_\Sigma \eta\, dA$, which implies that $A'(0) = 0$ too. If we now consider the second variation *at* S_r, since the mean curvature H is *constant* we have $0 = HV''(0) = \int_{S_r} (H^2\eta^2 + H\partial\eta/\partial t)\, dA$. Thus the second variation formula simplifies; if we also apply the divergence theorem to the first term, we then have

$$A''(0) = \int_{S_r} \big(|\nabla^\Sigma \eta|^2 - \eta^2\|\mathrm{II}\|^2 - \eta^2\,\mathrm{Ric}(\nu,\nu)\big)\, dA. \tag{5-1}$$

From the Appendix we have $\nu = \sqrt{1-2m/r}\;\partial_r$, $\mathrm{Ric}(\nu,\nu) = -2m/r^3$ and $\|\mathrm{II}\|^2 = (2/r^2)(1-2m/r)$. When we plug this into the preceding equation we get

$$A''(0) = \int_{S_r} \Big(|\nabla^\Sigma \eta|^2 - \eta^2 \frac{2}{r^2}\Big(1 - \frac{3m}{r}\Big)\Big)\, dA.$$

It is well known [Axler et al. 1992] that the lowest nonzero eigenvalue λ_1 for the Laplacian on a round two-sphere $\mathbb{S}^2_\kappa$ of curvature κ is $\lambda_1 = 2\kappa$, with eigenspace spanned by the restriction of the coordinate functions x, y, z to the sphere (isometrically embedded in $\mathbb{R}^3$ centered at the origin): e.g., $\Delta_{\mathbb{S}^2_\kappa}(x) = -2\kappa x$. We now invoke the Poincaré inequality we obtain from the decomposition of $L^2(\Sigma)$ by the eigenspaces of the Laplacian [Chavel 1993]: $\lambda_1 \int_\Sigma \eta^2\, dA \le \int_\Sigma |\nabla^\Sigma\eta|^2\, dA$, for all η with $\int_\Sigma \eta\, dA = 0$; equality holds precisely for functions in the λ_1-eigenspace. Applying this with $\Sigma = S_r$ we have $\lambda_1 = 2/r^2$, so that

$$A''(0) \ge \int_{S_r} \frac{6m}{r^3}\eta^2\, dA, \tag{5-2}$$

with equality if and only if η is in the λ_1-eigenspace. We see from this that in the *positive* mass Schwarzschild case, the second variation must be positive for (nontrivial) volume-preserving deformations (which we knew already from the isoperimetric profile) . But in the negative mass case, we see that for η a coordinate function, the right-hand side of (5-2) is negative. We note that $\eta(x,t)=x$ does not satisfy $V''(0)=0$. To satisfy this condition, we can let $\eta_0(x,t)=x+\alpha t$, where α is a constant chosen precisely so that $V''(0)=0$. Then η_0 generates a deformation that preserves volume to second order; from here it is not hard to modify the variation by a scaling to preserve volume, and so that the corresponding η has first-order Taylor expansion η_0. Another way to see that the spheres do not minimize area for a given volume is by considering the variation $\eta_0\nu$. Since this variation leaves the volume unchanged to second-order in t, the change in volume is $O(t^3)$. Now, the volume $V(S_r)$ enclosed by the radial spheres satisfies

$$\frac{dV(S_r)}{dr}=\frac{4\pi r^2}{\sqrt{1-2m/r}}>0,$$

so the radius $r(t)$ of the radial sphere with volume $V(t)$ is such that $(r(t)-r)=O(t^3)$. So the area $A(S_{r(t)})=A(S_r)+O(t^3)$, and thus $A''(0)<0$ implies that for some C and small $t>0$, the area $A(t)$ of Σ_t satisfies $A(t)<A(S_r)-Ct^2<A(S_{r(t)})$. This should not be surprising by considering the growth of the volume for small r:

$$V(S_r)=4\pi\int_0^r t^2\sqrt{\frac{1}{1-2m/t}}\,dt<4\pi\int_0^r t^2\sqrt{\frac{t}{2|m|}}\,dt=O(r^{7/2}).$$

This volume growth is slower than for the Euclidean metric $dr^2+r^2d\Omega^2$, but the radial spheres have the same area as in the Euclidean metric, so that it is more efficient to slide them off-center. It might be interesting to consider the isoperimetric problem in this singular space, and whether optimizing shapes tend to singular varieties that go *through* the singular point.

6. The Penrose inequality from isoperimetric techniques

The Riemannian Penrose inequality is a lower bound on the ADM mass of an asymptotically flat metric of nonnegative scalar curvature in terms of the areas of certain horizons. There are a host of partial results, including the isoperimetric approach of [Bray 1997], and then there are the proofs of [Huisken and Ilmanen 2001] and [Bray 2001]. We state the version from the latter reference.

Theorem 6.1 (Penrose Inequality). *Let (M,g) be asymptotically flat with $R(g)\geq 0$. Let m be the ADM mass of an end, and let A be the total surface area of the outermost minimal spheres with respect to this end. Then $m\geq\sqrt{A/(16\pi)}$.*

Various analogues of this inequality have been sought [Bray and Chruściel 2004], including asymptotically hyperbolic versions and versions with charge. We discuss an example each for both types, to illustrate that the beautiful arguments of Bray [1997] which connect the isoperimetric profiles to the Penrose inequality extend to the context of the isoperimetric profiles obtained above.

Variation of area along an isoperimetric profile. We again consider the isoperimetric problem of minimizing area for volume V between a horizon and competitor surfaces in the homology class of the horizon. We assume we have an isoperimetric profile $\Sigma(V)$, each surface of which is connected. The objective in the next sections will be to establish that a mass function $m(V)$ associated with the Hawking mass function $m_H(\Sigma(V))$ determined by the isoperimetric profile is nondecreasing, for which we now derive a key inequality. We compute the variation of the area function $A(V)$ of the profile, where we employ the harmless abuse of notation, $A(V) := A(\Sigma(V))$, and we note that $A(0) = A(\Sigma_0)$. The area function of the isoperimetric profile may not be smooth in V, so that this fact is established in a weak but sufficient form. To be precise, for each $V_0 > 0$, we let $A_{V_0}(V)$ be the area of the surface obtained by flowing $\Sigma(V_0)$ in the outward normal direction at unit speed until the volume enclosed with the horizon is V. A_{V_0} will be smooth for V near V_0. Moreover, $A_{V_0}(V_0) = A(V_0)$ and $A_{V_0}(V) \geq A(V)$. Thus if A were smooth, then $A'(V_0) = A'_{V_0}(V_0)$ and $A''(V_0) \leq A''_{V_0}(V_0)$; so an inequality for the derivatives of A_{V_0} at V_0 can be interpreted as a weak (distributional) inequality for the derivatives of A. We let $\Sigma^t_{V_0}$ be the surface obtained by flowing $\Sigma(V_0)$ for time t, and let $V(t)$ be the volume this surfaces encloses with the horizon. Then, by the equations of variation (as recalled in the preceding section), we have

$$\frac{d}{dt}(A_{V_0}(V(t))) = \int_{\Sigma^t_{V_0}} H\, dA, \qquad \frac{dV}{dt} = A_{V_0}(V(t)),$$

so that

$$\frac{d}{dV}(A_{V_0}(V)) = A'_{V_0}(V) = \frac{\displaystyle\int_{\Sigma^t_{V_0}} H\, dA}{A_{V_0}(V(t))}.$$

By the second variation of area formula we obtain (since $\eta = 1$)

$$A_{V_0}(V_0)^2 A''_{V_0}(V_0) = \int_{\Sigma(V_0)} \bigl(-\|\mathrm{II}\|^2 - \mathrm{Ric}(\nu,\nu) + H^2\bigr)\, dA.$$

Taking the trace of the Gauss equation gives $\mathrm{Ric}(\nu,\nu) = \frac{1}{2}R - K + \frac{1}{2}(H^2 - \|\mathrm{II}\|^2)$, where $R = R(g)$ is the scalar curvature of the ambient three-space and K is the Gauss curvature of the surface. We obtain

$$A_{V_0}(V_0)^2 A''_{V_0}(V_0) = \int_{\Sigma(V_0)} \bigl(-\tfrac{1}{2}R + K - \tfrac{1}{2}H^2 - \tfrac{1}{2}\|\mathrm{II}\|^2\bigr)\, dA.$$

Since $\Sigma(V_0)$ has only one component by assumption,

$$\int_{\Sigma(V_0)} K\, dA = 2\pi\chi(\Sigma(V_0)) \leq 4\pi$$

by the Gauss–Bonnet theorem. Since $\|\mathrm{II}\|^2 \geq \frac{1}{2}H^2$, we arrive at the inequality

$$A_{V_0}(V_0)^2 A''_{V_0}(V_0) \leq 4\pi + \int_{\Sigma(V_0)} (-\tfrac{1}{2}R - \tfrac{3}{4}H^2)\, dA, \tag{6-1}$$

which we apply below.

Penrose inequality for some solutions of the Einstein–Maxwell constraints. We now discuss the Penrose Inequality in the context of a certain class of solutions to the Einstein–Maxwell constraints. As noted in [Weinstein and Yamada 2005], in the case of a connected horizon, the Huisken–Ilmanen proof can be carried through to prove the Penrose inequality that we discuss below, under less restrictive assumptions. We also remark that Weinstein and Yamada [2005] showed that for multiple-component horizons, a natural related Penrose inequality fails.

Proposition 6.2. *Assume (M, g, E) is an asymptotically flat solution of the Einstein–Maxwell constraints $R(g) = 2|E|^2$, $div_g(E) = 0$, which outside a compact set agrees with Reissner–Nordstrom data on the exterior of a ball, with mass m and charge Q, and $m > |Q|$. Suppose furthermore that M has only one horizon Σ_0 and admits a connected isoperimetric profile (with respect to Σ_0) $\Sigma(V)$, so that for sufficiently large V, $\Sigma(V)$ is a spherically symmetric sphere in Reissner–Nordstrom. Then*

$$m \geq \sqrt{\frac{A(\Sigma_0)}{16\pi}} + \frac{Q^2}{2r_0} = \frac{1}{2}\Big(r_0 + \frac{Q^2}{r_0}\Big),$$

where r_0 is defined by $A(\Sigma_0) = 4\pi r_0^2$.

Proof. We have established the isoperimetric profile for Reissner–Nordstrom in Corollary 3.7. We discuss the calculations that relate the mass to the Hawking mass of the isoperimetric surfaces for the model. Since solutions (g, E) of the Einstein–Maxwell constraints have nonnegative scalar curvature $R(g) = 2|E|^2$, we have from (6-1)

$$\begin{aligned} A_{V_0}(V_0)^2 A''_{V_0}(V_0) &\leq 4\pi - \int_{\Sigma(V_0)} (|E|^2 + \tfrac{3}{4}H^2)\, dA \\ &= 4\pi - \tfrac{3}{4}H^2 A_{V_0}(V_0) - \int_{\Sigma(V_0)} |E|^2\, dA. \end{aligned}$$

Since E is divergence-free, the flux integral $\int_\Sigma E^i \nu_i\, dA$ is a homological invariant, and thus is just $4\pi Q$. The preceding inequality thus yields (using Cauchy–Schwarz)

$$A_{V_0}(V_0)^2 A''_{V_0}(V_0) \leq 4\pi - \tfrac{3}{4}H^2 A_{V_0}(V_0) - \frac{(4\pi Q)^2}{A(V_0)}.$$

Since $A'_{V_0}(V_0) = H$, this can, as noted above, be interpreted as a weak formulation of

$$A''(V) \le \frac{4\pi}{A(V)^2} - \frac{3A'(V)^2}{4A(V)} - \frac{(4\pi Q)^2}{A(V)^3}.$$

Equivalently, for $F = A^{3/2}$ we have

$$F''(V) \le \frac{36\pi - F'(V)^2 - 144\pi^2 Q^2 F(V)^{-2/3}}{6F(V)}. \tag{6-2}$$

We will work with the mass function $m(V)$, defined by

$$m(V) = \frac{F(V)^{1/3}}{144\pi^{3/2}}\bigl(36\pi - F'(V)^2\bigr) + \sqrt{\pi}\; Q^2 F(V)^{-1/3}.$$

If $F(V)$ were smooth, we would have

$$m'(V) = \frac{\frac{1}{3}F'(V)(F(V))^{-2/3}}{144\pi^{3/2}} \times \bigl((36\pi - (F'(V))^2) - 6F(V)F''(V) - \sqrt{\pi}\, Q^2 (F(V))^{-2/3}\bigr).$$

In view of (6-2), and since $F(V)$ is nondecreasing (there being only one horizon), $m(V)$ is a nondecreasing function. Actually this statement requires some care to prove, since the function $F(V)$ may fail to be smooth, so one would need to check directly that $m'(V) \ge 0$ in the sense of distributions, that is, by pairing with appropriate test functions. We omit the details.

If $Q = 0$, the mass function is the Hawking mass of the isoperimetric surface bounding a volume V, since $F'(V) = \frac{3}{2}A(V)^{1/2}A'(V) = \frac{3}{2}A(V)^{1/2}H$ implies

$$\begin{aligned} m(V) &= \frac{A(V)^{1/2}}{144\pi^{3/2}}\bigl(36\pi - \tfrac{9}{4}A(V)H^2\bigr) + \sqrt{\pi}\; Q^2 F(V)^{-1/3} \\ &= \sqrt{\frac{A(V)}{16\pi}}\left(1 - \int_{\Sigma(V)} \frac{H^2}{16\pi}\right) + \sqrt{\pi}\; Q^2 F(V)^{-1/3}. \end{aligned}$$

Since $H = 0$ at the horizon, we have

$$m(0) = \sqrt{\frac{A(0)}{16\pi}} + \frac{Q^2}{2r_0}.$$

For V sufficiently large $\Sigma(V)$ is a radial sphere S_r in Reissner–Nordstrom, so $m(V)$ is the Hawking mass of S_r plus the charge term:

$$m(V) = \sqrt{\frac{4\pi r^2}{16\pi}}\left(1 - 4\pi r^2 \frac{1 - 2m/r + Q^2/r^2}{4\pi r^2}\right) + \frac{Q^2}{2r} = m.$$

Hence $m = \lim_{V\to+\infty} m(V) \geq m(0)$, giving us a Penrose Inequality with charge:

$$m \geq \sqrt{\frac{A(0)}{16\pi}} + \frac{Q^2}{2r_0} = \frac{1}{2}\Big(r_0 + \frac{Q^2}{r_0}\Big). \qquad \square$$

We now briefly sketch how to use this result to conclude the Penrose inequality holds for more general asymptotically flat solutions (M, g, E) of the Einstein–Maxwell constraints. We cite a condition (C1) from [Bray 1997]: there is only one horizon, and for $V > 0$, if one or more isoperimetric surfaces exists for this volume V, then at least one of these surfaces has only one component. This condition is not required in [Bray 2001], [Huisken and Ilmanen 2001], for which if there is more than one horizon, one considers the *outermost* horizons in any end. We have an approximation result from [Corvino $\geq$ 2007] which allows us to normalize the asymptotics: asymptotically flat solutions (M, g, E) of the Einstein–Maxwell constraints admit approximations by data which agree with suitable Reissner–Nordstrom data in each end, where the perturbation is localized near infinity. Assuming condition (C1) holds after this perturbation, one shows that the isoperimetric surfaces $\Sigma(V)$ exist and agree with those of Reissner–Nordstrom for sufficiently large V. The proof of these claims should actually follow from the proofs in [Bray 1997] for the Schwarzschild case; much of the construction relies on the geometry being asymptotically flat and spherically symmetric near infinity, and a main technical theorem which is used in the proof is an inequality in Euclidean space, which carries over to Schwarzschild (as used by Bray) and Reissner–Nordstrom for large radii by perturbation. Since the Penrose inequality

$$m \geq \sqrt{\frac{A(0)}{16\pi}} + \frac{Q^2}{2r_0}$$

in this case also follows from [Huisken and Ilmanen 2001], we omit the technical details.

On the Penrose inequality for asymptotically Schwarzschild AdS spaces. We now show that the analogous mass function $m(V)$ (if it exists) will be nondecreasing in an asymptotically Schwarzschild AdS space. In general, the mass of asymptotically hyperbolic spaces is more subtle than for asymptotically flat spaces; compare [Chruściel and Herzlich 2003; Wang 2001; Zhang 2004]. We are only discussing below a class of asymptotically hyperbolic spaces with a spherical infinity and with special asymptotics.

In the next proposition, we mean by *horizon* that Σ_0 has (inward) mean curvature $H = 2$ [Bray and Chruściel 2004].

Proposition 6.3. *Assume (M, g) is a three-manifold with $R(g) \geq -6$, which outside a compact set is isometric to an exterior of a ball in Schwarzschild AdS space* *of mass $m > 0$. Suppose furthermore that M has only one horizon Σ_0 and admits a connected isoperimetric profile (with respect to Σ_0) $\Sigma(V)$, so that for sufficiently large V, $\Sigma(V)$ is the spherically symmetric sphere in Schwarzschild AdS of volume V. Then*

$$m \geq \sqrt{\frac{A(\Sigma_0)}{16\pi}}.$$

Proof. Schwarzschild AdS is asymptotic to hyperbolic three-space, so the definitions and computations change slightly from above. We begin by putting $R(g) \geq -6$ into inequality (6-1) to obtain

$$A_{V_0}(V_0)^2 A''_{V_0}(V_0) \leq 3A_{V_0}(V_0) + 4\pi - \int_{\Sigma(V_0)} \tfrac{3}{4} H^2 \, dA = 3A_{V_0}(V_0) + 4\pi - \tfrac{3}{4} H^2 A_{V_0}(V_0).$$

Hence

$$A''_{V_0}(V_0) \leq \frac{3}{A_{V_0}(V_0)} + \frac{4\pi}{A_{V_0}(V_0)^2} - \frac{3A'_{V_0}(V_0)^2}{4A_{V_0}(V_0)}.$$

Since by definition $A(V_0) = A_{V_0}(V_0)$ and $A(V) \leq A_{V_0}(V)$, we have the weak inequality

$$A''(V) \leq \frac{3}{A(V)} + \frac{4\pi}{A(V)^2} - \frac{3A'(V)^2}{4A(V)},$$

or equivalently, for $F = A^{3/2}$,

$$F''(V) \leq \frac{27F(V)^{2/3} + 36\pi - F'(V)^2}{6F(V)}. \tag{6-3}$$

We modify the Hawking mass in this setting with one extra term which accounts for the nonminimal horizon, so we get a corresponding $m(V)$ for the isoperimetric surfaces as follows:

$$\begin{aligned} m(V) &= \sqrt{\frac{A(V)}{16\pi}} \left(1 + \frac{A(V)}{4\pi} - \int_{\Sigma(V)} \frac{H^2}{16\pi} \, dA \right) \\ &= \frac{A(V)^{1/2}}{16\pi^{3/2}} \left(4\pi + A(V) - \tfrac{1}{4} A(V)(A'(V))^2 \right) \\ &= \frac{F(V)^{1/3}}{16\pi^{3/2}} \left(4\pi + F(V)^{2/3} - \tfrac{1}{4} F(V)^{2/3} (\tfrac{2}{3} F(V)^{-1/3} F'(V))^2 \right) \\ &= \frac{F(V)^{1/3}}{144\pi^{3/2}} \left(36\pi + 9F(V)^{2/3} - F'(V)^2 \right). \end{aligned}$$

The reason for the modification is that since the (inward) mean curvature of Σ_0 is 2, we again have $m(0) = \sqrt{A(0)/(16\pi)}$.

Since (6-3) holds (and again, $F(V)$ is nondecreasing since there is only one horizon), we have the distributional inequality

$$m'(V) = \frac{2}{144\pi^{3/2}} F(V)^{1/3} F'(V) \left(-F''(V) + \frac{36\pi + 27F(V)^{2/3} - F'(V)^2}{6F(V)} \right) \geq 0.$$

For V sufficiently large, $m(V)$ is the Hawking mass of some radially symmetric sphere $S_r = \Sigma(V)$ and thus

$$m(V) = \sqrt{\frac{4\pi r^2}{16\pi}} \left(1 + \frac{4\pi r^2}{4\pi} - \frac{4\pi r^2}{16\pi} \frac{4(1 + r^2 - 2m/r)}{r^2} \right) = m.$$

Hence

$$m = \lim_{V \to +\infty} m(V) \geq m(0) = \sqrt{\frac{A(0)}{16\pi}},$$

giving us the desired Penrose Inequality. □

7. Conclusions

We conjecture that there exists a reasonable class of spaces with $R(g) \geq -6$ which are asymptotically Schwarzschild AdS for which the above analysis will yield a Penrose Inequality. We hope to report on this in a future work. Although the class would be limited in several respects, it is interesting problem, following the work of Bray and in light of the recent work of Huisken [2005; 2006], to understand better the relationship of the mass to the isoperimetric properties of the space.

We mention that foliations near infinity of constant mean curvature (CMC) have appeared in the context of relativity; see [Huisken and Yau 1996; Metzger 2004; Qing and Tian 2004; Ye 1996]. It is tempting to conjecture that these uniquely determined foliations near infinity by constant mean curvature spheres give the isoperimetric profiles.

Appendix: Metric formulas

Consider a metric of the form

$$g = f(r)\, dr^2 + r^2\, d\Omega^2 = f(r)\, dr^2 + r^2\, d\phi^2 + r^2 \sin^2(\phi)\, d\theta^2,$$

with $f(r) > 0$. We collect here the basic geometric formulas which we apply to our three families of metrics above. We use the Einstein summation convention below, and the indices (1, 2, 3) correspond to the variables (r, ϕ, θ).

Christoffel symbols. We display the metric and its inverse in matrix form:

$$(g_{ij}) = \begin{pmatrix} f(r) & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\phi \end{pmatrix}, \quad (g^{ij}) = \begin{pmatrix} \frac{1}{f(r)} & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{1}{r^2\sin^2\phi} \end{pmatrix}.$$

We recall the formula $\Gamma^k_{ij} = \frac{1}{2}g^{mk}(g_{jm,i}+g_{mi,j}-g_{ij,m})$ for the Christoffel symbols. We can simplify our calculations by making two observations. Since g_{ij} and g^{ij} are diagonal, we have $\Gamma^k_{ij} = \frac{1}{2}g^{kk}(g_{jk,i}+g_{ki,j}-g_{ij,k})$, and $\Gamma^k_{ij}=0$ when i, j, and k are all distinct. For reference here are the nonzero Christoffel symbols:

$$\begin{aligned}
\Gamma^1_{11} &= \tfrac{1}{2}g^{11}(g_{11,1}+g_{11,1}-g_{11,1}) = \frac{f'(r)}{2f(r)},\\
\Gamma^1_{22} &= \tfrac{1}{2}g^{11}(g_{21,2}+g_{12,2}-g_{22,1}) = -\frac{r}{f(r)},\\
\Gamma^1_{33} &= \tfrac{1}{2}g^{11}(g_{31,3}+g_{13,3}-g_{33,1}) = -\frac{r\sin^2\phi}{f(r)},\\
\Gamma^2_{33} &= \tfrac{1}{2}g^{22}(g_{32,3}+g_{23,3}-g_{33,2}) = -\sin\phi\cos\phi,\\
\Gamma^2_{12} = \Gamma^2_{21} &= \tfrac{1}{2}g^{22}(g_{22,1}+g_{21,2}-g_{12,2}) = \frac{1}{r},\\
\Gamma^3_{13} = \Gamma^3_{31} &= \tfrac{1}{2}g^{33}(g_{33,1}+g_{31,3}-g_{13,3}) = \frac{1}{r},\\
\Gamma^3_{23} = \Gamma^3_{32} &= \tfrac{1}{2}g^{33}(g_{33,2}+g_{32,3}-g_{23,3}) = \cot\phi.
\end{aligned}$$

Second fundamental form and mean curvature of radial spheres S_r. We compute the second fundamental form and mean curvature of the coordinate spheres of constant r. Let Z^N be the normal projection of a vector Z. We have

$$\begin{aligned}
B(\partial_\phi,\partial_\phi) &= (\nabla_{\partial_\phi}\partial_\phi)^N = (\Gamma^1_{22}\partial_r+\Gamma^2_{22}\partial_\phi+\Gamma^3_{22}\partial_\theta)^N = -\frac{r}{f(r)}\partial_r,\\
B(\partial_\phi,\partial_\theta) &= (\nabla_{\partial_\phi}\partial_\theta)^N = (\Gamma^1_{23}\partial_r+\Gamma^2_{23}\partial_\phi+\Gamma^3_{23}\partial_\theta)^N = 0,\\
B(\partial_\theta,\partial_\theta) &= (\nabla_{\partial_\theta}\partial_\theta)^N = (\Gamma^1_{33}\partial_r+\Gamma^2_{33}\partial_\phi+\Gamma^3_{33}\partial_\theta)^N = -\frac{r\sin^2\phi}{f(r)}\partial_r.
\end{aligned}$$

Let $N=-\nu$ denote the inward unit normal vector field to S_r. Then $g(\partial_r,N) = -\|\partial_r\| = -\sqrt{f(r)}$, so the second fundamental form II, defined by $\mathrm{II}(V,W) = g(B(V,W),N)$, is given by

$$\mathrm{II}(\partial_\phi,\partial_\phi) = \frac{r}{\sqrt{f(r)}}, \qquad \mathrm{II}(\partial_\phi,\partial_\theta)=0, \qquad \mathrm{II}(\partial_\theta,\partial_\theta) = \frac{r\sin^2(\phi)}{\sqrt{f(r)}}.$$

Thus the mean curvature for S_r, which is constant by symmetry, is

$$H_{S_r} = g^{\phi\phi}\,\mathrm{II}(\partial_\phi, \partial_\phi) + g^{\theta\theta}\,\mathrm{II}(\partial_\theta, \partial_\theta) = \left(\frac{r/\sqrt{f(r)}}{r^2} + \frac{(r\sin^2\phi)/\sqrt{f(r)}}{r^2\sin^2\phi}\right)$$
$$= \frac{2}{r\sqrt{f(r)}}.$$

Ricci and scalar curvature. We use the formulas

$$R_{ij} = R^l_{ilj} \quad \text{and} R^l_{ikj} = \Gamma^l_{ij,k} - \Gamma^l_{ik,j} + \Gamma^m_{ij}\Gamma^l_{km} - \Gamma^m_{ik}\Gamma^l_{jm}.$$

A simple computation shows that the Ricci tensor is diagonal in this coordinate system, and the diagonal entries are given by

$$\begin{aligned}
R_{11} &= R^1_{111} + R^2_{121} + R^3_{131} = 0 + \frac{f'(r)}{2rf(r)} + \frac{f'(r)}{2rf(r)},\\
R_{22} &= R^1_{212} + R^2_{222} + R^3_{232} = \frac{rf'(r)}{2\,f(r)^2} + 0 + \left(1 - \frac{1}{f(r)}\right),\\
R_{33} &= R^1_{313} + R^2_{323} + R^3_{333} = \sin^2(\phi)\left(\frac{rf'(r)}{2\,f(r)^2} + \left(1 - \frac{1}{f(r)}\right) + 0\right).
\end{aligned}$$

Thus we find that the scalar curvature is

$$R(g) = g^{ij}R_{ij} = g^{11}R_{11} + g^{22}R_{22} + g^{33}R_{33} = \frac{2f'(r)}{r\,f(r)^2} + \frac{2}{r^2} - \frac{2}{r^2 f(r)}.$$

Acknowledgements

We thank Hubert Bray for several conversations about, and his encouragement for, this project. We also thank the National Science Foundation and Lafayette College for sponsoring the REU program in the Summer of 2005, at which much of the work for this project was completed. Corvino thanks Frank Morgan for advice on, and inspiration for, running an undergraduate research program in differential geometry.

References

[Axler et al. 1992] S. Axler, P. Bourdon, and W. Ramey, *Harmonic function theory*, Graduate Texts in Mathematics **137**, Springer, New York, 1992. MR 93f:31001 Zbl 0765.31001

[Bray 1997] H. L. Bray, *The Penrose inequality in general relativity and volume comparison theorems involving scalar curvature*, Ph.D. thesis, Stanford University, 1997.

[Bray 2001] H. L. Bray, "Proof of the Riemannian Penrose inequality using the positive mass theorem", *J. Differential Geom.* **59**:2 (2001), 177–267. MR 2004j:53046 Zbl 1039.53034

[Bray 2002] H. L. Bray, "Black holes, geometric flows, and the Penrose inequality in general relativity", *Notices Amer. Math. Soc.* **49**:11 (2002), 1372–1381. MR 2003j:83052 Zbl 02115028

[Bray and Chruściel 2004] H. L. Bray and P. T. Chruściel, "The Penrose inequality", pp. 39–70 in *The Einstein equations and the large scale behavior of gravitational fields: 50 years of the Cauchy problem in general relativity*, edited by P. T. Chruściel and H. Friedrich, Birkhäuser, Basel, 2004. MR 2005m:83014 Zbl 1058.83006

[Bray and Morgan 2002] H. Bray and F. Morgan, "An isoperimetric comparison theorem for Schwarzschild space and other manifolds", *Proc. Amer. Math. Soc.* **130**:5 (2002), 1467–1472. MR 2002i:53073 Zbl 0994.53029

[Chavel 1993] I. Chavel, *Riemannian geometry: a modern introduction*, Cambridge Tracts in Mathematics **108**, Cambridge University Press, Cambridge, 1993. MR 95j:53001 Zbl 0810.53001

[Chruściel and Herzlich 2003] P. T. Chruściel and M. Herzlich, "The mass of asymptotically hyperbolic Riemannian manifolds", *Pacific J. Math.* **212**:2 (2003), 231–264. MR 2005d:53052

[Corvino $\geq$ 2007] J. Corvino, "On the asymptotics of the Einstein–Maxwell constraint equations". In preparation.

[Geroch 1973] R. Geroch, "Energy extraction", *Ann. New York Acad. Sci.* **224** (1973), 108–117. Zbl 0942.53509

[Howards et al. 1999] H. Howards, M. Hutchings, and F. Morgan, "The isoperimetric problem on surfaces", *Amer. Math. Monthly* **106**:5 (1999), 430–439. MR 2000i:52027 Zbl 1003.52011

[Huisken 2005] G. Huisken, "Radial foliations of asymptotically flat 3-manifolds", Lecture given at the Einstein Constraint Equations Conference, Isaac Newton Institute, 2005, Available at http://www.newton.cam.ac.uk/webseminars/pg+ws/2005/gmr/gmrw03/1212/huisken.

[Huisken 2006] G. Huisken, "An isoperimetric concept for mass and quasilocal mass", *Oberwolfach Rep.* **3**:1 (2006), 87–88.

[Huisken and Ilmanen 2001] G. Huisken and T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose inequality", *J. Differential Geom.* **59**:3 (2001), 353–437. MR 2003h:53091 Zbl 1055.53052

[Huisken and Yau 1996] G. Huisken and S.-T. Yau, "Definition of center of mass for isolated physical systems and unique foliations by stable spheres with constant mean curvature", *Invent. Math.* **124**:1-3 (1996), 281–311. MR 96m:53037 Zbl 0858.53071

[Metzger 2004] J. Metzger, "Foliations of asymptotically flat three-manifolds with two-surfaces of prescribed mean curvature", Preprint, 2004. math.DG/0410413

[Qing and Tian 2004] J. Qing and G. Tian, "On the uniqueness of the foliation of spheres of constant mean curvature in asymptotically flat three-manifolds", Preprint, 2004. math.DG/0506005

[Taylor 1996] M. E. Taylor, *Partial differential equations, III: Nonlinear equations*, Applied Mathematical Sciences **116**, Springer, New York, 1996. MR 98b:35003 Zbl 0869.35001

[Wald 1984] R. M. Wald, *General relativity*, Univ. of Chicago Press, Chicago, 1984. MR 86a:83001 Zbl 0549.53001

[Wang 2001] X. Wang, "The mass of asymptotically hyperbolic manifolds", *J. Differential Geom.* **57**:2 (2001), 273–299. MR 2003c:53044 Zbl 1037.53017

[Weinstein and Yamada 2005] G. Weinstein and S. Yamada, "On a Penrose inequality with charge", *Comm. Math. Phys.* **257**:3 (2005), 703–723. MR 2007c:83016 Zbl 1078.83007

[Ye 1996] R. Ye, "Foliation by constant mean curvature spheres on asymptotically flat manifolds", pp. 369–383 in *Geometric analysis and the calculus of variations*, edited by J. Jost, Int. Press, Cambridge, MA, 1996. MR 98e:53040 Zbl 0932.53026

[Zhang 2004] X. Zhang, "A definition of total energy-momenta and the positive mass theorem on asymptotically hyperbolic 3-manifolds, I", *Comm. Math. Phys.* **249**:3 (2004), 529–548. MR 2006a:83012 Zbl 1073.83019

Received January 3, 2006. Revised March 2, 2007.

JUSTIN CORVINO
DEPARTMENT OF MATHEMATICS
LAFAYETTE COLLEGE
EASTON, PA 18042
UNITED STATES
corvinoj@lafayette.edu

AYDIN GEREK
DEPARTMENT OF MATHEMATICS
LAFAYETTE COLLEGE
EASTON, PA 18042
UNITED STATES
gereka@lafayette.edu

MICHAEL GREENBERG
DEPARTMENT OF MATHEMATICS
BROWN UNIVERSITY
BOX 1917
PROVIDENCE, RI 02912
UNITED STATES
mikeyg@math.brown.edu

BRIAN KRUMMEL
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CA 94305
UNITED STATES
bkrummel@math.stanford.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

IRREDUCIBLE REPRESENTATIONS FOR THE ABELIAN EXTENSION OF THE LIE ALGEBRA OF DIFFEOMORPHISMS OF TORI IN DIMENSIONS GREATER THAN 1

CUIPO JIANG AND QIFEN JIANG

We classify the irreducible weight modules of the abelian extension of the Lie algebra of diffeomorphisms of tori of dimension greater than 1, with finite-dimensional weight spaces.

1. Introduction

Let $W_{\nu+1}$ be the Lie algebra of diffeomorphisms of the $(\nu+1)$-dimensional torus. If $\nu = 0$, the universal central extension of the complex Lie algebra W_1 is the Virasoro algebra, which, together with its representations, plays a very important role in many areas of mathematics and physics [Belavin et al. 1984; Dotsenko and Fateev 1984; Di Francesco et al. 1997]. The representation theory of the Virasoro algebra has been studied extensively; see, for example, [Kac 1982; Kaplansky and Santharoubane 1985; Chari and Pressley 1988; Mathieu 1992].

If $\nu \geq 1$, however, the Lie algebra $W_{\nu+1}$ has no nontrivial central extension [Ramos et al. 1990]. But $W_{\nu+1}$ has abelian extensions whose abelian ideals are the central parts of the corresponding toroidal Lie algebras; see [Berman and Billig 1999], for example. There is a close connection between irreducible integrable modules of the toroidal Lie algebra and irreducible modules of the abelian extension $\mathscr{L}$; see [Berman and Billig 1999; Eswara Rao and Moody 1994; Jiang and Meng 2003], for instance. In fact, the classification of integrable modules of toroidal Lie algebras and their subalgebras depends heavily on the classification of irreducible representations of $\mathscr{L}$ and its subalgebras. See [Billig 2003] for the constructions of the abelian extensions for the group of diffeomorphisms of a torus.

In this paper we study the irreducible weight modules of $\mathscr{L}$, for $\nu \geq 1$. If V is an irreducible weight module of $\mathscr{L}$ some of whose central charges $c_0, \dots, c_\nu$ are nonzero, one can assume that $c_0, \dots, c_N$ are $\mathbb{Z}$-linearly independent and $c_{N+1} = \cdots = c_\nu = 0$, where $N \geq 0$. We prove that if $N \geq 1$, then V must have weight

MSC2000: primary 17B67, 17B65; secondary 17B68.
Keywords: irreducible representation, abelian extension, central charge.
Work supported in part by NSF of China, grants No. 10271076 and No. 10571119.

spaces which are infinite-dimensional. So if all the weight spaces of V are finite-dimensional, N vanishes. We classify the irreducible modules of $\mathcal{L}$ with finite-dimensional weight spaces and some nonzero central charges. We prove that such a module V is isomorphic to a highest weight module. The highest weight space T is isomorphic to an irreducible $(\mathcal{A}_\nu + W_\nu)$-module all of whose weight spaces have the same dimension, where $\mathcal{A}_\nu$ is the ring of Laurent polynomials in ν commuting variables, regarded as a commutative Lie algebra. An important step is to characterize the $\mathcal{A}_\nu$-module structure of T. It turns out that the action of $\mathcal{A}_\nu$ on T is essentially multiplication by polynomials in $\mathcal{A}_\nu$. Therefore T can be identified with Larsson's construction [1992] by a result in [Eswara Rao 2004]. That is, T is a tensor product of gl_ν-module with $\mathcal{A}_\nu$.

When all the central charges of V are zero, we prove that the abelian part acts on V as zero if V is a uniformly bounded $\mathcal{L}$-module. So the result in this case is not complete.

Throughout the paper, $\mathbb{C}$, $\mathbb{Z}_+$ and $\mathbb{Z}_-$ denote the sets of complex numbers, positive integers and negative integers.

2. Basic concepts and results

Let $\mathcal{A}_{\nu+1} = \mathbb{C}[t_0^{\pm 1}, t_1^{\pm 1}, \dots, t_\nu^{\pm 1}]$ $(\nu \geq 1)$ be the ring of Laurent polynomials in commuting variables $t_0, t_1, \dots, t_\nu$. For $\underline{n} = (n_1, n_2, \dots, n_\nu) \in \mathbb{Z}^\nu$, $n_0 \in \mathbb{Z}$, we denote $t_0^{n_0} t_1^{n_1} \cdots t_\nu^{n_\nu}$ by $t_0^{n_0} t^{\underline{n}}$. Let $\tilde{\mathcal{K}}$ be the free $\mathcal{A}_{\nu+1}$-module with basis $\{k_0, k_1, \dots, k_\nu\}$ and let $d\tilde{\mathcal{K}}$ be the subspace spanned by all elements of the form

$$\sum_{i=0}^{\nu} r_i t_0^{r_0} t^{\underline{r}} k_i, \quad \text{for } (r_0, \underline{r}) = (r_0, r_1, \dots, r_\nu) \in \mathbb{Z}^{\nu+1}.$$

Set $\mathcal{K} = \tilde{\mathcal{K}}/d\tilde{\mathcal{K}}$ and denote the image of $t_0^{r_0} t^{\underline{r}} k_i$ still by itself. Then $\mathcal{K}$ is spanned by the elements $\{t_0^{r_0} t^{\underline{r}} k_p \mid p = 0, 1, \dots, \nu, r_0 \in \mathbb{Z}, \underline{r} \in \mathbb{Z}^\nu\}$ with relations

$$\sum_{p=0}^{\nu} r_p t_0^{r_0} t^{\underline{r}} k_p = 0. \tag{2-1}$$

Let $\mathcal{D}$ be the Lie algebra of derivations on $\mathcal{A}_{\nu+1}$. Then

$$\mathcal{D} = \left\{ \sum_{p=0}^{\nu} f_p(t_0, t_1, \dots, t_\nu) d_p \mid f_p(t_0, t_1, \dots, t_\nu) \in \mathcal{A}_{\nu+1} \right\},$$

where $d_p = t_p \partial / \partial t_p$, $p = 0, 1, \dots, \nu$. From [Berman and Billig 1999] we know that the algebra $\mathcal{D}$ admits two nontrivial 2-cocycles with values in $\mathcal{K}$:

$$\tau_1(t_0^{m_0} t^{\underline{m}} d_a, t_0^{n_0} t^{\underline{n}} d_b) = -n_a m_b \sum_{p=0}^{\nu} m_p t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} k_p,$$

$$\tau_2(t_0^{m_0}t^{\underline{m}}d_a,\, t_0^{n_0}t^{\underline{n}}d_b) = m_a n_b \sum_{p=0}^{\nu} m_p t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} k_p.$$

Let $\tau = \mu_1\tau_1 + \mu_2\tau_2$ be an arbitrary linear combination of τ_1 and τ_2. Then the corresponding abelian extension of $\mathcal{D}$ is

$$\mathcal{L} = \mathcal{D} \oplus \mathcal{K},$$

with the Lie bracket

$$\begin{aligned}
(2\text{-}2)\quad [t_0^{m_0}t^{\underline{m}}d_a,\, t_0^{n_0}t^{\underline{n}}k_b] &= n_a t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} k_b + \delta_{ab}\sum_{p=0}^{\nu} m_p t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} k_p,\\
[t_0^{m_0}t^{\underline{m}}d_a,\, t_0^{n_0}t^{\underline{n}}d_b] &= n_a t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} d_b - m_b t_0^{m_0+n_0} t^{\underline{m}+\underline{n}} d_a\\
&\qquad + \tau(t_0^{m_0}t^{\underline{m}}d_a,\, t_0^{n_0}t^{\underline{n}}d_b).
\end{aligned}$$

The sum

$$\mathfrak{h} = \Big(\bigoplus_{i=0}^{\nu} \mathbb{C}k_i\Big) \oplus \Big(\bigoplus_{i=0}^{\nu} \mathbb{C}d_i\Big)$$

is an abelian Lie subalgebra of $\mathcal{L}$. An $\mathcal{L}$-module V is called a weight module if

$$V = \bigoplus_{\lambda\in\mathfrak{h}^*} V_\lambda,$$

where $V_\lambda = \{v \in V \mid h\cdot v = \lambda(h)v \text{ for all } h \in \mathfrak{h}\}$. Denote by $P(V)$ the set of all weights. Throughout the paper, we assume that V is an irreducible weight module of $\mathcal{L}$ with finite-dimensional weight spaces. Since V is irreducible, we have

$$k_i|_V = c_i,$$

where the constants c_i, for $i = 0, 1, \ldots, \nu$, are called the central charges of V.

Lemma 2.1. *Let $A = (a_{ij})$ $(0 \le i, j \le \nu)$ be a $(\nu+1)\times(\nu+1)$ matrix such that* $\det A = 1$ *and $a_{ij} \in \mathbb{Z}$. There exists an automorphism σ of $\mathcal{L}$ such that*

$$\sigma(t^{\bar m}k_j) = \sum_{p=0}^{\nu} a_{pj} t^{\bar m A^T} k_p, \quad \sigma(t^{\bar m}d_j) = \sum_{p=0}^{\nu} b_{jp} t^{\bar m A^T} d_p, \quad 0 \le j \le \nu,$$

where $t^{\bar m} = t_0^{m_0}t^{\underline{m}}$, $B = (b_{ij}) = A^{-1}$.

3. The structure of V with nonzero central charges

In this section, we discuss the weight module V which has nonzero central charges. It follows from Lemma 2.1 that we can assume that $c_0, c_1, \ldots, c_N$ are $\mathbb{Z}$-linearly independent, i.e., if $\sum_{i=0}^{N} a_i c_i = 0$, $a_i \in \mathbb{Z}$, then all $a_i (i = 0, \ldots, N)$ must be zero, and $c_{N+1} = c_{N+2} = \cdots = c_\nu = 0$, where $N \geq 0$. For $\bar{m} = (m_0, \underline{m})$, denote $t_0^{m_0} t^{\underline{m}}$ by $t^{\bar{m}}$ as in Lemma 2.1. It is easy to see that V has the decomposition

$$V = \bigoplus_{\bar{m} \in \mathbb{Z}^{\nu+1}} V_{\bar{m}},$$

where $V_{\bar{m}} = \{v \in V \mid d_i(v) = (\gamma_0(d_i) + m_i)v, i = 0, 1, \dots, \nu\}$, with $\gamma_0 \in P(V)$ a fixed weight, and $\bar{m} = (m_0, m_1, \dots, m_\nu) \in \mathbb{Z}^{\nu+1}$. If V has finite-dimensional weight spaces, the $V_{\bar{m}}$ are finite-dimensional, for $\bar{m} \in \mathbb{Z}^{\nu+1}$.

In Lemmas 3.1–3.6 we assume that V has finite-dimensional weight spaces.

Lemma 3.1. *For $p \in \{0, 1, \dots, \nu\}$ and $0 \neq t^{\bar{m}} k_p \in \mathcal{L}$, if there is a nonzero element v in V such that $t^{\bar{m}} k_p v = 0$, then $t^{\bar{m}} k_p$ is locally nilpotent on V.*

Lemma 3.2. *Let $t_0^{m_0} t^{\underline{m}} k_p \in \mathcal{L}$ be such that $\bar{m} = (m_0, \underline{m}) \neq \bar{0}$, and there exists $0 \leq a \leq N$ such that $m_a \neq 0$ if $N < p \leq \nu$. If $t_0^{m_0} t^{\underline{m}} k_p$ is locally nilpotent on V, then* $\dim V_{\bar{n}} > \dim V_{\bar{n}+\bar{m}}$ *for all $\bar{n} \in \mathbb{Z}^{\nu+1}$.*

Proof. Case 1: $p \in \{0, 1, \dots, N\}$. We first prove that $\dim V_{\bar{n}} \geq \dim V_{\bar{n}+\bar{m}}$ for all $\bar{n} \in \mathbb{Z}^{\nu+1}$. Suppose $\dim V_{\bar{n}} = m$, $\dim V_{\bar{n}+\bar{m}} = n$. Let $\{w_1, w_2, \dots, w_n\}$ be a basis of $V_{\bar{n}+\bar{m}}$ and $\{w'_1, w'_2, \dots, w'_m\}$ a basis of $V_{\bar{n}}$. We can assume that $m_a \neq 0$ for some $0 \leq a \leq \nu$ distinct from p, where $\bar{m} = (m_0, \underline{m}) = (m_0, m_1, \dots, m_\nu)$. Since $t^{\bar{m}} k_p$ is locally nilpotent on V and $V_{\bar{n}+\bar{m}}$ is finite-dimensional, there exists $k > 0$ such that $(t^{\bar{m}} k_p)^k V_{\bar{n}+\bar{m}} = 0$. Therefore

$$(t^{-\bar{m}} d_a)^k (t^{\bar{m}} k_p)^k (w_1, w_2, \dots, w_n) = 0.$$

On the other hand, by induction on k, we can deduce that

$$(t^{-\bar{m}} d_a)^k (t^{\bar{m}} k_p)^k = \sum_{i=0}^{k} \frac{k!\,k!}{i!\,(k-i)!\,(k-i)!} m_a^i c_p^i (t^{\bar{m}} k_p)^{k-i} (t^{-\bar{m}} d_a)^{k-i}.$$

Therefore

$$t^{\bar{m}} k_p \Big(\sum_{i=0}^{k-1} \frac{k!\,k!}{i!\,(k-i)!\,(k-i)!} m_a^i c_p^i (t^{\bar{m}} k_p)^{k-1-i} (t^{-\bar{m}} d_a)^{k-1-i} \Big) t^{-\bar{m}} d_a (w_1, w_2, \dots, w_n) \\ = -k!\, m_a^k c_p^k (w_1, w_2, \dots, w_n).$$

Assume that

$$\Big(\sum_{i=0}^{k-1} \frac{k!\,k!}{i!\,(k-i)!\,(k-i)!} m_a^i c_p^i (t^{\bar{m}} k_p)^{k-1-i} (t^{-\bar{m}} d_a)^{k-1-i} \Big) t^{-\bar{m}} d_a (w_1, w_2, \dots, w_n) \\ = (w'_1, w'_2, \dots, w'_m) C,$$

with $C \in \mathbb{C}^{m \times n}$, and that

$$t^{\bar{m}} k_p (w'_1, w'_2, \dots, w'_m) = (w_1, w_2, \dots, w_n) B, \tag{3-1}$$

with $B \in \mathbb{C}^{n\times m}$. Then

$$BC = -k!\, m_a^k c_p^k I.$$

This implies that $m \geq n$. So $\dim V_{\bar{n}} \geq \dim V_{\bar{n}+\bar{m}}$ for all $\bar{n} \in \mathbb{Z}^{\nu+1}$. Also, by (3-1) and the fact that $r(B) = n$, we know that $m > n$ if and only if there exists $v \in V_{\bar{n}}$ such that $t^{\bar{m}}k_p \cdot v = 0$. Since $t^{\bar{m}}k_p$ is locally nilpotent on V, there exist an integer $s \geq 0$ and $w \in V_{\bar{n}+s\bar{m}}$ such that

$$(t^{\bar{m}}k_p)\cdot w = 0.$$

Therefore $(t^{-\bar{m}}k_p)t^{\bar{m}}k_p \cdot w = t^{\bar{m}}k_p(t^{-\bar{m}}k_p \cdot w) = 0$. If $t^{-\bar{m}}k_p \cdot w = 0$, by the proof above, $\dim V_{\bar{n}+s\bar{m}-\bar{m}} < \dim V_{\bar{n}+s\bar{m}}$, contradicting the fact that $\dim V_{\bar{n}+s\bar{m}-\bar{m}} \geq \dim V_{\bar{n}+s\bar{m}}$. Therefore $(t^{-\bar{m}}k_p)^r \cdot w \neq 0$ for all $r \in \mathbb{N}$. Since

$$(t^{-\bar{m}}k_p)^s t^{\bar{m}}k_p \cdot w = t^{\bar{m}}k_p(t^{-\bar{m}}k_p)^s \cdot w = 0$$

and $(t^{-\bar{m}}k_p)^s \cdot w \in V_{\bar{n}}$, it follows that there is a nonzero element v in $V_{\bar{n}}$ such that $t^{\bar{m}}k_p \cdot v = 0$. Thus $n < m$.

Case 2: $N < p \leq \nu$. The proof is similar to that of case 1, but we have to consider $t^{-\bar{m}}d_p$ and $t^{\bar{m}}k_p$ instead and use the $\mathbb{Z}$-linear independence of $c_1, \ldots, c_N$. □

Lemma 3.3. *Let $0 \neq t^{\bar{m}}k_p \in \mathscr{L}$ and $0 \neq t^{\bar{n}}k_p \in \mathscr{L}$ be such that $(m_0, \ldots, m_N) \neq 0$, $(n_0, \ldots, n_N) \neq 0$ if $N < p \leq \nu$, where $\bar{m} = (m_0, m_1, \ldots, m_\nu)$.*

(1) *If $t^{\bar{m}}k_p$ is locally nilpotent on V, $t^{\bar{m}}k_q$ is locally nilpotent for $q = 0, 1, \ldots, \nu$.*

(2) *If both $0 \neq t^{\bar{m}}k_p$ and $0 \neq t^{\bar{n}}k_p$ are locally nilpotent on V, then $t^{\bar{m}+\bar{n}}k_p$ is locally nilpotent.*

(3) *If $0 \neq t^{\bar{m}+\bar{n}}k_p$ is locally nilpotent on V and $(m_0 + n_0, \ldots, m_N + n_N) \neq 0$ if $N < p \leq \nu$, then $t^{\bar{m}}k_p$ or $t^{\bar{n}}k_p$ is locally nilpotent.*

Lemma 3.4. *For $0 \leq p \leq \nu$, let $0 \neq t^{\bar{m}}k_p \in \mathscr{L}$ be such that $(m_0, \ldots, m_N) \neq 0$, where $\bar{m} = (m_0, m_1, \ldots, m_\nu)$. Then $t^{\bar{m}}k_p$ or $t^{-\bar{m}}k_p$ is locally nilpotent on V.*

Proof. The proof occupies the next few pages. We first deal with the case $0 \leq p \leq N$. Without losing generality, we can take $p = 0$.

Suppose the lemma is false. By Lemma 3.2, for any $\bar{r} \in \mathbb{Z}^{\nu+1}$ we have

$$\dim V_{\bar{r}+\bar{m}} = \dim V_{\bar{r}} = \dim V_{\bar{r}-\bar{m}}, \quad t^{\bar{m}}k_0 V_{\bar{r}} = V_{\bar{r}+\bar{m}}, \quad t^{-\bar{m}}k_0 V_{\bar{r}} = V_{\bar{r}-\bar{m}}.$$

Fix $\bar{r} = (r_0, \underline{r}) \in \mathbb{Z}^{\nu+1}$ such that $V_{\bar{r}} \neq 0$. Let $\{v_1, \ldots, v_n\}$ be a basis of $V_{\bar{r}}$ and set

$$v_i(k\bar{m}) = \frac{1}{c_0} t^{k\bar{m}}k_0 \cdot v_i, \quad i = 1, 2, \ldots, n,$$

where $k \in \mathbb{Z} \setminus \{0\}$. Then $\{v_1(k\bar{m}), v_2(k\bar{m}), \ldots, v_n(k\bar{m})\}$ is a basis of $V_{\bar{r}+k\bar{m}}$. Let $B^{(0)}_{-\bar{m},\bar{m}}, B^{(0)}_{\bar{m},-\bar{m}} \in \mathbb{C}^{n\times n}$ be such that

$$\frac{1}{c_0} t^{\bar{m}} k_0(v_1(-\bar{m}), v_2(-\bar{m}), \ldots, v_n(-\bar{m})) = (v_1, v_2, \ldots, v_n) B^{(0)}_{\bar{m},-\bar{m}},$$
$$\frac{1}{c_0} t^{-\bar{m}} k_0(v_1(\bar{m}), v_2(\bar{m}), \ldots, v_n(\bar{m})) = (v_1, v_2, \ldots, v_n) B^{(0)}_{-\bar{m},\bar{m}}.$$

Since $t^{\bar{m}}k_0$ and $t^{-\bar{m}}k_0$ are commutative, it is easy to deduce that

$$B^{(0)}_{\bar{m},-\bar{m}} = B^{(0)}_{-\bar{m},\bar{m}}.$$

By Lemma 3.1, $B^{(0)}_{\bar{m},-\bar{m}}$ is an $n \times n$ invertible matrix.

Claim. $B^{(0)}_{\bar{m},-\bar{m}}$ *does not have distinct eigenvalues.*

Proof. Set $c = 1/c_0$. To prove the claim, we need to consider $ct^{\bar{m}}k_0ct^{-\bar{m}}k_0 - \lambda\,\mathrm{id}$, where $\lambda \in \mathbb{C}^*$. As in the proof of Lemma 3.1, we can deduce that if there is a nonzero element v in V such that $(ct^{\bar{m}}k_0ct^{-\bar{m}}k_0 - \lambda\,\mathrm{id})v = 0$, then $ct^{\bar{m}}k_0ct^{-\bar{m}}k_0 - \lambda\,\mathrm{id}$ is locally nilpotent on V. On the other hand, we have

$$(ct^{\bar{m}}k_0ct^{-\bar{m}}k_0 - \lambda\,\mathrm{id})^l(v_1, v_2, \ldots, v_n) = (v_1, v_2, \ldots, v_n)(B^{(0)}_{\bar{m},-\bar{m}} - \lambda\,\mathrm{id})^l.$$

Therefore the claim holds. ☐

For $p \in \{1, 2, \ldots, \nu\}$, let $C^p_{\bar{m},\bar{0}}, C^p_{\bar{m},-\bar{m}} \in \mathbb{C}^{n\times n}$ be such that

$$t^{\bar{m}}k_p(v_1, v_2, \ldots, v_n) = (v_1(\bar{m}), \ldots, v_n(\bar{m}))C^{(p)}_{\bar{m},\bar{0}},$$
$$t^{\bar{m}}k_p(v_1(-\bar{m}), \ldots, v_n(-\bar{m})) = (v_1, v_2, \ldots, v_n)C^{(p)}_{\bar{m},-\bar{m}}.$$

Since

$$\frac{1}{c_0} t^{-\bar{m}}k_0 t^{\bar{m}}k_p(v_1, v_2, \ldots, v_n) = t^{\bar{m}}k_p \frac{1}{c_0} t^{-\bar{m}}k_0(v_1, v_2, \ldots, v_n),$$

we have

$$C^{(p)}_{\bar{m},-\bar{m}} = B^{(0)}_{-\bar{m},\bar{m}} C^{(p)}_{\bar{m},\bar{0}}. \tag{3-2}$$

Furthermore, by the fact that

$$\frac{1}{c_0} t^{\bar{m}}k_0 \frac{1}{c_0} t^{-\bar{m}}k_0 t^{\bar{m}}k_p(v_1, v_2, \ldots, v_n) = t^{\bar{m}}k_p \frac{1}{c_0} t^{\bar{m}}k_0 \frac{1}{c_0} t^{-\bar{m}}k_0(v_1, v_2, \ldots, v_n)$$

and

$$t^{\bar{m}}k_q \frac{1}{c_0} t^{-\bar{m}}k_0 t^{\bar{m}}k_p = t^{\bar{m}}k_p \frac{1}{c_0} t^{-\bar{m}}k_0 t^{\bar{m}}k_q,$$

we deduce that

$$B^{(0)}_{-\bar{m},\bar{m}}C^{(p)}_{\bar{m},\bar{0}} = C^{(p)}_{\bar{m},\bar{0}}B^{(0)}_{-\bar{m},\bar{m}}, \qquad C^{(p)}_{\bar{m},\bar{0}}C^{(q)}_{\bar{m},\bar{0}} = C^{(q)}_{\bar{m},\bar{0}}C^{(p)}_{\bar{m},\bar{0}}, \qquad 1 \le p, q \le \nu. \tag{3-3}$$

Hence there exists $D \in \mathbb{C}^{n\times n}$ such that $\{D^{-1}B^{(0)}_{-\bar{m},\bar{m}}D,\ D^{-1}C^{(p)}_{\bar{m},\bar{0}}D \mid 1 \le p \le \nu\}$ are all upper triangular matrices. If we set

$$(w_1, w_2, \dots, w_n) = (v_1, v_2, \dots, v_n)D$$

and

$$w_i(k\bar{m}) = \frac{1}{c_0}t^{k\bar{m}}k_0 w_i,\ 1 \le i \le n,\ k \in \mathbb{Z}\setminus\{0\},$$

then

$$\frac{1}{c_0}t^{k\bar{m}}k_0(w_1(-\bar{m}), w_2(-\bar{m}), \dots, w_n(-\bar{m})) = (w_1, \dots, w_n)D^{-1}B^{(0)}_{-\bar{m},\bar{m}}D,$$

$$t^{\bar{m}}k_p(w_1, w_2, \dots, w_n) = (w_1(\bar{m}), \dots, w_n(\bar{m}))D^{-1}C^{(p)}_{\bar{m},\bar{0}}D.$$

So we can assume that $B^{(0)}_{-\bar{m},\bar{m}}$, $C^{(p)}_{\bar{m},\bar{0}}$, and $C^{(p)}_{\bar{m},-\bar{m}}$, for $1 \le p \le \nu$ are all invertible upper triangular matrices. Furthermore, because

$$\Big(t^{\bar{m}}k_p\frac{1}{c_0}t^{-\bar{m}}k_0 - \lambda\,\mathrm{id}\Big)^l(v_1, v_2, \dots, v_n) = (v_1, v_2, \dots, v_n)(C^{(p)}_{\bar{m},-\bar{m}} - \lambda\,\mathrm{id})^l,$$

the argument used in the proof of the claim shows that $C^{(p)}_{\bar{m},-\bar{m}}$ also does not have distinct eigenvalues. For $1 \le p \le N$, set

$$B^{(p)}_{\bar{m},-\bar{m}} = \frac{1}{c_p}C^{(p)}_{\bar{m},-\bar{m}}$$

and for $0 \le p \le N$ denote by λ_p the eigenvalue of $B^{(p)}_{\bar{m},-\bar{m}}$.

Let $A^{(a)}_{k\bar{m},\bar{0}}$ and $A^{(a)}_{k_1\bar{m},k_2\bar{m}}$, for $0 \le a \le \nu$ and $k, k_1, k_2 \in \mathbb{Z}\setminus\{0\}$, be such that

$$\begin{aligned} t^{k\bar{m}}d_a(v_1, v_2, \dots, v_n) &= (v_1(k\bar{m}), v_2(k\bar{m}), \dots, v_n(k\bar{m}))A^{(a)}_{k\bar{m},\bar{0}}, \\ t^{k_1\bar{m}}d_a(v_1(k_2\bar{m}), &v_2(k_2\bar{m}), \dots, v_n(k_2\bar{m})) \\ &= (v_1(k_1\bar{m}+k_2\bar{m}), \dots, v_n(k_1\bar{m}+k_2\bar{m}))A^{(a)}_{k_1\bar{m},k_2\bar{m}}. \end{aligned}$$

Case 1: $\nu > 1$. Since $t^{\bar{m}}k_0 = t_0^{m_0}t^{\underline{m}}k_0 \neq 0$, it follows that there exists $1 \le a \le \nu$ such that $m_a \neq 0$, where $\underline{m} = (m_1, m_2, \dots, m_\nu)$. Let $b \in \{1, \dots, \nu\}$ be such that $a \neq b$. Consider

$$[t^{-\bar{m}}d_a, \frac{1}{c_0}t^{\bar{m}}k_0] = m_a\frac{1}{c_0}k_0, \qquad [t^{-\bar{m}}d_a, t^{\bar{m}}k_b] = m_a k_b. \tag{3-4}$$

Case 1.1: There exists $b \in \{0, 1, \dots, \nu\}$ such that $b \neq 0$, a and $c_b = 0$. Then

$$A^{(a)}_{-\bar{m},\bar{m}} = B^{(0)}_{\bar{m},-\bar{m}}A^{(a)}_{-\bar{m},\bar{0}} + m_a I, \qquad A^{(a)}_{-\bar{m},\bar{m}}C^{(b)}_{\bar{m},\bar{0}} = C^{(b)}_{\bar{m},-\bar{m}}A^{(a)}_{-\bar{m},\bar{0}}.$$

By (3-2) and (3-3),

$$A^{(a)}_{-\bar{m},\bar{0}} + m_a {B^{(0)}_{\bar{m},-\bar{m}}}^{-1} = C^{(b)}_{\bar{m},\bar{0}} A^{(a)}_{-\bar{m},\bar{0}} {C^{(b)}_{\bar{m},\bar{0}}}^{-1}.$$

But the sum on the left-hand side cannot be similar to $A^{(a)}_{-\bar{m},\bar{0}}$, since $m_a \neq 0$ and ${B^{(0)}_{\bar{m},-\bar{m}}}^{-1}$ is an invertible upper triangular matrix and does not have different eigenvalues. Thus this case is excluded.

Case 1.2: $c_b \neq 0$ *for all* $b \in \{0, 1, \dots, \nu\}$, $b \neq 0, a$. By (3-4) and (3-2), we have

$$\begin{aligned} B^{(0)}_{\bar{m},-\bar{m}} A^{(a)}_{-\bar{m},\bar{0}} {B^{(0)}_{\bar{m},-\bar{m}}}^{-1} + m_a {B^{(0)}_{\bar{m},-\bar{m}}}^{-1} - m_a {B^{(b)}_{\bar{m},-\bar{m}}}^{-1} \\ = B^{(0)}_{\bar{m},-\bar{m}} C^{(b)}_{\bar{m},\bar{0}} A^{(a)}_{-\bar{m},\bar{0}} {C^{(b)}_{\bar{m},\bar{0}}}^{-1} {B^{(0)}_{\bar{m},-\bar{m}}}^{-1}. \end{aligned}$$

(I) There exists $b \neq 0$ and a such that $\lambda_0 \neq \lambda_b$. Then $m_a {B^{(0)}_{\bar{m},-\bar{m}}}^{-1} - m_a {B^{(b)}_{\bar{m},-\bar{m}}}^{-1}$ is an invertible upper triangular matrix and does not have different eigenvalues. As in case 1.1, we deduce a contradiction.

(II) $\lambda_0 = \lambda_b$ for all $b \in \{1, \dots, \nu\}$ distinct from a.

(II.1) Suppose first that $c_a = 0$ (in this case $N = \nu - 1$, $a = \nu$) or $c_a \neq 0$ and $\lambda_a = \lambda_0$ (in this case $N = \nu$). Since $\sum_{p=0}^{\nu} m_p t^{\bar{m}} k_p = 0$, we have

$$\sum_{p=0}^{\nu} m_p t^{\bar{m}} k_p \frac{1}{c_0} t^{-\bar{m}} k_0 = 0.$$

So $\sum\limits_{p=0}^{\nu} m_p C^{(p)}_{\bar{m},-\bar{m}} = 0$, and therefore

$$\sum_{p=0}^{\nu} m_p c_p = 0,$$

which contradicts the assumption that $c_0, \dots, c_N$ are $\mathbb{Z}$-linearly independent.

(II.2) Now suppose $c_a \neq 0$, $\lambda_a \neq \lambda_0$ and there exists $b \neq 0$ and a such that $m_b \neq 0$. We deduce a contradiction as in case 1.2(I) by interchanging a by b.

(II.3) Suppose $c_a \neq 0$, $\lambda_a \neq \lambda_0$ and $m_b = 0$ for all $b \in \{1, \dots, \nu\}$ distinct from a. Then $m_0 c_0 \lambda_0 + m_a c_a \lambda_a = 0$. The proof of this case is the same as in case 2.2 below.

Case 2.: $\nu = 1$. In this case $a = 1$.

Case 2.1: $c_a = 0$. Since $[t^{-\bar{m}} d_0, t^{\bar{m}} k_0] = [t^{-\bar{m}} k_0, t^{\bar{m}} d_0] = 0$, we have

$$A^{(0)}_{-\bar{m},\bar{m}} = B^{(0)}_{\bar{m},-\bar{m}} A^{(0)}_{-\bar{m},\bar{0}}, \qquad A^{(0)}_{\bar{m},-\bar{m}} = B^{(0)}_{-\bar{m},\bar{m}} A^{(0)}_{\bar{m},\bar{0}}.$$

Therefore

$$[t^{-\bar{m}} d_0, t^{\bar{m}} d_0](v_1, v_2, \dots, v_n) = (v_1, v_2, \dots, v_n) B^{(0)}_{-\bar{m},\bar{m}} \big[A^{(0)}_{-\bar{m},\bar{0}}, A^{(0)}_{\bar{m},\bar{0}}\big].$$

At the same time, we have

$$[t^{-\bar m}d_0, t^{\bar m}d_0] = 2m_0d_0 + m_0^2(-\mu_1+\mu_2)(m_0k_0+m_1k_1),$$

where $\tau=\mu_1\tau_1+\mu_2\tau_2$ as above. So

$$B^{(0)}_{-\bar m,\bar m}[A^{(0)}_{-\bar m,\bar 0}, A^{(0)}_{\bar m,\bar 0}] = \big(2m_0(\gamma_0(d_0)+r_0)+m_0^2(-\mu_1+\mu_2)(m_0c_0+m_1c_1)\big)I, \tag{3-5}$$

where γ_0 is the weight fixed above. Since γ_0 is arbitrary, we can choose it such that

$$2m_0(\gamma_0(d_0)+r_0)+m_0^2(-\mu_1+\mu_2)(m_0c_0+m_1c_1)\neq 0.$$

But $B^{(0)}_{-\bar m,\bar m}$ is an invertible triangular matrix and does not have different eigenvalues, in contradiction with (3-5).

Case 2.2: $c_a\neq 0$. Since

$$[t^{-\bar m}d_0, t^{\bar m}k_0] = -m_1k_1,\ [t^{-\bar m}d_1, t^{\bar m}k_0] = m_1k_0 \text{ and}$$
$$[t^{\bar m}d_0, t^{-\bar m}k_0] = m_1k_1,\ [t^{\bar m}d_1, t^{-\bar m}k_0] = -m_1k_0,$$

we have

$$[k_0t^{-\bar m}d_0+k_1t^{-\bar m}d_1, t^{\bar m}k_0] = [k_0t^{\bar m}d_0+k_1t^{\bar m}d_1, t^{-\bar m}k_0] = 0.$$

Therefore

$$k_0A^{(0)}_{-\bar m,\bar m}+k_1A^{(1)}_{-\bar m,\bar m} = B^{(0)}_{\bar m,-\bar m}\big(k_0A^{(0)}_{-\bar m,\bar 0}+k_1A^{(1)}_{-\bar m,\bar 0}\big),$$
$$k_0A^{(0)}_{\bar m,-\bar m}+k_1A^{(1)}_{\bar m,-\bar m} = B^{(0)}_{-\bar m,\bar m}\big(k_0A^{(0)}_{\bar m,\bar 0}+k_1A^{(1)}_{\bar m,\bar 0}\big),$$

and

$$\begin{aligned}&[k_0t^{-\bar m}d_0+k_1t^{-\bar m}d_1, k_0t^{\bar m}d_0+k_1t^{\bar m}d_1](v_1,\dots,v_n)\\ &\qquad = (v_1,\dots,v_n)B^{(0)}_{\bar m,-\bar m}\big[k_0A^{(0)}_{-\bar m,\bar 0}+k_1A^{(1)}_{-\bar m,\bar 0}, k_0A^{(0)}_{\bar m,\bar 0}+k_1A^{(1)}_{\bar m,\bar 0}\big].\end{aligned}$$

At the same time, we have

$$\begin{aligned}&[k_0t^{-\bar m}d_0+k_1t^{-\bar m}d_1, k_0t^{\bar m}d_0+k_1t^{\bar m}d_1]\\ &\qquad = 2(m_0c_0+m_1c_1)(c_0d_0+c_1d_1)-(m_0c_0+m_1c_1)^3(\mu_1-\mu_2)\,\mathrm{id}.\end{aligned}$$

Since c_0 and c_1 are $\mathbb{Z}$-linearly independent, we know that $m_0c_0+m_1c_1\neq 0$. As in case 2.1, we deduce a contradiction.

This concludes the first part of the proof. We next turn to the second major case, $N<p\le \nu$.

If $N\ge 1$ or $N=0$, we have $(m_1,\dots,m_\nu)\neq 0$, and the lemma follows from the first part and Lemma 3.3. Otherwise, let $t^{\bar m}k_p=t_0^{m_0}k_p$. Set $\mathcal{L}_{\underline{0}}=\bigoplus_{m_0\in\mathbb{Z}}\mathbb{C}t_0^{m_0}d_0\oplus\mathbb{C}k_0$ and $W=U(\mathcal{L}_{\underline{0}})v$, where $v\in V_{\bar s}$ is a homogeneous element. Since $c_0\neq 0$, the sets $\{\dim W_{(n_0,0)+\bar s}\mid n_0\in\mathbb{Z}\}$ are not uniformly bounded. But if neither $t_0^{m_0}k_p$

nor $t_0^{-m_0}k_p$ is locally nilpotent, then t_0k_p and $t_0^{-1}k_p$ are not locally nilpotent. So by Lemmas 3.2 and 3.1, $\dim V_{(n_0,0)+\bar{s}} = \dim V_{\bar{s}}$ for all $n_0 \in \mathbb{Z}$, which is impossible since $\dim V_{(n_0,0)+\bar{s}} \geq \dim W_{(n_0,0)+\bar{s}}$. This proves Lemma 3.4 □

For $0 \leq p \leq N$, consider the direct sum

$$\bigoplus_{m_p \in \mathbb{Z}} \mathbb{C}t_p^{m_p}d_p \oplus \mathbb{C}k_p,$$

which is a Virasoro Lie subalgebra of $\mathscr{L}$. Since $c_p \neq 0$, it follows from [Mathieu 1992] that there is a nonzero $v_p \in V_{\bar{r}}$ for some $\bar{r} \in \mathbb{Z}^{\nu+1}$ such that

$$t_p^{m_p}d_pv_p = 0 \quad \text{for all } m_p \in \mathbb{Z}_+ \tag{3-6}$$

or

$$t_p^{m_p}d_pv_p = 0 \quad \text{for all } m_p \in \mathbb{Z}_-. \tag{3-7}$$

Lemma 3.5. *If $v_p \in V_{\bar{r}}$ satisfies* (3-6), *the sets*

$$\{t_p^{m_p}k_q \mid m_p \in \mathbb{Z}_+, q = 0, 1, 2, \ldots, \nu, q \neq p\}$$

are all locally nilpotent on V. Likewise for (3-7), *with $\mathbb{Z}_+$ replaced by $\mathbb{Z}_-$.*

Proof. We only prove the first statement. Suppose it is false; then by Lemma 3.3 t_pk_q is not locally nilpotent on V for some $q \in \{0, 1, \ldots, \nu\}$, $q \neq p$. By Lemma 3.4, $t_p^{-1}k_q$ is locally nilpotent. Therefore there exists $k \in \mathbb{Z}_+$ such that

$$(t_p^{-1}k_q)^{k-1}v_p \neq 0, \quad (t_p^{-1}k_q)^k v_p = 0.$$

So

$$\begin{aligned} t_p^2d_p(t_p^{-1}k_q)^k v_p &= -kt_pk_q(t_p^{-1}k_q)^{k-1}v_p + (t_p^{-1}k_q)^k t_p^2d_pv_p \\ &= -kt_pk_q(t_p^{-1}k_q)^{k-1}v_p = 0. \end{aligned}$$

This implies that t_pk_q is locally nilpotent, a contradiction. □

Lemma 3.6. *If $v_p \in V_{\bar{r}}$ satisfies* (3-6), *the sets*

$$\{t^{\bar{m}}k_p \mid \bar{m} = (m_0, \ldots, m_\nu) \in \mathbb{Z}^{\nu+1}, m_p \in \mathbb{Z}_+\}$$

are all locally nilpotent on V. Likewise for (3-7), *with $\mathbb{Z}_+$ replaced by $\mathbb{Z}_-$.*

Proof. Again we only prove the first statement. Without loss of generality, we assume that $p = 0$. Let $\mathscr{K}'$ be the subspace of $\mathscr{K}$ spanned by elements of $\mathscr{K}$ which are locally nilpotent on V. If $t^{\underline{m}}k_0$, for any $\underline{m} \in \mathbb{Z}^\nu \setminus \{0\}$, is not locally nilpotent on V, the lemma holds thanks to Lemmas 3.3 and 3.5. Suppose $\mathscr{K}' \cap \{t^{\underline{m}}k_0 \mid \underline{m} \in \mathbb{Z}^\nu\} \neq \{0\}$. By Lemmas 3.2, 3.3 and 3.5, if $t^{\underline{m}}k_0 \in \mathscr{K}'$, then $t^{-\underline{m}}k_0 \notin \mathscr{K}'$, and $t_0^{m_0}t^{\underline{m}}k_0 \in \mathscr{K}'$ for all $m_0 > 0$.

<u>Case 1</u>: Suppose $t_0^{m_0}t^{-\underline{m}}k_0 \in \mathscr{K}'$ for any $t^{\underline{m}}k_0 \in \mathscr{K}'$. Then the lemma is proved.

Case 2: Suppose there exists $0 \neq t^{\underline{m}}k_0 \in \mathcal{K}'$ such that $t_0t^{-\underline{m}}k_0 \notin \mathcal{K}'$. Since $\underline{m} = (m_1, \dots, m_\nu) \neq 0$, we can assume that $m_a \neq 0$ for some $a \in \{1, 2, \dots, \nu\}$. Let $V_{\bar{r}_0}$ be such that

$$\dim V_{\bar{r}_0} = \min\{\dim V_{\bar{s}} \mid V_{\bar{s}} \neq 0, \bar{s} \in \mathbb{Z}^{\nu+1}\}.$$

Case 2.1: Assume $t_0^it^{-\underline{m}}k_0 \notin \mathcal{K}'$ for any $i > 0$. Let $l \in \mathbb{Z}_+$ and consider

$$\sum_{i=0}^{l} a_i t_0^{-i} t^{-\underline{m}} k_0 t_0^i t^{-\underline{m}} k_0 v = 0, \tag{3-8}$$

where $v \in V_{\bar{r}_0} \setminus \{0\}$. By Lemma 3.4, $\{t_0^it^{\underline{m}}k_0, t_0^{-i}t^{\underline{m}}k_0 \mid i \in \mathbb{Z}_+\} \subseteq \mathcal{K}'$. So by Lemma 3.2, we have

$$t_0^it^{\underline{m}}k_0V_{\bar{r}_0} = t_0^{-i}t^{\underline{m}}k_0V_{\bar{r}_0} = t_0^it^{\underline{m}}d_pV_{\bar{r}_0} = t_0^{-i}t^{\underline{m}}d_pV_{\bar{r}_0} = 0, i \in \mathbb{Z}_+, 0 \leq p \leq \nu.$$

Let $j \in \{0, 1, \dots, l\}$. From (3-8) we have

$$t_0^{-j}t^{\underline{m}}d_at_0^jt^{\underline{m}}d_a\Big(\sum_{i=0}^{l} a_it_0^{-i}t^{-\underline{m}}k_0t_0^it^{-\underline{m}}k_0\Big)v = 0.$$

Therefore

$$\sum_{i=0}^{l} a_i(-m_a)t_0^{j-i}k_0(-m_a)t_0^{i-j}k_0v = a_jm_a^2c_0^2v = 0.$$

So $a_j = 0$, $j = 0, 1, \dots, l$. This means $\{t_0^{-i}t^{-\underline{m}}k_0t_0^it^{-\underline{m}}k_0)v \mid 0 \leq i \leq l\}$ are linearly independent. Since l can be any positive integer, it follows that $V_{\bar{r}_0-(0,2\underline{m})}$ is infinite-dimensional, a contradiction.

Case 2.2: Assume there exists $l \in \mathbb{Z}_+$ such that

$$t_0^{l-1}t^{-\underline{m}}k_0 \notin \mathcal{K}', \qquad t_0^lt^{-\underline{m}}k_0 \in \mathcal{K}'.$$

(I) Assume that $t_0^lt^{-i\underline{m}}k_0 \in \mathcal{K}'$ for any $i \in \mathbb{Z}_+$. Let $s > 0$ and consider

$$\sum_{i=1}^{s} a_it_0^{-l}t^{i\underline{m}}k_0t^{-i\underline{m}}k_0v = 0.$$

Similar to the proof above, we can deduce that $V_{\bar{r}_0-(l,\underline{0})}$ is infinite-dimensional, in contradiction with the assumption that V has finite-dimensional weight spaces.

(II) Assume there exists $s_1 \in \mathbb{Z}_+$ such that

$$t_0^lt^{-\underline{m}}k_0 \in \mathcal{K}', \quad t_0^lt^{-2\underline{m}}k_0 \in \mathcal{K}', \quad \dots, \quad t_0^lt^{-s_1\underline{m}}k_0 \in \mathcal{K}', \quad t_0^lt^{-(s_1+1)\underline{m}}k_0 \notin \mathcal{K}'.$$

Then there exist $s_2, s_3, \dots, s_k, \dots$ such that $s_i \geq s_1$ for $i = 2, 3, \dots, k, \dots$ and

$$t_0^{il}t^{(-s_1-s_2-\cdots-s_{i-1}-1)\underline{m}}k_0 \in \mathcal{K}',\ t_0^{il}t^{(-s_1-s_2-\cdots-s_{i-1}-2)\underline{m}}k_0 \in \mathcal{K}', \dots,$$

$$t_0^{il} t^{(-s_1-s_2-\cdots-s_{i-1}-s_i)\underline{m}} k_0 \in \mathcal{K}', \;\; t_0^{il} t^{(-s_1-s_2-\cdots-s_{i-1}-s_i-1)\underline{m}} k_0 \notin \mathcal{K}'.$$

Assume that

$$\begin{aligned}\Big(\sum_{i=1}^{s_1} a_i t_0^{-l} t^{i\underline{m}} k_0 t^{-i\underline{m}} k_0 + \sum_{i=1}^{s_2} a_{s_1+i} t_0^{-2l} t^{(s_1+i)\underline{m}} k_0 t_0^{l} t^{-(s_1+i)\underline{m}} k_0 \\ + \sum_{i=1}^{s_3} a_{s_1+s_2+i} t_0^{-3l} t^{(s_1+s_2+i)\underline{m}} k_0 t_0^{2l} t^{-(s_1+s_2+i)\underline{m}} k_0 + \cdots \\ + \sum_{i=1}^{s_k} a_{s_1+\cdots+s_{k-1}+i} t_0^{-kl} t^{(s_1+\cdots+s_{k-1}+i)\underline{m}} k_0 t_0^{(k-1)l} t^{-(s_1+\cdots+s_{k-1}+i)\underline{m}} k_0\Big) v = 0.\end{aligned}$$

Let

$$\begin{aligned}&t^{j\underline{m}} d_a t_0^{l} t^{-j\underline{m}} d_a, && 1 \le j \le s_1,\\ &t_0^{-l} t^{(s_1+j)\underline{m}} d_a t_0^{2l} t^{-(s_1+j)\underline{m}} d_a, && 1 \le j \le s_2,\\ &\ldots,\\ &t_0^{-(k-1)l} t^{(s_1+s_2+\cdots+s_{k-1}+j)\underline{m}} d_a t_0^{kl} t^{-(s_1+s_2+\cdots+s_{k-1}+j)\underline{m}} d_a, && 1 \le j \le s_k\end{aligned}$$

act on the two sides of the above equation respectively. By Lemma 3.4, we deduce that $a_i = 0$, for $i = 1, 2, \ldots, s_1$, and that

$$a_{s_1+\cdots+s_{j-1}+i} = 0 \quad \text{for } i = 1, 2, \ldots, s_j,\; 2 \le j \le k.$$

Since k can be any positive integer, it follows that $V_{\bar{r}_0-(l,\underline{0})}$ is infinite-dimensional, which contradicts our assumption. The lemma is proved. $\square$

Lemmas 3.1 through 3.6 immediately yield the following result.

Theorem 3.7. *Let V be an irreducible weight module of $\mathcal{L}$ such that $c_0, \ldots, c_N$ are $\mathbb{Z}$-linearly independent and $N \ge 1$. Then V has weight spaces that are infinite-dimensional.*

Let

$$\begin{aligned}\mathcal{L}_+ &= \sum_{p=0}^{\nu} t_0\mathbb{C}[t_0, t_1^{\pm1}, \ldots, t_\nu^{\pm1}]k_p \oplus \sum_{p=0}^{\nu} t_0\mathbb{C}[t_0, t_1^{\pm1}, \ldots, t_\nu^{\pm1}]d_p,\\ \mathcal{L}_- &= \sum_{p=0}^{\nu} t_0^{-1}\mathbb{C}[t_0^{-1}, t_1^{\pm1}, \ldots, t_\nu^{\pm1}]k_p \oplus \sum_{p=0}^{\nu} t_0^{-1}\mathbb{C}[t_0^{-1}, t_1^{\pm1}, \ldots, t_\nu^{\pm1}]d_p,\\ \mathcal{L}_0 &= \sum_{p=0}^{\nu} \mathbb{C}[t_1^{\pm1}, \ldots, t_\nu^{\pm1}]k_p \oplus \sum_{p=0}^{\nu} \mathbb{C}[t_1^{\pm1}, \ldots, t_\nu^{\pm1}]d_p.\end{aligned}$$

Then

$$\mathcal{L} = \mathcal{L}_+ \oplus \mathcal{L}_0 \oplus \mathcal{L}_-.$$

Definition 3.8. Let W be a weight module of $\mathscr{L}$. If there is a nonzero vector $v_0 \in W$ such that

$$\mathscr{L}_+ v_0 = 0,\ W = U(\mathscr{L}) v_0,$$

then W is called a highest weight module of $\mathscr{L}$. If there is a nonzero vector $v_0 \in W$ such that

$$\mathscr{L}_- v_0 = 0,\ W = U(\mathscr{L}) v_0,$$

then W is called a lowest weight module of $\mathscr{L}$.

From Lemmas 3.2 and 3.6, we obtain:

Theorem 3.9. *Let V be an irreducible weight module of $\mathscr{L}$ with finite-dimensional weight spaces and with central charges $c_0 \neq 0$, $c_1 = c_2 = \cdots = c_\nu = 0$. Then V is a highest or lowest weight module of $\mathscr{L}$.*

In the remainder of this section we assume that V is an irreducible weight module of $\mathscr{L}$ with finite-dimensional weight spaces and with central charges $c_0 \neq 0$, $c_1 = \cdots = c_\nu = 0$.

Set

$$T = \begin{cases} \{v \in V \mid \mathscr{L}_+ v = 0\} & \text{if} V \text{ is a highest weight module of } \mathscr{L}, \\ \{v \in V \mid \mathscr{L}_- v = 0\} & \text{if} V \text{ is a lowest weight module of } \mathscr{L}. \end{cases}$$

Then T is a $\mathscr{L}_0$-module and

$$V = U(\mathscr{L}_-) T \quad \text{or} \quad V = U(\mathscr{L}_+) T.$$

Since V is an irreducible $\mathscr{L}$-module, T is an irreducible $\mathscr{L}_0$-module. T has the decomposition

$$T = \bigoplus_{\underline{m} \in \mathbb{Z}^\nu} T_{\underline{m}},$$

where $\underline{m} = (m_1, m_2, \ldots, m_\nu)$, $T_{\underline{m}} = \{v \in T \mid d_i v = (m_i + \mu(d_i)) v, 1 \leq i \leq \nu\}$ and μ is a fixed weight of T. As in the proof in [Jiang and Meng 2003; Eswara Rao and Jiang 2005], we can deduce:

Theorem 3.10. (1) *For all $\underline{m}, \underline{n} \in \mathbb{Z}^\nu$, $p = 1, 2, \ldots, \nu$, we have*

$$\dim T_{\underline{m}} = \dim T_{\underline{n}},\ t^{\underline{m}} k_p \cdot T = 0,$$

$$t^{\underline{m}} k_0 (v_1(\underline{n}), \ldots, v_m(\underline{n})) = c_0 (v_1(\underline{m}+\underline{n}), v_2(\underline{m}+\underline{n}), \ldots, v_n(\underline{m}+\underline{n})),$$

$$t^{\underline{m}} d_0 (v_1(\underline{n}), v_2(\underline{n}), \ldots, v_n(\underline{n})) = \mu(d_0) (v_1(\underline{m}+\underline{n}), v_2(\underline{m}+\underline{n}), \ldots, v_n(\underline{m}+\underline{n})),$$

where $\{v_1(\underline{0}), \ldots, v_m(\underline{0})\}$ is a basis of $T_{\underline{0}}$ and $v_i(\underline{m}) = \dfrac{1}{c_0} t^{\underline{m}} k_0 v_i(\underline{0})$, for $i = 1, 2, \ldots, m$.

(2) *As an* $(\mathcal{A}_\nu \oplus \mathcal{D}_\nu)$*-module*, T *is isomorphic to*

$$F^\alpha(\psi, b) = V(\psi, b) \otimes \mathbb{C}[t_1^{\pm 1}, \dots, t_\nu^{\pm 1}]$$

for some $\alpha = (\alpha_1, \dots, \alpha_\nu)$, ψ, *and* b, *where* $\mathcal{A}_\nu = \mathbb{C}[t_1^{\pm 1}, \dots, t_\nu^{\pm 1}]$, $\mathcal{D}_\nu$ *is the derivation algebra of* $\mathcal{A}_\nu$, *and* $V(\psi, b)$ *is an* m*-dimensional, irreducible* $gl_\nu(\mathbb{C})$*-module satisfying* $\psi(I) = b\,\mathrm{id}_{V(\psi,b)}$ *and*

$$t^{\underline{r}} d_p(w \otimes t^{\underline{m}}) = (m_p + \alpha_p) w \otimes t^{\underline{r}+\underline{m}} + \sum_{i=1}^{\nu} r_i \psi(E_{ip}) w \otimes t^{\underline{r}+\underline{m}}$$

for $w \in V(\psi, b)$.

Let

$$M = \mathrm{Ind}_{\mathcal{L}_+ + \mathcal{L}_0}^{\mathcal{L}} T \quad \text{or} \quad M = \mathrm{Ind}_{\mathcal{L}_- + \mathcal{L}_0}^{\mathcal{L}} T.$$

Theorem 3.11. *Among the submodules of* M *intersecting* T *trivially, there is a maximal one, which we denote by* M^{rad}. *Moreover* $V \cong M/M^{\mathrm{rad}}$.

4. The structure of V with $c_0 = \cdots = c_\nu = 0$

Assume that V is an irreducible weight module of $\mathcal{L}$ with finite-dimensional weight spaces and $c_0 = \cdots = c_\nu = 0$.

Lemma 4.1. *For any* $t^{\bar{r}} k_p \in \mathcal{K}$, $t^{\bar{r}} k_p$ *or* $t^{-\bar{r}} k_p$ *is locally nilpotent on* V.

Lemma 4.2. *If* V *is uniformly bounded*, $t^{\bar{r}} k_p$ *is locally nilpotent on* V *for any* $t^{\bar{r}} k_p \in \mathcal{K}$.

Proof. For $t^{\bar{r}} k_p \in \mathcal{K}$, by Lemma 4.1, $t^{\bar{r}} k_p$ or $t^{-\bar{r}} k_p$ is nilpotent on $V_{\bar{m}}$ for all $\bar{m} \in \mathbb{Z}^{\nu+1}$. Since V is uniformly bounded, i.e., $\max\{\dim V_{\bar{m}} \mid \bar{m} \in \mathbb{Z}^{\nu+1}\} < \infty$, there exists $N \in \mathbb{Z}_+$ such that

$$(t^{\bar{r}} k_p t^{-\bar{r}} k_p)^N V = 0, (t^{\bar{r}} k_p t^{-\bar{r}} k_p)^{N-1} V \neq 0$$

If the lemma is false, we can assume that $t^{-\bar{r}} k_p$ is not locally nilpotent on V. Therefore for any $0 \neq v \in V$, we have $t^{-\bar{r}} k_p v \neq 0$. So

$$(t^{\bar{r}} k_p)^N V = 0.$$

Let $t^{-2\bar{r}} d_q \in \mathcal{K}$ be such that $p \neq q$ and $r_q \neq 0$. By the fact that $[t^{-2\bar{r}} d_q, t^{\bar{r}} k_p] = r_q t^{-\bar{r}} k_p$, we deduce that $t^{-\bar{r}} k_p (t^{\bar{r}} k_p)^{N-1} V = 0$, a contradiction. □

Lemma 4.3. *If there exists* $0 \neq v \in V$ *such that* $t^{\bar{m}} k_p v = 0$ *for all* $\bar{m} \in \mathbb{Z}^{\nu+1}$ *and* $0 \leq p \leq \nu$. *Then* $\mathcal{K}(V) = 0$.

Proof. This follows from (2-2), since $\mathcal{K}$ is commutative and V is an irreducible $\mathcal{L}$-module. □

Theorem 4.4. *If* V *is uniformly bounded*, $t^{\bar{r}} k_p V$ *vanishes for any* $t^{\bar{r}} k_p \in \mathcal{K}$.

Proof. Let $0 \neq t_i k_p \in \mathcal{K}$. If $t_i k_p V = 0$, it is easy to prove that $\mathcal{K}(V) = 0$. If $t_i k_p V \neq 0$. Since V is uniformly bounded, by Lemma 4.2, there exists $l \in \mathbb{Z}_+$ such that

$$(t_i k_p t_i^{-1} k_p)^l V = 0, \quad (t_1 k_p t_1^{-1} k_p)^{l-1} V \neq 0. \tag{4-1}$$

If there exists $s \in \mathbb{Z}_+$ such that $(t_i^{-1} k_p)^s V = 0$, $(t_i^{-1} k_p)^{s-1} V \neq 0$. By the fact that $[t^{\bar m} d_i, t_i^{-1} k_p] = -t_i^{-1} t^{\bar m} k_p$ and $[t^{\bar m} d_p, t_i^{-1} k_p] = t_i^{-1} t^{\bar m} k_i$, we have

$$t^{\bar r} k_p (t_i^{-1} k_p)^{s-1} V = t^{\bar r} k_i (t^{-\bar r} k_p)^{s-1} V = 0 \quad \text{for all } \bar r \in \mathbb{Z}^{\nu+1}.$$

If $(t_i^{-1} k_p)^s V \neq 0$ for all $s \in \mathbb{Z}_+$. Then by (4-1) there is $r \geq 0$ such that $(t_i k_p)^{l-i} (t_i^{-1} k_p)^{l+i} V = 0$ for all $0 \leq i \leq r$, and $(t_i k_p)^{l-r-1} (t_i^{-1} k_p)^{l+r+1} V \neq 0$. So for any $\bar m \in \mathbb{Z}^{\nu+1}$, we have

$$t^{-\bar m} d_i (t_i k_p)^{l-r} (t_i^{-1} k_p)^{l+r+1} V = 0, \quad t^{-\bar m} d_p (t_i k_p)^{l-r} (t_i^{-1} k_p)^{l+r+1} V = 0.$$

Therefore

$$\begin{aligned} t^{\bar r} k_p (t_i k_p)^{l-r-1} (t_i^{-1} k_p)^{l+r+1} V &= 0, \\ t^{\bar r} k_i (t_i k_p)^{l-r-1} (t_i^{-1} k_p)^{l+r+1} V &= 0, \end{aligned}$$

for all $\bar r \in \mathbb{Z}^{\nu+1}$.

<u>Case 1</u>: $\nu \in 2\mathbb{Z}_+ + 1$. By the preceding discussion, there exist nonnegative integers l_i and r_i, for $i = 0, 2, 4, \ldots, \nu - 1$, such that

$$(t_\nu k_{\nu-1})^{l_{\nu-1}} (t_\nu^{-1} k_{\nu-1})^{r_{\nu-1}} (t_{\nu-2} k_{\nu-3})^{l_{\nu-3}} (t_{\nu-2}^{-1} k_{\nu-3})^{r_{\nu-3}} \cdots (t_1 k_0)^{l_0} (t_1^{-1} k_0)^{r_0} V \neq 0$$

and

$$t^{\bar m} k_p (t_\nu k_{\nu-1})^{l_{\nu-1}} (t_\nu^{-1} k_{\nu-1})^{r_{\nu-1}} (t_{\nu-2} k_{\nu-3})^{l_{\nu-3}} (t_{\nu-2}^{-1} k_{\nu-3})^{r_{\nu-3}} \cdots (t_1 k_0)^{l_0} (t_1^{-1} k_0)^{r_0} V$$

vanishes for all $0 \leq p \leq \nu$ and $\bar m \in \mathbb{Z}^{\nu+1}$. By Lemma 4.3, the conclusion of the theorem holds.

<u>Case 2</u>: $\nu \in 2\mathbb{Z}$. Then there exist nonnegative integers l_i and r_i, for $i = 0, 2, 4, \ldots, \nu - 2$, such that

$$W = (t_{\nu-1} k_{\nu-2})^{l_{\nu-2}} (t_{\nu-1}^{-1} k_{\nu-2})^{r_{\nu-2}} (t_{\nu-3} k_{\nu-4})^{l_{\nu-4}} (t_{\nu-3}^{-1} k_{\nu-4})^{r_{\nu-4}} \cdots (t_1 k_0)^{l_0} (t_1^{-1} k_0)^{r_0} V$$

is nonzero and

$$t^{\bar m} k_p W = 0 \tag{4-2}$$

for all $0 \leq p \leq \nu - 1$ and $\bar m \in \mathbb{Z}^{\nu+1}$. By (2-1), we know that

$$t^{\bar m} k_\nu W = 0, \tag{4-3}$$

for $\bar{m} \in \mathbb{Z}^{\nu+1}$ such that $m_\nu \neq 0$. If there exists $t^{\bar{r}_0} k_\nu$ satisfying $t^{\bar{r}_0} k_\nu W \neq 0$, let

$$\begin{aligned} \mathcal{L}_\nu = \operatorname{span}\{ & t^{\underline{m}} d_i, t^{\bar{m}} d_\nu, t^{\underline{m}} k_\nu \mid t^{\underline{m}} = t_0^{m_0} t_1^{m_1} \cdots t_{\nu-1}^{m_{\nu-1}}, 0 \le i \le \nu - 1, \\ & \underline{m} = (m_0, \ldots, m_{\nu-1}) \in \mathbb{Z}^\nu, \bar{m} \in \mathbb{Z}^{\nu+1}\}, \\ W' = U(&\mathcal{L}_\nu) W. \end{aligned}$$

Then $W' \neq 0$ and

$$t^{\bar{m}} k_p W' = 0, \qquad t^{\bar{n}} k_\nu W' = 0,$$

for all $0 \le p \le \nu - 1$, $\bar{m} \in \mathbb{Z}^{\nu+1}$, and $\bar{n} \in \mathbb{Z}^{\nu+1}$ such that $n_\nu \neq 0$. If there exists $0 \neq t^{\underline{m}} k_\nu$ such that $t^{\underline{m}} k_\nu W' \neq 0$, we have

$$(t^{-\underline{m}} k_\nu)^l (t^{\underline{m}} k_\nu)^l W' = 0 \quad \text{and} \quad (t^{-\underline{m}} k_\nu)^{l-1} (t^{\underline{m}} k_\nu)^{l-1} W' \neq 0$$

for some $l \in \mathbb{Z}_+$. As in the preceding proof, we can deduce that there exists a nonzero $v \in W'$ such that

$$t^{\underline{n}} k_\nu v = 0$$

for all $\underline{n} \in \mathbb{Z}^\nu$. Therefore

$$t^{\bar{m}} k_p v = 0$$

for all $\bar{m} \in \mathbb{Z}^{\nu+1}$ and $0 \le p \le \nu$. We have proved that $\mathcal{K}(V) = 0$. □

References

[Belavin et al. 1984] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, "Infinite conformal symmetry in two-dimensional quantum field theory", *Nuclear Phys. B* **241**:2 (1984), 333–380. MR 86m:81097 Zbl 0661.17013

[Berman and Billig 1999] S. Berman and Y. Billig, "Irreducible representations for toroidal Lie algebras", *J. Algebra* **221**:1 (1999), 188–231. MR 2000k:17004 Zbl 0942.17016

[Billig 2003] Y. Billig, "Abelian extensions of the group of diffeomorphisms of a torus", *Lett. Math. Phys.* **64**:2 (2003), 155–169. MR 2004h:22012 Zbl 1079.58004

[Chari and Pressley 1988] V. Chari and A. Pressley, "Unitary representations of the Virasoro algebra and a conjecture of Kac", *Compositio Math.* **67**:3 (1988), 315–342. MR 89h:17025 Zbl 0661.17022

[Di Francesco et al. 1997] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal field theory*, Springer, New York, 1997. MR 97g:81062 Zbl 0869.53052

[Dotsenko and Fateev 1984] V. S. Dotsenko and V. A. Fateev, "Conformal algebra and multipoint correlation functions in 2D statistical models", *Nuclear Phys. B* **240**:3 (1984), 312–348. MR 85i:82061

[Eswara Rao 2004] S. Eswara Rao, "Partial classification of modules for Lie algebra of diffeomorphisms of d-dimensional torus", *J. Math. Phys.* **45**:8 (2004), 3322–3333. MR 2005d:17028 Zbl 1071.17020

[Eswara Rao and Jiang 2005] S. Eswara Rao and C. Jiang, "Classification of irreducible integrable representations for the full toroidal Lie algebras", *J. Pure Appl. Algebra* **200**:1-2 (2005), 71–85. MR 2006b:17038 Zbl 1070.17009

[Eswara Rao and Moody 1994] S. Eswara Rao and R. V. Moody, "Vertex representations for n-toroidal Lie algebras and a generalization of the Virasoro algebra", *Comm. Math. Phys.* **159**:2 (1994), 239–264. MR 94m:17028 Zbl 0808.17018

[Jiang and Meng 2003] C. Jiang and D. Meng, "Integrable representations for generalized Virasoro-toroidal Lie algebras", *J. Algebra* **270**:1 (2003), 307–334. MR 2005b:17053 Zbl 1037.17029

[Kac 1982] V. G. Kac, "Some problems on infinite-dimensional Lie algebras and their representations", pp. 117–126 in *Lie algebras and related topics* (New Brunswick, NJ, 1981), edited by D. J. Winter, Lecture Notes in Math. **933**, Springer, Berlin, 1982. MR 84e:17010 Zbl 0493.17011

[Kaplansky and Santharoubane 1985] I. Kaplansky and L. J. Santharoubane, "Harish-Chandra modules over the Virasoro algebra", pp. 217–231 in *Infinite dimensional groups with applications* (Berkeley, 1984), edited by V. Kac, Math. Sci. Res. Inst. Publ. **4**, Springer, New York, 1985. MR 87d:17013 Zbl 0589.17013

[Larsson 1992] T. A. Larsson, "Conformal fields: a class of representations of Vect(N)", *Internat. J. Modern Phys. A* **7**:26 (1992), 6493–6508. MR 93h:17053 Zbl 0972.17502

[Mathieu 1992] O. Mathieu, "Classification of Harish-Chandra modules over the Virasoro Lie algebra", *Invent. Math.* **107**:2 (1992), 225–234. MR 93d:17034 Zbl 0779.17025

[Ramos et al. 1990] E. Ramos, C.-H. Sah, and R. E. Shrock, "Algebras of diffeomorphisms of the N-torus", *J. Math. Phys.* **31**:8 (1990), 1805–1816. MR 91j:17043 Zbl 0733.17014

Received December 3, 2005.

CUIPO JIANG
DEPARTMENT OF MATHEMATICS
SHANGHAI JIAOTONG UNIVERSITY
SHANGHAI 200030
CHINA
cpjiang@sjtu.edu.cn

QIFEN JIANG
DEPARTMENT OF MATHEMATICS
SHANGHAI JIAOTONG UNIVERSITY
SHANGHAI 200030
CHINA
qfjiang@sjtu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

HIGHER HOMOTOPY COMMUTATIVITY OF H-SPACES AND HOMOTOPY LOCALIZATIONS

YUSUKE KAWAMOTO

Dedicated to Professor Takao Matumoto on his sixtieth birthday

In this paper, we prove that the homotopy localization of an AC_n-space is an AC_n-space so that the universal map is an AC_n-map. This result is used to study the higher homotopy commutativity of H-spaces with finitely generated cohomology over the Steenrod algebra $\mathscr{A}_p^*$. Our result shows that for any prime p, if X is a connected AC_p-space whose mod p cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$, then X has the mod p homotopy type of a Postnikov H-space.

1. Introduction

The theory of H-spaces has been studied in algebraic topology to understand homotopy properties of Lie groups. Given a prime p, a $\mathbb{Z}/p$-finite H-space is an H-space whose mod p cohomology is finite dimensional. In recent decades, many theorems have been proved about $\mathbb{Z}/p$-finite H-spaces [Kane 1988; Lin 1995], which suggest that they have many similar properties to those of Lie groups.

In this paper, we study an H-space which need not be $\mathbb{Z}/p$-finite but whose mod p cohomology is finitely generated as an algebra over the Steenrod algebra $\mathscr{A}_p^*$. For example, the n-connected covering $X\langle n\rangle$ of a $\mathbb{Z}/p$-finite H-space X is not $\mathbb{Z}/p$-finite for $n \geq 3$ but the mod p cohomology is finitely generated as an algebra over $\mathscr{A}_p^*$, by [Castellana et al. 2006, Corollary 4.3]. Eilenberg–Mac Lane spaces $K(\mathbb{Z}, n)$ and $K(\mathbb{Z}/p^i, n)$ are other examples for $n, i \geq 1$.

Using the homotopy localizations of Bousfield [1994] and Dror Farjoun [1996], Castellana, Crespo and Scherer have studied H-spaces with finitely generated cohomology over $\mathscr{A}_p^*$ [Castellana et al. 2007]. In their Theorem 7.3, these authors proved that if X is such an H-space, the $B\mathbb{Z}/p$-localization $L_{B\mathbb{Z}/p}(X)$ is a $\mathbb{Z}/p$-finite H-space and the homotopy fiber $F(\phi_X)$ of the universal map $\phi_X\colon X \to L_{B\mathbb{Z}/p}(X)$ is mod p homotopy equivalent to a Postnikov H-space (see Theorem

MSC2000: primary 55P45, 55P48; secondary 55P15, 55P65.
Keywords: higher homotopy commutativity, homotopy localizations, H-spaces, A_n-spaces, AC_n-spaces.

5.1). Here an H-space is called Postnikov if the homotopy groups are finitely generated over the p-adic integers $\mathbb{Z}_p^\wedge$ which vanish above some dimension, and mod p homotopy equivalence means homotopy equivalence up to p-completion in the sense of [Bousfield and Kan 1972].

Moreover, by combining their main result with the mod 2 torus theorem by Hubbuck [1969] and Lin [1985], Castellana et al. generalized results of Slack [1991] and Lin and Williams [1991], as follows:

Theorem 1.1 [Castellana et al. 2007, Corollary 7.4]. *If X is a connected homotopy commutative H-space whose* mod 2 *cohomology $H^*(X;\mathbb{Z}/2)$ is finitely generated as an algebra over $\mathscr{A}_2^*$, then X is* mod 2 *homotopy equivalent to a Postnikov H-space.*

On the other hand, the odd prime version of Theorem 1.1 does not hold. In fact, Iriye and Kono [1985] showed that for an odd prime p, any connected H-space is mod p homotopy equivalent to a homotopy commutative H-space. Moreover, $Sp(2)_3^\wedge$ for $p=3$ and $(S^3)_p^\wedge$ for $p \geq 5$ are examples of homotopy commutative loop spaces which are not Postnikov H-spaces by McGibbon [1984], where $Y_p^\wedge$ denotes the p-completion of a space Y.

To describe an odd prime version of Theorem 1.1, we use the higher homotopy commutativity of the multiplication. Such notions are first considered by Sugawara [1960] and Williams [1969] in the case of topological monoids. (The higher homotopy commutativity of the third order in the sense of Williams is illustrated by the left hexagon in Figure 1.)

Williams' definition was generalized to the case of A_n-spaces in [Hemmi and Kawamoto 2004] (see also [Hemmi 1991]). An A_n-space with a multiplication admitting the higher homotopy commutativity of the n-th order is called an AC_n-space. By [Hemmi and Kawamoto 2004, Example 3.2(1)], an AC_2-space is the same as a homotopy commutative H-space. Let X be an A_3-space admitting an AC_2-structure. Then by using the associating homotopy $M_3\colon I \times X^3 \to X$ and the

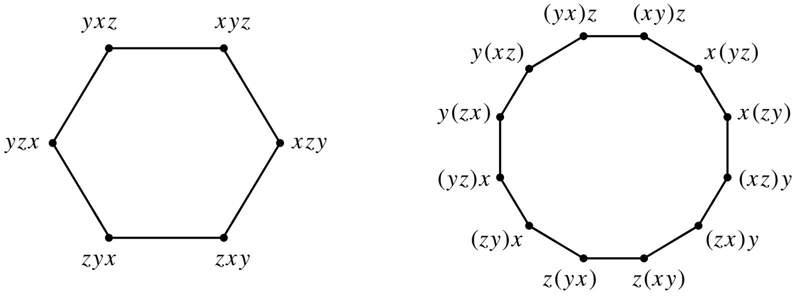


Figure 1. The higher homotopy commutativity of the third order.

commuting homotopy $Q_2 \colon I \times X^2 \to X$, we can define a map $\widetilde{Q}_3 \colon S^1 \times X^3 \to X$ illustrated by the right dodecagon in Figure 1. For example, the uppermost edge represents the commuting homotopy between xy and yx given by $Q_2(t, x, y)z$, and the next right edge is the associating homotopy between $(xy)z$ and $x(yz)$ given by $M_3(t, x, y, z)$. Then X is an AC_3-space if and only if $\widetilde{Q}_3$ is extended to a map $Q_3 \colon D^2 \times X^3 \to X$. In general, X is an AC_n-space if and only if there is a family of maps

$$\{Q_i \colon D^{i-1} \times X^i \to X\}_{1 \le i \le n}$$

with the relations in [Hemmi and Kawamoto 2004, Proposition 2.1].

To generalize Theorem 1.1 to the case of any prime p, we first show:

Theorem A. *Let A be a topological space and $n \ge 1$. If X is an AC_n-space, then the A-localization $L_A(X)$ is an AC_n-space so that the universal map $\phi_X \colon X \to L_A(X)$ is an AC_n-map.*

From Theorem A and [Castellana et al. 2007, Theorem 7.3], we can generalize the mod p torus theorem stated in [Hemmi and Kawamoto 2004, Corollary 1.1] to the case of AC_p-spaces with finitely generated cohomology over $\mathscr{A}_p^*$.

Theorem B. *Let p be a prime. If X is a connected AC_p-space whose* mod *p cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$, then X is* mod *p homotopy equivalent to a Postnikov H-space.*

Theorem B is a generalization of Theorem 1.1 to the case of any prime p since an AC_2-space is the same as a homotopy commutative H-space. In the above theorem, the assumption of AC_p-space cannot be relaxed to AC_{p-1}-space. In fact, by [Hemmi and Kawamoto 2004, Proposition 3.8], the $(2m-1)$-dimensional sphere $(S^{2m-1})^{\wedge}_p$ is an AC_{p-1}-space but not a Postnikov H-space for $m \ge 2$.

Moreover, since the loop space of an H-space admits an AC_∞-structure by [Hemmi and Kawamoto 2004, Example 3.2(3)], Theorem B implies:

Corollary 1.2 [Castellana et al. 2007, p. 17]. *Let p be a prime. Assume that X is a connected loop space whose* mod *p cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$. If the classifying space BX is an H-space, then X is* mod *p homotopy equivalent to a Postnikov H-space.*

There is an example of a Postnikov loop space Y admitting an AC_p-structure such that the classifying space BY is not an H-space by [McGibbon 1989, Example 5]. Corollary 1.2 is a generalization of results from [Aguadé and Smith 1986; Kawamoto 1999; Lin 1994].

Bousfield [2001] studied the $K(n)_*$-localizations of Postnikov H-spaces, where $K(n)_*$ denotes the Morava K-homology theory for $n \ge 1$. By Theorem B and [Bousfield 2001, Theorem 7.2], we have:

Corollary 1.3. *Let p be a prime and $n \geq 1$. If X is a connected AC_p-space whose* mod p *cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$, then the $K(n)_*$-localization $L_{K(n)_*}(X_p^\wedge)$ is* mod p *homotopy equivalent to the $\Sigma^n B\mathbb{Z}/p$-localization $L_{\Sigma^n B\mathbb{Z}/p}(X_p^\wedge)$. In particular, $X_p^\wedge$ is $K(n)_*$-local if and only if the n-fold loop space $\Omega^n X_p^\wedge$ is $B\mathbb{Z}/p$-local.*

We also generalize [Hemmi and Kawamoto 2007, Theorem B] to the case of A_p-spaces with finitely generated cohomology over $\mathscr{A}_p^*$.

Theorem C. *Let p be an odd prime. Assume that X is a connected A_p-space admitting an AC_n-structure with $n > (p-1)/2$ and the* mod p *cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$. If the Steenrod operations $\mathscr{P}^j$ act on the indecomposable module $QH^*(X; \mathbb{Z}/p)$ trivially for $j \geq 1$, then X is* mod p *homotopy equivalent to a finite product of $(S^1)_p^\wedge$s, $(\mathbb{C}P^\infty)_p^\wedge$s and $B\mathbb{Z}/p^i$s for $i \geq 1$.*

In Theorem C, the assumption $n > (p-1)/2$ is necessary. In fact, by [Hemmi 1991, Theorem 2.4], $(S^3)_p^\wedge$ is an A_p-space admitting an $AC_{(p-1)/2}$-structure for any odd prime p.

Outline of article. In Section 2, we recall the associahedra, the multiplihedra and the permuto-associahedra. Then we show that the permuto-associahedra are decomposed by using the multiplihedra in a combinatorial way (see Proposition 2.1). In Section 3, we give the definition of an AC_n-map between AC_n-spaces by using Proposition 2.1 (see Definition 3.1). Section 4 is devoted to the proof of Theorem A. We show that if $\phi\colon X \to Y$ is an A_n-map between A_n-spaces and Y is ϕ-local, then ϕ transmits an AC_n-structure from X to Y (see Proposition 4.1). By applying Proposition 4.1 to the universal map $\phi_X\colon X \to L_A(X)$ for the A-localization of X, we prove Theorem A. In Section 5, we first recall the result of [Castellana et al. 2007] on H-spaces with finitely generated cohomology over $\mathscr{A}_p^*$ (Theorem 5.1). From Theorem A, Theorem 5.1 and the results in [Hemmi 1991; Hemmi and Kawamoto 2004], we prove Theorem B. Next Corollary 1.3 is proved by Theorem B and the result from [Bousfield 2001] on the $K(n)_*$-localizations of Postnikov H-spaces. We finally give the proof of Theorem C by using Theorem A and [Hemmi and Kawamoto 2007, Theorem B].

The content of the paper was first presented in a conference on algebraic topology at Shinshu University in July 2005. The author is grateful to the organizers for their kind invitation and hospitality.

2. Decompositions of the permuto-associahedra

We first recall the associahedra $\{K_n\}_{n\geq 2}$ of Stasheff and the multiplihedra $\{J_n\}_{n\geq 1}$ of Iwase and Mimura.

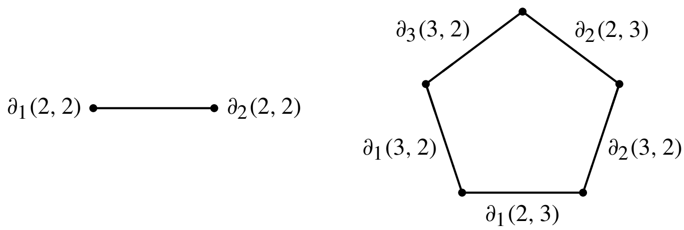


Figure 2. The associahedra K_3 and K_4.

Stasheff [1963, p. 283] constructed the associahedra $\{K_n\}_{n\geq 2}$ to introduce the concept of A_n-space (see Section 3). From the construction, the associahedron K_n is an $(n-2)$-dimensional polytope whose boundary ∂K_n is given by

$$\partial K_n = \bigcup_{r,s,k} K_k(r,s)$$

for $n \geq 2$, where $r, s \geq 2$ with $r+s = n+1$ and $1 \leq k \leq r$. Here the facet (codimension-one face) $K_k(r,s)$ is homeomorphic to the product $K_r \times K_s$ by a face operator $\partial_k(r,s)\colon K_r \times K_s \to K_k(r,s)$ with the relations in [Stasheff 1963, p. 278, 3(a),(b)]. There is a family of degeneracy operators $\{\theta_j : K_n \to K_{n-1}\}_{1\leq j\leq n}$ satisfying the relations in [Stasheff 1963, p. 278, Proposition 3].

The associahedra $\{K_n\}_{n\geq 2}$ are also used in [Stasheff 1970, Definition 11.9] to define an A_n-map from an A_n-space to a topological monoid.

Iwase and Mimura [1989, §2] introduced the multiplihedra $\{J_n\}_{n\geq 1}$ for the purpose of defining an A_n-map between A_n-spaces (see Section 3). From the properties in [Iwase and Mimura 1989, p. 200, (2-a) and (2-b)], the multiplihedron J_n is an $(n-1)$-dimensional polytope whose boundary ∂J_n is given by

$$\partial J_n = \bigcup_{k,r,s} J_k(r,s) \cup \bigcup_{q,r_1,\ldots,r_q} J(q, r_1, \ldots, r_q)$$

for $n \geq 1$, where $r \geq 1$, $s \geq 2$ with $r+s = n+1$ and $1 \leq k \leq r$, and $2 \leq q \leq n$, $r_1, \ldots, r_q \geq 1$ with $r_1 + \cdots + r_q = n$. As in the case of the associahedra, we have face

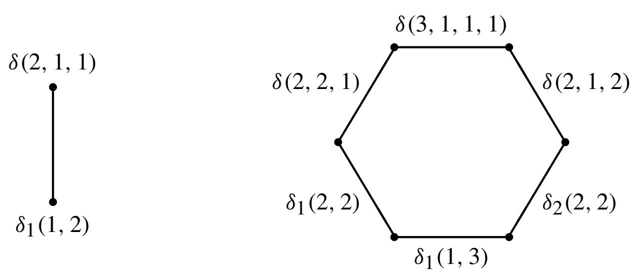


Figure 3. The multiplihedra J_2 and J_3.

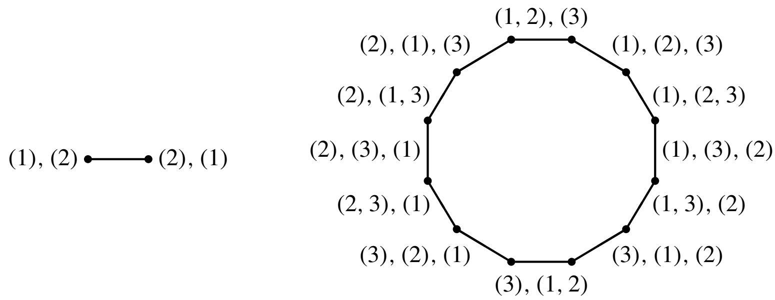


Figure 4. The permuto-associahedra Γ_2 and Γ_3.

operators $\delta_k(r, s)\colon J_r \times K_s \to J_k(r, s)$ and $\delta(q, r_1, \ldots, r_q)\colon K_q \times J_{r_1} \times \cdots \times J_{r_q} \to J(q, r_1, \ldots, r_q)$ with the relations in (2-c) of the same work. The degeneracy operators $\{\xi_j\colon J_n \to J_{n-1}\}_{1\le j\le n}$ satisfy the relations in (2-d).

We next recall the permuto-associahedra $\{\Gamma_n\}_{n\ge 1}$ constructed by Kapranov and by Reiner and Ziegler. By [Kapranov 1993, Theorem 2.5] and [Reiner and Ziegler 1994, Theorem 2], the permuto-associahedron Γ_n is an $(n-1)$-dimensional polytope whose faces are described in a combinatorial way for $n \ge 1$ (see also [Ziegler 1995, Definition 9.13, Example 9.14]). In particular, a facet of Γ_n is represented by a partition of the sequence $\mathbf{n} = (1, \ldots, n)$ into at least two parts. Here a partition of $\mathbf{n}$ of type $(t_1, \ldots, t_l)$ is an ordered sequence $(\alpha_1, \ldots, \alpha_l)$ consisting of disjoint subsequences α_i of $\mathbf{n}$ of length t_i with $\alpha_1 \cup \cdots \cup \alpha_l = \mathbf{n}$. See [Hemmi and Kawamoto 2004; Ziegler 1995] for the full details of the partitions.

Let $\Gamma(\alpha_1, \ldots, \alpha_l)$ denote the facet of Γ_n corresponding to a partition $(\alpha_1, \ldots, \alpha_l)$. The boundary of Γ_n is given by

$$\partial\Gamma_n = \bigcup_{(\alpha_1,\ldots,\alpha_l)} \Gamma(\alpha_1, \ldots, \alpha_l), \tag{2-1}$$

where the union covers all partitions $(\alpha_1, \ldots, \alpha_l)$ of $\mathbf{n}$ with $l \ge 2$. If $(\alpha_1, \ldots, \alpha_l)$ is of type $(t_1, \ldots, t_l)$, then the facet $\Gamma(\alpha_1, \ldots, \alpha_l)$ is homeomorphic to the product $K_l \times \Gamma_{t_1} \times \cdots \times \Gamma_{t_l}$ by a face operator $\epsilon^{(\alpha_1,\ldots,\alpha_l)}\colon K_l \times \Gamma_{t_1} \times \cdots \times \Gamma_{t_l} \to \Gamma(\alpha_1, \ldots, \alpha_l)$ with the relations in Proposition 2.1 of [Hemmi and Kawamoto 2004]. Moreover, there are degeneracy operators $\{\omega_j\colon \Gamma_n \to \Gamma_{n-1}\}_{1\le j\le n}$ satisfying the relations in Proposition 2.3 of the same reference.

In Definition 3.1, we need the following result:

Proposition 2.1. *Let* $n \ge 1$.

(1) *The permuto-associahedron* Γ_n *is decomposed by*

$$\Gamma_n = \bigcup_{(\beta_1,\ldots,\beta_m)} B(\beta_1, \ldots, \beta_m),$$

where the union covers all partitions $(\beta_1, \dots, \beta_m)$ *of* $\mathbf{n}$ *with* $m \geq 1$.

(2) *If* $(\beta_1, \dots, \beta_m)$ *is a partition of* $\mathbf{n}$ *of type* $(u_1, \dots, u_m)$, *then* $B(\beta_1, \dots, \beta_m)$ *is homeomorphic to the product* $J_m \times \Gamma_{u_1} \times \cdots \times \Gamma_{u_m}$ *by an operator*

$$\iota^{(\beta_1,\dots,\beta_m)} \colon J_m \times \Gamma_{u_1} \times \cdots \times \Gamma_{u_m} \to B(\beta_1, \dots, \beta_m).$$

By an inductive argument, we can show:

Lemma 2.2 [Stasheff 1963, p. 288, Proposition 25]. *There is a family of homeomorphisms* $\{\zeta_m \colon I \times K_m \to J_m\}_{m\geq 2}$ *with the relations*

$$\zeta_m(0, \sigma) = \delta_1(1, m)(*, \sigma),$$
$$\zeta_m(t, \partial_k(r, s)(\rho, \sigma)) = \delta_k(r, s)(\zeta_r(t, \rho), \sigma)$$

for $r, s \geq 2$ *with* $r + s = m + 1$ *and* $1 \leq k \leq r$.

Proof of Proposition 2.1. We work by induction on n. Since $\Gamma_1 = J_1 = *$, the result is clear for $n = 1$. We put

$$\mathcal{U}_n = \Gamma_n \cup_{\{0\}\times\partial\Gamma_n} I \times \partial\Gamma_n, \tag{2-2}$$

where I is the unit interval and $\{0\} \times \partial\Gamma_n$ is identified with $\partial\Gamma_n \subset \Gamma_n$. It is clear that $\mathcal{U}_n$ is homeomorphic to the $(n-1)$-dimensional ball.

Let $B(\mathbf{n}) = \Gamma_n \subset \mathcal{U}_n$. Then an operator $\iota^{(\mathbf{n})} \colon J_1 \times \Gamma_n \to B(\mathbf{n})$ is defined by $\iota^{(\mathbf{n})}(*, \tau) = \tau$ for $\tau \in \Gamma_n$. If $m \geq 2$, then by Lemma 2.2, we can identify J_m with

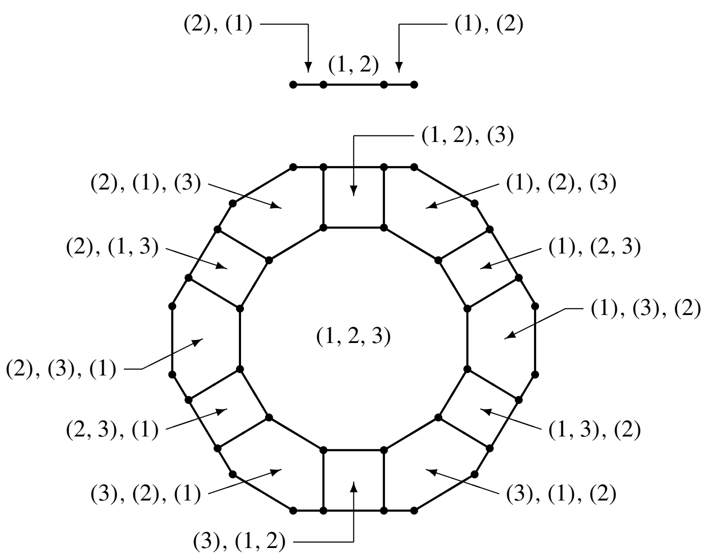


Figure 5. The decompositions of Γ_2 and Γ_3.

$I \times K_m$ by $\zeta_m : I \times K_m \to J_m$. Assume that $(\beta_1, \dots, \beta_m)$ is a partition of $\mathbf{n}$ of type $(u_1, \dots, u_m)$ with $m \geq 2$. Put

$$B(\beta_1, \dots, \beta_m) = I \times \Gamma(\beta_1, \dots, \beta_m) \subset \mathcal{U}_n,$$

and define an operator $\iota^{(\beta_1,\dots,\beta_m)} : J_m \times \Gamma_{u_1} \times \cdots \times \Gamma_{u_m} \to B(\beta_1, \dots, \beta_m)$ by

$$\iota^{(\beta_1,\dots,\beta_m)}(\zeta_m(t, \sigma), \tau_1, \dots, \tau_m) = (t, \epsilon^{(\beta_1,\dots,\beta_m)}(\sigma, \tau_1, \dots, \tau_m)).$$

Then by (2-1) and (2-2), we have that

$$\mathcal{U}_n = \bigcup_{(\beta_1,\dots,\beta_m)} B(\beta_1, \dots, \beta_m),$$

where the union covers all partitions $(\beta_1, \dots, \beta_m)$ of $\mathbf{n}$ with $m \geq 1$.

By Lemma 2.2, we see that

$$\zeta_m(\{1\} \times K_m) = \bigcup_{q, r_1, \dots, r_q} J(q, r_1, \dots, r_q)$$

for $2 \leq q \leq m$ and $r_1, \dots, r_q \geq 1$ with $r_1 + \cdots + r_q = m$. This implies that

$$\text{(2-3)} \quad \partial\mathcal{U}_n = \bigcup_{(\beta_1,\dots,\beta_m)} \iota^{(\beta_1,\dots,\beta_m)}\left(\left(\bigcup_{q, r_1, \dots, r_q} J(q, r_1, \dots, r_q)\right) \times \Gamma_{u_1} \times \cdots \times \Gamma_{u_m}\right)$$

for $2 \leq q \leq m$ and $r_1, \dots, r_q \geq 1$ with $r_1 + \cdots + r_q = m$, where $(\beta_1, \dots, \beta_m)$ are partitions of $\mathbf{n}$ of type $(u_1, \dots, u_m)$ with $m \geq 2$. If we define face operators on $\partial\mathcal{U}_n$ satisfying the relations in [Hemmi and Kawamoto 2004, Proposition 2.1], then $\partial\mathcal{U}_n$ is homeomorphic to $\partial\Gamma_n$, and it follows that $\mathcal{U}_n$ is homeomorphic to Γ_n, which implies the required conclusion.

Recall that $\partial\Gamma_n$ is given by

$$\partial\Gamma_n = \bigcup_{(\alpha_1,\dots,\alpha_l)} \Gamma(\alpha_1, \dots, \alpha_l),$$

where the union covers all partitions $(\alpha_1, \dots, \alpha_l)$ of $\mathbf{n}$ with $l \geq 2$.

Assume that $(\alpha_1, \dots, \alpha_l)$ is a partition of $\mathbf{n}$ of type $(t_1, \dots, t_l)$ with $l \geq 2$. Then by inductive hypothesis, we can assume that

$$\Gamma_{t_j} = \bigcup_{(\gamma_{j,1},\dots,\gamma_{j,h_j})} B(\gamma_{j,1}, \dots, \gamma_{j,h_j})$$

for $1 \leq j \leq l$, where the union covers all partitions $(\gamma_{j,1}, \dots, \gamma_{j,h_j})$ of $(1, \dots, t_j)$ with $h_j \geq 1$. If $(\gamma_{j,1}, \dots, \gamma_{j,h_j})$ is a partition of $(1, \dots, t_j)$ of type $(v_{j,1}, \dots, v_{j,h_j})$, then by inductive hypothesis, we have the operator $\iota^{(\gamma_{j,1},\dots,\gamma_{j,h_j})} : J_{h_j} \times \Gamma_{v_{j,1}} \times \cdots \times \Gamma_{v_{j,h_j}} \to B(\gamma_{j,1}, \dots, \gamma_{j,h_j})$ which is a homeomorphism. Put $m = h_1 + \cdots + h_l$.

We give a partition $(\beta_1, \dots, \beta_m)$ of $\mathbf{n}$ of type $(v_{1,1}, \dots, v_{1,h_1}, \dots, v_{l,1}, \dots, v_{l,h_l})$ by

$$\beta_i(t) = \alpha_j \gamma_{j,i-(h_1+\cdots+h_{j-1})}(t)$$

for $h_1 + \cdots + h_{j-1} + 1 \le i \le h_1 + \cdots + h_j$ and $1 \le t \le i - (h_1 + \cdots + h_{j-1})$. Define a face operator $\epsilon^{(\alpha_1,\dots,\alpha_l)} \colon K_l \times \Gamma_{t_1} \times \cdots \times \Gamma_{t_l} \to \partial \mathcal{U}_n$ by

$$\begin{aligned}\epsilon^{(\alpha_1,\dots,\alpha_l)}&(\sigma, \iota^{(\gamma_{1,1},\dots,\gamma_{1,h_1})}(\rho_1, \tau_{1,1}, \dots, \tau_{1,h_1}), \dots, \iota^{(\gamma_{l,1},\dots,\gamma_{l,h_l})}(\rho_l, \tau_{l,1}, \dots, \tau_{l,h_l}))\\ &= \iota^{(\beta_1,\dots,\beta_m)}(\delta(l, h_1, \dots, h_l)(\sigma, \rho_1, \dots, \rho_l), \tau_{1,1}, \dots, \tau_{1,h_1}, \dots, \tau_{l,1}, \dots, \tau_{l,h_l}).\end{aligned}$$

By (2-3) and the relation [Iwase and Mimura 1989, p. 201, (c-4)], the face operator satisfies the relations in [Hemmi and Kawamoto 2004, Proposition 2.1]. This implies that $\mathcal{U}_n$ is homeomorphic to Γ_n, and so we have the required conclusion. This completes the proof. □

Remark 2.3. The decomposition of Γ_n in Proposition 2.1 is compatible with the degeneracy operators $\{\omega_j \colon \Gamma_n \to \Gamma_{n-1}\}_{1 \le j \le n}$. Assume that $(\beta_1, \dots, \beta_m)$ is a partition of $\mathbf{n}$ of type $(u_1, \dots, u_m)$ for $u_1, \dots, u_m \ge 1$ with $u_1 + \cdots + u_m = n$. Let $1 \le j \le n$. Then $\beta_k(t) = j$ for some $1 \le k \le m$ and $1 \le t \le u_k$.

(i) If $u_k \ge 2$, then

$$\omega_j \iota^{(\beta_1,\dots,\beta_m)}(\sigma, \tau_1, \dots, \tau_m) = \iota^{(\tilde\beta_1,\dots,\tilde\beta_m)}(\sigma, \tau_1, \dots, \tau_{k-1}, \omega_t(\tau_k), \tau_{k+1}, \dots, \tau_m),$$

where $(\tilde\beta_1, \dots, \tilde\beta_m)$ is the partition of $(1, \dots, n-1)$ of type $(u_1, \dots, u_{k-1}, u_k - 1, u_{k+1}, \dots, u_m)$ given by

$$\tilde\beta_k(s) = \begin{cases} \beta_k(s) & \text{if } \beta_k(s) < j, \\ \beta_k(s+1) - 1 & \text{if } \beta_k(s) \ge j \end{cases}$$

and for $1 \le i \le n$ with $i \ne k$,

$$\tilde\beta_i(s) = \begin{cases} \beta_i(s) & \text{if } \beta_i(s) < j, \\ \beta_i(s) - 1 & \text{if } \beta_i(s) > j. \end{cases} \tag{2-4}$$

(ii) If $u_k = 1$, then

$$\omega_j \iota^{(\beta_1,\dots,\beta_m)}(\sigma, \tau_1, \dots, \tau_m) = \iota^{(\tilde\beta_1,\dots,\tilde\beta_{k-1},\tilde\beta_{k+1},\dots,\tilde\beta_m)}(\xi_k(\sigma), \tau_1, \dots, \tau_{k-1}, \tau_{k+1}, \dots, \tau_m),$$

where $(\tilde\beta_1, \dots, \tilde\beta_{k-1}, \tilde\beta_{k+1}, \dots, \tilde\beta_m)$ is the partition of $(1, \dots, n-1)$ of type

$$(u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_m)$$

given by (2-4) and ξ_k denotes the degeneracy operator of J_m.

3. Higher homotopy commutativity

We first recall the higher homotopy associativity of H-spaces and H-maps.

Sugawara [1957] gave a criterion for a topological space to have the homotopy type of a loop space. Later Stasheff [1963] expanded his definition, and introduced the concept of A_n-space by using the associahedra $\{K_i\}_{i\geq 2}$. Let X be an H-space whose multiplication is given by $\mu\colon X\times X\to X$ with $\mu(x,*)=\mu(*,x)=x$ for $x\in X$. Then an A_n-form on X is a family of maps $\{M_i\colon K_i\times X^i\to X\}_{2\leq i\leq n}$ with the relations

$$M_2(*,x,y)=\mu(x,y),$$

$$M_i(\partial_k(r,s)(\rho,\sigma),x_1,\dots,x_i)=M_r(\rho,x_1,\dots,x_{k-1},M_s(\sigma,x_k,\dots,x_{k+s-1}),x_{k+s},\dots,x_i)$$

for $r,s\geq 2$ with $r+s=i+1$ and $1\leq k\leq r$, and

$$M_i(\sigma,x_1,\dots,x_{j-1},*,x_{j+1},\dots,x_i)=M_{i-1}(\theta_j(\sigma),x_1,\dots,x_{j-1},x_{j+1},\dots,x_i)$$

for $1\leq j\leq i$.

An A_1-space is just a topological space, and an H-space which admits an A_n-form is called an A_n-space for $n\geq 2$. From the definition, an A_2-space and an A_3-space are an H-space and a homotopy associative H-space, respectively. Moreover, an A_∞-space X has the homotopy type of a loop space admitting the classifying space BX with $\Omega(BX)\simeq X$ (see [Kane 1988, §6-2]).

It is natural to consider the concept of A_n-map between A_n-spaces. Sugawara [1960] first considered such a concept for a map between topological monoids. Stasheff [1970] next studied an A_n-map from an A_n-space to a topological monoid by using the associahedra $\{K_i\}_{i\geq 2}$ used for the definition of an A_n-space.

The full generality was described by Iwase and Mimura [1989] by using the multiplihedra $\{J_i\}_{i\geq 1}$. Let X and Y be A_n-spaces with the A_n-forms $\{M_i\}_{2\leq i\leq n}$ and $\{N_i\}_{2\leq i\leq n}$, respectively. Assume that $\phi\colon X\to Y$ is a map. Then an A_n-form on ϕ is a family of maps $\{F_i\colon J_i\times X^i\to Y\}_{1\leq i\leq n}$ with the relations

$$F_1(*,x)=\phi(x),$$

$$F_i(\delta_k(r,s)(\rho,\sigma),x_1,\dots,x_i)=F_r(\rho,x_1,\dots,x_{k-1},M_s(\sigma,x_k,\dots,x_{k+s-1}),x_{k+s},\dots,x_i)$$

for $r\geq 1$, $s\geq 2$ with $r+s=i+1$ and $1\leq k\leq r$,

$$\begin{aligned}F_i(\delta(q,r_1,\dots,r_q)&(\rho,\sigma_1,\dots,\sigma_q),x_1,\dots,x_i)\\&=N_q(\rho,F_{r_1}(\sigma_1,x_1,\dots,x_{r_1}),\dots,F_{r_q}(\sigma_q,x_{r_1+\cdots+r_{q-1}+1},\dots,x_i))\end{aligned}$$

for $q\geq 2$ and $r_1,\dots,r_q\geq 1$ with $r_1+\cdots+r_q=n$, and

$$F_i(\rho,x_1,\dots,x_{j-1},*,x_{j+1},\dots,x_i)=F_{i-1}(\xi_j(\rho),x_1,\dots,x_{j-1},x_{j+1},\dots,x_i)$$

for $1\leq j\leq i$.

A map which admits an A_n-form is called an A_n-map for $n \geq 1$. An A_1-map is just a map, and by [Iwase and Mimura 1989, p. 195, P8], an A_2-map and an A_3-map are an H-map and an H-map preserving the homotopy associativity, respectively. Moreover, an A_∞-map ϕ is homotopic to a loop map which induces a map between the classifying spaces $B\phi\colon BX \to BY$ with $\Omega(B\phi) \simeq \phi$ (see [Kane 1988, §6-4]).

We next recall the higher homotopy commutativity of H-spaces.

Sugawara [1960] gave a criterion for the classifying space of a topological monoid to have the homotopy type of an H-space. His criterion is a higher homotopy commutativity of the multiplication. Later Williams [1969] considered another type of higher homotopy commutativity which is weaker than the one of Sugawara, and defined C_n-spaces.

In [Hemmi and Kawamoto 2004], we generalized the definition of Williams to the case of A_n-spaces, and defined AC_n-spaces by using the permuto-associahedra $\{\Gamma_i\}_{i\geq 1}$. Let X be an A_n-space with the A_n-form $\{M_i\}_{2\leq i\leq n}$. Then an AC_n-form on X is a family of maps $\{Q_i : \Gamma_i \times X^i \to X\}_{1\leq i\leq n}$ with the relations

$$Q_1(*, x) = x, \tag{3-1}$$

$$\begin{aligned} &Q_i(\epsilon^{(\alpha_1,\dots,\alpha_l)}(\sigma, \tau_1, \dots, \tau_l), x_1, \dots, x_i) \\ &\quad = M_l(\sigma, Q_{t_1}(\tau_1, x_{\alpha_1(1)}, \dots, x_{\alpha_1(t_1)}), \dots, Q_{t_l}(\tau_l, x_{\alpha_l(1)}, \dots, x_{\alpha_l(t_l)})) \end{aligned} \tag{3-2}$$

for a partition $(\alpha_1, \dots, \alpha_l)$ of $\mathbf{i}$ of type $(t_1, \dots, t_l)$ with $l \geq 2$, and

$$Q_i(\tau, x_1, \dots, x_{j-1}, *, x_{j+1}, \dots, x_i) = Q_{i-1}(\omega_j(\tau), x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_i) \tag{3-3}$$

for $1 \leq j \leq i$.

An A_n-space admitting an AC_n-form is called an AC_n-space for $n \geq 1$. By Example 3.2(1) in [Hemmi and Kawamoto 2004], X is an AC_2-space if and only if X is a homotopy commutative H-space. Moreover, if X is a topological monoid, then by Corollary 3.6 of the same work, X is an AC_n-space if and only if X is a C_n-space of Williams [1969].

Williams [1969, Definition 20], also considered the concept of C_n-map between C_n-spaces. We generalize his definition to the case of maps between AC_n-spaces.

Definition 3.1. Let X and Y be AC_n-spaces with the AC_n-forms $\{Q_i\}_{1\leq i\leq n}$ and $\{R_i\}_{1\leq i\leq n}$, respectively. Assume that $\phi\colon X \to Y$ is an A_n-map with the A_n-form $\{F_i\}_{1\leq i\leq n}$. Then an AC_n-form on ϕ is a family of maps

$$\{D_i : I \times \Gamma_i \times X^i \to Y\}_{1\leq i\leq n}$$

with the relations

$$D_1(t, *, x) = \phi(x), \tag{3-4}$$

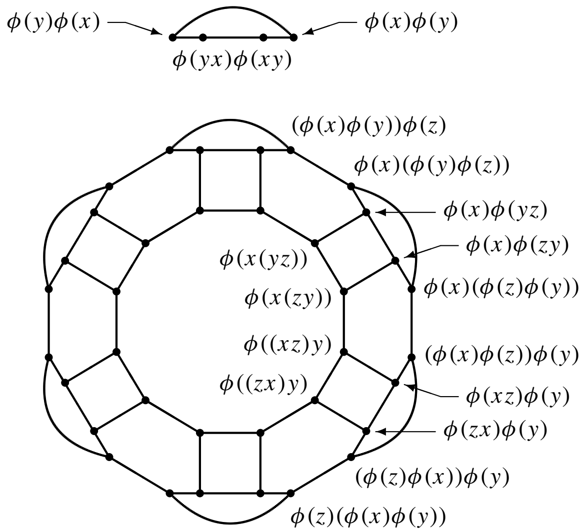


Figure 6. The AC_n-forms on ϕ for $n = 2, 3$.

$$D_i(t, \epsilon^{(\alpha_1,\dots,\alpha_l)}(\sigma, \tau_1, \dots, \tau_l), x_1, \dots, x_i) = N_l(\sigma, D_{t_1}(t, \tau_1, x_{\alpha_1(1)}, \dots, x_{\alpha_1(t_1)}), \dots, D_{t_l}(t, \tau_l, x_{\alpha_l(1)}, \dots, x_{\alpha_l(t_l)})) \tag{3-5}$$

for a partition $(\alpha_1, \dots, \alpha_l)$ of $\mathbf{i}$ of type $(t_1, \dots, t_l)$ with $l \geq 2$,

$$D_i(0, \iota^{(\beta_1,\dots,\beta_m)}(\sigma, \tau_1, \dots, \tau_m), x_1, \dots, x_i) = F_m(\sigma, Q_{u_1}(\tau_1, x_{\beta_1(1)}, \dots, x_{\beta_1(u_1)}), \dots, Q_{u_m}(\tau_m, x_{\beta_m(1)}, \dots, x_{\beta_m(u_m)})) \tag{3-6}$$

for a partition $(\beta_1, \dots, \beta_m)$ of $\mathbf{i}$ of type $(u_1, \dots, u_m)$ with $m \geq 1$,

$$D_i(1, \tau, x_1, \dots, x_i) = R_i(\tau, \phi(x_1), \dots, \phi(x_i)), \tag{3-7}$$

and

$$D_i(t, \tau, x_1, \dots, x_{j-1}, *, x_{j+1}, \dots, x_i) = D_{i-1}(t, \omega_j(\tau), x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_i) \tag{3-8}$$

for $1 \leq j \leq i$.

An A_n-map admitting an AC_n-form is called an AC_n-map for $n \geq 1$. If there is a family of maps $\{D_i\}_{i \geq 1}$ such that $\{D_i\}_{1 \leq i \leq n}$ is an AC_n-form on ϕ for any $n \geq 1$, then ϕ is called an AC_∞-map.

Example 3.2. (1) An AC_2-space is the same as a homotopy commutative H-space by [Hemmi and Kawamoto 2004, Example 3.2(1)]. Then an AC_2-map is a map between AC_2-spaces preserving the homotopy commutativity, and so it is the same as an HC-map of Zabrodsky [1976, p. 62].

(2) Let $\phi\colon X \to Y$ be a homomorphism for topological monoids X, Y. Then ϕ is an AC_n-map if and only if ϕ is a C_n-map of [Williams 1969, Definition 20].

(3) If $\phi\colon X \to Y$ is an H-map, then the loop map $\Omega\phi\colon \Omega X \to \Omega Y$ is an AC_∞-map.

4. Proof of Theorem A

Let $\mathcal{S}_*$ denote the category of pointed and connected topological spaces having the homotopy type of CW-complexes. Assume that $f\colon A \to B$ is a pointed map for $A, B \in \mathcal{S}_*$. According to Dror Farjoun [1996, p. 2, A.1], $Z \in \mathcal{S}_*$ is called f-local if the induced map

$$f^{\#}\colon \mathrm{Map}_*(B, Z) \longrightarrow \mathrm{Map}_*(A, Z)$$

is a homotopy equivalence. In the case that $B = *$ and $f\colon A \to *$ is the constant map, Z is called A-local, that is, the pointed mapping space $\mathrm{Map}_*(A, Z)$ is contractible.

Bousfield [1994, §2] and Dror Farjoun [1996, §1] constructed the A-localization $L_A(X)$ with the universal map $\phi_X\colon X \to L_A(X)$ for $X \in \mathcal{S}_*$ (see also [Chachólski 1996, §14]). By [Farjoun 1996, p. 4, A.4], $L_A(X)$ is A-local, and by [Bousfield 1994, Theorem 2.10(ii)], ϕ_X induces a homotopy equivalence

$$(\phi_X)^{\#}\colon \mathrm{Map}_*(L_A(X), Z) \longrightarrow \mathrm{Map}_*(X, Z) \tag{4-1}$$

for any A-local space Z (see also [Chachólski 1996, Theorem 14.1]).

To prove Theorem A, we first show:

Proposition 4.1. *Let $\phi\colon X \to Y$ be an A_n-map for A_n-spaces X, Y. If X is an AC_n-space and Y is ϕ-local, then Y is an AC_n-space so that ϕ is an AC_n-map.*

Lemma 4.2. *Let $\phi\colon X \to Y$ be a map. If Y is ϕ-local, then we have the homotopy equivalences*

$$(\phi^n)^{\#}\colon \mathrm{Map}_*(Y^n, Y) \longrightarrow \mathrm{Map}_*(X^n, Y) \tag{4-2}$$

$$(\phi^{(n)})^{\#}\colon \mathrm{Map}_*(Y^{(n)}, Y) \longrightarrow \mathrm{Map}_*(X^{(n)}, Y), \tag{4-3}$$

where $Z^{(n)}$ denotes the n-fold smash product of $Z \in \mathcal{S}_*$ for $n \geq 1$.

Proof. We first show (4-2). From the homotopy commutative diagram of fibrations

$$\begin{tikzcd} \mathrm{Map}_*(Y^n, Y) \arrow[r, "(\phi^n)^\#"] \arrow[d] & \mathrm{Map}_*(X^n, Y) \arrow[d] \\ \mathrm{Map}(Y^n, Y) \arrow[r, "(\phi^n)^\#"] \arrow[d, "e"'] & \mathrm{Map}(X^n, Y) \arrow[d, "e'"] \\ Y \arrow[r, equal] & Y, \end{tikzcd}$$

it is sufficient to show that the middle horizontal map is a homotopy equivalence for $n \geq 1$, where e and e' are the evaluation maps at the base points.

We work by induction on n. Since Y is ϕ-local, the result is clear for $n = 1$. Assume that $(\phi^{n-1})^\#\colon \mathrm{Map}(Y^{n-1}, Y) \to \mathrm{Map}(X^{n-1}, Y)$ is a homotopy equivalence. By [Farjoun 1996, p. 5, A.8, e.2], $\mathrm{Map}(Y^{n-1}, Y)$ is ϕ-local. From the homotopy commutative diagram

$$\begin{tikzcd} \mathrm{Map}(Y^n, Y) \arrow[r, "(\phi^n)^\#"] \arrow[d, "\simeq"'] & \mathrm{Map}(X^n, Y) \arrow[d, "\simeq"] \\ \mathrm{Map}(Y, \mathrm{Map}(Y^{n-1}, Y)) \arrow[d, "\simeq"', "\phi^\#"] & \mathrm{Map}(X, \mathrm{Map}(X^{n-1}, Y)) \arrow[d, equal] \\ \mathrm{Map}(X, \mathrm{Map}(Y^{n-1}, Y)) \arrow[r, "((\phi^{n-1})^\#)_\#", "\simeq"'] & \mathrm{Map}(X, \mathrm{Map}(X^{n-1}, Y)), \end{tikzcd}$$

we have that $(\phi^n)^\#\colon \mathrm{Map}(Y^n, Y) \to \mathrm{Map}(X^n, Y)$ is a homotopy equivalence.

In the case of (4-3), by similar arguments to the case of (4-2) and a homotopy equivalence

$$\mathrm{Map}_*(Z \wedge W, U) \simeq \mathrm{Map}_*(Z, \mathrm{Map}_*(W, U))$$

for $Z, W, U \in \mathcal{S}_*$, we have the required conclusion. This completes the proof. □

Lemma 4.3. *Let $\phi\colon Z \to W$ be a homotopy equivalence for $Z, W \in \mathcal{S}_*$, and let (K, L) be a relative CW-complex.*

(1) *If there are maps $f\colon K \to W$ and $g\colon L \to Z$ with $\phi g = f|_L$, then we have a map $h\colon K \to Z$ with $h|_L = g$ and $\phi h \simeq f$ rel L.*

(2) *If $h, k\colon K \to Z$ are maps with $h|_L = k|_L$ and $\phi h \simeq \phi k$ rel L, then $h \simeq k$ rel L.*

Proof of Proposition 4.1. Let $\{M_i\}_{2 \leq i \leq n}$ and $\{N_i\}_{2 \leq i \leq n}$ be the A_n-form on X and Y, respectively. Since $\phi\colon X \to Y$ is an A_n-map, there is an A_n-form $\{F_i\}_{1 \leq i \leq n}$ on ϕ. Moreover, we denote the AC_n-form on X by $\{Q_i\}_{1 \leq i \leq n}$.

We work by induction on n. By (3-1) and (3-4), the result is clear for $n = 1$. Assume that there are AC_{n-1}-forms $\{R_i\}_{1\le i\le n-1}$ and $\{D_i\}_{1\le i\le n-1}$ on Y and ϕ, respectively.

Put $\mathcal{V}_n(Z) = (I\times\partial\Gamma_n \cup \{0\}\times\Gamma_n)\times Z^n$ and $\mathcal{W}_n(Z) = I\times\Gamma_n\times Z^{[n]}$, where $Z^{[n]}$ denotes the n-fold fat wedge of $Z\in\mathcal{S}_*$ given by

$$Z^{[n]} = \{(z_1,\dots,z_n)\in Z^n \mid z_j = * \text{ for some } 1\le j\le n\}.$$

Let $E_n\colon \mathcal{V}_n(X)\cup\mathcal{W}_n(X)\to Y$ be the map defined by

$$\begin{aligned}E_n(t, \epsilon^{(\alpha_1,\dots,\alpha_l)}&(\sigma,\tau_1,\dots,\tau_l), x_1,\dots,x_n)\\ &= N_l(\sigma, D_{t_1}(t,\tau_1,x_{\alpha_1(1)},\dots,x_{\alpha_1(t_1)}),\dots,D_{t_l}(t,\tau_l,x_{\alpha_l(1)},\dots,x_{\alpha_l(t_l)}))\end{aligned}$$

for a partition $(\alpha_1,\dots,\alpha_l)$ of $\mathbf{n}$ of type $(t_1,\dots,t_l)$ with $l\ge 2$,

$$\begin{aligned}E_n(0, \iota^{(\beta_1,\dots,\beta_m)}&(\sigma,\tau_1,\dots,\tau_m), x_1,\dots,x_n)\\ &= F_m(\sigma, Q_{u_1}(\tau_1,x_{\beta_1(1)},\dots,x_{\beta_1(u_1)}),\dots,Q_{u_m}(\tau_m,x_{\beta_m(1)},\dots,x_{\beta_m(u_m)}))\end{aligned}$$

for a partition $(\beta_1,\dots,\beta_m)$ of $\mathbf{n}$ of type $(u_1,\dots,u_m)$ with $m\ge 1$ and

$$E_n(t,\tau,x_1,\dots,x_{j-1},*,x_{j+1},\dots,x_n) = D_{n-1}(t,\omega_j(\tau),x_1,\dots,x_{j-1},x_{j+1},\dots,x_n)$$

for $1\le j\le n$.

Since there is a map $\widetilde{E}_n\colon I\times\Gamma_n\times X^n\to Y$ with $\widetilde{E}_n|_{\mathcal{V}_n(X)\cup\mathcal{W}_n(X)} = E_n$ by the homotopy extension property, we define a map $S_n\colon \Gamma_n\times X^n\to Y$ by

$$S_n(\tau,x_1,\dots,x_n) = \widetilde{E}_n(1,\tau,x_1,\dots,x_n).$$

Let $\gamma_n\colon \Gamma_n\to \mathrm{Map}_*(X^n,Y)_{(\phi^n)^\#(\mu_n)}$ be the adjoint of S_n, where $\mu_n\colon Y^n\to Y$ is the map given by $\mu_n(y_1,\dots,y_n) = (\cdots(y_1y_2)\cdots)y_n$. If a map $\kappa_n\colon \partial\Gamma_n\to \mathrm{Map}_*(Y^n,Y)_{\mu_n}$ is defined by

$$\begin{aligned}\kappa_n(\epsilon^{(\alpha_1,\dots,\alpha_l)}&(\sigma,\tau_1,\dots,\tau_l), y_1,\dots,y_n)\\ &= N_l(\sigma, R_{t_1}(\tau_1,y_{\alpha_1(1)},\dots,y_{\alpha_1(t_1)}),\dots,R_{t_l}(\tau_l,y_{\alpha_l(1)},\dots,y_{\alpha_l(t_l)})),\end{aligned}$$

then $(\phi^n)^\#(\kappa_n) = \gamma_n|_{\partial\Gamma_n}$, and so by (4-2) and Lemma 4.3 (1), we have a map $\lambda_n\colon \Gamma_n\to\mathrm{Map}_*(Y^n,Y)_{\mu_n}$ with $\lambda_n|_{\partial\Gamma_n} = \kappa_n$ and $(\phi^n)^\#(\lambda_n)\simeq\gamma_n$ rel $\partial\Gamma_n$.

To construct a map $R_n\colon \Gamma_n\times Y^n\to Y$ with the relations (3-1)–(3-3), we need to show that the induced map

$$(4\text{-}4)\qquad (\phi^{[n]})^\#\colon \mathrm{Map}_*(Y^{[n]},Y)_{\nu_n}\longrightarrow \mathrm{Map}_*(X^{[n]},Y)_{(\phi^{[n]})^\#(\nu_n)}$$

is a homotopy equivalence, where $\nu_n\colon Y^{[n]}\to Y$ denotes the composite of μ_n with the inclusion $\iota_Y\colon Y^{[n]}\to Y^n$. Since Y is an H-space, it is sufficient to show the same homotopy equivalence on the components of the constant maps. Consider the following homotopy commutative diagram of fibrations:

$$\begin{array}{ccc}
\mathrm{Map}_*(Y^{(n)}, Y)_{C_1} & \xrightarrow[\simeq]{(\phi^{(n)})^\#} & \mathrm{Map}_*(X^{(n)}, Y)_{C_2} \\
{\scriptstyle (\pi_Y)^\#}\downarrow & & \downarrow{\scriptstyle (\pi_X)^\#} \\
\mathrm{Map}_*(Y^{n}, Y)_{c} & \xrightarrow[\simeq]{(\phi^{n})^\#} & \mathrm{Map}_*(X^{n}, Y)_{c} \\
{\scriptstyle (\iota_Y)^\#}\downarrow & & \downarrow{\scriptstyle (\iota_X)^\#} \\
\mathrm{Map}_*(Y^{[n]}, Y)_{c} & \xrightarrow{(\phi^{[n]})^\#} & \mathrm{Map}_*(X^{[n]}, Y)_{c},
\end{array}$$

where $C_1 = \{h : Y^{(n)} \to Y \mid (\pi_Y)^\#(h) \simeq c\}$ and $C_2 = \{k : X^{(n)} \to Y \mid (\pi_X)^\#(k) \simeq c\}$. Since the vertical arrows are fibrations, the bottom horizontal arrow is a homotopy equivalence, which implies (4-4).

Define a map $\rho_n : \Gamma_n \to \mathrm{Map}_*(Y^{[n]}, Y)_{\nu_n}$ by

$$\rho_n(\tau)(y_1, \dots, y_{j-1}, *, y_{j+1}, \dots, y_n) = R_{n-1}(\omega_j(\tau), y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_n)$$

for $1 \le j \le n$. Then $(\phi^{[n]})^\#(\iota_Y)^\#(\lambda_n) \simeq (\phi^{[n]})^\#(\rho_n)$ rel $\partial\Gamma_n$, and so by (4-4) and Lemma 4.3 (2), we have $(\iota_Y)^\#(\lambda_n) \simeq \rho_n$ rel $\partial\Gamma_n$, which implies that there is a map $\psi_n : I \times \Gamma_n \to \mathrm{Map}_*(Y^{[n]}, Y)_{\nu_n}$ with

$$\psi_n(t, \tau) = \begin{cases} (\iota_Y)^\#(\lambda_n)(\tau) & \text{if } (t, \tau) \in \{0\} \times \Gamma_n \cup I \times \partial\Gamma_n, \\ \rho_n(\tau) & \text{if } (t, \tau) \in \{1\} \times \Gamma_n. \end{cases}$$

If a map $G_n : \mathcal{V}_n(Y) \cup \mathcal{W}_n(Y) \to Y$ is given by

$$G_n(t, \tau, y_1, \dots, y_n) = \begin{cases} \lambda_n(\tau)(y_1, \dots, y_n) & \text{if } (t, \tau, y_1, \dots, y_n) \in \mathcal{V}_n(Y), \\ \psi_n(t, \tau)(y_1, \dots, y_n) & \text{if } (t, \tau, y_1, \dots, y_n) \in \mathcal{W}_n(Y), \end{cases}$$

there is an extension $\widetilde{G}_n : I \times \Gamma_n \times Y^n \to Y$ with $\widetilde{G}_n|_{\mathcal{V}_n(Y) \cup \mathcal{W}_n(Y)} = G_n$. Let $R_n : \Gamma_n \times Y^n \to Y$ be the map defined by $R_n(\tau, y_1, \dots, y_n) = \widetilde{G}_n(1, \tau, y_1, \dots, y_n)$. Then R_n satisfies the relations (3-1)–(3-3).

Since $R_n(1_{\Gamma_n} \times \phi^n) \simeq S_n$ rel $\partial\Gamma_n \times X^n$, we have a map $H_n : I \times \Gamma_n \times X^n \to Y$ with $H_n|_{\mathcal{V}_n(X)} = E_n|_{\mathcal{V}_n(X)}$ and $H_n(1, \tau, x_1, \dots, x_n) = R_n(\tau, \phi(x_1), \dots, \phi(x_n))$. Moreover, $H_n|_{\partial(I \times \Gamma_n) \times X^{[n]}} = E_n|_{\partial(I \times \Gamma_n) \times X^{[n]}}$, and so by [Williams 1969, Remark 10], we can choose a map $D_n : I \times \Gamma_n \times X^n \to Y$ with $D_n|_{\partial(I \times \Gamma_n) \times X^n} = H_n|_{\partial(I \times \Gamma_n) \times X^n}$ and $D_n|_{\mathcal{W}_n(X)} = E_n|_{\mathcal{W}_n(X)}$. Then D_n satisfies the relations (3-4)–(3-8), and we have the required conclusion. This completes the proof. $\square$

Let $\phi : X \to Y$ be a homotopy equivalence. Then Y is ϕ-local, and so by Proposition 4.1, we have:

Proposition 4.4. *Let X, Y be A_n-spaces. Assume that $\phi\colon X \to Y$ is an A_n-map which is a homotopy equivalence. If X is an AC_n-space, then Y is an AC_n-space so that ϕ is an AC_n-map.*

Remark 4.5. By Proposition 4.4, the property of being an AC_n-space is an invariant of A_n-homotopy type. This is a generalization of [Williams 1969, Proposition 8, Theorem 9] for C_n-spaces in the category of topological monoids.

Proof of Theorem A. If X is an AC_n-space, then the A-localization $L_A(X)$ is an A_n-space so that the universal map $\phi_X\colon X \to L_A(X)$ is an A_n-map, by [Kawamoto 2002, Theorem 2.1(1)]. Since $L_A(X)$ is ϕ_X-local by (4-1), we have the required conclusion by Proposition 4.1. This completes the proof of Theorem A. □

By [Farjoun 1996, p. 26, E.1], the S^{m+1}-localization $L_{S^{m+1}}(X)$ of X is the same as the m-th stage $P^m(X)$ of the Postnikov system of X for $m \geq 1$, where S^t denotes the t-dimensional sphere for $t \geq 1$. Then by Theorem A, we have:

Corollary 4.6. *Let $n \geq 1$. If X is an AC_n-space, then the m-th stage $P^m(X)$ of the Postnikov system of X is an AC_n-space so that the projection $\rho_m\colon X \to P^m(X)$ is an AC_n-map for $m \geq 1$.*

Let $A \in \mathcal{S}_*$. Dror Farjoun [1996, §2] constructed the A-colocalization $CW_A(X)$ with the universal map $\psi_X\colon CW_A(X) \to X$ for $X \in \mathcal{S}_*$ (see also [Chachólski 1996, §7]).

Theorem 4.7. *Let $A \in \mathcal{S}_*$. If X is an AC_n-space, then the A-colocalization $CW_A(X)$ is an AC_n-space so that the universal map $\psi_X\colon CW_A(X) \to X$ is an AC_n-map.*

Let $f\colon Z \to W$ be a pointed map for $Z, W \in \mathcal{S}_*$. According to [Farjoun 1996, p. 39, A.1], f is called an A-equivalence if the induced map

$$f_\#\colon \mathrm{Map}_*(A, Z) \longrightarrow \mathrm{Map}_*(A, W)$$

is a homotopy equivalence. From the proof of [Farjoun 1996, p. 53, E.1], the universal map $\psi_X\colon CW_A(X) \to X$ is a $CW_A(X)$-equivalence (see also [Chachólski 1996, p. 614]), and so we can prove Theorem 4.7 from the following result:

Proposition 4.8. *Let $\phi\colon X \to Y$ be an A_n-map for A_n-spaces X, Y. If Y is an AC_n-space and ϕ is an X-equivalence, then X is an AC_n-space so that ϕ is an AC_n-map.*

Proposition 4.8 is proved by similar arguments to the proof of Proposition 4.1, and so we omit the proof. In the proof of Proposition 4.8, we need the next lemma instead of Lemma 4.2:

Lemma 4.9. *Let $\phi\colon X \to Y$ be a map. If ϕ is an X-equivalence, then we have the homotopy equivalences*

$$\phi_\#\colon \mathrm{Map}_*(X^n, X) \longrightarrow \mathrm{Map}_*(X^n, Y),$$
$$\phi_\#\colon \mathrm{Map}_*(X^{(n)}, X) \longrightarrow \mathrm{Map}_*(X^{(n)}, Y)$$

for $n \geq 1$.

Proof. By [Farjoun 1996, p. 46, D.2.2],

$$\mathscr{E}(\phi) = \{A \in \mathscr{S}_* \mid \phi \text{ is an } A\text{-equivalence}\}$$

is a closed class.

Since ϕ is an X-equivalence, $X \in \mathscr{E}(\phi)$. If $A, B \in \mathscr{E}(\phi)$, then by [Farjoun 1996, p. 52, D.16], the product $A \times B \in \mathscr{E}(\phi)$. Since the wedge sum $A \vee B$ is represented by a homotopy colimit, we have $A \vee B \in \mathscr{E}(\phi)$ by [Farjoun 1996, p. 45, D.1] (see also [Chachólski 1996, Proposition 4.2]), and so $A \wedge B \in \mathscr{E}(\phi)$ by [Farjoun 1996, p. 45, D.1, 3,4]. From these properties, we have X^n, $X^{(n)} \in \mathscr{E}(\phi)$ for $n \geq 1$, which implies the required conclusion. This completes the proof. □

Dror Farjoun [1996, p. 39, A.3] proved that the S^{m+1}-colocalization $CW_{S^{m+1}}(X)$ of $X \in \mathscr{S}_*$ is the same as the m-connected covering $X\langle m\rangle$ of X for $m \geq 1$, and so by Theorem 4.7, we have:

Corollary 4.10. *Let $n \geq 1$. If X is an AC_n-space, then the m-connected covering $X\langle m\rangle$ of X is an AC_n-space so that the inclusion $\iota_m\colon X\langle m\rangle \to X$ is an AC_n-map for $m \geq 1$.*

Remark 4.11. In [Hemmi and Kawamoto 2004], we have shown that the universal covering inherits the property of being an AC_n-space. Corollary 4.10 is a generalization of Lemma 3.9 of that work to the case of any $m \geq 1$.

5. Proofs of Theorems B and C

Theorem 5.1 [Castellana et al. 2007, Theorem 7.3]. *Let p be a prime. If X is a connected H-space whose* mod *p cohomology $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathscr{A}_p^*$, then there is an H-fibration*

$$F(\phi_X) \longrightarrow X \xrightarrow{\phi_X} L_{B\mathbb{Z}/p}(X), \tag{5-1}$$

where $L_{B\mathbb{Z}/p}(X)$ is a connected $\mathbb{Z}/p$-finite H-space and $F(\phi_X)$ is mod *p homotopy equivalent to a Postnikov H-space.*

Remark 5.2. In Theorem 5.1, if $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathbb{Z}/p$, then $F(\phi_X)$ is mod p homotopy equivalent to a finite product of $(\mathbb{C}P^\infty)^\wedge_p$s and $B\mathbb{Z}/p^i$s for $i \geq 1$ by [Broto et al. 2001, Theorem 1.2, Theorem 1.3].

Proof of Theorem B. By Theorem A and Theorem 5.1, $L_{B\mathbb{Z}/p}(X)$ is a connected $\mathbb{Z}/p$-finite AC_p-space so that the universal map $\phi_X\colon X \to L_{B\mathbb{Z}/p}(X)$ is an AC_p-map. Then $H^*(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$ is an exterior algebra generated by odd dimensional generators by [Kane 1988, §12-3, Corollaries A and B]. Let Z be the universal covering of $L_{B\mathbb{Z}/p}(X)$. Then there is an H-fibration

$$Z \longrightarrow L_{B\mathbb{Z}/p}(X) \longrightarrow K(\pi_1(L_{B\mathbb{Z}/p}(X)), 1),$$

where $K(\pi_1(L_{B\mathbb{Z}/p}(X)), 1)$ has the mod p homotopy type of a torus. Since Z is a simply connected $\mathbb{Z}/p$-finite AC_p-space by [Hemmi and Kawamoto 2004, Lemma 3.9] and [Kane 1988, §3-1, Theorem B], we have $\widetilde{H}^*(Z; \mathbb{Z}/p) = 0$ by [Hemmi and Kawamoto 2004, Theorem A (1)] and [Hemmi 1991, Theorem 1.1]. Then $L_{B\mathbb{Z}/p}(X)$ has the mod p homotopy type of a torus, and so by Theorem 5.1, X is mod p homotopy equivalent to a Postnikov H-space. This completes the proof of Theorem B. $\square$

Remark 5.3. From Theorem B, we have the mod p torus theorem stated in [Hemmi and Kawamoto 2004, Corollary 1.1] since a result of McGibbon and Neisendorfer [1984] on a conjecture of Serre implies that a connected Postnikov H-space which is also $\mathbb{Z}/p$-finite has the mod p homotopy type of a torus.

The proof of Corollary 1.3 is given as follows:

Proof of Corollary 1.3. By Theorem B, X is mod p homotopy equivalent to a Postnikov H-space. Put $Y = X^\wedge_p$. Then, by [Bousfield 2001, Theorem 7.2], $L_{K(n)_*}(Y)$ is homotopy equivalent to the $(n+1)$-st stage $\widetilde{P}^{n+1}(Y)$ of the modified Postnikov system of Y given by

$$\pi_j(\widetilde{P}^{n+1}(Y)) \cong \begin{cases} \pi_j(Y) & \text{for } 1 \le j \le n, \\ \pi_{n+1}(Y)/T_{n+1}(p) & \text{for } j = n+1, \\ 0 & \text{for } j > n+1, \end{cases} \tag{5-2}$$

where $T_{n+1}(p)$ denotes the p-torsion subgroup of $\pi_{n+1}(Y)$. Since $\Omega^n L_{K(n)_*}(Y)$ is $B\mathbb{Z}/p$-local by (5-2), there is a map $f\colon L_{\Sigma^n B\mathbb{Z}/p}(Y) \to L_{K(n)_*}(Y)$ with $f\phi_Y \simeq \kappa_Y$, where $\kappa_Y\colon Y \to L_{K(n)_*}(Y)$ denotes the universal map for the $K(n)_*$-localization of Y.

Since $\Omega^n L_{\Sigma^n B\mathbb{Z}/p}(Y)$ is $B\mathbb{Z}/p$-local, there is an H-fibration

$$F(\phi_{L_{\Sigma^n B\mathbb{Z}/p}(Y)}) \longrightarrow L_{\Sigma^n B\mathbb{Z}/p}(Y) \xrightarrow{\phi_{L_{\Sigma^n B\mathbb{Z}/p}(Y)}} L_{B\mathbb{Z}/p}(L_{\Sigma^n B\mathbb{Z}/p}(Y)),$$

by [Castellana et al. 2007, Theorem 3.2], where $F(\phi_{L_{\Sigma^n B\mathbb{Z}/p}(Y)})$ is mod p homotopy equivalent to a Postnikov H-space which satisfies that $\pi_{n+1}(F(\phi_{L_{\Sigma^n B\mathbb{Z}/p}(Y)})^\wedge_p)$ has no p-torsion and $\pi_j(F(\phi_{L_{\Sigma^n B\mathbb{Z}/p}(Y)})^\wedge_p) = 0$ for $j > n+1$.

By [Farjoun 1996, p. 139, B.6], $L_{B\mathbb{Z}/p}(L_{\Sigma^n B\mathbb{Z}/p}(Y)) \simeq L_{B\mathbb{Z}/p}(Y)$, and by the proof of Theorem B, we have that $L_{B\mathbb{Z}/p}(Y)$ has the mod p homotopy type of a torus. Then $L_{\Sigma^n B\mathbb{Z}/p}(Y)^\wedge_p$ is $K(n)_*$-local by (5-2), and so there is a map

$$g\colon L_{K(n)_*}(Y) \to L_{\Sigma^n B\mathbb{Z}/p}(Y)^\wedge_p$$

with $g\kappa_Y \simeq (\phi_Y)^\wedge_p$. From the universality of the localizations we see that $L_{K(n)_*}(Y)$ is mod p homotopy equivalent to $L_{\Sigma^n B\mathbb{Z}/p}(Y)$. This completes the proof. □

To prove Theorem C, we need a lemma:

Lemma 5.4. *Let p be an odd prime. Assume that X is a connected H-space whose* mod p *cohomology $H^*(X;\mathbb{Z}/p)$ is finitely generated as an algebra over $\mathbb{Z}/p$. If $x \in QH^{2p^t}(X;\mathbb{Z}/p)$ is a generator of infinite height with $t \geq 2$, then $\mathcal{P}^1\beta(x) \neq 0$ in $QH^{2p^t+2p-1}(X;\mathbb{Z}/p)$ or there is a generator $y \in QH^{2p^{t-1}+1}(X;\mathbb{Z}/p)$ with $\mathcal{P}^{p^{t-1}}(y) = \beta(x) \neq 0$ in $QH^{2p^t+1}(X;\mathbb{Z}/p)$.*

Proof. Let $\widetilde{X}$ be the universal covering of X. Then there is an H-fibration

$$\widetilde{X} \xrightarrow{\ \iota\ } X \xrightarrow{\ \rho\ } K(\pi_1(X),1),$$

where $K(\pi_1(X),1)$ has the mod p homotopy type of a finite product of $(S^1)^\wedge_p$s and $B\mathbb{Z}/p^i$s for $i \geq 1$. According to [Browder 1959], the mod p cohomology $H^*(\widetilde{X};\mathbb{Z}/p)$ is finitely generated as an algebra over $\mathbb{Z}/p$ and

$$\iota^*\colon QH^s(X;\mathbb{Z}/p) \to QH^s(\widetilde{X};\mathbb{Z}/p) \tag{5-3}$$

is an isomorphism if $s \neq 2, 2p^j-1$ for $j \geq 1$.

Recall that the mod p cohomology of $B^2\mathbb{Z}/p$ is given by

$$H^*(B^2\mathbb{Z}/p;\mathbb{Z}/p) \cong \mathbb{Z}/p[u, \beta\mathcal{P}^1\beta(u), \dots, \beta\mathcal{P}^{\Delta_t}\beta(u), \dots] \\ \otimes \Lambda(\beta(u), \mathcal{P}^1\beta(u), \dots, \mathcal{P}^{\Delta_t}\beta(u), \dots),$$

where $u \in QH^2(B^2\mathbb{Z}/p;\mathbb{Z}/p)$ denotes the generator and $\mathcal{P}^{\Delta_t} = \mathcal{P}^{p^t}\cdots\mathcal{P}^1$ for $t \geq 0$. Let $\tilde{x} = \iota^*(x) \in QH^{2p^t}(\widetilde{X};\mathbb{Z}/p)$. By [Crespo 2001, Theorem 2.10, Proposition 5.7], there is an H-space Y and an H-fibration

$$\widetilde{X} \longrightarrow Y \longrightarrow B^2\mathbb{Z}/p \tag{5-4}$$

such that $\tau(\tilde{x}) = \mathcal{P}^{\Delta_{t-1}}\beta(u) \in H^{2p^t+1}(B^2\mathbb{Z}/p;\mathbb{Z}/p)$ and $\tau(\beta(\tilde{x})) = \beta\mathcal{P}^{\Delta_{t-1}}\beta(u) \in H^{2p^t+2}(B^2\mathbb{Z}/p;\mathbb{Z}/p)$ in the spectral sequence associated to the H-fibration (5-4), where $\tau\colon QH^s(\widetilde{X};\mathbb{Z}/p) \to H^{s+1}(B^2\mathbb{Z}/p;\mathbb{Z}/p)$ denotes the transgression of the spectral sequence for $s \geq 2$. Then by [Crespo 2001, Theorem 1.5], there is a generator $\tilde{y} \in QH^{2p^{t-1}+1}(\widetilde{X};\mathbb{Z}/p)$ with

$$\tau(\tilde{y}) = \beta\mathcal{P}^{\Delta_{t-2}}\beta(u) \in H^{2p^{t-1}+2}(B^2\mathbb{Z}/p;\mathbb{Z}/p) \tag{5-5}$$

or there is a generator $\tilde{z} \in QH^{2p^t+2p-1}(\widetilde{X}; \mathbb{Z}/p)$ with

$$\tau(\tilde{z}) = (\beta\mathcal{P}^{\Delta_{t-2}}\beta(u))^p \in H^{2p^t+2p}(B^2\mathbb{Z}/p; \mathbb{Z}/p). \tag{5-6}$$

If we have (5-5), then it follows that $\mathcal{P}^{p^{t-1}}(\tilde{y}) = \beta(\tilde{x})$ in $QH^{2p^t+1}(\widetilde{X}; \mathbb{Z}/p)$ by the choice of the generators in [Crespo 2001, p. 126] since $\mathcal{P}^{p^{t-1}}(\beta\mathcal{P}^{\Delta_{t-2}}\beta(u)) = \beta\mathcal{P}^{\Delta_{t-1}}\beta(u)$ by Lemma 3.2 of the same reference. Choose $y \in QH^{2p^{t-1}+1}(X; \mathbb{Z}/p)$ with $\iota^*(y) = \tilde{y}$. Then $\mathcal{P}^{p^{t-1}}(y) = \beta(x)$ in $QH^{2p^t+1}(X; \mathbb{Z}/p)$ by (5-3).

In the case of (5-6), since $\mathcal{P}^1(\beta\mathcal{P}^{\Delta_{t-1}}\beta(u)) = (\beta\mathcal{P}^{\Delta_{t-2}}\beta(u))^p$ by [Crespo 2001, Lemma 3.3], we have $\mathcal{P}^1\beta(\tilde{x}) = \tilde{z}$ in $QH^{2p^t+2p-1}(\widetilde{X}; \mathbb{Z}/p)$. Then by (5-3), we have $\mathcal{P}^1\beta(x) \neq 0$ in $QH^{2p^t+2p-1}(X; \mathbb{Z}/p)$. This completes the proof. $\square$

Proof of Theorem C. By Theorem A, [Kawamoto 2002, Theorem 2.1(1)] and Theorem 5.1, we have that $L_{B\mathbb{Z}/p}(X)$ is a connected $\mathbb{Z}/p$-finite A_p-space admitting an AC_n-form with $n > (p-1)/2$.

If $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathcal{A}_p^*$ and the operations $\mathcal{P}^j$ act on $QH^*(X; \mathbb{Z}/p)$ trivially for $j \geq 1$, then we see that $H^*(X; \mathbb{Z}/p)$ is finitely generated as an algebra over $\mathbb{Z}/p$, and so by Remark 5.2, $F(\phi_X)$ is mod p homotopy equivalent to a finite product of $(\mathbb{C}P^\infty)^\wedge_p$s and $B\mathbb{Z}/p^i$s for $i \geq 1$.

Consider the spectral sequence associated to the H-fibration (5-1) whose E_2-term is given by

$$E_2^{*,*} \cong H^*(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p) \otimes H^*(F(\phi_X); \mathbb{Z}/p).$$

Let us show that the spectral sequence collapses. If $w \in QH^1(F(\phi_X); \mathbb{Z}/p)$ is a generator, then $d_2(w) \in PH^2(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$ by [Kane 1988, §1-6], where PA denotes the primitive module of A. By [Kane 1988, §12-3, Corollary B], $H^*(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$ is an exterior algebra generated by odd dimensional generators. Since $PH^2(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p) = 0$, we have $d_2(w) = 0$. Assume that there is a generator $u \in QH^2(F(\phi_X); \mathbb{Z}/p)$ with $d_3(u) = v \neq 0$ in $QH^3(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$. Then $d_3(u^p) = \mathcal{P}^1(v) \neq 0$ in $QH^{2p+1}(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$ by [Hemmi and Kawamoto 2007, Theorem A (2)], and so by computing the spectral sequence, we have a generator $x \in QH^{2p^t}(X; \mathbb{Z}/p)$ with $t \geq 2$. Since the operations $\mathcal{P}^j$ act on $QH^*(X; \mathbb{Z}/p)$ trivially for $j \geq 1$, we have a contradiction by Lemma 5.4. Then the spectral sequence collapses, and so we have

$$H^*(X; \mathbb{Z}/p) \cong H^*(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p) \otimes H^*(F(\phi_X); \mathbb{Z}/p).$$

Since the operations $\mathcal{P}^j$ act on $QH^*(X; \mathbb{Z}/p)$ trivially for $j \geq 1$, they also act on $QH^*(L_{B\mathbb{Z}/p}(X); \mathbb{Z}/p)$ trivially, which implies that $L_{B\mathbb{Z}/p}(X)$ has the mod p homotopy type of a torus by [Hemmi and Kawamoto 2007, Theorem B] and Remark 5.5. Then there is a map $\zeta : L_{B\mathbb{Z}/p}(X) \times F(\phi_X) \to X$ which induces an isomorphism on the mod p cohomology, and so ζ is a mod p homotopy equivalence; compare [Mimura and Toda 1991, p. 157, Corollary 1.6]. This completes the proof of Theorem C. □

Remark 5.5. In [Hemmi and Kawamoto 2007], all spaces are assumed to be localized at p in the sense of [Bousfield and Kan 1972]. However, the proof of Theorem B in our paper with Hemmi is also available for $\mathbb{Z}/p$-finite A_p-spaces, even if they are not localized at p.

References

[Aguadé and Smith 1986] J. Aguadé and L. Smith, "On the mod p torus theorem of John Hubbuck", *Math. Z.* **191**:2 (1986), 325–326. MR 87e:57044 Zbl 0591.55003

[Bousfield 1994] A. K. Bousfield, "Localization and periodicity in unstable homotopy theory", *J. Amer. Math. Soc.* **7**:4 (1994), 831–873. MR 95c:55010 Zbl 0839.55008

[Bousfield 2001] A. K. Bousfield, "On the telescopic homotopy theory of spaces", *Trans. Amer. Math. Soc.* **353**:6 (2001), 2391–2426. MR 2001k:55030 Zbl 0971.55016

[Bousfield and Kan 1972] A. K. Bousfield and D. M. Kan, *Homotopy limits, completions and localizations*, Lecture Notes in Math. **304**, Springer, Berlin, 1972. MR 51 #1825 Zbl 0259.55004

[Broto et al. 2001] C. Broto, J. A. Crespo, and L. Saumell, "Non-simply connected H-spaces with finiteness conditions", *Math. Proc. Cambridge Philos. Soc.* **130** (2001), 475–488. MR 2002b:55015 Zbl 1002.55007

[Browder 1959] W. Browder, "The cohomology of covering spaces of H-spaces", *Bull. Amer. Math. Soc.* **65** (1959), 140–141. MR 22 #1891 Zbl 0088.38802

[Castellana et al. 2006] N. Castellana, J. A. Crespo, and J. Scherer, "On the cohomology of highly connected covers of finite complexes", preprint, 2006. math.AT/0607201

[Castellana et al. 2007] N. Castellana, J. A. Crespo, and J. Scherer, "Deconstructing Hopf spaces", *Invent. Math.* **167**:1 (2007), 1–18. MR MR2264802 Zbl 05118881

[Chachólski 1996] W. Chachólski, "On the functors CW_A and P_A", *Duke Math. J.* **84**:3 (1996), 599–631. MR 97i:55023 Zbl 0873.55014

[Crespo 2001] J. A. Crespo, "Structure of mod p H-spaces with finiteness conditions", pp. 103–130 in *Cohomological methods in homotopy theory* (Bellaterra, 1998), edited by J. Aguadé et al., Progr. Math. **196**, Birkhäuser, Basel, 2001. MR 2002g:55016 Zbl 0990.55004

[Farjoun 1996] E. D. Farjoun, *Cellular spaces, null spaces and homotopy localization*, Lecture Notes in Math. **1622**, Springer, Berlin, 1996. MR 98f:55010 Zbl 0842.55001

[Hemmi 1991] Y. Hemmi, "Higher homotopy commutativity of H-spaces and the mod p torus theorem", *Pacific J. Math.* **149**:1 (1991), 95–111. MR 92a:55010 Zbl 0691.55007

[Hemmi and Kawamoto 2004] Y. Hemmi and Y. Kawamoto, "Higher homotopy commutativity of H-spaces and the permuto-associahedra", *Trans. Amer. Math. Soc.* **356**:10 (2004), 3823–3839. MR 2005c:55020 Zbl 1064.55005

[Hemmi and Kawamoto 2007] Y. Hemmi and Y. Kawamoto, "Higher homotopy commutativity and cohomology of finite H-spaces", pp. 167–186 in *Proceedings of the Nishida Fest* (Kinosaki, 2003), edited by M. Ando et al., Geom. Topol. Monogr. **10**, Mathematical Science Publishers, Berkeley, CA, 2007.

[Hubbuck 1969] J. R. Hubbuck, "On homotopy commutative H-spaces", *Topology* **8** (1969), 119–126. MR 38 #6592 Zbl 0176.21301

[Iriye and Kono 1985] K. Iriye and A. Kono, "Mod p retracts of G-product spaces", *Math. Z.* **190**:3 (1985), 357–363. MR 88a:55017 Zbl 0554.55004

[Iwase and Mimura 1989] N. Iwase and M. Mimura, "Higher homotopy associativity", pp. 193–220 in *Algebraic topology* (Arcata, CA, 1986), edited by G. Carlsson et al., Lecture Notes in Math. **1370**, Springer, Berlin, 1989. MR 90f:55017 Zbl 0677.55010

[Kane 1988] R. M. Kane, *The homology of Hopf spaces*, Math. Library **40**, North-Holland, Amsterdam, 1988. MR 90f:55018 Zbl 0651.55001

[Kapranov 1993] M. M. Kapranov, "The permutoassociahedron, Mac Lane's coherence theorem and asymptotic zones for the KZ equation", *J. Pure Appl. Algebra* **85**:2 (1993), 119–142. MR 94b:52017 Zbl 0812.18003

[Kawamoto 1999] Y. Kawamoto, "Loop spaces of H-spaces with finitely generated cohomology", *Pacific J. Math.* **190**:2 (1999), 311–328. MR 2000i:55027 Zbl 1011.55006

[Kawamoto 2002] Y. Kawamoto, "Higher homotopy commutativity of H-spaces with finitely generated cohomology", *Pacific J. Math.* **204**:1 (2002), 145–161. MR 2003d:55010 Zbl 1053.55005

[Lin 1985] J. P. Lin, "A cohomological proof of the torus theorem", *Math. Z.* **190**:4 (1985), 469–476. MR 87c:55009 Zbl 0563.55007

[Lin 1994] J. P. Lin, "Loops of H-spaces with finitely generated cohomology rings", *Topology Appl.* **60**:2 (1994), 131–152. MR 95m:55022 Zbl 0837.55012

[Lin 1995] J. P. Lin, "H-spaces with finiteness conditions", pp. 1095–1141 in *Handbook of algebraic topology*, North-Holland, Amsterdam, 1995. MR 97c:55017 Zbl 0941.55002

[Lin and Williams 1991] J. P. Lin and F. Williams, "Homotopy-commutative H-spaces", *Proc. Amer. Math. Soc.* **113**:3 (1991), 857–865. MR 92b:55011 Zbl 0758.55008

[McGibbon 1984] C. A. McGibbon, "Homotopy commutativity in localized groups", *Amer. J. Math.* **106**:3 (1984), 665–687. MR 86a:55011 Zbl 0574.55004

[McGibbon 1989] C. A. McGibbon, "Higher forms of homotopy commutativity and finite loop spaces", *Math. Z.* **201**:3 (1989), 363–374. MR 90f:55019 Zbl 0682.55006

[McGibbon and Neisendorfer 1984] C. A. McGibbon and J. A. Neisendorfer, "On the homotopy groups of a finite-dimensional space", *Comment. Math. Helv.* **59**:2 (1984), 253–257. MR 86b:55015 Zbl 0538.55010

[Mimura and Toda 1991] M. Mimura and H. Toda, *Topology of Lie groups*, vol. I, Transl. Math. Monographs **91**, Amer. Math. Soc., Providence, RI, 1991. MR 92h:55001 Zbl 0757.57001

[Reiner and Ziegler 1994] V. Reiner and G. M. Ziegler, "Coxeter-associahedra", *Mathematika* **41**:2 (1994), 364–393. MR 95m:52023 Zbl 0822.52007

[Slack 1991] M. Slack, *A classification theorem for homotopy commutative H-spaces with finitely generated mod 2 cohomology rings*, Mem. Amer. Math. Soc. **449**, Amer. Math. Soc., Providence, RI, 1991. MR 92k:55015 Zbl 0755.55003

[Stasheff 1963] J. D. Stasheff, "Homotopy associativity of H-spaces, I, II", *Trans. Amer. Math. Soc.* **108** (1963), 275–292, 293–312. MR 28 #1623 Zbl 0114.39402

[Stasheff 1970] J. Stasheff, *H-spaces from a homotopy point of view*, Lecture Notes in Math. **161**, Springer, Berlin, 1970. MR 42 #5261 Zbl 0205.27701

[Sugawara 1957] M. Sugawara, "A condition that a space is group-like", *Math. J. Okayama Univ.* **7** (1957), 123–149. MR 20 #3546 Zbl 0091.37201

[Sugawara 1960] M. Sugawara, "On the homotopy-commutativity of groups and loop spaces", *Mem. Coll. Sci. Univ. Kyoto Ser. A Math.* **33** (1960), 257–269. MR 22 #11394 Zbl 0113.16903

[Williams 1969] F. D. Williams, “Higher homotopy-commutativity”, *Trans. Amer. Math. Soc.* **139** (1969), 191–206. MR 39 #2163 Zbl 0185.27103

[Zabrodsky 1976] A. Zabrodsky, *Hopf spaces*, Math. Studies **22**, North-Holland, Amsterdam, 1976. MR 55 #13416 Zbl 0339.55009

[Ziegler 1995] G. M. Ziegler, *Lectures on polytopes*, Graduate Texts in Math. **152**, Springer, New York, 1995. MR 96a:52011 Zbl 0823.52002

Received January 11, 2006. Revised December 17, 2006.

YUSUKE KAWAMOTO
DEPARTMENT OF MATHEMATICS
NATIONAL DEFENSE ACADEMY
YOKOSUKA 239-8686
JAPAN

yusuke@nda.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

CONDUCTORS AND NEWFORMS FOR SL(2)

JOSHUA M. LANSKY AND A. RAGHURAM

In this paper we develop a theory of newforms for $\mathrm{SL}_2(F)$ where F is a nonarchimedean local field whose residue characteristic is odd. This is analogous to results of Casselman for $\mathrm{GL}_2(F)$ and Jacquet, Piatetski-Shapiro, and Shalika for $\mathrm{GL}_n(F)$. To a representation π of $\mathrm{SL}_2(F)$ we attach an integer $c(\pi)$ that we call the conductor of π. The conductor of π depends only on the L-packet Π containing π. It is shown to be equal to the conductor of a minimal representation of $\mathrm{GL}_2(F)$ determining the L-packet Π. A newform is a vector in π which is essentially fixed by a congruence subgroup of level $c(\pi)$. For $\mathrm{SL}_2(F)$ we show that our newforms are always test vectors for some standard Whittaker functionals, and, in doing so, we give various explicit formulae for newforms.

1. Introduction

To introduce the main theme of this paper we recall the following theorem of Casselman [1973]. Let F be a nonarchimedean local field whose ring of integers is $\mathcal{O}_F$. Let $\mathcal{P}_F$ be the maximal ideal of $\mathcal{O}_F$. Let ψ_F be a nontrivial additive character of F which is normalized so that the maximal fractional ideal on which it is trivial is $\mathcal{O}_F$.

Theorem (Casselman). *Let (π, V) be an irreducible admissible infinite-dimensional representation of* $\mathrm{GL}_2(F)$. *Let ω_π denote the central character of π. Let*

$$\Gamma(m)=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in \mathrm{GL}_2(\mathcal{O}_F) : c\equiv 0 \pmod{\mathcal{P}_F^m}\right\}.$$

Let

$$V_m=\left\{v\in V : \pi\left(\begin{pmatrix} a & b\\ c & d\end{pmatrix}\right)v=\omega_\pi(d)v,\quad \forall \begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\Gamma(m)\right\}.$$

(i) *There exists a nonnegative integer m such that $V_m\neq(0)$. If $c(\pi)$ denotes the least nonnegative integer m with this property then the epsilon factor*

MSC2000: primary 22E50; secondary 22E35, 11S37, 11S40.
Keywords: conductor, newform, SL(2).

$\epsilon(s, \pi, \psi_F)$ of π is up to a constant multiple of the form $q^{-c(\pi)s}$. (Here q is the cardinality of the residue field of F.)

(ii) *For all $m \geq c(\pi)$ we have* $\dim V_m = m - c(\pi) + 1$.

The assertion $\dim V_{c(\pi)} = 1$ is sometimes referred to as *multiplicity one for newforms*, and the unique vector (up to scalars) in $V_{c(\pi)}$ is called the *newform* for π. This is closely related to the classical Atkin–Lehner theory of newforms for holomorphic cusp forms on the upper half-plane [Casselman 1973]. When $c(\pi) = 0$, π is a spherical representation and the newform is nothing but the spherical vector.

Newforms play an important role in the theory of automorphic forms. We cite two examples to illustrate this. First, the zeta integral corresponding to the newform is exactly the local L-factor associated to π (see [Jacquet et al. 1981] for instance). In addition, newforms frequently play the role of test vectors for interesting linear forms associated to π. For example, the newform is a test vector for an appropriate Whittaker linear functional. In showing this, explicit formulae for newforms are quite often needed. For instance, if π is a supercuspidal representation which is realized in its Kirillov model then the newform is the characteristic function of the unit group $\mathbb{O}_F^\times$. This observation is implicit in [Casselman 1973] and is explicitly stated and proved in [Shimizu 1977]. Since the Whittaker functional on the Kirillov model is given by evaluating functions at $1 \in F^*$, we get in particular that the functional is nonzero on the newform. In a related vein, it is shown in [Gross and Prasad 1991] that test vectors for trilinear forms for representations of $\mathrm{GL}_2(F)$ are often built from newforms. (See also [Schmidt 2002], where many of these results are documented.)

As far as we know, the only other p-adic groups for which there is a theory of newforms are $\mathrm{GL}_n(F)$ [Jacquet et al. 1981]; $\mathrm{GL}_2(D)$ for a p-adic division algebra D, [Prasad and Raghuram 2000]; and $\mathrm{GSp}_4(F)$ (unpublished work of Brooks Roberts and Ralf Schmidt). *In this paper, we propose a theory of newforms and conductors for* $\mathrm{SL}_2(F)$.

Let $G = \mathrm{SL}_2(F)$ where F is a nonarchimedean local field with odd residue characteristic. Crucial to our study of newforms are certain filtrations of maximal compact subgroups of G. Let $K = K_0 = \mathrm{SL}_2(\mathbb{O}_F)$. Let $K' = K_0' = \alpha^{-1} K_0 \alpha$ where $\alpha = \left(\begin{smallmatrix} \varpi_F & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then K_0 and K_0' are, up to conjugacy, the two maximal compact subgroups of $\mathrm{SL}_2(F)$. We define filtrations of these maximal compact subgroups as follows. For m an integer ≥ 1, let

$$K_m = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{O}_F) : c \equiv 0 \pmod{\mathcal{P}_F^m} \right\} \quad \text{and} \quad K_m' = \alpha^{-1} K_m \alpha.$$

Let (π, V) be an irreducible admissible infinite-dimensional representation of G. Let ω_π be the central character of π, i.e., the character of $\{\pm 1\}$ such that

$$\pi\left(\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right) = \omega_\pi(-1) 1_V.$$

Let η be any character of $\mathbb{O}_F^\times$ such that $\eta(-1) = \omega_\pi(-1)$. Let $c(\eta)$ denote the conductor of η. For any $m \geq c(\eta)$, η gives a character of K_m and K'_m defined by $\eta\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right) = \eta(d)$. We define for $m \geq 0$

$$\pi_\eta^{K_m} := \left\{ v \in V : \pi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) v = \eta(d) v, \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K_m \right\}.$$

Note that $\pi_\eta^{K_m} = (0)$ if $m < c(\eta)$. The space $\pi_\eta^{K'_m}$ is defined analogously. We define the *η-conductor* $c_\eta(\pi)$ of π as

$$c_\eta(\pi) = \min\{m \geq 0 : \pi_\eta^{K_m} \neq (0) \ \text{ or } \ \pi_\eta^{K'_m} \neq (0)\}.$$

We define the *conductor* $c(\pi)$ of π by $c(\pi) = \min\{c_\eta(\pi) : \eta\}$, where η runs over characters of F^* such that $\eta(-1) = \omega_\pi(-1)$.

We deal with the following basic issues in this paper.

(i) Given an irreducible representation π, we determine its conductor $c(\pi)$. A very easy consequence (almost built into the definition) is that the conductor depends only on the L-packet containing π.

(ii) We identify the conductor with some other invariants associated to the representation. For instance, for $\mathrm{SL}_2(F)$ we show that the conductor of π is same as the conductor of a minimal representation of $\mathrm{GL}_2(F)$ determining the L-packet containing π. We also determine an explicit relation between our conductor and the notion of depth due to Moy and Prasad (see Section 3.4).

(iii) We determine the growth of the space $\dim V_\eta^{K_m}$ as a function of m. This question is analogous to (ii) of Casselman's theorem quoted above. Computing such dimensions is of importance in local level raising. See [Mann 2001].

(iv) We address the question of whether there is a multiplicity one result for newforms. It turns out that quite often $\dim V_\eta^{K_{c(\pi)}} = 1$, but this fails in general (see Section 4). In these exceptional cases the dimension of the space of newforms is two.

(v) We prove appropriate Whittaker functionals are nonzero on spaces of newforms. This is of importance in global issues related to newforms. In the proofs, we often need explicit formulae for newforms in various models for the representations. These formulae are interesting for their own sake. For example, if ψ is a character of F of conductor $\mathbb{O}_F$ and (π, V) is a ψ-generic supercuspidal representation of G, then the newform can be taken as the characteristic function of $(\mathbb{O}_F^\times)^2$ where V is regarded as a subspace of the Kirillov model of a canonical minimal representation of $\mathrm{GL}_2(F)$ which determines the L-packet containing π; see [Shimizu 1977].

The paper is structured as follows. We briefly summarize preliminaries on representations and L-packets of $\mathrm{GL}_2(F)$ and $\mathrm{SL}_2(F)$ in Section 2. The main results of the paper are given in Section 3. In this section, after stating the definitions, we take up principal series constituents and supercuspidal representations in separate subsections. We also state results comparing the conductor with other invariants of representations. In Section 4 we discuss the multiplicity one issue for newforms for $\mathrm{SL}_2(F)$.

We now comment briefly on the proofs. A useful preliminary lemma (Lemma 3.1.3) is proved using a variant of an argument of Deligne [1973] based on Kirillov theory for representations of $\mathrm{GL}_2(F)$. This lemma bounds the growth of fixed vectors in representations of $\mathrm{SL}_2(F)$.

For subquotients of principal series representations and their L-packets, most of the proofs use Mackey theory, convenient double coset decompositions, and details regarding restriction of representations from GL_2 to SL_2. There are a few surprisingly difficult exceptions. In particular, for the L-packet corresponding to a quadratic unramified character we use three different realizations of principal series representations for which a general reference is [Gel'fand et al. 1969].

For supercuspidal representations and their L-packets, we make extensive use of Kutzko's construction [1978a; 1978b] of supercuspidal representations of $\mathrm{GL}_2(F)$, as well as the analysis of their restrictions to $\mathrm{SL}_2(F)$ due to Kutzko and Sally [1983]. In showing that certain vectors are newforms, as in the above mentioned SL_2 version of Shimizu's result, we use a combination of arguments involving Kutzko's constructions, the formal Mellin transforms as in Jacquet–Langlands, and the local Langlands correspondence for GL_2 (see Propositions 3.3.5 and 3.3.9).

We mention some further directions that arise naturally from this work. To begin with, we hope to show that our theory of newforms and conductors bears upon known results about local factors for $\mathrm{SL}_2(F)$. In particular, we believe that our conductors are closely related to the analytic conductors appearing in certain epsilon factors. We also believe that an appropriate zeta-integral corresponding to a newform of a representation is equal to a certain local L-factor for that representation. As a possible global application we would like to prove using our newforms that representations (or possibly L-packets) of SL_2 have a nice rationality field, akin to Waldspurger's [1985] result for GL_2.

A companion to this article [Lansky and Raghuram 2004] deals with newforms for the quasisplit unramified unitary group $U(1,1)$. It would be of interest to generalize these results to other groups, namely, to SL_n for higher n and for unitary groups in three variables (for instance the quasisplit unramified unitary group $U(2,1)$).

2. Preliminaries

2.1. ***Notation.*** In the following, F will be a fixed nonarchimedean local field whose residue characteristic is odd. Let $\mathcal{O}$ denote its ring of integers and let $\mathcal{P}$ be the maximal ideal of $\mathcal{O}$. Let ϖ be a uniformizer for F, i.e., $\mathcal{P} = \varpi\mathcal{O}$. Let $k = \mathcal{O}/\mathcal{P}$ be the residue field of F. Let p be the characteristic of k and denote by q the cardinality of k. Let ϵ be an element of $\mathcal{O}^*\backslash\mathcal{O}^{*2}$. We will denote by E a quadratic extension of F and by $\omega_{E/F}$ the quadratic character of F^* associated to E/F by local class field theory. Recall that the kernel of $\omega_{E/F}$ is $N_{E/F}(E^*)$, the norms from E^*.

If n is a positive integer, let U^n denote the nth filtration subgroup $1 + \mathcal{P}^n$ of $\mathcal{O}^\times$, and define $U^0 = \mathcal{O}^\times$. Let $\mathfrak{v}$ denote the additive valuation on F^* which takes the value 1 on ϖ. We let $|\cdot|$ denote the normalized multiplicative valuation given by $|x| = q^{-\mathfrak{v}(x)}$. If χ is a character of F^* we define the conductor $c(\chi)$ to be the smallest nonnegative integer n such that χ is trivial on U^n. Let ψ be a nontrivial additive character of F which is assumed to be trivial on $\mathcal{O}$ and nontrivial on $\mathcal{P}^{-1}$. For any $a \in F$ the character $x \mapsto \psi(ax)$ will be denoted as $\psi_{F,a}$ or simply by ψ_a.

Let $\widetilde{G}$ denote the group $\mathrm{GL}_2(F)$. Let $\widetilde{B} = \widetilde{T}N$ be the standard Borel subgroup of upper triangular matrices in $\widetilde{G}$ with Levi subgroup $\widetilde{T}$ and unipotent radical N. Let $\widetilde{Z}$ be the center of $\widetilde{G}$. Let $G = \mathrm{SL}_2(F)$. Let $B = TN$ be the standard Borel subgroup of upper triangular matrices in G with Levi subgroup T and unipotent radical N. Let w be a representative in the normalizer of T for the nontrivial element of the Weyl group of T. Set $K = \mathrm{SL}_2(\mathcal{O})$ and $\widetilde{K} = \mathrm{GL}_2(\mathcal{O})$. Let I and $\widetilde{I}$ respectively be the standard Iwahori subgroups of G and $\widetilde{G}$.

The following filtrations of maximal compact subgroups of G will be important in our study of newforms. Let $K_{-1} = G$ and $K_0 = K$. Let $K' = K_0' = \alpha^{-1}K_0\alpha$ where $\alpha = \left(\begin{smallmatrix} \varpi & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then K_0 and K_0' are, up to conjugacy, the two maximal compact subgroups of G. For m an integer ≥ 1, recall that K_m and K_m' stand for certain congruence subgroups as defined in the introduction.

In addition to α, we will also make frequent use of the matrices $\beta := \left(\begin{smallmatrix} 1 & 0 \\ 0 & \varpi \end{smallmatrix}\right)$ and $\gamma := \left(\begin{smallmatrix} \epsilon & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

For any subsets $A, B, C, D \subset F$ we let

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a \in A, b \in B, c \in C, d \in D \right\}.$$

We denote $\left[\begin{smallmatrix} 1 & \mathcal{P}^j \\ 0 & 1 \end{smallmatrix}\right]$ by $N(\mathcal{P}^j)$ or simply by $N(j)$. We let $\bar{N}$ denote the lower triangular unipotent subgroup of G and a similar meaning is given to $\bar{N}(\mathcal{P}^j)$ and $\bar{N}(j)$.

If $\mathcal{H}$ is a closed subgroup of a locally compact unimodular group $\mathcal{G}$ and if (σ, W) is a smooth representation of $\mathcal{H}$, then we let $\mathrm{Ind}_{\mathcal{H}}^{\mathcal{G}}(\sigma)$ denote the representation of

$\mathcal{G}$ induced from σ, i.e., the space of locally constant functions $f : \mathcal{G} \to W$ such that for all $h \in \mathcal{H}$ and $g \in \mathcal{G}$ we have

$$f(hg) = \Delta_{\mathcal{H}}^{-1/2}(h)\sigma(h) f(g),$$

where $\Delta_{\mathcal{H}}$ is the modulus character of $\mathcal{H}$. The group $\mathcal{G}$ acts on this space of functions via right translation. We let $\mathrm{ind}_{\mathcal{H}}^{\mathcal{G}}(\sigma)$ denote the subrepresentation of $\mathrm{Ind}_{\mathcal{H}}^{\mathcal{G}}(\sigma)$ consisting of those functions in $\mathrm{Ind}_{\mathcal{H}}^{\mathcal{G}}(\sigma)$ whose supports are compact modulo $\mathcal{H}$.

If π is any irreducible representation of $\mathcal{G}$ on which the center of $\mathcal{G}$ acts by a character, we will denote this character by ω_π. The symbol $\mathbb{1}$ will denote the trivial representation of the group in context.

For real ζ, let $\lceil \zeta \rceil$ denote the least integer greater than or equal to ζ and $\lfloor \zeta \rfloor = -\lceil -\zeta \rceil$.

2.2. *Some results on* $\mathbf{GL_2}(F)$. We briefly recall Kirillov theory for representations of $\mathrm{GL}_2(F)$. For details see [Casselman 1973; Jacquet and Langlands 1970; Prasad and Raghuram 2000]. Let (π, V) be an irreducible admissible infinite-dimensional representation of $\widetilde{G} = \mathrm{GL}_2(F)$. The representation space V may be uniquely realized as a certain space of functions $K(\pi)$, where $C_c^\infty(F^*) \subset K(\pi) \subset C^\infty(F^*)$. Moreover, the space $K(\pi)$ consists of locally constant functions on F^* which vanish outside compact subsets of F and the action of $\widetilde{B}$ on $K(\pi)$ is given by the formula

$$\left(\pi\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} f\right)(x) = \omega_\pi(d)\psi(d^{-1}bx) f(d^{-1}ax) \tag{2.2.1}$$

for all $a, d, x \in F^*$, for all $b \in F$, and for all $f \in K(\pi)$ (see [Casselman 1973]). This Kirillov model $K(\pi)$ has many nice properties, namely:

(i) For all $n \in N$ and for all $v \in V$, $\pi(n)\phi_v - \phi_v$ has compact support in F^* where ϕ_v is the function in $K(\pi)$ associated to the vector $v \in V$.

(ii) The space $K(\pi)$ contains $C_c^\infty(F^*)$ as a subspace of codimension at most two.

(iii) The $\mathbb{C}$-span of functions in (i) is $C_c^\infty(F^*)$. Or in other words the Jacquet module of π, denoted π_N, may be identified as a $\widetilde{T}$ module with $K(\pi)/C_c^\infty(F^*)$.

(iv) The representation π is supercuspidal if and only if $C_c^\infty(F^*) = K(\pi)$.

It is often of interest to know what the newform looks like in the Kirillov model. If π is a supercuspidal representation of $\widetilde{G}$ then the characteristic function of $\mathbb{O}^\times$ is a newform for π. (This was observed in [Shimizu 1977].) A similar result is known for supercuspidal representations of $\mathrm{GL}_2(D)$ [Prasad and Raghuram 2000, Proposition 5.5]. We prove analogous results for $\mathrm{SL}_2(F)$ in this paper.

2.3. ***L-packets for*** $\mathbf{SL_2}(\boldsymbol{F})$**.** In this section we collect statements about the structure of L-packets for $G = \mathrm{SL}_2(F)$. See [Labesse and Langlands 1979; Gelbart and Knapp 1982; Kutzko and Sally 1983; Shelstad 1979].

If $\widetilde{\pi}$ is an irreducible admissible representation of $\widetilde{G}$, its restriction $\mathrm{Res}_{\mathrm{SL}_2(F)}\widetilde{\pi}$ to G is a multiplicity-free finite direct sum of irreducible admissible representations $\pi_1 \oplus \cdots \oplus \pi_r$. On the other hand, if π is any irreducible admissible representation of G, then there exists an irreducible admissible representation $\widetilde{\pi}$ of $\widetilde{G}$ whose restriction to G contains π. The set $\{\pi_1, \ldots, \pi_r\}$ is an *L-packet* of G and $\widetilde{G}$ acts transitively on this set. It is known that the cardinality of an L-packet is $1, 2$ or 4 [Shelstad 1979].

Given such a pair $\widetilde{\pi}$ and π, let $X(\widetilde{\pi}) = \{\chi \in \widehat{F^*} : \widetilde{\pi} \otimes \chi \simeq \widetilde{\pi}\}$, where we identify a character χ of F^* with the character $\chi \circ \det$ of $\widetilde{G}$. Let $\widetilde{G}(\pi) = \{g \in \widetilde{G} : {}^g\pi \simeq \pi\}$. The representation ${}^g\pi$ is defined as ${}^g\pi(x) = \pi(gxg^{-1})$ for all $x \in G$. Clearly, if $\chi \in X(\widetilde{\pi})$ then χ, as a character of $\widetilde{G}$, is trivial on F^*G. (The center of $\widetilde{G}$ is identified with F^*.) Also F^*G is contained in $\widetilde{G}(\pi)$. In fact, given a character χ of F^* we have $\chi \in X(\widetilde{\pi})$ if and only if χ is trivial on $\widetilde{G}(\pi)$ [Labesse and Langlands 1979, Lemma 2.8].

3. Newforms for $\mathbf{SL_2}$

3.1. ***Definitions and the growth lemma.*** We now give our definition of the conductor of a representation of G. Let (π, V) be an admissible representation of G admitting a central character which we denote by ω_π.

We let η be any character of $\mathbb{O}^\times$ such that $\eta(-1) = \omega_\pi(-1)$. Let $c(\eta)$ denote the conductor of η. For any $m \geq c(\eta)$, η gives a character of K_m and K'_m defined by $\eta\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right) = \eta(d)$.

For any nonnegative integer m, we define

$$\pi_\eta^{K_m} := \left\{ v \in V : \pi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) v = \eta(d) v, \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K_m \right\}.$$

We note that $\pi_\eta^{K_m} = (0)$ if $m < c(\eta)$. The spaces $\pi_\eta^{K'_m}$ are defined analogously.

We define the *η-conductor* $c_\eta(\pi)$ of π as

$$c_\eta(\pi) := \min\{m \geq 0 : \pi_\eta^{K_m} \neq (0) \text{ or } \pi_\eta^{K'_m} \neq (0)\}. \tag{3.1.1}$$

We define the *conductor* $c(\pi)$ of π by

$$c(\pi) := \min\{c_\eta(\pi) : \eta\} \tag{3.1.2}$$

where η runs over characters of $\mathbb{O}^\times$ such that $\eta(-1) = \omega_\pi(-1)$. Suppose η satisfies $c_\eta(\pi) = c(\pi)$. If

$$\pi_\eta^{K_{c(\pi)}} \neq (0) \quad \text{or} \quad \pi_\eta^{K'_{c(\pi)}} \neq (0),$$

we call a nonzero element of these spaces a *newform* of π, and $\pi_\eta^{K_{c(\pi)}}$ or $\pi_\eta^{K'_{c(\pi)}}$ itself is then called a *space of newforms* of π.

The following *growth lemma* bounds the growth of the dimension of π^{K_m} for any irreducible representation π of G. It uses Kirillov theory for $\mathrm{GL}_2(F)$. The proof is modeled on Deligne's proof [1973] of a similar GL_2 statement of Casselman [1973]. Similar arguments have also been used in the context of $\mathrm{GL}_2(D)$ in [Prasad and Raghuram 2000].

Lemma 3.1.3. *Let $(\widetilde{\pi}, V)$ be an irreducible admissible representation of $\widetilde{G}$. Let (π, V) be the restriction of $\widetilde{\pi}$ to G. For any character η of $\mathbb{O}^\times$ such that $\eta(-1) = \omega_{\widetilde{\pi}}(-1)$, and for all $m \geq \max\{c(\eta), 1\}$, we have* $\dim(\pi_\eta^{K_m}) - \dim(\pi_\eta^{K_{m-1}}) \leq 2$.

Proof. If V is finite-dimensional, then it is one-dimensional so π is trivial and $\dim(\pi_\eta^{K_m}) = 0, 1$ for all m and we are done. We henceforth assume that V is infinite dimensional. We also assume that $(\widetilde{\pi}, V)$ is realized in its Kirillov model.

Note that $\beta\pi_\eta^{K_{m-1}} \subset \pi_\eta^{K_m}$ since

$$\beta K_{m-1}\beta^{-1} = \begin{bmatrix} \mathbb{O}^\times & \mathcal{P}^{-1} \\ \mathcal{P}^m & \mathbb{O}^\times \end{bmatrix} \cap G \supset K_m$$

It suffices to show that $\dim(\pi_\eta^{K_m}/\beta(\pi_\eta^{K_{m-1}})) \leq 2$ for $m \geq \max\{c(\eta), 1\}$.

Let $f \in \pi_\eta^{K_m}$. Since f is fixed by $N(\mathbb{O}) \subset K_m$, we get that $\mathrm{supp}(f) \subset \mathbb{O}$. (Recall that V is in Kirillov model for $\widetilde{\pi}$.) Indeed, if for $x \in F^*$ we have $f(x) \neq 0$ then

$$f(x) = \left(\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} f\right)(x) = \psi(ax) f(x)$$

for all $a \in \mathbb{O}$ which implies that $x \in \mathbb{O}$. (Recall that ψ is normalized to be trivial on $\mathbb{O}$ and nontrivial on $\mathcal{P}^{-1}$.) Since $T(\mathbb{O}) \subset K_m$ acts via η on f, we get for all $y \in F^*$ and all $u \in \mathbb{O}^\times$

$$\eta(u^{-1}) f(y) = \left(\begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix} f\right)(y) = \omega_{\widetilde{\pi}}(u^{-1}) f(u^2 y),$$

which gives $f(u^2 y) = (\omega_{\widetilde{\pi}}\eta^{-1})(u) f(y)$. This implies that on $\mathbb{O}^\times$, f is completely determined by its values on 1 and ϵ.

Now suppose $f \in \pi_\eta^{K_m}$ is such that $\mathrm{supp}(f) \subset \mathcal{P}$. Then we claim that $f \in \beta(\pi_\eta^{K_{m-1}})$. For this, we show that $\beta^{-1} f \in \pi_\eta^{K_{m-1}}$. Note that $\beta^{-1} f$ is η-fixed by

$$\beta^{-1} K_m \beta \supset \begin{bmatrix} \mathbb{O}^\times & \mathcal{P} \\ \mathcal{P}^{m-1} & \mathbb{O}^\times \end{bmatrix} \cap G$$

and so it suffices to show that $\beta^{-1}f$ is also fixed by $N(\mathbb{O})$. Thus we need to show that for all $y \in F^*$ and all $a \in \mathbb{O}$

$$\left(\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \beta^{-1} f\right)(y) = \left(\beta^{-1} f\right)(y).$$

This reduces to $\psi(ay) f(\varpi y) = f(\varpi y)$, which is true from the assumptions on the support of f and the normalization on ψ.

Suppose now that $f_1, f_2, f_3 \in \pi_\eta^{K_m}$. Then there exist constants a_1, a_2, a_3 such that $\mathrm{supp}(a_1 f_1 + a_2 f_2 + a_3 f_3) \subset \mathcal{P}$. By the arguments above, we obtain $a_1 f_1 + a_2 f_2 + a_3 f_3 \in \beta(\pi_\eta^{K_{m-1}})$. This implies that $\dim(\pi_\eta^{K_m}/\beta(\pi_\eta^{K_{m-1}})) < 3$. $\square$

3.2. ***Principal series representations.*** Let χ be a character of F^*. Then χ inflates to a character of B. Let $\pi(\chi)$ stand for the (unitarily) induced representation $\mathrm{Ind}_B^G(\chi)$. The representation space of $\pi(\chi)$ consists of locally constant complex valued functions f on G such that for all $a \in F^*$, $b \in F$ and $g \in G$, we have

$$f\left(\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} g\right) = |a|\chi(a) f(g).$$

The action of G on such functions is by right translation. It is well known that $\pi(\chi)$ is reducible if and only if χ is either $|\cdot|^{\pm}$ or a nontrivial quadratic character.

There is an essential difference between the two kinds of reducibilities. In the case $\chi = |\cdot|^{\pm}$, $\pi(\chi)$ is the restriction to G of a reducible principal series representation of $\widetilde{G}$. Hence $\pi(\chi)$ will have two representations in its Jordan–Hölder series, namely the trivial representation and the Steinberg representation which we will denote by St_G.

If χ is a nontrivial quadratic character, then $\pi(\chi)$ is the restriction to G of an irreducible principal series representation of $\widetilde{G}$ and breaks up as a direct sum of two irreducible representations, which constitute an L-packet of G. If $\chi = \omega_{E/F}$ we denote $\pi(\chi)$ by π_E and let $\pi_E \simeq \pi_E^1 \oplus \pi_E^2$. We denote the L-packet by $\xi_E = \{\pi_E^1, \pi_E^2\}$.

To begin, we need a lemma on double coset decompositions to analyze the space of fixed vectors in principal series representations.

Lemma 3.2.1. *Let $m \geq 1$ and set $x_i = \left(\begin{smallmatrix} 1 & 0 \\ \varpi^i & 1 \end{smallmatrix}\right)$ and $y_i = \left(\begin{smallmatrix} 1 & 0 \\ \epsilon\varpi^i & 1 \end{smallmatrix}\right)$ for $1 \leq i \leq m-1$. A complete set of representatives for the double coset space $K_m \backslash K / B(\mathbb{O})$ is given by $\{1, w, x_i, y_i\}_{1 \leq i \leq m-1}$.*

Proof. Let $k = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in K$. The coset representatives are determined by considering the cases where c is in $\mathcal{P}^m$, where c is in $\mathbb{O}^\times$, or for $1 \leq i \leq m$, where $\mathfrak{v}(c) = i$ and $\varpi^i c^{-1} d$ is or is not a square. We leave the routine details to the reader. $\square$

Let χ be a character of F^*. Let η be a character of $\mathbb{O}^\times$ such that $\eta(-1) = \chi(-1)$. Let $m \geq c(\eta)$. We note that the space $\pi(\chi)_\eta^{K_m}$ is isomorphic to $\mathrm{Hom}_{K_m}(\eta, \pi(\chi))$.

In light of the Lemma 3.2.1, standard Mackey theory yields for $m \geq 2$

$$\pi(\chi)_\eta^{K_m} = \mathrm{Hom}_{B(\mathbb{O})}(\eta, \chi) \oplus \mathrm{Hom}_{w^{-1}K_m w \cap B(\mathbb{O})}({}^w\eta, \chi) \oplus \bigoplus_{i=1}^{m-1} \mathrm{Hom}_{x_i^{-1}K_m x_i \cap B(\mathbb{O})}({}^{x_i}\eta, \chi) \oplus \bigoplus_{i=1}^{m-1} \mathrm{Hom}_{y_i^{-1}K_m y_i \cap B(\mathbb{O})}({}^{y_i}\eta, \chi). \tag{3.2.2}$$

If $m = 1$, only the first two terms appear, while if $m = 0$, only the first appears. We use this result extensively in the computations of this section.

As mentioned in the introduction, one of the applications of newforms we have in mind is that they are test vectors for Whittaker functionals. For principal series representations and in fact all their subquotients we consider the following ψ-Whittaker functional [Schmidt 2002]. For any function f in a principal series representation $\pi(\chi)$ we define

$$\Lambda_\psi f := \lim_{r\to\infty} \int_{\mathcal{P}^{-r}} f\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\right)\overline{\psi}(x)\,dx, \tag{3.2.3}$$

where the Haar measure dx is normalized such that $\mathrm{vol}(\mathbb{O}) = 1$.

Proposition 3.2.4 (Unramified principal series representations). *Let χ be an unramified character of F^* and let $\pi = \pi(\chi)$ be the corresponding principal series representation of G. Then $c(\pi) = 0$, and moreover, $c_\eta(\pi) = c(\pi)$ only when η is trivial. The dimension of the space of fixed vectors under K_m is given by*

$$\dim \pi(\chi)^{K_m} = \begin{cases} 1 & \text{if } m = 0, \\ 2m & \text{if } m \geq 1. \end{cases}$$

Proof. Note that Lemma 3.2.1 and the fact that $G = BK$ implies that

$$|K_m\backslash G/B| = |K_m\backslash K/B(\mathbb{O})| = \begin{cases} 1 & \text{if } m = 0, \\ 2m & \text{if } m \geq 1. \end{cases}$$

This, together with (3.2.2) proves the proposition. □

Corollary 3.2.5 (Test vectors for unramified principal series representations). *For an unramified character χ of F^* such that $\chi \neq |\cdot|^{-1}$, let f_{new} be any nonzero K-fixed vector of the representation $\pi(\chi)$. Then we have $\Lambda_\psi f_{\mathrm{new}} = L(1, \chi)^{-1} \neq 0$, where $L(s, \chi)$ is the standard local abelian L-factor associated to χ.*

Proof. This is a standard computation in the theory of spherical representations. We merely give a sketch of the details. We can take the newform $f = f_{\mathrm{new}}$ to be given by

$$f(g) = \chi(a)|a|\ , \quad g = \begin{pmatrix} a & * \\ 0 & a^{-1} \end{pmatrix} k \in BK = G.$$

We have

$$\begin{aligned}\Lambda_\psi f_{\text{new}} &= 1 + \lim_{r\to\infty}\int_{\mathcal{P}^{-r}\setminus\mathcal{O}} f\left(\begin{pmatrix} x^{-1} & -1 \\ 0 & x\end{pmatrix}\begin{pmatrix} 1 & 0 \\ x^{-1} & 1\end{pmatrix}\right)\overline{\psi}(x)\,dx \\ &= 1 + \sum_{m=1}^{\infty}\int_{\varpi^{-m}\mathcal{O}^\times} \chi(x^{-1})|x^{-1}|\overline{\psi}(x)\,dx \\ &= (1-\chi(\varpi)q^{-1}) = L(1,\chi)^{-1} \neq 0. \qquad \square\end{aligned}$$

Proposition 3.2.6 (Steinberg representation). *Let* St_G *be the Steinberg representation of* G. *Then* $c(\mathrm{St}_G) = 1$, *and moreover*, $c_\eta(\mathrm{St}_G) = c(\mathrm{St}_G)$ *only when* η *is trivial. The dimension of the space of fixed vectors under* K_m *is given by*

$$\dim(\mathrm{St}_G^{K_m}) = \begin{cases} 0 & \text{if } m = 0, \\ 2m-1 & \text{if } m \geq 1.\end{cases}$$

Proof. The result follows from Proposition 3.2.4 and (3.2.2). □

Corollary 3.2.7 (Test vectors for the Steinberg representation). *Let the Steinberg representation* St_G *be realized as the unique irreducible subrepresentation of* $\pi(|\cdot|)$. *Then the* ψ*-Whittaker functional* Λ_ψ *is nonzero on the space of newforms* $(\mathrm{St}_G)_{\text{new}} = \mathrm{St}_G^{K_1}$.

Proof. We consider the standard intertwining operator $M : \pi(|\cdot|) \to \pi(|\cdot|^{-1})$ given by

$$(Mf)(g) = \int_F f\left(\begin{pmatrix} 0 & -1 \\ 1 & 0\end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix} g\right) dx$$

for all $f \in \pi(|\cdot|)$ and for all $g \in G$. The representation space of the Steinberg representation is simply the kernel V_M of M [Bump 1997, §4.5].

Note that a function $f \in \pi(|\cdot|)^{K_1}$ is determined by its values on the elements $1, w \in G$. Let f_{new} be an element of $\pi(|\cdot|)^{K_1}$ determined by $f_{\text{new}}(1) = q$ and $f_{\text{new}}(w) = -1$. It is easy to see that $Mf_{\text{new}} = 0$ and so f_{new} is indeed a newform for the Steinberg representation. A computation very much like that in the proof of Corollary 3.2.5 shows that $\Lambda_\psi f_{\text{new}} = -(1+q^{-1}) \neq 0$. We leave the details to the reader. See also [Schmidt 2002]. □

Proposition 3.2.8 (Ramified principal series representations). *Let* χ *be a ramified character of* F^*. *Let* $\pi = \pi(\chi)$ *be the corresponding principal series representation of* G. *Let* $c(\chi)$ *denote the conductor of* χ.

(i) *We have* $c(\pi) = c(\chi)$ *and further* $c_\eta(\pi) = c(\pi)$ *only for those characters* η *such that* $\eta = \chi^{\pm}$ *on the group of units* $\mathcal{O}^\times$.

(ii) *If* $\chi^2|_{(\mathbb{O}^\times)^2} \neq \mathbb{1}$ *and* $\eta = \chi^{\pm}|_{\mathbb{O}^\times}$ *then*

$$\dim \pi(\chi)_\eta^{K_m} = \begin{cases} 0 & \text{if } m < c(\chi), \\ 1 & \text{if } m = c(\chi), \\ 2(m - c(\chi)) + 1 & \text{if } m > c(\chi). \end{cases}$$

(iii) *If* $\chi^2|_{(\mathbb{O}^\times)^2} = \mathbb{1}$ *and* $\eta = \chi^{\pm}|_{\mathbb{O}^\times}$ *then*

$$\dim \pi(\chi)_\eta^{K_m} = \begin{cases} 0 & \text{if } m = 0, \\ 2m & \text{if } m \geq 1 = c(\chi). \end{cases}$$

Proof. Let η be a character of $\mathbb{O}^\times$ such that $\eta(-1) = \chi(-1)$. Let $m \geq c(\eta)$. We must determine the dimensions of the Hom-spaces in (3.2.2). The space $\mathrm{Hom}_{B(\mathbb{O})}(\eta, \chi)$ is nonzero if and only if $\eta = \chi^{-1}$ as characters of $\mathbb{O}^\times$. The space $\mathrm{Hom}_{w^{-1}K_m w \cap B(\mathbb{O})}({}^w\eta, \chi)$ is nonzero if and only if $\eta = \chi$ as characters of $\mathbb{O}^\times$. For the summands corresponding to x_i we observe that $\mathrm{Hom}_{x_i^{-1}K_m x_i \cap B(\mathbb{O})}({}^{x_i}\eta, \chi) \neq (0)$ if and only if $\eta = \chi$ on $1 + \mathcal{P}^{\min\{i, m-i\}}$ and $m - i \geq c(\eta)$. An identical statement holds for the summands corresponding to y_i. All the assertions in the proposition follow easily from these observations. We leave the details to the reader. □

Corollary 3.2.9 (Test vectors for ramified principal series representations). *Let χ be a ramified character of F^*. Let $\pi = \pi(\chi)$ be the corresponding principal series representation of G. Assume that π is irreducible. Let $m = c(\chi) \geq 1$ denote the conductor of χ. The space of newforms $\pi(\chi)_{\mathrm{new}} = \pi(\chi)_\eta^{K_{c(\chi)}}$ is one-dimensional where η is χ restricted to $\mathbb{O}^\times$, and the Whittaker functional Λ_ψ is nonzero on this space of newforms.*

Proof. From the proof of Proposition 3.2.8 it follows that a newform may be taken as the function f_{new} that is supported on the double coset BwK_m and on this coset it is given by

$$f\left(\begin{pmatrix} t & * \\ 0 & t^{-1} \end{pmatrix} w \begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \chi(t)|t|\chi(d).$$

As in the proof of Corollary 3.2.5 it can be shown that $\Lambda_\psi(f_{\mathrm{new}}) = \chi(-1) \neq 0$. □

We now consider the L-packets $\xi_E = \{\pi_E^1, \pi_E^2\}$ where E/F is a quadratic extension. To begin, we take up the case where E/F is ramified.

Proposition 3.2.10 (Ramified principal series L-packets). *Let $\xi_E = \{\pi_E^1, \pi_E^2\}$ with E/F ramified. We have $c(\pi_E^1) = c(\pi_E^2) = 1$. Further, we have for $\eta = \omega_{E/F}|_{\mathbb{O}^\times}$*

$$\dim(\pi_E^1)_\eta^{K_m} = \dim(\pi_E^2)_\eta^{K_m} = \begin{cases} 0 & \text{if } m = 0, \\ m & \text{if } m \geq 1. \end{cases}$$

Proof. Let $E = F(\sqrt{-\lambda})$ where λ is either ϖ or $\epsilon\varpi$. Note that $\omega_{E/F}$ is trivial on λ. Let $\widetilde{\pi}_E$ be the principal series representation of $\widetilde{G}$ unitarily induced from the character $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \mapsto \omega_{E/F}(a)$. Then it is easily seen that $\widetilde{\pi}_E \otimes \omega_{E/F} = \widetilde{\pi}_E$. (Note that $\widetilde{\pi}_E$ restricts to the representation $\pi(\omega_{E/F}) = \pi_E = \pi_E^1 \oplus \pi_E^2$ of G.) From Section 2.3 we get that $\omega_{E/F}$ is trivial on $G(\pi_E^1)$. Hence $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & 1 \end{smallmatrix}\right) \in G(\pi_E^1)$, which implies that $\gamma \notin G(\pi_E^1)$ or in other words γ conjugates π_E^1 into π_E^2. This also gives that γ conjugates $(\pi_E^1)_\eta^{K_m}$ into $(\pi_E^2)_\eta^{K_m}$ for all m and for all permissible η. Since $(\pi_E)_\eta^{K_m} = (\pi_E^1)_\eta^{K_m} \oplus (\pi_E^2)_\eta^{K_m}$ the proof follows from (i) and (iii) of Proposition 3.2.8 for $\eta = \omega_{E/F}$ on the units. □

Corollary 3.2.11 (Test vectors for ramified principal series L-packets). *Let $\xi_E = \{\pi_E^1, \pi_E^2\}$ with E/F ramified. Then one and only one of the two representations in the packet is ψ-generic, say π_E^1 (so π_E^2 is ψ_ϵ-generic). The Whittaker functional Λ_ψ is nonzero on the one-dimensional space of newforms $(\pi_E^1)_{\text{new}} = (\pi_E^1)_{\omega_{E/F}}^{K_1}$. Any ψ_ϵ-Whittaker functional is nonzero on the one-dimensional space of newforms for π_E^2.*

Proof. The assertions for π_E^1 follow exactly as in the proof of Corollary 3.2.9. Conjugating by γ proves the assertions for π_E^2. □

Proposition 3.2.12 (Unramified principal series L-packet). *Let $\xi_E = \{\pi_E^1, \pi_E^2\}$ with E/F unramified. Then $c(\pi_E^1) = c(\pi_E^2) = 0$ and η is trivial as in Proposition 3.2.4. One and only one of the two representations, say π_E^1, has a nonzero vector fixed by K_0. The dimensions of the space of fixed vectors under K_m and K'_m for the two representations are as follows*:

(i) *For* $r \geq 0$, $\dim((\pi_E^1)^{K_r}) = 2\left\lfloor \frac{r}{2} \right\rfloor + 1 = \dim((\pi_E^2)^{K'_r})$.

(ii) *For* $r \geq 0$, $\dim((\pi_E^1)^{K'_r}) = \max\left\{2\left\lfloor \frac{r-1}{2} \right\rfloor + 1, 0\right\} = \dim((\pi_E^2)^{K_r})$.

Proof. Recall our notation that $\pi(\omega_{E/F}) = \pi_E = \pi_E^1 \oplus \pi_E^2$. As in the ramified principal series L-packets case, we see that α conjugates π_E^1 to π_E^2. The assumption that it is π_E^1 which has a nonzero vector fixed under K_0 gives all the dimensions when $r = 0$. Since K'_1 can be conjugated by an element of G inside K_0, and K_1 inside K'_0, it follows from Proposition 3.2.4 that all the spaces $(\pi_E^1)^{K_1}$, $(\pi_E^1)^{K'_1}$, $(\pi_E^2)^{K_1}$ and $(\pi_E^2)^{K'_1}$ are one-dimensional.

Assume now that π_E is realized in the space $\overline{V}$ of locally constant functions $f \in L^2(F)$ such that $\omega_{E/F}(x)|x|f(x)$ is constant for sufficiently large $x \in F$. For this realization [Gel'fand et al. 1969, Chapter 2, §3.1], the action of $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ on an $f \in L^2(F)$ as above is given by

$$\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} f\right)(x) = \omega_{E/F}(bx+d)|bx+d|^{-1} f((ax+c)/(bx+d)).$$

In this realization, the spherical vector $f_0 \in (\pi_E^1)^{K_0}$ is given by

$$f_0(x) = \begin{cases} 1 & \text{if } x \in \mathbb{O}, \\ \omega_{E/F}(x)|x|^{-1} & \text{if } x \notin \mathbb{O}. \end{cases}$$

We define two elements f_1 and f_2 by

$$f_1(x) = (\beta f_0)(x) = \begin{cases} -q & \text{if } x \in \mathcal{P}, \\ \omega_{E/F}(x)|x|^{-1} & \text{if } x \notin \mathcal{P} \end{cases}$$

and

$$f_2(x) = \left(\begin{pmatrix} \varpi^{-1} & 0 \\ 0 & \varpi \end{pmatrix} f_0 \right)(x) = \begin{cases} -q & \text{if } x \in \mathcal{P}^2, \\ -q^{-1}\omega_{E/F}(x)|x|^{-1} & \text{if } x \notin \mathcal{P}^2. \end{cases}$$

Note that $f_1 \in (\pi_E^2)^{K_1}$ and $f_2 \in (\pi_E^1)^{K_2}$. Analogous to f_1 we define for every $c \in \mathbb{C}$ an element f_1^c in π_E given by

$$f_1^c(x) = \begin{cases} c & \text{if } x \in \mathcal{P}, \\ \omega_{E/F}(x)|x|^{-1} & \text{if } x \notin \mathcal{P}. \end{cases}$$

We claim that for every c the element f_1^c is fixed by K_2. This can be seen using the Iwahori factorization: $K_2 = \overline{N}(\mathcal{P}^2)T(\mathbb{O})N(\mathbb{O})$. Clearly, both $T(\mathbb{O})$ and $\overline{N}(\mathcal{P}^2)$ fix f_1^c. Moreover, $N(\mathbb{O})$ fixes f_1^c if and only if wf_1^c is fixed by $\overline{N}(\mathbb{O})$. The observation that

$$(wf_1^c)(x) = \begin{cases} 1 & \text{if } x \in \mathbb{O}, \\ c\omega_{E/F}(x)|x|^{-1} & \text{if } x \notin \mathbb{O} \end{cases}$$

shows that wf_1^c is indeed fixed by $\overline{N}(\mathbb{O})$. Hence, for every c, f_1^c is fixed by K_2. Note that for any fixed c the elements f_0, f_2, f_1^c are all linearly independent. Further, as c varies the elements f_1^c span a two-dimensional space. Therefore, we have shown that $\dim \pi_E^{K_2}$ is at least 4, and, applying Lemma 3.1.3, we see that the dimension is at most four, hence equals 4. We now need to determine the dimensions of $(\pi_E^1)^{K_2}$ and $(\pi_E^2)^{K_2}$ given that their sum is 4. For this, we need yet another realization of principal series representations for G. We will show that there is some c such that f_1^c is in π_E^1 which will then force $\dim((\pi_E^1)^{K_2}) = 3$ and $\dim((\pi_E^2)^{K_2}) = 1$.

We refer the reader to [Gel'fand et al. 1969, Chapter 2, §3.2] for this third realization, which is obtained by taking Fourier transforms $\widehat{f}$ of functions $f \in \overline{V}$ with respect to $\psi_{\varpi^{-1}}$. We will let $\widehat{V}$ denote the space of all $\widehat{f}$ as f varies over $\overline{V}$.

The representation space of π_E^1 can be recognized in $\widehat{V}$ as the space of functions supported only on the norms $N_{E/F}(E)$, and π_E^2 as that of functions supported only on nonnorms (see [Gel'fand et al. 1969, Chapter 2, §3.5]). Since E/F is unramified, $N_{E/F}(E)$ is the set of elements of F^* with has even valuation together with 0, and the nonnorms are those elements with odd valuation.

Computing the Fourier transform of the spherical vector f_0 we get that π_E^1 consists of functions supported on nonnorms. We will show now that there is a c such that f_1^c has support inside nonnorms. Indeed, the formula for the Fourier transform of f_1^c is

$$\widehat{f_1^c}(\varpi^n) = \begin{cases} 1 + cq^{-1} & \text{if } n \text{ is odd,} \\ (c-1)q^{-1} & \text{if } n \text{ is even.} \end{cases}$$

Hence for $c = 1$ we get that $f_1^c \in \pi_E^1$ and for $c = -q$ we get that $f_1^c \in \pi_E^2$. We have now computed the dimensions of fixed vectors under K_r and K_r' for $r = 0, 1, 2$. From this point onwards an induction argument takes over.

If the dimensions are known for all $r \leq 2m$ then using the fact that K_{r+1}' can be conjugated inside K_r and K_{r+1} inside K_r', we get using Lemma 3.1.3 that all the spaces $(\pi_E^1)^{K_{2m+1}}$, $(\pi_E^1)^{K_{2m+1}'}$, $(\pi_E^2)^{K_{2m+1}}$ and $(\pi_E^2)^{K_{2m+1}'}$ have dimensions equal to $2m+1$. Using the same lemma it suffices now to show that the dimension of $(\pi_E^1)^{K_{2m+2}}$ is at least $2m+3$.

Using $(\pi_E^1)^{K_{2m+1}} + (\pi_E^1)^{K_{2m+1}'} \subset (\pi_E^1)^{K_{2m+2}}$ and that the subgroup of G generated by K_{2m+1} and K_{2m+1}' is K_{2m}' (which can be seen by Iwahori factorization) we get

$$\begin{aligned} \dim((\pi_E^1)^{K_{2m+2}}) &\geq \dim((\pi_E^1)^{K_{2m+1}}) + \dim((\pi_E^1)^{K_{2m+1}'}) \\ &= (2m+1) + (2m+1) - \dim\big((\pi_E^1)^{K_{2m+1}} \cap (\pi_E^1)^{K_{2m+1}'}\big) \\ &= 4m+2 - \dim((\pi_E^1)^{K_{2m}'}) = 4m+2-(2m-1) = 2m+3. \qquad \square \end{aligned}$$

Corollary 3.2.13 (Test vectors for unramified principal series L-packet). *For E/F unramified let $\xi_E = \{\pi_E^1, \pi_E^2\}$. Then one and only one of the two representations in the packet is ψ-generic, namely π_E^1. Moreover, a ψ-Whittaker functional is nonzero on the K_0-fixed vector in π_E^1. The representation π_E^2 is not ψ'-generic for any ψ' of conductor $\mathbb{O}$. It is ψ_ϖ-generic and any ψ_ϖ-Whittaker functional is nonzero on the K_0'-fixed vectors in π_E^2.*

Proof. The result follows from multiplicity one for Whittaker models, the fact that α conjugates π_E^1 to π_E^2, and Corollary 3.2.5. $\square$

3.3. ***Supercuspidal representations.*** We now consider supercuspidal representations of $G = \mathrm{SL}_2(F)$. For this we need some preliminaries on how they are constructed. A direct approach is found in Manderscheid's papers [1984]. We, however, use Kutzko's construction [1978a; 1978b] of supercuspidal representations for $\widetilde{G}$ and then the results of [Kutzko and Sally 1983] to obtain information on the supercuspidal representations (L-packets) for G.

We begin by briefly recalling Kutzko's construction of supercuspidal representations of $\widetilde{G}$ via compact induction from very cuspidal representations of maximal open compact-mod-center subgroups. For $l \geq 1$, let $\widetilde{K}(l) = 1 + \mathcal{P}^l M_{2\times 2}(\mathbb{O})$ be the principal congruence subgroup of $\widetilde{K}$ of level l. Let $\widetilde{K}(0) = \widetilde{K}$. Let $\widetilde{I}$ be the standard Iwahori subgroup consisting of all elements in $\widetilde{K}$ that are upper triangular modulo $\mathcal{P}$. For $l \geq 1$ let $\widetilde{I}(l) = \left[\begin{smallmatrix} 1+\mathcal{P}^l & \mathcal{P}^l \\ \mathcal{P}^{l+1} & 1+\mathcal{P}^l \end{smallmatrix}\right]$, and let $\widetilde{I}(0) = \widetilde{I}$. We will let $\widetilde{H}$ denote either $Z\widetilde{K}$ or $N_{\widetilde{G}}\widetilde{I}$, while $\widetilde{J}$ will denote either $\widetilde{K}$ or $\widetilde{I}$ Here $N_{\widetilde{G}}\widetilde{I}$ is the normalizer in $\widetilde{G}$ of $\widetilde{I}$. In either case we let $\widetilde{J}(l)$ denote the corresponding filtration subgroup.

Definition 3.3.1 (Kutzko). An irreducible (and necessarily finite-dimensional) representation $(\widetilde{\sigma}, W)$ of $\widetilde{H}$ is called a very cuspidal representation of level $l \geq 1$ if $\widetilde{\sigma}$ is trivial on $\widetilde{J}(l)$ and $\widetilde{\sigma}$ does not contain the trivial character of $N(\mathcal{P}^{l-1})$.

Remark 3.3.2. One easy consequence of the definition is that, if $\widetilde{\sigma}$ is a very cuspidal representation of $Z\widetilde{K}$ (resp. $N_{\widetilde{G}}\widetilde{I}$) of level l then $\mathrm{Hom}_{N(l-1)}(\mathbb{1}, \widetilde{\sigma}) = \mathrm{Hom}_{\overline{N}(l-1)}(\mathbb{1}, \widetilde{\sigma}) = (0)$ (resp. $\mathrm{Hom}_{N(l-1)}(\mathbb{1}, \widetilde{\sigma}) = \mathrm{Hom}_{\overline{N}(l)}(\mathbb{1}, \widetilde{\sigma}) = (0)$).

We say that an irreducible admissible representation $\widetilde{\pi}$ of $\widetilde{G}$ is minimal if for every character χ of F^* we have $c(\widetilde{\pi}) \leq c(\widetilde{\pi} \otimes \chi)$, for $c(\widetilde{\pi})$ as in Casselman's Theorem (page 127).

Theorem 3.3.3 [Kutzko 1978a; 1978b]. *Let π be a minimal irreducible supercuspidal representation of $\widetilde{G}$. Then π is compactly induced from a very cuspidal representation $\widetilde{\sigma}$ of one of the two maximal open compact-mod-center subgroups $\widetilde{H}$ of $\widetilde{G}$. Moreover, $\widetilde{H}$ and the equivalence class of σ are uniquely determined by π. If the conductor of π is* $2l$ (*resp.* $2l + 1$), *then π comes from a very cuspidal representation of $Z\widetilde{K}$* (*resp. $N_{\widetilde{G}}\widetilde{I}$*) *of level l.*

Following Kutzko we use the terminology that a supercuspidal representation of $\widetilde{G}$ is said to be *unramified* if it comes via compact induction from a representation of $Z\widetilde{K}$ and is said to be *ramified* otherwise. The theorem assures us that the ramified ones come via compact induction from representations of $N_{\widetilde{G}}\widetilde{I}$. We now take up both types of supercuspidal representations and briefly review how they break up on restriction to G (see [Kutzko and Sally 1983]).

We begin with the unramified case. Let $\widetilde{\sigma}$ be an irreducible very cuspidal representation of $Z\widetilde{K}$ of level l (≥ 1). Let $\widetilde{\pi}$ be the corresponding supercuspidal representation of $\widetilde{G}$. Let $\sigma = \mathrm{Res}_K(\widetilde{\sigma})$. Then we have $\mathrm{Res}_G(\widetilde{\pi}) = \mathrm{ind}_K^G(\sigma) \oplus {}^{\alpha}\big(\mathrm{ind}_K^G(\sigma)\big)$ where $\alpha = \left(\begin{smallmatrix} \varpi & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If $l \geq 2$, or if $l = 1$ and σ is irreducible, then $\pi = \pi(\sigma) = \mathrm{ind}_K^G(\sigma)$ is irreducible, hence so is $\pi' = {}^{\alpha}\pi$. We thus have an *unramified supercuspidal L-packet* $\{\pi, \pi'\}$. If $l = 1$ and σ is reducible, which happens when $\widetilde{\sigma}$ comes from the unique (up to twists) cuspidal representation of $\mathrm{GL}_2(\mathbb{F}_q)$ whose restriction to $\mathrm{SL}_2(\mathbb{F}_q)$ is reducible and in this case it breaks up into the direct sum of the two cuspidal representations of $\mathrm{SL}_2(\mathbb{F}_q)$ of dimension $(q-1)/2$.

Correspondingly, we have $\sigma = \sigma_1 \oplus \sigma_2$, and if we let $\pi_i = \mathrm{ind}_K^G(\sigma_i)$ and $\pi_i' = {}^{\alpha}(\pi_i)$, then we obtain the unique supercuspidal L-packet $\{\pi_1, \pi_1', \pi_2, \pi_2'\}$ of G containing 4 elements.

For the ramified case let $\widetilde{\sigma}$ be a very cuspidal representation of $N_{\widetilde{G}}\widetilde{I}$ of level l (≥ 1) and let $\widetilde{\pi}$ be the corresponding supercuspidal representation of $\widetilde{G}$. Let $\sigma = \mathrm{Res}_I(\widetilde{\sigma})$. Then $\sigma = \sigma_1 \oplus \sigma_2$ for two irreducible representations σ_i $(i = 1, 2)$ of I and γ conjugates one to the other, i.e., $\sigma_2 = {}^{\gamma}\sigma_1$. Let $\pi_i = \mathrm{ind}_I^G(\sigma_i)$ and so $\pi_2 = {}^{\gamma}\pi_1$. Then the restriction of $\widetilde{\pi}$ to G breaks up into the direct sum of two irreducible supercuspidal representations as $\mathrm{Res}_G(\widetilde{\pi}) = \pi_1 \oplus \pi_2$. We call $\{\pi_1, \pi_2\}$ a *ramified supercuspidal L-packet* of G.

Proposition 3.3.4 (Unramified supercuspidal L-packets of cardinality two). *Consider an unramified supercuspidal L-packet $\{\pi, \pi'\}$ determined by a very cuspidal representation $\widetilde{\sigma}$ of level l of $Z\widetilde{K}$ as above. Then the conductors $c(\pi), c(\pi')$ are both equal to $2l$. For any character η such that $\eta(-1) = \omega_\pi(-1)$ we have*

(i) $\pi_\eta^{K_{2l-1}} = \pi_\eta^{K'_{2l-1}} = (\pi')_\eta^{K_{2l-1}} = (\pi')_\eta^{K'_{2l-1}} = (0)$.

(ii) *If $c(\eta) \leq l$ and l is odd then for all $m \geq 2l$*
  - (a) $\dim \pi_\eta^{K'_m} = \dim (\pi')_\eta^{K_m} = 2\left\lceil \frac{m-2l+1}{2} \right\rceil$.
  - (b) $\dim \pi_\eta^{K_m} = \dim (\pi')_\eta^{K'_m} = 2\left\lfloor \frac{m-2l+1}{2} \right\rfloor$.

(iii) *If $c(\eta) \leq l$ and l is even then for all $m \geq 2l$*
  - (a) $\dim \pi_\eta^{K_m} = \dim (\pi')_\eta^{K'_m} = 2\left\lceil \frac{m-2l+1}{2} \right\rceil$.
  - (b) $\dim \pi_\eta^{K'_m} = \dim (\pi')_\eta^{K_m} = 2\left\lfloor \frac{m-2l+1}{2} \right\rfloor$.

Proof. The statement about the conductors of π, π' follows immediately from (i), (ii), and (iii). To prove (i) it suffices to prove that $\mathrm{Hom}_{K_{2l-1}}(\eta, \widetilde{\pi}) = (0)$ where $\widetilde{\pi} = \mathrm{ind}_{Z\widetilde{K}}^{\widetilde{G}}(\widetilde{\sigma})$ is the unramified supercuspidal representation of $\widetilde{G}$ given by the very cuspidal representation $\widetilde{\sigma}$ of level m. Using Frobenius reciprocity and Mackey theory [Kutzko 1977] we have

$$\begin{aligned}\mathrm{Hom}_{K_{2l-1}}(\eta, \widetilde{\pi}) &= \mathrm{Hom}_{\widetilde{G}}(\mathrm{ind}_{K_{2l-1}}^{\widetilde{G}}(\eta), \mathrm{Ind}_{Z\widetilde{K}}^{\widetilde{G}}(\widetilde{\sigma})) \\ &= \prod_{x \in K_{2l-1}\backslash \widetilde{G}/Z\widetilde{K}} \mathrm{Hom}_{Z\widetilde{K} \cap x^{-1}K_{2l-1}x}({}^{x}\eta, \widetilde{\sigma}).\end{aligned}$$

For brevity, let $I_x = \mathrm{Hom}_{Z\widetilde{K} \cap x^{-1}K_{2l-1}x}({}^{x}\eta, \widetilde{\sigma})$. For every x we will show that $I_x = (0)$.

To this end we need a set of representatives for the double cosets. We leave it to the reader to check that this is given by

$$\begin{aligned}K_{2l-1}\backslash \widetilde{G}/Z\widetilde{K} = \{&\bar{n}(s)h(u)a_r : s \in \mathcal{P}/\mathcal{P}^{2l-1}, u \in \mathbb{O}^\times, r \geq 0\} \\ &\cup \{\bar{n}(t)wh(u)a_r : t \in \mathbb{O}/\mathcal{P}^{2l-1}, u \in \mathbb{O}^\times, r \geq 0\},\end{aligned}$$

where $\bar{n}(s) = \left(\begin{smallmatrix} 1 & 0 \\ s & 1 \end{smallmatrix}\right)$, $h(u) = \left(\begin{smallmatrix} u & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $a_r = \alpha^r = \left(\begin{smallmatrix} \varpi^r & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

We begin with the case where $x = \bar{n}(s)h(u)a_r$. If $s = 0$, then $x = h(u)a_r$, which implies that $N(\mathbb{O}) \subset Z\widetilde{K} \cap x^{-1}K_{2l-1}x$ and so

$$I_x \subset \mathrm{Hom}_{N(\mathbb{O})}({}^x\eta, \widetilde{\sigma}) = \mathrm{Hom}_{N(\mathbb{O})}(\mathbb{1}, \widetilde{\sigma}) = (0)$$

since $\widetilde{\sigma}$ is very cuspidal. If $s \neq 0$, then $x = h(u)a_r\bar{n}(u\varpi^r s) = h(u)a_r k$. Since $\bar{n}(u\varpi^r s) \in \widetilde{K}$ we get $I_x \simeq I_{h(u)a_r}$, and we are reduced to the case $s = 0$. Thus $I_x = (0)$.

Now let $x = \bar{n}(t)wh(u)a_r$. If $r \geq l$, then $N(\mathcal{P}^{l-1}) \subset Z\widetilde{K} \cap x^{-1}K_{2l-1}x$, and so

$$I_x \subset \mathrm{Hom}_{N(\mathcal{P}^{l-1})}({}^x\eta, \widetilde{\sigma}) = \mathrm{Hom}_{N(\mathcal{P}^{l-1})}(\mathbb{1}, \widetilde{\sigma}) = (0)$$

by Remark 3.3.2. If $r = 0$, then $x \in K$, and we have

$$I_x \simeq \mathrm{Hom}_{K_{2l-1} \cap xZ\widetilde{K}x^{-1}}(\eta, {}^x\widetilde{\sigma}) \subset \mathrm{Hom}_{N(\mathbb{O})}(\mathbb{1}, \widetilde{\sigma}) = (0).$$

If $0 < r < l$ and $t \in \mathcal{P}^r$, then rewrite x as $x = g(u)b_r k$ where $g(u) = wh(u)w^{-1}$ and $b_r = wa_rw^{-1}$ and for some $k \in \widetilde{K}$. Here we finish the argument with

$$I_x \simeq \mathrm{Hom}_{(g(u)b_r)^{-1}K_{2l-1}g(u)b_r \cap Z\widetilde{K}}({}^{g(u)b_r}\eta, \widetilde{\sigma}) \subset \mathrm{Hom}_{N(l-1)}(\mathbb{1}, \widetilde{\sigma}) = (0).$$

We are finally left with the case where $0 < r < l$ and $t \in \mathbb{O} - \mathcal{P}^r$. For this, rewrite x as $x = yw$ with $y = \bar{n}(t)b_r g(u)$. Since $w \in \widetilde{K}$, as before, we have that $I_x = (0)$ if and only if $I_y = (0)$. Let $j = j(r, t) = \max\{0, r - 2\mathfrak{v}(t)\}$. Then we get $y^{-1}N(\mathcal{P}^j)y \subset y^{-1}K_{2l-1}y \cap Z\widetilde{K}$, which gives

$$I_y \subset \mathrm{Hom}_{y^{-1}N(\mathcal{P}^j)y}({}^y\eta, \widetilde{\sigma}) = \mathrm{Hom}_{y^{-1}N(\mathcal{P}^j)y}(\mathbb{1}, \widetilde{\sigma}).$$

We claim that $\mathrm{Hom}_{y^{-1}N(\mathcal{P}^j)y}(\mathbb{1}, \widetilde{\sigma}) = (0)$. If not there is a nonzero vector $v \in W$ that is fixed by $y^{-1}N(\mathcal{P}^j)y$. Let $v' = \widetilde{\sigma}(n(b))v$ where $b = \varpi^r ut^{-1} \in \mathbb{O}$. Then v' is fixed by $n(b)y^{-1}N(\mathcal{P}^j)yn(-b)$. Note that

$$n(b)y^{-1}N(\mathcal{P}^j)yn(-b) = \begin{cases} \overline{N}(\mathbb{O}) & \text{if } r \geq 2\mathfrak{v}(t) \\ \overline{N}(\mathcal{P}^{2\mathfrak{v}(t)-r}) & \text{if } r < 2\mathfrak{v}(t). \end{cases}$$

If $r < 2\mathfrak{v}(t)$, then $2\mathfrak{v}(t) - r \leq l - 1$ (by the hypothesis of this case). Hence v' is fixed by $\overline{N}(\mathcal{P}^{l-1})$, which contradicts that $\widetilde{\sigma}$ is very cuspidal by Remark 3.3.2. This proves (i).

For the proof of (ii) and (iii), we begin by proving that if l is odd, then $(\pi')^{K_{2l}}$ has dimension 2, and if l is even, then $\dim(\pi^{K_{2l}}) = 2$. To this end, let

$$U = \left\{ \begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix} : u \in \mathbb{O}^\times \right\}.$$

Note that $\dim \sigma_\eta^U \geq 2$. (Recall that $\sigma = \mathrm{Res}_K \widetilde{\sigma}$. We use the notation σ_η^U to stand for the set of all vectors in σ on which U acts via η.) This can be seen as follows. Let χ be an eigencharacter of $N(\mathbb{O})$ occurring in σ. Since $\widetilde{\sigma}$ is very cuspidal, χ has conductor l. Let $v_\chi \in W$ be a nonzero eigenvector with eigencharacter χ. Let

$$w_1 = \sum_{u \in \mathbb{O}^\times / \pm(1+\mathcal{P}^l)} \eta(u)\sigma\left(\begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix}\right) v_\chi, \qquad w_2 = \widetilde{\sigma}(\gamma) w_1.$$

These sums make sense by the hypothesis on η and are nonzero since the summands lie in distinct eigenspaces. It is easy to see that $\mathbb{C}w_1 \oplus \mathbb{C}w_2 \subset \sigma_\eta^U$.

For any $w \in \sigma_\eta^U$ we define two elements $g_w \in \mathrm{ind}_K^G(\sigma)$ and $f_w \in \widetilde{\pi}$ given by

$$g_w(x) = \begin{cases} \sigma(x)w & \text{if } x \in K, \\ 0 & \text{if } x \notin K, \end{cases} \qquad f_w(x) = \widetilde{\pi}\left(\begin{pmatrix} \varpi^{-l} & 0 \\ 0 & 1 \end{pmatrix}\right) g_w.$$

It is easy to see that K_{2l} acts via η on f_w, that the map $w \mapsto f_w$ is injective, and finally that $f_w \in \pi$ if l is even and $f_w \in \pi'$ if l is odd. Now applying Lemma 3.1.3, we get that if l is even, then $\dim \pi_\eta^{K_{2l}} = 2$ and $(\pi')_\eta^{K_{2l}} = (0)$, whereas if l is odd then $\dim(\pi')_\eta^{K_{2l}} = 2$ and $\pi_\eta^{K_{2l}} = (0)$. (This also shows that $\dim \sigma_\eta^U = 2$.)

Now we use induction to obtain all the dimensions. Let l be even. (We leave the case of odd l to the reader since it is entirely analogous to the even case.) Since for every $m \geq 0$ we have that $K_{m+2} \subset K'_{m+1} \subset K_m$ up to G-conjugacy, it follows from Lemma 3.1.3 that the dimensions of the spaces $\pi_\eta^{K_{2l+1}}$, $(\pi')_\eta^{K_{2l+1}}$, $\pi_\eta^{K'_{2l+1}}$, and $(\pi')_\eta^{K'_{2l+1}}$ are all equal to 2. In fact, the same argument shows that if we know the dimensions of η-fixed vectors under K_{2m} and K'_{2m}, then we would know the dimensions for those under K_{2m+1} and K'_{2m+1}. Let us now suppose that we know $\dim \pi_\eta^{K_i}$ for $i \leq 2m+1$. In order to calculate $\dim \pi_\eta^{K_{2m+2}}$, we claim that

$$\pi_\eta^{K_{2m}} \oplus \begin{pmatrix} \varpi^{-1} & 0 \\ 0 & \varpi \end{pmatrix}^{m-l+1} \pi_\eta^{K_{2l}} \subset \pi_\eta^{K_{2m+2}}.$$

Clearly, both the summands in the left hand side are contained in the right hand side. We show that the sum is indeed a direct sum. Let v be a vector in the intersection of the two subspaces on the left hand side. Then both

$$K_{2l} \quad \text{and} \quad \left(\begin{smallmatrix} \varpi & 0 \\ 0 & \varpi^{-1} \end{smallmatrix}\right)^{m-l+1} K_{2m} \left(\begin{smallmatrix} \varpi^{-1} & 0 \\ 0 & \varpi \end{smallmatrix}\right)^{m-l+1},$$

and hence K_{2l-1}, act via the character η on the vector $\left(\begin{smallmatrix} \varpi & 0 \\ 0 & \varpi^{-1} \end{smallmatrix}\right)^{m-l+1} v$. Hence by (i) we get that $v = 0$. This implies that

$$\dim \pi_\eta^{K_{2m+2}} \geq 2 + \dim \pi_\eta^{K_{2m}} = 2 + \dim \pi_\eta^{K_{2m+1}}$$

by the induction hypothesis. Lemma 3.1.3 now says that this is an equality. Conjugating by α gives the dimensions for π'. $\square$

Proposition 3.3.5 (Test vectors for unramified supercuspidal L-packets of cardinality two). *Let $\widetilde{\sigma}$ be a very cuspidal representation of $Z\widetilde{K}$ of level l, which determines an unramified supercuspidal L-packet $\{\pi, \pi'\}$ as above. Assume that $\widetilde{\pi} = \mathrm{ind}_{Z\widetilde{K}}^{\widetilde{G}}(\widetilde{\sigma})$ is realized in its Kirillov model with respect to ψ. Define two elements ϕ_1 and ϕ_ϵ in the Kirillov model as follows*:

$$\phi_1(x) = \begin{cases} 1 & \text{if } x \in (\mathbb{O}^\times)^2, \\ 0 & \text{if } x \notin (\mathbb{O}^\times)^2, \end{cases} \qquad \phi_\epsilon(x) = \widetilde{\pi}(\gamma)\phi_1.$$

Let $\eta = \omega_{\widetilde{\pi}}$. We have

(i) $\mathbb{C}\phi_1 \oplus \mathbb{C}\phi_\epsilon = \widetilde{\pi}_\eta^{K_{2l}}$.

(ii) *If l is even then $\pi_\eta^{K_{2l}} = \widetilde{\pi}_\eta^{K_{2l}}$. In addition, π is ψ-generic and any ψ-Whittaker functional is nonzero on ϕ_1 and vanishes on ϕ_ϵ. Furthermore, π' is not ψ'-generic for any character ψ' of conductor $\mathbb{O}$. It is however ψ_ϖ-generic and any ψ_ϖ-Whittaker functional is nonvanishing on $\widetilde{\pi}(\alpha^{-1})\phi_1$ which is a newform for π'.*

(iii) *If l is odd, then (ii) holds with π and π' interchanged.*

Proof. We show that K_{2l} acts via η on ϕ_1. Given this, the rest of the assertions are all quite easy to show using the facts that α conjugates π to π', γ conjugates K_{2l} to itself, and the ψ-Whittaker functional on the ψ-Kirillov model is given by evaluation at 1.

To prove that K_{2l} acts on ϕ_1 via η, it is enough to prove, as in [Casselman 1973], that $B(\mathbb{O})$ acts on ϕ_1 via η and that $N(\mathbb{O})$ fixes $\widetilde{\pi}\left(\begin{smallmatrix} 0 & 1 \\ -\varpi^{2l} & 0 \end{smallmatrix}\right)\phi_1$. The former is easy to verify using the definition of ϕ_1. To address the latter, note that

$$\widetilde{\pi}\begin{pmatrix} 0 & 1 \\ -\varpi^{2l} & 0 \end{pmatrix}\phi_1 = \widetilde{\pi}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \varpi^{2l} & 0 \\ 0 & 1 \end{pmatrix}\right)\phi_1 = \widetilde{\pi}(w)\tau,$$

where $\tau = \widetilde{\pi}\left(\begin{smallmatrix} \varpi^{2l} & 0 \\ 0 & 1 \end{smallmatrix}\right)\phi_1$. It follows that $\tau(x) = \phi_1(\varpi^{2l}x)$. To show that $N(\mathbb{O})$ fixes $\widetilde{\pi}(w)\tau$, it suffices to show that the support of the function $\widetilde{\pi}(w)\tau$ is in $\mathbb{O}$. For this we need some information on the action of the Weyl group element w on functions in the Kirillov model. This is given in terms of the formal Mellin transform of Jacquet and Langlands [1970].

The formal Mellin transform of any function ξ in the Kirillov model for $\widetilde{\pi}$ is a formal power series in t defined for every character ν of $\mathbb{O}^\times$ as

$$\widehat{\xi}(\nu, t) := \sum_{n\in\mathbb{Z}} \xi_n(\nu)t^n := \sum_{n\in\mathbb{Z}} \left(\int_{u\in\mathbb{O}^\times} \xi(\varpi^n u)\nu(u)\,du\right)t^n.$$

Here we normalize the Haar measure du so that $\mathbb{O}^\times$ has volume 1. For every ν there is a formal series $c(\nu, t)$ such that

$$\widehat{\widetilde{\pi}(w)\xi}(\nu, t) = c(\nu, t)\widehat{\xi}(\omega_0^{-1}\nu^{-1}, z_0^{-1}t^{-1})$$

where ω_0 is the restriction of the central character $\omega_{\widetilde{\pi}}$ to $\mathbb{O}^\times$ and $z_0 = \omega_{\widetilde{\pi}}(\varpi)$ (see Proposition 2.10 of [Jacquet and Langlands 1970]). Since $\widetilde{\pi}$ is supercuspidal, it follows from Equation 2.18.1, Proposition 2.23, and the proof of Theorem 2.18 of the same paper that $c(\nu, t)$ is a monomial in t of the form

$$c(\nu, t) = c_0(\nu)t^{n_\nu}, \qquad n_\nu = -c(\widetilde{\pi} \otimes \nu^{-1}) \le -2.$$

Using the definitions of ϕ_1 and τ and the orthogonality of characters we get

$$\widehat{\tau}(\nu, t) = \begin{cases} 0 & \text{if } \nu \ne \mathbb{1} \text{ on } (\mathbb{O}^\times)^2, \\ c_1 t^{-2l} & \text{if } \nu = \mathbb{1} \text{ on } (\mathbb{O}^\times)^2. \end{cases}$$

where $c_1 = \mathrm{vol}((\mathbb{O}^\times)^2)$. Hence we get

$$\widehat{\widetilde{\pi}(w)\tau}(\nu, t) = \begin{cases} 0 & \text{if } \nu \ne \omega_0^{-1} \text{ on } (\mathbb{O}^\times)^2, \\ c_2(\nu)t^{n_\nu+2l} & \text{if } \nu = \omega_0^{-1} \text{ on } (\mathbb{O}^\times)^2. \end{cases}$$

for some nonzero constant $c_2(\nu)$. Now $N(\mathbb{O})$ fixes $\widetilde{\pi}(w)\tau$ if the function $\widetilde{\pi}(w)\tau$ is supported on $\mathbb{O}$ (since the conductor of ψ is $\mathbb{O}$), and the latter is true if we show that $n_\nu + 2l \ge 0$ for any character ν which is ω_0^{-1} on $(\mathbb{O}^\times)^2$. In other words, we need to show that $c(\widetilde{\pi} \otimes \nu^{-1}) \le c(\widetilde{\pi})$. (In fact, minimality of $\widetilde{\pi}$ then forces equality, which would imply that the function $\widetilde{\pi}(w)\tau$ is supported on $\mathbb{O}^\times$.)

To prove this inequality, using the local Langlands correspondence (see [Kudla 1994] for instance), we consider the Langlands parameter $\varphi = \varphi(\widetilde{\pi})$ of $\widetilde{\pi}$ which is a two-dimensional irreducible representation of the Weil group W_F of F. Since the residue characteristic is not 2, we get that φ is induced from a Galois regular character χ of E^* for the unramified quadratic extension E/F. If $c(\varphi)$ denotes the local Artin conductor of φ, then we have, using Proposition 4(b) of §4.3 in [Serre 1967], $c(\widetilde{\pi}) = c(\varphi) = 2c(\chi)$. Since ν is ω_0^{-1} on $(\mathbb{O}^\times)^2$, we have $c(\nu) \le c(\omega_{\widetilde{\pi}})$ unless $c(\omega_{\widetilde{\pi}}) = 0$ and $c(\nu) = 1$. Suppose the former condition holds. The central character is the determinant of the Langlands parameter and hence we get $\omega_{\widetilde{\pi}} = \det(\varphi(\widetilde{\pi})) = \det(\mathrm{ind}_{W_E}^{W_F}(\chi)) = \omega_{E/F}\chi|_{F^*}$. Since E/F is unramified we have $2c(\nu^{-1}) \le 2c(\omega_{\widetilde{\pi}}) = 2c(\chi|_{F^*}) \le 2c(\chi) = c(\widetilde{\pi})$. For every character κ of F^* we have the inequality $c(\widetilde{\pi} \otimes \kappa) \le \max\{c(\widetilde{\pi}), 2c(\kappa)\}$ (see §4 of [Gross 1988]). Applying this to $\kappa = \nu^{-1}$ and using the preceding information on the conductor of ν^{-1}, we get the required inequality $-n_\nu = c(\widetilde{\pi} \otimes \nu^{-1}) \le \max\{c(\widetilde{\pi}), 2c(\nu^{-1})\} = c(\widetilde{\pi}) = 2l$. If, on the other hand, $c(\omega_{\widetilde{\pi}}) = 0$ and $c(\nu) = 1$, then $c(\widetilde{\pi} \otimes \nu^{-1}) \le c(\widetilde{\pi})$ follows easily. $\square$

We now state the results for the supercuspidal L-packets of cardinality four. We omit the proofs since they are minor modifications of the corresponding statements for the unramified supercuspidal L-packets that we just dealt with.

Proposition 3.3.6 (Unramified supercuspidal L-packet of cardinality four). *Let $\widetilde{\sigma}$ denote a very cuspidal representation of $Z\widetilde{K}$ of level $l = 1$ such that* $\mathrm{Res}_K(\widetilde{\sigma}) = \sigma = \sigma_1 \oplus \sigma_2$. *Let $\{\pi_1, \pi_1', \pi_2, \pi_2'\}$ be the corresponding L-packet of G. Then $c(\pi_1) = c(\pi_1') = c(\pi_2) = c(\pi_2') = 2$. Moreover,*

(i) *Let η be any character such that $\eta(-1) = \omega_\sigma(-1)$. If π denotes any representation in the L-packet, then $\pi_\eta^{K_1} = \pi_\eta^{K_1'} = (0)$.*

(ii) *Let η be any character such that $\eta(-1) = \omega_\sigma(-1)$ and $c(\eta) \le 1$. Then for all $m \ge 2$ we have*
   (a) $\dim(\pi_1)_\eta^{K_m'} = \dim(\pi_2)_\eta^{K_m'} = \dim(\pi_1')_\eta^{K_m} = \dim(\pi_2')_\eta^{K_m} = \lceil \frac{m-1}{2} \rceil$.
   (b) $\dim(\pi_1)_\eta^{K_m} = \dim(\pi_1')_\eta^{K_m'} = \dim(\pi_2)_\eta^{K_m} = \dim(\pi_2')_\eta^{K_m'} = \lfloor \frac{m-1}{2} \rfloor$.

Corollary 3.3.7 (Test vectors for unramified supercuspidal L-packets of cardinality four). *With notation as above let $\{\pi_1, \pi_1', \pi_2, \pi_2'\}$ be the unramified supercuspidal L-packet of cardinality four. Let $\overline{\psi}$ be the character of $\mathbb{F}_q$ induced by ψ by identifying $\mathbb{F}_q$ with $\mathcal{P}^{-1}/\mathbb{O}$. Without loss of generality assume that σ_1 is $\overline{\psi}$-generic. Then*

(i) *π_1' is ψ-generic, π_1 is ψ_ϖ-generic, π_2' is ψ_ϵ-generic, and π_2 is $\psi_{\epsilon\varpi}$-generic.*

(ii) *Assume that $\widetilde{\pi}$ is realized in its ψ-Kirillov model. The function ϕ_1 of Proposition 3.3.5 is a newform for π_1'. This further implies that $\widetilde{\pi}(\alpha)(\phi_1)$ is a newform for π_1, $\widetilde{\pi}(\gamma)(\phi_1)$ is a newform for π_2' and $\widetilde{\pi}(\alpha\gamma)(\phi_1)$ is a newform for π_2. Finally, each of these newforms is a test vector for an appropriate Whittaker functional coming from* (*i*).

We now consider the ramified supercuspidal L-packets. Let $\widetilde{\sigma}$ be a very cuspidal representation of $N_{\widetilde{G}}(\widetilde{I})$ of level $l \ge 1$. Let $\mathrm{Res}_I \widetilde{\sigma} = \sigma_1 \oplus \sigma_2$ and let $\pi_i = \mathrm{ind}_I^G(\sigma_i)$. We call $\{\pi_1, \pi_2\}$ a *ramified supercuspidal L-packet* of G.

Proposition 3.3.8 (Ramified supercuspidal L-packets). *Let $\{\pi_1, \pi_2\}$ be a ramified supercuspidal L-packet as above. Then $c(\pi_1) = c(\pi_2) = 2l + 1$. We have*

(i) *For any character η of F^* such that $\eta(-1) = \omega_\sigma(-1)$ we have $(\pi_1)_\eta^{K_{2l}} = (\pi_2)_\eta^{K_{2l}} = (\pi_1)_\eta^{K_{2l}'} = (\pi_2)_\eta^{K_{2l}'} = (0)$.*

(ii) *Let $\eta(-1) = \omega_\sigma(-1)$ and $c(\eta) \le l$. For all $m \ge 2l + 1$ we have* $\dim(\pi_1)_\eta^{K_m} = \dim(\pi_2)_\eta^{K_m} = \dim(\pi_1)_\eta^{K_m'} = \dim(\pi_2)_\eta^{K_m'} = m - 2l$.

Proof. Since γ conjugates π_1 to π_2, $\dim(\pi_1)_\eta^{K_m} = \dim(\pi_2)_\eta^{K_m}$ for all m. Furthermore, conjugation by α stabilizes both π_1 and π_2, which implies this same equality for K_m'.

That $(\pi_1)_\eta^{K_{2l}} = (0)$ can be seen using Mackey theory [Kutzko 1977] as in the proof of Proposition 3.3.4 by considering the set of representatives for

$$K_{2l}\backslash\widetilde{G}/N_{\widetilde{G}}(\widetilde{I})$$

given by

$$\{\bar{n}(s)h(u)a_r : s \in \mathcal{P}/\mathcal{P}^{2l}, u \in \mathbb{O}^\times, r \geq 0\} \cup \{\bar{n}(t)wh(u)a_r : t \in \mathbb{O}/\mathcal{P}^{2l}, u \in \mathbb{O}^\times, r \geq 0\}.$$

To prove (ii), we use induction on the level m of the congruence subgroups K_m or K'_m. To begin the induction, we show that for $m = 2l+1$, all the relevant spaces are one-dimensional. It suffices, using Lemma 3.1.3, to prove that $\dim(\pi_1)_\eta^{K_{2l+1}} \geq 1$. This is done as in the proof of Proposition 3.3.4. Note that $\widetilde{\sigma}_\eta^U$ is at least two-dimensional and contains the span of w_1 and w_2. Since γ conjugates σ_1 to σ_2, we get that both $(\sigma_i)_\eta^U$ are nonzero. Let us say that $w_i \in \sigma_i^U$. Then the corresponding f_{w_i} is in $(\pi_i)_\eta^{K_{2l+1}}$. To proceed with the induction argument, we note that, by Lemma 3.1.3, we need only show that for each $m \geq 2l+2$, the dimension of $(\pi_1)_\eta^{K_m}$ is at least $m - 2l$. This follows from $(\pi_1)_\eta^{K_{m-1}} \oplus \widetilde{\pi}(\beta)^{m-2l-1}(\pi_1)_\eta^{K_{2l+1}} \subset (\pi_1)_\eta^{K_m}$. This inclusion and the fact that the sum is direct is proved exactly as in the proof of Proposition 3.3.4. □

Proposition 3.3.9 (Test vectors for ramified supercuspidal L-packets). *Let $\{\pi_1, \pi_2\}$ be a ramified supercuspidal L-packet coming from a very cuspidal representation $\widetilde{\sigma}$ of $N_{\widetilde{G}}(\widetilde{I})$ of level $l \geq 1$. One and only one of the π_i is ψ-generic, say π_1. Then π_2 is ψ_ϵ-generic. Let $\eta = \omega_\sigma$. Let ϕ_1 and ϕ_ϵ be as in Proposition 3.3.5. We have*

(i) $(\pi_1)_\eta^{K_{2l+1}} = \mathbb{C}\phi_1$ *and* $(\pi_2)_\eta^{K_{2l+1}} = \mathbb{C}\phi_\epsilon$.

(ii) *Any ψ-Whittaker functional is nonzero on ϕ_1 and similarly any ψ_ϵ-Whittaker functional is nonzero on ϕ_ϵ.*

Proof. The proof is entirely analogous to the proof of Proposition 3.3.5. In fact, using the notation in that proof, it suffices now to show that the support of $\widetilde{\pi}(w)\tau$ is in $\mathbb{O}$, where $\tau = \widetilde{\pi}(\alpha^{2l+1})\phi_1$. We can see as before that

$$\widehat{\widetilde{\pi}(w)\tau}(\nu, t) = \begin{cases} 0 & \text{if } \nu \neq \omega_0^{-1} \text{ on } (\mathbb{O}^\times)^2, \\ c_1(\nu)t^{n_\nu+2l+1} & \text{if } \nu = \omega_0^{-1} \text{ on } (\mathbb{O}^\times)^2 \end{cases}$$

where $\omega_0 = \omega_{\widetilde{\pi}}|_{\mathbb{O}^\times}$. As before, we need to show that if $\nu = \omega_0^{-1}$ on $(\mathbb{O}^\times)^2$, then in fact $n_\nu + 2l + 1 = 0$. Since $\widetilde{\pi}$ is a ramified supercuspidal representation, its Langlands parameter $\varphi = \varphi(\widetilde{\pi})$ is a two dimensional irreducible representation of the Weil group W_F of F that is induced from a Galois regular character χ of E^* for a ramified quadratic extension E/F. From Proposition 4(b) of §4.3 in [Serre 1967], we get that $c(\widetilde{\pi}) = c(\varphi) = c(\chi)+1$. By Theorem 3.3.3, $c(\widetilde{\pi}) = 2l+1$ which implies that $c(\chi) = 2l \geq 2$. This together with the fact that E/F is ramified gives

$2c(\chi|_{F^*}) \leq c(\chi)$. As in Proposition 3.3.5 we have $\omega_{\widetilde{\pi}} = \omega_{E/F}\chi|_{F^*}$. Hence we get $2c(\nu^{-1}) \leq 2c(\omega_{\widetilde{\pi}}) \leq \max\{2, 2c(\chi|_{F^*})\} \leq c(\chi) < c(\chi)+1 = c(\widetilde{\pi})$. We deduce (using for instance §4 of [Gross 1988]) that $-n_\nu = c(\widetilde{\pi}\otimes\nu^{-1}) = c(\widetilde{\pi}) = 2l+1$. □

3.4. ***Comparison of conductor with other invariants.*** We begin by recording the following theorem relating the conductor of a representation π of G to the conductor of a minimal representation of $\widetilde{G}$ that determines the L-packet containing π.

Theorem 3.4.1 (Relation between $c(\pi)$ and $c(\widetilde{\pi})$). *Let π be an irreducible admissible representation of $G = \mathrm{SL}_2(F)$. Let $\widetilde{\pi}$ be a representation of $\widetilde{G} = \mathrm{GL}_2(F)$ whose restriction to G contains π. Assume that $\widetilde{\pi}$ is minimal, i.e., $c(\widetilde{\pi}\otimes\chi) \geq c(\widetilde{\pi})$ for all characters χ of F^*. Then $c(\pi) = c(\widetilde{\pi})$.*

Proof. If π is a subquotient of a principal series representation $\pi(\chi)$ then the theorem follows from Propositions 3.2.4, 3.2.6, 3.2.8, 3.2.10 and 3.2.12 together with the easily verifiable fact that $\widetilde{\pi}$ may be taken as $\mathrm{Ind}_{\widetilde{B}}^{\widetilde{G}}(\chi\otimes\mathbb{1})$. If π is a supercuspidal representation then the theorem follows from Propositions 3.3.4, 3.3.6 and 3.3.8 while keeping in mind that Kutzko's construction (Theorem 3.3.3) actually gives minimal supercuspidal representations $\widetilde{\pi}$ of $\widetilde{G}$. □

Now we relate the conductor of a representation π of G with the depth $\rho(\pi)$ of π (a notion due to A. Moy and G. Prasad [1994]). We urge the reader to compare this theorem with a result from [Lansky and Raghuram 2003] where we determine such a relation for all discrete series representations of D^*, $\mathrm{GL}_n(F)$ and $\mathrm{GL}_2(D)$ for a central division algebra D over F. We also mention in passing that considering the action of $\mathrm{GL}_n(F)$ on the Bruhat–Tits building of $\mathrm{SL}_n(F)$ we get that the depth of every representation in an L-packet of $\mathrm{SL}_n(F)$ is the same.

Theorem 3.4.2 (Relation between the conductor $c(\pi)$ and the depth $\rho(\pi)$ for G). *Let π be an irreducible representation of G. Let $\rho(\pi)$ be the depth of π.*

(i) *If π is a subquotient of a principal series $\pi(\chi)$, then $\rho(\pi) = \max\{c(\pi)-1, 0\}$.*

(ii) *If π is an irreducible supercuspidal representation, then*

$$\rho(\pi) = \max\left\{\frac{c(\pi)-2}{2},\ 0\right\}.$$

Proof. The first statement is proved by the equalities $\rho(\pi) = \rho(\pi(\chi)) = \rho(\chi) = \max\{c(\chi)-1, 0\} = \max\{c(\pi)-1, 0\}$. He the first and second equality follow from [Moy and Prasad 1996] and the third from [Lansky and Raghuram 2003]. We omit the details of the proof of the second statement which can be proved almost exactly as in §4 of the latter paper. □

4. Towards multiplicity one for newforms

Given an irreducible representation π of $G = \mathrm{SL}_2(F)$ and a character η of F^* such that $c_\eta(\pi) = c(\pi)$, one might ask if we have $\dim V_\eta^{K_{c(\pi)}} = 1$. The answer is that this is often the case but not true in general. In fact we have $\dim V_\eta^{K_{c(\pi)}} = 1$ unless π is a representation in what we have called an unramified supercuspidal L-packet of cardinality two or if π is an irreducible principal series representations $\pi(\chi)$ such that χ is not quadratic but $\chi|_{\mathbb{O}^\times}$ is quadratic. For representations in these packets we get $\dim V_\eta^{K_{c(\pi)}} = 2$.

Nevertheless, in all cases we have proved that an appropriate Whittaker functional is nonvanishing on some newform. This can be used to formulate a kind of a multiplicity one result if we consider the quotient of the space $V_\eta^{K_{c(\pi)}}$ of newforms by the kernel of this Whittaker functional. More precisely, if η is such that $c_\eta(\pi) = c(\pi)$, Ψ is a nontrivial additive character of F of conductor either $\mathbb{O}$ or $\mathcal{P}^{-1}$ such that π is Ψ-generic, and Λ_Ψ is a Ψ-Whittaker functional for π, then we have

$$\dim \frac{V_\eta^{K_{c(\pi)}}}{V_\eta^{K_{c(\pi)}} \cap \mathrm{kernel}(\Lambda_\Psi)} = 1.$$

Another possibility is to consider some canonical nondegenerate bilinear form on the space $V^{K_{c(\pi)}}$ and consider the orthogonal complement of the subspace

$$V_\eta^{K_{c(\pi)}} \cap \mathrm{kernel}(\Lambda_\Psi)$$

as a candidate for a one-dimensional space of newforms. Then the multiplicity one result is formulated as $\dim\big(V_\eta^{K_{c(\pi)}} \cap \mathrm{kernel}(\Lambda_\Psi)\big)^\perp = 1$.

Acknowledgments

We thank Benedict Gross, Dipendra Prasad, Brooks Roberts and Paul Sally Jr. for some helpful correspondence.

References

[Bump 1997] D. Bump, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics **55**, Cambridge University Press, Cambridge, 1997. MR 97k:11080 Zbl 0868.11022

[Casselman 1973] W. Casselman, "On some results of Atkin and Lehner", *Math. Ann.* **201** (1973), 301–314. MR 49 #2558 Zbl 0239.10015

[Deligne 1973] P. Deligne, "Formes modulaires et représentations de GL(2)", pp. 55–105 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math **349**, Springer, Berlin, 1973. MR 50 #240 Zbl 0271.10032

[Gelbart and Knapp 1982] S. S. Gelbart and A. W. Knapp, "L-indistinguishability and R groups for the special linear group", *Adv. in Math.* **43**:2 (1982), 101–121. MR 83j:22009 Zbl 0493.22005

[Gel'fand et al. 1969] I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskii-Shapiro, *Generalized functions, VI: Representation theory and automorphic functions*, Academic Press, 1969. MR 91g:11052 Zbl 0235.43012

[Gross 1988] B. H. Gross, "Local orders, root numbers, and modular curves", *Amer. J. Math.* **110**:6 (1988), 1153–1182. MR 90b:11053 Zbl 0675.12011

[Gross and Prasad 1991] B. H. Gross and D. Prasad, "Test vectors for linear forms", *Math. Ann.* **291**:2 (1991), 343–355. MR 92k:22028 Zbl 0768.22004

[Jacquet and Langlands 1970] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Lecture Notes in Mathematics **114**, Springer, Berlin, 1970. MR 53 #5481 Zbl 0236.12010

[Jacquet et al. 1981] H. Jacquet, I. I. Piatetski-Shapiro, and J. Shalika, "Conducteur des représentations du groupe linéaire", *Math. Ann.* **256**:2 (1981), 199–214. MR 83c:22025 Zbl 0443.22013

[Kudla 1994] S. S. Kudla, "The local Langlands correspondence: the non-Archimedean case", pp. 365–391 in *Motives* (Seattle, WA, 1991), vol. 2, edited by U. Jannsen et al., Proc. Sympos. Pure Math. **55**, Amer. Math. Soc., Providence, RI, 1994. MR 95d:11065 Zbl 0811.11072

[Kutzko 1977] P. C. Kutzko, "Mackey's theorem for nonunitary representations", *Proc. Amer. Math. Soc.* **64**:1 (1977), 173–175. MR 56 #533 Zbl 0375.22005

[Kutzko 1978a] P. C. Kutzko, "On the supercuspidal representations of Gl_2", *Amer. J. Math.* **100**:1 (1978), 43–60. MR 58 #22411a Zbl 0417.22012

[Kutzko 1978b] P. C. Kutzko, "On the supercuspidal representations of Gl_2. II", *Amer. J. Math.* **100**:4 (1978), 705–716. MR 58 #22411b Zbl 0421.22012

[Kutzko and Sally 1983] P. C. Kutzko and P. J. Sally, Jr., "All supercuspidal representations of SL_l over a p-adic field are induced", pp. 185–196 in *Representation theory of reductive groups* (Park City, UT, 1982), edited by P. C. Trombi, Progr. Math. **40**, Birkhäuser, Boston, 1983. MR 85i:22029 Zbl 0538.22011

[Labesse and Langlands 1979] J.-P. Labesse and R. P. Langlands, "L-indistinguishability for SL(2)", *Canad. J. Math.* **31**:4 (1979), 726–785. MR 81b:22017 Zbl 0421.12014

[Lansky and Raghuram 2003] J. Lansky and A. Raghuram, "On the correspondence of representations between GL(n) and division algebras", *Proc. Amer. Math. Soc.* **131**:5 (2003), 1641–1648. MR 2003m:22021 Zbl 1027.22017

[Lansky and Raghuram 2004] J. Lansky and A. Raghuram, "Conductors and newforms for $U(1, 1)$", *Proc. Indian Acad. Sci. Math. Sci.* **114**:4 (2004), 319–343. MR 2005k:11099 Zbl 1059.11039

[Manderscheid 1984] D. Manderscheid, "On the supercuspidal representations of SL_2 and its two-fold cover", *Math. Ann.* **266**:3 (1984), 287–295, 297–305. MR 85d:22032 Zbl 0507.22014

[Mann 2001] W. R. Mann, *Local level raising for* GL(n), Ph.D. thesis, Harvard University, Cambridge, MA, 2001.

[Moy and Prasad 1994] A. Moy and G. Prasad, "Unrefined minimal K-types for p-adic groups", *Invent. Math.* **116**:1-3 (1994), 393–408. MR 95f:22023 Zbl 0804.22008

[Moy and Prasad 1996] A. Moy and G. Prasad, "Jacquet functors and unrefined minimal K-types", *Comment. Math. Helv.* **71**:1 (1996), 98–121. MR 97c:22021 Zbl 0860.22006

[Prasad and Raghuram 2000] D. Prasad and A. Raghuram, "Kirillov theory for $\mathrm{GL}_2(\mathfrak{D})$ where $\mathfrak{D}$ is a division algebra over a non-Archimedean local field", *Duke Math. J.* **104**:1 (2000), 19–44. MR 2001i:22024 Zbl 0958.22010

[Schmidt 2002] R. Schmidt, "Some remarks on local newforms for GL(2)", *J. Ramanujan Math. Soc.* **17**:2 (2002), 115–147. MR 2003g:11056 Zbl 0997.11040

[Serre 1967] J.-P. Serre, "Local class field theory", pp. 128–161 in *Algebraic number theory: Proceedings of the instructional conference* (Brighton, Sussex, 1965), Thompson, Washington, DC, 1967. Reprinted Academic Press, London, 1986. MR 36 #3753 Zbl 0959.13010

[Shelstad 1979] D. Shelstad, "Notes on L-indistinguishability", pp. 193–203 in *Automorphic forms, representations and L-functions* (Corvallis, OR., 1977), vol. 2, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, 1979. Based on a lecture of R. P. Langlands. MR 81m:12018 Zbl 0421.12015

[Shimizu 1977] H. Shimizu, "Some examples of new forms", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **24**:1 (1977), 97–113. MR 56 #5436 Zbl 0359.10023

[Waldspurger 1985] J.-L. Waldspurger, "Quelques propriétés arithmétiques de certaines formes automorphes sur GL(2)", *Compositio Math.* **54**:2 (1985), 121–171. MR 87g:11061a Zbl 0567.10022

Received August 30, 2005.

JOSHUA M. LANSKY
DEPARTMENT OF MATHEMATICS AND STATISTICS
AMERICAN UNIVERSITY
4400 MASSACHUSETTS AVENUE, NW
WASHINGTON, DC 20016
UNITED STATES

lansky@american.edu
http://www.american.edu/faculty/lansky/

A. RAGHURAM
DEPARTMENT OF MATHEMATICS
OKLAHOMA STATE UNIVERSITY
401 MATHEMATICAL SCIENCES
STILLWATER, OK 74078
UNITED STATES

araghur@math.okstate.edu
http://www.math.okstate.edu/∼raghuram/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

AN ABSOLUTE ESTIMATE OF THE HOMOGENEOUS EXPANSIONS OF HOLOMORPHIC MAPPINGS

TAISHUN LIU AND JIANFEI WANG

Let $f : \Omega \to \Omega$ be a holomorphic mapping, where Ω is one of the four classical domains in $\mathbb{C}^{m\times n}$. We show that, if $P = f(0)$, we have

$$\sum_{k=0}^{\infty} \frac{\left\| D\varphi_P(P)[D^k f(0)(Z^k)] \right\|_\Omega}{k!\, \left\| D\varphi_P(P) \right\|} < 1$$

for $\|Z\|_\Omega < \frac{1}{3}$ and $\varphi_P \in \operatorname{Aut}\Omega$ such that $\varphi_P(P) = 0$. This generalizes to higher dimensions a classical result of Bohr, which corresponds to the case $\Omega = \{z : |z| < 1\} \subset \mathbb{C}$. The constant $\frac{1}{3}$ is the best possible.

Let f be a holomorphic function from the unit disc $D \subset \mathbb{C}$ to itself, with Taylor expansion

$$f(z) = \sum_{k=0}^{\infty} a_k z^k.$$

Then

$$\sum_{k=0}^{\infty} |a_k z^k| < 1 \quad \text{for } |z| < \tfrac{1}{3}. \tag{0}$$

This result, known as Bohr's theorem, was originally obtained in [Bohr 1914] for $|z| < \frac{1}{6}$. That $\frac{1}{6}$ can be improved to $\frac{1}{3}$ and that this is the best possible constant was quickly realized independently by M. Riesz, I. Schur, and N. Wiener. New proofs were given in [Sidon 1927; Tomić 1962]. More recently, attention has been paid to multidimensional generalizations of Bohr's theorem [Boas and Khavinson 1997; Boas 2000; Defant et al. 2003; Dineen and Timoney 1989; 1991]. Such generalizations were obtained by studying the power series of a holomorphic function defined in

$$B_{\ell_p^n} := \Big\{ z \in \mathbb{C}^n : \|z\|_p = \Big(\sum_{k=1}^{n} |z_k|^p \Big)^{1/p} < 1 \Big\}$$

MSC2000: primary 30H05; secondary 32A05.
Keywords: holomorphic mapping, homogeneous expansion, classical domains, Bohr's theorem.
This work was supported by the National Natural Science Foundation of China Grant No. 10571164 and Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) Grant No. 20050358052.

with modulus less than 1. They can be summarized as follows:

$$\frac{1}{3\sqrt[3]{e}}\frac{1}{n^{1-1/p}} \le K < 3\Big(\frac{\log n}{n}\Big)^{1-1/p} \qquad \text{if } 1 \le p \le 2,$$

$$\frac{1}{3}\frac{1}{\sqrt{n}} \le K < 2\sqrt{\frac{\log n}{n}} \qquad \text{if } 2 \le p \le \infty,$$

where K is the supremum of $r \in [0, 1]$ such that $\sum_{\alpha \ge 0} |c_\alpha z^\alpha| < 1$ for $z \in rB_{\ell_p^n}$ whenever $\big|\sum_{\alpha\ge 0} c_\alpha z^\alpha\big| < 1$ for $z \in B_{\ell_p^n}$. Here the sum is taken over multi-indices $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, where the α_j are nonnegative integers. Aizenberg [2000, Theorem 9] established these inequalities for $p = 1$. Dineen and Timoney [1989] investigated the case $p = \infty$ and their result was clarified in [Boas and Khavinson 1997]. Boas [2000, Theorem 3] then generalized to $1 < p < \infty$.

The result of Aizenberg and Boas does not, strictly speaking, reduce to Bohr's classical theorem, as consideration of the case $n = 1$ shows. In this article, we give a new generalization of Bohr's theorem to higher dimensions. We investigate holomorphic mappings from Ω to Ω, where Ω is one of the four classical domains in $\mathbb{C}^n$ (see below), and demonstrate a result analogous to Bohr's, which reduces to it when $n = 1$. We also prove that the constant $\frac{1}{3}$ is best possible in higher dimensions. In the proof we use homogeneous expansions of holomorphic mappings, which replace multiple power series. The Minkowski norm in each of the four classical domains replaces the Euclidean norm, and certain properties of the automorphisms of these domains play an important role.

We first recall the definition of the four classical domains in the sense of Hua [1963]. Let $\mathbb{C}^{m\times n}$ denote the set of $m \times n$ matrices $Z = \big(z_{ij}\big)_{1\le i\le m,\, 1\le j\le n}$, with $z_{ij} \in \mathbb{C}$ and $1 \le m \le n$; denote by Z' and $\overline{Z}$, respectively, the transpose and the complex conjugate of Z.

The first classical domain, $\mathfrak{R}_I(m, n) \subset \mathbb{C}^{m\times n}$, consists of matrices Z such that $I_m - Z\overline{Z}' > 0$, where I_m is the identity matrix of rank m and the inequality sign means that the left-hand side is positive definite.

The second classical domain, $\mathfrak{R}_{II}(n) \subset \mathbb{C}^{n\times n}$, consists of Z such that $Z = Z'$ and $I_n - Z\overline{Z}' > 0$.

The third classical domain, $\mathfrak{R}_{III}(n) \subset \mathbb{C}^{n\times n}$, consists of Z such that $Z = -Z'$ and $I_n - Z\overline{Z}' > 0$.

The fourth classical domain, $\mathfrak{R}_{IV}(n) \subset \mathbb{C}^n$, is the set of $Z = (z_1, z_2, \dots, z_n)$ satisfying

$$|ZZ'|^2 + 1 - 2|Z|^2 > 0, \quad |ZZ'| < 1.$$

Let Ω denote one of the four classical domains or the unit polydisc $D^n \subset \mathbb{C}^n$. The span of Ω in the ambient space ($\mathbb{C}^{m\times n}$, $\mathbb{C}^{n\times n}$ or $\mathbb{C}^n$, as the case may be) is provided with a Minkowski functional $\|\cdot\|_\Omega$ arising from Ω [Liu and Ren 1998].

By results in [Liu 1989] and [Gong 1998], we know that

$$\|Z\|_\Omega = \sup\left\{|\alpha Z\beta'| : \alpha \in \partial B^m,\ \beta \in \partial B^n\right\}$$

if $\Omega = \mathcal{R}_I(m,n)$, $\mathcal{R}_{II}(n)$, or $\mathcal{R}_{III}(n)$ (with $m = n$ in the latter two cases), and this supremum equals the square root of the largest characteristic root of $Z\overline{Z}'$; if $\Omega = D^n$, then $\|Z\|_\Omega = \max\{|z_k| : 1 \le k \le n\}$; and if $\Omega = \mathcal{R}_{IV}(n)$, then

$$\|Z\|_\Omega = \sqrt{|Z|^2 + \sqrt{|Z|^4 - |ZZ'|^2}},$$

where $|Z|$ is the Euclidean norm in $\mathbb{C}^n$. Hence $\Omega = \mathcal{R}_I(m,n)$ is the unit ball of the complex Banach space $\mathbb{C}^{m\times n}$ with respect to the norm $\|\cdot\|_\Omega$. The subspaces $\{Z \in \mathbb{C}^{n\times n} : Z = Z'\}$ and $\{Z \in \mathbb{C}^{n\times n} : Z = -Z'\}$ are complex Banach spaces with respect to the norm $\|\cdot\|_\Omega$, for $\Omega = \mathcal{R}_{II}(n)$ and $\mathcal{R}_{III}(n)$ respectively, and Ω is the unit ball for that norm. $\mathbb{C}^n$ is a complex Banach space whose unit ball is $\Omega = \mathcal{R}_{IV}(n)$ for the norm $\|\cdot\|_\Omega$.

Let $\partial\Omega$ and $\partial_0\Omega$ denote the topological boundary and distinguished boundary of Ω. Denote by $H(\Omega,\Omega)$ the space of holomorphic mappings from Ω to Ω, and by $\operatorname{Aut}\Omega$ the group of holomorphic automorphisms of Ω. Let $\overline{\Omega}$ denote the closure of Ω. If T is a linear operator between normed linear spaces, we denote by $\|T\|$ its norm. Finally, $D^k f(Z)$ will mean the k-th Fréchet derivative of f at Z, where $f \in H(\Omega,\Omega)$ and k is a nonnegative integer.

Theorem. *Let $f : \Omega \to \Omega$ be holomorphic, where Ω is one of the classical domains, and set $P = f(0)$. Then*

$$\sum_{k=0}^{\infty} \frac{\left\|D\varphi_P(P)[D^k f(0)(Z^k)]\right\|_\Omega}{k!\,\left\|D\varphi_P(P)\right\|} < 1 \tag{1}$$

for $\|Z\|_\Omega < \frac{1}{3}$ and $\varphi_P \in \operatorname{Aut}\Omega$ such that $\varphi_P(P) = 0$.

If $\|Z\|_\Omega > \frac{1}{3}$, there exists a holomorphic map $f : \Omega \to \Omega$ such that (1) *is not valid.*

As already mentioned, if $\Omega = D \subset \mathbb{C}$, inequality (1) reduces to the relation (0) of page 155, recovering Bohr's classical theorem in one complex variable.

The proof of the theorem requires some lemmas, the first two of which are well known.

Lemma 1 [Liu 1989]. *Let $P \in \mathcal{R}_I(m,n)$. There is an $m \times m$ unitary matrix U and an $n \times n$ unitary matrix V for which P has the polar decomposition*

$$P = U \begin{pmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m & 0 & \cdots & 0 \end{pmatrix} V,$$

where $1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$ *and* $\|P\|_\Omega = \lambda_1$. *Set*

$$Q = U \begin{pmatrix} \frac{1}{\sqrt{1-\lambda_1^2}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{1-\lambda_m^2}} \end{pmatrix} \overline{U}', \quad R = \overline{V}' \begin{pmatrix} \frac{1}{\sqrt{1-\lambda_1^2}} & & & 0 \\ & \ddots & & \\ & & \frac{1}{\sqrt{1-\lambda_m^2}} & \\ 0 & & & I_{n-m} \end{pmatrix} V.$$

Then

$$\varphi_P(Z) = Q^{-1}(I_m - Z\overline{P}')^{-1}(P - Z)\,R \in \operatorname{Aut}\mathfrak{R}_I(m, n)$$

for $Z \in \overline{\mathfrak{R}_I(m, n)}$, *and hence*

$$D\varphi_P(P)(W) = -QWR$$

for $W \in \mathbb{C}^{m\times n}$.

Lemma 2 [Liu 1989]. *Given any* $A \in \mathfrak{R}_{IV}(n)$, *there exist a real orthogonal* $n \times n$ *matrix* T *and* $1 > \lambda_1 \geq \lambda_2 \geq 0$ *such that*

$$A = e^{i\theta}\left(\frac{\lambda_1+\lambda_2}{2},\, i\frac{\lambda_1-\lambda_2}{2},\, 0, \dots, 0\right) T \in \mathfrak{R}_{IV}(n) \tag{2}$$

and $\|A\|_\Omega = \lambda_1$, *where* $i = \sqrt{-1}$ *and* $\theta \in \mathbb{R}$. *Let*

$$Q = T' \begin{pmatrix} 1+\lambda_1\lambda_2 & 0 & 0 \\ 0 & 1-\lambda_1\lambda_2 & 0 \\ 0 & 0 & \sqrt{(1-\lambda_1^2)(1-\lambda_2^2)}\,I_{n-2} \end{pmatrix} T. \tag{3}$$

Then

$$\varphi_A(Z) = \frac{A + ZZ'\overline{A} - ZQ}{1 - 2Z\overline{A}' + ZZ'\overline{A}\overline{A}'} \in \operatorname{Aut}\mathfrak{R}_{IV}(n)$$

for any $Z \in \overline{\mathfrak{R}_{IV}(n)}$, *and hence*

$$D\varphi_A(A)(X) = X\frac{2A'\overline{A} - Q}{1 - 2|A|^2 + |AA'|^2}$$

for $X \in \mathbb{C}^n$.

Lemma 3. *Let* Ω *be one of the four classical domains. Then*

$$\|D\varphi_P(P)\| = \frac{1}{1 - \|P\|_\Omega^2}$$

for any $P \in \Omega$.

Proof. Case 1: Ω is one of $\mathfrak{R}_I(m,n)$, $\mathfrak{R}_{II}(n)$, $\mathfrak{R}_{III}(n)$. We assume without loss of generality that $\Omega=\mathfrak{R}_I(m,n)$. From Lemma 1 and the definition of $\|\cdot\|_\Omega$, we get

$$\begin{aligned}\|D\varphi_P(P)(W)\|_\Omega &= \sup\left\{|\alpha QWR\beta'| : \alpha\in\partial B^m,\ \beta\in\partial B^n\right\}\\ &\le \sup\left\{\frac{|\alpha W\beta'|}{1-\lambda_1^2} : \alpha\in\partial B^m,\ \beta\in\partial B^n\right\}=\frac{\|W\|_\Omega}{1-\lambda_1^2}\end{aligned}$$

for $W\in\mathbb{C}^{m\times n}$. This implies that

$$\|D\varphi_P(P)\|_\Omega\le\frac{1}{1-\lambda_1^2}=\frac{1}{1-\|P\|_\Omega^2}.$$

If we take $Z_0\in\overline{\mathfrak{R}_I(m,n)}$ with $\|Z_0\|_\Omega=1$ such that

$$\overline{U}'Z_0\overline{V}'=\begin{pmatrix}1&0&&0\\0&0&&0\\&&\ddots&\\0&0&&0\end{pmatrix},$$

we obtain

$$\begin{aligned}\|D\varphi_P(P)(Z_0)\|_\Omega&=\sup\left\{\left\|\alpha U\begin{pmatrix}(1-\lambda_1^2)^{-1}&0&0\\0&0&0\\&&\ddots\\0&0&0\end{pmatrix}V\beta'\right\|_\Omega : \alpha\in\partial B^m,\beta\in\partial B^n\right\}\\ &=\frac{1}{1-\lambda_1^2}.\end{aligned}$$

This shows that $\|D\varphi_P(P)\|_\Omega\ge\dfrac{1}{1-\lambda_1^2}=\dfrac{1}{1-\|P\|_\Omega^2}$, which leads to the desired conclusion.

Case 2: $\Omega=\mathfrak{R}_{IV}(n)$. Taking $A=P\in\Omega$ in Lemma 2 and expressing it as in (2), we see from the lemma that

$$D\varphi_A(A)(Z)=\frac{W}{1-2|A|^2+|AA'|^2},\tag{4}$$

where $W=Z(2A'\overline{A}-Q)$, with Q as in (3). It is clear that

$$1-2|A|^2+|AA'|=(1-\lambda_1^2)(1-\lambda_2^2),\tag{5}$$

where $1>\lambda_1\ge\lambda_2\ge0$, and that

$$W=ZT'\begin{pmatrix}\frac12(\lambda_1^2+\lambda_2^2-2)&-\frac{i}{2}(\lambda_1^2-\lambda_2^2)&0\\ \frac{i}{2}(\lambda_1^2-\lambda_2^2)&\frac12(\lambda_1^2+\lambda_2^2-2)&0\\0&0&-\sqrt{(1-\lambda_1^2)(1-\lambda_2^2)}I_{n-2}\end{pmatrix}T.$$

If $Z \in \partial_0\Omega$, that is, $Z = e^{i\theta}(x_1, x_2, \dots, x_n)$ with $x_k \in \mathbb{R}$ for $k = 1, 2, \dots, n$, it is clear that

$$|Z|^2 = |ZZ'| = x_1^2 + \cdots + x_n^2 = 1.$$

A simple computation then shows that

$$W\overline{W}' = \\ ZT' \begin{pmatrix} \frac{1}{2}\big((1-\lambda_1^2)^2+(1-\lambda_2^2)^2\big) & 0 & 0 \\ 0 & \frac{1}{2}\big((1-\lambda_1^2)^2+(1-\lambda_2^2)^2\big) & 0 \\ 0 & 0 & (1-\lambda_1^2)(1-\lambda_2^2)I_{n-2} \end{pmatrix} T\overline{Z}'.$$

Since T is a real orthogonal matrix we obtain

$$|W|^2 = W\overline{W}' \le \frac{(1-\lambda_1^2)^2 + (1-\lambda_2^2)^2}{2}|Z\overline{Z}'| = \frac{(1-\lambda_1^2)^2 + (1-\lambda_2^2)^2}{2},$$

where we have used the Schwarz inequality on the coefficient of I_{n-2}. Clearly,

$$WW' = (1-\lambda_1^2)(1-\lambda_2^2)ZZ'.$$

Hence

$$\|W\|_\Omega = \sqrt{|W|^2 + \sqrt{|W|^4 - |WW'|^2}} \le 1 - \lambda_2^2,$$

which together with (4) and (5) yields

$$\|D\varphi_A(A)(Z)\|_\Omega \le \frac{1}{1-\lambda_1^2} \quad \text{for } Z \in \partial_0\Omega.$$

If $Z \in \partial\Omega$, there exists a linear functional f satisfying

$$f(Z) = \|D\varphi_A(A)(Z)\|_\Omega, \quad \|f\| = 1.$$

The function g defined by $g(\xi) = f(D\varphi_A(A)(\xi))$ is holomorphic on $\overline{\Omega}$, so we obtain from the preceding inequality

$$|g(\xi)| \le \|D\varphi_A(A)(\xi)\|_\Omega \le \frac{1}{1-\lambda_1^2} \quad \text{for } \xi \in \partial_0\Omega.$$

On the other hand, the maximum principle gives

$$|g(Z)| = \|D\varphi_A(A)(Z)\|_\Omega \le \|D\varphi_A(A)(\xi)\|_\Omega \le \frac{1}{1-\lambda_1^2} \quad \text{for } \xi \in \partial_0\Omega.$$

Therefore

$$\|D\varphi_A(A)\| \le \frac{1}{1-\lambda_1^2} = \frac{1}{1-\|A\|_\Omega^2}. \tag{6}$$

There remains to show the reverse inequality,

$$\|D\varphi_A(A)\| \geq \frac{1}{1-\|A\|_\Omega^2}. \tag{7}$$

Take $Z_0 = (1, 0, 0, \dots, 0) \in \partial_0\Omega$. Then $\|Z_0\|_\Omega = 1$, and

$$\|D\varphi_A(A)(Z_0)\|_\Omega = \frac{1}{1-\lambda_1^2} = \frac{1}{1-\|A\|_\Omega^2}.$$

But this immediately implies (7), completing the proof. □

Proof of the Theorem. <u>*Case 1*</u>*:* $\Omega = D^n$. For $\|Z\|_\Omega < \frac{1}{3}$ it is easy to show that

$$\sum_{k=0}^{\infty} \frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} < 1.$$

On the other hand, when $\|Z\|_\Omega > \frac{1}{3}$, we let $\|Z_0\|_\Omega = |z_j^0| = \max_k\{|z_k^0|\} > \frac{1}{3}$ and define

$$f(Z) = \frac{p_j - z_j}{1 - p_j z_j},$$

where $1 > p_j > \frac{1}{2}(|z_j^0|^{-1} - 1)$. Then

$$\sum_{k=0}^{\infty} \frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} > 1.$$

<u>*Case 2*</u>*:* Ω is one of $\mathfrak{R}_I(m,n)$, $\mathfrak{R}_{II}(n)$, $\mathfrak{R}_{III}(n)$. We assume without loss of generality that $\Omega = \mathfrak{R}_I(m,n)$. Take $P = f(0) \in \Omega$ and express it as in Lemma 1, defining Q and R accordingly. The lemma then says that

$$\varphi_P\,(Z) = Q^{-1}(I_m - Z\overline{P}')^{-1}(P-Z)R$$

for any $Z \in \mathfrak{R}_I(m,n)$. From Lemma 3, we get

$$\|D\varphi_P(P)\|_\Omega = \frac{1}{1-\|P\|_\Omega^2}.$$

Since Ω is a convex domain, for a fixed k we can define

$$f_k(Z) = \sum_{j=1}^{k} \frac{f(e^{i2\pi j/k}Z)}{k}.$$

Then $f_k \in H(\Omega,\Omega)$. It is clear that

$$\frac{1}{k}\sum_{j=1}^{k} e^{i2\pi jl/k} = \begin{cases} 1 & \text{if } l \equiv 0 \pmod{k}, \\ 0 & \text{otherwise.} \end{cases}$$

From the homogeneous expansion of the holomorphic mapping f, we get

$$f_k(Z)=\frac{1}{k}\bigg(\sum_{j=1}^{k}\bigg(f(0)+\sum_{l=1}^{\infty}e^{i2\pi jl/k}\frac{D^l f(0)(Z^l)}{l!}\bigg)\bigg).$$

This implies that

$$\begin{aligned}\varphi_P\circ f_k(Z)&=\varphi_P\bigg(P+\sum_{l=1}^{\infty}\frac{D^{lk}f(0)(Z^{lk})}{(lk)!}\bigg)\\&=\varphi_P(P)+D\varphi_P(P)\bigg(\sum_{l=1}^{\infty}\frac{D^{lk}f(0)(Z^{lk})}{(lk)!}\bigg)+\cdots\\&=\frac{D\varphi_P(P)[D^k f(0)(Z^k)]}{k!}+\frac{D\varphi_P(P)[D^{2k}f(0)(Z^{2k})]}{(2k)!}+\cdots,\end{aligned}$$

and hence

$$\frac{D\varphi_P(P)[D^k f(0)(Z^k)]}{k!}=\frac{1}{2\pi}\int_0^{2\pi}\varphi_P\circ f_k(Ze^{i\theta})e^{-ik\theta}d\theta$$

since $\varphi_P\circ f_k$ is holomorphic and maps 0 to 0. Again because $\varphi_P\circ f_k\in H(\Omega,\Omega)$, we have

$$\frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!}<1$$

for any $Z\in\Omega$. Thus

$$\frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!}\le 1$$

for any $Z\in\overline{\Omega}$. This shows that

$$\frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!}=\|Z\|_\Omega^k\frac{\|D\varphi_P(P)[D^k f(0)(Z^k/\|Z\|^k)]\|_\Omega}{k!}\le\|Z\|_\Omega^k.$$

Using the equality $\|D\varphi_P(P)\|=\dfrac{1}{1-\|P\|_\Omega^2}$ from Lemma 3, we then get

$$\begin{aligned}\sum_{k=0}^{\infty}\frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|}&\le\|P\|_\Omega+(1-\|P\|_\Omega^2)\sum_{k=1}^{\infty}\|Z\|_\Omega^k\\&<\|P\|_\Omega+(1-\|P\|_\Omega^2)\sum_{k=1}^{\infty}\Big(\frac{1}{3}\Big)^k\\&=\|P\|_\Omega+\frac{1-\|P\|_\Omega^2}{2}=1-\frac{(1-\|P\|_\Omega)^2}{2}<1\end{aligned}$$

for $\|Z\|_\Omega<\frac{1}{3}$.

There remains to show that $\frac{1}{3}$ is the best possible constant. In fact, if $Z \in \Omega$ with $\|Z\|_\Omega > \frac{1}{3}$, we take

$$Z_0 = \begin{pmatrix} \mu_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_m & 0 & \cdots & 0 \end{pmatrix} \in \mathfrak{R}_I(m,n),$$

where $1 > \mu_1 \geq \mu_2 \geq \cdots \geq \mu_m \geq 0$ and $\|Z_0\|_\Omega = \mu_1 > \frac{1}{3}$. Take $p_{11} \in \mathbb{R}$ such that

$$\frac{1}{2}\Big(\frac{1}{\mu_1} - 1\Big) < p_{11} < 1. \tag{8}$$

If we define $f \in H(\Omega, \Omega)$ by

$$f(Z) = \begin{pmatrix} \frac{p_{11} - z_{11}}{1 - p_{11} z_{11}} & 0 & & 0 \\ 0 & 0 & & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \end{pmatrix},$$

we obtain successively

$$P = f(0) = \begin{pmatrix} p_{11} & 0 & & 0 \\ 0 & 0 & & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \end{pmatrix},$$

$$Q = \begin{pmatrix} (1 - p_{11}^2)^{-1/2} & 0 \\ 0 & I_{m-1} \end{pmatrix}, \quad R = \begin{pmatrix} (1 - p_{11}^2)^{-1/2} & 0 \\ 0 & I_{n-1} \end{pmatrix},$$

$$\|D\varphi_P(P)P\|_\Omega = \frac{p_{11}}{1 - p_{11}^2}, \quad \frac{D^k f(0)(Z^k)}{k!} = (p_{11}^2 - 1)p_{11}^{k-1} z_{11}^k \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 0 & & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \end{pmatrix}$$

for $k \geq 1$. This implies that

$$\frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!} = p_{11}^{k-1}|z_{11}|^k = p_{11}^{k-1}\mu_1^k$$

when $k \geq 1$. In view of the definition of Z_0, we get

$$\|D\varphi_P(P)\| = \frac{1}{1 - p_{11}^2}.$$

Therefore

$$\sum_{k=0}^{\infty}\frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} = p_{11}+(1-p_{11}^2)\sum_{k=1}^{\infty}p_{11}^{k-1}\mu_1^k$$
$$= p_{11}+(1-p_{11}^2)\frac{\mu_1}{1-p_{11}\mu_1}.$$

Then we immediately get from (8) the desired inequality

$$\sum_{k=0}^{\infty}\frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} > 1.$$

Case 3: $\Omega=\mathfrak{R}_{IV}$. The proof of (1) for $\|Z\|_\Omega<\frac{1}{3}$ changes little from Case 2. To show that $\frac{1}{3}$ is best possible, let $Z_0=(\mu_1,0,\dots,0)\in\mathfrak{R}_{IV}(n)$, with $1>\mu_1>\frac{1}{3}$. We have $\|Z_0\|_\Omega=\mu_1$, by [Liu 1989]. Therefore

$$Q=\begin{pmatrix}1+\mu_1^2 & 0 & 0\\ 0 & 1-\mu_1^2 & 0\\ 0 & 0 & (1-\mu_1^2)I_{n-2}\end{pmatrix}.$$

Take $p_{11}\in\mathbb{R}$ with $\frac{1}{2}\Big(\frac{1}{\mu_1}-1\Big)<p_{11}<1$ and define

$$f(Z)=\left(\frac{p_{11}-z_1}{1-p_{11}z_1},0,\dots,0\right)\in H(\Omega,\Omega).$$

Then $P=f(0)=(p_{11},0,\dots,0)$. From Lemma 2 we obtain

$$\frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!} = p_{11}^{k-1}|\mu_1|^k$$

when $k\ge 1$. Hence, as required,

$$\sum_{k=0}^{\infty}\frac{\|D\varphi_P(P)[D^k f(0)(Z_0^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} = p_{11}+(1-p_{11}^2)\sum_{k=1}^{\infty}p_{11}^{k-1}\mu_1^k$$
$$= p_{11}+(1-p_{11}^2)\frac{\mu_1}{1-p_{11}\mu_1}>1. \qquad \square$$

From the proof of the theorem, we have obtained in addition:

Corollary. *Let $P\in\Omega$ be given, where Ω is one of the four classical domains, and define*

$$\gamma_1=\frac{1}{2+\|P\|_\Omega},\qquad \gamma_2=\frac{1}{1+2\|P\|_\Omega}.$$

If $f : \Omega \to \Omega$ is a holomorphic mapping taking 0 to P, the inequality

$$\sum_{k=0}^{\infty} \frac{\|D\varphi_{P(P)}[D^k f(0)(Z^k)]\|_\Omega}{k!\,\|D\varphi_{P(P)}\|} < 1$$

holds for all Z such that $\|Z\|_\Omega < \gamma_1$. If $\|Z\|_\Omega > \gamma_2$, there exists $f \in H(\Omega, \Omega)$ with $f(0) = P$ such that the inequality fails.

This leads naturally to the following problem:

Question. *What is the best constant γ_P, depending on $\|P\|_\Omega$, such that*

$$\sum_{k=0}^{\infty} \frac{\|D\varphi_P(P)[D^k f(0)(Z^k)]\|_\Omega}{k!\,\|D\varphi_P(P)\|} < 1$$

whenever $\|Z\|_\Omega < \gamma_P$? According to the Corollary, $\gamma_P \in \left[\frac{1}{2+\|P\|_\Omega}, \frac{1}{1+2\|P\|_\Omega}\right]$.

References

[Aizenberg 2000] L. Aizenberg, "Multidimensional analogues of Bohr's theorem on power series", *Proc. Amer. Math. Soc.* **128**:4 (2000), 1147–1155. MR 2000i:32001 Zbl 0948.32001

[Boas 2000] H. P. Boas, "Majorant series", *J. Korean Math. Soc.* **37**:2 (2000), 321–337. Several complex variables (Seoul, 1998). MR 2001j:32001 Zbl 0965.32001

[Boas and Khavinson 1997] H. P. Boas and D. Khavinson, "Bohr's power series theorem in several variables", *Proc. Amer. Math. Soc.* **125**:10 (1997), 2975–2979. MR 98i:32002 Zbl 0888.32001

[Bohr 1914] H. Bohr, "A theorem concerning power series", *Proc. London. Math. Soc. (2)* **13** (1914), 1–5. JFM 44.0289.01

[Defant et al. 2003] A. Defant, D. García, and M. Maestre, "Bohr's power series theorem and local Banach space theory", *J. Reine Angew. Math.* **557** (2003), 173–197. MR 2004d:46012 Zbl 1031.46014

[Dineen and Timoney 1989] S. Dineen and R. M. Timoney, "Absolute bases, tensor products and a theorem of Bohr", *Studia Math.* **94**:3 (1989), 227–234. MR 91e:46006 Zbl 0713.46005

[Dineen and Timoney 1991] S. Dineen and R. M. Timoney, "On a problem of H. Bohr", *Bull. Soc. Roy. Sci. Liège* **60**:6 (1991), 401–404. MR 93e:46050 Zbl 0795.46034

[Gong 1998] S. Gong, *Convex and starlike mappings in several complex variables*, Mathematics and its Applications **435**, Kluwer, Dordrecht, 1998. MR 2000c:32054 Zbl 0926.32007

[Hua 1963] L. K. Hua, *Harmonic analysis of functions of several complex variables in the classical domains*, Translations of Mathematical Monographs **6**, Amer. Math. Soc., Providence, RI, 1963. MR 30 #2162 Zbl 0507.32025

[Liu 1989] T. S. Liu, *The growth theorems and covering theorems for biholomorphic mappings on classical domains*, Dissertation, University of Science and Technology of China, Hefei, 1989. In Chinese.

[Liu and Ren 1998] T. Liu and G. Ren, "Decomposition theorem of normalized biholomorphic convex mappings", *J. Reine Angew. Math.* **496** (1998), 1–13. MR 99e:32034 Zbl 0894.32009

[Sidon 1927] S. Sidon, "Über einen Satz von Herrn Bohr", *Math. Z.* **26**:1 (1927), 731–732. MR 1544888 JFM 53.0281.04

[Tomić 1962] M. Tomić, "Sur un théorème de H. Bohr", *Math. Scand.* **11** (1962), 103–106. MR 31 #316 Zbl 0109.30202

Received December 23, 2005. Revised September 25, 2006.

TAISHUN LIU
DEPARTMENT OF MATHEMATICS
HUZHOU TEACHERS COLLEGE
HUZHOU, ZHEJIANG, 313000
CHINA
lts@ustc.edu.cn

JIANFEI WANG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI, ANHUI, 230026
CHINA
wangjf@mail.ustc.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

A VARIATIONAL FORMULA FOR FLOATING BODIES

JOHN MCCUAN

Well known first order necessary conditions for a liquid mass to be in equilibrium in contact with a fixed solid surface declare that the free surface interface has mean curvature prescribed in terms of the bulk accelerations acting on the liquid and meets the solid surface in a materially dependent contact angle. We derive first order necessary conditions for capillary surfaces in equilibrium in contact with solid surfaces which may also be allowed to move. These conditions consist of the same prescribed mean curvature equation for the interface, the same prescribed contact angle condition on the boundary, and an additional integral condition which may be said to involve, somewhat surprisingly, only the wetted region.

An example of the kind of system under consideration is that of a floating ball in a fixed container of liquid. We apply our first order conditions to this particular problem.

1. Introduction: expressions for energy and volume

In the calculus of variations, one is interested in minimizing an energy functional over some class of admissible functions. The energy is usually assumed to be an integral operator. For example, one often considers

$$\mathcal{F}[u] = \int_{\Omega} F(x, u, Du)\, dx + \int_{\partial\Omega} \bar{F}(x, u, Du)\, dx \tag{1}$$

where Ω is a fixed domain in $\mathbb{R}^n$ and the competing functions u are all defined on $\bar{\Omega}$.

Following Gauss, a standard approach to the derivation of equations governing equilibrium configurations for liquid masses in contact with solid bounding surfaces is via the calculus of variations. In this context, due to the geometric and in particular parametric nature of the situation, it is more natural to consider energy functionals which are integral operators on a given Riemann surface M and which involve a parametric mapping $X : M \to \mathbb{R}^3$. Though the value of the energy involves

---

MSC2000: primary 76B45, 76D45, 49K20; secondary 49Q05, 53A10.
Keywords: calculus of variations, capillarity, minimal surfaces, constant mean curvature.
The author is grateful to the Max Planck Institute for Mathematics in the Sciences, which supported this research.

universally the inclusion of surface area terms for the liquid *free surface interface* and the *wetted region* on the solid support surface, the precise identification in the literature of the Riemann surface M and the mapping X is somewhat more obscure.

In most, if not all, cases it is tacitly assumed that the entire configuration is determined by the free surface interface and, hence, that the functional $\mathscr{E} = \mathscr{E}[X]$ has as argument the mapping X which parameterizes the free surface interface. More precisely, it is assumed that the mapping X which determines the free surface interface also determines the contact line (the boundary of the free interface); the contact line determines a particular wetted region on the solid surface; the free surface interface together with the wetted region determine the enclosed volume or liquid mass, and this is enough information to compute the energy. Competing admissible mappings are then naturally considered as those mappings defined on the image Λ of X, or equivalently on its preimage M. Each such *free surface interface* leads to an admissible configuration defining a liquid mass as just described. From this point of view, the objective may be stated as finding a piecewise C^1 mapping X_0 determining a free surface interface for which $\mathscr{E}(X_0) \le \mathscr{E}(X)$ for all piecewise C^1 competing or admissible mappings X. It should be noted that in this formulation, the energy is given implicitly at least to the extent that a particular integral describing, say, the wetted area, is never written down. Since the first variation formula in this context only involves integrals over Λ and $\partial\Lambda$, this point of view seems at first adequately justified, at least formally.

We give here an alternative approach which is somewhat more explicit at least to the extent that the initial energy is given in terms of a well defined integral operator. In situations in which portions of the solid boundary are free to move, or partially free to move, the image of the free surface does not always determine the wetted region or the volume; consider a cylindrical container

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 2,\ -3 \le z \le 1\} \cup \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 2,\ z = -3\}$$

with the floating solid cylinder

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 1,\ -2 + h \le z \le 1 + h\}$$

for h fixed with $-1 < h < 2$. An admissible interface for every value of h is given by the annulus $\{(x, y, 0) \in \mathbb{R}^3 : 1 < x^2 + y^2 < 2\}$; while this interface determines the contact line, the wetted region and enclosed volume both depend explicitly on h. In such situations, our alternative and more general approach is more natural, being further justified by the fact that additional integral terms, not expressible as integrals over the free surface interface or its boundary, appear in the first order necessary conditions. Furthermore, the explicit appearance of an independent parameter h in this type of problem suggests an analogous class of problems in the nonparametric case different from (1). Namely, one may consider

$$\mathcal{F}[u, h] = \int_{\Omega} F(x, u, Du, h)\, dx + \int_{\partial\Omega} \bar{F}(x, u, Du, h)\, dx,$$

where the admissible functions u are defined on $\Omega \times \tilde{\Omega}$ for some set $\tilde{\Omega} \in \mathbb{R}^m$. We are unaware of any treatment of problems of this sort in the literature, but one can think of some interesting examples with little difficulty.

We now give formal mathematical assumptions used to model capillary systems in somewhat more detail than we have found in the literature. We model solid surfaces by the boundaries of initially prescribed closed sets in $\mathbb{R}^3$. The union of all such solid surfaces is denoted by $\Sigma = \partial\mathscr{C}$ where $\mathscr{C}$ denotes the union of the closed sets. We assume, for the moment, that all solid surfaces are fixed, and an *interior liquid mass*, modeled by an open set $\mathcal{M}$, occupies some portion of $\mathbb{R}^3 \backslash \mathscr{C}$. The boundary of $\mathcal{M}$ consists of a portion $\mathscr{W}$ in Σ, which we call the *wetted region*, and a portion $\Lambda = \partial\mathcal{M} \backslash \mathscr{W}$, the *free surface interface*.

Given a particular liquid mass $\mathcal{M}_0$, we assume $\mathcal{M}_0$ and the associated surfaces Λ_0 and $\mathscr{W}_0$ admit the structure of abstract Riemannian manifolds (of dimensions 3, 2, and 2 respectively).

Relative to $\mathcal{M}_0$, we consider the admissible class of liquid mass configurations $\mathcal{M}$ obtained as parameterized images

$$X : \overline{\mathcal{M}}_0 \to \mathbb{R}^3.$$

More precisely, we consider liquid masses $\mathcal{M} = X(\mathcal{M}_0)$ with smooth parameterization $X\big|_{\mathcal{M}_0}$; it is assumed that $\partial\mathcal{M} = X(\Lambda_0) \cup X(\mathscr{W}_0)$ with free surface interface $\Lambda = X(\Lambda_0)$, parameterized by $X\big|_{\Lambda_0}$, and wetted region $\mathscr{W} = X(\mathscr{W}_0) = X(\overline{\mathcal{M}}_0) \cap \Sigma$ with parameterization $X\big|_{\mathscr{W}_0}$.

Given an admissible liquid mass configuration $\mathcal{M}$, we assume the energy of the system is given by

$$\mathscr{E} = A(\Lambda) - \beta A(\mathscr{W}) \tag{2}$$

where A denotes the area functional. The constant β is called the *adhesion coefficient*. See [Finn 1986] for further details. For simplicity, we have not included in (2) the possibility that bulk accelerations are acting on the liquid. The applications given in this paper do not involve gravity or other accelerations, but for reference we will briefly indicate below what occurs in this more general setting.

We will assume that each global parameterization, such as $X\big|_{\Lambda_0} : \Lambda_0 \to \Lambda$, extends smoothly to an open neighborhood of the boundary. In particular, we can parameterize a small neighborhood of every point in the closure of Λ locally by some map defined in a neighborhood of $\mathbb{R}^2$. We will employ the usual abuse of

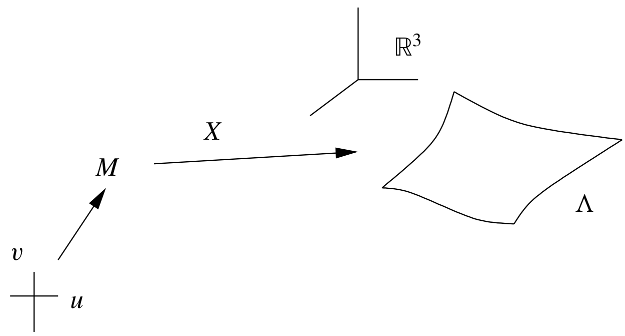


Figure 1. A free surface interface parameterized on an abstract manifold

notation in denoting this local parameterization also by X. More generally, we may emphasize the role of either parameter domain Λ_0 or $\mathcal{W}_0$ (or their union) as an abstract manifold by denoting it as M and referring to the appropriate restriction of $X : \overline{\mathcal{M}}_0 \to \mathbb{R}^3$ simply as $X : M \to \mathbb{R}^3$. The context is usually adequate to indicate which portion of $\partial\mathcal{M}$ is under consideration. If we wish to emphasize the distinction between the free surface Λ and wetted region $\mathcal{W}$, we will use a superscript "$\mathcal{W}$" in reference to the wetted region.

Under these assumptions, the area functional can be written explicitly as an integral over a fixed Riemann surface M of the Jacobian scaling factor $\sqrt{\det(dX^T dX)}$ where the superscript "T" denotes the transpose (or adjoint) of the linear transformation; see [Evans and Gariepy 1992]. We can write, for example,

$$A(\Lambda) = \int_\Lambda 1 = \int_{\Lambda_0} \sqrt{\det(dX^T dX)} = \int_M \sqrt{\det(dX^T dX)}.$$

Using a similar expression for the wetting energy, we see that the energy of admissible configurations is well defined as a functional on pairs of restrictions $X : M = \Lambda_0 \to \mathbb{R}^3$, $X^{\mathcal{W}} : M^{\mathcal{W}} = \mathcal{W}_0 \to \mathbb{R}^3$ on a pair of fixed Riemann surfaces (or alternatively on admissible maps X on the union $M = \Lambda_0 \cup \mathcal{W}_0$):

$$\mathcal{E}(X) = \int_M \sqrt{\det(dX^T dX)} - \beta \int_{M^{\mathcal{W}}} \sqrt{\det(dX^{\mathcal{W}T} dX^{\mathcal{W}})} \tag{3}$$

The volume of the liquid mass can also be expressed in terms of integrals over the same parameter domains. In fact, using the divergence theorem

$$V = \int_{\mathcal{M}} 1 = \frac{1}{3} \int_{\mathcal{M}} \operatorname{div}_{\mathbb{R}^3} x = \frac{1}{3} \int_\Lambda X \cdot N + \frac{1}{3} \int_{\mathcal{W}} X_0^{\mathcal{W}} \cdot N_0^{\mathcal{W}}.$$

Thus, we may write

$$(4)\quad V(X) = \frac{1}{3}\int_M X \cdot N\sqrt{\det(dX^T dX)} + \frac{1}{3}\int_{M^{\mathscr{W}}} X_0^{\mathscr{W}} \cdot N_0^{\mathscr{W}} \sqrt{\det(dX^{\mathscr{W}T} dX^{\mathscr{W}})}.$$

Expressions (3) and (4) provide the basis for the calculations in the next section.

When bulk accelerations are considered, one encounters in the energy an additional term of the form

$$\mathscr{G} = \int_{\mathscr{M}} g$$

where g is a given scalar function of position. I am unaware of a means to express this term as an integral over $\partial\mathscr{M}_0$. Thus, we are apparently forced to use the full information of the mapping $X : \overline{\mathscr{M}}_0 \to \mathbb{R}^3$ associated with an admissible variation of the bulk liquid and admit the dependence of the energy on the third restriction $X\big|_{\mathscr{M}_0} : \mathscr{M}_0 \to \mathbb{R}^3$, which we also denote by X:

$$\mathscr{G}(X) = \int_{\mathscr{M}_0} g \circ X |\det DX|.$$

2. General formulae

Given a liquid mass configuration as described above, we wish to consider a variation $X : \overline{\mathscr{M}}_0 \times (-\epsilon, \epsilon) \to \mathbb{R}^3$ such that $X(\,\cdot\,; 0) = \mathrm{id}_{\overline{\mathscr{M}}_0}(\,\cdot\,)$ and for each fixed $t \in (-\epsilon, \epsilon)$, the two restrictions

$$X\big|_{\Lambda_0 \times \{t\}} : \Lambda_0 = M \to \mathbb{R}^3 \quad \text{and} \quad X\big|_{\mathscr{W}_0 \times \{t\}} : \mathscr{W}_0 = M^{\mathscr{W}} \to \mathbb{R}^3,$$

which parameterize images Λ and $\mathscr{W}$ respectively, arise as restrictions to Λ_0 and $\mathscr{W}_0$ of the global map $X : \overline{\mathscr{M}}_0 \to \mathbb{R}^3$ associated to an admissible liquid mass $\mathscr{M} = \mathscr{M}_t$. The variations under consideration here are required to be smooth so that the functions

$$\mathscr{E}(t) = A(\Lambda) - \beta A(\mathscr{W}) \quad \text{and} \quad V(t) = \frac{1}{3}\int_\Lambda X \cdot N + \frac{1}{3}\int_{\mathscr{W}} X^{\mathscr{W}} \cdot N^{\mathscr{W}}$$

are differentiable on $(-\epsilon, \epsilon)$.

Finally, we note the compatibility of the restrictions to $\mathscr{W}_0$ and Λ_0 along their common boundary:

$$(5)\qquad X^{\mathscr{W}}(q; t) = X(q; t) \quad \text{for every } q \in \partial M^{\mathscr{W}} = \partial M$$

This simple relation will play a key role in our derivation of a variational formula for floating bodies.

With no bulk accelerations, no condition on the interior restriction $X\big|_{\mathscr{M}_0 \times \{t\}}$ need be considered, and this part of the global map can be ignored.

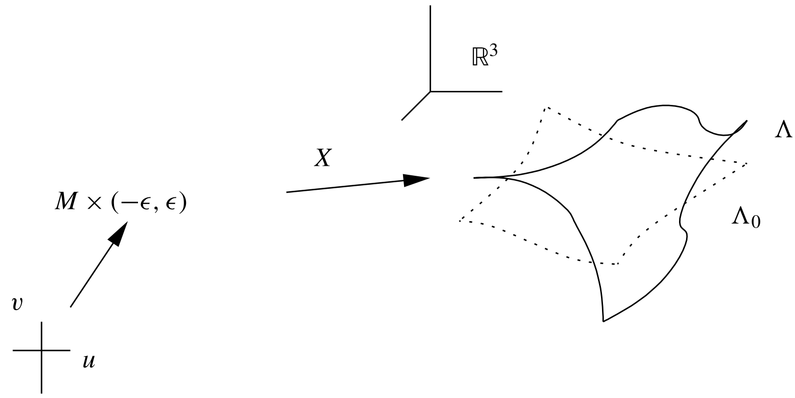


Figure 2. Variation of Λ_0.

We wish to compute $\dot{\mathscr{E}} = d\mathscr{E}/dt$. We begin with $A(\Lambda)$.

$$A(\Lambda) = \int_\Lambda 1 = \sum_U \int_U \sqrt{EG - F^2},$$

where we have written the abstract manifold M (up to a set of measure zero) as a disjoint union of images of coordinate neighborhoods U in the u, v-plane and E, F, and G are the coefficients of the first fundamental form on a given neighborhood U. That is, taking X to be locally defined on U, we have

$$E = |X_u|^2, \quad F = X_u \cdot X_v, \quad G = |X_v|^2, \quad \sqrt{EG - F^2} = |X_u \times X_v|.$$

Thus we obtain

$$\begin{aligned} \frac{d}{dt} \int_U \sqrt{EG - F^2} &= \int_U \frac{(X_u \times X_v) \cdot (\dot{X}_u \times X_v + X_u \times \dot{X}_v)}{\sqrt{EG - F^2}} \\ &= \int_U \frac{(X_u \cdot \dot{X}_u)G - (X_v \cdot \dot{X}_u)F + (X_v \cdot \dot{X}_v)E - (X_u \cdot \dot{X}_v)F}{\sqrt{EG - F^2}}. \end{aligned}$$

Now, if we write $\dot{X} = \phi N + \psi X_u + \eta X_v = \phi N + \dot{X}^{\mathrm{T}}$, where N is a unit normal to Λ, then

$$\begin{aligned} \dot{X}_u &= \phi_u N + \phi N_u + \psi_u X_u + \psi X_{uu} + \eta_u X_v + \eta X_{uv}, \\ \dot{X}_v &= \phi_v N + \phi N_v + \psi_v X_u + \psi X_{uv} + \eta_v X_v + \eta X_{vv}, \end{aligned}$$

so that

$$X_u \cdot \dot{X}_u = -\phi e + \psi_u E + \tfrac{1}{2}\psi E_u + \eta_u F + \tfrac{1}{2}\eta E_v,$$

$$X_v \cdot \dot{X}_u = -\phi f + \psi_u F + \psi F_u - \tfrac{1}{2}\psi E_v + \eta_u G + \tfrac{1}{2}\eta G_u,$$

$$X_u \cdot \dot{X}_v = -\phi f + \psi_v E + \tfrac{1}{2}\psi E_v + \eta_v F + \eta F_v - \tfrac{1}{2}\eta G_u,$$

$$X_v \cdot \dot{X}_v = -\phi g + \psi_v F + \tfrac{1}{2}\psi G_u + \eta_v G + \tfrac{1}{2}\eta G_v,$$

where $e = X_{uu} \cdot N$, $f = X_{uv} \cdot N$, and $g = X_{vv} \cdot N$ are the coefficients of the second fundamental form. Multiplying the expressions above by the appropriate factors, adding them together and grouping like terms we obtain

$$\begin{aligned}(X_u \cdot \dot{X}_u)G - (X_v \cdot \dot{X}_u)F + (X_v \cdot \dot{X}_v)E - (X_u \cdot \dot{X}_v)F \\ = -\phi(eG - 2fF + gE) + \psi_u(EG - F^2) + \tfrac{1}{2}\psi(EG - F^2)_u \\ + \eta_v(EG - F^2) + \tfrac{1}{2}\eta(EG - F^2)_v.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\frac{d}{dt}\int_U \sqrt{EG - F^2} \\ &\quad = -\int_U \phi\,\frac{eG - 2fF + gE}{\sqrt{EG - F^2}} + \int_U \psi_u\sqrt{EG - F^2} + \psi\,\frac{(EG - F^2)_u}{2\sqrt{EG - F^2}} \\ &\qquad\qquad + \int_U \eta_v\sqrt{EG - F^2} + \eta\,\frac{(EG - F^2)_v}{2\sqrt{EG - F^2}} \\ &\quad = -\int_U \phi\,\frac{eG - 2fF + gE}{\sqrt{EG - F^2}} + \int_U (\psi\sqrt{EG - F^2})_u + (\eta\sqrt{EG - F^2})_v.\end{aligned}$$

Finally, summing over the neighborhoods U we have

$$\begin{aligned}\dot{A}(\Lambda) &= -\int_\Lambda 2H\phi + \sum_U \int_U \frac{(\psi\sqrt{EG - F^2})_u + (\eta\sqrt{EG - F^2})_v}{\sqrt{EG - F^2}}\sqrt{EG - F^2} \\ &= -\int_\Lambda 2H\phi + \int_\Lambda \operatorname{div}_\Lambda \dot{X}^{\mathrm{T}} \\ &= -\int_\Lambda 2HX \cdot N + \int_{\partial\Lambda} \dot{X}^{\mathrm{T}} \cdot \vec{n},\end{aligned}$$

where H is the mean curvature of Λ, $\vec{n}$ is the outward conormal to $\partial\Lambda$ and the last term comes from the divergence theorem. We note finally, that $\dot{X}^{\mathrm{T}} \cdot \vec{n} = \dot{X} \cdot \vec{n}$ because, $N \cdot \vec{n} = 0$.

This formula should be recognized as the general formula for the first variation of area for an arbitrary surface in $\mathbb{R}^3$ under an arbitrary vector-valued variation. While the derivation above is somewhat more general than that of [Wente 1966] or [Finn 1986], since we allow arbitrary variations, Spivak [1979] derives an even more general version which allows arbitrary dimension and also immersion in an arbitrary submanifold. We felt it was worthwhile, however, to derive the generalization in the limited context of the present discussion.

We may apply the same calculation to the variation of $\mathcal{W}$ by $X^{\mathcal{W}}$ and obtain

$$\dot{A}(\mathcal{W}) = \int_{\partial\Lambda} \dot{X} \cdot \vec{\nu} \tag{6}$$

where ν is the outward conormal to $\partial\mathcal{W} = \partial\Lambda$ with respect to $\mathcal{W}$. Notice from (5) that

$$\dot{X}^{\mathcal{W}}(q;t) = \dot{X}(q;t)$$

for every $q \in \partial M^{\mathcal{W}}$, and in the case of a fixed rigid bounding surface Σ, we are requiring $\dot{X}^{\mathcal{W}} = \dot{X}$ to be tangent to Σ along $\partial\Lambda$. This is why the surface integral of $2H^{\mathcal{W}}\dot{X} \cdot N^{\mathcal{W}}$ does not appear in (6). Moreover, along $\partial\Lambda$ we may write $\dot{X} = (\dot{X} \cdot \vec{\nu})\vec{\nu} + \delta\vec{t}$ where $\vec{t}$ is a unit tangent to $\partial\Lambda$, so that

$$\dot{\mathscr{E}} = -\int_{\Lambda} 2HX \cdot N + \int_{\partial\Lambda} \dot{X} \cdot \vec{\nu}(\vec{\nu} \cdot \vec{n} - \beta).$$

If γ is the angle measured inside $\mathcal{M}$ between Λ and Σ along $\partial\Lambda$, then we obtain the familiar form

$$\dot{\mathscr{E}} = -\int_{\Lambda} 2HX \cdot N + \int_{\partial\Lambda} (\cos\gamma - \beta)\dot{X} \cdot \vec{\nu}. \tag{7}$$

We next proceed to derive an expression for

$$\dot{V} = \frac{d}{dt}V(t),$$

where

$$V = V(t) = \int_{\mathcal{M}} 1 = \frac{1}{3}\int_{\mathcal{M}} \operatorname{div}_{\mathbb{R}^3} x = \frac{1}{3}\int_{\Lambda} X \cdot N + \frac{1}{3}\int_{\mathcal{W}} X^{\mathcal{W}} \cdot N^{\mathcal{W}}$$

by the divergence theorem.

Note. The expression used for the change in $V(\mathcal{M})$ in [Barbosa et al. 1988] and referenced by [Ros and Souam 1997] apparently assumes that the variation mapping X (on $M \times [0, \epsilon]$) gives a parameterization of the volume under consideration. By following Wente [1966] (who followed Blaschke [1930]) we get an expression which is valid under this assumption and also without it. The computation, however, is a little more involved.

It is interesting to note that earlier, Barbosa and do Carmo [1984] use also Blaschke's expression for the volume. Presumably, the formulation in their latter paper [1988] is a consequence of working in a curved ambient manifold. This suggests the following question: *Is there a generalization of the trivial formula* $\operatorname{div}_{\mathbb{R}^n} x = n$ *to situations in which* $\mathbb{R}^n$ *is replaced by an n-dimensional manifold?*

The question in one lower dimension, also has relevance for the main problem under consideration in this paper, because if this were the case, it might be possible to replace the wetting energy term with an integral around $\partial\Lambda$ and, hence, simplify the overall construction. It should be noted, of course that in the case of a floating object, that vector field would be expected to depend on the position of the floating object.

Finally, in using Blaschke's expression for the volume, we have effectively replaced an integral over $\mathcal{M}$ by one over $\partial\mathcal{M}$. If we are willing to consider variation of the entire liquid mass (which we are apparently forced to do below to accommodate bulk accelerations), then an alternative approach is possible. This approach is presented in the context of the more complicated situation below.

Returning to the variation of Blaschke's volume expression, we begin as before, writing Λ as a union of parameter neighborhoods $X(U)$. Then

$$\frac{d}{dt}\int_\Lambda X\cdot N = \sum_U \frac{d}{dt}\int_U X\cdot(X_u\times X_v).$$

For each local domain U,

$$\frac{d}{dt}\int_U X\cdot(X_u\times X_v) = \int_U \dot{X}\cdot(X_u\times X_v)+X\cdot(\dot{X}_u\times X_v)+X\cdot(X_u\times\dot{X}_v).$$

Considering the last two terms in this expression,

$$\begin{aligned}
&X\cdot(\dot{X}_u\times X_v)+X\cdot(X_u\times\dot{X}_v)\\
&\quad=[X\cdot(\dot{X}\times X_v)]_u+[X\cdot(X_u\times\dot{X})]_v-X_u\cdot(\dot{X}\times X_v)-X\cdot(X_{uv}\times\dot{X})\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad-X_v\cdot(X_u\times\dot{X})-X\cdot(\dot{X}\times X_{uv})\\
&\quad=[X\cdot(\dot{X}\times X_v)]_u+[X\cdot(X_u\times\dot{X})]_v+2\dot{X}\cdot(X_u\times X_v).
\end{aligned}$$

In the last line, we have used the identity $-X_v\cdot(X_u\times\dot{X})=\mathrm{vol}(X_u,X_v,\dot{X})=-X_u\cdot(\dot{X}\times X_v)$ where $\mathrm{vol}(X_u,X_v,\dot{X})=\dot{X}\cdot(X_u\times X_v)$ is the volume of the parallelepiped spanned by $\dot{X}$, X_u and X_v (up to a sign). Substituting this into the integral under consideration, we find

$$\frac{d}{dt}\int_U X\cdot(X_u\times X_v)=3\int_U\dot{X}\cdot(X_u\times X_v)+\int_U[X\cdot(\dot{X}\times X_v)]_u+[X\cdot(X_u\times\dot{X})]_v.$$

We wish to express the last integrand as the local expression of the divergence of a globally defined vector field on Λ. Substituting the expression

$$\dot{X}=\phi N+\psi X_u+\eta X_v=\phi N+\dot{X}^{\mathrm{T}}$$

for $\dot{X}$, we find that the integrand can be written as

$$[X\cdot(\dot{X}\times X_v)]_u+[X\cdot(X_u\times\dot{X})]_v=\left[X\cdot\frac{\phi(X_u\times X_v)\times X_v}{\sqrt{EG-F^2}}\right]_u+\left[\sqrt{EG-F^2}\,\psi X\cdot N\right]_u$$
$$+\left[X\cdot\frac{\phi X_u\times(X_u\times X_v)}{\sqrt{EG-F^2}}\right]_v+\left[\sqrt{EG-F^2}\,\eta X\cdot N\right]_v.$$

The second and fourth terms sum to precisely

$$\sqrt{EG-F^2}\,\mathrm{div}_\Lambda[(X\cdot N)\dot{X}^{\mathrm{T}}].$$

Applying the triple cross product rule, we find

$$(X_u\times X_v)\times X_v=-GX_u+FX_v \quad\text{and}\quad X_u\times(X_u\times X_v)=FX_u-EX_v.$$

Thus, the first and third terms sum to

$$\left[\sqrt{EG-F^2}\,\phi\frac{-GX\cdot X_u+FX\cdot X_v}{EG-F^2}\right]_u$$
$$+\left[\sqrt{EG-F^2}\,\phi\frac{-FX\cdot X_u-EX\cdot X_v}{EG-F^2}\right]_v.$$

Comparing this with the expression of a general vector field in local coordinates, we see that this is

$$-\sqrt{EG-F^2}\,\mathrm{div}_\Lambda[\phi X^{\mathrm{T}}].$$

Going back to the original integral, we have

$$\frac{d}{dt}\int_U X\cdot(X_u\times X_v)=3\int_{X(U)}\dot{X}\cdot N+\int_{X(U)}\mathrm{div}_\Lambda[\phi\dot{X}^{\mathrm{T}}-(X\cdot N)X^{\mathrm{T}}].$$

Properly speaking, the expressions X and $\dot{X}$ in the integrals on the right side of this equation represent respectively $X(X^{-1}(x;t);t)$ and $\dot{X}(X^{-1}(x;t);t)$ for $x\in X(U)\subset\Lambda$. The function $v(x;t):=\dot{X}(X^{-1}(\,\cdot\,;t);t)$ will play an important role in the case of bulk accelerations below.

Finally summing over all neighborhoods U and recalling that $\phi=\dot{X}\cdot N$, we have

$$\begin{aligned}\frac{1}{3}\frac{d}{dt}\int_\Lambda X\cdot N&=\int_\Lambda\dot{X}\cdot N+\frac{1}{3}\int_\Lambda\mathrm{div}_\Lambda[(X\cdot N)\dot{X}^{\mathrm{T}}-(\dot{X}\cdot N)X^{\mathrm{T}}]\\&=\int_\Lambda\dot{X}\cdot N+\frac{1}{3}\int_{\partial\Lambda}[(X\cdot N)\dot{X}^{\mathrm{T}}-(\dot{X}\cdot N)X^{\mathrm{T}}]\cdot\vec{n}\\&=\int_\Lambda\dot{X}\cdot N+\frac{1}{3}\int_{\partial\Lambda}[(X\cdot N)\dot{X}-(\dot{X}\cdot N)X]\cdot\vec{n}.\end{aligned}$$

Again, we have used the divergence theorem. The last line results from the fact that the vectors $(X\cdot N)\dot{X}$ and $(\dot{X}\cdot N)X$ have the same component in the direction N.

The same calculation applies to the integral over $\mathcal{W}$ appearing in the expression for V so that

$$\dot{V} = \int_\Lambda \dot{X} \cdot N + \int_{\mathcal{W}} \dot{X}^{\mathcal{W}} \cdot N^{\mathcal{W}} + J,$$

where

$$3J = \int_{\partial\Lambda} [(X \cdot N)\dot{X} - (\dot{X} \cdot N)X] \cdot \vec{n} + [(X \cdot N^{\mathcal{W}})\dot{X} - (\dot{X} \cdot N^{\mathcal{W}})X] \cdot \vec{\nu}.$$

It would be usual at this point to assume $\dot{X} \cdot N^{\mathcal{W}} = 0$ along $\partial\Lambda$ so that one of the boundary integral terms vanishes. In keeping with the use of quite general vector-valued variations, we will avoid such an assumption. We will substitute rather

$$\dot{X} = \phi N + \delta\vec{n} + \eta\vec{t}, \quad \vec{\nu} = (\vec{\nu} \cdot N)N + (\vec{\nu} \cdot \vec{n})\vec{n}, \quad N^{\mathcal{W}} = -(\vec{\nu} \cdot \vec{n})N + (\vec{\nu} \cdot N)\vec{n},$$

where $\vec{t}$ is a vector tangent to $\partial\Lambda$. In this way, we find

$$\begin{aligned}
&[(X \cdot N)\dot{X} - (\dot{X} \cdot N)X] \cdot \vec{n} + [(X \cdot N^{\mathcal{W}})\dot{X} - (\dot{X} \cdot N^{\mathcal{W}})X] \cdot \vec{\nu} \\
&\qquad = \phi[-X \cdot \vec{n} + (X \cdot N^{\mathcal{W}})(\vec{\nu} \cdot N) + (\vec{\nu} \cdot \vec{n})(X \cdot \vec{\nu})] \\
&\qquad\qquad\qquad\qquad + \delta[X \cdot N + (X \cdot N^{\mathcal{W}})(\vec{\nu} \cdot \vec{n}) - (\vec{\nu} \cdot N)(X \cdot \vec{\nu})] \\
&\qquad = \phi[-(X \cdot \vec{n})(1 - (\vec{\nu} \cdot \vec{n})^2 - (\vec{\nu} \cdot N)^2)] + \delta[(X \cdot N)(1 - (\vec{\nu} \cdot \vec{n})^2 - (\vec{\nu} \cdot N)^2)] \\
&\qquad = 0.
\end{aligned}$$

Thus, we have a general variational formula for volume with respect to an arbitrary admissible vector-valued variation:

$$\dot{V} = \int_\Lambda \dot{X} \cdot N + \int_{\mathcal{W}} \dot{X}^{\mathcal{W}} \cdot N^{\mathcal{W}}. \tag{8}$$

In the situation presently under consideration, we also have that $\dot{X}^{\mathcal{W}}$ is in the tangent space of Σ, so that the second integral also vanishes, and we obtain the familiar

$$\dot{V} = \int_\Lambda \dot{X} \cdot N. \tag{9}$$

If the additional term $\mathcal{G} = \int_{\mathcal{M}} g$ is included in the energy, we require in addition that the third restriction of our variation $X\big|_{\mathcal{M}_0 \times \{t\}} : \mathcal{M}_0 \to \mathbb{R}^3$ parameterizes its image which is required to be the admissible liquid mass $\mathcal{M} = \mathcal{M}_t$ bounded by Λ and $\mathcal{W}$. In this more complicated situation, we first write

$$\mathcal{G} = \int_{\mathcal{M}_0} g \circ X \det DX,$$

where we have omitted the absolute values on the Jacobian factor for small t under the assumption of smoothness in t on the third restriction and the observation that $\det DX(\,\cdot\,; 0) = 1 > 0$.

This puts us in a position to apply Euler's first kinematical theorem — identity (4.1) in [Serrin 1959] — to obtain

$$\dot{\mathscr{G}} = \int_{\mathcal{M}} [Dg \cdot v + g \ \mathrm{div}_{\mathbb{R}^3} v] = \int_{\mathcal{M}} \mathrm{div}_{\mathbb{R}^3}[gv] = \int_{\Lambda} g \, \dot{X} \cdot N + \int_{\mathcal{W}} g \, \dot{X}^{\mathcal{W}} \cdot N^{\mathcal{W}}$$

by the divergence theorem, where $v(x;t) = \dot{X}(X^{-1}(x;t);t)$ is the *spatial velocity* of the flow induced by the variation. (In the last expression, we have returned to the use of $\dot{X}$ to represent $v(\cdot\,;t)$.) For a detailed derivation of the kinematical identity in the special case under consideration, see the appendix to [McCuan 2006].

If Σ is a fixed rigid surface, $\dot{X}^{\mathcal{W}} \cdot N^{\mathcal{W}} = 0$ as before, and we obtain, in complete generality, the formula

$$\dot{\mathscr{G}} = \int_{\Lambda} g \, \dot{X} \cdot N$$

which was obtained in special cases in [Wente 1966] and [Finn 1986] but not considered in [Ros and Souam 1997] or [Spivak 1979].

Note: If one is willing to admit the full variation of the interior liquid mass, this approach gives also, in the special case $g \equiv 1$, formulae (8) and (9).

The next step in developing the variational theory of these surfaces would be to introduce a Lagrange parameter corresponding to the volume constraint of the problem. We will have the opportunity to consider this topic under more general circumstances below, and hence move directly to those considerations.

3. Movable boundaries

We now wish to allow portions of the rigid bounding surface Σ to assume the character of free floating or partially free floating rigid obstacles. To this end, we assume Σ_0 (and correspondingly the relevant abstract manifold $M^{\mathcal{W}} = \mathcal{W}_0$) to be composed of disjoint components Σ_s and Σ_m. The former portions remain fixed, or stationary, as in the discussion above, while the latter are movable in the sense we now describe.

Let $w = w(x;h)$ be a family of rigid motions of $\mathbb{R}^3$ indexed by and smoothly depending on the parameter h which for simplicity we will assume is real-valued. Correspondingly, we will generalize our notion of variation to that of smooth maps $X : \overline{\mathcal{M}}_0 \times (-\delta,\delta) \times (-\epsilon,\epsilon) \to \mathbb{R}^3$, which for each fixed $h \in (-\delta,\delta)$ satisfy the conditions set forth in the previous section, with one slight modification. The new parameter h lying between $-\delta$ and δ will index the position of the movable portion of Σ, so that for fixed h the image of $X_m^{\mathcal{W}}$ is required to lie in $w(\Sigma_m;h)$. Otherwise, the parameter t on $(-\epsilon,\epsilon)$ is the variation parameter as before. When convenient, we will continue to refer to the union $M_s^{\mathcal{W}} \cup M_m^{\mathcal{W}} = \mathcal{W}_{0s} \cup \mathcal{W}_{0m}$ simply as $M^{\mathcal{W}}$. As we have denoted derivatives with respect to t with the customary dot above the letter, we will denote derivatives with respect to h with an acute accent: ´ . For each

h and t, the energy of and volume associated with the configuration are given by the same expressions as before.

We will assume that $\mathscr{E}(X_0, X_0^{\mathscr{W}}) \leq \mathscr{E}(X, X^{\mathscr{W}})$ for all C^1 admissible variations $(X, X^{\mathscr{W}})$ in a C^0 neighborhood of $(X_0, X_0^{\mathscr{W}})$ whose associated volume is also constrained to match that of the original minimizing configuration.

We give now a brief justification for the existence of a Lagrange multiplier in this problem. The argument is slightly more general than that seen for integral operators of the form (1) because we avoid the use of "additive variations," i.e., the usual $u + \epsilon\phi$. This is essentially an avoidance of assuming a linear structure in the class of admissible variations. The argument is also somewhat different than that usually given for Fréchet differentiable functionals as in [Gelfand and Fomin 1963] or for Gâteaux differentiable functionals as in [Sagan 1969] or [Ekeland and Temam 1976] for essentially the same reason.

We begin with a standard single parameter variation $Y : \overline{\mathscr{M}}_0 \times (-\epsilon, \epsilon) \to \mathbb{R}^3$ of $\mathscr{M}_0$ for which the volume has nonvanishing first variation. Examination of the variational formula for volume indicates that this is always possible; we can arrange for example that the first restriction (to Λ_0) satisfies $\dot{Y} \cdot N \geq 0$ with strict inequality on some open subset of Λ_0. We denote the parameter for this variation by τ so that $Y = Y(p; \tau)$. Corresponding to each fixed τ, we consider a further variation $X = X(p; h; t; \tau)$ with $X(p; 0; 0; \tau) = Y(p; \tau)$. This construction results in two smooth real-valued functions of three variables,

$$f(h, t, \tau) = \mathscr{E}(X, X^{\mathscr{W}}) \quad \text{and} \quad g(h, t, \tau) = V(X, X^{\mathscr{W}}),$$

having the property that $f(0, 0, 0)$ is a minimum for $f(h, t, \tau)$ subject to the constraint $g(h, t, \tau) = V(X_0, X_0^{\mathscr{W}})$. The usual principle of Lagrange applied to these functions asserts the gradients of f and g must be parallel at the origin; that is,

$$Df(0, 0, 0) = \lambda Dg(0, 0, 0)$$

for some constant λ. This results in three equations, the first two of which are

$$\acute{\mathscr{E}} - \lambda \acute{V} = 0 \quad \text{and} \quad \dot{\mathscr{E}} - \lambda \dot{V} = 0, \tag{10}$$

where everything is evaluated at $h = t = 0$. We have computed expressions (7) and (9) for $\dot{\mathscr{E}}$ and $\dot{V}$ above, so that the second equation becomes

$$\dot{\mathscr{E}} - \lambda \dot{V} = -\int_{\Lambda_0} (2H + \lambda)\dot{X} \cdot N + \int_{\partial\Lambda_0} (\cos\gamma - \beta)\dot{X} \cdot \vec{n} = 0.$$

By taking first interior variations (for which $\dot{X} \equiv 0$ on $\partial\Lambda_0$) we find

$$2H = -\lambda = \text{ constant} \quad \text{on } \Lambda_0.$$

Thus, the first integral vanishes, and we can take $\dot{X} \cdot \vec{n}$ arbitrary on $\partial\Lambda_0$, so that

$$\cos\gamma = \beta = \text{ constant} \quad \text{on } \partial\Lambda_0.$$

These are the usual first order necessary conditions for a capillary surface in zero gravity. The inclusion of $\mathcal{G}$ evidently leads also to the usual prescribed mean curvature equation $2H = g - \lambda$ where g is a volumetric potential energy density appropriate to any accelerations present.

We now turn to the first equation in (10). Applying the reasoning that led to $\dot{\mathcal{E}}$ but differentiating instead with respect to h, we note that there is no reason to believe $\acute{X}^{\mathcal{W}}$ lies in the tangent space to Σ. Thus, we obtain a variational expression with all terms present:

$$\begin{aligned}\acute{\mathcal{E}} - \lambda\acute{V} = -\int_\Lambda 2H\acute{X}\cdot N + \int_{\partial\Lambda} \acute{X}\cdot\vec{n} + \beta\int_{\mathcal{W}} 2H^{\mathcal{W}}\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}} - \beta\int_{\partial\Lambda}\acute{X}\cdot\vec{\nu}\\ -\lambda\int_\Lambda \acute{X}\cdot N - \lambda\int_{\mathcal{W}}\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}},\end{aligned}$$

where $H^{\mathcal{W}}$ is the mean curvature on the wetted region $\mathcal{W}$. We note, that we could, of course eliminate some of the integrals over portions $\mathcal{W}_s$ of $\mathcal{W}$ which are stationary, but we will save on subscripts by including them for the moment. We can use the prescribed mean curvature equation to cancel the integrals over Λ, and we may also substitute the contact angle boundary condition so that for $\Lambda = \Lambda_0$ and $\Sigma = \Sigma_0$ we have

$$\int_{\partial\Lambda}\acute{X}\cdot(\vec{n} - \cos\gamma\vec{\nu}) + \cos\gamma\int_{\mathcal{W}} 2H^{\mathcal{W}}\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}} + \int_{\mathcal{W}} 2H\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}} = 0.$$

We have moved the constant $2H$ inside the integral; it may be easily checked that this same formula holds in the case of bulk accelerations where $2H = g - \lambda$ may be nonconstant. Noting that $\vec{n} - \cos\gamma\vec{\nu} = \vec{n} - (\vec{n}\cdot\vec{\nu})\vec{\nu}$ and that $N^{\mathcal{W}}$ and $\vec{\nu}$ form an orthonormal basis, we see that $\vec{n} - \cos\gamma\vec{\nu} = \sin\gamma N^{\mathcal{W}}$. Thus the condition becomes

$$\int_{\partial\Lambda}\sin\gamma\acute{X}\cdot N^{\mathcal{W}} + \cos\gamma\int_{\mathcal{W}} 2H^{\mathcal{W}}\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}} + \int_{\mathcal{W}} 2H\acute{X}^{\mathcal{W}}\cdot N^{\mathcal{W}} = 0. \tag{11}$$

This is a first version of our new first order condition which applies to situations in which the solid support structures are deformable. We have not used, however, any special properties of the deformation $X^{\mathcal{W}}$ and of the smooth family of rigid motions $w : \mathbb{R}^3 \times (-\delta, \delta) \to \mathbb{R}^3$. Using these properties should allow us to express the formula in a manner that is essentially independent of $\acute{X}^{\mathcal{W}}$ in the interior of $\mathcal{W}$. The crucial observation is that because

$$\Sigma_{m;h} = \Sigma_h = w(\Sigma_m; h)$$

we have

$$w^{-1}(X^{\mathcal{W}}) \in \Sigma_0 \tag{12}$$

where we have suppressed the subscript "m," and $X^{\mathcal{W}}$ denotes $X_m^{\mathcal{W}}(x; h, t); h$. (We note that in the following discussion, we will also consistently suppress t dependence.)

The inclusion (12) implies

$$Dw^{-1}(X^{\mathcal{W}})\acute{X}^{\mathcal{W}} + \acute{w}^{-1}(X^{\mathcal{W}}) \in T_{w^{-1}(X^{\mathcal{W}})}\Sigma_0.$$

Recall, however, that for fixed h, the transformations w and w^{-1} are rigid motions. Therefore, we know that, for example, $w - w(0)$ is linear and has a corresponding matrix Dw which depends only on h. Since multiplication by this matrix also corresponds to the differential mapping $dw : T_{w^{-1}(X^{\mathcal{W}})}\Sigma_0 \to T_{X^{\mathcal{W}}}\Sigma$ of tangent spaces where $\Sigma = \Sigma_h$, we find that

$$\acute{X}^{\mathcal{W}} + Dw\acute{w}^{-1}(X^{\mathcal{W}}) \in T_{X^{\mathcal{W}}}\Sigma. \tag{13}$$

Differentiating the defining relation $w(w^{-1}(x; h); h) = x$ with respect to h, we find

$$Dw(w^{-1}(x; h); h)\acute{w}^{-1}(x; h) = -\acute{w}(w^{-1}(x; h); h).$$

Therefore, (13) can also be written as

$$\acute{X}^{\mathcal{W}} - \acute{w}(w^{-1}(X^{\mathcal{W}})) \in T_{X^{\mathcal{W}}}\Sigma.$$

From this it follows that each of the terms $\acute{X}^{\mathcal{W}} \cdot N^{\mathcal{W}}$ appearing in (11) may be replaced with $\acute{w} \cdot N^{\mathcal{W}} = \acute{w}(\,\cdot\,; 0) \cdot N^{\mathcal{W}}$.

$$\sin\gamma \int_{\partial\Lambda} \acute{w} \cdot N^{\mathcal{W}} + \cos\gamma \int_{\mathcal{W}} 2H^{\mathcal{W}}\acute{w} \cdot N^{\mathcal{W}} + \int_{\mathcal{W}} 2H\acute{w} \cdot N^{\mathcal{W}} = 0. \tag{14}$$

(We are still suppressing the subscript "m" on all the domains of integration; technically, $\partial\Lambda = \partial\mathcal{W}_s \cup \partial\mathcal{W}_m$ and the stationary portion of the first integral vanishes because $\acute{X}^{\mathcal{W}} \cdot N^{\mathcal{W}} \equiv 0$ there.)

Next, we temporarily fix h_0 and consider a specific motion of the wetted region $\mathcal{W}_{h_0}$ as follows. Let $\tilde{\mathcal{W}} = \tilde{\mathcal{W}}_h$ be parameterized on M_m by

$$p \mapsto w(w^{-1}(X^{\mathcal{W}}(p; h_0); h_0); h).$$

Notice that for $h = h_0$ the surface $\tilde{\mathcal{W}}$ is exactly $\mathcal{W}_{h_0}$ and for $h > h_0$ the surface $\tilde{\mathcal{W}}$ is just a rigid motion of the same wetted region and, therefore, has the same area. Thus, using the general formula for the variation of surface area in this special case, we have

$$0 = \frac{d}{dh}A(\tilde{\mathcal{W}}) = -\int_{\tilde{\mathcal{W}}} 2\tilde{H}^{\mathcal{W}}\acute{w} \cdot \tilde{N}^{\mathcal{W}} + \int_{\partial\tilde{\mathcal{W}}} \acute{w} \cdot \tilde{\nu},$$

where $\tilde{H}^{\mathcal{W}}$ is the mean curvature on $\tilde{\mathcal{W}}$, $\tilde{N}^{\mathcal{W}}$ is the normal on $\tilde{\mathcal{W}}$, $\tilde{\nu}$ is the conormal on $\partial\tilde{\mathcal{W}}$, and the vectors $\acute{w}$ appearing in the integrals are evaluated at

$$(w^{-1}(w(w^{-1}(x; h_0); h); h_0); h).$$

We now change variables in the integrals from the previous page using the map $x \mapsto w(w^{-1}(x; h); h_0)$ which maps $\tilde{\mathcal{W}}_h$ back to $\tilde{\mathcal{W}}_{h_0} = \mathcal{W}_{h_0}$. We then have

$$-\int_{\mathcal{W}_{h_0}} 2H^{\mathcal{W}} Dw(w^{-1}(\,\cdot\,; h); h_0)\, Dw^{-1}(\,\cdot\,; h_0)\, \acute{w}(w^{-1}(\,\cdot\,; h_0); h) \cdot N^{\mathcal{W}} \\ + \int_{\partial\mathcal{W}_{h_0}} Dw(w^{-1}(\,\cdot\,; h); h_0)\, Dw^{-1}(\,\cdot\,; h_0)\, \acute{w}(w^{-1}(\,\cdot\,; h_0); h) \cdot \vec{\nu} = 0.$$

Setting $h = h_0$ and suppressing all the h's again so that $\acute{w}(w^{-1}) = \acute{w}(w^{-1}(\,\cdot\,; h); h)$, we get

$$-\int_{\mathcal{W}} 2H^{\mathcal{W}} \acute{w}(w^{-1}) \cdot N^{\mathcal{W}} + \int_{\partial\mathcal{W}} \acute{w}(w^{-1}) \cdot \vec{\nu} = 0.$$

Using this identity to replace the middle integral on the left in (14), we find at $h = 0$ the condition

$$\int_{\partial\mathcal{W}} \acute{w} \cdot (\sin\gamma\, N^{\mathcal{W}} + \cos\gamma\, \vec{\nu}) + \int_{\mathcal{W}} 2H\acute{w} \cdot N^{\mathcal{W}}.$$

Finally, we note that $\sin\gamma\, N^{\mathcal{W}} + \cos\gamma\, \vec{\nu} = \vec{n}$ and modify the notation to emphasize once again the significance of the movable portion Σ_m of the solid support surface to obtain the fundamental formula.

Theorem 3.1. *Assume X_0 parameterizes a capillary surface Λ_0 that meets solid support surfaces at a constant contact angle γ and defines a liquid mass $\mathcal{M}$ as described above. If moreover, Λ_0 provides a minimum among variations compatible with the motion of a portion Σ_m of the bounding surfaces allowed by the rigid motions $w = w(x; h)$, then*

$$\int_{\mathcal{W}_m} 2H\acute{w} \cdot N^{\mathcal{W}} + \int_{\partial\mathcal{W}_m} \acute{w} \cdot \vec{n} = 0,$$

where $\acute{w} = \partial w/\partial h(\,\cdot\,; 0)$. It should be noted in this formula that $\vec{n}$ is the conormal to $\partial\Lambda \supset \partial\mathcal{W}_m$ and H is the mean curvature of Λ. The formula holds both when H is constant and when $2H = g - \lambda$ is prescribed by bulk accelerations.

We now apply this result to a particular physical situation.

4. The floating ball

Finn [2005] considers the configuration shown in Figure 3 in which a planar interface $z \equiv 0$ for $\alpha = a\sin\gamma \le r \le R$ meets a cylindrical container of radius R at an angle $\pi/2$ and a sphere held rigidly on the axis of the container at an angle of

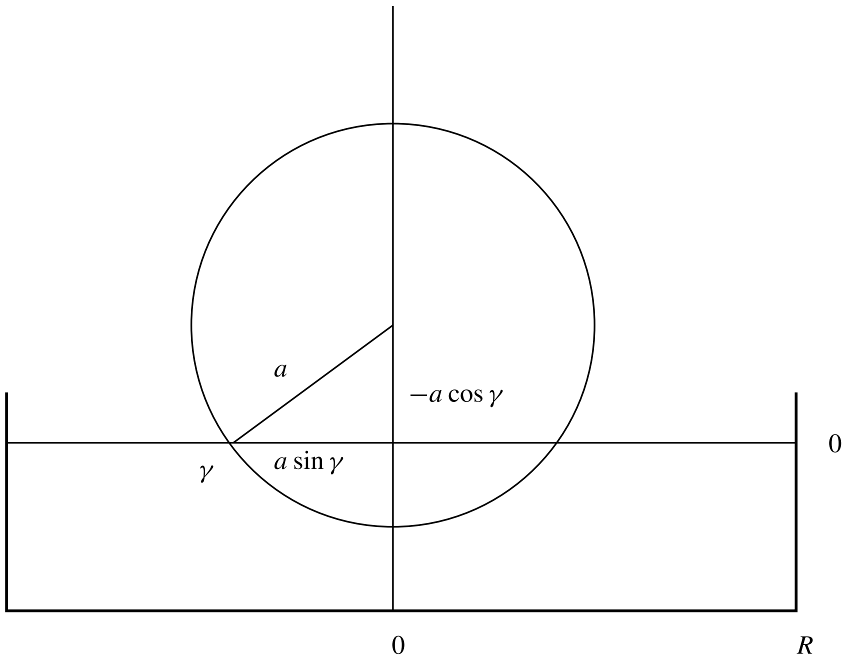


Figure 3. A planar interface determining a liquid mass of given volume and prescribed contact angle.

contact γ. If the ball remains rigidly fixed with center at height $-a\cos\gamma$, then we may apply the stability theory of Vogel [2000] to obtain the following result whose proof we give at the end of the section.

Theorem 4.1. *For the fixed ball problem, the planar interface is linearly stable and is a strong local minimum in the sense of Vogel.*

We conclude from this, that such an initial configuration is almost certainly experimentally reproducible, though with some external support to hold the ball in place. One would like to know if the ball will remain in this position and the planar interface will persist if the external support were to be removed. Will the ball float in this position as suggested by Finn?

A theory allowing such a decisive conclusion is apparently unavailable in the situation of moving boundaries. Our calculation leading to Theorem 3.1 however allows us to make an argument that stability with respect to vertical motion of the ball in this problem is also likely.

Finn shows in his paper that if the ball is moved (or suggestively pushed) downward by a small distance h, then there is a unique rotationally symmetric constant mean curvature surface that meets the ball still at the angle γ, the container still at angle $\pi/2$ and determines a liquid mass of the same volume. The same assertions

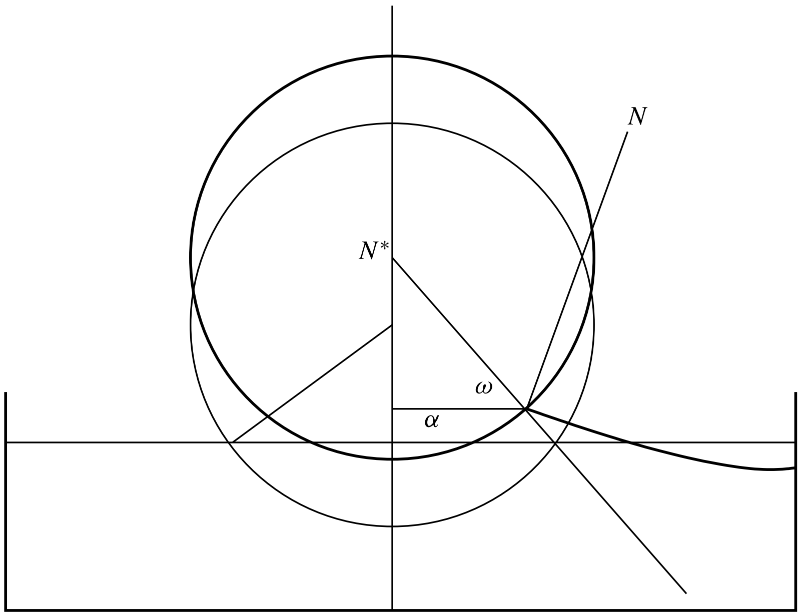


Figure 4. Nodoid interface.

are true if one raises the ball a small distance h (see Figure 4), and the surfaces are, in each case, nodoids whose meridians are obtained as the path of the focal point of a hyperbola as it rolls along the vertical axis through the center of the ball. If for each $h \in (-\delta, \delta)$, these auxiliary configurations have free surface interface Λ_h and wetted area $\mathscr{W}_h$, where the center of the ball is located at $h - a\cos\gamma$, then presumably the overall stability of the configuration (at least with respect to vertical motion of the ball) should be characterized by the values

$$\mathscr{E}(h) = A(\Lambda_h) - \cos\gamma\, A(\mathscr{W}_h).$$

We wish to compute, in particular, $\acute{\mathscr{E}} = d\mathscr{E}/dh$; the volume term can be ignored, since all the volumes are the same. The quantity $\mathscr{E}(h)$ can be written explicitly in terms of integrals as

$$\mathscr{E} = 2\pi \int_\alpha^R r\sqrt{1+u'(r)^2}\, dr - \cos\gamma\, 2\pi \int_0^\alpha \frac{ar}{\sqrt{a^2-r^2}}\, dr$$

where u and α are determined as follows. The function u solves the ordinary differential equation of constant mean curvature for rotationally symmetric surfaces.

This equation is

$$\frac{1}{r}\left(\frac{ru'}{\sqrt{1+u'^2}}\right)' = 2H_h. \tag{15}$$

Denoting by $\alpha = \alpha(h)$ the radius at which the nodoid surface meets the ball, we can integrate to find

$$u(r) = u(\alpha) + \int_\alpha^r \frac{\alpha q + H_h(t^2-\alpha^2)}{|H_h|\sqrt{(\rho_{\text{out}}^2 - t^2)(t^2-\rho_{\text{in}}^2)}}\, dt, \tag{16}$$

where $u(\alpha) = h - a\cos\gamma - \sqrt{a^2-\alpha^2}$, and

$$q = q(h) = \frac{u'(\alpha)}{\sqrt{1+u'(\alpha)^2}} = -\left(\frac{\alpha}{a}\cos\gamma + \sqrt{1-\left(\frac{\alpha}{a}\right)^2}\sin\gamma\right) = -\cos(\gamma-\omega)$$

where ω is the smaller angle made by a segment from the center of the ball to the contact line with the horizontal (see Figure 4),

$$H_h = -\frac{\alpha q}{R^2-\alpha^2} = \frac{\alpha\cos(\gamma-\omega)}{R^2-\alpha^2}$$

which follows from the fact that $u'(R) = 0$ and the first integral of equation (15), and $0 < \rho_{\text{in}} < \rho_{\text{out}}$ are the positive roots of $r^2 - [\alpha q + H_h(r^2-\alpha^2)]^2 = 0$. Finally, it remains to describe $\alpha = \alpha(h)$ which is determined implicitly by the equation

$$\begin{aligned} V &= \pi R^2 l + 2\pi\int_0^\alpha r(h - a\cos\gamma - \sqrt{a^2-r^2})\, dr + 2\pi\int_\alpha^R ru(r)\, dr \\ &= \pi R^2(l + h - a\cos\gamma) - 2\pi\int_0^\alpha r\sqrt{a^2-r^2}\, dr \\ &\qquad -2\pi\sqrt{a^2-\alpha^2}\int_\alpha^R r\,dr + 2\pi\int_\alpha^R r\int_\alpha^r u'(t)\, dt\, dr, \end{aligned} \tag{17}$$

where l denotes the initial depth of the liquid near the wall of the cylindrical container. In order to show that $\alpha = \alpha(h)$ is well defined for h in some neighborhood of zero, we think of the expression above as defining $V = V(\alpha, h)$ and show that

$$\frac{\partial V}{\partial\alpha}(a\sin\gamma, 0) \neq 0.$$

This calculation is contained in the proof of Lemma 4.3 below.

The expression for u may also be expressed in terms of standard elliptic integrals, as indicated in [Finn 2005]. Then it will be noted that the volume above as well as $\mathscr{E}$ will depend on integrals of complicated expressions involving elliptic integrals. This should suffice to indicate the difficulty in calculating $\acute{\mathscr{E}}$ directly. Since we know that $\acute{V} = 0$ for this variation, however, we have direct recourse to the formula of Theorem 3.1 which is completely independent of the value of $u = u(r)$ for $\alpha < r < R$.

Lemma 4.2. *For the floating ball problem described above*

$$\acute{\mathscr{E}} = \frac{d\mathscr{E}}{dh} = \frac{2\pi\alpha R^2}{R^2-\alpha^2}\cos(\gamma-\omega).$$

Proof. In this case, the rigid motion is simply translation in the vertical direction $w(x; h) = x + h\mathbf{e}_3$ and $\acute{w} = \mathbf{e}_3$. The wetted region (on the ball) $\mathscr{W}$ is the portion of lower hemisphere $\{(x, y, h-a\cos\gamma-\sqrt{a^2-r^2}) : 0 \le r^2 = x^2+y^2 \le \alpha^2\}$, and can be parameterized in spherical coordinates (θ, ϕ) for $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi/2-\omega$. Thus, from the formula in Theorem 3.1,

$$\begin{aligned}\acute{\mathscr{E}} &= 2\pi\alpha\mathbf{e}_3\cdot(\sin\gamma N^{\mathscr{W}} + \cos\gamma\vec{\nu}) + 2H_h 2\pi\int_0^{\pi/2-\omega}(\mathbf{e}_3\cdot N^{\mathscr{W}})a^2\sin\phi\,d\phi\\ &= 2\pi\alpha(\sin\gamma\sin\omega+\cos\gamma\cos\omega) + \frac{2\alpha\cos(\gamma-\omega)}{R^2-\alpha^2}\,2\pi a^2\int_0^{\pi/2-\omega}\cos\phi\sin\phi\,d\phi\\ &= 2\pi\alpha\cos(\gamma-\omega)\Big(1+\frac{a^2}{R^2-\alpha^2}\sin^2(\pi/2-\omega)\Big).\end{aligned}$$

Since $a^2\sin^2(\pi/2-\omega) = \alpha^2$, this is the formula given above. $\square$

Notice that α and $R^2-\alpha^2$ are both positive. Thus, the sign of $\acute{\mathscr{E}}$ in this case is determined by the factor $\cos(\gamma-\omega)$. We know that when $h = 0$ we have $\omega = \gamma-\pi/2$ so that $\acute{E}(0) = 0$. In order to determine the value of $\acute{\mathscr{E}}$ for other values of h we must examine more closely the volume equation for $\alpha = a\cos\omega$.

Lemma 4.3. *The function* $\alpha = \alpha(h)$ *defined in a neighborhood of* $h = 0$ *satisfies* $\alpha'(0) < 0$. *Consequently, given h in some neighborhood of zero, since* $\cos\omega = \alpha/a$, *if* $h < 0$, *then* $\omega < \gamma-\pi/2$; *if* $h > 0$, *then* $\omega > \gamma-\pi/2$.

Proof. In this proof, we consider the expressions in the definition of the nodoid meridians (16) to be functions of both α and h as independent variables. From this point of view, we note that the expressions for q, H and u' have no explicit dependence on h. Furthermore, setting

$$v = v(\alpha; r) = \alpha q + H(r^2-\alpha^2),$$

we obtain an alternative expression for $u' = u'(r; \alpha)$:

$$u' = \frac{v}{\sqrt{r^2-v^2}}.$$

Differentiating the expression for the fixed volume (17) with respect to h, we find

$$0 = \pi R^2 + \left(\pi\Big(\frac{\alpha}{\sqrt{a^2-\alpha^2}} - u'(\alpha)\Big)(R^2-\alpha^2) + 2\pi\int_\alpha^R r\int_\alpha^r \frac{\partial u'}{\partial\alpha}(t)\,dt\,dr\right)\alpha'(h).$$

Thus, we wish to show that

$$\lim_{h\to 0}\left(\frac{\alpha(R^2-\alpha^2)}{\sqrt{a^2-\alpha^2}}-u'(\alpha)(R^2-\alpha^2)+2\int_\alpha^R r\int_\alpha^r \frac{\partial u'}{\partial\alpha}(t)\,dt\,dr\right)>0. \tag{18}$$

We consider each term individually; each term has a completely nonsingular limit. First,

$$\frac{\alpha(R^2-\alpha^2)}{\sqrt{a^2-\alpha^2}} \longrightarrow -\tan\gamma(R^2-a^2\sin^2\gamma)>0 \quad \text{as } h\to 0,$$

where we have used that $-\tan\gamma>0$. Since q and H tend to 0 with h, we see that v tends uniformly to zero for r in some neighborhood of $[a\sin\gamma,\, R]$. In particular, $\alpha u'(\alpha)$ tends to 0 with h, so the second term vanishes in the limit.

The last term requires some preliminary calculations and limits:

$$\frac{\partial u'}{\partial\alpha}=\frac{r^2}{(r^2-v^2)^{3/2}}\frac{\partial v}{\partial\alpha}=\frac{r^2}{(r^2-v^2)^{3/2}}\left(q+\alpha\frac{\partial q}{\partial\alpha}+\frac{\partial H}{\partial\alpha}(r^2-\alpha^2)-2\alpha H\right).$$

Inside the parentheses we have, as $h\to 0$,

$$\frac{\partial q}{\partial\alpha}=-\frac{1}{a}\cos\gamma+\frac{\alpha}{a\sqrt{a^2-\alpha^2}}\sin\gamma \longrightarrow -\frac{1}{a}(\cos\gamma+\tan\gamma\sin\gamma)=-\frac{1}{a\cos\gamma}$$

and

$$\frac{\partial H}{\partial\alpha}=-\frac{\partial}{\partial\alpha}\left(\frac{\alpha}{R^2-\alpha^2}\right)q-\left(\frac{\alpha}{R^2-\alpha^2}\right)\frac{\partial q}{\partial\alpha} \longrightarrow \frac{\tan\gamma}{R^2-a^2\sin^2\gamma}$$

Therefore,

$$\frac{\partial u'}{\partial\alpha} \longrightarrow \frac{1}{r}\left(-\tan\gamma+\tan\gamma\frac{r^2-a^2\sin^2\gamma}{R^2-a^2\sin^2\gamma}\right)=-\frac{\tan\gamma}{r(R^2-a^2\sin^2\gamma)}(R^2-r^2)$$

as $h\to 0$. Since $-\tan\gamma>0$, this shows the last term in (18) also has a positive limit. □

We recall that each of the nodoid interfaces is considered a critical interface for the fixed ball problem. That is, they satisfy the necessary conditions of having constant mean curvature and constant contact angle equal to γ on the contact line with the ball. With the addition of our third necessary condition in the floating ball problem, we can now state the following.

Theorem 4.4. *With respect to the variation of the planar interface through nodoid interfaces parameterized by the height of the ball h, the plane is the only equilibrium for h in some neighborhood of $h=0$ and provides a strict minimum of energy among those interfaces.*

We were unable to analyze precisely the nodoid interfaces (even concerning existence and connectedness) for h far from zero, but since the formula in Lemma 4.2 results only from boundary considerations and we can say that $\cos(\gamma-\omega)\neq 0$ for every nodoid interface, we can make the following assertion:

Theorem 4.5. *The unique rotationally symmetric equilibrium for the floating ball problem is the planar interface*; *each nodoid interface is unstable to first order with respect to motion of the ball.*

It is, of course, unlikely that any of the other nodoid interfaces would have lower energy than the plane. Presumably, an analytic argument could be given to cover the component of those interfaces connected via h to the neighborhood of the planar interface considered in Theorem 4.4, and the global picture could be understood numerically if not analytically. For now, we have illustrated the use of Theorem 3.1.

Lastly, we return to the fixed ball problem and show that the planar interface is a strong local minimum in the sense of Vogel.

Proof of Theorem 4.1. This configuration satisfies the conditions stated in [Vogel 2000]. Therefore, the stability for the interface is determined by the eigenvalue problem

$$\begin{cases} \Delta\phi+\lambda\phi=0, & r=\sqrt{x^2+y^2}\in(\alpha,\,R),\\ \left.\dfrac{\partial\phi}{\partial r}\right|_{r=\alpha}=\dfrac{1}{\alpha}\,\phi\big|_{r=\alpha}; & \left.\dfrac{\partial\phi}{\partial r}\right|_{r=R}\equiv 0. \end{cases}$$

Vogel shows that if all eigenvalues for this problem are positive, the interface is strictly stable and a strong local minimum with respect to a large class of perturbations.

In polar coordinates, the equation becomes

$$\frac{1}{r}\left(\frac{\partial}{\partial r}\Big(r\,\frac{\partial\phi}{\partial r}\Big)+\frac{\partial}{\partial\theta}\Big(\frac{1}{r}\frac{\partial\phi}{\partial\theta}\Big)\right)+\lambda\phi=0.$$

Setting

$$\phi=\sum_{k=1}^{\infty}g_k(r)\sin(k\theta)+\sum_{k=0}^{\infty}f_k(r)\cos(k\theta),$$

we are led to the Sturm–Liouville problems

$$\begin{cases} rf''+f'+\big(\lambda r-k^2/r\big)\,f=0, & \alpha<r<R,\\ f'(\alpha)=\dfrac{1}{\alpha}\,f(\alpha); \quad f'(R)=0, \end{cases}$$

where f denotes $f_k,\ k=0,1,\ldots$ or $g_k,\ k=1,2,\ldots.$

We wish to show that there are no nontrivial solutions for $\lambda \le 0$. For $\lambda < 0$ we set $\lambda = -\mu^2$ and change variables ($t = \mu r$) to obtain

$$\begin{cases} f'' + (1/t)f' - \left(1 + k^2/t^2\right) f = 0, & \alpha\mu < t < R\mu, \\ f'(\mu\alpha) = \dfrac{1}{\mu\alpha} f(\mu\alpha); \quad f'(\mu R) = 0. \end{cases}$$

This is a regular problem which we can transform using the change of variables $\dot{\tau} = \log t$. Then we get

$$f(t) = v(\log t); \qquad f' = \frac{v'}{t}, \qquad f'' = \frac{v''}{t^2} - \frac{v'}{t^2},$$

so that

$$\begin{cases} v'' - (e^{2\tau} + k^2)v = 0, & l_1 = \log(\mu\alpha) < \tau < \log(\mu R) = l_2, \\ v'(l_1) = v(l_1); \quad v'(l_2) = 0. \end{cases}$$

We claim that this problem has no nontrivial solution. In fact, if $v(l_1) < 0$, then $v'(l_1) = v(l_1) < 0$ and $v''(l_1) = (e^{2l_1} + k^2)v(l_1) < 0$. Since each of $v(\tau)$, $v'(\tau)$, and $v''(\tau)$ must remain negative, there can be no $l_2 > l_1$ for which $v'(l_2) = 0$. (Assume that one of v, v', v'' vanishes first at $\tau = \tau_1$. If $v(\tau_1) = 0$, there is some $\tau_2 < \tau_1$ for which $v'(\tau_2) > 0$, a contradiction. The same argument holds if $v'(\tau_1) = 0$. If $v''(\tau_1) = 0$, then

$$v(\tau_1) = \frac{1}{\left(e^{2l_1} + k^2\right)} v'(\tau_2) = 0,$$

also a contradiction.)

If $v(l_1) = 0$, then uniqueness of solutions gives $v \equiv 0$.

If $v(l_1) > 0$, then v, v', and v'' must remain forever positive, so we cannot have $v'(l_2) = 0$.

If $\lambda = 0$, the problem becomes

$$\begin{cases} r^2 f'' + r f' - k^2 f = 0, & \alpha < r < R, \\ f'(\alpha) = \dfrac{1}{\alpha} f(\alpha); \quad f'(R) = 0. \end{cases}$$

This is an Euler equation that transforms as above to

$$\begin{cases} v'' = k^2 v, & l_1 = \log\alpha < \tau < \log R = l_2, \\ v'(l_1) = v(l_1); \quad v'(l_2) = 0. \end{cases}$$

An argument similar to that in the case $\lambda < 0$, shows that only the trivial solution is possible. In fact, more explicitly we have $v = ce^{kt} + de^{-kt}$, so that

$$\begin{cases} ce^{kl_1} - de^{-kl_1} = ce^{kl_1} + de^{-kl_1}, \\ ce^{kl_2} - de^{-kl_2} = 0 \end{cases}$$

or

$$\begin{cases} (k-1)e^{kl_1}c - (k+1)e^{-kl_1}d = 0, \\ e^{kl_2}c - e^{-kl_2}d = 0, \end{cases}$$

which has only the trivial solution since

$$\begin{aligned} -(k-1)e^{k(l_1-l_2)} + (k+1)e^{k(l_2-l_1)} &= k(e^{k(l_2-l_1)} - e^{-k(l_2-l_1)}) + e^{k(l_2-l_1)} + e^{-k(l_2-l-1)} \\ &= 2k \sinh k(l_2-l_1) + 2\cosh k(l_2-l_1) \neq 0. \end{aligned}$$

(If this were to vanish, we would have

$$\frac{\sinh k(l_2 - l_1)}{\cosh k(l_2 - l_1)} = -\frac{1}{k} < 0, \qquad k \neq 0$$

or

$$\cosh k(l_2 - l_1) = 0, \qquad k = 0,$$

both of which are impossible.)

Thus, the interface under consideration is stable and also a strong local minimum by the work of Vogel. □

References

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

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

[Blaschke 1930] W. Blaschke, *Vorlesungen über Differentialgeometrie und geometrische Grundlagen von Einsteins Relativitätstheorie, I: Elementare Differentialgeometrie*, 3rd ed., Grundlehren der Math. Wiss. **1**, Springer, Berlin, 1930. Reprinted Dover, New York, 1945. JFM 56.0588.07

[Ekeland and Temam 1976] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Studies in Mathematics and its Applications **1**, North-Holland, Amsterdam, 1976. MR 57 #3931b Zbl 0322.90046

[Evans and Gariepy 1992] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 1992. MR 93f:28001 Zbl 0804.28001

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

[Finn 2005] R. Finn, “Floating and partly immersed balls in a weightless environment”, *Funct. Differ. Equ.* **12**:1-2 (2005), 167–173. MR 2005m:76072 Zbl 1085.76016

[Gelfand and Fomin 1963] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Prentice-Hall, Englewood Cliffs, NJ, 1963. Reprinted Dover, New York, 2000. MR 28 #3353 Zbl 0127.05402

[McCuan 2006] J. McCuan, "Variation of energy for capillary surfaces", Preprint, MPI-MIS, Leipzig, 2006, Available at http://www.mis.mpg.de/preprints/2006/prepr2006_18.html.

[Ros and Souam 1997] A. Ros and R. Souam, "On stability of capillary surfaces in a ball", *Pacific J. Math.* **178**:2 (1997), 345–361. MR 98c:58029 Zbl 0930.53007

[Sagan 1969] H. Sagan, *Introduction to the calculus of variations*, McGraw-Hill, New York, 1969. Reprinted Dover, New York, 1992. MR 94h:49001

[Serrin 1959] J. Serrin, "Mathematical principles of classical fluid mechanics", pp. 125–263 in *Handbuch der Physik,* **8/1***: Strömungsmechanik I*, edited by S. Flügge and C. Truesdell, Springer, Berlin, 1959. MR 21 #6836b

[Spivak 1979] M. Spivak, *A comprehensive introduction to differential geometry*, vol. 4, 2nd ed., Publish or Perish, Wilmington, DE, 1979. MR 82g:53003d Zbl 0439.53004

[Vogel 2000] T. I. Vogel, "Sufficient conditions for capillary surfaces to be energy minima", *Pacific J. Math.* **194**:2 (2000), 469–489. MR 2001j:58023 Zbl 1021.58014

[Wente 1966] H. C. Wente, *Existence theorems for surfaces of constant mean curvature and perturbations of a liquid globule in equilibrium*, Ph.D. thesis, Harvard University, Cambridge, MA, 1966.

Received January 11, 2006. Revised June 9, 2006.

JOHN MCCUAN
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
686 CHERRY STREET
ATLANTA, GA 30332
UNITED STATES

mccuan@math.gatech.edu
http://www.math.gatech.edu/~mccuan

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

HOPFISH ALGEBRAS

XIANG TANG, ALAN WEINSTEIN AND CHENCHANG ZHU

We introduce a notion of "hopfish algebra" structure on an associative algebra, allowing the structure morphisms (coproduct, counit, antipode) to be bimodules rather than algebra homomorphisms. We prove that quasi-Hopf algebras are hopfish algebras. We find that a hopfish structure on the algebra of functions on a finite set G is closely related to a "hypergroupoid" structure on G. The Morita theory of hopfish algebras is also discussed.

1. Introduction

When the multiplication on a (discrete, topological, smooth, algebraic) group G is encoded in an appropriate algebra $A = A(G)$ of functions on G with values in a commutative ring k, it becomes a coproduct, that is, an algebra homomorphism $\Delta : A \to A \otimes_k A$. The inclusion of the unit and the inversion map are also encoded as homomorphisms: the counit $\epsilon : A \to k$ and the antipode $S : A \to A$. The group properties (associativity, unit, inverse) become statements about these homomorphisms which constitute the axioms for a (commutative) *Hopf algebra*; any noncommutativity of the underlying group appears as non<u>co</u>commutativity of the coproduct.

In noncommutative geometry, a noncommutative algebra A is thought of as the functions on a "noncommutative space" or "quantum space" X. If X is to be a "quantum group", the algebra A should have the additional structure of a Hopf algebra. We note that, for noncommutative Hopf algebras, the antipode has to be an antihomomorphism rather than a homomorphism of algebras. For this reason, a Hopf algebra is not quite a group in the category of algebras; this anomaly will come back to haunt us later.

One type of quantum space is a quantum torus, whose function algebra is the crossed product algebra A_α associated to an action of $\mathbb{Z}$ on the circle $S^1 = \mathbb{R}/\mathbb{Z}$ generated by an irrational rotation r_α. This irrational rotation algebra is generally taken as a surrogate for the algebra of continuous functions on the "bad quotient

MSC2000: primary 16W30; secondary 81R50.
Keywords: Hopf algebra, hopfish algebra, groupoid, bimodule, Morita equivalence, hypergroupoid.
Weinstein's research is partially supported by NSF Grant DMS-0204100. Tang and Zhu worked respectively at UC Davis and ETH Zürich during the period of this research.

space" $S^1/\alpha\mathbb{Z}$ because, for nice quotients, the crossed product algebra is Morita equivalent to the algebra of functions on the quotient. Since $S^1/\alpha\mathbb{Z}$ is a group, one might expect A_α to have a Hopf algebra structure, but this is not so. In particular, there can be no counit, since there are no algebra homomorphisms $A_\alpha \to \mathbb{C}$. In geometric language, "the quantum torus has no points".

Additionally, in noncommutative geometry, Morita equivalent algebras are often thought of as representing the "same space", but the notion of Hopf algebra, and even that of biunital bialgebra, is far from Morita invariant.

In this paper, we propose a new algebraic approach to "group structure" based on the idea that the appropriate morphisms between algebras are bimodules (perhaps with extra structure, or satisfying extra conditions) rather than algebra homomorphisms. Our immediate inspiration to use bimodules was [Tseng and Zhu 2006], in which leaf spaces of foliations are treated as differential stacks for the purpose of putting group(oid)-like structures on them. This means that the structure morphisms of the groupoids are themselves bibundles [Mrčun 1996] (with respect to foliation groupoids, which play in this geometric story the role of the crossed product algebras above) rather than ordinary mappings of leaf spaces. We were also motivated by previous uses of bimodules as generalized morphisms of algebras, C^*-algebras, groupoids, and Poisson manifolds, a point of view which has been extensively developed by Landsman and others (see, for instance, [Bursztyn and Weinstein 2005; Landsman 2001a; 2001b]).

We call our new objects *hopfish algebras*, the suffix "oid" and prefixes like "quasi" and "pseudo" having already been appropriated for other uses. Also, our term retains a hint of the Poisson geometry which inspired some of our work.

Outline of the paper. We begin with a discussion of the category in which objects are algebras and morphisms are bimodules, emphasizing the functor, which we call modulation, from the usual category to this one. We then look at the analogues of semigroups and groups in this category, which we call sesquialgebras and hopfish algebras. What turns out to be especially delicate is the definition of the antipode. We next show that Hopf algebras, and the more general quasi-Hopf algebras, become hopfish algebras upon modulation. In the following section, we study the Morita invariance of the hopfish property, showing that a sufficient condition for this to hold is that a Morita equivalence bimodule be compatible with the antipode of a hopfish algebra. Finally, we study hopfish structures on finite dimensional commutative algebras. We show that these correspond to "multiple-valued groupoid structures" and give examples of hopfish algebras which do not correspond under Morita equivalence to Hopf algebras.

Outlook. In this present paper, we restrict ourselves to the purely algebraic situation; in particular, our tensor products do not involve any completion. We do not require finite dimensionality of our algebras, although some of our examples do have this property. We hope to develop a theory of hopfish C^*-algebras in the future, with a treatment of irrational rotation algebras as a first goal. Even without this theory, it has been possible in [Blohmann et al. 2006] to construct a sesquiunital sesquialgebra structure on the "polynomial part" of the irrational rotation algebras. These algebras are not quite hopfish, since the candidate antiautomorphism satisfies only a weakened version of our antipode axiom. (We hope that this will be remedied when we go on to the C^*-algebras.) Nevertheless, our structure is sufficient to induce an interesting monoid structure on isomorphism classes of modules.

Finally, we remark that all of our examples of hopfish examples are either weak Hopf algebras or Morita equivalent to quasi-Hopf algebras. It would be interesting to find completely new examples. The irrational rotation algebras are probably not of either of these special types, but, as we have already noted, they are not quite hopfish.

2. The modulation functor

Fixing a commutative ring k as our ring of scalars, we will work mostly in a category Alg whose objects are unital k-algebras. The morphism space $\mathrm{Hom}(A, B)$ is taken to be the set of isomorphism classes of biunital (A, B)-bimodules. We will almost always consider these morphisms as going from right to left, i.e.- from B to A (or, better, "to A from B"). The composition $XY \in \mathrm{Hom}(A, C)$ of $X \in \mathrm{Hom}(A, B)$ and $Y \in \mathrm{Hom}(B, C)$ is defined (on representative bimodules) as $X \otimes_B Y$, with the residual actions of A and C providing the bimodule structure.

We will frequently fail to distinguish between morphisms in Alg and their representative bimodules, as long as we can do so without causing confusion. It is also possible to work in the more refined 2-category whose morphisms are bimodules and whose 2-morphisms are bimodule isomorphisms, but we leave this for the future.

We will denote by Alg_0 the "usual" category whose objects are again unital k algebras but whose morphisms are unital homomorphisms. Thus, $\mathrm{Hom}_0(A, B)$ will denote the homomorphisms to A from B. There is an important functor from Alg_0 to Alg which we will call *modulation*.[1] The modulation of $f \in \mathrm{Hom}_0(A, B)$ is the isomorphism class of A_f, which is the k-module A with the (A, B)-bimodule structure

$$a \cdot x \cdot b = axf(b). \tag{1}$$

[1] We are indebted to Yvette Kosmann-Schwarzbach for suggesting this apt name for a functor which is ubiquitous in the literature on Morita equivalence, but which does not seem to have acquired a standard designation.

We will often denote the modulation of a morphism by the same symbol, but in bold face, e.g. $\boldsymbol{f} \in \mathrm{Hom}(A, B)$. The modulation functor is not necessarily faithful, as the next lemma shows.

Lemma 2.1. *For $f, g \in \mathrm{Hom}_0(A, B)$, their modulations $\boldsymbol{f}$ and $\boldsymbol{g}$ are equal (i.e. the bimodules A_f and A_g are isomorphic) if and only if $f = \phi g \phi^{-1}$ for some invertible $\phi \in A$.*

Proof. If $f = \phi g \phi^{-1}$, $\boldsymbol{f}$ and $\boldsymbol{g}$ are both represented by A, with the same left A-module structures. To correct for the difference between the right actions of B, we introduce the bijective map $\mathbf{\Phi} : A_f \to A_g$ defined by $x \mapsto x\phi$, which is a bimodule isomorphism because

$$\mathbf{\Phi}(axf(b)) = axf(b)\phi = ax\phi\phi^{-1}f(b)\phi = ax\phi g(b) = a\mathbf{\Phi}(x)g(b).$$

For the converse, given a bimodule isomorphism $\mathbf{\Phi} : A_f \to A_g$, we define ϕ to be $\mathbf{\Phi}(1_A)$. By setting $x = 1_A$ in the bimodule morphism identities $\mathbf{\Phi}(ax) = a\mathbf{\Phi}(x)$ and $\mathbf{\Phi}(xf(b)) = \mathbf{\Phi}(x)g(b)$, we find first that $\mathbf{\Phi}(a) = a\phi$, so that ϕ is invertible because $\mathbf{\Phi}$ is, and then that $f(b)\phi = \phi g(b)$, or $f = \phi g \phi^{-1}$. □

Lemma 2.2. *A morphism $X \in \mathrm{Hom}(A, B)$ is the modulation of $f \in \mathrm{Hom}_0(A, B)$ if and only if it is isomorphic to A as a left A module.*

Proof. If X represents f, then clearly X is isomorphic to A as a left A module. For the converse, if $X = A$ as a left A module then X is isomorphic to A_f where $f(b) = 1_A \cdot b$. □

An invertible morphism in $\mathrm{Hom}(A, B)$ is called a *Morita equivalence* between A and B, and the group of Morita self-equivalences of A is called its *Picard group*. The modulation functor clearly takes algebra isomorphisms to Morita equivalences. In fact, we have:

Lemma 2.3. *The modulation of $f \in \mathrm{Hom}_0(A, B)$ is invertible if and only if f is invertible.*

Proof. A standard fact about Morita equivalence is that, if $X \in \mathrm{Hom}(A, B)$ is invertible, the natural homomorphisms from A and B to the B- and A-endomorphisms of X are isomorphisms. When $X = A_f$, the map which takes $b \in B$ to the operator of right multiplication by $f(b)$ is injective if and only if f is injective. On the other hand, all of the left A-module endomorphisms of A are the right multiplications, so they are all realized by the action of B if and only if f is surjective. □

Remark 2.4. It is also possible to modulate a nonunital f. In this case, the underlying k-module should be taken to be the left ideal I in A generated by $f(1_B)$, so that the bimodule structure (1) is still biunital. The three lemmas above change to the following statements, whose proofs are similar, so we only sketch them.

Lemma 2.1′. *If f and g are algebra homomorphisms $A \leftarrow B$ not necessarily unital, then their modulations $\boldsymbol{f}$ and $\mathbf{g}$ are equal if and only if there are elements $\phi \in A \cdot f(1_B)$ and $\psi \in A \cdot g(1_B)$ such that $\phi\psi = g(1_B)$, $\psi\phi = f(1_B)$, $g = \phi f \psi$, satisfying the two additional conditions that $x\phi\psi = 0$ implies $x\phi = 0$ and $x\psi\phi = 0$ implies $x\psi = 0$.*

Sketch of proof. Given an isomorphism Φ to $\boldsymbol{f}$ from $\boldsymbol{g}$, let $\phi = \Phi(g(1_B))$ and $\psi = \Phi^{-1}(f(1_B))$. Then $\Phi(xg(1_B)) = x\phi$ and $\Phi^{-1}(xf(1_B)) = x\psi$. All this gives us the desired equations and properties. For the converse, the morphism $\Phi(xg(1_B)) := x\phi$ is an isomorphism from $A \cdot g(1_B)$ to $A \cdot f(1_B)$ with inverse $\Phi^{-1}(xf(1_B)) := x\psi$. The two additional conditions make Φ and Φ^{-1} well defined. □

Lemma 2.2′. *A morphism $X \in \mathrm{Hom}(A, B)$ is the modulation of a (not necessarily unital) map $f : A \leftarrow B$ if and only if it is represented by a principal left ideal in A.*

Sketch of proof. If X is the modulation of f, then $X = A \cdot f(1_B)$. For the converse, if X is isomorphic to a left A ideal $A \cdot c$, then X is the modulation of $f : b \mapsto c \cdot b$, where $b \in B$ and $\cdot$ is the right action of B on $X = A \cdot c$. □

Lemma 2.3′. *When $f(1_B)$ is in the center of A, the modulation of a morphism $f : A \leftarrow B$ (not necessarily unital) is invertible if and only if f is an isomorphism from B to $A \cdot f(1_B)$ and $f(1_B)$ is not a zero divisor.*

Sketch of proof. One applies the same argument. If $\boldsymbol{f}$ is invertible, notice that $A \to \mathrm{End}_B(X)$ by $a \mapsto a\cdot$ is an isomorphism, therefore $af(1_B) \neq a' f(1_B)$ if $a \neq a'$. This implies that $f(1_B)$ is not a zero divisor. As before f has to be injective. For any $a \in A$, right multiplication by a is in $\mathrm{End}_A(X)$, therefore there is $b \in B$ such that $f(1_B)a = f(b)$. It is not hard to prove the converse. □

Finally, we recall that every (A, B) bimodule gives rise (via tensor product over B) to a k-linear functor from the category of left B-modules to that of left A-modules, that isomorphisms between bimodules produce naturally equivalent functors, and that invertible elements of $\mathrm{Hom}(A, B)$ correspond to homotopy classes of equivalences of categories. (The Eilenberg–Watts theorem characterizes the functors arising from bimodules as those which commute with finite limits and colimits.)

Sesquialgebras. To make the notion of biunital bialgebra Morita invariant, we introduce the following definition. For simplicity of notation, we omit the subscript $_k$ on tensor products over k, and the unadorned asterisk * will denote the k-dual.

Definition 2.4. A *sesquiunital sesquialgebra* over a commutative ring k is a unital k-algebra A equipped with an $(A \otimes A, A)$-bimodule $\boldsymbol{\Delta}$ (the *coproduct*) and a (k, A)-module (that is, a right A module) $\boldsymbol{\epsilon}$ (the *counit*), satisfying the following properties.

(1) (coassociativity) The $(A \otimes A \otimes A, A)$-bimodules

$$(A \otimes \mathbf{\Delta}) \otimes_{A \otimes A} \mathbf{\Delta} \quad \text{and} \quad (\mathbf{\Delta} \otimes A) \otimes_{A \otimes A} \mathbf{\Delta}$$

are isomorphic.

(2) (counit) The $(k \otimes A, A) = (A \otimes k, A) = (A, A)$-bimodules

$$(\boldsymbol{\epsilon} \otimes A) \otimes_{A \otimes A} \mathbf{\Delta} \quad \text{and} \quad (A \otimes \boldsymbol{\epsilon}) \otimes_{A \otimes A} \mathbf{\Delta}$$

are isomorphic to A.

For example, if (A, Δ, ϵ) is a biunital bialgebra, then its modulation $(A, \mathbf{\Delta}, \boldsymbol{\epsilon})$ is a sesquiunital sesquialgebra. If we have a Morita equivalence X between A and another algebra B, we can use composition with X and $X \otimes X$ to put a biunital sesquialgebra structure on B. See Section 5 below for more details.

3. The antipode and hopfish algebras

Our definition of sesquiunital sesquialgebra expresses (with arrows reversed) the usual axioms of a monoid (semigroup with identity) in the category Alg. A monoid is a group when all its elements have inverses, so it is natural to look for a sesquialgebraic analogue of the inverse. In a Hopf algebra, the antipode, which encodes inversion, is an algebra *anti*homomorphism $S : A \to A$. The properties of inversion ($gg^{-1} = e = g^{-1}g$ for every group element) are then expressed as commutativity of two diagrams, or equality of compositions

$$1 \circ \epsilon = \mu \circ \beta \circ \Delta, \tag{2}$$

where $1 : k \to A$ is inclusion of the scalars, $\mu : A \otimes A \to A$ is algebra multiplication, and $\beta : A \otimes A \to A \otimes A$ is either $I \otimes S$ or $S \otimes I$ (I being the identity morphism on A).

When A is noncommutative, the maps μ and β are k-linear but not algebra homomorphisms. One can consider S as a homomorphism from A to the opposite algebra A^{op}, or vice versa, but there is no way to correct μ in such a manner. As a result, we see no way to rewrite (2) in the category Alg. Instead, we take an alternate approach, which may also be useful elsewhere in the theory of Hopf algebras.

We keep in mind the example where A is the algebra of k-valued functions on a group G.

One way to characterize groups among monoids without explicitly postulating the existence of inverses is to consider the subset

$$J = \{(g, h) \mid gh = e\} \subset G \times G$$

and require that it project bijectively to one factor in the product. To represent J algebraically, even when A is noncommutative, we borrow an idea from Poisson geometry [Lu 1993], where coisotropic submanifolds become one-sided ideals when a Poisson manifold is quantized to become a noncommutative algebra.

We begin, then, with the space $Z' = \mathrm{Hom}_A(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$ of right module homomorphisms. (In the group case, Z' plays the role of measures on $G \times G$ which are supported on J.) Using the left $A \otimes A$ module structure on $\boldsymbol{\Delta}$, we define a right $A \otimes A$ module structure on Z' by $(gb)(u) = g(bu)$ for g in Z', b in $A \otimes A$ and u in $\boldsymbol{\Delta}$. Note that Z' is completely determined by $\boldsymbol{\epsilon}$ and $\boldsymbol{\Delta}$ and is not an extra piece of data.

For the algebraic model of functions on J, we must take a *predual* of Z', that is, a left $A \otimes A$-module Z whose k-dual Z^* is equipped with a right $A \otimes A$-module isomorphism with Z'.

Definition 3.1. A *preantipode* for a sesquiunital sesquialgebra A over k is a left $A \otimes A$ module $\boldsymbol{S}$ together with an isomorphism of its k-dual with the right $A \otimes A$ module $\mathrm{Hom}_A(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$.

Since a left A module is also a right A^{op} module, we may consider $\boldsymbol{S}$ as an (A, A^{op}) bimodule, where $(A, \cdot)$ is from the left A in $A \otimes A$ and $(\cdot, A^{\mathrm{op}})$ is from the right one, i.e. as an Alg morphism in $\mathrm{Hom}(A, A^{\mathrm{op}})$.

The following is our way of expressing algebraically that the first projection from J to G is bijective.

Definition 3.2. Let A be a sesquiunital sesquialgebra. If a preantipode $\boldsymbol{S}$, considered as an (A, A^{op}) bimodule, is a free left A module of rank 1, we call $\boldsymbol{S}$ an *antipode* and say that A along with $\boldsymbol{S}$ is a *hopfish algebra*.

By Lemma 2.2, $\boldsymbol{S}$ is the modulation of an algebra homomorphism $A \leftarrow A^{\mathrm{op}}$. Thus, the definition is effectively that there is a homomorphism S to A from A^{op} such that the full k-dual of the modulation of S is isomorphic to $\mathrm{Hom}_A(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$.

4. Hopf and quasi-Hopf algebras as hopfish algebras

As we observed earlier, the modulation of a biunital bialgebra is a sesquiunital sesquialgebra. In this section, we will give an explicit description of a preantipode in this case, and we will show that the modulation of a Hopf algebra is hopfish. Although this is a special case of the quasi-Hopf algebras treated later in this section, we deal separately with the Hopf case because the proof is much simpler.

Let (A, Δ, ϵ) be a biunital bialgebra. Considering the modulations $\boldsymbol{\epsilon} = k$ and $\boldsymbol{\Delta} = A \otimes A$ as right A modules respectively, one may identify Z' with the subspace of $(A \otimes A)^* = \mathrm{Hom}_k(k, A \otimes A)$ consisting of those linear functionals which annihilate the left ideal W generated by

$$\{\epsilon(a)(1\otimes 1)-\Delta(a)|a\in A\},$$

i.e. with the k-module dual to $(A\otimes A)/W$. We may therefore take the (cyclic) left $A\otimes A$ module $\boldsymbol{S}_1=(A\otimes A)/W$ as a preantipode.

We will use the following lemma later. Its straightforward proof is left to the reader.

Lemma 4.1. *W is equal to the left ideal generated by* $\Delta(\ker\epsilon)$.

Now suppose that A is equipped with an antipode S making it into a Hopf algebra. We will consider S as a homomorphism $A\leftarrow A^{\mathrm{op}}$, with modulation $\boldsymbol{S}$. As a k-module, $\boldsymbol{S}$ is A; its (A,A^{op}) bimodule structure is $a\cdot x\cdot b=axS(b)$.

If we can show that the preantipode $\boldsymbol{S}_1$ is isomorphic to $\boldsymbol{S}$ as a bimodule, then since $\boldsymbol{S}$ is isomorphic to A as a left A-module, $\boldsymbol{S}=\boldsymbol{S}_1$ is an antipode, making the modulation of A into a hopfish algebra.

We define a map $\phi: A\otimes A\to A$ by

$$a\otimes b\mapsto aS(b),$$

This map is obviously a morphism of (A,A^{op}) bimodules because

$$\begin{aligned}\phi(c\cdot(a\otimes b))&=\phi(ca\otimes b)=caS(b)=c\cdot(aS(b)),\\ \phi((a\otimes b)\cdot c)&=\phi(a\otimes cb))=aS(b)S(c)=(aS(b))\cdot c.\end{aligned}$$

Hence this map descends to $\boldsymbol{S}_1=(A\otimes A)/W$ because

$$\phi(\epsilon(a)(1\otimes 1)-\Delta(a))=1\circ\epsilon(a)-(\mathrm{id}\otimes S)\circ\Delta(a)=0.$$

The induced map from $\boldsymbol{S}_1$ to A, which we also denote by ϕ, is also a morphism of (A,A^{op}) bimodules.

Moreover ϕ is surjective, since it has a left inverse $a\mapsto[a\otimes 1]$, where $[\,]$ denotes the equivalence class modulo W. This map is also a right inverse, and ϕ is injective, if and only if the equation

$$1\otimes a-S(a)\otimes 1\in W \tag{3}$$

is satisfied for all $a\in A$. Notice that $aS(b)\otimes 1-a\otimes b=(a\otimes 1)(S(b)\otimes 1-1\otimes b)$ and W is a left ideal. Since $\mathrm{id}=m\circ(\mathrm{id}\otimes\epsilon)\circ\Delta$, composing with S we have $\sum S(a_1)\epsilon(a_2)=\sum S(a_1\epsilon(a_2))=S(a)$. (Here we use Sweedler's notation $\Delta(a)=\sum a_1\otimes a_2$ and $\sum\Delta(a_1)\otimes a_2=\sum a_{1,1}\otimes a_{1,2}\otimes a_2$, etc.) On the other hand, we have

$$\begin{aligned}\sum(S(a_1)\otimes 1)\cdot\Delta(a_2)&=\sum(S(a_1)a_{2,1})\otimes a_{2,2}\\ &=\sum(S(a_{1,1})a_{1,2})\otimes a_2=\sum 1\otimes\epsilon(a_1)a_2=1\otimes a.\end{aligned}$$

We explain the equalities above as follows. The first equality just comes from the notation and the multiplication in the tensor product algebra. For the second, we consider the map $s : A \otimes A \otimes A \to A \otimes A$ defined by $s(a \otimes b \otimes c) = S(a)b \otimes c$. Coassociativity and evaluation of s give

$$\begin{aligned}\sum s(a_1 \otimes a_{2,1} \otimes a_{2,2}) &= \sum s(a_{1,1} \otimes a_{1,2} \otimes a_2)\\ &= \sum (S(a_1)a_{2,1}) \otimes a_{2,2} = \sum (S(a_{1,1})a_{1,2}) \otimes a_2.\end{aligned}$$

For the third equality, we have used the property of S that $\mu \circ (S \otimes \mathrm{id}) \circ \Delta = 1 \circ \epsilon$. Therefore,

$$1 \otimes a - S(a) \otimes 1 = \sum (S(a_1) \otimes 1)(-\epsilon(a_2) + \Delta(a_2)) \in W.$$

So (3) is proved, hence $\boldsymbol{S}$ and $\boldsymbol{S}_1$ are isomorphic as (A, A^{op}) bimodules.

We have thus proved the following theorem.

Theorem 4.2. *Let (A, Δ, ϵ) be a biunital bialgebra. Then $(A \otimes A)/W$, where W is the left ideal generated by*

$$\{\epsilon(a)(1 \otimes 1) - \Delta(a) \mid a \in A\},$$

is a preantipode for the modulation of A. If A is a Hopf algebra, with antipode S, then $(A \otimes A)/W$ is isomorphic to the modulation $\boldsymbol{S}$, and $(A, \boldsymbol{\Delta}, \boldsymbol{\epsilon}, \boldsymbol{S})$ is a hopfish algebra.

Remark 4.3. The hopfish antipode $\boldsymbol{S}$ is also isomorphic to A^{op} as a right A^{op}-module if and only if the Hopf antipode S is invertible. This is why we use a "one sided" criterion for a preantipode to be an antipode.

We turn now to quasi-Hopf algebras. Recall that a quasibialgebra (A, ϵ, Δ, S) is nearly a bialgebra, except that the coproduct does not satisfy associativity exactly; instead, there is an invertible element $\Phi \in A \otimes A \otimes A$ (the coassociator), satisfying

$$(\mathrm{id} \otimes \Delta)(\Delta(a)) = \Phi^{-1}(\Delta \otimes \mathrm{id})(\Delta(a))\Phi \quad \text{for all } a \in A, \tag{4}$$

and further coherence conditions,

$$\begin{aligned}(\Delta \otimes \mathrm{id} \otimes \mathrm{id})(\Phi) \cdot (\mathrm{id} \otimes \mathrm{id} \otimes \Delta)(\Phi) &= (\Phi \otimes 1) \cdot (\mathrm{id} \otimes \Delta \otimes \mathrm{id})(\Phi) \cdot (1 \otimes \Phi),\\ (\epsilon \otimes \mathrm{id}) \circ \Delta &= \mathrm{id} = (\mathrm{id} \otimes \epsilon) \circ \Delta,\\ (\mathrm{id} \otimes \epsilon \otimes \mathrm{id})(\Phi) &= 1.\end{aligned} \tag{5}$$

Since the modulation functor "kills" inner automorphisms (Lemma 2.1), the modulation of a quasibialgebra is a sesquialgebra.

Now A is a quasi-Hopf algebra if there is an anti-homomorphism $S : A \to A$ and elements α, β in A, such that

$$\sum S(a_1)\alpha a_2 = \epsilon(a)\alpha, \quad \sum a_1\beta S(a_2) = \epsilon(a)\beta \quad \text{for all } a \in A, \tag{6}$$

where we use Sweedler's notation: $\Delta(a) = \sum a_1 \otimes a_2$. There are also higher coherence conditions for α and β; see [Drinfel'd 1989] for details.

The following proposition is a slight modification of [Drinfel'd 1989, Proposition 1.5]. Unlike Drinfel'd, we do not assume that S is invertible, so we can not obtain the "right" part of his proposition, but his "left" part can be proved under weaker hypotheses.

Proposition 4.4. *Let $(A, \Delta, \epsilon, \Phi, S, \alpha, \beta)$ be a quasi-Hopf algebra, with $\Phi = \sum_i X_i \otimes Y_i \otimes Z_i$ and $\Phi^{-1} = \sum_j P_j \otimes Q_j \otimes R_j$. Define*

$$\omega = \sum_j S(P_j)\alpha Q_j \otimes R_j \in A \otimes A.$$

Denote by W the left ideal of $A \otimes A$ generated by $\Delta(\ker \epsilon)$. Then

(1) *the k-linear mappings $\phi, \psi : A \otimes A \to A \otimes A$, given by*

$$\phi(a \otimes b) = (a \otimes 1)\omega\Delta(b), \quad \psi(a \otimes b) = \sum_i aX_i\beta S(Y_i)S(b_1) \otimes b_2 Z_i,$$

are bijective, where we have used Sweedler's notation $\Delta b = b_1 \otimes b_2$;

(2) *the mapping $a \otimes b \mapsto (\mathrm{id} \otimes \epsilon)(\phi^{-1}(a \otimes b))$ induces a bijection $(A \otimes A)/W \to A$, and $(\mathrm{id} \otimes \epsilon)(\phi^{-1}(a \otimes b)) = a\beta S(b)$;*

Proof. First, $\phi\psi = \mathrm{id} = \psi\phi$. We will prove only that $\phi\psi = \mathrm{id}$; the other equation can be derived by the same method, as in [Drinfel'd 1989]. We have

$$\begin{aligned}
\phi\psi(a \otimes b) &= \sum_i \phi(aX_i\beta S(Y_i)S(b_1) \otimes b_2 Z_i) \\
&= \sum_i (aX_i\beta S(Y_i)S(b_1) \otimes 1)\omega\Delta(b_2)\Delta(Z_i) \\
&= \sum_i (a \otimes 1)(X_i\beta S(Y_i) \otimes 1)\big((S(b_1) \otimes 1)\omega\Delta b_2\big)\Delta Z_i \\
&= \sum_i (a \otimes 1)(X_i\beta S(Y_i) \otimes 1)(B)\Delta(Z_i),
\end{aligned}$$

where $B = (S(b_1) \otimes 1)\omega\Delta b_2$.

We insert the definition of ω in B, and have

$$
\begin{aligned}
&(S(b_1)\otimes 1)\omega\Delta b_2\\
&\qquad = \sum_j \big(S(b_1)S(P_j)\alpha Q_j\otimes R_j\big)\Delta b_2\\
&\qquad = \sum_j (m\otimes \mathrm{id})\big((S\otimes\alpha\cdot\otimes \mathrm{id})\big((P_jb_1\otimes Q_j\otimes R_j)(1\otimes\Delta b_2)\big)\big)\\
&\qquad = \sum_j (m\otimes \mathrm{id})\big((S\otimes\alpha\cdot\otimes \mathrm{id})\big((P_j\otimes Q_j\otimes R_j)(b_1\otimes 1\otimes 1)(1\otimes\Delta b_2)\big)\big)\\
&\qquad = (m\otimes \mathrm{id})\big((S\otimes\alpha\cdot\otimes \mathrm{id})\big(\Phi^{-1}(1\otimes\Delta)\Delta(b)\big)\big),
\end{aligned}
$$

where $m : A\otimes A\to A$ is the multiplication on A and $\alpha\cdot\ : A\to A$ is the left multiplication by α.

Using the twisted coassociativity $(\mathrm{id}\otimes\Delta)\Delta=\Phi(\Delta\otimes\mathrm{id})(\Delta)\Phi^{-1}$ we continue the calculation above to find that B is equal to

$$
\begin{aligned}
&(m\otimes \mathrm{id})\big((S\otimes\alpha\cdot\otimes \mathrm{id})\big((\Delta\otimes\mathrm{id})\Delta(b)\Phi^{-1}\big)\big)\\
&\qquad = \sum_j (m\otimes \mathrm{id})\big((S\otimes\alpha\cdot\otimes \mathrm{id})\big(b_{11}P_j\otimes b_{12}Q_j\otimes b_2R_j\big)\big)\\
&\qquad = \sum_j (m\otimes \mathrm{id})\big(S(P_j)S(b_{11})\otimes\alpha b_{12}Q_j\otimes b_2R_j\big)\\
&\qquad = \sum_j S(P_j)S(b_{11})\alpha b_{12}Q_j\otimes b_2R_j\\
&\qquad = \sum_j S(P_j)\alpha\epsilon(b_1)Q_j\otimes b_2R_j = \sum_j S(P_j)\alpha Q_j\otimes\epsilon(b_1)b_2R_j\\
&\qquad = \sum_j S(P_j)\alpha Q_j\otimes bR_j = (1\otimes b)\sum_j (S(P_j)\alpha Q_j\otimes R_j),
\end{aligned}
$$

where in the fourth equality we have used a property of the antipode S, and at the fifth we have used a property of ϵ.

Substituting the expression above for B in the calculation of $\phi\psi$, we have

$$
\begin{aligned}
\phi\psi(a\otimes b) &= \sum_{i,j}(a\otimes 1)(X_i\beta S(Y_i)\otimes 1)(1\otimes b)(S(P_j)\alpha Q_j\otimes R_j)\Delta(Z_i)\\
&= (a\otimes b)\sum_{i,j}(X_i\beta S(Y_i)\otimes 1)(S(P_j)\alpha Q_j\otimes R_j)\Delta(Z_i).
\end{aligned}
$$

Next, we show that $U=\sum_{i,j}(X_i\beta S(Y_i)\otimes 1)(S(P_j)\alpha Q_j\otimes R_j)\Delta(Z_i)$ is equal to 1. We define the k-linear map $\Psi : A\otimes A\otimes A\otimes A\to A\otimes A$ by

$$\Psi(a\otimes b\otimes c\otimes d)=a\beta S(b)\alpha c\otimes f,$$

so U can be written as

$$\begin{aligned}
&\sum_{i,j}(X_i\beta S(Y_i)\otimes 1)(S(P_j)\alpha Q_j\otimes R_j)\Delta(Z_i)\\
&\qquad\qquad=\sum_{i,j} X_i\beta S(Y_i)S(P_j)\alpha Q_j Z_{i1}\otimes R_j Z_{i2}\\
&\qquad\qquad=\sum_{i,j}\Psi((1\otimes P_j\otimes Q_j\otimes R_j)(X_i\otimes Y_i\otimes Z_{i1}\otimes Z_{i2}))\\
&\qquad\qquad=\Psi\big((1\otimes\Phi^{-1})(\mathrm{id}\otimes\mathrm{id}\otimes\Delta)(\Phi)\big).
\end{aligned}$$

Using the coherence condition

$$(\mathrm{id}\otimes\mathrm{id}\otimes\Delta)(\Phi)(\Delta\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)=(1\otimes\Phi)(\mathrm{id}\otimes\Delta\otimes\mathrm{id})(\Phi)(\Phi\otimes 1), \tag{7}$$

we get

$$\begin{aligned}
(1\otimes\Phi^{-1})(\mathrm{id}\otimes\mathrm{id}\otimes\Delta)(\Phi)&=(\mathrm{id}\otimes\Delta\otimes\mathrm{id})(\Phi)(\Phi\otimes 1)(\Delta\otimes\mathrm{id}\otimes\mathrm{id})(\Phi^{-1})\\
&=\sum_{i,j,k} X_iX_jP_{k1}\otimes Y_{i1}Y_jP_{k2}\otimes Y_{i2}Z_jQ_k\otimes Z_iR_k.
\end{aligned}$$

Hence $\Psi\big((1\otimes\Phi^{-1})(\mathrm{id}\otimes\mathrm{id}\otimes\Delta)(\Phi)\big)$ is equal to

$$\begin{aligned}
&\sum_{i,j,k}\Psi(X_iX_jP_{k1}\otimes Y_{i1}Y_jP_{k2}\otimes Y_{i2}Z_jQ_k\otimes Z_iR_k)\\
&\qquad\qquad=\sum_{i,j,k} X_iX_jP_{k_1}\beta S(P_{k2})S(Y_j)S(Y_{i1})\alpha Y_{i2}Z_jQ_k\otimes Z_iR_k\\
&\qquad\qquad=\sum_{i,j,k} X_iX_j\beta\epsilon(P_k)S(Y_j)\epsilon(Y_i)\alpha Z_jQ_k\otimes Z_iR_k\\
&\qquad\qquad=\sum_{i,j,k} X_i\epsilon(P_k)(X_j\beta S(Y_j)\alpha Z_j)\epsilon(Y_i)Q_k\otimes Z_iR_k\\
&\qquad\qquad=\sum_{i,k} X_i\epsilon(P_k)\epsilon(Y_i)Q_k\otimes Z_iR_k.
\end{aligned}$$

In the second equality, we used properties of the antipode: $P_{k1}\beta S(P_{k2})=\beta\epsilon(P_k)$ and $S(Y_{i1})\alpha Y_{i2}=\alpha\epsilon(Y_i)$. In the last equality, we used $\sum_j X_j\beta S(Y_j)\alpha Z_j=1$.

We evaluate $\mathrm{id}\otimes\epsilon\otimes\mathrm{id}\otimes\mathrm{id}$ on both sides of (7), and since ϵ is an homomorphism from A to k, we obtain

$$\begin{aligned}
&(\mathrm{id}\otimes\epsilon\otimes\Delta)(\Phi)((\mathrm{id}\otimes\epsilon)\Delta\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)\\
&\qquad=(\mathrm{id}\otimes((\epsilon\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)))(\mathrm{id}\otimes(\epsilon\otimes\mathrm{id})\Delta\otimes\mathrm{id})(\Phi)((\mathrm{id}\otimes\epsilon\otimes\mathrm{id})(\Phi)\otimes\mathrm{id}).
\end{aligned}\tag{8}$$

In the definition of a quasi-Hopf algebra, we assumed that $\mathrm{id}\otimes\epsilon\otimes\mathrm{id}(\Phi)=1$. Therefore, $(\mathrm{id}\otimes\epsilon\otimes\Delta)(\Phi)=(\mathrm{id}\otimes\mathrm{id}\otimes\Delta)(\mathrm{id}\otimes\epsilon\otimes\mathrm{id})(\Phi)=1$. Hence, by

$(\mathrm{id}\otimes\epsilon))\Delta=\mathrm{id}\otimes 1$, the left-hand side of (8) is equal to

$$((\mathrm{id}\otimes\epsilon)\Delta\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)=\sum_i X_i\otimes 1\otimes Y_i\otimes Z_i.$$

The right-hand side of (8) is equal to

$$(\epsilon\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)\Big(\sum_i X_i\otimes 1\otimes Y_i\otimes Z_j\Big).$$

Therefore, we have

$$\sum_i X_i\otimes 1\otimes Y_i\otimes Z_i=(\epsilon\otimes\mathrm{id}\otimes\mathrm{id})(\Phi)\Big(\sum_i X_i\otimes 1\otimes Y_i\otimes Z_j\Big). \tag{9}$$

We multiply both sides of (9) by $\sum_i P_j\otimes 1\otimes Q_j\otimes R_j$ and obtain

$$\epsilon\otimes\mathrm{id}\otimes\mathrm{id}(\Phi)=1.$$

So we have $\epsilon\otimes\mathrm{id}\otimes\mathrm{id}(\Phi^{-1})=\epsilon\otimes\mathrm{id}\otimes\mathrm{id}(\Phi^{-1}\Phi)=1$.

Finally,

$$\begin{aligned}\sum_{i,j} X_i\epsilon(P_k)\epsilon(Y_i)Q_k\otimes Z_iR_k&=\sum_{i,k}(m\otimes\mathrm{id})(X_i\otimes\epsilon(Y_i)\otimes Z_i)(\epsilon(P_k)\otimes Q_k\otimes R_k)\\&=(m\otimes\mathrm{id})\big((\mathrm{id}\otimes\epsilon\otimes\mathrm{id})(\Phi)(\epsilon\otimes\mathrm{id}\otimes\mathrm{id})(\Phi^{-1})\big)\\&=1.\end{aligned}$$

In conclusion, we have shown that $\phi\psi(a\otimes b)=a\otimes b$ and similarly $\psi\phi(a\otimes b)=a\otimes b$. Therefore, ϕ and ψ are invertible. This completes the proof of the first statement of Proposition 4.4.

Now we calculate $(\mathrm{id}\otimes\epsilon)\phi^{-1}(a\otimes b)$. By the proof above, ψ is the inverse of ϕ, and

$$\begin{aligned}(\mathrm{id}\otimes\epsilon)\phi^{-1}(a\otimes b)&=(\mathrm{id}\otimes\epsilon)\Big(\sum_i aX_i\beta S(Y_i)S(b_1)\otimes b_2Z_i\Big)\\&=\sum_i aX_i\beta S(Y_i)S(b_1)\epsilon(b_2)\epsilon(Z_i)\\&=\sum_i aX_i\beta S(Y_i)S(b_1\epsilon(b_2))\epsilon(Z_i)\\&=\sum_i aX_i\beta S(Y_i)S(b)\epsilon(Z_i).\end{aligned}$$

To show that the last term is equal to $a\beta S(b)$, we consider the k-linear map $\Upsilon: A\otimes A\otimes A\to A$ defined by $\Upsilon(a_1\otimes a_2\otimes a_3)=a_1\beta S(a_2)a_3$. Accordingly, we have $\sum_i X_i\beta S(Y_i)\epsilon(Z_i)=\Upsilon((\mathrm{id}\otimes\mathrm{id}\otimes\epsilon)(\Phi))$. By applying $\mathrm{id}\otimes\mathrm{id}\otimes\epsilon\otimes\mathrm{id}$ to (8), we have similarly $(\mathrm{id}\otimes\mathrm{id}\otimes\epsilon)(\Phi)=1\otimes 1\otimes 1$. So $\sum_i X_i\beta S(Y_i)\epsilon(Z_i)=\Upsilon(1)=\beta$, and $\sum_i aX_i\beta S(Y_i)S(b)\epsilon(Z_i)$ is equal to $a\beta S(b)$.

Therefore if there is an element in W, which can be written as $\Delta(\mu)$, where μ is in the kernel of ϵ, then $(\mathrm{id}\otimes\epsilon)\phi^{-1}(\Delta(\mu))=\mu_1\beta S(\mu_2)=\epsilon(\mu)\beta=0$. This shows that W is contained in the kernel of the map $(\mathrm{id}\otimes\epsilon)\phi^{-1}: A\otimes A\to A$. Finally, we show that $(\mathrm{id}\otimes\epsilon)\phi^{-1}$ is a bijection from $A\otimes A/W$ to A. If $\sum_i x_i\otimes y_i$ is in the kernel of $(\mathrm{id}\otimes\epsilon)\phi^{-1}$. We define $\sum_j a_j\otimes b_j$ to be equal to $\phi\big(\sum_i x_i\otimes y_i\big)$, and $(\mathrm{id}\otimes\epsilon)\big(\sum_j a_j\otimes b_j\big)=\sum_j a_j\epsilon(b_j)=0$. Then $\sum_i x_i\otimes y_i$ is equal to

$$\begin{aligned}\sum_i x_i\otimes y_i&=\sum_j\phi(a_j\otimes b_j)=\sum_j(a_j\otimes 1)\omega\Delta(b_j)\\&=\sum_j(a_j\otimes 1)\omega(\Delta(b_j)-\epsilon(b_j))\in W,\end{aligned}$$

where in the third equality, we have used that

$$\sum_j(a_j\otimes 1)\omega\epsilon(b_j)=\sum_j(a_j\epsilon(b_j)\otimes 1)\omega=0.\qquad\square$$

By the same arguments used in Theorem 4.2, we obtain:

Theorem 4.5. *Let (A,Δ,ϵ,Φ) be a biunital quasibialgebra, and let W be the left ideal generated by $\{\epsilon(a)(1\otimes 1)-\Delta(a)\mid a\in A\}$. Then $(A\otimes A)/W$ is a preantipode for the modulation of A.*

If A is a quasi-Hopf algebra, with antipode (S,α,β), then $(A\otimes A)/W$ is isomorphic to the modulation $\boldsymbol{S}$, and $(A,\boldsymbol{\Delta},\boldsymbol{\epsilon},\boldsymbol{S})$ is a hopfish algebra.

5. Morita invariance

The following theorem shows that, with our definition of hopfish algebra, we are on the right track toward defining a Morita invariant notion.

Theorem 5.1. *Let A be a quasi-Hopf algebra and B an algebra Morita equivalent to A. Then B is a sesquiunital sesquialgebra with a preantipode.*

Proof. Let P be an (A,B)-bimodule, and Q a (B,A)-bimodule, inverse to one another in the category Alg. We recall the hopfish structure on A defined in Theorem 4.5, with

$$\boldsymbol{\epsilon}^A=k,\quad \boldsymbol{\Delta}^A=A\otimes A,\quad \boldsymbol{S}^A=A\otimes A/W.$$

We use the bimodules P and Q to define

$$\boldsymbol{\epsilon}^B:=\boldsymbol{\epsilon}^A\otimes_A P,\qquad \boldsymbol{\Delta}^B:=\big(Q\otimes Q\big)\otimes_{A\otimes A}\boldsymbol{\Delta}^A\otimes_A P,$$

These data make B into a sesquiunital sesquilinear algebra.

Now we define

$$S^B := \big(Q \otimes Q\big) \otimes_{A \otimes A} S^A.$$

Remark 5.2. Our definition of the antipode S^B only uses the bimodule Q, not P. This is because Q is a (B, A) bimodule, and therefore is also an $(A^{\mathrm{op}}, B^{\mathrm{op}})$ bimodule naturally. Since S^A is an (A, A^{op}) bimodule, $Q \otimes_A S^B \otimes_{A^{\mathrm{op}}} Q$ defines a (B, B^{op}) bimodule, which is isomorphic to $(Q \otimes Q) \otimes_{A \otimes A} S^A$.

In the following, we will show that S^B is a preantipode:

$$\mathrm{Hom}_k(k, S^B) \cong \mathrm{Hom}_B(\epsilon^B, \Delta^B).$$

According to our definitions, we have

$$\mathrm{Hom}_B(\epsilon^B, \Delta^B) = \mathrm{Hom}_B\big(\epsilon^A \otimes_A P, \big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A \otimes_A P\big).$$

Since the Morita equivalence between A and B defines an equivalence of right-module categories, we have a natural isomorphism

$$\mathrm{Hom}_B\big(\epsilon^A \otimes_A P, \big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A \otimes_A P\big) \cong \mathrm{Hom}_A\big(\epsilon^A, \big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A\big).$$

The space $\mathrm{Hom}_A\big(\epsilon^A, \big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A\big)$ consists of k-linear morphisms from $\big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A$ to k, vanishing on the A-submodule $\widetilde{W}$ spanned by

$$(q_1 \otimes q_2) \otimes_{A \otimes A} (a_1 \otimes a_2)(\epsilon(a)1 \otimes 1 - \Delta(a)),\ q_i \in Q,\ a,\ a_i \in A,\ i = 1, 2.$$

The A-submodule $\widetilde{W}$ is isomorphic to $(Q \otimes Q) \otimes_{A \otimes A} W$, where W is defined as in Theorem 4.5. Therefore, $\mathrm{Hom}_A\big(\epsilon^A, \big(Q \otimes Q\big) \otimes_{A \otimes A} \Delta^A\big)$ is isomorphic to the k-dual of the quotient

$$(Q \otimes Q) \otimes_{A \otimes A} \Delta^A / \widetilde{W} \cong (Q \otimes Q) \otimes_{A \otimes A} (A \otimes A / W). \tag{10}$$

Replacing $A \otimes A / W$ by S^A in (10), we have

$$\mathrm{Hom}_B(\epsilon^B, \Delta^B) \cong \big((Q \otimes Q) \otimes_{A \otimes A} S^A\big)^* \cong (S^B)^*.$$

Therefore, S^B defines a preantipode on $(B, \Delta^B, \epsilon^B)$. □

Now we study when the sesquiunital sesquialgebra just defined is a hopfish algebra, i.e. when S^B is isomorphic to B as a left B-module.

We introduce the following special type of module over a hopfish algebra.

Definition 5.3. Let be A be a hopfish algebra with antipode bimodule S, and let X be a right A-module and therefore a left A^{op}-module. Then X is *self-conjugate* if $\mathrm{Hom}_A(A, X)$ is isomorphic to $S \otimes_{A^{\mathrm{op}}} X$ as a left A-module.

Remark 5.4. The category of finite dimensional left modules over a quasi-Hopf algebra is a rigid monoidal category. A self-dual module X of a quasi-Hopf algebra A is a self-dual object in the category of finite dimensional modules, i.e. $\mathrm{Hom}_k(k, X)$ is isomorphic to $\boldsymbol{S} \otimes_{A^{\mathrm{op}}} X$.

We can understand the definition of a self-conjugate module geometrically as follows. A hopfish algebra A can be thought as functions on a "noncommutative space with group structure" G. If we view a finite projective right A-module X as the space of sections of a "vector bundle" E over G, $\mathrm{Hom}_A(A, X)$ corresponds to the space of sections of the dual bundle E^*, and $\boldsymbol{S} \otimes_{A^{\mathrm{op}}} X$ is the pullback of the bundle E by the "inversion" map ι of G. The self-conjugacy condition on E says that E^* is isomorphic to $\iota^* E$.

Proposition 5.5. *With the same assumptions and notation as in Theorem 5.1, if the (B, A)-Morita equivalence bimodule Q is self-conjugate as a right A-module, then B is a hopfish algebra with antipode $\boldsymbol{S}^B$ defined in Theorem 5.1.*

Proof. Recall that the preantipode on B defined in Theorem 5.1 is equal to $(Q \otimes Q) \otimes_{A\otimes A} \boldsymbol{S}^A$. Since Q is a right A-module, it is also a left A^{op}-module, and the preantipode $\boldsymbol{S}^B$ can be rewritten as $Q \otimes_A \boldsymbol{S}^A \otimes_{A^{\mathrm{op}}} Q$.

Since Q is self-conjugate, we have $\boldsymbol{S}^A \otimes_{A^{\mathrm{op}}} Q \cong \mathrm{Hom}_A(A, Q)$, and so

$$Q \otimes_A \boldsymbol{S}^A \otimes_{A^{\mathrm{op}}} Q \cong Q \otimes_A \mathrm{Hom}_A(A, Q).$$

When Q is a Morita equivalence bimodule between A and B, Q is a finitely generated projective A-module and $B \cong \mathrm{Hom}_A(Q, Q) = Q \otimes_A \mathrm{Hom}_A(A, Q)$. This shows that $Q \otimes_A \boldsymbol{S}^A \otimes_{A^{\mathrm{op}}} Q$ is isomorphic to B as a left B-module. □

The following example is a special case of Proposition 5.5. We remark that given a (quasi)-Hopf algebra A, the matrix algebra $M_n(A)$ of $n \times n$ matrices with coefficients in A is not a (quasi-)Hopf algebra when $n \geq 2$.

Example 5.6. Let A be a quasi-Hopf algebra with $\boldsymbol{\epsilon}^A = k$, $\boldsymbol{\Delta}^A = A \otimes A$, and $\boldsymbol{S}^A = A$. Then the $n \times n$ matrix algebra $M_n(A) = B$ with coefficients in A is a hopfish algebra. We consider $Q = A^n$ as a space of column vectors, so that it has the structure of an $(M_n(A), A)$-bimodule, The counit $\boldsymbol{\epsilon}^B$ is A^n viewed as row vectors, i.e. as a $(k, M_n(A))$-bimodule. The coproduct $\boldsymbol{\Delta}^B$ is isomorphic to

$$\big(A^n \otimes A^n\big) \otimes_{A\otimes A} (A \otimes A) \otimes_A (A^n)^T = \big(A^n \otimes A^n\big) \otimes_{A\otimes A} (A^n)^T.$$

$\boldsymbol{S}^B$ is equal to $A^n \otimes_A A \otimes_{A^{\mathrm{op}}} A^n$. $A^n \otimes_A A \otimes_{A^{\mathrm{op}}} A^n$ is isomorphic to $M_n(A)$ as an $(M_n(A), M_n(A)^{\mathrm{op}})$-bimodule, where the left $M_n(A)$ module structure is from the standard left multiplication, while the right $M_n(A)^{\mathrm{op}}$ module structure is the composition of the left multiplication, transposition of matrices, and the antipode on A. Therefore, $B = M_n(A)$ is a hopfish algebra.

The following example shows that the self-conjugacy condition in Proposition 5.5 can not be eliminated.

Example 5.7. Consider the cyclic group $\mathbb{Z}/3\mathbb{Z}$ with elements 0, 1, 2. The algebra A of functions on $\mathbb{Z}/3\mathbb{Z}$ is a commutative Hopf algebra spanned by the characteristic functions e_0, e_1, and e_2. We notice that the e_i's are projections in A, and denote the submodule $e_i A$ by A_i. Now consider the following projective module over A

$$Q = A_0^r \oplus A_1^s \oplus A_2^t,$$

where r, s, t are nonnegative integers. Then

$$B = \mathrm{Hom}_A(Q, Q) = A_0^{r^2} \oplus A_1^{s^2} \oplus A_2^{t^2}.$$

It is not difficult to see that Q is self-conjugate if and only if $s = t$.

We calculate the expression for $\boldsymbol{S}^B$ in Theorem 5.1:

$$\begin{aligned}(Q \otimes Q) \otimes_{A\otimes A} \boldsymbol{S}^A &= \big((A_0^r \oplus A_1^s \oplus A_2^t) \otimes (A_0^r \oplus A_1^s \oplus A_2^t)\big) \otimes_{A\otimes A} \boldsymbol{S}^A \\ &= \big(A_0^r \otimes (A_0^r \oplus A_1^s \oplus A_2^t)\big) \otimes_{A\otimes A} \boldsymbol{S}^A \\ &\qquad \oplus \big(A_1^s \otimes (A_0^r \oplus A_1^s \oplus A_2^t)\big) \otimes_{A\otimes A} \boldsymbol{S}^A \\ &\qquad\quad \oplus \big(A_2^t \otimes (A_0^r \oplus A_1^s \oplus A_2^t)\big) \otimes_{A\otimes A} \boldsymbol{S}^A.\end{aligned}$$

We look at the tensor product $(A_i \otimes A_j) \otimes_{A\otimes A} \boldsymbol{S}^A$. By Theorem 4.2, the antipode bimodule $\boldsymbol{S}^A$ is isomorphic to A. Therefore $(A_i \otimes A_j) \otimes_{A\otimes A} \boldsymbol{S}^A$ is equal to

$$(A_i \otimes A_j) \otimes_{A\otimes A} A = A_i \otimes_A A_j,$$

where the left A-module structure on A_j is the composition of the right multiplication with the antipode map $S : A \to A$.

We notice that $S(e_i)e_j = 0$ if $S(e_i) \neq e_j$. Therefore,

$$A_i \otimes_A A_j = \begin{cases} 0 & \text{if } S(e_i) \neq e_j, \\ A_i & \text{if } S(e_i) = e_j. \end{cases}$$

We conclude that $\boldsymbol{S}^B = A_0^{r^2} \oplus A_1^{st} \oplus A_2^{st}$. We observe that $\boldsymbol{S}^B$ is isomorphic to B as a left B module if and only if $s = t$.

Therefore, $\boldsymbol{S}^B$ is isomorphic to B if and only if Q is a self-conjugate A-module.

We define a notion of Morita equivalence between hopfish algebras.

Definition 5.8. Let $(A, \boldsymbol{\epsilon}^A, \boldsymbol{\Delta}^A, \boldsymbol{S}^A)$ and $(B, \boldsymbol{\epsilon}^B, \boldsymbol{\Delta}^B, \boldsymbol{S}^B)$ be two hopfish algebras. Then $(A, \boldsymbol{\epsilon}^A, \boldsymbol{\Delta}^A, \boldsymbol{S}^A)$ is Morita equivalent to $(B, \boldsymbol{\epsilon}^B, \boldsymbol{\Delta}^B, \boldsymbol{S}^B)$ if there is an (A, B)-bimodule ${}_A P_B$ and a (B, A)-bimodule ${}_B Q_A$ satisfying

(1) $P \otimes_B Q = A$, and $Q \otimes_A P = B$.

(2) $\boldsymbol{\epsilon}^B = \boldsymbol{\epsilon}^A \otimes_A P$,

(3) $\mathbf{\Delta}^B = (Q \otimes Q) \otimes_{A\otimes A} \mathbf{\Delta}^A \otimes_A P$,

(4) $\boldsymbol{S}^B = (Q \otimes Q) \otimes_{A\otimes A} \boldsymbol{S}^A$.

Proposition 5.9. *Definition 5.8 defines an equivalence relation among hopfish algebras.*

The proof is a straightforward check.

6. Hopfish structures on k^n

In this section, we give examples of hopfish algebras which are not Morita equivalent to modulations of Hopf algebras. In particular, we will describe hopfish structures on the commutative algebra k^G of k-valued functions on a finite set G which do not correspond to group structures on G.

We may identify the r-th tensor power of k^G with k^{G^r}. Since this algebra is commutative, we can also identify $(k^g)^{\mathrm{op}}$ with k^g.

If G is a semigroup, k^G is a bialgebra with coproduct $\Delta(a)(g,h) = a(gh)$, with a counit $\epsilon(a) = a(e)$ when G has an identity element e. When G is a group, we get a Hopf algebra structure by setting $S(a)(g) = a(g^{-1})$.

Now let G be a groupoid. We may make the same definitions as above, adding that $\Delta(a)(g,h)$ should be 0 when gh is not defined, and $\epsilon(a)$ is the sum of the values of a on all the identity elements. When G is not a group, k^G is no longer a Hopf algebra, but rather a *weak Hopf algebra* [Nikshych 2002, Example 2.3], since Δ is not unital and ϵ is not even an algebra homomorphism. When G is a groupoid, we have two algebra morphisms $\alpha, \beta: k^{G^0} \to k^G$ as the lifts of the source and target maps. The coproduct Δ is defined on $k^G \otimes_{k^{G^0}} k^G$ by $\Delta(a)(g,h) = a(gh)$, and counit $\epsilon : k^G \to k^{G^0}$ by $\epsilon(a)(e) = a(e)$, and the antipode S is defined by $S(a)(g) = a(g^{-1})$. $(k^G, \alpha, \beta, \Delta, \epsilon, S)$ is a *quantum groupoid* [Lu 1996]. It turns out that we can still form the modulation of these operators, and we still get a hopfish algebra because of the commutativity of the algebras k^G and k^{G^0}. To prove this, we will look at a more general situation.

Any sesquialgebra coproduct on $A = k^G$ is an $(A \otimes A, A)$-bimodule, i.e. a module $\mathbf{\Delta}$ over $k^{G\times G\times G}$. Such a module decomposes into submodules supported at the points of G^3. For our purposes, we will assume that these are free modules of finite rank. Then $\mathbf{\Delta}$ is determined up to isomorphism by the dimensions d^g_{hk} of the components $\mathbf{\Delta}^g_{hk}$, for $(g,h,k) \in G^3$. It is straightforward to check that the condition for coassociativity is precisely that the d^g_{hk}'s be the structure constants of an associative algebra $A' = \mathbb{Z}^G$ over $\mathbb{Z}$, i.e. a ring. Namely, identifying each element of G with its characteristic function, we have $gh = \sum_k d^k_{gh} k$. Similarly, a (k, A)-bimodule $\boldsymbol{\epsilon}$ with free submodules $\boldsymbol{\epsilon}^g$ as components is determined by the dimensions e^g of $\boldsymbol{\epsilon}^g$, and this bimodule is a counit precisely when $e := \sum_g e^g g$ is a unit element for A'. We say that such sesquiunital sesquialgebras are of *finite free type*. Thus we have shown:

Proposition 6.1. *There is a one to one correspondence between sesquiunital sesquialgebra structures of finite free type on k^G and unital ring structures on $\mathbb{Z}^G$ for which the structure constants and the components of the unit are nonnegative.*

The best known examples of such rings are the monoid algebras. If G is a monoid, then we may define δ^g_{hk} to be the characteristic function of the graph $g = hk$ of multiplication and e^g to be the characteristic function of the identity element. The corresponding sesquialgebra is just the modulation of the dual to the monoid bialgebra A'.

With this example in mind, we may think of a general structure of convolution type on $\mathbb{Z}^G$ as corresponding to a "product" operation on G in which the product of any two elements is a (possibly empty) subset of G whose elements are provided with positive integer "multiplicities". We will call such a subset a "multiple element"; the identity is also such a multiple element. (Of course, any ring structure may be viewed in this way, if we allow the multiplicities to be arbitrary integers).

To begin our analysis of these structures, we show that there are restricted possibilities for the unit.

Proposition 6.2. *Each e^g is either* 0 *or* 1.

Proof. Given g, by the counit property $\sum_k e^k d^g_{gk} = \delta_{gg} = 1$, we see that there is at least one $k \in G$ such that $d^g_{gk} \neq 0$. By the counit property again, we have

$$e^g \leq e^g d^g_{gk} \leq \sum_h e^h d^g_{hk} = \delta_{gk} \leq 1. \qquad \square$$

We will denote by G^0 the support of the unit, that is, the set of $g \in G$ for which $e^g = 1$. This set will play the role of identity elements in G.

As long as G is nonempty, so is G^0. In fact, we have the following:

Proposition 6.3. *Given any g in G, there are unique elements $l(g)$ and $r(g)$ in G^0 such that, for all $h \in G^0$, $d^k_{hg} = \delta_{hl(g)}\delta_{gk}$ and $d^k_{gh} = \delta_{r(g)h}\delta_{gk}$.*

Proof. This is again a straightforward corollary of the counit property. We obtain from $\sum_k e^k d^h_{gk} = \delta_{gh}$, $\sum_g e^g d^h_{gk} = \delta_{kh}$ that $\sum_{g \in G^0} d^k_{gh} = \delta_{kh}$. So $d^k_{gh} = 0$ when $k \neq h$ and there exists a unique element $g_0 \in G^0$ such that d^h_{gh} equals 1 for $g = g_0$ and 0 for all other g. We let $l(h)$ be this g_0. This proves the first equation; the second is proved by a similar argument. $\square$

Since the sum of the elements of G^0 is the unit of k^G, it is idempotent, from which it follows that k^{G_0} is a subalgebra. In fact, one may show:

Proposition 6.4. *The elements of G^0 form a set of orthogonal idempotents in $\mathbb{Z}^G$. In other words, the algebra structure on the subalgebra $\mathbb{Z}^{G^0}$ of A' is just pointwise multiplication.*

Proof. This follows from uniqueness in Proposition 6.3. □

Proposition 6.5. *For all g and h in G, if $d^k_{gh} \neq 0$, $l(k) = l(g)$ and $r(k) = r(h)$. If $r(g)$ is not equal to $l(h)$, then $gh = 0$ in G. In particular, $l(h) = h = r(h)$ for all $h \in G^0$.*

Proof. Coassociativity gives us

$$\sum_s d^k_{l(g)s} d^s_{gh} = \sum_s d^s_{l(g)g} d^k_{sh}.$$

By Proposition 6.3, $d^s_{l(g)g} = \delta_{gs}$. Therefore, the right-hand side of the equation is equal to $d^k_{gh} \neq 0$.

On the left-hand side, according to Proposition 6.3, $d^k_{l(g)s} \neq 0$ only when $l(s) = l(g)$ and $k = s$. Therefore, if $d^k_{gh} \neq 0$, then $d^k_{l(g)k} = 1$, so $l(k) = l(g)$. Similar arguments show that $r(k) = r(h)$.

If $r(g) \neq l(h)$, again by coassociativity, we have

$$d^k_{gh} = \sum_s d^s_{gr(g)} d^k_{sh} = \sum_s d^k_{gs} d^s_{r(g)h}.$$

According to Proposition 6.3, if $r(g) \neq l(h)$, $d^s_{r(g)h} = 0$ for all s; therefore, $d^k_{gh} = 0$. □

We now have retractions l and r from G onto G^0 which are like the target and source maps from a category to its set of identity elements. In fact, in terms of the multiplicative structure on G corresponding to the algebra structure on A', we have $l(g)g = gr(g) = g$; in particular, these products are single valued and without multiplicities. We might call G a "hypercategory". The composition of morphisms is a "multiple morphism" between two definite objects.

We will show next that, when k^G has an antipode and is hence a hopfish algebra, the underlying multiplicative structure on G has inverses and the property that gh is nonzero whenever $r(g) = l(h)$. We will call such a structure a "hypergroupoid" (see Definition 6.9).[2]

According to Definition 3.2, an antipode is a (k^G, k^G)-bimodule $\boldsymbol{S}$ whose dual is isomorphic to $\mathrm{Hom}_{k^G}(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$.

We recall the definition of $\boldsymbol{\epsilon}$ and $\boldsymbol{\Delta}$:

$$\boldsymbol{\epsilon} = \oplus_g \boldsymbol{\epsilon}^g, \quad \boldsymbol{\Delta} = \oplus_{g,h,t} \boldsymbol{\Delta}^t_{gh}.$$

[2]The notion of group with multiple-valued multiplication has a long history. The reader may start with [Kuntzmann 1939], which cites even earlier work.

Therefore we can write

$$\mathrm{Hom}_{k^G}(\boldsymbol{\epsilon}, \boldsymbol{\Delta}) = \big(\oplus_s \boldsymbol{\epsilon}^s\big) \otimes_{k^G} \big(\oplus_{g,h,t} \boldsymbol{\Delta}^t_{gh}\big)^*.$$

We notice that k^G acts via the upper indices of $\boldsymbol{\epsilon}$ and $\boldsymbol{\Delta}$ by componentwise multiplication. Therefore, this expression for $\mathrm{Hom}_{k^G}(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$ can be simplified to

$$\oplus_{g,h}\big(\oplus_t \boldsymbol{\epsilon}^{t*} \otimes \boldsymbol{\Delta}^t_{gh}\big)^*,$$

which is isomorphic to

$$\oplus_{g,h}\big(\oplus_t \mathrm{Hom}_k(k, \boldsymbol{\epsilon}^{t*} \otimes \boldsymbol{\Delta}^t_{gh})\big) \cong \mathrm{Hom}_k\big(k, \oplus_{g,h}(\oplus_t \boldsymbol{\epsilon}^{t*} \otimes \boldsymbol{\Delta}^t_{gh})\big).$$

Therefore, $\boldsymbol{S}$ is isomorphic to $\oplus_{g,h}(\oplus_t \boldsymbol{\epsilon}^{t*} \otimes \boldsymbol{\Delta}^t_{gh})$ as a (k^G, k^G) bimodule.

When $\boldsymbol{S}$ is an antipode, $\boldsymbol{S}$ is by definition isomorphic to k^G as a left k^G-module. Therefore, if we write $\boldsymbol{S}$ as $\oplus_{g,h}\boldsymbol{S}_{gh}$, for any fixed g there exists a unique element $h \in G$ such that $\dim(\boldsymbol{S}_{gh}) = 1$, and $\dim(\boldsymbol{S}_{gh'}) = 0$ for $h' \neq h$. Hence, we may define a map $\sigma : G \to G$ by $g \mapsto h$.

Since $\mathrm{Hom}_k(k, \boldsymbol{S}) \cong \mathrm{Hom}_{k^G}(\boldsymbol{\epsilon}, \boldsymbol{\Delta})$, we know that $\dim\big(\bigoplus_t \boldsymbol{\epsilon}^{t*} \otimes \boldsymbol{\Delta}^t_{gh}\big) = \delta_{\sigma(g)h}$, that is,

$$\sum_t e^t d^t_{gh} = \delta_{\sigma(g)h}. \tag{11}$$

Proposition 6.6. *For any $g \in G$, define g^{-1} to be $\sigma(g)$. Then there is a unique $s \in G^0$ such that*

$$d^s_{gg^{-1}} = 1.$$

In fact, $s = l(g) = r(g^{-1})$.

Proof. By (11), we have

$$\sum_t e^t d^t_{gg^{-1}} = \delta_{\sigma(g)g^{-1}} = 1.$$

Thus, there is a unique $s \in G^0$ such that $d^s_{gg^{-1}} = 1$ and $d^t_{gg^{-1}} = 0$ for all other $t \in G^0$. By Proposition 6.5, $s = l(g) = r(g^{-1})$. □

We also have another characteristic property of categories (though here we can only prove it in the presence of an antipode).

Proposition 6.7. *If $r(g) = l(h)$, there exists $s \in G$, such that $d^s_{gh} \neq 0$.*

Proof. Using coassociativity, we have

$$\sum_s d^g_{gs} d^s_{hh^{-1}} = \sum_s d^s_{gh} d^g_{sh^{-1}}.$$

Since $d^{l(h)}_{hh^{-1}} = 1$ and $d^g_{gl(h)} = 1$(since $r(g) = l(h)$), the left-hand side of the preceding equation is not equal to 0. This implies that, on the right-hand side, there is at least one term which is not equal to 0. Therefore, there exists $s \in G$, such that $d^s_{gh} \neq 0$. $\square$

Question 6.8. Is the inversion map $\sigma : G \to G$ involutive?

To summarize the arguments above, we define the combinatorial objects associated to hopfish structures on k^G:

Definition 6.9. A *hypergroupoid* is a set G with data $(\,\cdot\,, l, r, ^{-1}\,)$ as follows.

(1) There is a multivalued associative binary operation $\cdot$ on G. More precisely, for all $g, h \in G$, $g \cdot h$ is an element of $\mathbb{Z}_+^G$, where $\mathbb{Z}_+$ is the semiring of nonnegative integers. When this product is linearly extended to a product on $\mathbb{Z}_+^G$, we have $g \cdot (h \cdot k) = (g \cdot h) \cdot k$.

(2) There is a subset $G^0 \subset G$ with maps $l, r : G \to G^0$ such that $l(g) \cdot g = g \cdot r(g) = g$, for all $g \in G$ The product of g and h is nonzero if and only if $r(g) = l(h)$.

(3) There is an inverse operation $g \mapsto g^{-1}$ on G such that $g \cdot g^{-1}|_{G^0} = l(g)$ for all $g \in G$, and further $g \cdot h|_{G^0} = 0$ if $h \neq g^{-1}$.

Note that the inverse operation is determined by the product operation and G^0.

We now look at the commutative algebra k^G of k-valued functions on a hypergroupoid G. The duals of the maps $l, r : G \to G^0$ define morphisms from k^G to k^{G^0}. The multiplication on G defines a (nonunital) coproduct homomorphism $k^{G \times G} \leftarrow k^G$ whose modulation is a coproduct bimodule, the embedding map from G^0 to G makes k^{G^0} into a counit bimodule, and the inversion map defines an antipode. These define a hopfish algebra structure on k^G.

Theorem 6.10. *The hopfish structures of free finite type on k^G are in one to one correspondence with the hypergroupoid structures on G.*

Example 6.11. Here is an example of a hypergroupoid which is not a groupoid, based on [Etingof et al. 2005, Example 8.19].[3] Let $G = \{e, g\}$, with multiplication and inversion given by

$$eg = ge = g, \quad ee = e, \quad gg = e + ng, \quad e^{-1} = e, \quad g^{-1} = g,$$

where n is a nonnegative integer. $G^0 = \{e\}$ and $l(g) = r(g) = e$. The algebra A' associated to this hypergroupoid is $\mathbb{Z}[x]/\{x^2 = 1 + nx\}$. The corresponding hopfish algebra k^G is not a quasi-Hopf algebra when $n = 1$ and k a field. We explain the reason below.

The hopfish algebra structure of k^G is in fact a weak Hopf algebra, with $\epsilon(e) = 1$, $\epsilon(g) = 0$, $\Delta(e) = g \otimes g + e \otimes e$ and $\Delta(g) = e \otimes g + g \otimes e + g \otimes g$. Since a k^G module can be decomposed into submodules supported at points of G, the representation ring of k^G is generated by two elements 1 and X corresponding to

[3] The hypergroupoid itself, when $n = 1$, already appears as the first example in [Kuntzmann 1939]!

the 1-dimensional k^G module supported at e and g respectively. Since k is a field, 1 and X are just 1-dimensional k-vector spaces. Using the formulas for the coproduct and counit, it is easy to check that this representation ring is the Grothendieck ring of what is called Yang–Lee fusion rules in [Ostrik 2003], namely it is generated by 1 and X with the relation $X \otimes X = 1 \oplus X$. The Frobenius–Perron dimension of the element 1 is 1, while the Frobenius–Perron dimension of the element X is the irrational number $(1+\sqrt{5})/2$. According to Theorem 8.33 of [Etingof et al. 2005], the Frobenius–Perron dimension of any finite dimensional module over a finite dimensional quasi-Hopf algebra must be a positive integer, which is equal to the dimension of the module. This shows that k^G is not Morita equivalent to a quasi-Hopf algebra.

Acknowledgements

This work began in July, 2004, when Tang and Weinstein were participants in the trimester on K-Theory and Noncommutative Geometry at the Centre Émile Borel, and was carried to completion while Tang was affiliated with the University of California, Davis and Zhu with the Eidgenössische Technische Hochschule, Zürich. We had further opportunities for collaboration, and received further stimulation, during a number of other conferences and short-term visits.

We would like to thank Max Karoubi, Yvette Kosmann-Schwarzbach, Ryszard Nest, Tudor Ratiu, Pierre Schapira, and Michel van den Bergh for their invitations and hospitality. For useful comments in conversation and correspondence, we thank Paul Baum, Christian Blohmann, Micho Durdevic, Piotr Hajac, Yvette Kosmann-Schwarzbach, Olivier Mathieu, Ryszard Nest, Radu Popescu, Jean Renault, Earl Taft, and Boris Tsygan. We would especially like to thank Pavel Etingof for helping us to overcome a stumbling block in the characterization of the antipode, as well as Noah Snyder for providing the example in the last section.

References

[Blohmann et al. 2006] C. Blohmann, X. Tang, and A. Weinstein, "Hopfish structure and modules over irrational rotation algebras", 2006. To appear in proceedings of the CINVESTAV workshop on Deformation Quantization and Noncommutative Structures, Mexico City, 2005, to be published by the AMS in the Contemporary Mathematics series. math.QA/0604405

[Bursztyn and Weinstein 2005] H. Bursztyn and A. Weinstein, "Poisson geometry and Morita equivalence", pp. 1–78 in *Poisson geometry, deformation quantisation and group representations*, edited by S. Gutt et al., London Math. Soc. Lecture Note Ser. **323**, Cambridge Univ. Press, Cambridge, 2005. MR 2166451 Zbl 1075.53082

[Drinfel'd 1989] V. G. Drinfel'd, "Quasi-Hopf algebras", *Algebra i Analiz* **1**:6 (1989), 114–148. In Russian; translation in *Leningrad Math. J.* **1**:6, (1990), 1419–1457. MR 91b:17016 Zbl 0718.16033

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR 2006m:16051 Zbl 05042683

[Kuntzmann 1939] J. Kuntzmann, "Contribution à l'étude des systèmes multiformes", *Ann. Fac. Sci. Univ. Toulouse* (4) **3** (1939), 155–194. MR 8,439h Zbl 0024.29803

[Landsman 2001a] N. P. Landsman, "Bicategories of operator algebras and Poisson manifolds", pp. 271–286 in *Mathematical physics in mathematics and physics* (Siena, 2000), edited by R. Longo, Fields Inst. Commun. **30**, Amer. Math. Soc., Providence, RI, 2001. MR 2002h:46099 Zbl 1023.18009

[Landsman 2001b] N. P. Landsman, "Operator algebras and Poisson manifolds associated to groupoids", *Comm. Math. Phys.* **222**:1 (2001), 97–116. MR 2002f:46142 Zbl 1013.46060

[Lu 1993] J.-H. Lu, "Moment maps at the quantum level", *Comm. Math. Phys.* **157**:2 (1993), 389–404. MR 94m:58096 Zbl 0801.17019

[Lu 1996] J.-H. Lu, "Hopf algebroids and quantum groupoids", *Internat. J. Math.* **7**:1 (1996), 47–70. MR 97a:16073 Zbl 0884.17010

[Mrčun 1996] J. Mrčun, *Stability and invariants of Hilsum–Skandalis maps*, Ph.D. thesis, Utrecht University, 1996. math.DG/0506484

[Nikshych 2002] D. Nikshych, "On the structure of weak Hopf algebras", *Adv. Math.* **170**:2 (2002), 257–286. MR 2003f:16063 Zbl 1010.16041

[Ostrik 2003] V. Ostrik, "Fusion categories of rank 2", *Math. Res. Lett.* **10**:2-3 (2003), 177–183. MR 2004c:18015 Zbl 1040.18003

[Tseng and Zhu 2006] H.-H. Tseng and C. Zhu, "Integrating Lie algebroids via stacks", *Compos. Math.* **142**:1 (2006), 251–270. MR 2006h:58023 Zbl 05017588

Received November 2, 2005.

XIANG TANG
DEPARTMENT OF MATHEMATICS
WASHINGTON UNIVERSITY
ST. LOUIS, MO 63130
UNITED STATES
xtang@math.wustl.edu

ALAN WEINSTEIN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720
UNITED STATES
alanw@math.berkeley.edu

CHENCHANG ZHU
INSTITUT FOURIER
UNIVERSITÉ JOSEPH FOURIER GRENOBLE I
100, RUE DES MATHS - BP 74
38402 SAINT MARTIN D'HÈRES CEDEX
FRANCE
zhu@ujf-grenoble.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

GUTZMER'S FORMULA AND POISSON INTEGRALS ON THE HEISENBERG GROUP

SUNDARAM THANGAVELU

Dedicated to Alladi Sitaram on his sixtieth birthday

In 1978 M. Lassalle obtained an analogue of the Laurent series for holomorphic functions on the complexification of a compact symmetric space and proved a Plancherel type formula for such functions. In 2002 J. Faraut established such a formula, which he calls Gutzmer's formula, for all noncompact Riemannian symmetric spaces. This was immediately put into use by B. Krotz, G. Olafsson and R. Stanton to characterise the image of the heat kernel transform. In this article we prove an analogue of Gutzmer's formula for the Heisenberg motion group and use it to characterise Poisson integrals associated to the sublaplacian. We also use the Gutzmer's formula to study twisted Bergman spaces.

1. Introduction

Consider the Laplace–Beltrami operator on a Riemannian manifold M and the associated heat semigroup T_t. The problem of characterising the image of $L^2(M)$ under T_t has received considerable attention starting with Bargmann [1961]. He treated the case of $\mathbb{R}^n$ and showed that the image is a weighted Bergman space of entire functions. Similar results were obtained for compact Lie groups by Hall [1994] and for compact symmetric spaces by Stenzel [1999]. For the case of Hermite, Laguerre and Jacobi expansions see [Karp 2005] and the references there.

Contrary to the general expectation, such a characterisation is not true for the Laplace operator on the Heisenberg group $\mathbb{H}^n$, as shown by Krötz, Thangavelu and Xu in [Krötz et al. 2005b]. Specifically, we proved there that the image of $L^2(\mathbb{H}^n)$ is not a weighted Bergman space but a sum of two such spaces defined by signed weights.

Recently Krötz, Olafsson and Stanton [Krötz et al. 2005a] showed that when X is a noncompact Riemannian symmetric space the image of $L^2(X)$ cannot be

MSC2000: primary 43A85, 43A90; secondary 22E25, 22E30.
Keywords: Heisenberg group, sublaplacian, Fourier transform, Poisson integrals, Gutzmer's formula, Laguerre functions.

described as a weighted Bergman space. In the same paper a different characterisation of the image was obtained, using orbital integrals. In the case of noncompact Riemannian symmetric spaces the functions $T_t f$ do not extend as entire functions but only as holomorphic functions on a domain called the *complex crown*. The behaviour of $T_t f$ is therefore similar to that of Poisson integrals of L^2 functions on $\mathbb{R}^n$. Recall that the Poisson integrals extend only as holomorphic functions on a tube domain. The main ingredient used in [Krötz et al. 2005a] is a result of Faraut [2002] which he calls Gutzmer's formula. Our aim in this paper is to show that, using Gutzmer's formula, Poisson integrals can be characterised as certain spaces of holomorphic functions.

In Section 2 we treat the Poisson integrals on $\mathbb{R}^n$, where the results are easy to obtain. Section 3 recapitulates necessary results on special Hermite functions. In Section 4 we prove an analogue of Gutzmer's formula for the Heisenberg group, and we use this in Section 5 to give a characterisation of Poisson integrals on $\mathbb{H}^n$. Since Gutzmer's formula is available on all noncompact Riemannian symmetric spaces, a similar characterisation of Poisson integrals on them should be possible. In Section 6 we revisit twisted Bergman spaces and give a new proof of their characterisation as the image of $L^2(\mathbb{C}^n)$ under the special Hermite semigroup. (This was proved in [Krötz et al. 2005b] by a different method.)

2. Poisson integrals on Euclidean spaces

Throughout this paper x^2 stands for $|x|^2$, for $x \in \mathbb{R}^n$. Let

$$p_t(x) = c_n t (t^2 + x^2)^{-(n+1)/2}$$

be the Poisson kernel on $\mathbb{R}^n$, where c_n is a suitable constant. The Poisson integral of a function $f \in L^2(\mathbb{R}^n)$ is defined by

$$f * p_t(x) = \int_{\mathbb{R}^n} f(u) p_t(x-u)\, du,$$

which is also given in terms of the Fourier transform by

$$f * p_t(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{i x\cdot\xi} \hat{f}(\xi) e^{-t|\xi|}\, d\xi.$$

From any of these expressions it is clear that the function $F(x) = f * p_t(x)$ extends to the tube domain

$$\Omega_t = \{z = x + iy \in \mathbb{C}^n : |y| < t\}$$

as a holomorphic function. We are interested in knowing exactly what kind of holomorphic functions F arise as Poisson integrals.

To state our result, we recall the definition of the spherical function φ_λ on $\mathbb{R}^n$. For each $\lambda \in \mathbb{C}$ we define

$$\varphi_\lambda(x) = \int_{S^{n-1}} e^{-i\lambda x\cdot\omega}\, d\omega.$$

It is clear that φ_λ extends to $\mathbb{C}^n$ as an entire function. Moreover, when λ is real or purely imaginary it is given in terms of the Bessel function of order $n/2-1$. More precisely,

$$\varphi_\lambda(x) = c_n J_{n/2-1}(\lambda|x|)(\lambda|x|)^{-(n/2-1)}.$$

Using the Plancherel theorem for the Fourier transform, it is easy to see that for any r with $0 \le r < t$,

$$\int_{S^{n-1}} \int_{\mathbb{R}^n} |f * p_t(x+ir\omega)|^2\, dx\, d\omega = \int_{\mathbb{R}^n} |\hat{f}(\xi)|^2 e^{-2t|\xi|} \varphi_{2ir}(\xi)\, d\xi. \tag{2-1}$$

Following [Faraut 2002] we call this *Gutzmer's formula* for Euclidean spaces.

The right-hand side of the formula is finite even if $r = t$, as long as

$$\int_{\mathbb{R}^n} |\hat{f}(\xi)|^2 e^{-2t|\xi|} \varphi_{2it}(\xi)\, d\xi < \infty.$$

This happens precisely when

$$\int_{\mathbb{R}^n} |\hat{f}(\xi)|^2 (1+|\xi|^2)^{-(n-1)/2}\, d\xi < \infty$$

as can be seen using the asymptotic properties of the Bessel functions; see [Szegö 1967], for example. Recall that the Sobolev space $H^s(\mathbb{R}^n)$, for $s \in \mathbb{R}$, is the space of tempered distributions f for which

$$\|f\|_{(s)}^2 = \int_{\mathbb{R}^n} |\hat{f}(\xi)|^2 (1+|\xi|^2)^s\, d\xi$$

is finite. We have the following characterisation of Poisson integrals on $\mathbb{R}^n$.

Theorem 2.1. *A holomorphic function F on the tube domain Ω_t is the Poisson integral of a function $f \in H^{-(n-1)/2}(\mathbb{R}^n)$ if and only if*

$$\lim_{r\to t} \int_{S^{n-1}} \int_{\mathbb{R}^n} |F(x+ir\omega)|^2\, dx\, d\omega$$

is finite. Moreover, the limit is equivalent to the norm of $f \in H^{-(n-1)/2}(\mathbb{R}^n)$.

When $n=1$, this says that $F = f * p_t$ with $f \in L^2(\mathbb{R})$ if and only if both integrals $\int_{\mathbb{R}} |F(x+it)|^2\, dx$ and $\int_{\mathbb{R}} |F(x-it)|^2\, dx$ are finite. We are interested in finding an analogue of Theorem 2.1 for the Heisenberg group. This will be achieved via Gutzmer's formula for the Heisenberg motion group.

We can rewrite Gutzmer's formula (2-1) in terms of the group $M(n)$ of Euclidean motions, which is the semidirect product of $\mathrm{SO}(n)$ and $\mathbb{R}^n$. The action of $M(n)$ on $\mathbb{R}^n$ is given by $(\sigma, u)x = u + \sigma x$. This action has a natural extension to $\mathbb{C}^n$: simply define $(\sigma, u)(x + iy) = u + \sigma x + i\sigma y$. Gutzmer's formula then takes the following form.

Proposition 2.2. *Let F be holomorphic in Ω_t and let the restriction of F to $\mathbb{R}^n$ be such that $\hat{F}$ is square integrable with respect to the measure $\varphi_{2it}(\xi)\, d\xi$. Then*

$$\lim_{y^2 \to t^2} \int_{M(n)} |F(g.(x+iy))|^2 dg = \int_{\mathbb{R}^n} |\hat{F}(\xi)|^2 \varphi_{2it}(\xi)\, d\xi.$$

To see why this is true, we observe that

$$\int_{M(n)} |F(g.(x+iy))|^2 dg = \int_{\mathrm{SO}(n)} \int_{\mathbb{R}^n} |F(u + \sigma x + i\sigma y)|^2 du d\sigma.$$

Plancherel theorem for the Fourier transform shows that the $\mathbb{R}^n$ integral is

$$\int_{\mathbb{R}^n} |\hat{F}(\xi)|^2 e^{-2\sigma y \cdot \xi}\, d\xi.$$

The proposition follows since

$$\int_{\mathrm{SO}(n)} e^{-2\sigma y \cdot \xi}\, d\sigma = c_n \varphi_{2i|y|}(\xi).$$

In Section 4 we prove an analogue of this proposition for the Heisenberg motion group, which is then used in Section 5 to characterise Poisson integrals on the Heisenberg group.

3. Some results on special Hermite functions

We collect here relevant information about special Hermite functions and prove some estimates required in the next section. We closely follow the notations used in [Thangavelu 1998; 2004]; see those monographs for more details.

We will denote by Φ_α, $\alpha \in \mathbb{N}^n$, the Hermite functions on $\mathbb{R}^n$, normalised so that their L^2 norms are 1. On finite linear combinations of such functions we can define certain operators $\pi(z, w)$, where $z, w \in \mathbb{C}^n$, by setting

$$\pi(z, w)\Phi_\alpha(\xi) = e^{i(z\cdot\xi + (z/2)\cdot w)} \Phi_\alpha(\xi + w),$$

where $\cdot$ denotes the Euclidean inner product. Note that $\Phi_\alpha(\xi) = H_\alpha(\xi)e^{-|\xi|^2/2}$, where H_α is a polynomial on $\mathbb{R}^n$ and for $z \in \mathbb{C}^n$ we define $\Phi_\alpha(z)$ to be $H_\alpha(z)e^{-z^2/2}$, where $z^2 = z \cdot z$. The special Hermite functions $\Phi_{\alpha,\beta}(z, w)$ are then defined by

$$\Phi_{\alpha,\beta}(z, w) = (2\pi)^{-n/2}(\pi(z, w)\Phi_\alpha, \Phi_\beta). \tag{3-1}$$

The restrictions of $\Phi_{\alpha,\beta}(z, w)$ to $\mathbb{R}^n \times \mathbb{R}^n$ are usually called *special Hermite functions*. The family $\{\Phi_{\alpha,\beta}(x, u) : \alpha, \beta \in \mathbb{N}^n\}$ forms an orthonormal basis for $L^2(\mathbb{R}^n \times \mathbb{R}^n)$. The functions $\Phi_{\alpha,\beta}(z, w)$ are just the holomorphic extensions of $\Phi_{\alpha,\beta}(x, u)$ to $\mathbb{C}^n \times \mathbb{C}^n$. An easy calculation shows that

$$(\pi(z, w)\Phi_\alpha, \Phi_\beta) = (\Phi_\alpha, \pi(-\bar{z}, -\bar{w})\Phi_\beta). \tag{3-2}$$

This means that for $x, u \in \mathbb{R}^n$ the operators $\pi(x, u)$ are unitary on $L^2(\mathbb{R}^n)$ (related to the Schrödinger representation π_1 of the Heisenberg group) and

$$(\pi(ix, iu)\Phi_\alpha, \Phi_\beta) = (\Phi_\alpha, \pi(ix, iu)\Phi_\beta).$$

We can also verify that

$$\pi(z, w)\pi(z', w') = e^{(i/2)(z'\cdot w - z\cdot w')}\pi(z + z', w + w'). \tag{3-3}$$

For $x, u \in \mathbb{R}^n$ this gives

$$(\pi(2ix, 2iu)\Phi_\alpha, \Phi_\alpha) = \|\pi(ix, iu)\Phi_\alpha\|_2^2. \tag{3-4}$$

Let L_k^{n-1} be the Laguerre polynomials of type $n - 1$ and define the Laguerre functions φ_k by

$$\varphi_k(x, u) := L_k^{n-1}\big(\tfrac{1}{2}(x^2 + u^2)\big)e^{-(x^2+u^2)/4} = \sum_{|\alpha|=k} \Phi_{\alpha,\alpha}(x, u),$$

where the second equality is classical. The Laguerre functions have a natural holomorphic extension to $\mathbb{C}^n \times \mathbb{C}^n$, which we denote by the same symbol:

$$\varphi_k(z, w) = \sum_{|\alpha|=k} \Phi_{\alpha,\alpha}(z, w). \tag{3-5}$$

From this expression we obtain the following estimate for the complexified Laguerre functions.

Proposition 3.1. *For $z, w \in \mathbb{C}^n$ and $k \in \mathbb{N}$,*

$$|\varphi_k(z, w)|^2 \le C\frac{(k + n - 1)!}{k!(n-1)!}e^{(u\cdot y - v\cdot x)}\varphi_k(2iy, 2iv),$$

where $z = x + iy$ and $w = u + iv$.

Proof. We have

$$\Phi_{\alpha,\alpha}(x + iy, u + iv) = (2\pi)^{-n/2}(\pi(x + iy, u + iv)\Phi_\alpha, \Phi_\alpha)$$

and this gives in view of (3-3)

$$\Phi_{\alpha,\alpha}(x + iy, u + iv) = (2\pi)^{-n/2}e^{(u\cdot y - v\cdot x)/2}(\pi(iy, iv)\Phi_\alpha, \pi(-x, -u)\Phi_\alpha).$$

Since $\pi(-x,-u)$ is unitary we get the estimate

$$|\Phi_{\alpha,\alpha}(x+iy,u+iv)| \le (2\pi)^{-n/2} e^{(u\cdot y-v\cdot x)/2} \|\pi(iy,iv)\Phi_\alpha\|_2.$$

Applying the Cauchy–Schwarz inequality in (3-5), recalling that

$$\sum_{|\alpha|=k} 1 = \frac{(k+n-1)!}{k!(n-1)!},$$

and using (3-4) we get the proposition. □

The functions $\varphi_k(z,w)$ are members of the twisted Bergman spaces $\mathcal{B}_t$ studied in [Krötz et al. 2005b]. Since evaluations are continuous functionals on reproducing kernel Hilbert spaces of holomorphic functions and as $\mathcal{B}_t$ is one such space we get the estimate

$$|\varphi_k(z,w)| \le C_t \|\varphi_k\|_{\mathcal{B}_t} \le C_t \left(\frac{(k+n-1)!}{k!(n-1)!}\right)^{1/2} e^{(2k+n)t}.$$

However, we can greatly improve this estimate using the proposition above.

Proposition 3.2. *For each $r>0$ we have the uniform estimates*

$$|\varphi_k(z,w)| \le C_r e^{(u\cdot y-v\cdot x)/2} \left(\frac{(k+n-1)!}{k!(n-1)!}\right)^{1/2} k^{n/4-3/8} \exp\big(r(2k+n)^{1/2}\big),$$

valid for $y^2+v^2\le r^2$.

Proof. In view of Proposition 3.1 it is enough to estimate $\varphi_k(2iy,2iv)$. In the region $\delta\le y^2+v^2\le r^2$, with $\delta>0$, the required inequality follows using Perron's estimate [Szegö 1967, Theorem 8.22.3]:

$$L_k^\alpha(s) = \tfrac{1}{2}\pi^{-1/2} e^{s/2}(-s)^{-\alpha/2-1/4} k^{\alpha/2-1/4} e^{2(-ks)^{1/2}}\big(1+O(k^{-1/2})\big),$$

valid for s in the complex plane cut along the positive real axis (we require the estimate when $s<0$). We now use the representation

$$L_k^\alpha(s) = \frac{(-1)^k\pi^{-1/2}\Gamma(k+\alpha+1)}{\Gamma(\alpha+\frac{1}{2})(2k)!}\int_{-1}^{1}(1-t^2)^{\alpha-1/2} H_{2k}(s^{1/2}t)\,dt$$

for the Laguerre polynomials in terms of Hermite polynomials, along with the estimates given in [Szegö 1967, Theorem 8.22.6] for Hermite polynomials, to get the uniform estimates even when $\delta=0$. □

We need one more result on special Hermite functions. If $F(z,w)$ is a function on $\mathbb{C}^n\times\mathbb{C}^n$ we define its twisted translation by

$$\tau(z',w')F(z,w) = e^{-(i/2)(z'\cdot w-z\cdot w')} F(z-z',w-w'). \tag{3-6}$$

In view of (3-3) we have

$$\tau(z', w')\Phi_{\alpha,\alpha}(z, w) = (2\pi)^{-n/2}(\pi(z, w)\pi(-z', -w')\Phi_\alpha, \Phi_\alpha).$$

Using (3-2) this can be written as

$$\tau(z', w')\Phi_{\alpha,\alpha}(z, w) = (2\pi)^{-n/2}(\pi(-z', -w')\Phi_\alpha, \pi(-\bar{z}, -\bar{w})\Phi_\alpha).$$

Expanding $\pi(-z', -w')\Phi_\alpha$ and $\pi(-\bar{z}, -\bar{w})\Phi_\alpha$ in terms of Hermite functions we obtain

$$\tau(z', w')\Phi_{\alpha,\alpha}(z, w) = (2\pi)^{n/2}\sum_\beta \Phi_{\alpha,\beta}(-z', -w')\Phi_{\beta,\alpha}(z, w). \tag{3-7}$$

By taking $z = -z' = iy$ and $w = -w' = iv$ we obtain

$$\Phi_{\alpha,\alpha}(2iy, 2iv) = (2\pi)^{n/2}\sum_\beta |\Phi_{\alpha,\beta}(iy, iv)|^2, \tag{3-8}$$

since we have the relation

$$\Phi_{\beta,\alpha}(iy, iv) = (2\pi)^{-n/2}(\pi(iy, iv)\Phi_\beta, \Phi_\alpha) = \overline{\Phi_{\alpha,\beta}(iy, iv)}.$$

Proposition 3.3. *For $y, v \in \mathbb{R}^n$,*

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} |\varphi_k(x+iy, u+iv)|^2 e^{-(u\cdot y - v\cdot x)}\,dx\,du = (2\pi)^{n/2}\varphi_k(2iy, 2iv).$$

Proof. In view of (3-8) we have

$$\varphi_k(2iy, 2iv) = (2\pi)^{-n/2}\sum_{|\alpha|=k}\sum_\beta |\Phi_{\alpha,\beta}(iy, iv)|^2.$$

On the other hand from (3-6) we get

$$\Phi_{\alpha,\alpha}(x+iy, u+iv)e^{-(u\cdot y - v\cdot x)/2} = \tau(-iy, -iv)\Phi_{\alpha,\alpha}(x, u),$$

which gives, in view of (3-7),

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} |\tau(-iy, -iv)\Phi_{\alpha,\alpha}(x, u)|^2 dxdu = (2\pi)^n\sum_\beta |\Phi_{\alpha,\beta}(iy, iv)|^2.$$

Summing over all α with $|\alpha| = k$ we get the proposition. □

4. Gutzmer's formula on the Heisenberg group

Let $\mathbb{H}^n = \mathbb{C}^n \times \mathbb{R}$ be the Heisenberg group with multiplication defined by

$$(z, t)(z', t') = \big(z + z',\ t + t' + \tfrac{1}{2}\operatorname{Im}(z\cdot\bar{z}')\big).$$

More often we write (x, u, t) in place of (z, t) and the group law takes the form

$$(x, u, t)(x', u', t') = \big(x + x',\ u + u',\ t + t' + \tfrac{1}{2}(u \cdot x' - x \cdot u')\big),$$

where $x, u, x', u' \in \mathbb{R}^n$. For a function f on $\mathbb{H}^n$ we define

$$f^\lambda(z) = \int_{\mathbb{R}} f(z, t) e^{i\lambda t}\, dt.$$

For each nonzero $\lambda \in \mathbb{R}$ the Schrödinger representation π_λ of $\mathbb{H}^n$ is defined by

$$\pi_\lambda(x, u, t)\varphi(\xi) = e^{i\lambda t} e^{i\lambda(x\cdot\xi + x\cdot u/2)} \varphi(\xi + u),$$

where $\varphi \in L^2(\mathbb{R}^n)$. We define $\Phi_\alpha^\lambda(x) = |\lambda|^{n/4}\Phi(|\lambda|^{1/2}x)$ and

$$E_{\alpha,\beta}^\lambda(x, u, t) = (2\pi)^{-n/2}(\pi_\lambda(x, u, t)\Phi_\alpha^\lambda, \Phi_\beta^\lambda).$$

Note that

$$E_{\alpha,\beta}^\lambda(x, u, t) = e^{i\lambda t}\Phi_{\alpha,\beta}^\lambda(x, u),$$

where $\Phi_{\alpha,\beta}^\lambda(x, u) = E_{\alpha,\beta}^\lambda(x, u, 0)$. We write $\varphi_k^\lambda(x, u) = \varphi_k(|\lambda|^{1/2}(x, u))$, so in view of (3-5) we have

$$\varphi_k^\lambda(x, u) = \sum_{|\alpha|=k} \Phi_{\alpha,\alpha}^\lambda(x, u) = \varphi_k(|\lambda|^{1/2}(x, u))$$

and let $e_k^\lambda(x, u, t) = e^{i\lambda t}\varphi_k^\lambda(x, u)$. The results proved in the previous section are all valid for these functions for every nonzero $\lambda \in \mathbb{R}$.

The inversion formula for the group Fourier transform on $\mathbb{H}^n$ can be written in the form

$$f(x, u, t) = \int_{-\infty}^{\infty} \bigg(\sum_{k=0}^{\infty} f * e_k^\lambda(x, u, t) \bigg) d\mu(\lambda), \tag{4-1}$$

where $d\mu(\lambda) = (2\pi)^{-n-1}|\lambda|^n\, d\lambda$ is the Plancherel measure on the Heisenberg group. Define the λ-twisted convolution of two functions F and G on $\mathbb{C}^n$ by

$$F *_\lambda G(x, u) = \int_{\mathbb{R}^{2n}} F(x', u') G(x - x', u - u') e^{-(i/2)\lambda(u\cdot x' - x\cdot u')}\, dx'\, du'.$$

Then (4-1) takes the form

$$f(x, u, t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \bigg(\sum_{k=0}^{\infty} f^\lambda *_\lambda \varphi_k^\lambda(x, u) \bigg) d\mu(\lambda).$$

We would like to rewrite this inversion formula in terms of certain representations of the Heisenberg motion group, whose definition we recall. The unitary group $U(n)$ acts on the Heisenberg group as automorphisms, the action being defined by $\sigma(z,t)=(\sigma.z,t)$, where $\sigma\in U(n)$. The Heisenberg motion group G_n is the semidirect product of $U(n)$ and $\mathbb{H}^n$ with group law

$$(\sigma,z,t)(\tau,w,s)=(\sigma\tau,(z,t)(\sigma.w,s)).$$

Functions on $\mathbb{H}^n$ can be considered as right $U(n)$-invariant functions on G_n. Hence the inversion formula for such functions on G_n will involve only certain class-one representations of G_n. We now proceed to describe the relevant representations.

For each $k\in\mathbb{N}$ and nonzero $\lambda\in\mathbb{R}$ let $\mathcal{H}_k^\lambda$ be the Hilbert space for which the functions $E_{\alpha,\beta}^\lambda$, with $\alpha,\beta\in\mathbb{N}^n$, $|\alpha|=k$, form an orthonormal basis. The inner product in $\mathcal{H}_k^\lambda$ is defined by

$$(F,G)=|\lambda|^n\int_{\mathbb{C}^n}F(z,0)\overline{G(z,0)}\,dz.$$

On this Hilbert space we define a representation ρ_k^λ of the Heisenberg motion group by

$$\rho_k^\lambda(\sigma,z,t)F(w,s)=F((\sigma,z,t)^{-1}(w,s)).$$

Then it is known (from [Ratnakumar et al. 1997], for example) that ρ_k^λ is an irreducible unitary representation of G_n. As $(G_n,U(n))$ is a Gelfand pair, ρ_k^λ has a unique $U(n)$ fixed vector, which is none other than e_k^λ.

Given $f\in L^1(\mathbb{H}^n)$ we can define its group Fourier transform by

$$\rho_k^\lambda(f)=\int_{G_n}f(z,t)\rho_k^\lambda(\sigma,z,t)\,d\sigma\,dz\,dt,$$

which is a bounded operator acting on $\mathcal{H}_k^\lambda$. In this integral $d\sigma$ stands for the normalised Haar measure on $U(n)$. From calculations done in [Thangavelu 1998, Chapter 3] we infer that

$$\operatorname{tr}(\rho_k^\lambda(\sigma,z,t)^*\rho_k^\lambda(f))=\frac{k!(n-1)!}{(k+n-1)!}f*e_k^\lambda(z,t)$$

and the inversion formula for a right $U(n)$-invariant function on G_n takes the form

$$f(z,t)=(2\pi)^{-n-1}\int_{-\infty}^{\infty}\biggl(\sum_{k=0}^{\infty}\operatorname{tr}(\rho_k^\lambda(\sigma,z,t)^*\rho_k^\lambda(f))\frac{(k+n-1)!}{k!(n-1)!}\biggr)|\lambda|^n\,d\lambda.$$

Also the Plancherel theorem can be written as

$$\int_{\mathbb{H}^n}|f(z,t)|^2\,dz\,dt=(2\pi)^{-n-1}\int_{-\infty}^{\infty}\biggl(\sum_{k=0}^{\infty}\|\rho_k^\lambda(f)\|_{HS}^2\frac{(k+n-1)!}{k!(n-1)!}\biggr)|\lambda|^n\,d\lambda.$$

From this we can read off the Plancherel measure for G_n when dealing with right $U(n)$-invariant functions.

It can be shown that the operator $\rho_k^\lambda(f)$ is of rank one. Indeed, $\rho_k^\lambda(f)E_{\alpha,\beta}=0$ if $\alpha\neq\beta$ and

$$\rho_k^\lambda(f)E_{\alpha,\alpha}^\lambda=\frac{k!(n-1)!}{(k+n-1)!}f*e_k^\lambda.$$

From this we infer that $\rho_k^\lambda(f)e_k^\lambda=f*e_k^\lambda$. Hence it follows that for any $F\in\mathcal{H}_k^\lambda$ we have

$$\rho_k^\lambda(f)F=\frac{k!(n-1)!}{(k+n-1)!}(F,e_k^\lambda)\rho_k^\lambda(f)e_k^\lambda.$$

Proposition 4.1. *For every Schwartz class function f on $\mathbb{H}^n$ the inversion formula*

$$f(z,t)=(2\pi)^{n/2}\int_{-\infty}^{\infty}\left(\sum_{k=0}^{\infty}\Big(\rho_k^\lambda(f)e_k^\lambda,\,\rho_k^\lambda(1,z,t)\frac{k!(n-1)!}{(k+n-1)!}e_k^\lambda\Big)\frac{(k+n-1)!}{k!(n-1)!}\right)d\mu(\lambda)$$

holds, where 1 *stands for the identity matrix in* $U(n)$.

Proof. In view of the inversion formula (4-1) it is enough to show that

$$\big(\rho_k^\lambda(f)e_k^\lambda,\,\rho_k^\lambda(1,z,t)e_k^\lambda\big)=f*e_k^\lambda(z,t).$$

As $e_k^\lambda(z,t)=\sum_{|\alpha|=k}E_{\alpha,\alpha}^\lambda(z,t)$ we consider

$$\rho_k^\lambda(1,z,t)E_{\alpha,\alpha}^\lambda(w,s)=E_{\alpha,\alpha}^\lambda((-z,-t)(w,s)).$$

Recall that

$$E_{\alpha,\alpha}^\lambda((-z,-t)(w,s))=(2\pi)^{-n/2}(\pi_\lambda(-z,-t)\pi_\lambda(w,s)\Phi_\alpha^\lambda,\Phi_\alpha^\lambda).$$

Expanding $\pi_\lambda(w,s)\Phi_\alpha^\lambda$ and $\pi_\lambda(z,t))\Phi_\alpha^\lambda$ in terms of Φ_β^λ we get

$$\rho_k^\lambda(1,z,t)E_{\alpha,\alpha}^\lambda(w,s)=(2\pi)^{n/2}\sum_\beta E_{\alpha,\beta}^\lambda(w,s)\,\overline{E_{\alpha,\beta}^\lambda(z,t)}.$$

Since $\rho_k^\lambda(f)e_k^\lambda=f*e_k^\lambda$ and $\{E_{\alpha,\beta}^\lambda(w,s):\alpha,\beta\in\mathbb{N}^n,|\alpha|=k\}$ is an orthogonal basis for $\mathcal{H}_k^\lambda$ we get the proposition. □

From now on we identify $\mathbb{H}^n$ with $\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}$ and use the notation (x,u,t) rather than $(x+iu,t)$ to denote elements of $\mathbb{H}^n$. The action of $U(n)$ on $\mathbb{H}^n$ then takes the form $\sigma.(x,u,t)=(a.x-b.u,\,b.x+a.u,\,t)$, where a and b are the real and imaginary parts of σ. This action has a natural extension to $\mathbb{C}^n\times\mathbb{C}^n\times\mathbb{C}$ given by $\sigma.(z,w,\zeta)=(a.z-b.w,\,b.z+a.w,\,\zeta)$. With this definition we can extend the action of G_n on $\mathbb{H}^n$ to $\mathbb{C}^n\times\mathbb{C}^n\times\mathbb{C}$:

$$(a+ib,x',u',t')(z,w,\zeta)=(x',u',t')(a.z-b.w,b.z+a.w,\zeta).$$

This action is then extended to functions defined on $\mathbb{C}^n\times\mathbb{C}^n\times\mathbb{C}$:

$$\rho(g)f(z,w,\zeta)=f(g^{-1}.(z,w,\zeta)),\quad g\in G_n.$$

We are now ready to prove Gutzmer's formula for the Heisenberg group. Suppose f is a Schwartz class function on $\mathbb{H}^n$ such that $f^\lambda = 0$ for all $|\lambda| > A$ and $\rho_k^\lambda(f) = 0$ for all λ, k such that $(2k+n)|\lambda| > B$. We say that the Fourier transform of f is compactly supported if this condition is satisfied for some A and B. Now the inversion formula

$$f(g.(x,u,\xi)) = \int_{-A}^{A} \sum_{(2k+n)|\lambda|\le B} \big(\rho_k^\lambda(f)e_k^\lambda, \rho_k^\lambda(g)\rho_k^\lambda(1,x,u,\xi)e_k^\lambda\big)\, d\mu(\lambda)$$

is valid for any $g \in G_n$. Moreover, as each of $\rho_k^\lambda(1,x,u,\xi)e_k^\lambda$ extends to $\mathbb{C}^{2n+1}$ as an entire function, the same is true of $f(g.(x,u,\xi))$ and we have

$$f(g.(z,w,\zeta)) = \int_{-A}^{A} e^{\lambda\eta} \sum_{(2k+n)|\lambda|\le B} \big(\rho_k^\lambda(f)e_k^\lambda, \rho_k^\lambda(g)\rho_k^\lambda(1,x,u,\xi)e_k^\lambda\big)\, d\mu(\lambda),$$

where $\zeta = \xi + i\eta$. We then have the following Gutzmer's formula for the action of Heisenberg motion group on $\mathbb{C}^{2n+1}$, which is the complexification of $\mathbb{H}^n$.

Theorem 4.2. *Let f be Schwartz function whose Fourier transform is compactly supported in the sense above. Then f extends to $\mathbb{C}^{2n+1}$ as an entire function and we have the identity*

$$\int_{G_n} \big|f(g.(z,w,\zeta))\big|^2 dg =$$

$$(2\pi)^{n/2} \int_{-\infty}^{\infty} e^{2\lambda\eta} e^{-\lambda(u\cdot y - v\cdot x)} \bigg(\sum_{k=0}^{\infty} \|f^\lambda *_\lambda \varphi_k^\lambda\|_2^2 \frac{k!\,(n-1)!}{(k+n-1)!} \varphi_k^\lambda(2iy, 2iv) \bigg)\, d\mu(\lambda),$$

where $\|f^\lambda *_\lambda \varphi_k^\lambda\|_2$ is the $L^2(\mathbb{C}^n)$ norm of $f^\lambda *_\lambda \varphi_k^\lambda$.

We will prove this theorem by appealing to a general result on locally compact unimodular groups. Gutzmer's formula in the case of the circle group S^1 is just the Plancherel formula for the Fourier series applied to the Laurent series expansion of a function holomorphic in an annulus containing S^1. Lassalle [1978] has made an extensive study of Laurent series expansion for functions holomorphic in certain domains D contained in the complexification $X_{\mathbb{C}}$ of compact Riemannian symmetric spaces X. He obtained a Plancherel formula for such a series, which he later used in studying analogues of Hardy spaces over tube domains associated to compact symmetric spaces; see [Lassalle 1985]. Faraut [2002] treated the case of noncompact Riemannian symmetric spaces and proved a formula which he calls Gutzmer's formula. We recall the general setup for the reader's convenience.

Let G be a locally compact unimodular group with unitary dual $\hat{G}$. Let Λ be a Borel subset of $\hat{G}$ and let $d\mu$ be the Plancherel measure. For each $\lambda \in \Lambda$ choose a unitary representation (π_λ, V_λ) of class λ. Let $A(\lambda)$ be a family of trace class operators for which $\int_\Lambda \operatorname{tr}(|A(\lambda)|)\,d\mu$ and $\int_\Lambda \operatorname{tr}(A(\lambda)^*A(\lambda))\,d\mu$ are finite. Define a function f on G by

$$f(g) = \int_\Lambda \operatorname{tr}(A(\lambda)\pi_\lambda(g^{-1}))\,d\mu.$$

Then f lies in $L^2(G)$, and by the Plancherel Theorem,

$$\int_G |f(g)|^2\,dg = \int_\Lambda \operatorname{tr}(A(\lambda)^*A(\lambda))\,d\mu.$$

This sets up a one-to-one correspondence between subspaces of $L^2(G)$ whose Fourier transforms are supported on Λ and families $(A(\lambda))$ of Hilbert–Schmidt operators satisfying

$$\int_\Lambda \operatorname{tr}(A(\lambda)^*A(\lambda))\,d\mu < \infty.$$

The isometry just described takes a particularly simple form when $(A(\lambda))$ is a family of rank-one operators. Let $A(\lambda)v = (v, \eta(\lambda))\xi(\lambda)$, where $v \in V_\lambda$ and $\eta(\lambda), \xi(\lambda)$ are measurable functions taking values in V_λ. We then have the following result, whose proof can be found in [Faraut 2002; 2003].

Proposition 4.3. *Assume that $\eta(\lambda)$ and $\xi(\lambda)$ satisfy*

$$\int_\Lambda \|\eta(\lambda)\|^2\|\xi(\lambda)\|^2\,d\mu < \infty.$$

Then the function f defined by $f(g) = \int_\Lambda (\pi_\lambda(g^{-1})\xi(\lambda), \eta(\lambda))\,d\mu$ belongs to $L^2(G)$ and satisfies

$$\int_G |f(g)|^2\,dg = \int_\Lambda \|\eta(\lambda)\|^2\|\xi(\lambda)\|^2\,d\mu.$$

Proof of Theorem 4.2. Take $G = G_n$ and $\Lambda = \mathbb{R}\times\mathbb{N}$. The relevant representations are ρ_k^λ acting on the Hilbert spaces $\mathcal{H}_k^\lambda$. As already seen, the operators $\rho_k^\lambda(f)$ have rank one. We take $\xi(\gamma) := \rho_k^\lambda(f)e_k^\lambda$ and

$$\eta(\gamma) := \frac{k!\,(n-1)!}{(k+n-1)!}\rho_k^\lambda(1, z, w, \zeta)e_k^\lambda$$

when $\gamma = (\lambda, k) \in \Lambda$. (The first factor on the right is used because its reciprocal appears in the Plancherel measure for G_n.) We wish to appeal to Proposition 4.3 to complete the proof of Theorem 4.2.

We are therefore left with proving the equality

$$\|\rho_k^\lambda(1, z, w, \zeta)e_k^\lambda\|^2 = e^{2\lambda\eta}e^{-\lambda(u\cdot y - v\cdot x)}\varphi_k^\lambda(2iy, 2iv).$$

Recall that $e_k^\lambda(x', u', t') = e^{i\lambda t'}\varphi_k^\lambda(x, u)$ and this can be extended to $\mathbb{C}^n\times\mathbb{C}^n\times\mathbb{C}$ as a holomorphic function. The action of $\rho_k^\lambda(1, x, u, t)$ on e_k^λ is given by

$$\rho_k^\lambda(1, x, u, t)e_k^\lambda(x', u', t') = e_k^\lambda((x, u, t)^{-1}(x', u', t'))$$

which reduces to

$$e^{i\lambda(t'-t-(u\cdot x'-x\cdot u')/2)}\varphi_k^\lambda(x'-x,u'-u).$$

The holomorphic extension of this is given by

$$\rho_k^\lambda(1,z,w,\zeta)e_k^\lambda(x',u',t')=e^{i\lambda(t'-\zeta-(w\cdot x'-z\cdot u')/2)}\varphi_k^\lambda(x'-z,u'-w).$$

In terms of real and imaginary parts of $z=x+iy$, $w=u+iv$, and $\zeta=\xi+i\eta$, this becomes

$$\rho_k^\lambda(1,z,w,\zeta)e_k^\lambda(x',u',t')=e^{\lambda\eta}e^{i\lambda(t'-\xi-(u\cdot x'-x\cdot u')/2)}e^{-\lambda(u'\cdot y-x'\cdot v)/2}\varphi_k^\lambda(x'-z,u'-w).$$

From the definition of the Hilbert space $\mathcal{H}_k^\lambda$, the norm of $\rho_k^\lambda(1,z,w,\zeta)e_k^\lambda$ in $\mathcal{H}_k^\lambda$ is

$$\|\rho_k^\lambda(1,z,w,\zeta)e_k^\lambda\|^2=|\lambda|^n e^{2\lambda\eta}\int_{\mathbb{R}^n\times\mathbb{R}^n}e^{-\lambda(u'\cdot y-x'\cdot v)}|\varphi_k^\lambda(x'-z,u'-w)|^2\,dx'\,du'.$$

Without loss of generality we can assume $\lambda>0$. By a change of variables the integral reduces to

$$e^{-\lambda(u\cdot y-x\cdot v)}\int_{\mathbb{R}^n\times\mathbb{R}^n}e^{-\lambda^{1/2}(u'\cdot y-x'\cdot v)}\left|\varphi_k(x'-i\lambda^{1/2}y,u'-i\lambda^{1/2}v)\right|^2dx'\,du'.$$

Using Proposition 3.3 we see that the integral is equal to $e^{-\lambda(u\cdot y-x\cdot v)}\varphi_k^\lambda(2iy,2iv)$, as required. □

Remark. We have stated Gutzmer's formula for functions whose Fourier transforms are compactly supported. This condition is not necessary for the validity of the formula. Consider the inversion formula stated in Proposition 4.1, namely

$$f(z,w,\zeta)=(2\pi)^{n/2}\int_{-\infty}^{\infty}\left(\sum_{k=0}^{\infty}\left(\rho_k^\lambda(f)e_k^\lambda,\ \rho_k^\lambda(1,z,w,\zeta)e_k^\lambda\right)\right)d\mu(\lambda).$$

In view of the calculations made above the series converges as long as

$$\sum_{k=0}^{\infty}\|f^\lambda*_\lambda\varphi_k^\lambda\|_2\,e^{-\lambda(u\cdot y-v\cdot x)/2}(\varphi_k^\lambda(2iy,2iv))^{1/2}<\infty.$$

If we assume that $f^\lambda(x,u)$ is compactly supported in λ and the norms $\|f^\lambda*_\lambda\varphi_k^\lambda\|_2$ have enough decay, then the inversion formula is valid and $f(z,w,\zeta)$ is holomorphic. For example, when f belongs to the image of $L^2(\mathbb{H}^n)$ under the heat semigroup associated to the full Laplacian then $\|f^\lambda*_\lambda\varphi_k^\lambda\|_2\le Ce^{-t(2k+n)|\lambda|}$ and f extends as an entire function (see [Krötz et al. 2005b]). For such functions Gutzmer's formula is valid, as can be proved by using a density argument. We refer to [Faraut 2003] for some details in the case of noncompact Riemannian symmetric spaces.

5. Poisson integrals on the Heisenberg group

On the Heisenberg group we consider the sublaplacian $\mathscr{L}$ defined by

$$\mathscr{L} = -\sum_{j=1}^{n}(X_j^2 + Y_j^2),$$

where $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ together with $T = \partial/\partial t$ form a basis for the Heisenberg Lie algebra. (See [Thangavelu 1998; 2004] for explicit expressions for these vector fields.) The operator $\mathscr{L}$ is nonnegative, so using the spectral theorem we can define the Poisson semigroup $e^{-a(\mathscr{L})^{1/2}}$, $a > 0$. This is explicitly given by the spectral representation

$$e^{-a(\mathscr{L})^{1/2}} f(x, u, t) = \int_{-\infty}^{\infty} e^{-i\lambda t}\left(\sum_{k=0}^{\infty} e^{-(2k+n)^{1/2}|\lambda|^{1/2}a} f^{\lambda} *_{\lambda} \varphi_k^{\lambda}(x, u)\right) d\mu(\lambda).$$

We may suppose that f is a Schwartz class function whose Fourier transform is compactly supported in the sense defined in the previous section. For such functions the series above converges pointwise. We denote $\exp(-a(\mathscr{L})^{1/2}) f$ by $P_a f$ and call it the *Poisson integral of f*. For each $r > 0$, define

$$\Omega_r = \left\{(x + iy, u + iv, \zeta) \in \mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C} : y^2 + v^2 < r^2\right\}.$$

Theorem 5.1. *Let f be a Schwartz class function on $\mathbb{H}^n$ and assume that $f^{\lambda}(x, u)$ is compactly supported in $(-b, b)$ as a function of λ. Then for $0 < r < a$, we can extend $P_a f(x, u, t)$ to Ω_r as a holomorphic function of (z, w, ζ), and*

$$\int_{G_n} |P_a f(g.(z, w, \zeta))|^2 dg = c_n \times$$
$$\int_{-b}^{b} e^{2\lambda\eta} e^{-\lambda(u\cdot y - v\cdot x)}\left(\sum_{k=0}^{\infty} \| f^{\lambda} *_{\lambda} \varphi_k^{\lambda}\|_2^2 e^{-2(2k+n)^{1/2}|\lambda|^{1/2}a} \frac{k!(n-1)!}{(k+n-1)!}\varphi_k^{\lambda}(2iy, 2iv)\right) d\mu(\lambda).$$

Proof. Once we show that $P_a f$ extends as a holomorphic function of (z, w, ζ) on Ω_r the theorem will follow from Gutzmer's formula. Consider now the expansion

$$P_a f(x, u, t) = (2\pi)^{-n-1}\int_{-\infty}^{\infty} e^{-i\lambda t}\left(\sum_{k=0}^{\infty} f^{\lambda} *_{\lambda} \varphi_k^{\lambda}(x, u) e^{-(2k+n)^{1/2}|\lambda|^{1/2}a}\right)|\lambda|^n d\lambda$$

and define the Poisson kernel by

$$P_a^{\lambda}(x, u) = \sum_{k=0}^{\infty} \varphi_k^{\lambda}(x, u) e^{-(2k+n)^{1/2}|\lambda|^{1/2}a}.$$

Then we have

$$P_a f(x,u,t) = \int_{\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}} e^{-i\lambda t} f^\lambda(x',u') P_a^\lambda(x-x',u-u') e^{-i(\lambda/2)(x\cdot u'-u\cdot x')}\,dx'\,du'\,d\lambda.$$

It is therefore enough to show that $P_a^\lambda(x,u)$ extends as a holomorphic function of (z,w). Fix λ. In view of Proposition 3.1, the series

$$\sum_{k=0}^{\infty} \varphi_k^\lambda(x+iy,u+iv) e^{-(2k+n)^{1/2}|\lambda|^{1/2}a}$$

is bounded by a constant times

$$\left(\sum_{k=0}^{\infty}\left(\frac{(k+n-1)!}{k!(n-1)!}\right)^{1/2} (\varphi_k^\lambda(2iy,2iv))^{1/2} e^{-(2k+n)^{1/2}|\lambda|^{1/2}a}\right) e^{|\lambda|(u\cdot y-v\cdot x)/2}.$$

For any fixed (y,v) with $y^2+v^2\le r^2<a^2$, Perron's formula gives the estimate

$$(\varphi_k^\lambda(2iy,2iv))^{1/2} \le C_r k^{(n-1)/4-1/8} e^{(2k+n)^{1/2}|\lambda|^{1/2}r}$$

and hence the series is dominated by

$$\sum_{k=1}^{\infty} k^{(n-1)/2} k^{(n-1)/4-1/8} e^{-(2k+n)^{1/2}|\lambda|^{1/2}(a-r)}$$

which certainly converges as long as $r<a$.

Moreover, using the asymptotic estimates given by Perron's formula the integral

$$\int_{-\infty}^{\infty} (\varphi_k^\lambda(2iy,2iv))^{1/2} e^{-(2k+n)^{1/2}|\lambda|^{1/2}a} |\lambda|^n\,d\lambda$$

is bounded by a constant multiple of

$$k^{(n-1)/4-1/8}\int_0^\infty e^{-(2k+n)^{1/2}t(a-r)} t^{(6n+9)/4-1}\,dt,$$

so for $y^2+v^2\le r^2$ and $|\lambda|\le b$, the sum

$$\sum_{k=0}^{\infty}\int_{-\infty}^{\infty} |\varphi_k^\lambda(x+iy,u+iv)| e^{-(2k+n)^{1/2}|\lambda|^{1/2}a} e^{|\lambda|(u\cdot y-v\cdot x)/2} |\lambda|^n\,d\lambda$$

is dominated by a constant times

$$e^{b|u\cdot y-v\cdot x|}\sum_{k=1}^{\infty} k^{(n-1)/2} k^{(n-1)/4-1/8} k^{-(6n+9)/8},$$

which is finite. Hence standard arguments show that $P_a^\lambda(x,u)$ extends to Ω_r as a holomorphic function. $\square$

We now use Theorem 5.1 to get a characterisation of functions that arise as Poisson integrals. Let $\mathcal{G}(\Omega_a)$ be the space of functions $F(z,w,\zeta)$ holomorphic on Ω_a and such that $F^\lambda(x,u)$ is compactly supported in λ and

$$\|F\|_{\Omega_a}^2 = \lim_{y^2+v^2\to a^2} \int_{G_n} |F(g.(iy,iv,t))|^2\,dg < \infty.$$

We define $\mathcal{B}(\Omega_a)$ to be its completion. For $F \in \mathcal{G}(\Omega_a)$, Gutzmer's formula shows that the restriction of F to $\mathbb{H}^n$ satisfies

$$\int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \left(\|F^\lambda *_\lambda \varphi_k^\lambda\|_2^2 \frac{k!(n-1)!}{(k+n-1)!} \varphi_k^\lambda(2iy,2iv) \right) |\lambda|^n\,d\lambda < \infty$$

whenever $y^2+v^2 \le a^2$.

We also need to recall the definition of the Sobolev spaces $H^s(\mathbb{H}^n)$. This is the space of all tempered distributions for which $(1+\mathcal{L})^{s/2} f \in L^2(\mathbb{H}^n)$. The norm is given by the expression

$$\|f\|_{(s)}^2 = c_n \int_{-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \|f^\lambda *_\lambda \varphi_k^\lambda\|_2^2 \left((2k+n)|\lambda|\right)^s \right) |\lambda|^n\,d\lambda.$$

The asymptotic formula we have used for $\varphi_k^\lambda(2iy,2iv)$ reads as

$$\frac{k!(n-1)!}{(k+n-1)!} \varphi_k^\lambda(2iy,2iv) \le C((2k+n)|\lambda|)^{-(2n-1)/4} e^{-2(2k+n)^{1/2}|\lambda|^{1/2} r},$$

and therefore f lies in $H^{-(2n-1)/4}(\mathbb{H}^n)$ precisely when

$$\int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \left(\|f^\lambda *_\lambda \varphi_k^\lambda\|_2^2 \frac{k!(n-1)!}{(k+n-1)!} \varphi_k^\lambda(2iy,2iv) \right) |\lambda|^n\,d\lambda < \infty$$

for all y, v with $y^2+v^2 \le a^2$.

Theorem 5.2. *A function F lies in $\mathcal{B}(\Omega_a)$ if and only if $F = P_a f$ for some $f \in H^{-(2n-1)/4}(\mathbb{H}^n)$. The norm of f in $H^{-(2n-1)/4}(\mathbb{H}^n)$ is equivalent to*

$$\lim_{y^2+v^2\to a^2} \int_{G_n} |F(g.(iy,iv,t))|^2\,dg.$$

Proof. If $f \in H^{-(2n-1)/4}(\mathbb{H}^n)$ then f can be approximated by a sequence f_n of functions whose Fourier transforms in the central variable are compactly supported. For such functions we have verified that $P_a f_n$ extends to a function in $\mathcal{G}(\Omega_a)$. This proves half the theorem. Since all the steps in our calculations are reversible, the converse also follows in light of the remarks preceding the theorem. $\square$

6. Revisiting twisted Bergman spaces

Consider the full Laplacian $\Delta = \mathscr{L} - T^2$ on the Heisenberg group. In [Krötz et al. 2005b] we studied the problem of characterising the image of $L^2(\mathbb{H}^n)$ under the semigroup $e^{-t\Delta}$ as a weighted Bergman space over the domain $\mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}$, which is just the complexification of $\mathbb{H}^n$. It turned out that the image is not a weighted Bergman space, contrary to expectation. What is true is that the image is a direct integral of certain weighted Bergman spaces $\mathscr{B}_t^\lambda$, which are the images of $L^2(\mathbb{C}^n)$ under the semigroup e^{-tL_λ} generated by the special Hermite operators L_λ, which are related to $\mathscr{L}$. In this section we give a different proof of this characterisation of $\mathscr{B}_t^\lambda$ using Gutzmer's formula on the Heisenberg group.

We briefly recall the definitions of L_λ and the twisted Bergman spaces $\mathscr{B}_t^\lambda$, referring to [Krötz et al. 2005b] for more details. For each $\lambda \neq 0$, the operator L_λ is defined by $\mathscr{L}(e^{-i\lambda t} f(z)) = e^{-i\lambda t} L_\lambda f(z)$. The spectral decomposition of L_λ is given by

$$L_\lambda f(x, u) = (2\pi)^{-n} \sum_{k=0}^{\infty} (2k+n)|\lambda| f *_\lambda \varphi_k^\lambda(x, u).$$

The operator L_λ generates a diffusion semigroup e^{-tL_λ} given by twisted convolution with the kernel

$$p_t^\lambda(x, u) = c_n \left(\frac{\lambda}{\sinh(\lambda t)} \right)^n e^{-\lambda \coth(\lambda t)(x^2+u^2)/4}.$$

For each $f \in L^2(\mathbb{C}^n)$ the function

$$f *_\lambda p_t^\lambda(x, u) = \int_{\mathbb{R}^{2n}} f(x - x', u - u') p_t^\lambda(x', u') e^{i/2\lambda(u \cdot x' - x \cdot u')} \, dx' \, du'$$

can be extended to $\mathbb{C}^n \times \mathbb{C}^n$ as an entire function. We let

$$W_t^\lambda(x + iy, u + iv) = 4^n e^{\lambda(u \cdot y - v \cdot x)} p_{2t}^\lambda(2y, 2v)$$

and define $\mathscr{B}_t^\lambda$ to be the space of all entire functions on $\mathbb{C}^n \times \mathbb{C}^n$ that are square integrable with respect to the weight function $W_t^\lambda(z, w)$.

Theorem 6.1 [Krötz et al. 2005b]. *An entire function F on $\mathbb{C}^n \times \mathbb{C}^n$ belongs to $\mathscr{B}_t^\lambda$ if and only if $F(x, u) = f *_\lambda p_t^\lambda(x, u)$ for some $f \in L^2(\mathbb{C}^n)$.*

In this section we give a different and more transparent proof that $f *_\lambda p_t^\lambda(z, w)$ belongs to $\mathscr{B}_t^\lambda$, based on Gutzmer's formula for special Hermite expansions. In proving this we assume $\lambda = 1$ and simply write p_t in place of p_t^1.

Consider the reduced Heisenberg group (or Heisenberg group with compact center) $\mathbb{H}_{\mathrm{red}}^n$ defined to be $\mathbb{C}^n \times S^1$ with group law

$$(z, e^{it})(w, e^{is}) = (z + w, e^{i(t+s+\frac{1}{2}\operatorname{Im}(z \cdot \bar{w}))}).$$

The infinite-dimensional irreducible unitary representations of $\mathbb{H}^n_{\text{red}}$ that are nontrivial at the center are given (up to unitary equivalence) by the Schrödinger representations π_j, where now j is a nonzero integer. The case $j = 0$ corresponds to the one-dimensional representations that factor through characters of $\mathbb{C}^n$. For functions f on $\mathbb{H}^n_{\text{red}}$ having mean value zero, i.e., $\int_{S^1} f(z, e^{it})\, dt = 0$, the relevant representations are just π_j with $j \neq 0$. Let G_n^{red} be the Heisenberg motion group formed with $\mathbb{H}^n_{\text{red}}$ in place of $\mathbb{H}^n$. Then with the same notations as in Section 4 we have the following result. Let Ω_r be defined as before.

Theorem 6.2. *Let f be a function on $\mathbb{H}^n_{\text{red}}$ having mean value zero and satisfying the conditions stated in Theorem 4.2. For all $(z, w) \in \Omega_r$, we have*

$$\int_{G_n^{\text{red}}} |f(g.(iy, iv, e^{it}))|^2 dg = (2\pi)^{-n-1} \sum_{j\neq 0} \left(\sum_{k=0}^{\infty} \| f^j *_j \varphi_k^j \|_2^2 \frac{k!(n-1)!}{(k+n-1)!} \varphi_k^j(2iy, 2iv) \right) |j|^n.$$

When we take functions of the form $f(x, u)e^{-it}$, exactly one term (corresponding to $j = 1$) survives in this sum over j (since f^j is the $(-j)$-th Fourier coefficient of f), and we get

$$\int_{\mathbb{R}^{2n}} |f(x+iy, u+iv)|^2 e^{(u\cdot y - v\cdot x)} dx\, du = c_n \sum_{k=0}^{\infty} \| f \times \varphi_k \|_2^2 \frac{k!(n-1)!}{(k+n-1)!} \varphi_k(2iy, 2iv),$$

which we refer to as *Gutzmer's formula for special Hermite expansions*. Here $f \times \varphi_k$ is the twisted convolution, which is just the λ-twisted convolution when $\lambda = 1$. The equality is valid for all functions f for which the right-hand side converges. This is so if the norms of the projections $f \times \varphi_k$ decay fast enough. In particular, the preceding formula is valid if f is replaced by $e^{-tL} f$ with f in $L^2(\mathbb{C}^n)$.

Applying Gutzmer's formula to the function $F(z, w) = e^{-tL} f(z, w)$ we obtain

$$\int_{\mathbb{R}^{2n}} |F(x+iy, u+iv)|^2 e^{(u\cdot y - v\cdot x)} dx\, du = c_n \sum_{k=0}^{\infty} \| f \times \varphi_k \|_2^2 e^{-2(2k+n)t} \frac{k!(n-1)!}{(k+n-1)!} \varphi_k(2iy, 2iv).$$

If we can show that

$$\int_{\mathbb{R}^{2n}} \varphi_k(2iy, 2iv) p_{2t}(2y, 2v)\, dy\, dv = \frac{(k+n-1)!}{k!(n-1)!} e^{2(2k+n)t},$$

we can integrate Gutzmer's formula against $p_{2t}(2y, 2v)\, dy\, dv$ to get

$$\int_{\mathbb{C}^{2n}} |F(z, w)|^2 W_t^1(z, w)\, dz\, dw = c_n \int_{\mathbb{R}^{2n}} |f(x, u)|^2 dx\, du,$$

which will prove our claim and hence Theorem 6.1. So it remains to prove the following lemma.

Lemma 6.3. $\displaystyle\int_{\mathbb{R}^{2n}} \varphi_k(iy, iv)p_t(y, v)\,dy\,dv = \frac{(k+n-1)!}{k!(n-1)!}e^{(2k+n)t}.$

Before proving the lemma we make some remarks. Since the heat kernel p_t is given by the expansion

$$p_t(y, v) = (2\pi)^{-n}\sum_{k=0}^{\infty} e^{-(2k+n)t}\varphi_k(y, v)$$

it follows, in view of the orthogonality properties of φ_k, that

$$\int_{\mathbb{R}^{2n}} p_t(y, v)\varphi_k(y, v)\,dy\,dv = \frac{(k+n-1)!}{k!(n-1)!}e^{-(2k+n)t}. \tag{6-1}$$

As φ_k are the spherical functions associated to the Gelfand pair $(G_n, U(n))$, the formula in the lemma is the analogue of the formula

$$\int_{\mathbb{R}^n} \varphi_\lambda(iy)e^{-y^2/(4t)}\,dy = c_n e^{t\lambda^2},$$

where φ_λ are the spherical functions on $\mathbb{R}^n$, namely, the Bessel functions. This was the key formula used in characterising Bergman spaces associated to the Laplacian on $\mathbb{R}^n$.

Proof of the lemma. Recall from [Szegö 1967] that

$$L_k^{n-1}(s) = \sum_{j=0}^{k} c_{k,j}(-s)^j,$$

where the $c_{k,j}$ are constants whose exact values are immaterial. Equation (6-1) now reads as

$$(\sinh t)^{-n}\int_{\mathbb{R}^{2n}}\sum_{j=0}^{k} c_{k,j}(-1)^j 2^{-j}(y^2+v^2)^j e^{-(1+\coth t)(y^2+v^2)/4}\,dy\,dv = (2\pi)^n\frac{(k+n-1)!}{k!(n-1)!}e^{-(2k+n)t}.$$

This can be rewritten as

$$(\cosh t)^{-n}\sum_{j=0}^{k} c_{k,j}(-1)^j 2^{-j}a_j(\tanh t)^j(1+\tanh t)^{-j-n} = (2\pi)^n\frac{(k+n-1)!}{k!(n-1)!}e^{-(2k+n)t}, \tag{6-2}$$

where the a_j are constants defined by

$$a_j = \int_{\mathbb{R}^{2n}} (y^2+v^2)^j e^{-(y^2+v^2)/4}\,dy\,dv.$$

Now both sides of (6-2) are holomorphic functions of t in a strip containing the real line, so the equation is true for t negative as well. This leads to

$$(\cosh t)^{-n}\sum_{j=0}^{k} c_{k,j}2^{-j}a_j(\tanh t)^j(1-\tanh t)^{-j-n}=(2\pi)^n\frac{(k+n-1)!}{k!\,(n-1)!}e^{(2k+n)t}.$$

The left-hand side now is simply

$$(\sinh t)^{-n}\int_{\mathbb{R}^{2n}}\sum_{j=0}^{k} c_{k,j}2^{-j}(y^2+v^2)^j e^{-(-1+\coth t)(y^2+v^2)/4}\,dy\,dv,$$

which is the same as

$$(2\pi)^n\int_{\mathbb{R}^{2n}}\varphi_k(iy,iv)p_t(y,v)\,dy\,dv.$$

This completes the proof of the lemma and hence of Theorem 6.1. □

Acknowledgements

We are immensely thankful to one of the referees for pointing out a missing factor in the main theorem. He made a very careful study of the paper and pointed out several inaccuracies. The revised version owes a lot to both referees for pointing out several typos and demanding details wherever our arguments were brief.

References

[Bargmann 1961] V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform", *Comm. Pure Appl. Math.* **14** (1961), 187–214. MR 28 #486 Zbl 0107.09102

[Faraut 2002] J. Faraut, "Formule de Gutzmer pour la complexification d'un espace riemannien symétrique", *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **13**:3-4 (2002), 233–241. MR 2004k:43015 Zbl 05019686

[Faraut 2003] J. Faraut, "Analysis on the crown of a Riemannian symmetric space", pp. 99–110 in *Lie groups and symmetric spaces*, Amer. Math. Soc. Transl. Ser. 2 **210**, Amer. Math. Soc., Providence, RI, 2003. MR 2004k:43016 Zbl 1045.43014

[Hall 1994] B. C. Hall, "The Segal-Bargmann "coherent state" transform for compact Lie groups", *J. Funct. Anal.* **122**:1 (1994), 103–151. MR 95e:22020 Zbl 0838.22004

[Karp 2005] D. Karp, "Square summability with geometric weight for classical orthogonal expansions", pp. 407–421 in *Advances in analysis*, edited by H. G. W. Begehr et al., World Sci. Publ., Hackensack, NJ, 2005. MR MR2184956 Zbl 05018246

[Krötz et al. 2005a] B. Krötz, G. Ólafsson, and R. J. Stanton, "The image of the heat kernel transform on Riemannian symmetric spaces of the noncompact type", *Int. Math. Res. Not.* 22 (2005), 1307–1329. MR 2006e:58036 Zbl 1079.22007

[Krötz et al. 2005b] B. Krötz, S. Thangavelu, and Y. Xu, "The heat kernel transform for the Heisenberg group", *J. Funct. Anal.* **225**:2 (2005), 301–336. MR 2006b:22008 Zbl 1075.22005

[Lassalle 1978] M. Lassalle, "Séries de Laurent des fonctions holomorphes dans la complexification d'un espace symétrique compact", *Ann. Sci. École Norm. Sup.* (4) **11**:2 (1978), 167–210. MR 80f:22006 Zbl 0452.43011

[Lassalle 1985] M. Lassalle, "L'espace de Hardy d'un domaine de Reinhardt généralisé", *J. Funct. Anal.* **60**:3 (1985), 309–340. MR 86g:32049 Zbl 0578.32009

[Ratnakumar et al. 1997] P. K. Ratnakumar, R. Rawat, and S. Thangavelu, "A restriction theorem for the Heisenberg motion group", *Studia Math.* **126**:1 (1997), 1–12. MR 98m:42056 Zbl 0978.42016

[Stenzel 1999] M. B. Stenzel, "The Segal-Bargmann transform on a symmetric space of compact type", *J. Funct. Anal.* **165**:1 (1999), 44–58. MR 2000j:58053 Zbl 0929.22007

[Szegö 1967] G. Szegö, *Orthogonal polynomials*, 3rd ed., Colloq. Publ. **23**, Amer. Math. Soc., Providence, RI, 1967. Zbl 0305.42011

[Thangavelu 1998] S. Thangavelu, *Harmonic analysis on the Heisenberg group*, Progress in Mathematics **159**, Birkhäuser, Boston, 1998. MR 99h:43001 Zbl 0892.43001

[Thangavelu 2004] S. Thangavelu, *An introduction to the uncertainty principle: Hardy's theorem on Lie groups*, Progress in Mathematics **217**, Birkhäuser, Boston, 2004. MR 2004j:43007 Zbl 02039097

Received March 3, 2006. Revised June 4, 2006.

SUNDARAM THANGAVELU
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560 012
INDIA
veluma@math.iisc.ernet.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 231, No. 1, 2007

EQUIVARIANT NIELSEN INVARIANTS FOR DISCRETE GROUPS

JULIA WEBER

For discrete groups G, we introduce equivariant Nielsen invariants. They are equivariant analogs of the Nielsen number and give lower bounds for the number of fixed point orbits in the G-homotopy class of an equivariant endomorphism $f : X \to X$. Under mild hypotheses, these lower bounds are sharp.

We use the equivariant Nielsen invariants to show that a G-equivariant endomorphism f is G-homotopic to a fixed point free G-map if the generalized equivariant Lefschetz invariant $\lambda_G(f)$ is zero. Finally, we prove a converse of the equivariant Lefschetz fixed point theorem.

1. Introduction

The Lefschetz number is a classical invariant in algebraic topology. If the Lefschetz number $L(f)$ of an endomorphism $f : X \to X$ of a compact CW-complex is nonzero, then f has a fixed point. This is the famous Lefschetz fixed point theorem. The converse does not hold: If the Lefschetz number of f is zero, we cannot conclude f to be fixed point free.

A more refined invariant which allows to state the converse is the Nielsen number: The Nielsen number $N(f)$ is zero if and only if f is homotopic to a fixed point free map. More generally, the Nielsen number is used to give precise minimal bounds for the number of fixed points of maps homotopic to f. Its development was started by Nielsen [1920]; a comprehensive treatment can be found in [Jiang 1983].

We are interested in the equivariant generalization of these results. Given a discrete group G and a G-equivariant endomorphism $f : X \to X$ of a finite proper G-CW-complex, we introduce equivariant Nielsen invariants called $N_G(f)$ and $N^G(f)$. They are equivariant analogs of the Nielsen number and are derived from the generalized equivariant Lefschetz invariant $\lambda_G(f)$ [Weber 2005, Definition 5.13].

MSC2000: primary 55M20, 57R91; secondary 54H25, 57S99.
Keywords: Nielsen number, discrete groups, equivariant, Lefschetz fixed point theorem.

We proceed to show that these Nielsen invariants give minimal bounds for the number of orbits of fixed points in the G-homotopy class of f. One even obtains results concerning the type and "location" (connected component of the relevant fixed point set) of these fixed point orbits. These lower bounds are sharp if X is a cocompact proper smooth G-manifold satisfying the standard gap hypotheses.

Finally, we prove a converse of the equivariant Lefschetz fixed point theorem: If X is a G-Jiang space as defined in Definition 5.2, then $L_G(f) = 0$ implies that f is G-homotopic to a fixed point free map. Here, $L_G(f)$ is the equivariant Lefschetz class [Lück and Rosenberg 2003a, Definition 3.6], the equivariant analog of the Lefschetz number.

These results were motivated by work of Lück and Rosenberg [2003a; 2003b]. For G a discrete group and an endomorphism f of a cocompact proper smooth G-manifold M, they prove an equivariant Lefschetz fixed point theorem [2003a, Theorem 0.2]. The converse of that theorem is proven here.

Another motivation for the present article is the fact that the algebraic approach to the equivariant Reidemeister trace provides a good framework for computation. The connection to the machinery used in the study of transformation groups [Lück 1989; tom Dieck 1987] allows results to translate more readily from transformation groups to geometric equivariant topology and vice-versa.

When G is a compact Lie group, Wong [1993] obtains results on equivariant Nielsen numbers which strongly influenced us. The main difference between our work and Wong's is that we treat possibly infinite discrete groups. Another difference is that our approach is more structural. We can read off the equivariant Nielsen invariants from the generalized equivariant Lefschetz invariant $\lambda_G(f)$.

In case G is a finite group, Ferrario [2003] studies a collection of generalized Lefschetz numbers which can be thought of as an equivariant generalized Lefschetz number. In contrast to the generalized equivariant Lefschetz invariant $\lambda_G(f)$ these do not incorporate the WH-action on the fixed point set X^H. A generalized Lefschetz trace for equivariant maps has also been defined by Wong, as mentioned in [Hart 1999].

For compact Lie groups, earlier definitions of generalized Lefschetz numbers for equivariant maps were made in [Wilczyński 1984; Fadell and Wong 1988]. These authors used the collection of generalized Lefschetz numbers of the maps f^H, for $H < G$. In general, these numbers are not sufficient since they do not take the equivariance into account adequately. For further reading on equivariant fixed point theory, see [Ferrario 2005], where an extensive list of references is given.

Organization of paper. In Section 2, we introduce the generalized equivariant Lefschetz invariant. We briefly assemble the concepts and definitions which are needed for the definition of equivariant Nielsen invariants in Section 3.

The equivariant Nielsen invariants give lower bounds for the number of fixed point orbits of f. This is shown in Section 4. The standard gap hypotheses are introduced, and it is shown that under these hypotheses these lower bounds are sharp.

In the nonequivariant case, we know that the generalized Lefschetz invariant is the right element when looking for a precise count of fixed points. We read off the Nielsen number from this invariant. In general, the Lefschetz number contains too little information.

But under certain conditions, we can conclude facts about the Nielsen number from the Lefschetz number directly. These are called the Jiang conditions [1983, Definition II.4.1] (see also [Brown 1971, Chapter VII]). In Section 5, we introduce the equivariant version of these conditions. We also give examples of G-Jiang spaces.

In Section 6, we derive equivariant analogs of statements about Nielsen numbers found in [Jiang 1983], generalizing results from [Wong 1993] to infinite discrete groups. In particular, if X is a G-Jiang space, the converse of the equivariant Lefschetz fixed point theorem holds.

2. The generalized equivariant Lefschetz invariant

Classically, the Nielsen number is defined geometrically by counting essential fixed point classes [Brown 1971, Chapter VI; Jiang 1983, Definition I.4.1]. Alternatively one defines it using the generalized Lefschetz invariant.

Let X be a finite CW-complex, let $f : X \to X$ be an endomorphism, let x be a basepoint of X, and let

$$\lambda(f) = \sum_{\overline{\alpha} \in \pi_1(X,x)_\phi} n_{\overline{\alpha}} \cdot \overline{\alpha} \in \mathbb{Z}\pi_1(X,x)_\phi$$

be the *generalized Lefschetz invariant* associated to f [Reidemeister 1936; Wecken 1941], where

$$\mathbb{Z}\pi_1(X,x)_\phi := \mathbb{Z}\pi_1(X,x)/\phi(\gamma)\alpha\gamma^{-1} \sim \alpha, \text{ with } \gamma, \alpha \in \pi_1(X,x).$$

Here ϕ is the map induced by f on the fundamental group $\pi_1(X,x)$. We have $\phi(\gamma) = wf(\gamma)w^{-1}$, where w is a path from x to $f(x)$. The generalized Lefschetz invariant is also called *Reidemeister trace* in the literature, which goes back to the original name “Reidemeistersche Spureninvariante” used by Wecken.

The set $\pi_1(X,x)_\phi$ is often denoted by $\mathcal{R}(f)$ and called set of *Reidemeister classes* of f. Since we will introduce a variation of this set in Definition 2.3, we prefer to stick with the notation used in [Weber 2005].

Definition 2.1. The *Nielsen number* of f is defined by

$$N(f) := \#\{\overline{\alpha} \mid n_{\overline{\alpha}} \neq 0\}.$$

The Nielsen number is the number of classes in $\pi_1(X, x)_\phi$ with nonzero coefficients. A class α with nonzero coefficient corresponds to an essential fixed point class in the geometric sense.

In the equivariant setting, the fundamental category replaces the fundamental group. The fundamental category of a topological space X with an action of a discrete group G is defined as follows [Lück 1989, Definition 8.15].

Definition 2.2. Let G be a discrete group, and let X be a G-space. Then the *fundamental category* $\Pi(G, X)$ is the following category:

- The objects $\mathrm{Ob}\big(\Pi(G, X)\big)$ are G-maps $x : G/H \to X$, where the $H \leq G$ are subgroups.
- The morphisms $\mathrm{Mor}\big(x(H), y(K)\big)$ are pairs $(\sigma, [w])$, where
  - σ is a G-map $\sigma : G/H \to G/K$
  - $[w]$ is a homotopy class of G-maps $w : G/H \times I \to X$ relative $G/H \times \partial I$ such that $w_1 = x$ and $w_0 = y \circ \sigma$.

The fundamental category is a combination of the orbit category of G and the fundamental groupoid of X. If X is a point, then the fundamental category is just the orbit category of G, whereas when G is the trivial group, the definition reduces to the definition of the fundamental groupoid of X.

We often view x as the point $x(1H)$ in the fixed point set X^H. We call $X^H(x)$ the connected component of X^H containing $x(1H)$. We also consider the relative fixed point set, the pair $\big(X^H(x), X^{>H}(x)\big)$. Here $X^{>H}(x) = \{z \in X^H(x) \mid G_z \neq H\}$ is the singular set, where G_z denotes the isotropy group of z. In order to simplify notation, we use $f^H(x)$ to denote $f|_{X^H(x)}$, and we use $f_H(x)$ instead of $f|_{\left(X^H(x), X^{>H}(x)\right)}$.

Fixed points of f can only exist in $X^H(x)$ when $X^H(f(x)) = X^H(x)$, i.e, when the points $f(x)$ and x lie in the same connected component of X^H.

Definition 2.3. For $x \in \mathrm{Ob}\,\Pi(G, X)$ with $X^H(f(x)) = X^H(x)$ and a morphism $v = (\mathrm{id}, [w]) \in \mathrm{Mor}(f(x), x)$, set

$$\mathbb{Z}\pi_1\big(X^H(x), x\big)_{\phi'} := \mathbb{Z}\pi_1\big(X^H(x), x\big)/\phi(\gamma)\alpha\gamma^{-1} \sim \alpha,$$

where $\alpha \in \pi_1(X^H(x), x)$, $\gamma \in \mathrm{Aut}(x)$ and $\phi(\gamma) = v\phi(\gamma)v^{-1} \in \mathrm{Aut}(x)$.

The automorphism group $\mathrm{Aut}(x)$ of the object x in the category $\Pi(G, X)$ is a group extension lying in the short exact sequence

$$1 \to \pi_1(X^H(x), x) \to \mathrm{Aut}(x) \to WH_x \to 1.$$

Here $WH := N_G H/H$ is the Weyl group of H, it acts on X^H. We call WH_x the subgroup of WH which fixes the connected component $X^H(x)$.

Groups obtained from different choices of the path w and of the point x in its isomorphism class $\bar{x}$ are canonically isomorphic, so those choices do not play a role. The group $\mathbb{Z}\pi_1\big(X^H(x), x\big)_{\phi'}$ generalizes the group $\mathbb{Z}\pi_1(X, x)_\phi$ defined above. So it can be seen as the free abelian group generated by equivariant Reidemeister classes of $f^H(x)$ with respect to the action of WH_x on $X^H(x)$.

A map $(\sigma, [w]) \in \operatorname{Mor}(x, y)$ induces a group homomorphism

$$(\sigma, [w])^* : \mathbb{Z}\pi_1\big(X^K(y), y\big)_{\phi'} \to \mathbb{Z}\pi_1\big(X^H(x), x\big)_{\phi'}$$

by twisted conjugation, and we know that the induced group homomorphism is the same for every map in $\operatorname{Mor}(x, y)$ [Weber 2005, Lemma 5.2].

The *generalized equivariant Lefschetz invariant* [Weber 2005, Definition 5.13], $\lambda_G(f)$, is an element in the group

$$\Lambda_G(X, f) := \bigoplus_{\substack{\bar{x} \in \operatorname{Is}\Pi(G,X),\\ X^H(f(x)) = X^H(x)}} \mathbb{Z}\pi_1(X^H(x), x)_{\phi'}.$$

Here $\operatorname{Is}\Pi(G, X)$ denotes the set of isomorphism classes of the category $\Pi(G, X)$. Geometrically, it corresponds to the set of WH-orbits of connected components $X^H(x)$ of the fixed point sets X^H, for $(H) \in \operatorname{consub}(G)$, i.e., for a set of representatives of conjugacy classes of subgroups of G. There is a bijection $\operatorname{Is}\Pi(G, X) \xrightarrow{\simeq} \coprod_{(H)\in\operatorname{consub}(G)} WH \setminus \pi_0(X^H)$ which sends $x : G/H \to X$ to the orbit under the WH-action on $\pi_0(X^H)$ of the component $X^H(x)$ of X^H which contains the point $x(1H)$ [Lück and Rosenberg 2003a, Equation 3.3].

Let $\widetilde{f^H(x)}$ and $\widetilde{f^{>H}(x)}$ denote the lift of $f^H(x)$ to the universal covering space $\widetilde{X^H(x)}$ and to the subset $\widetilde{X^{>H}(x)} \subseteq \widetilde{X^H(x)}$ that projects to $X^{>H}(x)$ under the covering map.

At the summand indexed by $\bar{x}$, the generalized equivariant Lefschetz invariant is given by

$$\lambda_G(f)_{\bar{x}} := L^{\mathbb{Z}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}, \widetilde{f^{>H}(x)}\big) \in \mathbb{Z}\pi_1(X^H(x), x)_{\phi'},$$

where the *refined equivariant Lefschetz number* [Weber 2005, Definition 5.7] appears on the right hand side. It is defined by

$$L^{\mathbb{Z}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}, \widetilde{f^{>H}(x)}\big) := \sum_{p\geq 0} (-1)^p \operatorname{tr}_{\mathbb{Z}\operatorname{Aut}(x)}(C^c_p(\widetilde{f^H(x)}, \widetilde{f^{>H}(x)})),$$

where the trace map $\operatorname{tr}_{\mathbb{Z}\operatorname{Aut}(x)}$ [Weber 2005, Definition 5.4] is induced by the projection $\mathbb{Z}\operatorname{Aut}(x) \to \mathbb{Z}\pi_1(X^H(x), x)_{\phi'}$, $\sum_{g\in\operatorname{Aut}(x)} r_g \cdot g \mapsto \sum_{g\in\pi_1(X^H(x),x)} r_g \cdot \bar{g}$. Instead of $\mathbb{Z}$, other rings can be used.

This trace map generalizes the trace map used in [Lück and Rosenberg 2003a], and the refined equivariant Lefschetz number is a generalization of the orbifold Lefschetz number (Definition 1.4 of that reference).

The refined equivariant Lefschetz number $L^{\mathbb{Q}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}\big)$ will be particularly important to us, so we give some formulas describing it. For a finite proper G-CW-complex X we have [Weber 2005, Lemma 5.9]

$$L^{\mathbb{Q}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}\big)=\sum_{p\geq 0}(-1)^p\sum_{G\cdot e\in G\backslash I_p(X)}|G_e|^{-1}\cdot\operatorname{inc}_\phi(f,e)\in\mathbb{Q}\pi_1(X^H(x),x)_{\phi'}.$$

Here $I_p(X)$ denotes the set of p-cells of X, e runs through the equivariant cells of X, and G_e is its isotropy group. The *refined incidence number* [Weber 2005, Definition 5.8] $\operatorname{inc}_\phi(f,e)\in\mathbb{Z}\pi_1(X^H(x),x)_{\phi'}$ for a p-cell $e\in I_p(X)$ is defined to be the “degree” of the composition

$$\begin{aligned}\bar{e}/\partial e\xrightarrow{i_e}\bigvee_{e'\in I_p(X)}\overline{e'}/\partial e'\xrightarrow{h^{\sim}}X_p/X_{p-1}\xrightarrow{f}X_p/X_{p-1}\xrightarrow{h^{-1}\sim}\bigvee_{e'\in I_p(X)}\overline{e'}/\partial e'\\ \xrightarrow{\operatorname{pr}_{\pi\cdot\bar{e}/\partial e}}\pi\cdot\bar{e}/\partial e\xrightarrow{\bar{\cdot}}\pi_{\phi'}\cdot\bar{e}/\partial e.\end{aligned}$$

Here $\bar{e}$ is the closure of the open p-cell e and $\partial e=\bar{e}\backslash e$. The map i_e is the inclusion, h is a homeomorphism and $\operatorname{pr}_{\pi\cdot\bar{e}/\partial e}$ is the projection.

If $X=M$ is a cocompact proper G-manifold, we have [Weber 2005, Theorem 6.6]

$$L^{\mathbb{Q}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}\big)=\sum_{\substack{WH_x\cdot z\in\\ WH_x\backslash\operatorname{Fix}(f^H(x))}}\big|(WH_x)_z\big|^{-1}\deg\Big(\big(\operatorname{id}_{T_zM^H(x)}-T_z(f^H(x))\big)^c\Big)\cdot\overline{\alpha_z}.$$

Here the map on the tangent space is extended to the one-point compactification $\big(T_zM^H(x)\big)^c$. The relative versions of these formulas also hold.

We have $L^{\mathbb{Q}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}\big)=\operatorname{ch}_G(X,f)\big(\lambda_G(f)\big)_{\bar{x}}$ [Weber 2005, Lemma 6.4], where $\operatorname{ch}_G(X,f):\Lambda_G(X,f)\to\bigoplus_{\bar{y}\in\operatorname{Is}\Pi(G,X)}\mathbb{Q}\pi_1(X^K(y),y)_{\phi'}$ is the character map [Weber 2005, Definition 6.2]. So we can derive $L^{\mathbb{Q}\operatorname{Aut}(x)}\big(\widetilde{f^H(x)}\big)$ from $\lambda_G(f)$.

The equivariant analog of the Lefschetz number is the *equivariant Lefschetz class* $L_G(f)\in\bigoplus_{\bar{x}\in\operatorname{Is}\Pi(G,X),X^H(f(x))=X^H(x)}\mathbb{Z}$, whose value at $\bar{x}$ is

$$L_G(f)_{\bar{x}}=L^{\mathbb{Z}WH_x}\big(f^H(x),f^{>H}(x)\big)$$

[Lück and Rosenberg 2003a, Definition 3.6]. The projection of $\pi_1(X^H(x),x)_{\phi'}$ to the trivial group $\{1\}$ induces an augmentation map sending $\lambda_G(f)$ to $L_G(f)$:

$$s:\bigoplus_{\substack{\bar{x}\in\operatorname{Is}\Pi(G,X),\\ X^H(f(x))=X^H(x)}}\mathbb{Z}\pi_1(X^H(x),x)_{\phi'}\;\to\;\bigoplus_{\substack{\bar{x}\in\operatorname{Is}\Pi(G,X),\\ X^H(f(x))=X^H(x)}}\mathbb{Z}.$$

3. Equivariant Nielsen invariants

Given an element $\sum_{\overline{\alpha}} n_{\overline{\alpha}} \cdot \overline{\alpha} \in \mathbb{Z}\pi_1(X^H(x), x)_{\phi'}$, we call a class $\overline{\alpha} \in \pi_1(X^H(x), x)_{\phi'}$ *essential* if the coefficient $n_{\overline{\alpha}}$ is nonzero.

Let G be a discrete group and let X be a cocompact proper smooth G-manifold. Let $f : X \to X$ be a smooth G-equivariant map such that $\operatorname{Fix}(f) \cap \partial X = \varnothing$ and such that for every $z \in \operatorname{Fix}(f)$ the determinant of the map $(\mathrm{id}_{T_z X} - T_z f)$ is different from zero. One can always find a representative in the G-homotopy class of f which satisfies this assumption. Since the generalized equivariant Lefschetz invariant is G-homotopy invariant, we can replace f by this representative if necessary.

Definition 3.1. The *equivariant Nielsen class* of f is

$$\nu_G(f) = \sum_{Gz \in G\backslash \operatorname{Fix}(f)} \frac{\det\bigl(\mathrm{id}_{T_z X} - T_z(f)\bigr)}{\bigl|\det\bigl(\mathrm{id}_{T_z X} - T_z(f)\bigr)\bigr|} \cdot \overline{\alpha_z}.$$

Here $\alpha_z \in \pi_1(X^{G_z}(z), x)$ is the loop given by $[t * f(t)^{-1} * w]$, where x is a basepoint in $X^{G_z}(z)$, t is a path from x to z and w is a path from $f(x)$ to x. The basepoint x may differ from z, e.g., if we have more than one fixed point in a connected component of X^{G_z}. If $x = z$, we may choose t and w to be constant. The equivalence relation assures that this definition is independent of the choices involved.

We can also derive the equivariant Nielsen class $\nu_G(f)$ from the generalized equivariant Lefschetz invariant $\lambda_G(f)$.

Lemma 3.2. *The invariant $\nu(f)$ is the image of the generalized equivariant Lefschetz invariant $\lambda(f)$ under the quotient map where we divide out the images of nonisomorphisms*:

$$\nu_G(f) = \overline{\lambda_G(f)} \in \bigoplus_{\substack{\bar{x} \in \operatorname{Is}\Pi(G,X),\\ X^H(f(x)) = X^H(x)}} \mathbb{Z}\pi_1(X^H(x), x)_{\phi'} / \{\operatorname{Im}(\sigma, [w])^* \mid \sigma \textit{ nonisom.}\}$$

Proof. We consider the equation obtained in the refined equivariant Lefschetz fixed point theorem [Weber 2005, Theorem 0.2]. We have

$$\lambda_G(f) = \sum_{Gz \in G\backslash \operatorname{Fix}(f)} \Lambda_G(z, f) \circ \mathrm{ind}_{G_z \subseteq G}(\mathrm{Deg}_0^{G_z}((\mathrm{id}_{T_z X} - T_z f)^c)).$$

Here $\mathrm{Deg}_0^{G_z}$ is the equivariant degree [Lück and Rosenberg 2003a], it has values in the Burnside ring $A(G_z)$. On basis elements $[G_z/L] \in A(G_z)$, the map $\Lambda_G(z, f) \circ \mathrm{ind}_{G_z \subseteq G}$ is given by

$$\Lambda_G(z, f) \circ \mathrm{ind}_{G_z \subseteq G}([G_z/L]) = (\mathrm{pr}, [\mathrm{cst}])^* \overline{\alpha_z},$$

where cst denotes the constant map and

$$(\mathrm{pr}, [\mathrm{cst}])^* : \mathbb{Z}\pi_1(X^{G_z}(z), x)_{\phi'} \to \mathbb{Z}\pi_1(X^L(z \circ \mathrm{pr}), x \circ \mathrm{pr})_{\phi'}$$

is the map induced by the projection $\mathrm{pr} : G_z/L \to G_z/G_z$.

We know that $\mathrm{Deg}_0^{G_z}((\mathrm{id}_{T_z X} - T_z f)^c)$ is a unit of the Burnside ring $A(G_z)$ since $\big(\mathrm{Deg}_0^{G_z}((\mathrm{id}_{T_z X} - T_z f)^c)\big)^2 = 1$ [Lück and Rosenberg 2003a, Example 4.7]. In general a unit of the Burnside ring $A(G_z)$ may consist of more than one summand [tom Dieck 1979]. The summand $[G_z/G_z]$ is always included with a coefficient $+1$ or -1, but there might be summands $[G_z/L]$ for $L < G_z$ appearing. So one fixed point might give more than one class with nonzero coefficients.

If we divide out the images of nonisomorphisms, then we divide out the image of $(\mathrm{pr}, [\mathrm{cst}])^*$ for all $L \neq G_z$. We are left with the summand $\pm\overline{\alpha_z}$ coming from $\pm 1[G_z/G_z]$. This cannot lie in the image of any nonisomorphism. So each fixed point leads to exactly one summand. The sign is the sign of the determinant $\det\big(\mathrm{id}_{T_z X} - T_z(f)\big)$, so the claim follows. □

We set

$$\mathbb{Z}\pi_1(X^H(x), x)_{\phi''} := \mathbb{Z}\pi_1(X^H(x), x)_{\phi'}/\{\mathrm{Im}(\sigma, [w])^* \mid \sigma \text{ nonisom.}\}.$$

We use the equation established in Lemma 3.2 to define $\nu_G(f)$ directly for all endomorphisms of finite proper G-CW-complexes.

Definition 3.3. Let X be a finite proper G-CW-complex, and let $f : X \to X$ be an equivariant endomorphism. Then the *equivariant Nielsen class* of f is

$$\nu_G(f) := \overline{\lambda_G(f)} \in \bigoplus_{\substack{\bar{x} \in \mathrm{Is}\,\Pi(G,X),\\ X^H(f(x)) = X^H(x)}} \mathbb{Z}\pi_1(X^H(x), x)_{\phi''}.$$

We define equivariant Nielsen invariants by counting the essential classes $\overline{\alpha}$ of $\nu_G(f)_{\bar{x}}$ in $\mathbb{Z}\pi_1(X^H(x), x)_{\phi''}$ and of $L^{\mathbb{Q}\,\mathrm{Aut}(x)}\big(\widetilde{f^H(x)}\big)$ in $\mathbb{Q}\pi_1(X^H(x), x)_{\phi'}$.

Definition 3.4. Let G be a discrete group, let X be a finite proper G-CW-complex, and let $f : X \to X$ be a G-equivariant map.

Then the *equivariant Nielsen invariants* of f are elements

$$N_G(f), N^G(f) \in \bigoplus_{\bar{x} \in \mathrm{Is}\,\Pi(G,X)} \mathbb{Z}$$

defined for $\bar{x}$ with $X^H(f(x)) = X^H(x)$ by

$$N_G(f)_{\bar{x}} := \#\big\{\text{essential classes of } \nu_G(f)_{\bar{x}}\big\},$$

$$N^G(f)_{\bar{x}} := \min\Big\{\#\mathscr{C} \;\Big|\; \mathscr{C} \subseteq \bigcup_{y \geq x} \pi_1(X^K(y), y)_{\phi'} \text{ such that for all } \bar{z} \geq \bar{x} \text{ and}$$
$$\text{for all essential classes } \overline{\alpha} \text{ of } L^{\mathbb{Q}\,\mathrm{Aut}(z)}\big(\widetilde{f^{G_z}(z)}\big) \text{ there are}$$
$$\overline{\beta} \in \mathscr{C} \text{ and } (\sigma, [t]) \in \mathrm{Mor}(z, y_{\overline{\beta}}) \text{ such that } (\sigma, [t])^*(\overline{\beta}) = \alpha\Big\}.$$

We continue them by 0 to $\bar{x} \in \mathrm{Is}\,\Pi(G, X)$ with $X^H(f(x)) \neq X^H(x)$.

Note that $N_G(f)_{\bar{x}} = N_G(f_H(x))$ and $N^G(f)_{\bar{x}} = NO_G(f^H(x))$ in the notation of [Wong 1993]. Thus the invariants defined here using the algebraic approach are equivalent to the invariants defined using the classical covering space approach of Wong.

An essential class $\overline{\alpha}$ of $L^{\mathbb{Q}\,\mathrm{Aut}(x)}\big(\widetilde{f^H(x)}\big)$ corresponds to an essential fixed point class of $f^H(x)$, a WH_x-orbit of fixed points which one cannot get rid of under any G-homotopy, as can be seen from the refined orbifold Lefschetz fixed point theorem [Weber 2005, Theorem 6.6]. An essential class $\overline{\alpha}$ of $\nu_G(f)_{\bar{x}}$ corresponds to an essential fixed point class of $f_H(x)$, an orbit of fixed points on $X^H(x) \setminus X^{>H}(x)$ that cannot be moved into $X^{>H}(x)$. Counting the essential classes will give us information on the number of fixed points and fixed point orbits.

The equivariant Nielsen invariants are G-homotopy invariant since they are derived from $\lambda_G(f)$, which is itself G-homotopy invariant.

Proposition 3.5. *Given a G-homotopy $f \simeq_G f'$, we have*

$$N_G(f) = N_G(f'), \qquad N^G(f) = N^G(f').$$

Proof. If $f \simeq_G f'$, with a homotopy $H : X \times I \to X$ such that $H_0 = f$ and $H_1 = f'$, then by invariance under homotopy equivalence [Weber 2005, Theorem 5.14] we have an isomorphism $\Lambda_G(i_1)^{-1}\Lambda_G(i_0) : \Lambda_G(X, f) \xrightarrow{\sim} \Lambda_G(X, f')$ which sends $\lambda_G(f)$ to $\lambda_G(f')$. The isomorphisms $\Lambda_G(i_1)$ and $\Lambda_G(i_0)$ are given by composition of maps, so they do not change the number of essential classes. They also do not change the property of a class to lie in the image of a nonisomorphism. So we have $N_G(f) = N_G(f')$.

An isomorphism $i_{0*} : \mathbb{Q}\Pi(G, X)_{\phi,\bar{y}} \to \mathbb{Q}\Pi(G, X \times I)_{\Phi,\overline{i_0(y)}}$ is induced by the inclusion i_0, and analogously i_1 induces an isomorphism. These isomorphisms do not change the number of essential classes. We have $\mathrm{ch}_G(X, f)(\lambda_G(f)) = (i_{0*})^{-1} i_{1*}\,\mathrm{ch}_G(X, f')(\lambda_G(f'))$, so $N^G(f) = N^G(f')$. □

4. Lower bound property

The equivariant Nielsen invariants give a lower bound for the number of fixed point orbits on $X^H(x) \setminus X^{>H}(x)$ and on $X^H(x)$, for maps lying in the G-homotopy class of f. Under mild hypotheses, this is even a sharp lower bound.

Definition 4.1. Let G be a discrete group, let X be a finite proper G-CW-complex, and let $f : X \to X$ be a G-equivariant map. For every $\bar{x} \in \operatorname{Is}\Pi(G, X)$, with $x : G/H \to X$, we set

$$M_G(f)_{\bar{x}} := \min\bigl\{\#\text{ fixed point orbits of } \varphi_H(x) \bigm| \varphi \simeq_G f\bigr\},$$
$$M^G(f)_{\bar{x}} := \min\bigl\{\#\text{ fixed point orbits of } \varphi^H(x) \bigm| \varphi \simeq_G f\bigr\}.$$

When speaking of fixed point orbits of $f^H(x)$, we can either look at the WH_x-orbits $WH_x \cdot z \subseteq X^H(x)$ or at the G-orbits $G \cdot z \subseteq X^{(H)}(x)$, for a fixed point z in $X^H(x)$. These two notions are of course equivalent.

We now proceed to show the first important property of the equivariant Nielsen invariants, the lower bound property.

Proposition 4.2. *For every $\bar{x} \in \operatorname{Is}\Pi(G, X)$ we have*

$$N_G(f)_{\bar{x}} \le M_G(f)_{\bar{x}}, \qquad N^G(f)_{\bar{x}} \le M^G(f)_{\bar{x}}.$$

Proof. (1) If $\overline{\alpha} \in \pi_1(X^H(x), x)_{\phi''}$ is an essential class of $\nu_G(f)_{\bar{x}}$, there has to be at least one fixed point orbit in $X^H(x) \setminus X^{>H}(x)$ that corresponds to $\overline{\alpha}$ and that cannot be moved into $X^{>H}(x)$. So, for any $\varphi \simeq_G f$, the restriction φ_H must have at least $N_G(f)_{\bar{x}}$ fixed point orbits in $X^H(x) \setminus X^{>H}(x)$. We arrive at $N_G(f)_{\bar{x}} \le$ {#fixed point orbits of φ_H} for all $\varphi \simeq_G f$, so $N_G(f)_{\bar{x}} \le M_G(f)_{\bar{x}}$.

(2) Let $\bar{x} \in \operatorname{Is}\Pi(G, X)$. Suppose that $\varphi \simeq_G f$ such that $\varphi^H(x)$ has $M^G(f)_{\bar{x}}$ fixed point orbits in $X^H(x)$. Let $\mathscr{C} \subseteq \bigcup_{\bar{x} \le \bar{y}} \pi_1(X^K(y), y)_{\phi'}$ such that $N^G(\varphi)_{\bar{x}} = N^G(f)_{\bar{x}} = \#\mathscr{C}$. If there were less than $\#\mathscr{C}$ fixed point orbits in $X^H(x)$, there would be less that $\#\mathscr{C}$ essential classes and we could have chosen a smaller $\mathscr{C}$. So there are at least $\#\mathscr{C}$ essential classes, and thus $\varphi^H(x)$ has at least $\#\mathscr{C}$ fixed point orbits. □

To prove the sharpness of this lower bound, we need certain hypotheses, which are usually introduced when dealing with these problems. Such conditions were first used in [Fadell and Wong 1988]. Some authors treat slightly weakened assumptions [Ferrario 1999; Ferrario 2003; Jezierski 1995; Wilczyński 1984]. We do not weaken the standard gap hypotheses in the context of functorial equivariant Lefschetz invariants since the standard gap hypotheses are not homotopy invariant. So an analog of Theorem 6.3 would not hold.

Definition 4.3. Let G be a discrete group and let X be a cocompact smooth G-manifold. We say that X satisfies the *standard gap hypotheses* if for each $\bar{x} \in \operatorname{Is}\Pi(G, X)$, with $x : G/H \to X$, the inequalities $\dim X^H(x) \ge 3$ and $\dim X^H(x) - \dim X^{>H}(x) \ge 2$ hold.

Under these hypotheses, we can use an equivariant analog of Wecken's classical method [1941] to coalesce fixed points.

Lemma 4.4. *Let G be a discrete group and let X be a cocompact proper smooth G-manifold satisfying the standard gap hypotheses. Let $f : X \to X$ be a G-equivariant map. Let $\mathbb{O}_1 = Gx_1$ and $\mathbb{O}_2 = Gx_2$ be two distinct isolated G-fixed point orbits, where $x_1 : G/H \to X$ and $x_2 : G/K \to X$ with $x_1 \leq x_2$. Suppose that there are paths $(\sigma_1, [t_1]) \in \mathrm{Mor}(x, x_1)$ and $(\sigma_2, [t_2]) \in \mathrm{Mor}(x, x_2)$ for an $\bar{x} \in \Pi(G, X)$, with $x : G/H \to X$, such that $(\sigma_1, [t_1])^*\overline{1_{x_1}} = \overline{\alpha} = (\sigma_2, [t_2])^*\overline{1_{x_2}}$, i.e., that the fixed point orbits induce the same $\overline{\alpha} \in \pi_1(X^H(x), x)_{\phi'}$. Then there exists a G-homotopy $\{f_t\}$ relative to $X^{>(H)}$ such that $f_0 = f$ and $\mathrm{Fix}\, f_1 = \mathrm{Fix}\, f_0 - G\mathbb{O}_1$.*

Proof. Suppose first that $\overline{x_1} < \overline{x_2}$. Then $\mathrm{Mor}(x_1, x_2) \neq \varnothing$. By replacing x_1 and x_2 with other points in the orbit if necessary, we can suppose that there exists a morphism $(\tau, [v]) \in \mathrm{Mor}(x_1, x_2)$, where v is a path in $X^H(x)$ with $v_1 = x_1$ and $v_0 = x_2 \circ \tau$ and $\tau : G/H \to G/K$ is a projection. We know that $v \simeq f^H \circ v$ (relative endpoints). (This is an equivalent characterization of x_1 and x_2 belonging to the same fixed point class [Jiang 1983, I.1.10].) Since $x_1 \in X^H(x) \setminus X^{>H}(x)$ and $x_2 \in X^{>H}(x)$ and $\dim X^H(x) - \dim X^{>H}(x) \geq 2$, we may assume that v can be chosen such that $v((0, 1]) \subseteq X^H(x) \setminus X^{>H}(x)$. We coalesce x_1 and x_2 along v as in [Wong 1991b, 1.1] and [Schirmer 1986, 6.1]. We can do this by only changing f in a (cone-shaped) neighborhood $U(v)$ of v. Because of the proper action of G on X and the free action of WH on $X^H \setminus X^{>H}$, this neighborhood $U(v)$ can be chosen such that in $X^H \setminus X^{>H}$ it does not intersect its g-translates for $g \notin H \leq G$. Taking the G-translates of $U(v)$, we move $\mathbb{O}_1$ to $\mathbb{O}_2$ along the paths Gv in $GU(v)$, not changing the map f outside $GU(v)$.

Now suppose $\overline{x_1} = \overline{x_2}$. In this case, the result follows from [Wong 1991a, 5.4], since $X^H(x) \setminus X^{>H}(x)$ is a free and proper WH_x-space, where again the proper action of G on X ensures that we can find a neighborhood of a path from x_1 to x_2 such that the G/H-translates do not intersect. □

From Lemma 4.4, we can conclude the sharpness of the lower bound given by the equivariant Nielsen invariants.

Theorem 4.5. *Let G be a discrete group. Let X be a cocompact proper smooth G-manifold satisfying the standard gap hypotheses. Let $f : X \to X$ be a G-equivariant endomorphism. Then*

$$M_G(f)_{\bar{x}} = N_G(f)_{\bar{x}}, \qquad M^G(f)_{\bar{x}} = N^G(f)_{\bar{x}}$$

for all $\bar{x} \in \mathrm{Is}\, \Pi(G, X)$.

Proof. (1) Since X is a cocompact smooth G-manifold, there is a G-map f' which is G-homotopic to f and which has only finitely many fixed point orbits. We apply Lemma 4.4 to f' to coalesce fixed point orbits in $X^H(x) \setminus X^{>H}(x)$ with others of the same class $\overline{\alpha} \in \mathbb{Z}\pi_1(X^H(x), x)_{\phi'}$. We move them into $X^{>H}(x)$ whenever possible. (We might need to create a fixed point orbit in the inessential fixed point class beforehand; see [Wong 1991b, 1.1].) We remove the inessential fixed point orbits. We arrive at a map $h \simeq_G f$ such that $N_G(f)_{\bar{x}} = \#\{\text{fixed point orbits of } h_H(x)\} \geq M_G(f)_{\bar{x}}$. Using Proposition 4.2, we obtain equality.

(2) Since X is a cocompact smooth G-manifold, there is a map f' which is G-homotopic to f and which only has finitely many fixed point orbits. We have a partial ordering on the $\bar{y} \geq \bar{x}$ given by $\bar{y} \geq \bar{z} \Leftrightarrow \text{Mor}(z, y) \neq \varnothing$. We apply Lemma 4.4 to f' to coalesce fixed point orbits of the same class, starting from the top. Note that when we remove fixed point orbits, we can only move them up in this partial ordering. That is why the definition has to be so complicated. We remove the inessential fixed point orbits. We are left with one fixed point orbit for every essential class.

We now look at a class $\mathscr{C}$ such that $N^G(f)_{\bar{x}} = \#\mathscr{C}$, and we coalesce the essential fixed point orbits with the corresponding classes appearing in $\mathscr{C}$. (If the corresponding class in $\mathscr{C}$ is inessential, we might need to create a fixed point orbit in this inessential fixed point class beforehand.) We obtain a map $h \simeq_G f$ which has exactly $\#\mathscr{C}$ fixed point orbits. Hence

$$N^G(f)_{\bar{x}} = \#\{\text{fixed point orbits of } h^H(x)\} \geq M^G(f)_{\bar{x}}.$$

Using Proposition 4.2, we obtain equality. □

In general, it is not possible to find a map $h \simeq_G f$ realizing all minima simultaneously. As an example, one can take $G = \mathbb{Z}/2$ acting on $X = S^4$ as an involution so that $X^{\mathbb{Z}/2} = S^3$. One obtains $M_G(\text{id}_{S^4})_{\bar{x}} = 0$ for all $\bar{x} \in \text{Is}\,\Pi(\mathbb{Z}/2, S^4)$, but the minimal number of fixed points in the G-homotopy class of the identity id_{S^4} is equal to 1 [Wong 1993, Remark 3.4]. In this example, the standard gap hypotheses are not satisfied. Other examples where the standard gap hypotheses do not hold and where the converse of the equivariant Lefschetz theorem is false are given in [Ferrario 1999, Section 5].

5. The *G*-Jiang condition

In the nonequivariant case, we know that the generalized Lefschetz invariant is the right element when looking for a precise count of fixed points. We read off the Nielsen numbers from this invariant. In general, the Lefschetz number contains too little information. But under certain conditions, we can conclude facts about the Nielsen numbers from the Lefschetz numbers directly, and thus obtain a converse of the Lefschetz fixed point theorem.

These conditions are called Jiang conditions. See [Jiang 1983, Definition II.4.1], where one can find a thorough treatment, and [Brown 1971, Chapter VII]. The Jiang group is a subgroup of $\pi_1(X, f(x))$ [Jiang 1983, Definition II.3.5]. We generalize its definition to the equivariant case.

Definition 5.1. Let G be a discrete group, let X be a finite proper G-CW-complex, and let $f : X \to X$ be a G-equivariant endomorphism. Then a G-equivariant self-homotopy $h : f \simeq_G f$ of f determines a path $h(x, -) \in \pi_1(X^H(x), f(x))$ for every $\bar{x} \in \operatorname{Is}\Pi(G, X)$, with $x : G/H \to X$. Define the *G-Jiang group* of (X, f) to be

$$\begin{aligned} J_G(X, f) &:= \left\{ \sum_{\bar{x}\in\Pi(G,X)} \big[h(x, -)\big] \,\middle|\, h : f \simeq_G f \text{ G-equivariant self-homotopy} \right\} \\ &\le \bigoplus_{\bar{x}\in\Pi(G,X)} \pi_1\big(X^H(x), f(x)\big), \end{aligned}$$

and define the *G-Jiang group* of X to be

$$\begin{aligned} J_G(X) &:= \left\{ \sum_{\bar{x}\in\Pi(G,X)} \big[h(x, -)\big] \,\middle|\, h : \mathrm{id} \simeq_G \mathrm{id} \text{ G-equivariant self-homotopy} \right\} \\ &\le \bigoplus_{\bar{x}\in\Pi(G,X)} \pi_1\big(X^H(x), x\big). \end{aligned}$$

In the nonequivariant case, we know that the Jiang group $J(X, f, x)$ is a subgroup of the centralizer of $\pi_1(f, x)\big(\pi_1(X, x)\big)$ in $\pi_1\big(X, f(x)\big)$. In particular,

$$J(X) \le Z\big(\pi_1(X, x)\big),$$

where $Z\big(\pi_1(X, x)\big)$ denotes the center of $\pi_1(X, x)$ [Jiang 1983, Lemma II.3.7]. Furthermore, the isomorphism $(f \circ w)_* : \pi_1\big(X, f(x_1)\big) \to \pi_1\big(X, f(x_0)\big)$ induced by a path w from x_0 to x_1 induces an isomorphism $(f \circ w)_* : J(X, f, x_1) \to J(X, f, x_0)$ which does not depend on the choice of w. So the definition does not depend on the choice of the basepoint [Jiang 1983, Lemma II.3.9]. It is also known that $J(X) \le J(X, f) \le \pi_1(X)$ for all f [Jiang 1983, Lemma II.3.8]. This leads to the consideration of spaces with $J(X) = \pi_1(X)$ in the definition of a Jiang space. All these lemmata also make sense in the equivariant case. Thus we make the following definition.

Definition 5.2. Let G be a discrete group and let X be a cocompact G-CW-complex. Then X is called a *G-Jiang space* if for all $\bar{x} \in \operatorname{Is}\Pi(G, X)$ we have

$$J_G(X)_{\bar{x}} = \pi_1\big(X^H(x), x\big).$$

The group $J_G(X, f)$ acts on $\Lambda_G(X, f)$ as follows: If $X^H(f(x)) = X^H(x)$ and $X^H(f(x)) = X^H(x)$, then $J_G(X, f)_{\bar{x}}$ acts on $\mathbb{Z}\pi_1(X^H(x), x)$. The element $u = [h(x, -)] \in J_G(X, f)_{\bar{x}}$ acts as composition with $[wuw^{-1}]$, where $v = (\mathrm{id}, [w]) \in \operatorname{Mor}(f(x), x)$.

Since $J_G(X, f)_{\bar{x}}$ is contained in the centralizer of $\pi_1(f^H(x), x)\big(\pi_1(X^H(x), x)\big)$ in $\pi_1\big(X^H(x), f(x)\big)$, this action induces an action on $\mathbb{Z}\pi_1(X^H(x), x)_{\phi'}$ by composition, whence on $\Lambda_G(X, f)_{\bar{x}}$. Thus $J_G(X, f)$ acts on $\Lambda_G(X, f)$, and by invariance of $\lambda_G(f)$ under homotopy equivalence, we see that

$$\lambda_G(f) \in \big(\Lambda_G(X, f)\big)^{J_G(X,f)}.$$

Examples of G-Jiang spaces can be obtained from Jiang spaces. It is known [Jiang 1983, Theorem II.3.11] that the class of Jiang spaces is closed under homotopy equivalence and the topological product operation and contains

- simply connected spaces,
- generalized lens spaces,
- H-spaces,
- homogeneous spaces of the form A/A_0 where A is a topological group and A_0 is a subgroup which is a connected compact Lie group.

Hence we obtain many examples of G-Jiang spaces using the following proposition, analogous to [Wong 1993, Proposition 4.9].

Proposition 5.3. *Let G be a discrete group, and let X be a free cocompact connected proper G-space. If X/G is a Jiang space, then X is a G-Jiang space.*

Proof. Since X is connected and free, the set $\operatorname{Is}\Pi(G, X)$ consists of one element. Let x be a basepoint of X. We need to check that $J_G(X)_{\bar{x}} = \pi_1(X, x)$. Let $X \xrightarrow{p} X/G$ be the projection. The Jiang subgroup of X/G is given by

$$\begin{aligned} J(X/G) :=& \Big\{\big[h(p(x), -)\big] \,\Big|\, h : \mathrm{id}_{X/G} \simeq \mathrm{id}_{X/G} \text{ self-homotopy}\Big\} \\ &\leq \pi_1\big(X/G, p(x)\big). \end{aligned}$$

Let $\alpha \in \pi_1(X, x)$. Since $X \xrightarrow{p} X/G$ is a discrete cover, $\widetilde{X} = \widetilde{X/G}$. There is a map $p_\# : \pi_1(X, x) \to \pi_1(X/G, p(x))$ induced by the projection. Since X/G is a Jiang space, $J(X/G) = \pi_1(X/G, p(x))$, so there is a homotopy $h : \mathrm{id}_{X/G} \simeq \mathrm{id}_{X/G}$ such that $p_\#(\alpha) = [h(p(x), -)]$. Because of the free and proper action of G on X, this homotopy h can be lifted to a G-equivariant homotopy $h' : \mathrm{id}_X \simeq_G \mathrm{id}_X$ such that $\alpha = [h'(x, -)]$. Thereby $\alpha \in J_G(X)$. □

6. The converse of the equivariant Lefschetz Fixed Point Theorem

One can derive equivariant analogs of statements about Nielsen numbers found in [Jiang 1983], generalizing results from [Wong 1993] to infinite discrete groups. In particular, if X is a G-Jiang space, the converse of the equivariant Lefschetz fixed point theorem holds. The next theorem can be compared with [Jiang 1983, Theorem II.4.1].

Theorem 6.1. *Let G be a discrete group, and let X be a finite proper G-CW-complex which is a G-Jiang space. Then for any G-map $f : X \to X$ and $\bar{x} \in \operatorname{Is}\Pi(G, X)$ with $x : G/H \to X$ we have*:

$$L_G(f)_{\bar{x}} = 0 \implies \lambda_G(f)_{\bar{x}} = 0 \text{ and } N_G(f)_{\bar{x}} = 0,$$
$$L_G(f)_{\bar{x}} \neq 0 \implies \lambda_G(f)_{\bar{x}} \neq 0 \text{ and } N_G(f)_{\bar{x}} = \#\{\pi_1(X^H(x), x)_{\phi''}\}.$$

Here $L_G(f)$ is the equivariant Lefschetz class [Lück and Rosenberg 2003a, Definition 3.6], the equivariant analog of the Lefschetz number.

Proof. Since X is a G-Jiang space, the G-Jiang group $J_G(X)$ acts transitively on $\pi_1(X^H(x), x)$ for all $\bar{x} \in \operatorname{Is}\Pi(G, X)$. This implies that

$$\lambda_G(f)_{\bar{x}} = \sum_{\overline{\alpha}} n_{\overline{\alpha}} \cdot \overline{\alpha} = n \cdot \sum_{\overline{\alpha}} \overline{\alpha}$$

for some $n \in \mathbb{Z}$. This leads to $L_G(f)_{\bar{x}} = n \cdot \#\{\pi_1(X^H(x), x)_{\phi'}\}$ by the augmentation map. We see that

$$\begin{aligned} L_G(f)_{\bar{x}} = 0 &\implies n = 0 \\ &\implies \lambda_G(f)_{\bar{x}} = 0 \\ &\implies \nu_G(f)_{\bar{x}} = 0 \\ &\implies N_G(f)_{\bar{x}} = 0, \end{aligned}$$

$$\begin{aligned} L_G(f)_{\bar{x}} \neq 0 &\implies n \neq 0 \\ &\implies \lambda_G(f)_{\bar{x},\overline{\alpha}} \neq 0 \text{ for all } \overline{\alpha} \in \mathbb{Z}(\pi_1(X^H(x), x))_{\phi'} \\ &\implies \nu_G(f)_{\bar{x},\overline{\alpha}} \neq 0 \text{ for all } \overline{\alpha} \in \mathbb{Z}(\pi_1(X^H(x), x))_{\phi''} \\ &\implies N_G(f)_{\bar{x}} = \#\{\pi_1(X^H(x), x)_{\phi''}\}. \end{aligned}$$

□

The proof of Theorem 6.1 already works if $J_G(X, f)$ acts transitively on every summand of $\Lambda_G(X, f)$. We could have called X a G-Jiang space if the condition that $J_G(X, f)$ acts transitively on every summand of $\Lambda_G(X, f)$ is satisfied. But this condition is less tractable. It is implied by $J_G(X, f)_{\bar{x}} = \pi_1(X^H(x), f(x))$ for all $\bar{x}$, which is implied by $J_G(X)_{\bar{x}} = \pi_1(X^H(x), x)$ for all $\bar{x}$.

We now show that f is G-homotopic to a fixed point free G-map if the generalized equivariant Lefschetz invariant $\lambda_G(f)$ is zero.

Theorem 6.2. *Let G be a discrete group. Let X be a cocompact proper smooth G-manifold satisfying the standard gap hypotheses. Let $f : X \to X$ be a G-equivariant endomorphism. If $\lambda_G(f) = 0$, then f is G-homotopic to a fixed point free G-map.*

Proof. If $\lambda_G(f)=0$, then $\mathrm{ch}_G(X, f)(\lambda_G(f))=0$, and therefore we have $N^G(f)_{\bar{x}} = 0$ for all $\bar{x} \in \operatorname{Is}\Pi(G, X)$. We know from Theorem 4.5 that $N^G(f)_{\bar{x}} = M^G(f)_{\bar{x}} = \min\{\#\text{ fixed point orbits of } \varphi^H(x) \mid \varphi \simeq_G f\}$. In particular, for $x : G/\{1\} \to X$ we obtain a map φ such that $\varphi^{\{1\}}(x)$ is fixed point free and $\varphi \simeq_G f$. Thus we obtain our result on every connected component of X, and combining these we arrive at a map $h \simeq_G f$ which is fixed point free. □

These two theorems, Theorem 6.1 and Theorem 6.2, combine to give the main theorem of this paper, the converse of the equivariant Lefschetz fixed point theorem.

Theorem 6.3. *Let G be a discrete group. Let X be a cocompact proper smooth G-manifold satisfying the standard gap hypotheses which is a G-Jiang space. Let $f : X \to X$ be a G-equivariant endomorphism. Then the following holds*:

If $L_G(f) = 0$, then f is G-homotopic to a fixed point free G-map.

Proof. We know that $L_G(f) = 0$ means that $L_G(f)_{\bar{x}} = 0$ for all $\bar{x} \in \operatorname{Is}\Pi(G, X)$. Since X is a G-Jiang space, by Theorem 6.1 this implies that $\lambda_G(f)_{\bar{x}} = 0$ for all $\bar{x} \in \operatorname{Is}\Pi(G, X)$, so we have $\lambda_G(f) = 0$. We apply Theorem 6.2 to arrive at the desired result. □

Remark 6.4. As another corollary of Theorem 6.2, we obtain: If G is a discrete group and X is a cocompact proper smooth G-manifold satisfying the standard gap hypotheses, then $\chi^G(X) = 0$ implies that the identity id_X is G-homotopic to a fixed point free G-map. This was already stated in [Lück and Rosenberg 2003a, Remark 6.8]. Here $\chi^G(X)$ is the universal equivariant Euler characteristic of X [Lück and Rosenberg 2003a, Definition 6.1] defined by $\chi^G(X)_{\bar{x}} = \chi\big(WH_x \backslash X^H(x), WH_x \backslash X^{>H}(x)\big) \in \mathbb{Z}$, we have $\chi^G(X) = L_G(\mathrm{id}_X)$. We calculate that

$$\begin{aligned}\lambda_G(\mathrm{id}_X)_{\bar{x}} &= \sum_{p\geq 0}(-1)^p \sum_{\substack{\mathrm{Aut}(x)\cdot e\in \\ \mathrm{Aut}(x)\backslash I_p(\widetilde{X^H(x)},\widetilde{X^{>H}(x)})}} \mathrm{inc}_\phi\big(\mathrm{id}_{\widetilde{X^H(x)}}, e\big) \\ &= \chi\big(WH_x \backslash X^H(x), WH_x \backslash X^{>H}(x)\big)\cdot \bar{1} \in \mathbb{Z}\pi_1\big(X^H(x), x\big)_{\phi'}.\end{aligned}$$

So we have $\chi^G(X) = 0$ if and only if $\lambda_G(\mathrm{id}_X) = 0$, and with Theorem 6.2 we conclude that there is an endomorphism G-homotopic to the identity which is fixed point free.

References

[Brown 1971] R. F. Brown, *The Lefschetz fixed point theorem*, Scott, Foresman, Glenview, IL, 1971. MR 44 #1023

[tom Dieck 1979] T. tom Dieck, *Transformation groups and representation theory*, Lecture Notes in Mathematics **766**, Springer, Berlin, 1979. MR 82c:57025 Zbl 0445.57023 [tom Dieck 1987] T. tom Dieck, *Transformation groups*, Studies in Mathematics **8**, de Gruyter, Berlin, 1987. MR 89c:57048 Zbl 0611.57002

[Fadell and Wong 1988] E. Fadell and P. Wong, "On deforming G-maps to be fixed point free", *Pacific J. Math.* **132**:2 (1988), 277–281. MR 89d:55003

[Ferrario 1999] D. L. Ferrario, "A fixed point index for equivariant maps", *Topol. Methods Nonlinear Anal.* **13**:2 (1999), 313–340. MR 2001b:55004 Zbl 0959.55004

[Ferrario 2003] D. L. Ferrario, "A Möbius inversion formula for generalized Lefschetz numbers", *Osaka J. Math.* **40**:2 (2003), 345–363. MR 2004d:55001

[Ferrario 2005] D. L. Ferrario, "A note on equivariant fixed point theory", pp. 287–300 in *Handbook of topological fixed point theory*, edited by F. Brown et al., Springer, Dordrecht, 2005. MR 2006e:55003 Zbl 1082.55001

[Hart 1999] E. L. Hart, "The Reidemeister trace and the calculation of the Nielsen number", pp. 151–157 in *Nielsen theory and Reidemeister torsion* (Warsaw, 1996), edited by J. Jezierski, Banach Center Publ. **49**, Polish Acad. Sci., Warsaw, 1999. MR 1734729 Zbl 0943.55001

[Jezierski 1995] J. Jezierski, "A modification of the relative Nielsen number of H. Schirmer", *Topology Appl.* **62**:1 (1995), 45–63. MR 95m:55002

[Jiang 1983] B. J. Jiang, *Lectures on Nielsen fixed point theory*, Contemporary Mathematics **14**, American Mathematical Society, Providence, R.I., 1983. MR 84f:55002 Zbl 0512.55003

[Lück 1989] W. Lück, *Transformation groups and algebraic K-theory*, Lecture Notes in Mathematics **1408**, Springer, Berlin, 1989. MR 91g:57036 Zbl 0679.57022

[Lück and Rosenberg 2003a] W. Lück and J. Rosenberg, "The equivariant Lefschetz fixed point theorem for proper cocompact G-manifolds", pp. 322–361 in *High-dimensional manifold topology* (Trieste, 2001), edited by F. T. Farrell and W. Lück, World Scientific, River Edge, NJ, 2003. MR 2005b:57064 Zbl 1046.57026

[Lück and Rosenberg 2003b] W. Lück and J. Rosenberg, "Equivariant Euler characteristics and K-homology Euler classes for proper cocompact G-manifolds", *Geom. Topol.* **7** (2003), 569–613. MR 2004k:19005 Zbl 1034.19003

[Nielsen 1920] J. Nielsen, "Über die Minimalzahl der Fixpunkte bei den Abbildungstypen der Ringflächen", *Math. Ann.* **82**:1-2 (1920), 83–93. MR 1511973 JFM 47.0527.03

[Reidemeister 1936] K. Reidemeister, "Automorphismen von Homotopiekettenringen", *Math. Ann.* **112**:1 (1936), 586–593. MR 1513064 JFM 62.0659.05

[Schirmer 1986] H. Schirmer, "A relative Nielsen number", *Pacific J. Math.* **122**:2 (1986), 459–473. MR 87e:55002 Zbl 0553.55001

[Weber 2005] J. Weber, "The universal functorial equivariant Lefschetz invariant", *K-Theory* **36**:1-2 (2005), 169–207. MR 2274162 Zbl 05125362

[Wecken 1941] F. Wecken, "Fixpunktklassen, II: Homotopieinvarianten der Fixpunkttheorie", *Math. Ann.* **118** (1941), 216–234. MR 5,275a JFM 67.1094.01

[Wilczyński 1984] D. Wilczyński, "Fixed point free equivariant homotopy classes", *Fund. Math.* **123**:1 (1984), 47–60. MR 86i:57046 Zbl 0548.55002

[Wong 1991a] P. Wong, "Equivariant Nielsen fixed point theory for G-maps", *Pacific J. Math.* **150**:1 (1991), 179–200. MR 92i:55005 Zbl 0691.55004

[Wong 1991b] P. Wong, "On the location of fixed points of G-deformations", *Topology Appl.* **39**:2 (1991), 159–165. MR 92c:55004 Zbl 0722.55002

[Wong 1993] P. Wong, "Equivariant Nielsen numbers", *Pacific J. Math.* **159**:1 (1993), 153–175. MR 94e:55006 Zbl 0739.55001

Received November 16, 2005. Revised June 21, 2006.

JULIA WEBER
MAX-PLANCK-INSTITUT FÜR MATHEMATIK
VIVATSGASSE 7
D-53111 BONN
GERMANY
jweber@mpim-bonn.mpg.de

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 231 No. 1 May 2007
