# Pacific Journal of Mathematics

Volume 268 No. 1 March 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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



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

The subscription price for 2014 is US \$410/year for the electronic version, and \$535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.1

# DEFORMATION OF THREE-DIMENSIONAL HYPERBOLIC CONE STRUCTURES: THE NONCOLLAPSING CASE

ALEXANDRE PAIVA BARRETO

*Dedicated to my wife Cynthia*

**This work is devoted to the study of deformations of hyperbolic cone structures under the assumption that the length of the singularity remains uniformly bounded over the deformation. Let $(M_i, p_i)$ be a sequence of pointed hyperbolic cone manifolds with cone angles of at most $2\pi$ and topological type $(M, \Sigma)$, where $M$ is a closed, orientable and irreducible 3-manifold and $\Sigma$ an embedded link in $M$. Assuming that the length of the singularity remains uniformly bounded, we prove that either the sequence $M_i$ collapses and $M$ is Seifert fibered or a Sol manifold, or the sequence $M_i$ does not collapse and, in this case, a subsequence of $(M_i, p_i)$ converges to a complete three dimensional Alexandrov space endowed with a hyperbolic metric of finite volume on the complement of a finite union of quasigeodesics. We apply this result to a question proposed by Thurston and to provide universal constants for hyperbolic cone structures when $\Sigma$ is a small link in $M$.**

## 1. Introduction

This text focuses on deformations of hyperbolic cone structures on a closed, orientable and irreducible 3-manifold $M$ which are singular along a fixed embedded link $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_l$. Unlike complete hyperbolic structures, which are rigid by Mostow's theorem, the hyperbolic cone structures can be deformed (see [Hodgson and Kerckhoff 1998]). The difficulty in understanding these deformations lies in the possibility that the structure degenerates. In other words, the Hausdorff–Gromov limit (see Section 2 for the definition) of the deformation is only an Alexandrov space which may have dimension strictly smaller than 3, although its curvature remains bounded from below by $-1$ (see [Kojima 1998]).

From [Kojima 1998; Hodgson and Kerckhoff 2005; Fujii 2000] it is known that the degeneration of the hyperbolic cone structures occurs if and only if the

This work is part of Barreto's doctoral thesis, made at Université Paul Sabatier–Toulouse under the supervision of Professor Michel Boileau and supported by CAPES and FAPESP (2009/16234-5).
*MSC2010:* primary 57M50; secondary 57N16, 53C23.
*Keywords:* hyperbolic 3-manifolds, cone manifolds, Alexandrov spaces.

singular link of these structures intersects itself during the deformation. When the cone angles vary between 0 and $\pi$, the Dirichlet polyhedron of the hyperbolic cone structures is convex and we can use this fact to avoid self intersections of the singular link over deformations (see [Kojima 1998]). In this article we will not use this restrictive assumption and allow the cone angles vary until $2\pi$.

We are interested in studying the following question that was proposed by W. Thurston in the 1980s:

**Question 1.** *Let $M$ be a closed and orientable hyperbolic 3-manifold and suppose there exists a simple closed geodesic $\Sigma$ in $M$. Can the hyperbolic structure of $M$ be deformed to the complete hyperbolic structure on $M-\Sigma$ through a path $M_\alpha$ of hyperbolic cone structures with topological type $(M, \Sigma)$ and parametrized by the cone angles $\alpha \in [0, 2\pi]$?*

If the deformation proposed by Thurston exists, it is a consequence of his hyperbolic Dehn surgery theorem that the length of the singular link must converge to zero. In particular, we have that its length remains uniformly bounded over the deformation. This conclusion give us a necessary condition for the existence of Thurston's desired deformation. For this reason, we will focus only on deformations of hyperbolic cone structures with this additional hypothesis on the singularity's length. We remark that this assumption is automatically verified when the holonomy representations of the hyperbolic cone structures are convergent.

We started studying this question in [Barreto 2012]. In that paper we obtained the following result (see Section 3 for the definition of collapse):

**Theorem 2.** *Let $M$ be a closed, orientable and irreducible 3-manifold and let $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_l$ be an embedded link in $M$. Suppose there exists a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$ and having cone angles $\alpha_{ij} \in (0, 2\pi]$ along $\Sigma_j$ for $i \in \mathbb{N}$. Denote by $\mathscr{L}_{M_i}(\Sigma_j)$ the length of the connected component $\Sigma_j$ of $\Sigma$ in the hyperbolic cone manifold $M_i$. If*

$$\sup\bigl\{\mathscr{L}_{M_i}(\Sigma_j) \bigm| i \in \mathbb{N} \text{ and } j \in \{1, \dots, l\}\bigr\} < \infty \tag{1-1}$$

*and the sequence $M_i$ collapses, then $M$ is Seifert fibered or a* Sol *manifold.*

As a consequence of this theorem, we obtained the following result yielding some information on Thurston's question:

**Corollary 3.** *Let $M$ be a closed and orientable hyperbolic 3-manifold and suppose there exists a finite union of disjoint simple closed geodesics $\Sigma$ in $M$. Let $M_\alpha$ be a (angle decreasing) deformation of this structure along a continuous path of hyperbolic cone structures with topological type $(M, \Sigma)$ and having cone angles $\alpha \in (L, 2\pi] \subset [0, 2\pi]$ (the same for all components of $\Sigma$). If*

$$\sup\bigl\{\mathscr{L}_{M_\alpha}(\Sigma_j) \bigm| \alpha \in (L, 2\pi] \text{ and } j \in \{1, \dots, l\}\bigr\} < \infty, \tag{1-2}$$

*then every convergent sequence $M_{\alpha_i}$, with $\alpha_i$ converging to $L$, does not collapse.*

In this article, we will focus on noncollapsing deformations of hyperbolic cone structures. The principal result of this paper is the following one:

**Theorem 4.** *Let $M$ be a closed, orientable and irreducible 3-manifold and let $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_l$ be an embedded link in $M$. Suppose there exists a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$ and having cone angles $\alpha_{ij} \in (0, 2\pi]$ along $\Sigma_j$ for $i \in \mathbb{N}$. Denote by $\mathcal{L}_{M_i}(\Sigma_j)$ the length of the connected component $\Sigma_j$ of $\Sigma$ in the hyperbolic cone manifold $M_i$. If*

$$\sup\left\{\mathcal{L}_{M_i}(\Sigma_j) \;\middle|\; i \in \mathbb{N} \text{ and } j \in \{1, \ldots, l\}\right\} < \infty,$$

*then one of the following statements holds:*

(i) *The sequence $M_i$ collapses and $M$ is Seifert fibered or a* Sol *manifold.*

(ii) *The sequence $M_i$ does not collapse and there exists a sequence of points $p_{i_k} \in M - \Sigma$ such that the sequence $(M_{i_k}, p_{i_k})$ converges to a three-dimensional pointed Alexandrov space $(Z, z_0)$. The Alexandrov space $Z$ is endowed with a (noncomplete) hyperbolic metric of finite volume on the complement of a finite union $\Sigma_Z$ of quasigeodesics. Moreover, $Z$ is homeomorphic to $M$ (in particular, $Z$ is compact) if there exists $\varepsilon \in (0, 2\pi)$ such that the cone angles $\alpha_{ij}$ belong to $(\varepsilon, 2\pi]$. Further, the following statements are equivalent:*

   (a) *$Z$ is compact.*

   (b) $\inf\left\{\text{cone angle}_{M_{i_k}}(\Sigma_j) \;\middle|\; k \in \mathbb{N} \text{ and } \Sigma_j \subset \Sigma\right\} > 0$.

   (c) $\inf\left\{\mathcal{L}_{M_{i_k}}(\Sigma_j) \;\middle|\; k \in N\right\} > 0$ *for each component $\Sigma_j$ of $\Sigma$.*

**Remark 5.** A byproduct of this theorem is that the length of a connected component $\Sigma_j$ of $\Sigma$ shrinks down to zero if and only if the same arises for its cone angles $\alpha_{ij}$ (when $i$ goes to infinity). If the cone angles are supposed to be the same on all connected components of $\Sigma$, it follows from this (see Corollary 23) that the sequence of cone angles converges to zero if and only if the following statements hold:

(i) $\sup\left\{\mathcal{L}_{M_i}(\Sigma) \;\middle|\; i \in \mathbb{N}\right\} < \infty$.

(ii) $\lim_{i\to\infty} \operatorname{diam}(M_i) = \infty$.

(iii) The sequence $M_i$ does not collapse.

In general, the limiting singular locus $\Sigma_Z$ need not be a disjoint union of quasi-geodesics since the singular link could intersect itself as cone angles are changed. It seems possible that the components of $\Sigma_Z$ are continuous geodesics and that the limit is a hyperbolic cone manifold in a more general sense allowing singularities along a graph instead of a link. The main problem in understanding the limiting singular locus lies in the possibility that the singularity intersects itself infinitely many times at the limit. More precisely, $\Sigma_Z$ may be a graph with infinite degree vertices. A better comprehension of the limiting singular locus is an interesting problem for further investigation.

As an application of Theorem 4, we obtain the following result related to Question 1.

**Corollary 6.** *Let $M$ be a closed and orientable hyperbolic 3-manifold and suppose there exists a finite union of disjoint simple closed geodesics $\Sigma$ in $M$. Let $M_\alpha$ be a deformation of this structure along a continuous path of hyperbolic cone structures with topological type $(M, \Sigma)$ and having cone angles $\alpha \in (\theta, 2\pi] \subset [0, 2\pi]$ (the same for all components of $\Sigma$). The following statements are equivalent*:

(i) $\theta = 0$ *and the path $M_\alpha$ extends continuously to* $[0, 2\pi]$, *where $M_0$ denotes $M - \Sigma$ with the complete hyperbolic metric.*

(ii) $\lim_{\alpha \to \theta} \mathscr{L}_{M_\alpha}(\Sigma) = \lim_{\alpha \to \theta} \sum_{i=1}^{l} \mathscr{L}_{M_\alpha}(\Sigma_j) = 0.$

(iii) *There exists a sequence $\alpha_i \in (\theta, 2\pi]$ converging to $\theta$ satisfying*

$$\sup\left\{\mathscr{L}_{M_\alpha}(\Sigma_j) \;\middle|\; \alpha \in (\theta, 2\pi] \text{ and } j \in \{1, \dots, l\}\right\} < \infty$$

*and such that the sequence* $\operatorname{diam}(M_{\alpha_i})$ *goes to infinity with $i$.*

**Remark 7.** Corollary 6 provides a necessary and sufficient condition for the existence of the deformation proposed by Thurston. Using the notation in Question 1,

$$\theta = 0 \quad \text{if and only if} \quad \lim_{\alpha \to \theta} \mathscr{L}_{M_\alpha}(\Sigma) = 0.$$

Supposing in addition that $M$ is not Seifert fibered and that $\Sigma$ is a small link in $M$, we have also the following theorem (see Corollaries 25 and 26) providing universal constants for hyperbolic cone structures with topological type $(M, \Sigma)$.

**Theorem 8.** *Let $M$ be a closed, orientable, irreducible and non-Seifert fibered 3-manifold and let $\Sigma$ be a small link in $M$. There exists a constant $V = V(M, \Sigma) > 0$ and a constant $K = K(M, \varepsilon) > 0$, for each $\varepsilon \in (0, 2\pi)$, such that*

(i) $\operatorname{Vol}(\mathcal{M}) > V$ *for every hyperbolic cone manifold $\mathcal{M}$ with topological type $(M, \Sigma)$, and*

(ii) $\operatorname{diam}(\mathcal{M}) < K$ *for every hyperbolic cone manifold $\mathcal{M}$ with topological type $(M, \Sigma)$ and having cone angles in the interval $(\varepsilon, 2\pi]$.*

## 2. Metric geometry

In this section, we recall some definitions about Alexandrov spaces and Hausdorff–Gromov convergence. We refer to [Burago et al. 2001; Burago et al. 1992; Gromov 1981; Perelman and Petrunin 1994] for details.

Given a metric space $Z$, the metric on $Z$ will always be denoted by $d_Z(\cdot,\cdot)$. The open ball of radius $r>0$ about a subset $A$ of $Z$ will be denoted by

$$B_Z(A,r)=\bigcup_{a\in A}\{z\in Z \mid d_Z(z,a)<r\}.$$

A metric space $Z$ is called a *length space* (and its metric is called *intrinsic*) when the distance between every pair of points in $Z$ is given by the infimum of the lengths of all rectifiable curves connecting them. When a minimizing geodesic between every pair of points exists, we say that $Z$ is *complete*.

For all $k\in\mathbb{R}$, denote by $\mathbb{M}^2_k$ the complete and simply connected two-dimensional Riemannian manifold of constant sectional curvature equal to $k$.

Let $\triangle(x,y,z)\subset Z$ be a geodesic triangle in $Z$ with vertices $x,y,z\in Z$. The angle of $\triangle(x,y,z)$ at vertex $x$, for example, will be denoted by $\measuredangle_\triangle(x)$. A *comparison triangle* for $\triangle(x,y,z)\subset Z$ in $\mathbb{M}^2_k$ is a geodesic triangle $\overline{\triangle}_k(\bar{x},\bar{y},\bar{z})\subset\mathbb{M}^2_k$ satisfying

$$d_{\mathbb{M}^2_k}(\bar{x},\bar{y})=d_Z(x,y),\quad d_{\mathbb{M}^2_k}(\bar{y},\bar{z})=d_Z(y,z),\quad\text{and}\quad d_{\mathbb{M}^2_k}(\bar{z},\bar{x})=d_Z(z,x).$$

**Definition 9.** A length space $Z$ is called an Alexandrov space of curvature not smaller than $k\in\mathbb{R}$ if every point of $Z$ has a neighborhood $U$ such that, the angles of every triangle $\triangle(x,y,z)\subset U$ are well defined and satisfy the inequalities

$$\measuredangle_\triangle(x)\ge\measuredangle_{\overline{\triangle}_k}(\bar{x}),\quad \measuredangle_\triangle(y)\ge\measuredangle_{\overline{\triangle}_k}(\bar{y}),\quad\text{and}\quad \measuredangle_\triangle(z)\ge\measuredangle_{\overline{\triangle}_k}(\bar{z})$$

for every comparison triangle $\overline{\triangle}_k(\bar{x},\bar{y},\bar{z})\subset\mathbb{M}^2_k$ of $\triangle$.

Suppose from now on that $Z$ is an $n$-dimensional Alexandrov space of curvature not smaller than $k\in\mathbb{R}$ and fix a point $O\in\mathbb{M}^2_k$. We next recall the definition of quasigeodesics on an Alexandrov space (see [Perelman and Petrunin 1994]). Let $\gamma:[a,b]\to Z$ be a 1-Lipschitz curve and let $z\in Z$ be a point satisfying

$$0<d_Z(z,\gamma(t))<\frac{\pi}{\sqrt{k}} \tag{2-1}$$

for all $t\in[a,b]$. We say that a curve $\widetilde{\gamma}:[a,b]\to\mathbb{M}^2_k$ is a *development* of $\gamma$ with respect to $z\in Z$ when

$$d_Z(z,\gamma(t))=d_{\mathbb{M}^2_k}(O,\widetilde{\gamma}(t))$$

for all $t\in[a,b]$.

**Definition 10.** A 1-Lipschitz curve $\gamma:[a,b]\to Z$ is a quasigeodesic of $Z$ if it is parametrized by arc length and, for every point $z\in Z$ satisfying (2-1) and every development $\widetilde{\gamma}:[a,b]\to\mathbb{M}^2_k$ of $\gamma$ with respect to $z\in Z$, the curvilinear triangle bounded by the segments $O\widetilde{\gamma}(t\pm\delta)$ and the arc $\widetilde{\gamma}|_{[t-\delta,t+\delta]}$, where $t\in(a,b)$ and $\delta>0$ sufficiently small, is convex.

Given three points $x, y, z \in Z$, let $\overline{\triangle}_k(\bar{x}, \bar{y}, \bar{z})$ be a triangle in $\mathbb{M}_k^2$ satisfying

$$d_{\mathbb{M}_k^2}(\bar{x}, \bar{y}) = d_Z(x, y), \quad d_{\mathbb{M}_k^2}(\bar{y}, \bar{z}) = d_Z(y, z), \quad \text{and} \quad d_{\mathbb{M}_k^2}(\bar{z}, \bar{x}) = d_Z(z, x).$$

We denote by $\measuredangle_k(x; y, z)$ the angle of $\overline{\triangle}_k(\bar{x}, \bar{y}, \bar{z})$ at $\bar{x}$. Note that this definition does not depend on the choice of the triangle $\overline{\triangle}_k(\bar{x}, \bar{y}, \bar{z})$.

Consider $z \in Z$ and $\lambda \in (0, \pi)$. The point $z$ is said to be $\lambda$-*strained* if there exists a set $\{(a_i, b_i) \in Z \times Z \mid i \in \{1, \dots, n\}\}$, called a $\lambda$-strainer at $z$, such that $\measuredangle_k(z; a_i, b_i) > \pi - \lambda$ and

$$\max\left\{\left|\measuredangle_k(z; a_i, a_j) - \frac{\pi}{2}\right|, \left|\measuredangle_k(z; b_i, b_j) - \frac{\pi}{2}\right|, \left|\measuredangle_k(z; a_i, b_j) - \frac{\pi}{2}\right|\right\} < \lambda$$

for all $i \neq j \in \{1, \dots, n\}$. The set $R_\lambda(Z)$ of $\lambda$-strained points of $Z$ is called the *set of $\lambda$-regular points of $Z$*. It is a remarkable fact that $R_\lambda(Z)$ is an open and dense subset of $Z$.

We now recall the notion of (pointed) Hausdorff–Gromov convergence:

**Definition 11** [Burago et al. 2001]. Let $(Z_i, z_i)$ be a sequence of (pointed) metric spaces. We say that the sequence $(Z_i, z_i)$ converges in the (pointed) Hausdorff–Gromov sense to a (pointed) metric space $(Z, z_0)$, if the following holds: For every $r > \varepsilon > 0$, there exist $i_0 \in \mathbb{N}$ and a sequence of (maybe noncontinuous) maps $f_i : B_{Z_i}(z_i, r) \to Z$ $(i > i_0)$ such that

(i) $f_i(z_i) = z_0$,

(ii) $\sup\bigl\{d_{Z'}\bigl(f_i(z_1), f_i(z_2)\bigr) - d_Z(z_1, z_2) \bigm| z_1, z_2 \in Z\bigr\} < \varepsilon$,

(iii) $B_Z(z_0, r - \varepsilon) \subset B_Z\bigl(f_i(B_{Z_i}(z_i, r)), \varepsilon\bigr)$,

(iv) $f_i(B_{Z_i}(z_i, r)) \subset B_Z(z_0, r + \varepsilon)$.

*For the rest of the paper, the term "converges" will stand for "converges in the (pointed) Hausdorff–Gromov sense".*

Let $(Z_i, z_i)$ be a convergent sequence of Alexandrov spaces with the same lower curvature bound $k \in \mathbb{R}$ and the same dimension $n \in \mathbb{N}$. The limit Alexandrov space must have the same lower curvature bound $k$, but can have dimension less than or equal to $n$ (see [Burago et al. 2001, Corollary 10.8.25]). When the limit Alexandrov space has dimension $n$, Perelman's stability theorem (see [Kapovitch 2007]) assures that it is homeomorphic to $Z_i$, for sufficiently large indexes.

It is a fundamental fact that the class of Alexandrov spaces of curvature not smaller than $k \in \mathbb{R}$ is precompact with respect to the Hausdorff–Gromov convergence (see [Gromov 1981, Proposition 5.2] and [Burago et al. 2001, Corollary 10.8.25]). More precisely, every sequence of pointed Alexandrov spaces of curvature not smaller than $k \in \mathbb{R}$ admits a convergent subsequence to an Alexandrov space with the same lower bound for the curvature.

Another important fact concerning Alexandrov spaces is that the Hausdorff–Gromov limit of quasigeodesics is a quasigeodesic (see [Perelman and Petrunin 1994]). More precisely, if $\gamma_i : [a, b] \to Z_i$ is a convergent sequence of quasigeodesics, then the limit curve is a quasigeodesic on the limit space.

## 3. Sequences of hyperbolic cone manifolds

Let $M$ be a closed, orientable and irreducible differential manifold of dimension 3 and let $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_l$ be an embedded link in $M$. A *hyperbolic cone structure* with topological type $(M, \Sigma)$ is a complete intrinsic metric on $M$ such that every nonsingular point (i.e., every point in $M - \Sigma$) has a neighborhood isometric to an open set of $\mathbb{H}^3$, the hyperbolic space of dimension 3, and that every singular point (i.e., every point in $\Sigma$) has a neighborhood isometric to an open neighborhood of a singular point of $\mathbb{H}^3(\alpha)$, the space obtained by identifying the sides of a wedge of angle $\alpha \in (0, 2\pi]$ in $\mathbb{H}^3$ by a rotation about the axis of the wedge. The angles $\alpha$ are called *cone angles* and they may vary from one connected component of $\Sigma$ to the other. We emphasize that we only allow cone angles of at most $2\pi$ in this paper. By convention, the complete hyperbolic structure $M_0$ on $M - \Sigma$ (see [Kojima 1996]) is considered as a hyperbolic cone structure with topological type $(M, \Sigma)$ and cone angles equal to zero.

We point out that every hyperbolic cone manifold is an Alexandrov space of curvature not smaller than $-1$. Furthermore, every geodesic on it is a quasigeodesic.

A natural way to study degenerating deformations of hyperbolic cone structures on $(M, \Sigma)$ is to consider sequences of hyperbolic cone structures converging (in the pointed Hausdorff–Gromov sense) to the limit Alexandrov space. To study these kind of sequences, we need the important notion of collapse which illustrates the intuitive fact that the volume of the sequence may or may not go to zero.

**Definition 12.** We say that a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$ collapses if, for every sequence of points $p_i \in M - \Sigma$, the sequence $r_{\mathrm{inj}}^{M_i - \Sigma}(p_i)$ consisting of their Riemannian injectivity radii in $M_i - \Sigma$ converges to zero. Otherwise, we say that the sequence $M_i$ does not collapse.

When a convergent sequence of hyperbolic cone manifolds collapses, most of the geometric information can be lost. This happens because the dimension of the limit Alexandrov space is strictly smaller than 3 (see [Barreto 2012]). On the noncollapsing case, however, the limit Alexandrov space must have dimension 3 and, in this case, many kinds of geometric information are preserved and can be used to study the deformation.

Given a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$, fix indices $i \in \mathbb{N}$ and $j \in \{1, \ldots, l\}$. For sufficiently small radius $R > 0$, the metric neighborhood

$$B_{M_i}(\Sigma_j, R) = \left\{x \in M_i \;\middle|\; d_{M_i}(x, \Sigma_j) < R\right\}$$

of $\Sigma$ is a solid torus embedded in $M_i$. The supremum of the radius $R > 0$ satisfying the above property will be called *normal injectivity radius of* $\Sigma_j$ *in* $M_i$ and it is going to be denoted by $R_i(\Sigma_j)$. Analogously we can define $R_i(\Sigma)$, the *normal injectivity radius of* $\Sigma$. It is a remarkable fact (see [Fujii 2000; Hodgson and Kerckhoff 2005]) that the existence of a uniform lower bound for $R_i(\Sigma)$ ensures the existence of a sequence of points $p_{i_k} \in M$ such that the sequence $(M_{i_k}, p_{i_k})$ converges to a pointed hyperbolic cone manifold $(M_\infty, p_\infty)$ with topological type $(M, \Sigma)$. Moreover, $M_\infty$ must be compact provided that the cone angles of $M_{i_k}$ are uniformly bounded from below.

Let us also emphasize that the sequence $\mathrm{Vol}(M_i)$ consisting of the Riemannian volumes of the hyperbolic manifolds $M_i - \Sigma$ is always uniformly bounded. More precisely (see [Dunfield 1999; Francaviglia 2004]), we have

$$\mathrm{Vol}(M_i) < \mathrm{Vol}(M_0), \tag{3-1}$$

where $M_0$ denotes the complete hyperbolic manifold that is homeomorphic to $M - \Sigma$.

The purpose of this section is to prove Theorem 4. It is divided into two parts. The first part contains some preliminary results whereas the remaining part deals with the proof of Theorem 4.

Let us point out that, throughout the rest of the paper, the term “component” is going to stand for “connected component”.

**3.1.** ***Preliminary results.*** Let us recall some definitions and elementary results which will be important for the proof of Theorem 4. We will begin with the classification of two-dimensional embedded tori in $M - \Sigma$ (see [Barreto 2012]).

**Lemma 13.** *Suppose that $M - \Sigma$ is hyperbolic and let $T$ be a two-dimensional torus embedded in $M - \Sigma$. Then $T$ separates $M$. Moreover, one and only one of the following statements holds*:

(i) *$T$ is parallel to a component of $\Sigma$ (hence it bounds a solid torus in $M$).*

(ii) *$T$ is not parallel to a component of $\Sigma$ and it bounds a solid torus in $M - \Sigma$.*

(iii) *$T$ is not parallel to a component of $\Sigma$ and it is contained in a ball $B$ of $M - \Sigma$. Furthermore, $T$ bounds a region in $B$ which is homeomorphic to the exterior of a knot in $S^3$.*

We turn to the geometric classification of the thin part of a hyperbolic manifold.

**Definition 14.** Fix $\delta > 0$ and let $M$ be a hyperbolic manifold of dimension 3 (without boundary and perhaps noncomplete). Define the $\delta$-thin part $M_{\mathrm{thin}}(\delta)$ of $M$ by

$$M_{\mathrm{thin}}(\delta) = \left\{q \in M \;\middle|\; r^M_{\mathrm{inj}}(q) < \delta \text{ and } \exp_q \text{ is defined on } B_{T_qM}(0, 3\delta)\right\}.$$

The following result concerning the thin part of hyperbolic manifolds will be needed later.

**Proposition 15.** *Let $M$ be a hyperbolic manifold of dimension* 3 (*without boundary and perhaps noncomplete*) *of finite volume. If $\delta > 0$ is small enough, then each component of $M_{\text{thin}}(\delta)$ contains a maximal region which is isometric to either*

(i) *the quotient of a metric neighborhood of a geodesic $\gamma$ in $\mathbb{H}^3$ by a loxodromic element of* $\mathrm{PSL}_2(\mathbb{C})$ *leaving $\gamma$ invariant and whose translation length is not bigger than $\delta$, or*

(ii) *a parabolic cusp of rank* 2.

*In addition, when* $\mathrm{Vol}(M) < \infty$, *it follows that $M$ has finitely many ends.*

This proposition is a consequence of the existence of a Margulis foliation for the thin part of a hyperbolic manifold. A proof for this proposition is given in [Boileau et al. 2005, Theorem 5.3 and Corollary 5.5] where the authors study the thin part of hyperbolic cone manifolds with topological type $(M, \Sigma)$ and whose cone angles are not bigger than $\pi$. Note that the condition imposed on the cone angles is used only in the description of the singular components of the thin part. We summarize below their proof for the first part of the proposition which, indeed, makes unnecessary the angle condition.

Consider a hyperbolic manifold $M$ and denote by $\pi : \widetilde{M} \to M$ the universal cover of $M$. Let $\delta > 0$ be the constant given by the Margulis lemma (see [Každan and Margulis 1968; Ballmann et al. 1985; Boileau et al. 2005]). Then for every component $\mathcal{P}$ of $M_{\text{thin}}(\delta)$, the stabilizer of a component of $\pi^{-1}(\mathcal{P}) \subset \widetilde{M}$ is an elementary subgroup of $\mathrm{PSL}_2(\mathbb{C})$ generated by a loxodromic element or by at most two parabolic elements. Associated to this group we have a canonical foliation of $\mathbb{H}^3$. The pull-back of this foliation by a developing map gives a foliation on $\pi^{-1}(\mathcal{P})$ which is equivariant by the action of $\pi_1 M$. The quotient of this foliation is the Margulis foliation on $\mathcal{P}$.

To finish the proof, it is sufficient to show that the leaves of this foliation are two-dimensional tori.

First, we remark that the leaves are complete. This is a consequence of the fact that injectivity radius is constant on them (see [Barreto 2009]). When the stabilizer of a component of $\pi^{-1}(\mathcal{P})$ is generated by a loxodromic element, the conclusion follows immediately. In the second case, we need to use the fact that the leaves are flat (they were obtained from horospheres) and the Gauss–Bonnet theorem. The hypothesis that the volume of the manifold is finite excludes undesirable euclidean surfaces other than torus.

**3.2.** ***Proof of Theorem 4.*** The purpose of this section is to study a noncollapsing sequence $M_i$. Without loss of generality, this hypothesis implies the existence of a sequence $p_i \in M - \Sigma$ satisfying

$$r_0 = \inf\left\{r_{\mathrm{inj}}^{M_i}(p_i) \,\middle|\, i \in \mathbb{N}\right\} > 0,$$

and such that the sequence $(M_i, p_i)$ converges to a pointed Alexandrov space $(Z, z_0)$. By definition of the pointed Hausdorff–Gromov convergence, the ball $B_Z(z_0, r_0)$ is isometric to a ball of radius $r_0$ in $\mathbb{H}^3$ and this implies that $Z$ has dimension equal to 3.

We are interested in the case where the length of the singularity remains uniformly bounded, i.e., where

$$\sup\left\{\mathscr{L}_{M_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \text{ and } j \in \{1, \ldots, l\}\right\} < \infty.$$

Fix $j \in \{1, \ldots, l\}$. We can suppose (passing to a subsequence if necessary) that

$$\sup\left\{d_{M_i}(p_i, \Sigma_j) \,\middle|\, i \in \mathbb{N}\right\} < \infty \quad \text{or} \quad \lim_{i\to\infty} d_{M_i}(p_i, \Sigma_j) = \infty.$$

In the first case, we can use again the precompactness to suppose that the component $\Sigma_j \subset M_i$, viewed as a sequence of Alexandrov spaces of dimension 1, converges to a closed curve $\Sigma_j^Z$ in $Z$. Since $Z$ has dimension 3 and it is the limit of a sequence of Alexandrov spaces with same dimension 3 and same lower curvature bound $-1$, we can conclude that $\Sigma_j^Z$ is a quasigeodesic in $Z$ (see [Perelman and Petrunin 1994]).

Summarizing, each component $\Sigma_j$ of $\Sigma$ satisfies one, and only one, of the following statements:

(1) $\sup\left\{d_{M_i}(p_i, \Sigma_j) \,\middle|\, i \in \mathbb{N}\right\} < \infty$ and $\Sigma_j$ converges to a quasigeodesic $\Sigma_j^Z \subset Z$.

(2) $\lim_{i\to\infty} d_{M_i}(p_i, \Sigma_j) = \infty$.

This dichotomy allows us to write $\Sigma = \Sigma_0 \sqcup \Sigma_\infty$, where $\Sigma_0$ contains the components $\Sigma_j$ of $\Sigma$ which satisfy item (1) and $\Sigma_\infty$ those that satisfy item (2).

The following lemma shows that the hypothesis of noncollapsing imposes restrictions on the length and on the cone angles of the singular components of $\Sigma$ contained in $\Sigma_0$.

**Lemma 16.** *Suppose that the sequence $M_i$ does not collapse and let $p_i \in M - \Sigma$ be a sequence of points such that $r_0 = \inf\{r_{\mathrm{inj}}^{M_i}(p_i) \mid i \in \mathbb{N}\} > 0$. If*

$$L = \sup\left\{\mathscr{L}_{M_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \textit{ and } j \in \{1, \ldots, l\}\right\} < \infty,$$

*the following inequalities hold*:

(i) $\inf\left\{\mathscr{L}_{M_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \textit{ and } \Sigma_j \subset \Sigma_0\right\} > 0$.

(ii) $\inf\left\{\alpha_{ij} \,\middle|\, i \in \mathbb{N} \textit{ and } \Sigma_j \subset \Sigma_0\right\} > 0$.

(iii) $\sup\left\{R_i(\Sigma_j) \,\middle|\, i \in \mathbb{N} \textit{ and } \Sigma_j \subset \Sigma_0\right\} < \infty$.

*Proof.* Consider $\mathcal{R} > \sup\{d_{M_i}(p_i, \Sigma_j) \mid i \in \mathbb{N} \text{ and } \Sigma_j \subset \Sigma_0\} + r_0$. By construction, $\mathcal{R} < \infty$ and $B_{M_i}(p_i, r_0) \subset B_{M_i}(\Sigma_j, \mathcal{R})$, for all $i \in \mathbb{N}$ and all components $\Sigma_j$ of $\Sigma_0$.

Fix $i \in \mathbb{N}$ and fix a component $\Sigma_j$ of $\Sigma_0$. Let $\mathcal{A}$ be a region of $\mathbb{H}^3(\alpha_{ij})$ which is bounded by two planes orthogonal to the singular geodesic $\sigma$ of $\mathbb{H}^3(\alpha_{ij})$ and having distance $\mathcal{L}_{M_i}(\Sigma_j)$ between them. Using a developing map for $M_i - \Sigma$ and the minimizing geodesics leaving $\Sigma_j$ orthogonally, the manifold $M_i$ can be developed in a compact domain $K \subset \mathcal{A}$ such that $\mathrm{Vol}(K) = \mathrm{Vol}(M_i)$.

Since $B_{M_i}(p_i, r_0) \subset B_{M_i}(\Sigma_j, \mathcal{R})$, the development of $B_{M_i}(p_i, r_0)$ in $K$ is contained in $B_{\mathbb{H}^3(\alpha_{ij})}(\sigma, \mathcal{R}) \cap \mathcal{A}$. If $V_0$ represents the volume of a ball of radius $r_0$ in $\mathbb{H}^3$, we have

$$V_0 = \mathrm{Vol}\big(B_{M_i}(p_i, r_0)\big) \le \mathrm{Vol}\big(B_{\mathbb{H}^3(\alpha_{ij})}(\sigma, \mathcal{R}) \cap \mathcal{A}\big) = \frac{\alpha_{ij}}{2} \mathcal{L}_{M_i}(\Sigma_j) \sinh^2(\mathcal{R})$$

and therefore

$$\mathcal{L}_{M_i}(\Sigma_j) \ge \frac{V_0}{\pi \sinh^2(\mathcal{R})} > 0 \quad \text{and} \quad \alpha_{ij} \ge \frac{2V_0}{L \sinh^2(\mathcal{R})} > 0.$$

Finally, item (iii) follows from the fact that the sequence $\mathrm{Vol}(M_i)$ is uniformly bounded from above (see (3-1)). □

With the preceding notations, set

$$\Sigma_Z = \bigcup_{\Sigma_j \subset \Sigma_0} \Sigma_j^Z \subset Z.$$

We present now the main result for the noncollapsing:

**Theorem 17** (noncollapsing). *Suppose that there exists a sequence $p_i \in M - \Sigma$ satisfying*

$$r_0 = \inf\big\{r_{\mathrm{inj}}^{M_i}(p_i) \,\big|\, i \in \mathbb{N}\big\} > 0$$

*and such that the sequence $(M_i, p_i)$ converges to a pointed Alexandrov space $(Z, z_0)$ of dimension* 3. *If*

$$\sup\big\{\mathcal{L}_{M_i}(\Sigma_j) \,\big|\, i \in \mathbb{N} \text{ and } j \in \{1, \dots, l\}\big\} < \infty.$$

*Then*:

(i) *$Z - \Sigma_Z$ is a hyperbolic* 3*-manifold of finite volume whose convex and unbounded ends are finite in number and are parabolic cusps of rank* 2.

(ii) *$Z$ is compact* (*and therefore homeomorphic to $M$*) *if and only if $\Sigma_\infty = \varnothing$.*

(iii) *If $Z$ is not compact, there is a bijection between the connected components of $\Sigma_\infty$ and the complete ends of $Z - \Sigma_Z$. In fact, each unbounded end $C_j$ of $Z - \Sigma_Z$ is the Hausdorff–Gromov limit of metric neighborhoods* (*homeomorphic to solid tori*) *$B_{M_i}(\Sigma_j, r_i)$ of a component $\Sigma_j$ of $\Sigma_\infty$, where $r_i > 0$ is an* *increasing sequence going off to infinity. In addition, the cone angles* $\alpha_{ij}$ *and the lengths of these components converge to* 0*.*

*Proof of* (i). According to [Fujii 2000, Lemma 2], every point of $Z-\Sigma_Z$ is the limit of a sequence of points of $M_i-\Sigma$ whose injectivity radius is uniformly bounded from below. This implies that $Z-\Sigma_Z$ is a (without boundary and noncomplete) hyperbolic manifold. Note that the unbounded ends of $Z$ are those of $Z-\Sigma_Z$. In view of Proposition 15, to prove item (i) it is sufficient to show the following:

**Claim.** $\mathrm{Vol}(Z-\Sigma_Z)<\infty$.

*Proof of claim.* Suppose for contradiction the statement is false. Let $K_\infty$ be a compact set of $Z-\Sigma_Z$ whose Riemannian volume is strictly greater than $\mathrm{Vol}(M_0)$, where $M_0$ is $M-\Sigma$ with its complete hyperbolic metric. Since the convergence is bilipschitz on compact subsets (see [Cooper et al. 2000, Theorem 6.20]), there exists an index $i_0\in\mathbb{N}$ and a compact subset $K_{i_0}$ of $M_{i_0}-\Sigma$ (near $K_\infty$) such that

$$\mathrm{Vol}(M_0)<\mathrm{Vol}_{M_{i_0}}(K_{i_0})\le\mathrm{Vol}(M_{i_0}).$$

This is however impossible since $\mathrm{Vol}(M_{i_0})<\mathrm{Vol}(M_0)$ (see (3-1)). This proves the claim, and thus completes the proof of item (i) of Theorem 17. □

*Proof of* (ii) *and* (iii). If $Z$ is compact then $\Sigma_\infty=\varnothing$. Suppose now that $Z$ is not compact. By Lemma 16 we can choose $R>0$ such that

$$B_{M_i}\big(\Sigma_j, R_i(\Sigma_j)\big)\subset B_{M_i}(p_i, R/2)$$

for all connected components $\Sigma_j$ of $\Sigma_0$ and all $i\in N$. Let $K$ be a compact subset of $Z$ which contains the ball $B_Z(z_0,R)$ (and hence $\Sigma_Z$) in its interior and satisfies

$$\mathscr{X}=Z-\mathrm{int}(K)=C_1\sqcup\cdots\sqcup C_m,$$

where each $C_k\approx T^2\times[0,\infty)$ is a cuspidal end of $Z$.

Consider a sequence $C_{1i}=T^2\times[0,t_i]$ of compact subsets of $C_1$, where $t_i>0$ is an unbounded and strictly increasing sequence.

Let $\varepsilon_i>0$ be a sequence converging to zero. Without loss of generality, there exists (according to [Cooper et al. 2000, Theorem 6.20]) a sequence of $(1+\varepsilon_i)$-bilipschitz embeddings $f_{1i}:C_{1i}\to M_i-\Sigma$ onto their images. Therefore, the sequence $B_{1i}=f_{1i}(C_{11})$ converges in the bilipschitz sense to the compact set $C_{11}$.

Consider now a sequence of holonomy representations $\zeta_{1i}:\mathbb{Z}\times\mathbb{Z}\to\mathrm{PSL}_2(\mathbb{C})$ for the hyperbolic structures on the interior sets $B_{1i}$. According to [Cooper et al. 2000, Theorem 6.22], we can assume that

$$\zeta_{1i}\circ(f_{1i})_*\longrightarrow\varphi_1, \tag{3-2}$$

where $\varphi_1 : \mathbb{Z}\times\mathbb{Z} \to \mathrm{PSL}_2(\mathbb{C})$ is a holonomy representation of the hyperbolic structure in the interior of $C_1$ and where $(f_{1i})_* : \mathbb{Z}\times\mathbb{Z} \to \pi_1(M-\Sigma)$ is the canonical homomorphism induced by the map $f_{1i}$.

Consider the torus $T_{1i} = f_{1i}(T^2\times\{0\})$ embedded in $M-\Sigma$. Since $K$ contains the ball $B_Z(z_0, R)$, the torus $T_{1i}$ cannot be parallel to a component $\Sigma_j$ of $\Sigma_0$. For $i$ sufficiently large, the torus $T_{1i}$ cannot be contained in a ball of $M-\Sigma$. To see this, consider a homotopically nontrivial loop $\gamma_1$ on $T^2\times\{0\}\subset C_{11}$. Since $C_1$ is a parabolic cusp, $\varphi_1(\gamma_1)$ is a nontrivial parabolic element of $\mathrm{PSL}_2(\mathbb{C})$ and therefore the convergence (3-2) implies that $\zeta_{1i}\circ(f_{1i})_*(\gamma_1)$ is not trivial for $i$ very large. The same then holds for the sequence $(f_{1i})_*(\gamma_1)$.

According to Lemma 13, we can suppose that the torus $T_{1i}$ bounds a solid torus $W_{1i}$ in $M$ (with perhaps a singular soul). Note that

$$\lim_{i\to\infty} \mathrm{diam}_{M_i}(W_{1i}) = \infty, \tag{3-3}$$

because $f_{1i}(C_{1i})\subset W_{1i}$, for all $i\in\mathbb{N}$.

We can repeat the same construction for each cusp $C_k$ of $\mathcal{X}$ in order to obtain sequences of embedded tori $T_{ki}\subset M-\Sigma$ ($k\in\{1,\dots,m\}$ and $i\in\mathbb{N}$), each of then bounds solid torus $W_{ki}$ in $M-\Sigma_0$. Furthermore whose diameters become infinite with $i$. This yields a sequence of 3-manifolds with torus boundary

$$\mathcal{M}_i = M_i - \bigcup_{k=1}^{m} W_{ki}$$

such that $M$ can be obtained by Dehn filling on their boundary components. By construction, the sequence $\mathcal{M}_i$ converges to the compact $K$ and then (by Perelman's stability theorem [Kapovitch 2007]), we can assume that the manifolds $\mathcal{M}_i$ are all homeomorphic to $K$.

For all $i\in\mathbb{N}$ and all $k\in\{1,\dots,m\}$, fix a homotopically nontrivial loop $\mu_{ki}$ in $T^2\times\{0\}\subset C_k$ satisfying:

- the loop $f_{ki}\circ\mu_{ki}$ bounds a disc in $W_{ki}$,
- if, for some index $j\in\mathbb{N}$, a loop $\mu_{kj}$ belongs to the same homotopy class of the loop $\mu_{ki}$, then $\mu_{kj}=\mu_{ki}$.

The rest of the proof is going to be divided in two cases depending on whether or not $\Sigma_0$ is empty.

<u>First case:</u> $\Sigma_0=\varnothing$. Since the link $\Sigma$ was supposed to be nonempty, it follows that $\Sigma_\infty\neq\varnothing$. Since the distance between $p_i$ and $\Sigma_\infty$ becomes infinite, we can assume that $\Sigma_\infty$ is contained in the complement of $\mathcal{M}_i$. More precisely, we can also assume (see Lemma 13) that each solid torus of $M_i-\mathcal{M}_i$ contains at most one component

of $\Sigma_\infty$ and, in the latter case, this component corresponds to the soul of the solid torus in question.

The singular set $\Sigma_\infty$ has a finite number of components. Passing to a subsequence if necessary, we obtain an one-to-one map which associates each component $\Sigma_j$ of $\Sigma_\infty$ to a component $C_{k_j}$ of $\mathscr{X}$, that is, the component $\Sigma_j$ is contained in the component $W_{k_j i}$ of $M_i - \mathcal{M}_i$, for all $i \in \mathbb{N}$.

Recall that $\lim_{i\to\infty} d_{M_i}(p_i, \Sigma_j) = \infty$ for every connected component $\Sigma_j$ of $\Sigma_\infty$. Since the tori $T_{k_j i}$ remain at a finite distance to the points $p_i$ and they are parallel to the components $\Sigma_j$, we must have $\lim_{i\to\infty} R_i(\Sigma_j) = \infty$.

Since $\Sigma_0 = \varnothing$ and thanks to [Fujii 2000, Theorem 1], the cone angles of $\Sigma$ converge to zero and $Z$ has a complete hyperbolic structure whose ends are associated with components of $\Sigma_\infty$. In other words, the injection defined above between the components of $\Sigma_\infty$ and the components of $\mathscr{X}$ is, indeed, a bijection.

<u>Second case:</u> $\Sigma_0 \neq \varnothing$. Denote by $\Lambda$ the subset of $\{1, \dots, m\}$ containing the indices that are not associated with components of $\Sigma_\infty$. Denote also by $\Omega$ the subset of $\{1, \dots, m\}$ containing the indices that are associated with components of $\Sigma_\infty$ whose sequence of cone angles does not converge to zero.

**Lemma 18.** *There exist $i_0 \in \mathbb{N}$ satisfying: for each $k \in \Lambda \cup \Omega$, the homotopy classes of loops $\mu_{ki}$ $(i > i_0)$ are pairwise distinct.*

*Proof.* Suppose for a contradiction that the statement of the lemma does not hold. Without loss of generality, there exists $k_0 \in \Lambda \cup \Omega$ such that all loops $\mu_{k_0 i}$ $(i \in \mathbb{N})$ belongs to the same homotopy class. By construction, this implies that the loops $\mu_{k_0 i}$ $(i \in \mathbb{N})$ are the same loop, say $\mu$.

Suppose first that $k_0 \in \Lambda$. By construction,

$$\zeta_{k_0 i} \circ (f_{k_0 i})_*(\mu) = \zeta_{k_0 i}(f_{k_0 i} \circ \mu) = 1_{\mathrm{PSL}_2(\mathbb{C})}, \tag{3-4}$$

for all $i \in \mathbb{N}$. Because $\varphi_{k_0}([\mu])$ is a nontrivial parabolic element of $\mathrm{PSL}_2(\mathbb{C})$, we have a contradiction.

Suppose now that $k_0 \in \Omega$. Then $k_0 = k_j$, for some component $\Sigma_j$ of $\Sigma_\infty$ whose sequence of cone angles converges to $\alpha_{\infty j} \neq 0$. Since the maps $f_{k_0 i}$ are $(1+\varepsilon_i)$-bilipschitz embeddings (with $\varepsilon_i$ shrinks down to zero), the loops $f_{k_0 i} \circ \mu$ must have bounded lengths.

As noted in the preceding case, the sequence $R_i(\Sigma_j)$ of the normal injectivity radii of the component $\Sigma_j$ goes off to infinity. Since $\alpha_{\infty j} \neq 0$, the sequence $\mathscr{L}_{M_i}(f_{k_0 i} \circ \mu)$ formed by the lengths of the loops $f_{k_0 i} \circ \mu$ cannot be bounded. This is a contradiction with the above paragraph. □

As a consequence of the above lemma, we will show that the set $\Lambda \cup \Omega$ is empty. To do this, the following lemma will be needed:

**Lemma 19.** *Given $k \in \Lambda$, there exists $i_0 = i_0(k) \in N$ such that the solid tori $W_{ki}$ contains a simple closed geodesic $\sigma_{ki}$, for every $i > i_0$.*

*Proof.* Fix $k \in \Lambda$ and let

$$\delta = \tfrac{1}{2} \inf\bigl\{ r_{\text{inj}}^{Z-\Sigma_Z}(z) \bigm| z \in C_{k1} \bigr\} > 0.$$

Since the map $f_{ki}|_{C_{k1}} : C_{k1} \to B_{ki}$ becomes closer and closer to isometries, there exists $i_1 \in \mathbb{N}$ such that

$$r_{\text{inj}}^{M_i}(q) > \delta,$$

for all $i > i_1$ and for all $q \in B_{ki}$ (in particular, for all $q \in T_{ki}$).

**Claim.** *There is $i_2 \in \mathbb{N}$ such that, for all $i > i_2$, we can find a loop $\gamma_{ki}$ in $W_{ki}$ which is homotopically nontrivial in the interior $M - \Sigma$ and has length smaller than $\delta$.*

*Proof of claim.* Consider the loops consisting of two geodesic segments with same ends and equal lengths which, furthermore, are smaller than $\delta/2$. These loops are always homotopically nontrivial; otherwise we would obtain, after development, two distinct geodesic arcs with the same ends and equal lengths in $\mathbb{H}^3$, which is not possible.

The fact that $W_{ki}$ does not admit this type of loop in its interior is equivalent to saying that all points of $W_{ki}$ have injectivity radius not smaller than $\delta/2$. This is a contradiction because the sequence $\mathrm{Vol}(M_i)$ is uniformly bounded from above (see (3-1)) and the diameter of components $W_{ki}$ becomes infinite. This proves the claim.

Consider $i_o = \max\{i_1, i_2\}$ and fix $i > i_0$. Let $\gamma_{ki} \subset W_{ki}$ be a loop as above. By [Kojima 1998, Lemma 1.2.4], the loop $\gamma_{ki}$ is freely homotopic (in $M - \Sigma$) to a closed geodesic $\sigma_{ki} \subset M - \Sigma$. Moreover, the length of $\sigma_{ki}$ is smaller than $\delta$ because the length of loops is strictly decreasing along this homotopy. Because the points of the torus $T_{ki}$ have injectivity radius bigger than $\delta$, all the loops involved in this homotopy must lie entirely in the interior of $W_{ki}$. In particular, $\sigma_{ki} \subset W_{ki}$.

If $\sigma_{ki}$ is not simple, then it gives rise to a loop $\gamma'_{ki}$ consisting of two geodesic segments with same ends and equal lengths which are smaller than $\delta/4$. This implies that the injectivity radius of the ends of $\gamma'_{ki}$ is smaller than $\delta/4$. We can apply the same construction for the loop $\gamma'_{ki}$ in order to obtain a new closed geodesic $\sigma_{ki} \subset W_{ki}$ whose length is smaller than $\delta/4$. Since the injectivity radius of points of $W_{ki}$ bounded from below by compactness, this process must end after a finite number of steps and therefore we can suppose that $\sigma_{ki}$ is simple. This completes the proof of Lemma 19. □

The following lemma shows that $\Sigma_\infty$ is not empty and the cone angles of its components goes to zero. Moreover the map between the components of $\Sigma_\infty$ and the components of $\mathfrak{X}$ must be a bijection.

**Lemma 20.** *The set $\Lambda \cup \Omega$ is empty.*

*Proof.* According to the above lemma, we can suppose there exists a simple closed geodesic $\sigma_{ki}$ in the solid torus $W_{ki}$, for every $i \in \mathbb{N}$ and every $k \in \Lambda$. If the manifolds $M_i$ are regarded as hyperbolic cone manifolds with topological type $(M, \Sigma')$, where

$$\Sigma' = \Sigma \cup \bigcup_{k \in \Lambda} \sigma_{ki}$$

and the cone angles on the geodesics $\sigma_{ki}$ are equal to $2\pi$, it follows from Lemma 13 that the tori $T_{ki}$ are parallel to the geodesics $\sigma_{ki}$. In addition, $M - \Sigma'$ admits a complete hyperbolic structure (see [Kojima 1996]) that will be denoted by $\mathcal{M}_0$.

For all $i \in \mathbb{N}$ and all $k \in \Lambda$, denote the homotopy class of the loop $\mu_{ik}$ by $(p_{ki}, q_{ki}) \in \mathbb{Z} \times \mathbb{Z} \approx \pi_1 C_k$. Without loss of generality, the Thurston's hyperbolic Dehn surgery [Cooper et al. 2000, Theorem 1.13] gives a sequence of complete hyperbolic manifolds $\mathcal{M}(p_{i1}, q_{i1}, \ldots, p_{im}, q_{im})$ diffeomorphic to $M - \Sigma$ and such that

$$V_i := \mathrm{Vol}(\mathcal{M}(p_{1i}, q_{1i}, \ldots, p_{mi}, q_{mi})) < \mathrm{Vol}(\mathcal{M}_0), \tag{3-5}$$

where $(p_{ki}, q_{ki}) = \infty$, for all $i \in \mathbb{N}$ and all $k \in \{1, \ldots, m\} - \Lambda$.

Since, for each $k \in \Lambda$, the pairs $(p_{ki}, q_{ki})_{i \in \mathbb{N}}$ are pairwise distinct (the homotopy classes of $\mu_{ik}$ are pairwise distinct), a subsequence $\mathcal{M}(p_{1i_s}, q_{1i_s}, \ldots, p_{mi_s}, q_{mi_s})$ such that

$$\lim_{s \to \infty} \|(p_{ki_s}, q_{ki_s})\| = \lim_{s \to \infty} (p_{ki_s})^2 + (q_{ki_s})^2 = \infty \quad \text{for every } k \in \Lambda$$

always exists. Thurston's hyperbolic Dehn surgery then gives

$$\lim_{s \to \infty} V_{i_s} = \mathrm{Vol}(\mathcal{M}_0). \tag{3-6}$$

Recall that the Riemannian volume of a complete hyperbolic manifold with finite volume is a topological invariant (Mostow's theorem). Since the manifolds $\mathcal{M}(p_{i1}, q_{i1}, \ldots, p_{im}, q_{im})$ are diffeomorphic, the sequence $V_i$ must be constant. This contradicts the statements (3-5) and (3-6). Hence $M_i - \mathcal{M}_i$ cannot have nonsingular components. Therefore, $\Sigma_\infty \neq \varnothing$ and the map between the components of $\Sigma_\infty$ and the components of $\mathcal{X}$ is a bijection. This proves Lemma 20, and thus completes the proof of items (ii) and (iii) of Theorem 17. $\square$

**Corollary 21.** *Suppose that the sequence $M_i$ does not collapse and verifies*

$$\sup\left\{\mathcal{L}_{M_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \text{ and } j \in \{1, \ldots, l\}\right\} < \infty.$$

*If there is $\varepsilon \in (0, 2\pi)$ such that the cone angles $\alpha_{ij}$ belong to $(\varepsilon, 2\pi]$, then there exists a sequence of points $p_{i_k} \in M - \Sigma$ such that the sequence $(M_{i_k}, p_{i_k})$ converges* *to a compact and 3-dimensional pointed Alexandrov space $(Z, z_0)$ (in fact homeomorphic to $M$). Moreover, there exists a finite union of quasigeodesics $\Sigma_Z$ such that $Z - \Sigma_Z$ is a noncomplete hyperbolic manifold of finite volume.*

**Remark 22.** Suppose that $\Sigma$ is not connected. If $(M_i, p_i)$ is a sequence as in the statement of Theorem 17, then the inequality

$$\sup\left\{\operatorname{diam}_{M_i}(\Sigma) \;\middle|\; i \in \mathbb{N}\right\} < \infty$$

is a necessary and sufficient condition to ensure that the sequence $\operatorname{diam}(M_i)$ remains bounded.

We have also the following less immediate corollary:

**Corollary 23.** *Let $M$ be a closed, orientable and irreducible 3-manifold and let $\Sigma$ be an embedded link in $M$. Assume that there exists a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$ and having the same cone angles $\alpha_i \in (0, 2\pi]$ for all components of $\Sigma$. Then there is a pointed subsequence $M_{i_k}$ converging to $M_0$ ($M - \Sigma$ with its complete hyperbolic metric) if and only if the following conditions hold*:

(i) $\sup\left\{\mathcal{L}_{M_i}(\Sigma) \;\middle|\; i \in \mathbb{N}\right\} < \infty$.

(ii) $\sup\left\{\operatorname{diam}(M_i) \;\middle|\; i \in \mathbb{N}\right\} = \infty$.

(iii) *The sequence $M_i$ does not collapse.*

*Proof.* By Kojima's result [1998], the existence of a subsequence $M_{i_k}$ converging to $M_0$ is equivalent to the convergence of the cone angles $\alpha_{i_k}$ to zero.

Suppose that the sequence $\alpha_i$ converges to zero. Without loss of generality, we can assume that $\alpha_i \in (0, \pi]$, for every $i \in \mathbb{N}$. According to [Kojima 1998], there exists a continuous path (parametrized by cone angles) of hyperbolic cone structures with topological type $(M, \Sigma)$ which connects the hyperbolic cone structure of $M_0$ to the complete hyperbolic structure on $M - \Sigma$. Moreover, by uniqueness of the hyperbolic cone structures with cone angles not bigger than $\pi$ (see [Kojima 1998]), this path contains the hyperbolic cone structures of $M_i$, for every $i \in \mathbb{N}$. Then for every point $p \in M$, the sequence $(M_i, p)$ converges to $(M - \Sigma, p)$ with the complete hyperbolic structure. This implies items (ii) and (iii). Item (i) is a consequence of Thurston's hyperbolic Dehn surgery theorem which implies that the sequence $\mathcal{L}_{M_i}(\Sigma)$ converges to zero.

Conversely, suppose now that items (i), (ii) and (iii) are true. Then there exists a sequence of points $p_{i_k} \in M - \Sigma$ satisfying

$$\inf\left\{r_{\mathrm{inj}}^{M_i}(p_{i_k}) \;\middle|\; k \in \mathbb{N}\right\} > 0$$

and such that the sequence $(M_{i_k}, p_{i_k})$ converges to a noncompact and 3-dimensional pointed Alexandrov space $(Z, z_0)$. Corollary 21 then shows that the sequence $\alpha_i$ must converge to zero. □

## 4. Applications

**4.1.** ***Small links.*** An embedded link $\Sigma$ in a 3-manifold $M$ is called small (in $M$) if it has an open tubular neighborhood $U$ such that $M - U$ does not contain an embedded essential surface whose boundary is empty or an union of meridians of $\Sigma$. An important fact due to W. Thurston and A. Hatcher [1985, Lemma 3] is that every 3-manifold containing a small link does not admit an embedded essential surface.

Given a 3-manifold $M$, let $\Sigma$ be an embedded link in $M$. Suppose there exists a sequence $M_i$ of hyperbolic cone manifolds with topological type $(M, \Sigma)$ and consider the sequence $\mathscr{L}_{M_i}(\Sigma)$ formed by the lengths of the singular set $\Sigma$ in $M_i$. As a consequence of the Culler–Shalen theory [1983], the holonomy representations of $M_i$ are convergent. Therefore, we have the following proposition:

**Proposition 24.** *Let $M_i$ be a sequence of hyperbolic cone manifolds with topological type $(M, \Sigma)$. If $\Sigma$ is a small link in $M$, then*

$$\sup\left\{\mathscr{L}_{M_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \text{ and } \Sigma_j \text{ component of } \Sigma\right\} < \infty.$$

When $\Sigma$ is a small link in $M$, Theorem 4 yields the following corollaries:

**Corollary 25.** *Suppose that $M$ is a closed, orientable, irreducible and non-Seifert fibered 3-manifold and let $\Sigma$ be an embedded small link in $M$. Then there exists a constant $V = V(M, \Sigma) > 0$ such that $\mathrm{Vol}(\mathcal{M}) > V$, for every hyperbolic cone manifold $\mathcal{M}$ with topological type $(M, \Sigma)$ and having cone angles of at most $2\pi$.*

*Proof.* First note that $M$ is not a Sol manifold. In fact every Sol manifold is foliated by essential two-dimensional tori and this is not possible since $\Sigma$ is small (see [Hatcher and Thurston 1985, Lemma 3]).

Suppose that the lower bound $V$ does not exist. Since $\Sigma$ is small in $M$, the nonexistence of $V$ implies the existence of a sequence of hyperbolic cone manifolds $\mathcal{M}_i$ with topological type $(M, \Sigma)$ satisfying

- $\sup\left\{\mathscr{L}_{\mathcal{M}_i}(\Sigma_j) \,\middle|\, i \in \mathbb{N} \text{ and } \Sigma_j \text{ component of } \Sigma\right\} < \infty$,
- the sequence $\mathrm{Vol}(\mathcal{M}_i - \Sigma)$ formed by the Riemannian volumes of the hyperbolic manifolds $\mathcal{M}_i - \Sigma$ shrinks down to zero (and therefore the sequence $\mathcal{M}_i$ collapses).

According to Theorem 4, $M$ must be Seifert fibered, contradicting our hypothesis. □

**Corollary 26.** *Suppose that $M$ is a closed, orientable, irreducible and non-Seifert fibered 3-manifold and let $\Sigma$ be an embedded small link in $M$. Given $\varepsilon \in (0, 2\pi)$, there is a constant $K = K(M, \varepsilon) > 0$ such that* $\operatorname{diam}(\mathcal{M}) < K$, *for every hyperbolic cone manifold $\mathcal{M}$ with topological type $(M, \Sigma)$ and having cone angles belonging to $(\varepsilon, 2\pi]$.*

*Proof.* As seen in the previous corollary, $M$ is not a Sol manifold. Fix $\varepsilon \in (0, 2\pi)$ and suppose that the upper bound $K$ does not exist. Since $\Sigma$ is small in $M$, the nonexistence of $K$ implies the existence of a sequence of hyperbolic cone manifolds $\mathcal{M}_i$ with topological type $(M, \Sigma)$, having cone angles $\alpha_{ji} \in (\varepsilon, 2\pi]$ and satisfying these conditions:

(i) $\sup\{\mathcal{L}_{\mathcal{M}_i}(\Sigma_j) \mid i \in \mathbb{N} \text{ and } \Sigma_j \text{ component of } \Sigma\} < \infty$.

(ii) The sequence $\operatorname{diam}(\mathcal{M}_i)$ formed by the diameters of the hyperbolic cone manifolds $\mathcal{M}_i$ go to infinity.

Since $M$ is neither Seifert fibered nor a Sol manifold, it follows from item (i) and Theorem 4 that the sequence $\mathcal{M}_i$ does not collapse. Moreover, since the cone angles $\alpha_{ji}$ belong to $(\varepsilon, 2\pi]$, it follows that the sequence $\operatorname{diam}(\mathcal{M}_i)$ is bounded and this yields a contradiction with item (ii). □

**4.2. *Proof of Corollary 6.*** First, we would like to recall that the existence of a deformation $M_\alpha$ as in Corollary 6 is a consequence of the local deformation theorem due to [Hodgson and Kerckhoff 1998].

*Proof.* The implication (i) ⇒ (ii) is immediate (see [Kojima 1998]). Suppose now that the sequence $\mathcal{L}_{M_\alpha}(\Sigma)$ converges to 0 when $\alpha$ converges to $\theta$. Then

$$\sup\{\mathcal{L}_{M_{\alpha_i}}(\Sigma_j) \mid i \in \mathbb{N} \text{ and } \Sigma_j \text{ component of } \Sigma\} < \infty,$$

for every sequence $\alpha_i \in (\theta, 2\pi]$ converging to $\theta$. Consider such a sequence $\alpha_i$. Since $M$ is hyperbolic (and therefore is neither Seifert fibered nor a Sol manifold), it follows from Theorem 4 that the sequence $M_{\alpha_i}$ does not collapse. Moreover, since the sequence $\mathcal{L}_{M_{\alpha_i}}(\Sigma)$ converges to zero, we must have $\lim_{i\to\infty} \operatorname{diam}(M_{\alpha_i}) = \infty$. This concludes the proof of the implication (ii) ⇒ (iii).

To prove (iii) ⇒ (i) take a sequence $\alpha_i$ satisfying item (iii). Again by Theorem 4, it follows that the sequence $M_{\alpha_i}$ does not collapse. Moreover, since the sequence $\operatorname{diam}(M_{\alpha_i})$ is not bounded, we must have $\theta = 0$ because all the components of $\Sigma$ have the same cone angle. Then, by [Kojima 1998], it follows that $M_i$ converges to $M_0$. □

## References

[Ballmann et al. 1985] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of nonpositive curvature*, Progress in Mathematics **61**, Birkhäuser, Boston, 1985. MR 87h:53050 Zbl 0591.53001

[Barreto 2009] A. P. Barreto, *Déformations de structures hyperboliques coniques*, thesis, Université de Toulouse, 2009, Available at http://thesesups.ups-tlse.fr/838/1/Paiva-Barreto_Alexandre.pdf.

[Barreto 2012] A. P. Barreto, "Deformation of hyperbolic cone-structures: study of the collapsing case", preprint, 2012. arXiv 1201.2923

[Boileau et al. 2005] M. Boileau, B. Leeb, and J. Porti, "Geometrization of 3-dimensional orbifolds", *Ann. of Math.* (2) **162**:1 (2005), 195–290. MR 2007f:57028 Zbl 1087.57009

[Burago et al. 1992] Y. Burago, M. Gromov, and G. Perel'man, "пространства А. Д. Александрова с ограниченными снизу кривизнами", *Uspekhi Mat. Nauk* **47**:2(284) (1992), 3–51. Translated as "A. D. Alexandrov spaces with curvature bounded below" in *Russ. Math. Surv.* **47**:2 (1992), 1–58. MR 93m:53035 Zbl 0802.53018

[Burago et al. 2001] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, Amer. Math. Soc., Providence, RI, 2001. MR 2002e:53053 Zbl 0981.51016

[Cooper et al. 2000] D. Cooper, C. D. Hodgson, and S. P. Kerckhoff, *Three-dimensional orbifolds and cone-manifolds*, MSJ Memoirs **5**, Math. Soc. Japan, Tokyo, 2000. MR 2002c:57027 Zbl 0955.57014

[Culler and Shalen 1983] M. Culler and P. B. Shalen, "Varieties of group representations and splittings of 3-manifolds", *Ann. of Math.* (2) **117**:1 (1983), 109–146. MR 84k:57005 Zbl 0529.57005

[Dunfield 1999] N. M. Dunfield, "Cyclic surgery, degrees of maps of character curves, and volume rigidity for hyperbolic manifolds", *Invent. Math.* **136**:3 (1999), 623–657. MR 2000d:57022 Zbl 0928.57012

[Francaviglia 2004] S. Francaviglia, "Hyperbolic volume of representations of fundamental groups of cusped 3-manifolds", *Int. Math. Res. Not.* **2004**:9 (2004), 425–459. MR 2004m:57032 Zbl 1088.57015

[Fujii 2000] M. Fujii, "A cone angle condition on strong convergence of hyperbolic 3-cone-manifolds", *RIMS Kōkyūroku* **1163** (2000), 126–131. MR 1799375 Zbl 0969.57509

[Gromov 1981] M. Gromov, *Structures métriques pour les variétés Riemanniennes*, Textes Mathématiques **1**, CEDIC, Paris, 1981. MR 85e:53051 Zbl 0509.53034

[Hatcher and Thurston 1985] A. Hatcher and W. Thurston, "Incompressible surfaces in 2-bridge knot complements", *Invent. Math.* **79**:2 (1985), 225–246. MR 86g:57003 Zbl 0602.57002

[Hodgson and Kerckhoff 1998] C. D. Hodgson and S. P. Kerckhoff, "Rigidity of hyperbolic cone-manifolds and hyperbolic Dehn surgery", *J. Differential Geom.* **48**:1 (1998), 1–59. MR 99b:57030 Zbl 0919.57009

[Hodgson and Kerckhoff 2005] C. D. Hodgson and S. P. Kerckhoff, "Universal bounds for hyperbolic Dehn surgery", *Ann. of Math.* (2) **162**:1 (2005), 367–421. MR 2006g:57031 Zbl 1087.57011

[Kapovitch 2007] V. Kapovitch, "Perelman's stability theorem", pp. 103–136 in *Metric and comparison geometry*, edited by J. Cheeger and K. Grove, Surv. Differ. Geom. **11**, International Press, Somerville, MA, 2007. MR 2009g:53057 Zbl 1151.53038

[Każdan and Margulis 1968] D. A. Każdan and G. A. Margulis, "Доказательство гипотезы Сельберга", *Mat. Sb.* (*N.S.*) **75**(117):1 (1968), 163–168. Translated as "A proof of Selberg's hypothesis" in *Math. USSR Sb.* **4**:1 (1968), 147–152. MR 36 #6535

[Kojima 1996] S. Kojima, "Nonsingular parts of hyperbolic 3-cone-manifolds", pp. 115–122 in *Topology and Teichmüller spaces* (Katinkulta, 1995), edited by S. Kojima et al., World Scientific, River Edge, NJ, 1996. MR 99i:57028 Zbl 0928.57010

[Kojima 1998] S. Kojima, "Deformations of hyperbolic 3-cone-manifolds", *J. Differential Geom.* **49**:3 (1998), 469–516. MR 2000d:57023 Zbl 0990.57004

[Perelman and Petrunin 1994] G. Perelman and A. Petrunin, "Quasigeodesics and gradient curves in Alexandrov spaces", 1994, Available at http://www.math.psu.edu/petrunin/papers/qg_ams.pdf.

Received January 19, 2012. Revised October 23, 2013.

ALEXANDRE PAIVA BARRETO
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE SÃO CARLOS
ROD. WASHINGTON LUÍS, KM 235
C.P. 676
13565-905 SÃO CARLOS
BRAZIL
alexandre@dm.ufscar.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.23

# A TRANSPORT INEQUALITY ON THE SPHERE OBTAINED BY MASS TRANSPORT

DARIO CORDERO-ERAUSQUIN

**Using McCann's transportation map, we establish a transport inequality on compact manifolds with positive Ricci curvature. This inequality contains the sharp spectral comparison estimates.**

## 1. Introduction

Extending the mass transportation approach to sharp Sobolev-type inequalities from Euclidean space to curved geometries remains a challenging problem. In the present note, we propose a new twist in the classical transportation technique that allows for a transport inequality which contains sharp Poincaré inequalities.

The method applies to a (compact) Riemannian manifold of dimension $n \geq 2$ having a lower bound on the Ricci curvature of the form $\mathrm{Ric} \geq (n-1)k^2\mathrm{g}$ with $k > 0$ and g the Riemannian metric. By scaling the distances, we can always assume that $k = 1$.

So, in the rest of the paper $M = (M, \mathrm{g})$ will stand for an $n$-dimensional Riemannian manifold satisfying

$$\mathrm{Ric} \geq (n-1)\mathrm{g}. \tag{1}$$

The main example is the usual sphere $S^n \subset \mathbb{R}^{n+1}$. The interest, perhaps, in stating a result under the condition (1), even if one aims at the sphere only, is that it makes it clear that we will not use any of the algebraic properties of the sphere. Our computations are modeled on the sphere case; the extension to the situation given by (1) relies on Bishop comparison's estimates only. We will denote by $d\sigma = d\,\mathrm{vol}\,/\,\mathrm{vol}(M)$ the Riemannian volume measure normalized to be a probability measure. The distance will be denoted $d$; recall as well that $M$ has diameter smaller than $\pi$.

A simple but important result is that, on such manifold $M$, the spectral gap for the Laplacian satisfies $\lambda_1 \geq n$. Equivalently, one has the following Wirtinger–Poincaré

*MSC2010:* primary 49Q20, 53C21; secondary 60E15.
*Keywords:* transport inequality, Ricci curvature, optimal mass transport.

inequality: for every Lipschitz function $g$ on $M$,

$$\mathrm{Var}_\sigma(g) := \int \left(g - \int g \, d\sigma\right)^2 d\sigma \le \frac{1}{n} \int |\nabla g|^2 \, d\sigma. \tag{2}$$

The $L^2$ proof of this inequality as done by Lichnerovicz using Bochner's formula is rather short and elementary. In the particular case of the sphere, one can also use the expansion of $g$ in the spherical harmonics basis; moreover, in this case, equality holds for linear functions, which are eigenfunctions for the spherical Laplacian.

It is well known that Poincaré inequalities are not well suited to mass transport techniques. However, in the Euclidean case and under appropriate curvature assumptions, one can prove very easily using mass transport (Brenier map) *stronger* inequalities such as *transport inequalities* or *logarithmic Sobolev inequalities* (see [Cordero-Erausquin 2002]). So it is quite annoying that no mass transport proof of the sharp log-Sobolev inequality (see [Ledoux 2000]), say, is available on $M$. Indeed, the straightforward adaptation of the techniques from Euclidean space leads to a log-Sobolev inequality with a constant $(n-1)$ in place of the expected constant $n$. Similarly, the transport inequality (definitions are recalled below) that one gets by standard techniques is as follows: for every $f \ge 0$ on $M$ with $\int f \, d\sigma = 1$,

$$\mathcal{W}_c(f \, d\sigma, \sigma) \le \int f \log f \, d\sigma \tag{3}$$

for the cost $c(d) := \frac{1}{2}(n-1)d^2$. Linearization of this inequality gives only a weak form of (2) with $1/(n-1)$ in place of the correct $1/n$. Let us note that by an abstract result of Otto and Villani [2000], the log-Sobolev inequality mentioned above with the sharp constant $n$ implies that the transport inequality (3) holds with the cost $c(d) := \frac{1}{2}nd^2$. As for the log-Sobolev inequality, it is not known how to reach this inequality using mass transport.

The difficulty is to properly quantify the interplay between dimension and nonzero curvature in the mass transportation techniques.

This was partly overcome in [Lott and Villani 2007] (see also [Villani 2009, Chapter 20 and 21]). There, the authors manage to prove some Sobolev-like inequalities under the so called "curvature-dimension condition $CD(K,n)$" that imply, after linearization, sharp spectral bounds. To be precise, their assumption is that the metric measure space $(M, d, \sigma)$ satisfies a curvature-dimension lower bound which is defined in terms of uniform convexity along optimal transport of a class of entropy functionals. From this assumption, they deduce a (not very natural) Sobolev-like inequality. This inequality has no reason to be sharp when the curvature is nonzero, but after linearization it gives the correct Poincaré inequality (2) (so in a sense it is sharp at first order). Of course, it is known, by the properties of optimal transport on manifolds (McCann's map), that a Riemannian manifold with condition (1) satisfies the curvature-dimension criterion. So putting all together, we see that Lott and Villani's work is already an answer to the question on how to use mass transport to derive some sharp dimensional inequalities. But of course, it is rather indirect, and no standard inequality that one could prove using optimal transport on a manifold is easy to extract from it. Actually, this is somehow the content of Open Problem 21.11 in [Villani 2009].

Our original motivation was to provide, in the particular case of a manifold, a different, more direct, approach based on the geometric properties of McCann's transport map. The aim was to find an inequality that contained the sharp bound (2). Eventually, we managed to establish a new, suitable, transport inequality, that is an inequality between an entropy functional and a transportation cost functional (we recommend the survey [Gozlan and Léonard 2010] for background on transport inequalities). The question of obtaining the sharp log-Sobolev inequality using mass transport remains.

Let us introduce the following classical dimensional entropy: given a probability density $f$ on $M$, meaning a Borel nonnegative function on $M$ with $\int f\, d\sigma = 1$, we put

$$H_{n,\sigma}(f) := n \int \left(f - f^{1-1/n}\right) d\sigma = n - n \int f^{1-1/n}\, d\sigma.$$

Note that $H_{n,\sigma}$ is a nonnegative convex functional of $f$.

We will consider transportation costs given by functions of the distance $d$ on $M$. Given a function $c : \mathbb{R} \to \mathbb{R}^+$ (or rather $c : [0, \pi] \to \mathbb{R}^+$ in our case), the associated Kantorovich transportation cost between two Borel probability measures $\mu$ and $\nu$ on $M$ is defined by

$$\mathcal{W}_c(\mu, \nu) := \inf_{\pi} \iint c(d(x, y))\, d\pi(x, y)$$

where the infimum is taken over all probability measures $\pi$ on $M \times M$ projecting on $\mu$ and $\nu$, respectively.

In the proof of the Theorem below, will use McCann's map, which arises from an optimizer in the functional $\mathcal{W}_c$ when $c$ is the quadratic cost, $c(d) = d^2/2$; we shall recall McCann's result in detail later. However, let us emphasize that, although we will use this quadratic-optimal map, the cost in our transport inequality will be a different function of the distance.

Our cost function is defined for $d \in [0, \pi)$ by

$$c_n(d) := n - \frac{\sin^{n-1} d}{\mathrm{S}_n(d)^{n-1}} - (n-1)\frac{\mathrm{S}_n(d)}{\tan d}$$

and at the limit by $c_n(\pi) = +\infty$, where $\mathrm{S}_n$ is the familiar function defined for

$d \in [0, \pi]$ by

$$\mathrm{S}_n(d) := \left( n \int_0^d \sin^{n-1} s \, ds \right)^{1/n}.$$

We have, as expected, $c_n(0) = 0$ (since $\mathrm{S}_n(t) \sim t$ at 0) and $c_n(d) > 0$ for $d > 0$.

We now state the transport inequality satisfied by the uniform measure $\sigma$ on $M$.

**Theorem.** *Let $M$ be an $n$-dimensional Riemannian manifold with positive Ricci curvature satisfying (1) and let $\sigma$ be its normalized Riemannian volume. Then, for every probability density $f$ on $M$ we have*

$$\mathcal{W}_{c_n}(f \, d\sigma, \sigma) \le H_{n,\sigma}(f).$$

We will see that the cost $c_n(d(x, y))$ behaves like $(n-1)d(x, y)^2/2$ for small distances, so it may seem that we are back to the bad situation (3) where we were stuck with the constant $(n-1)$. However, the entropy $H_{n,\sigma}$ is *better*, i.e., smaller, than the usual entropy $\int f \log f \, d\sigma$ (note that $H_{n,\sigma}(f) \nearrow \int f \log(f) \, d\sigma$ as $n \to +\infty$) , and as a matter of fact we will reproduce the sharp Poincaré inequality. So there is an interesting trade-off between the cost and the entropy. Incidentally, both sides of our inequality are zero when $n = 1$ (which is a good sign), meaning that we don't derive any result on the torus $S^1$, although it might be possible, by looking at first orders when $n \to 1$ and analyzing the proof below, to guess what one should get in this case.

The next section contains the proof of the Theorem. In the last section we give some properties of the cost $c_n$ and we explain how to derive the sharp spectral gap inequality (2) from the Theorem.

## 2. Proof of the theorem

We start by recalling the result of [McCann 2001]. Given two (compactly supported) probability densities $f$ and $g$ on a manifold $M$ with respect to $d\mathrm{vol}$, the Riemannian volume, there exists a Lipschitz function $\theta : M \to \mathbb{R}$ such that $-\theta$ is *c-concave* and the map

$$T(x) = \exp_x(\nabla\theta(x))$$

pushes forward $f \, d\mathrm{vol}$ to $g \, d\mathrm{vol}$. The latter means that for every (bounded or nonnegative) Borel function $u$ on $M$,

$$\int u(y) g(y) \, d\mathrm{vol}(y) = \int u(T(x)) f(x) \, d\mathrm{vol}(x).$$

The $c$-concavity of $-\theta$ is defined by the property that there exists a Lipschitz function $\psi$ such that $-\theta(x) = \inf_y\{\psi(y) + d(x, y)^2/2\}$. This implies (and is formally equivalent to) that at every point $x$ where $\theta$ is differentiable, and thus

$y := T(x)$ is uniquely defined, the function $v \to \theta(v) + \frac{1}{2}d(v, y)^2 - \frac{1}{2}d(x, y)^2$ achieves its minimum at $v = x$.

Following a classical approach, the map $T$ is constructed by establishing that $\pi = (\mathrm{Id} \times T) f \, d\mathrm{vol}$ is the optimizer for $\mathcal{W}_c(f \, d\mathrm{vol}, g \, d\mathrm{vol})$ when $c$ is the quadratic cost. We will not use this property, though.

As explained in [Cordero-Erausquin et al. 2001, 2006], it is possible to do, in a weak sense, the change of variable $y = T(x)$ and to establish a pointwise Jacobian change of variable equation. To be precise, let us set, whenever it makes sense,

$$dT_x := Y(H + \mathrm{Hess}_x \theta)$$

where, for fixed $x \in M$, the linear operators $Y : T_x M \to T_{T(x)} M$ and $H : T_x M \to T_x M$ are defined by

$$Y := d(\exp_x)_{\nabla\theta(x)} \quad \text{and} \quad H := \mathrm{Hess}_x \, d^2_{T(x)}/2,$$

with the notation $d_y(\cdot) = d(y, \cdot)$ for fixed $y \in M$. Then, one has

$$f(x) = g(T(x)) \det dT_x \quad (f \, d\mathrm{vol})\text{-a.e.}$$

The set of points where this equation holds is contained in the set of $x \in M$ where $\theta$ is differentiable at $x$ with $\gamma(t) := \exp_x(t\nabla\theta(x))$ being the unique minimizing geodesic between $x = \gamma(0)$ and $T(x) = \gamma(1) \notin \mathrm{cut}(x)$, and such that $\mathrm{Hess}_x \theta$ exists, in the sense of Aleksandrov for the Lipschitz (and locally semiconvex) function $\theta$; later we shall use that $\mathrm{tr}\,\mathrm{Hess}\,\theta =: \Delta\theta \le \Delta_{\mathcal{D}}\theta$, where $\Delta_{\mathcal{D}}\theta$ is the distributional Laplacian of the Lipschitz function $\theta$. The $c$-concavity of $-\theta$ then implies the following, crucial monotonicity property of $T$, which holds $(f \, d\mathrm{vol})$-a.e.:

$$H + \mathrm{Hess}\,\theta \ge 0. \tag{4}$$

In Euclidean space, $H = \mathrm{Id}$ and we recover that $T(x) = x + \nabla\theta$ is the gradient of the convex function $|x|^2/2 + \theta(x)$ — the Brenier map.

We refer the interested (or worried) reader to [Cordero-Erausquin et al. 2001, 2006] where these facts are carefully stated and proved.

So, under the assumptions of the theorem, let $T(x) = \exp_x(\nabla\theta)$ be the McCann map pushing $\sigma$ forward to $f \, d\sigma$. Denote the displacement distance by

$$\alpha(x) := d(x, T(x)) = |\nabla\theta(x)| \in [0, \pi].$$

The Jacobian equation satisfied almost everywhere is then

$$f(T(x))^{-1} = \det\big(Y(H + \mathrm{Hess}_x \theta)\big) \tag{5}$$

with $Y := d(\exp_x)_{\nabla\theta(x)}$ and $H := \mathrm{Hess}_x \, d^2_{T(x)}/2$.

For $x \in M$ a point where Equation (5) holds, let $E_1 := \nabla\theta/|\nabla\theta|$ be the direction of transport, completed by $E_2, \dots, E_n$ in order to have an orthonormal frame. In this basis, the symmetric operator $H$ takes the form

$$\begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix}$$

and the classical Bishop comparison estimates (see [Petersen 1998], for example) ensure that under (1) we have

$$\det Y \le \left(\frac{\sin\alpha}{\alpha}\right)^{n-1} =: v_n(\alpha)^n \quad \text{and} \quad \operatorname{tr} K \le (n-1)\frac{\alpha}{\tan\alpha} =: w_n(\alpha).$$

Of course, these inequalities are equalities when $M = S^n$, a case where $Y$ and $K$ can be computed explicitly (see [Cordero-Erausquin 1999]).

If we write $\operatorname{Hess}_x \theta = \left(\begin{smallmatrix} a & b^t \\ b & M \end{smallmatrix}\right)$, where $M$ is a symmetric $(n-1)\times(n-1)$ matrix and $a := \operatorname{Hess}_x \theta(E_1)\cdot E_1$ (all the quantities depend on $x$, of course), then we have

$$\begin{aligned} f(T(x))^{-1} &= \det\left[Y\begin{pmatrix} 1+a & b^t \\ b & K+M \end{pmatrix}\right] \le v_n(\alpha)^n \det\begin{pmatrix} 1+a & b^t \\ b & K+M \end{pmatrix} \\ &\le v_n(\alpha)^n \det\begin{pmatrix} 1+a & 0 \\ 0 & K+M \end{pmatrix} \\ &= v_n(\alpha)^n \det\begin{pmatrix} (1+a)\mu(\alpha)^{-(n-1)} & 0 \\ 0 & \mu(\alpha)K+\mu(\alpha)M \end{pmatrix}, \end{aligned}$$

where $\mu$ is a numerical $C^1$ positive function defined on $[0,\pi]$ that will be fixed later. Note that $1+a \ge 0$ and $K+M \ge 0$ by (4). Using the arithmetic-geometric inequality, namely $\det^{1/n} \le \operatorname{tr}/n$ on nonnegative matrices, we then get that

$$nf(T(x))^{-1/n} \le v_n(\alpha)\big((1+a)\mu(\alpha)^{-(n-1)} + \mu(\alpha)w_n(\alpha) + \mu(\alpha)(\Delta\theta - a)\big).$$

We integrate this inequality with respect to $\sigma$. Integration by parts gives

$$\int v_n(\alpha)\mu(\alpha)\Delta\theta\, d\sigma \le -\int (v_n\mu)'(\alpha)\nabla\alpha\cdot\nabla\theta\, d\sigma.$$

When $\theta$ is smooth, the previous equation is an equality, but as we explained above, the Laplacian we used is smaller than the distributional Laplacian in general.

By construction, $\nabla\alpha\cdot\nabla\theta = \alpha \operatorname{Hess}\theta(E_1)\cdot E_1 = \alpha a$ (that this property should be used to improve mass transportation techniques on manifolds was suggested to us by Michael Schmuckenschläger: personal communication, 2001). So we find

$$n\int f^{1-1/n}\,d\sigma \le \int \big(v_n(\alpha)\mu(\alpha)^{-(n-1)} - \mu(\alpha)v_n(\alpha) - \alpha\cdot(v_n\mu)'(\alpha)\big)a\,d\sigma \\ + \int \big(\mu(\alpha)^{-(n-1)} + \mu(\alpha)w_n(\alpha)\big)v_n(\alpha)\,d\sigma.$$

We now want to choose the numerical function $\mu$ such that for all $t\in[0,\pi)$,

$$v_n(t)\mu(t)^{-(n-1)} - \mu(t)v_n(t) - t(v_n\mu)'(t) = 0. \tag{6}$$

Setting $h(t) := t\mu(t)v_n(t)$, the previous equation rewrites as

$$h'(t) = v_n(t)\big(h(t)/tv_n(t)\big)^{-(n-1)} = v_n(t)^n t^{n-1} h(t)^{-(n-1)},$$

or equivalently

$$\frac{1}{n}(h^n)'(t) = \sin^{n-1} t,$$

which suggests the choice $h = \mathrm{S}_n$. So the function *defined* by $\mu(t) := \mathrm{S}_n(t)/tv_n(t)$ satisfies (6), and consequently we have the desired inequality:

$$n\int f^{1-1/n}\,d\sigma \le \int \left(\frac{\sin^{n-1}\alpha(x)}{\mathrm{S}_n(\alpha(x))^{n-1}} + (n-1)\frac{\mathrm{S}_n(\alpha(x))}{\tan\alpha(x)}\right)d\sigma(x). \qquad \square$$

## 3. Further remarks

We start with some properties of the function

$$c_n(\alpha) = n - \frac{\sin^{n-1}\alpha}{\mathrm{S}_n(\alpha)^{n-1}} - (n-1)\frac{\mathrm{S}_n(\alpha)}{\tan\alpha}, \quad \alpha\in[0,\pi).$$

First, observe that for $\alpha\in[0,\pi]$,

$$\int_0^\alpha \sin^{n-1}s\cos s\,ds \le \int_0^\alpha \sin^{n-1}s\,ds \le \int_0^\alpha s^{n-1}\,ds$$

so that

$$\sin\alpha \le \mathrm{S}_n(\alpha) \le \alpha.$$

This implies that $c_n \ge 0$. It also gives that $0 \le (\alpha - \mathrm{S}_n(\alpha))/\alpha^2 \le (\alpha - \sin\alpha)/\alpha^2$ and consequently, for $\alpha\to 0$,

$$\mathrm{S}_n(\alpha) = \alpha + o(\alpha^2).$$

In turn, this gives the behavior of $c_n(\alpha)$ when $\alpha \to 0$:

$$c_n(\alpha) \sim (n-1)\alpha^2/2. \tag{7}$$

To perform this series expansion of $c_n$, write $\mathrm{S}_n(\alpha) = \alpha + a\alpha^3 + o(\alpha^3)$; the coefficient $a$ indeed disappears in the second order. We believe (from numerical examples) that the function $c_n$ is convex on $[0, \pi]$. But since we don't need this property (which seems a bit more technical), we leave this question for another time.

It is well known that the property (7) of the cost is sufficient to derive by linearization, from the corresponding transport inequality, a Poincaré-type inequality. The standard procedure is to first state an infimal convolution inequality (for the Hamilton–Jacobi semigroup), obtained by dualizing the transportation cost and the entropy, and then to linearize (see [Gozlan and Léonard 2010]). Actually, it is enough to dualize only the transportation cost (we don't want to dualize the entropy, since eventually we will linearize it).

Recall the classical Kantorovich duality: for two probability measures $\mu$ and $\nu$ on $M$ and for a cost $c$,

$$\mathcal{W}_c(\mu, \nu) = \sup_{\varphi} \left\{ \int Q_c(\varphi)\, d\mu - \int \varphi\, d\nu \right\}$$

where the supremum is taken over all (Lipschitz) functions $\varphi : M \to \mathbb{R}$ and

$$Q_c(\varphi)(x) := \inf_{y \in M} \bigl\{\varphi(y) + c(d(x, y))\bigr\} \quad \text{for all } x \in M.$$

Note that $Q_c(\varphi) \le \varphi$ (provided $c \ge 0$ and $c(0) = 0$) and that the bigger the cost is in terms of $d(x, y)$, the closer $Q_c(\varphi)$ is to $\varphi$.

Let $g$ be a smooth function on $M$ with $\int g\, d\sigma = 0$, and $\varepsilon > 0$ small. Applying our transport inequality to the probability density $f = 1 + \varepsilon\lambda g$ where $\lambda > 0$ is a constant to be fixed later, and using the above-mentioned duality with the test function $\varphi = \varepsilon g$ we get

$$\int Q_{c_n}(\varepsilon g)(1 + \varepsilon\lambda g)\, d\sigma - \int (\varepsilon g)\, d\sigma \le H_{n,\sigma}(1 + \varepsilon\lambda g). \tag{8}$$

On one hand we have, for the entropy term, uniformly on $M$,

$$n\bigl((1 + \varepsilon\lambda g) - (1 + \varepsilon\lambda g)^{1-1/n}\bigr) = \varepsilon\lambda g + \varepsilon^2 \frac{n-1}{2n} (\lambda g)^2 + o(\varepsilon^2).$$

On the other hand, because of (7) we have

$$Q_{c_n}(\varepsilon g) = \varepsilon \left( g - \varepsilon \frac{1}{2(n-1)} |\nabla g|^2 + o(\varepsilon) \right).$$

Putting these two expansions in (8), we see that the orders 0 and 1 vanish (they have to, since the constant function **1** is an equality case in the transport inequality), and the inequality between the second orders reads as

$$\left( \lambda - \frac{n-1}{2n} \lambda^2 \right) \int g^2\, d\sigma \le \frac{1}{2(n-1)} \int |\nabla g|^2\, d\sigma.$$

Picking $\lambda = \frac{n}{n-1}$ we get the sharp Poincaré inequality $\int g^2\, d\sigma \le \frac{1}{n} \int |\nabla g|^2\, d\sigma$.

## References

[Cordero-Erausquin 1999] D. Cordero-Erausquin, "Inégalité de Prékopa–Leindler sur la sphère", *C. R. Acad. Sci. Paris Sér. I Math.* **329**:9 (1999), 789–792. MR 2000k:26022 Zbl 0945.26026

[Cordero-Erausquin 2002] D. Cordero-Erausquin, "Some applications of mass transport to Gaussian-type inequalities", *Arch. Ration. Mech. Anal.* **161**:3 (2002), 257–269. MR 2003h:49076 Zbl 0998.60080

[Cordero-Erausquin et al. 2001] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, "A Riemannian interpolation inequality à la Borell, Brascamp and Lieb", *Invent. Math.* **146**:2 (2001), 219–257. MR 2002k:58038 Zbl 1026.58018

[Cordero-Erausquin et al. 2006] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, "Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport", *Ann. Fac. Sci. Toulouse Math.* (6) **15**:4 (2006), 613–635. MR 2008j:49111 Zbl 1125.58007

[Gozlan and Léonard 2010] N. Gozlan and C. Léonard, "Transport inequalities: a survey", *Markov Process. Related Fields* **16**:4 (2010), 635–736. MR 2895086 Zbl 1229.26029 arXiv 1003.3852

[Ledoux 2000] M. Ledoux, "The geometry of Markov diffusion generators", *Ann. Fac. Sci. Toulouse Math.* (6) **9**:2 (2000), 305–366. MR 2002a:58045 Zbl 0980.60097

[Lott and Villani 2007] J. Lott and C. Villani, "Weak curvature conditions and functional inequalities", *J. Funct. Anal.* **245**:1 (2007), 311–333. MR 2008f:53039 Zbl 1119.53028

[McCann 2001] R. J. McCann, "Polar factorization of maps on Riemannian manifolds", *Geom. Funct. Anal.* **11**:3 (2001), 589–608. MR 2002g:58017 Zbl 1011.58009

[Otto and Villani 2000] F. Otto and C. Villani, "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality", *J. Funct. Anal.* **173**:2 (2000), 361–400. MR 2001k:58076 Zbl 0985.58019

[Petersen 1998] P. Petersen, *Riemannian geometry*, Graduate Texts in Mathematics **171**, Springer, New York, 1998. 2nd ed. published in 2006. MR 98m:53001 Zbl 0914.53001

[Villani 2009] C. Villani, *Optimal transport: old and new*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **338**, Springer, Berlin, 2009. MR 2010f:49001 Zbl 1156.53003

Received February 6, 2013. Revised August 3, 2013.

DARIO CORDERO-ERAUSQUIN
INSTITUT DE MATHÉMATIQUES DE JUSSIEU
UNIVERSITÉ PIERRE ET MARIE CURIE (PARIS 6)
4 PLACE JUSSIEU
75252 PARIS CEDEX 5
FRANCE
dario.cordero@imj-prg.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.33

# A COHOMOLOGICAL INJECTIVITY RESULT FOR THE RESIDUAL AUTOMORPHIC SPECTRUM OF $\mathrm{GL}_n$

HARALD GROBNER

**Let $\Pi$ be a cohomological residual automorphic representation of $\mathrm{GL}_n/F$, for $F$ an arbitrary number field. Let $q_{\min}$ be the lowest degree in which $\Pi$ has nonvanishing cohomology. We prove that the cohomology of $\Pi$ always injects into the cohomology of the corresponding locally symmetric space in degree $q_{\min}$. This extends the well-known result of Borel for cuspidal automorphic representations to all square-integrable automorphic representations in this certain degree. Moreover, we thereby improve a result of Rohlfs and Speh and confirm an idea of Harder.**

## Introduction

Let $F$ be any number field and let $G = \mathrm{GL}_n/F$. As it is well-known, the space of square-integrable automorphic forms of $G(\mathbb{A})$ decomposes into the space $\mathscr{A}_{\mathrm{cusp}}(G)$ of cuspidal automorphic forms, and a natural complement, the space $\mathscr{A}_{\mathrm{res}}(G)$ of residual automorphic forms. The latter are given by square-integrable residues of Eisenstein series and described in terms of representation theory by [Mœglin and Waldspurger 1989]. Let $\Pi$ be a residual automorphic representation of $G(\mathbb{A})$. We say that $\Pi$ is cohomological, if the ring of relative Lie algebra cohomology of $\Pi$ is nonvanishing with respect to some irreducible, finite-dimensional algebraic representation $\mathscr{M}$ of $G$. See also Section 1C. Assume now that $\Pi$ is cohomological and let $q_{\min}$ be the lowest degree in which $\Pi$ has nonvanishing cohomology.

In this paper we prove that, in degree $q_{\min}$, the cohomology of $\Pi$ always injects into the cohomology of the locally symmetric space attached to $G$. This extends the well-known result of Borel [1981] for cuspidal automorphic representations to all square-integrable automorphic representations in this certain degree. The precise result reads as follows (see Theorem 4.1, also for unexplained notation):

**Theorem.** *Let $G = \mathrm{GL}_n/F$ and let $\mathscr{M}$ be an irreducible, finite-dimensional, algebraic representation of $G$ on a complex vector space. Let $\{P\}$ be an associate class of proper parabolic $F$-subgroups of $G$ and let $\varphi_P$ be an associate class of*

Grobner is supported by the Austrian Science Fund, project number P 25974-N25.
*MSC2010:* primary 11F70, 11F75, 22E47; secondary 11F67.
*Keywords:* automorphic cohomology, residual representation, general linear group.

*cuspidal automorphic representations of* $L_P(\mathbb{A})$. *Let* $\Pi \hookrightarrow \mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ *be a residual automorphic representation of* $G(\mathbb{A})$ *with cuspidal support* $\pi \in \varphi_P$, *spanned by iterated residues of Eisenstein series at a point* $\nu \in \check{\mathfrak{a}}^G_{P,\mathbb{C}}$ *for which* $\nu + \chi_{\tilde{\pi}}$ *is annihilated by* $\mathcal{J}$. *The map in cohomology*

$$H^{q_{\min}}\big(\mathfrak{m}_G, K, \Pi \otimes \mathcal{M}\big) \longrightarrow H^{q_{\min}}\big(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G) \otimes \mathcal{M}\big),$$

*induced from the natural inclusion* $\Pi \hookrightarrow \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G)$, *is injective. In other words, the* $(\mathfrak{m}_G, K)$*-cohomology of a residual automorphic representation of* $\mathrm{GL}_n(\mathbb{A})$ *always embeds into* $H^\bullet\big(G(F)\backslash G(\mathbb{A})/A^{\mathbb{R}}_G K, \widetilde{\mathcal{M}}\big)$ *in its lowest, nonvanishing degree.*

This improves a result of Rohlfs and Speh [2011] (see also our Remark 4.2) and confirms an idea of Harder. Moreover, it may be viewed as a refinement of one of our own results in [Grobner 2013]. Although we believe that it is interesting in its own right, we hope that it will also be of use in a forthcoming work of Harder and Raghuram on special values of Ranking–Selberg $L$-functions.

## 1. Notation

**1A.** ***Number fields and adeles.*** Let $F$ denote an arbitrary number field with set of places $S$. We write $S_\infty = S_{\mathbb{R}} \cup S_{\mathbb{C}}$ for the subset of archimedean places, where $S_{\mathbb{R}}$ denotes the set of real archimedean places and $S_{\mathbb{C}}$ denotes the set of complex archimedean places of $F$. We use $F_v$ for the topological completion of $F$ at $v \in S$. As usual, $\mathbb{A}$ stands for its ring of adeles.

**1B.** ***Algebraic groups.*** In this paper, $G := \mathrm{GL}_n/F$ denotes the general linear group over $F$. We fix the usual Borel subgroup $B$ of upper triangular matrices with Levi decomposition $B = TU$. This choice defines the standard parabolic $F$-subgroups $P$ with Levi decomposition $P = L_P N_P$, where $L_P \supseteq T$ and $N_P \subseteq U$. Clearly, $L_P \cong \mathrm{GL}_{k_1} \times \cdots \times \mathrm{GL}_{k_\ell}$, with $\sum_{i=1}^{\ell} k_i = n$. We let $A_P = Z_{L_P}$ be the maximal $F$-split torus of $L_P$, satisfying $A_P \subseteq T$ and denote by $\mathfrak{a}_P$ (resp., $\mathfrak{a}_{P,\mathbb{C}}$) its Lie algebra (resp., its complexification $\mathfrak{a}_{P,\mathbb{C}} = \mathfrak{a}_P \otimes \mathbb{C}$). The respective duals are denotes $\check{\mathfrak{a}}_P$ and $\check{\mathfrak{a}}_{P,\mathbb{C}}$. The inclusion $A_P \subseteq T$ (resp., the restriction to $P$) defines $\mathfrak{a}_P \to \mathfrak{t}$ (resp., $\check{\mathfrak{a}}_P \to \check{\mathfrak{t}}$), which leads to direct sum decompositions $\mathfrak{t} = \mathfrak{a}_P \oplus \mathfrak{a}^P$ and $\check{\mathfrak{t}} = \check{\mathfrak{a}}_P \oplus \check{\mathfrak{a}}^P$. We let $\mathfrak{a}^Q_P := \mathfrak{a}_P \cap \mathfrak{a}^Q$ and $\check{\mathfrak{a}}^Q_P := \check{\mathfrak{a}}_P \cap \check{\mathfrak{a}}^Q$ for parabolic $F$-subgroups $Q$ and $P$. We write $H_P : L_P(\mathbb{A}) \to \mathfrak{a}_{P,\mathbb{C}}$ for the standard Harish-Chandra height function [Franke 1998, p. 185]. The group $L_P(\mathbb{A})^1 := \ker H_P$, admits a direct complement $A^{\mathbb{R}}_P \cong \mathbb{R}_+^{\dim \mathfrak{a}_P} = \mathbb{R}^\ell_+$ in $L_P(\mathbb{A})$ whose Lie algebra is isomorphic to $\mathfrak{a}_P \cong \mathbb{R}^\ell$. With respect to a maximal compact subgroup $K_{\mathbb{A}} \subseteq G(\mathbb{A})$ in good position (see [Mœglin and Waldspurger 1995, I.1.4]), we obtain an extension $H_P : G(\mathbb{A}) \to \mathfrak{a}_{P,\mathbb{C}}$ to all of $G(\mathbb{A})$.

**1C.** ***Lie groups and Lie algebras.*** The Lie algebra of a real Lie group is denoted by the same letter in gothic lowercase; thus $\mathfrak{g}_\infty = \mathfrak{gl}_n(\mathbb{R})^{|S_\mathbb{R}|} \oplus \mathfrak{gl}_n(\mathbb{C})^{|S_\mathbb{C}|}$ is the real Lie algebra of $G_\infty := R_{F/\mathbb{Q}}(G)(\mathbb{R})$, and so on. We set $\mathfrak{m}_G := \mathfrak{g}_\infty/\mathfrak{a}_G = \mathrm{Lie}(G(\mathbb{A})^1 \cap G_\infty)$ and denote by $\mathfrak{Z}(\mathfrak{g}_\infty)$ the center of the universal enveloping algebra $\mathcal{U}(\mathfrak{g}_\infty)$ of $\mathfrak{g}_{\infty,\mathbb{C}} := \mathfrak{g}_\infty \otimes_\mathbb{R} \mathbb{C}$. We will also use the notation $G_v$ for $G(F_v)$, $v \in S_\infty$, and similar for other local groups (such as $L_{P,v}$ etc.).

Let $K_\infty \subset G_\infty$ be a maximal compact subgroup (the archimedean factor of the maximal compact subgroup $K_\mathbb{A}$ of $G(\mathbb{A})$ in good position) and set once and for all $K := K_\infty^\circ$, the connected component of the identity element. We refer the reader to [Borel and Wallach 1980, Chapter I] for the basic facts and notations concerning $(\mathfrak{m}_G, K)$-cohomology. If $H$ is any subgroup of $G_\infty$, we denote by $K_H$ the intersection $K \cap H$.

**1D.** ***Algebraic representations.*** In this paper, $\mathcal{M}$ will always be a finite-dimensional irreducible algebraic representation of $G$ on a complex vector space. For simplicity, we will assume that $A_G^\mathbb{R}$ (and so $\mathfrak{a}_G$) acts trivially on $\mathcal{M}$. There is hence no difference between the $(\mathfrak{g}_\infty, K)$-module and the $(\mathfrak{m}_G, K)$-module defined by $\mathcal{M}$.

## 2. Automorphic representations

**2A.** ***Automorphic forms.*** Our notion of an *automorphic form* $f : G(\mathbb{A}) \to \mathbb{C}$ and of an *automorphic representation* of $G(\mathbb{A})$ is the one from [Borel and Jacquet 1979, 4.2 and 4.6]. Let $\mathcal{A}(G)$ be the space of all automorphic forms $f : G(\mathbb{A}) \to \mathbb{C}$ that are constant on the real Lie subgroup $A_G^\mathbb{R}$. By its very definition, every automorphic form is annihilated by some power of an ideal $\mathcal{J} \lhd \mathfrak{Z}(\mathfrak{g}_\infty)$ of finite codimension. We fix such an ideal $\mathcal{J}$ once and for all; as we will only be interested in cohomological automorphic forms, we take $\mathcal{J}$ to be the ideal which annihilates the contragredient representation $\mathcal{M}^\vee$ of $\mathcal{M}$ (see Section 1C) and denote by

$$\mathcal{A}_\mathcal{J}(G) \subset \mathcal{A}(G)$$

the space consisting of those automorphic forms that are annihilated by some power of $\mathcal{J}$. Clearly, $\mathcal{A}_\mathcal{J}(G)$ carries commuting $(\mathfrak{g}_\infty, K)$- and $G(\mathbb{A}_f)$-actions and hence defines an $(\mathfrak{m}_G, K, G(\mathbb{A}_f))$-module. As such a module, any irreducible subquotient (that is, any automorphic representation) $\Pi$ decomposes as $\Pi \cong \Pi_\infty \otimes \Pi_f$.

**2B.** ***$L^2$-automorphic forms.*** The $(\mathfrak{m}_G, K, G(\mathbb{A}_f))$-submodule of all square-integrable automorphic forms in $\mathcal{A}_\mathcal{J}(G)$ is denoted $\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G)$. An irreducible subrepresentation of $\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G)$ will be called an *$L^2$-automorphic* representation (see [Borel 2007, 9.6]). If $\omega : Z_G(F)\backslash Z_G(\mathbb{A}) \to \mathbb{C}^*$ is a continuous character of the center $Z_G$ of $G$, we let $\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G, \omega)$ be the space of square-integrable automorphic forms with central character $\omega$.

We further recall that $\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G,\omega)$ decomposes as a direct sum of automorphic representations $\Pi$

$$\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G,\omega) \cong \bigoplus \Pi,$$

which can be described as follows: According to [Mœglin and Waldspurger 1989], every summand $\Pi$ in the above decomposition is of the form $\Pi \cong J(P,\pi,\nu)$, where the latter stands for the (smooth, $K$-finite vectors in the) unique irreducible quotient of the (normalized) induced representation $I_{P(\mathbb{A})}^{G(\mathbb{A})}[\pi \otimes \nu]$, with inducing data $\pi$, a cuspidal automorphic representation of $L_P(\mathbb{A})$, and $\nu \in \check{\mathfrak{a}}_{P,\mathbb{C}}$. In fact, as is well-known, by [Mœglin and Waldspurger 1989, Théorème, p. 606], more can be said:

**Theorem 2.1.** *Any $L^2$-automorphic representation of $G(\mathbb{A})$ is given by a triple $(L_P,\sigma,\nu)$, where*

(1) $L_P \cong \mathrm{GL}_k \times \cdots \times \mathrm{GL}_k$, *with* $\ell k = n$;

(2) $\pi \cong \sigma \otimes \cdots \otimes \sigma$, *with $\sigma$ a cuspidal automorphic representation of* $\mathrm{GL}_k(\mathbb{A})$;

(3) $\nu = ((\ell-1)/2,\ldots,(1-\ell)/2)$ *in the coordinates given by the absolute value of the determinant of* $\mathrm{GL}_k(\mathbb{A})$;

*and no other triples determine an $L^2$-automorphic representation. The datum $(L_P,\sigma,\nu)$ is unique.*

As a matter of fact, the space of $L^2$-automorphic forms decomposes as a direct sum

$$\mathcal{A}_{\mathrm{dis},\mathcal{J}}(G) \cong \mathcal{A}_{\mathrm{cusp},\mathcal{J}}(G) \oplus \mathcal{A}_{\mathrm{res},\mathcal{J}}(G),$$

where $\mathcal{A}_{\mathrm{cusp},\mathcal{J}}(G)$ is the space of cuspidal automorphic forms in $\mathcal{A}_{\mathcal{J}}(G)$ and $\mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ denotes the space of residual automorphic forms in $\mathcal{A}_{\mathcal{J}}(G)$. More precisely, adding a central character $\omega$ to this datum, according to the theorem above, $\mathcal{A}_{\mathrm{res},\mathcal{J}}(G,\omega)$ is the direct sum of all $L^2$-automorphic representations given by a triple $(L_P,\sigma,\nu)$, with $P$ proper.

**2C. *Parabolic supports.*** Let $\{P\}$ be the associate class of the parabolic $F$-subgroup $P$. It consists by definition of all parabolic $F$-subgroups $Q = L_Q N_Q$ of $G$ for which $L_Q$ and $L_P$ are conjugate by an element in $G(F)$. We denote by $\mathcal{A}_{\mathcal{J},\{P\}}(G)$ the space of all $f \in \mathcal{A}_{\mathcal{J}}(G)$ that are negligible along every parabolic $F$-subgroup $Q \notin \{P\}$. (For the sake of completeness, we recall that the latter condition means that for all $g \in G(\mathbb{A})$, the function $L_Q(\mathbb{A}) \to \mathbb{C}$ given by $l \mapsto f_Q(lg)$ is orthogonal to the space of cuspidal functions on $L_P(F)A_G^{\mathbb{R}}\backslash L_P(\mathbb{A})$.) There is the following decomposition of $\mathcal{A}_{\mathcal{J}}(G)$ as an $(\mathfrak{m}_G, K, G(\mathbb{A}_f))$-module (see [Borel et al. 1996, Theorem 2.4] or [Borel 2007, 10.3]), first established by Langlands:

$$\mathcal{A}_{\mathcal{J}}(G) \cong \bigoplus_{\{P\}} \mathcal{A}_{\mathcal{J},\{P\}}(G).$$

**2D.** ***Cuspidal supports.*** The various summands $\mathcal{A}_{\mathcal{J},\{P\}}(G)$ can be decomposed even further. To this end, recall from [Franke and Schwermer 1998, 1.2] the notion of an *associate class* $\varphi_P$ of cuspidal automorphic representations of the Levi subgroups of the elements in the class $\{P\}$. Therefore, let $\{P\}$ be represented by $P = LN$. Then the associate classes $\varphi_P$ may be parametrized by pairs of the form $(\Lambda, \tilde{\pi})$, where

(1) $\tilde{\pi}$ is a unitary cuspidal automorphic representation of $L(\mathbb{A})$, whose central character vanishes on the group $A_P^{\mathbb{R}}$;

(2) $\Lambda : A_P^{\mathbb{R}} \to \mathbb{C}^*$ is a Lie group character; and

(3) the infinitesimal character $\chi_{\tilde{\pi}}$ of $\tilde{\pi}_\infty$ and the derivative $d\Lambda \in \check{\mathfrak{a}}_{P,\mathbb{C}}$ of $\Lambda$ are compatible with the action of $\mathcal{J}$ (see [loc. cit.]).

Each associate class $\varphi_P$ may thus be represented by a cuspidal automorphic representation

$$\pi := \tilde{\pi} \otimes e^{\langle d\Lambda, H_P(\cdot)\rangle}$$

of $L(\mathbb{A})$. Given $\varphi_P$, represented by a cuspidal representation $\pi$ of the above form, an $(\mathfrak{m}_G, K, G(\mathbb{A}_f))$-submodule

$$\mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G)$$

of $\mathcal{A}_{\mathcal{J},\{P\}}(G)$ was defined in Section 1.3 of [Franke and Schwermer 1998] as the span of all possible holomorphic values or residues of all Eisenstein series attached to $\tilde{\pi}$, evaluated at the point $\lambda = d\Lambda$, together with all their derivatives. This definition is independent of the choice of the representatives $P$ and $\pi$, thanks to the functional equations satisfied by the Eisenstein series considered. For details, we refer the reader to Sections 1.2–1.4 of the same paper.

The following refined decomposition as $(\mathfrak{m}_G, K, G(\mathbb{A}_f))$-modules of the spaces $\mathcal{A}_{\mathcal{J},\{P\}}(G)$ of automorphic forms was obtained in [Franke and Schwermer 1998, Theorem 1.4]:

$$\mathcal{A}_{\mathcal{J},\{P\}}(G) \cong \bigoplus_{\varphi_P} \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G).$$

**2E.** ***Quadruples in the refined version of Franke's filtration.*** A definition of the integer-valued function $T$ on the set of automorphic exponents is given in [Franke 1998, p. 233]. Because the technicalities are of little consequence to this paper, we won't repeat this definition here, but refer the reader to the original paper. The important fact is that we may assume a fixed choice of $T$ making the length $m = m(\{P\})$ of the corresponding filtration of $\mathcal{A}_{\mathcal{J},\{P\}}(G)$ minimal, as in our paper [Grobner 2013, 3.1 on p. 1072].

Given a cuspidal support $\varphi_P$, we will need the following collection of data, as was already introduced in [Grobner 2013, 3.2]. Let $M_{\mathcal{J},\{P\},\varphi_P}$ be the set of quadruples $(R, \Pi, \nu, \lambda)$, with

(1) $R$ a standard parabolic $F$-subgroup of $G$ containing a representative of $\{P\}$;

(2) $\Pi$ a unitary discrete series automorphic representation of $L_R(\mathbb{A})$ with cuspidal support determined by $\varphi_P$, spanned by iterated residues of Eisenstein series at the point $\nu \in \check{\mathfrak{a}}^R_{P,\mathbb{C}}$; and

(3) $\lambda \in \check{\mathfrak{a}}_{R,\mathbb{C}}$ such that $\Re e(\lambda) \in \overline{\check{\mathfrak{a}}^{G+}_R}$, the closed positive Weyl chamber in $\check{\mathfrak{a}}^G_R$, and such that $\lambda + \nu + \chi_{\tilde{\pi}}$ is annihilated by $\mathcal{J}$.

We point out that with this definition, although not entirely obvious, one can show that $T$ is well-defined on $\Re e(\lambda)_+$; [Franke 1998, p. 233]. Therefore, taking this for granted, it makes sense to define

$$M^{(j)}_{\mathcal{J},\{P\},\varphi_P} := \{(R, \Pi, \nu, \lambda) \mid T(\Re e(\lambda)_+) = j\}.$$

These sets of quadruples $M^{(j)}_{\mathcal{J},\{P\},\varphi_P}$ originate from [Franke 1998, pp. 218, 233–234]. There, however, only the parabolic support $\{P\}$ and not the cuspidal support $\varphi_P$ was taken into account.

## 3. Automorphic cohomology

**3A.** ***Cohomology of locally symmetric spaces.*** We let

$$S := G(F)A^{\mathbb{R}}_G\backslash G(\mathbb{A})/K$$

be the projective limit of the "locally symmetric spaces" attached to $G$. Starting from the algebraic representation $\mathcal{M}$, one obtains a sheaf $\widetilde{\mathcal{M}}$ on $S$ by letting $\widetilde{\mathcal{M}}$ be the sheaf with espace étalé $G(\mathbb{A})/A^{\mathbb{R}}_G K \times_{G(F)} \mathcal{M}$ with the discrete topology on $\mathcal{M}$. We write $H^q(S, \widetilde{\mathcal{M}})$ for the corresponding space of sheaf cohomology (in degree $q$).

**3B.** ***Automorphic cohomology.*** We recall that the $G(\mathbb{A}_f)$-module

$$H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J}}(G) \otimes \mathcal{M})$$

is called the *automorphic cohomology* of $G$ in degree $q$. From Sections 2C and 2D we know that it inherits a direct sum decomposition

$$\begin{aligned} H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J}}(G) \otimes \mathcal{M}) &\cong \bigoplus_{\{P\}} H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{P\}}(G) \otimes \mathcal{M}) \\ &\cong \bigoplus_{\{P\}} \bigoplus_{\varphi_P} H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G) \otimes \mathcal{M})\,. \end{aligned}$$

The summand $H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{G\}}(G) \otimes \mathcal{M})$ attached to $\{G\}$ consists precisely of all cuspidal automorphic forms in $\mathcal{A}_{\mathcal{J}}(G)$.

Conjectured by Harder and Borel and proved by Franke [1998, Theorem 18], the following result which links automorphic cohomology with the sheaf cohomology of $S$:

**Theorem 3.1.** *There is an isomorphism of $G(\mathbb{A}_f)$-modules*

$$H^q(S, \widetilde{\mathcal{M}}) \cong H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J}}(G) \otimes \mathcal{M}).$$

The latter results brings us back to the more geometric point of view of cohomology, presented in Section 3A.

**3C.** ***Certain bounds in cohomology.*** Let $R = L_R N_R$ be a standard parabolic subgroup of $G$ and $v$ an archimedean place of $F$. We write

$$\mathfrak{l}_{R,v} \cap \mathfrak{m}_G = \mathfrak{l}^{ss}_{R,v} \oplus (\mathfrak{a}_{R,v} \cap \mathfrak{m}_G) \quad \text{and} \quad \mathfrak{k}^{ss}_{L_R} := \mathfrak{k}_{L_{R,v}} \cap \mathfrak{l}^{ss}_{R,v}.$$

Now, given an irreducible, admissible $L_{R,v}$-representation $\pi_v$, let $q(L_{R,v}, \pi_v)$ be the smallest degree in which $\pi_v$ has nontrivial $(\mathfrak{l}^{ss}_{R,v}, \mathfrak{k}^{ss}_{L_{R,v}})$-cohomology, twisted by an irreducible, finite-dimensional, algebraic representation of $L_{R,v}$. If there is no such coefficient module, then we let $q(L_{R,v}, \pi_v) = 0$. (This number was denoted "$m(L_{R,v}, \pi_v)$" in [Grobner 2013].) Similarly, we write $q(L_{R,\infty}, \pi_\infty) := \sum_{v \in S_\infty} q(L_{R,v}, \pi_v)$.

Let $\{P\}$ be an associate class of proper parabolic $F$-subgroups of $G$, and $\varphi_P$ an associate class of cuspidal automorphic representations of $L_P(\mathbb{A})$. We define

$$q_{\mathsf{res}} := \min_{0 \le j < m} \left( \min_{(R,\Pi,\nu,\lambda) \in M^{(j)}_{\mathcal{J},\{P\},\varphi_P}} \left( \sum_{v \in S_\infty} \left\lceil \tfrac{1}{2} \dim_{\mathbb{R}} N_R(F_v) \right\rceil + q(L_{R,v}, \Pi_v) \right) \right).$$

Of course, although not reflected in the notation, $q_{\mathsf{res}}$ depends on the support $\{P\}$ and $\varphi_P$. This rather complicatedly defined number (see [Grobner 2013, 6.1] for the original source) serves as a certain bound of degrees of cohomology, as we proved in the same paper. Indeed, boiled down to the case of $G = \mathrm{GL}_n$ here, in Corollary 17 of that paper, we showed the following result:

**Theorem 3.2.** *Let $G = \mathrm{GL}_n/F$ and let $\mathcal{M}$ be an irreducible, finite-dimensional, algebraic representation of $G$ on a complex vector space. Let $\{P\}$ be an associate class of proper parabolic $F$-subgroups of $G$ and let $\varphi_P$ be an associate class of cuspidal automorphic representations of $L_P(\mathbb{A})$. Let $\Pi \hookrightarrow \mathcal{A}_{\mathsf{res},\mathcal{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$ with cuspidal support $\pi \in \varphi_P$, spanned by iterated residues of Eisenstein series at a point $\nu \in \check{\mathfrak{a}}^G_{P,\mathbb{C}}$, for which $\nu + \chi_{\tilde{\pi}}$ is annihilated by $\mathcal{J}$. Then, the map in cohomology*

$$H^q(\mathfrak{m}_G, K, \Pi \otimes \mathcal{M}) \longrightarrow H^q(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G) \otimes \mathcal{M}),$$

*induced from the natural inclusion $\Pi \hookrightarrow \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G)$, is injective in all degrees $0 \le q < q_{\mathsf{res}} = q_{\mathsf{res}}(\{P\}, \varphi_P)$.*

The latter theorem will be the key result for the proof of our main result of this article in the next section.

## 4. The main result

**4A.** Let $\Pi \hookrightarrow \mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$. Recall from Section 3C our notation $q(G_\infty, \Pi_\infty)$ for the minimal degree in which $\Pi_\infty$ has nontrivial $(\mathfrak{m}_G, K)$-cohomology with respect to an irreducible, finite-dimensional, algebraic representation of $G$. For sake of simplicity, since the group $G$ and the representation $\Pi$ are clear from the context, we will write $q_{\mathsf{min}} := q(G_\infty, \Pi_\infty)$ for this minimal degree.

**Theorem 4.1.** *Let $G = \mathrm{GL}_n/F$ and let $\mathcal{M}$ be an irreducible, finite-dimensional, algebraic representation of $G$ on a complex vector space. Let $\{P\}$ be an associate class of proper parabolic $F$-subgroups of $G$ and let $\varphi_P$ be an associate class of cuspidal automorphic representations of $L_P(\mathbb{A})$. Let $\Pi \hookrightarrow \mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$ with cuspidal support $\pi \in \varphi_P$, spanned by iterated residues of Eisenstein series at a point $\nu \in \check{\mathfrak{a}}^G_{P,\mathbb{C}}$, for which $\nu + \chi_{\tilde{\pi}}$ is annihilated by $\mathcal{J}$. The map in cohomology*

$$H^{q_{\mathsf{min}}}(\mathfrak{m}_G, K, \Pi \otimes \mathcal{M}) \longrightarrow H^{q_{\mathsf{min}}}(\mathfrak{m}_G, K, \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G) \otimes \mathcal{M}),$$

*induced from the natural inclusion $\Pi \hookrightarrow \mathcal{A}_{\mathcal{J},\{P\},\varphi_P}(G)$, is injective. In other words, the $(\mathfrak{m}_G, K)$-cohomology of a residual automorphic representation of $\mathrm{GL}_n(\mathbb{A})$ always embeds into $H^\bullet(S, \widetilde{\mathcal{M}})$ in its lowest, nonvanishing degree.*

**Remark 4.2.** The reader should not confuse this theorem with [Rohlfs and Speh 2011, Theorem IV.4] and with [Grobner 2013, Theorem 22], where seemingly similar results were shown. In fact, Theorem 4.1 above is a improvement as well as a refinement of both of these theorems: First of all, here we show that the cohomology of a residual automorphic representation of $\mathrm{GL}_n(\mathbb{A})$ *always injects into* $H^q(S, \widetilde{\mathcal{M}})$ in its lowest nonvanishing degree and hence give a precise description of its nontrivial contribution. (Moreover, in contrast to [Rohlfs and Speh 2011], we allow any number field $F$ and any coefficient module $\mathcal{M}$.) Secondly, we also obtain an improvement of the bound of degrees of cohomology given in [Grobner 2013, Theorem 22].

The proof of this theorem consists of two steps. First, we determine the minimal degree $q_{\mathsf{min}}$ explicitly for all cohomological residual automorphic representations of $G(\mathbb{A})$. Secondly, we make the effort and calculate our bound $q_{\mathsf{res}} = q_{\mathsf{res}}(\{P\}, \varphi_P)$ for given support $\{P\}$, $\varphi_P$ and show that it is always strictly greater than $q_{\mathsf{min}}$. The theorem is then a consequence from Theorem 3.2. As the reader will see, we will have to distinguish the case of a real archimedean place and a complex archimedean place.

## 5. Proof of main theorem: determination of $q_{\mathsf{min}}$

**5A.** Let $\Pi \hookrightarrow \mathscr{A}_{\mathrm{res},\mathscr{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$. By Theorem 2.1 it is given by a triple $(L_P, \pi, \nu)$, where $L_P \cong \mathrm{GL}_k \times \cdots \times \mathrm{GL}_k$, $\ell k = n$, and $\pi \cong \sigma \otimes \cdots \otimes \sigma$ is a cuspidal automorphic representation of $L_P(\mathbb{A})$. If $\Pi_\infty$ is cohomological with respect to $\mathscr{M}$, then $\pi_\infty$ is cohomological, too. As the only cohomological generic representations of $G_\infty$ are essentially tempered, we see that $\pi_\infty$ is essentially tempered. Hence, by its very construction, $\Pi_\infty$ is the Langlands quotient given by the triple $(L_{P,\infty}, \pi_\infty, \nu)$. Of course this also holds locally at $v \in S_\infty$.

Let now be $v \in S_\mathbb{R}$. Then $\Pi_v$ comes under the purview of the Vogan–Zuckerman classification of cohomological representations in terms of $A_\mathfrak{q}(\lambda)$-modules. We assume that the reader is familiar with this theory and refer to [Vogan and Zuckerman 1984] and [Knapp and Vogan 1995]. We write $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{u}$ for the Levi decomposition of the complex parabolic subalgebra $\mathfrak{q}$ of $\mathfrak{g}_{v,\mathbb{C}}$. By [Knapp and Vogan 1995, Chapter IV, Proposition 4.76], $\mathfrak{u}$ is the direct sum of certain root-eigenspaces, all of them one-dimensional. Hence, $\dim_\mathbb{C} \mathfrak{u}$ is the number of roots appearing in $\mathfrak{u}$. Moreover, from $\Pi_v$ being the Langlands quotient given by the triple $(L_P(\mathbb{R}), \pi_v, \nu)$, we derive that

$$\mathfrak{l} \cong \begin{cases} \mathfrak{gl}_\ell(\mathbb{C})^{k/2} & \text{for } k \text{ even,} \\ \mathfrak{gl}_\ell(\mathbb{C})^{(k-1)/2} \oplus \mathfrak{gl}_\ell(\mathbb{R}) & \text{for } k \text{ odd;} \end{cases}$$

see [Vogan and Zuckerman 1984, Theorem 6.16]. It is now an easy combinatorial exercise, using [Knapp and Vogan 1995, IV, Proposition 4.76] to show that the number of roots appearing in $\mathfrak{u}$ (and hence $\dim_\mathbb{C} \mathfrak{u}$) equals

$$\dim_\mathbb{C} \mathfrak{u} = \begin{cases} \frac{1}{4} n(n-\ell+1) & \text{for } k \text{ even,} \\ \frac{1}{4}(n(n-\ell+1)-\ell) & \text{for } k \text{ odd.} \end{cases}$$

Because in the case of $\mathfrak{g} = \mathfrak{gl}_n(\mathbb{R})$ all roots showing up in $\mathfrak{u}$ are noncompact, Theorem 5.5 of [Vogan and Zuckerman 1984] implies that the minimal degree in which $\Pi_v \cong A_\mathfrak{q}(\lambda)$ has nontrivial cohomology is precisely $\dim_\mathbb{C} \mathfrak{u}$.

Now let $v \in S_\mathbb{C}$. Then, by [Enright 1979], $\Pi_v$ is fully induced and so the minimal degree in which $\Pi_v$ has nontrivial cohomology is readily computed using [Borel and Wallach 1980, Chapter III, Theorem 3.3]. Summarizing, we have shown:

**Proposition 5.1.** *Let $\Pi \hookrightarrow \mathscr{A}_{\mathrm{res},\mathscr{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$ that is $(\mathfrak{m}_G, K)$-cohomological with respect to $\mathscr{M}$. Assume that $\Pi$ is given by the triple $(L_P, \pi, \nu)$, where $L_P \cong \mathrm{GL}_k \times \cdots \times \mathrm{GL}_k$, $\ell k = n$. Then*

$$q_{\mathsf{min}} = \begin{cases} |S_\mathbb{R}| \cdot \frac{1}{4} n(n-\ell+1) + |S_\mathbb{C}| \cdot \frac{1}{2} n(n-\ell) & \text{for } k \text{ even,} \\ |S_\mathbb{R}| \cdot \frac{1}{4}(n(n-\ell+1)-\ell) + |S_\mathbb{C}| \cdot \frac{1}{2} n(n-\ell) & \text{for } k \text{ odd.} \end{cases}$$

## 6. Proof of main theorem: determination of $q_{\mathsf{res}}$

**6A. *Reduction to maximal parabolics.*** This step is much more technical in nature. We will have to make many case-by-case distinctions to actually calculate $q_{\mathsf{res}}$. As a first result towards the determination of $q_{\mathsf{res}}$, we shall need the following result:

**Lemma 6.1.** *For every proper support $\{P\}$, $L_P \cong \mathrm{GL}_k^\ell$, $\varphi_P$ and every $j$, $0 \le j < m$, the minimum*

$$\min_{(R,\Pi,\nu,\lambda)\in M^{(j)}_{\mathcal{J},\{P\},\varphi_P}} \left( \sum_{v\in S_\infty} \left\lceil \tfrac{1}{2} \dim_{\mathbb{R}} N_R(F_v) \right\rceil + q(L_{R,v}, \Pi_v) \right)$$

*is obtained at a maximal parabolic subgroup $R$.*

*Proof.* We will prove this by checking that the number

$$n(R_v) := \left\lceil \tfrac{1}{2} \dim_{\mathbb{R}} N_R(F_v) \right\rceil + q(L_{R,v}, \Pi_v) \tag{6.2}$$

decreases for all $v \in S_\infty$ as we increase the parabolic subgroup $R$, that is, as we form the union of two diagonal blocks $a \cdot k$ and $b \cdot k$ to a block of size $(a+b)\cdot k$. We will write $R(a,b)$ for the first parabolic subgroup, that is, the one having two diagonal blocks of size $a \cdot k$ and $b \cdot k$, and $R(a+b)$ for the second parabolic subgroup, that is, the one having a diagonal block of size $(a+b)\cdot k$, containing the two diagonal blocks of size $a \cdot k$ and $b \cdot k$, instead.

Now, let $v \in S_{\mathbb{R}}$. We have to distinguish several cases. The first three, where both $a$ and $b$ are assumed to be greater than or equal to one, are checked using Proposition 5.1: As cohomology satisfies the Künneth rule, we computed the degree $q(L_{R,v}, \Pi_v)$ for a given quadruple $(R, \Pi, \nu, \lambda) \in M^{(j)}_{\mathcal{J},\{P\},\varphi_P}$ in this proposition. The dimension of the unipotent radical is easily computed for each parabolic subgroups $R(a,b)$ and $R(a+b)$. Putting this together, we obtain:

**Case 1:** $k$ even, $a, b \ge 2$.

$$\begin{aligned} n(R(a,b)_v) - n(R(a+b)_v) &= \tfrac{1}{4}ak(ak-a+1) + \tfrac{1}{4}bk(bk-b+1) + \left\lceil \tfrac{1}{2}abk^2 \right\rceil \\ &\qquad - \tfrac{1}{4}(a+b)k\big((a+b)k-(a+b)+1\big) \\ &= \tfrac{1}{2}abk. \end{aligned}$$

**Case 2:** $k$ odd, $a, b \ge 2$, $a$ or $b$ even.

$$n(R(a,b)_v) - n(R(a+b)_v) = \tfrac{1}{2}abk.$$

**Case 3:** $k$ odd, $a, b \ge 2$, $a$ and $b$ odd.

$$n(R(a,b)_v) - n(R(a+b)_v) = \tfrac{1}{2}(abk+1).$$

The remaining cases, namely when $b = 1$, have a cuspidal automorphic component at the single $k$-block of $L_R$. This cuspidal automorphic representation has to be cohomological, whence its archimedean component at $v$ is tempered. The degree $q(L_{R,v}, \Pi_v)$ is now computed by Proposition 5.1 (for the residual representation of the block of size $a \cdot k$) and using [Borel and Wallach 1980, III, Proposition 5.3] (for the cuspidal representation of the block of size $k$), where the lowest degree of cohomology of tempered representations is determined. Finally, we obtain:

**Case 4:** $k$ even, $a \geq 2$, $b = 1$.

$$\begin{aligned} n(R(a,b)_v) - n(R(a+b)_v) &= \tfrac{1}{4}ak(ak-a+1) + \tfrac{1}{2}\bigl(\tfrac{1}{2}k(k+1) - k + \lfloor \tfrac{1}{2}k \rfloor\bigr) \\ &\qquad + \lceil \tfrac{1}{2}ak^2 \rceil - \tfrac{1}{4}(a+1)k\bigl((a+1)k-(a+1)+1\bigr) \\ &= \tfrac{1}{2}ak. \end{aligned}$$

**Case 5:** $k$ odd, $a \geq 2$ even, $b = 1$.

$$n(R(a,b)_v) - n(R(a+b)_v) = \tfrac{1}{2}ak.$$

**Case 6:** $k$ odd, $a \geq 2$ odd, $b = 1$.

$$n(R(a,b)_v) - n(R(a+b)_v) = \tfrac{1}{2}(ak+1).$$

**Case 7:** $k$ even, $a = b = 1$.

$$\begin{aligned} n(R(a,b)_v) - n(R(a+b)_v) &= \bigl(\tfrac{1}{2}k(k+1) - k + \lfloor \tfrac{1}{2}k \rfloor + \lceil \tfrac{1}{2}ak^2 \rceil\bigr) - \tfrac{1}{4}2k(2k-1) \\ &= \tfrac{1}{2}k. \end{aligned}$$

**Case 8:** $k$ odd, $a = b = 1$.

$$n(R(a,b)_v) - n(R(a+b)_v) = \tfrac{1}{2}(k+1).$$

Summarizing all eight cases, we see that

$$n(R(a,b)_v) - n(R(a+b)_v) > 0;$$

that is, $n(R_v)$ decreases, if $R$ increases.

Now, let $v \in S_{\mathbb{C}}$. This is the simple case, since $\Pi_v$ is fully induced and the cohomology of such representations is determined in [Borel and Wallach 1980, III, Theorem 3.3]. This is what we used in the proof of Proposition 5.1, where we also computed $q(L_{R,v}, \Pi_v)$ for a given quadruple $(R, \Pi, \nu, \lambda) \in M^{(j)}_{\mathcal{J},\{P\},\varphi_P}$. Again, the dimension of the unipotent radical of the parabolic subgroups $R(a,b)$ and $R(a+b)$ is easily calculated. We obtain

$$\begin{aligned} n(R(a,b)_v) - n(R(a+b)_v) &= \tfrac{1}{2}ak(ak-a) + \tfrac{1}{2}bk(bk-b) + abk^2 \\ &\qquad - \tfrac{1}{2}(a+b)k\bigl((a+b)k-(a+b)\bigr) \\ &= abk, \end{aligned}$$

now really for all cases of $a$ and $b$. Therefore, $n(R(a,b)_v)-n(R(a+b)_v)>0$, that is, $n(R_v)$ decreases, if $R$ increases also for $v\in S_{\mathbb{C}}$. This proves the lemma. $\square$

**Proposition 6.3.** *For every proper support* $\{P\}$, $L_P\cong \mathrm{GL}_k^{\ell}$, $\varphi_P$ *and every* $j$, $0\le j<m$, *the minimum*

$$\min_{(R,\Pi,\nu,\lambda)\in M^{(j)}_{\mathcal{J},\{P\},\varphi_P}}\left(\sum_{v\in S_\infty}\left\lceil \tfrac{1}{2}\dim_{\mathbb{R}} N_R(F_v)\right\rceil + q(L_{R,v},\Pi_v)\right)$$

*occurs at the standard parabolic subgroup* $R=L_RN_R$ *with* $L_R\cong \mathrm{GL}_{(\ell-1)k}\times\mathrm{GL}_k$.

*Proof.* By Lemma 6.1, we only need to check that for $R=R((\ell-1)k,k)$, the number $n(R_v)$ is minimal among all maximal parabolic subgroups $R((\ell-a)k,ak)$ for all places $v\in S_\infty$. Precisely as in the proof of Lemma 6.1, this is again a lengthy exercise using Proposition 5.1 and [Borel and Wallach 1980, III, Proposition 5.3]. Their use is justified step-by-step, as in the proof of Lemma 6.1.

Let $v\in S_{\mathbb{R}}$. First of all, we check that

$$n\big(R((\ell-1)k,k)\big)=\begin{cases}\frac14(n^2+(3-\ell)n-2k) & \text{for } k \text{ even},\\ \frac14(n^2+(3-\ell)n-2k-\ell) & \text{for } k \text{ odd, } \ell \text{ odd},\\ \frac14(n^2+(3-\ell)n-2k-\ell+2) & \text{for } k \text{ odd, } \ell \text{ even}.\end{cases}$$

Moreover, if $a\ge 2$, then $n\big(R((\ell-a)k,ak)\big)$ is given by

$$\tfrac14\big(a^2k^2-a^2k+ak+(\ell-a)^2k^2-(\ell-a)^2k+(\ell-a)k+2a(\ell-a)k^2\big)$$

for $k$ even, by

$$\tfrac14\big(a^2k^2-a^2k+ak+(\ell-a)^2k^2-(\ell-a)^2k+(\ell-a)k-\ell+2a(\ell-a)k^2\big)$$

for $k$ odd and $a$ or $\ell-a$ even, and by

$$\tfrac14\big(a^2k^2-a^2k+ak+(\ell-a)^2k^2-(\ell-a)^2k+(\ell-a)k-\ell+2a(\ell-a)k^2+2\big)$$

for $k$, $a$ and $\ell-a$ odd.

The expression $n\big(R((\ell-a)k,ak)\big)$ is a quadratic polynomial in $a$, with strictly negative leading coefficient. Hence, for $a\ge 2$, $n\big(R((\ell-a)k,ak)\big)$ is minimal at $a=2$ (and $a=\ell-2$). We obtain

$$n\big(R((\ell-2)k,2k)\big)=\begin{cases}\frac14(n^2+(5-\ell)n-8k) & \text{for } k \text{ even},\\ \frac14(n^2+(5-\ell)n-8k-\ell) & \text{for } k \text{ odd}.\end{cases}$$

Comparing $n\big(R((\ell-2)k,2k)\big)$ to $n\big(R((\ell-1)k,k)\big)$, in the cases when either $k$ is even or $k$ is odd and $\ell$ is odd, we see that $n\big(R((\ell-2)k,2k)\big)\ge n\big(R((\ell-1)k,k)\big)$ if and only if $\ell\ge 3$. But this is fine without loss of generality, since for $\ell=2$ the result holds trivially. If $k$ is odd and $\ell$ is even, $n\big(R((\ell-2)k,2k)\big)\ge n\big(R((\ell-1)k,k)\big)$ if and only if $\ell \geq 4$. This is satisfied by the same reason, since $\ell \geq 3$ is assumed to be even, hence without loss of generality already $\ell \geq 4$.

Now, let $v \in S_{\mathbb{C}}$. Then

$$n\big(R((\ell - a)k, ak)\big) = \frac{n^2 - (2a - \ell)n - 2a^2 k}{2},$$

for all $a \geq 1$. Clearly, this is minimal at $a = 1$ (and $a = \ell - 1$). $\square$

**Proposition 6.4.** *Let $\Pi \hookrightarrow \mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ be a residual automorphic representation of $G(\mathbb{A})$, which is $(\mathfrak{m}_G, K)$-cohomological with respect to $\mathcal{M}$. Assume that $\Pi$ is given by the triple $(L_P, \pi, \nu)$, where $L_P \cong \mathrm{GL}_k \times \cdots \times \mathrm{GL}_k$, $\ell k = n$. Then $q_{\mathsf{res}}$ is given by*

$$|S_{\mathbb{R}}| \cdot \tfrac{1}{4}(n^2 + (3 - \ell)n - 2k) + |S_{\mathbb{C}}| \cdot \tfrac{1}{2}(n^2 - (2 - \ell)n - 2k)$$

*for $k$ even, by*

$$|S_{\mathbb{R}}| \cdot \tfrac{1}{4}(n^2 + (3 - \ell)n - 2k - \ell) + |S_{\mathbb{C}}| \cdot \tfrac{1}{2}(n^2 - (2 - \ell)n - 2k)$$

*for $k$ odd and $\ell$ odd, and by*

$$|S_{\mathbb{R}}| \cdot \tfrac{1}{4}(n^2 + (3 - \ell)n - 2k - \ell + 2) + |S_{\mathbb{C}}| \cdot \tfrac{1}{2}(n^2 - (2 - \ell)n - 2k)$$

*for $k$ odd and $\ell$ even.*

*Proof.* This holds by the definition of $q_{\mathsf{res}}$ and Proposition 6.3. $\square$

**6B.** ***End of the proof of the Theorem 4.1.*** A direct comparison of $q_{\mathsf{min}}$ and $q_{\mathsf{res}}$ shows that if $\Pi \hookrightarrow \mathcal{A}_{\mathrm{res},\mathcal{J}}(G)$ is a residual automorphic representation of $G(\mathbb{A})$ that is $(\mathfrak{m}_G, K)$-cohomological with respect to $\mathcal{M}$, then $q_{\mathsf{min}} < q_{\mathsf{res}}$. Hence, Theorem 4.1 follows from our Theorem 3.2. $\square$

## Acknowledgements

We would like to thank G. Harder for many inspiring discussions on this topic at the Erwin Schrödinger Institute in Vienna.

## References

[Borel 1981] A. Borel, "Stable real cohomology of arithmetic groups II", pp. 21–55 in *Manifolds and Lie groups* (Notre Dame, In, 1980), edited by J. Hano et al., Progr. Math. **14**, Birkhäuser, Boston, 1981. MR 83h:22023 Zbl 0483.57026

[Borel 2007] A. Borel, "Automorphic forms on reductive groups", pp. 5–40 in *Automorphic forms and applications*, edited by P. Sarnak and F. Shahidi, IAS/Park City Math. Ser. **12**, Amer. Math. Soc., Providence, RI, 2007. MR 2008k:11045

[Borel and Jacquet 1979] A. Borel and H. Jacquet, "Automorphic forms and automorphic representations", pp. 189–202 in *Automorphic forms, representations and L-functions* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33,** Part 1, Amer. Math. Soc., Providence, RI, 1979. MR 81m:10055 Zbl 0414.22020

[Borel and Wallach 1980] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Annals of Mathematical Studies **94**, Princeton University Press, New Jersey, 1980. MR 83c:22018 Zbl 0443.22010

[Borel et al. 1996] A. Borel, J.-P. Labesse, and J. Schwermer, "On the cuspidal cohomology of $S$-arithmetic subgroups of reductive groups over number fields", *Compositio Math.* **102**:1 (1996), 1–40. MR 97j:11026 Zbl 0853.11044

[Enright 1979] T. J. Enright, "Relative Lie algebra cohomology and unitary representations of complex Lie groups", *Duke Math. J.* **46**:3 (1979), 513–525. MR 81i:22007 Zbl 0427.22010

[Franke 1998] J. Franke, "Harmonic analysis in weighted $L_2$-spaces", *Ann. Sci. École Norm. Sup.* (4) **31**:2 (1998), 181–279. MR 2000f:11065 Zbl 0938.11026

[Franke and Schwermer 1998] J. Franke and J. Schwermer, "A decomposition of spaces of automorphic forms, and the Eisenstein cohomology of arithmetic groups", *Math. Ann.* **311**:4 (1998), 765–790. MR 99k:11077 Zbl 0924.11042

[Grobner 2013] H. Grobner, "Residues of Eisenstein series and the automorphic cohomology of reductive groups", *Compos. Math.* **149**:7 (2013), 1061–1090. MR 3078638

[Knapp and Vogan 1995] A. W. Knapp and D. A. Vogan, Jr., *Cohomological induction and unitary representations*, Princeton Mathematical Series **45**, Princeton University Press, 1995. MR 96c:22023 Zbl 0863.22011

[Mœglin and Waldspurger 1989] C. Mœglin and J.-L. Waldspurger, "Le spectre résiduel de $\mathrm{GL}(n)$", *Ann. Sci. École Norm. Sup.* (4) **22**:4 (1989), 605–674. MR 91b:22028 Zbl 0696.10023

[Mœglin and Waldspurger 1995] C. Mœglin and J.-L. Waldspurger, *Spectral decomposition and Eisenstein series*, Cambridge Tracts in Mathematics **113**, Cambridge University Press, Cambridge, 1995. MR 97d:11083 Zbl 0846.11032

[Rohlfs and Speh 2011] J. Rohlfs and B. Speh, "Pseudo Eisenstein forms and the cohomology of arithmetic groups III: residual cohomology classes", pp. 501–523 in *On certain $L$-functions* (Purdue University, IN, 2007), edited by J. Arthur et al., Clay Math. Proc. **13**, Amer. Math. Soc., Providence, RI, 2011. MR 2012d:11123

[Vogan and Zuckerman 1984] D. A. Vogan, Jr. and G. J. Zuckerman, "Unitary representations with nonzero cohomology", *Compositio Math.* **53**:1 (1984), 51–90. MR 86k:22040 Zbl 0692.22008

Received December 20, 2013. Revised March 4, 2014.

HARALD GROBNER
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT WIEN
OSKAR-MORGENSTERN-PLATZ 1
1090 WIEN
AUSTRIA
harald.grobner@univie.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.47

# GRADIENT ESTIMATES AND ENTROPY FORMULAE OF POROUS MEDIUM AND FAST DIFFUSION EQUATIONS FOR THE WITTEN LAPLACIAN

GUANGYUE HUANG AND HAIZHONG LI

**We study gradient estimates for the positive solutions of the porous medium equations and the fast diffusion equations**

$$u_t = \Delta_\phi(u^p)$$

**associated with the Witten Laplacian on Riemannian manifolds. Under the assumption that the $m$-dimensional Bakry–Emery Ricci curvature is bounded from below, we obtain some gradient estimates which generalize some previous results of Lu et. al. and Huang et. al. As applications, several parabolic Harnack inequalities are obtained. Moreover, inspired by X.-D. Li's work, we also extend the entropy formulae introduced by Lu et. al. to the porous medium equations and the fast diffusion equations associated with the Witten Laplacian. We prove some monotonicity theorems for such entropy on compact Riemannian manifolds with nonnegative $m$-dimensional Bakry–Emery Ricci curvature.**

## 1. Introduction

Let $(M^n, g)$ be an $n$-dimensional complete Riemannian manifold. P. Li and Yau [1986] considered positive solutions of the heat equation

$$u_t = \Delta u \tag{1-1}$$

and proved the following gradient estimates.

**Theorem A** [Li and Yau 1986]. *Let $(M^n, g)$ be a complete Riemannian manifold with $\mathrm{Ric}(B_p(2R)) \geq -K$, where $\mathrm{Ric}(B_p(2R))$ denotes the Ricci curvature on the geodesic ball $B_p(2R)$ with radius $2R$ and $K$ is a nonnegative constant. Let $u$ be a*

Huang was supported by the NSFC (grant numbers 11001076, 11171091, 11371018). Li was supported by the NSFC (grant number 11271214).
*MSC2010:* primary 35B45; secondary 35K55.
*Keywords:* porous medium equation, fast diffusion equation, entropy formulae, Witten Laplacian.

*positive solution of (1-1) on $B_p(2R)\times[0,T]$. Then, on $B_p(R)$, we have*

$$\frac{|\nabla u|^2}{u^2}-\alpha\frac{u_t}{u}\le\frac{C(n)\alpha^2}{R^2}\left(\frac{\alpha^2}{\alpha-1}+\sqrt{K}R\right)+\frac{n\alpha^2K}{2(\alpha-1)}+\frac{n\alpha^2}{2t},\tag{1-2}$$

*where $\alpha>1$ is a constant and $C(n)$ is a constant depending only on $n$. Moreover, taking $R\to\infty$, (1-2) yields the following estimate on $(M^n,g)$:*

$$\frac{|\nabla u|^2}{u^2}-\alpha\frac{u_t}{u}\le\frac{n\alpha^2K}{2(\alpha-1)}+\frac{n\alpha^2}{2t}.\tag{1-3}$$

J. F. Li and X. J. Xu [2011] obtained new Li–Yau-type gradient estimates for positive solutions of the heat equation (1-1) on complete Riemannian manifolds. For related research and some improvements on Li–Yau-type gradient estimates of (1-1), see [Yau 1994; 1995; Bakry and Qian 1999; Hamilton 1993; Li 2005; Davies 1989]. The equation

$$u_t=\Delta(u^p)\tag{1-4}$$

with $p>1$ is called the porous medium equation, which is a nonlinear extension of the classical heat equation. For various values of $p>1$, it has appeared in different applications to model diffusive phenomena (see [Vázquez 2007; Aronson and Bénilan 1979; Lu et al. 2009] and the references therein). Equation (1-4) with $p\in(0,1)$ is called the fast diffusion equation, which appears in plasma physics and in geometric flows. However, there are remarkable differences between the porous medium equations and the fast diffusion equation; see [Vázquez 2006; Daskalopoulos and Kenig 2007]. For the study of gradient estimates of (1-4), see [Huang et al. 2013; Aronson and Bénilan 1979; Vázquez 2007; Xu 2012].

Lu, Ni, Vázquez, and Villani studied gradient estimates of (1-4) and proved the following results.

**Theorem B** [Lu et al. 2009, Theorem 3.3]. *Let $(M^n,g)$ be a complete Riemannian manifold with $\mathrm{Ric}(B_p(2R))\ge -K$, where $\mathrm{Ric}(B_p(2R))$ denotes the Ricci curvature on the geodesic ball $B_p(2R)$ with radius $2R$ and $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-4) with $p>1$. Let $v=(p/(p-1))u^{p-1}$ and $M=(p-1)\max_{B_p(2R)\times[0,T]}v$. Then, for any $\alpha>1$, on $B_p(R)$, we have*

$$\frac{|\nabla v|^2}{v}-\alpha\frac{v_t}{v}\le\frac{C(n)Ma\alpha^2}{R^2}\left(\frac{\alpha^2}{\alpha-1}\frac{ap^2}{p-1}+(1+\sqrt{K}R)\right)+\frac{\alpha^2}{\alpha-1}aMK+\frac{a\alpha^2}{t},\tag{1-5}$$

*where* $a = n(p-1)/(n(p-1)+2)$. *Moreover, taking* $R \to \infty$, (1-5) *yields the following estimate on* $(M^n, g)$:

$$\frac{|\nabla v|^2}{v} - \alpha \frac{v_t}{v} \leq \frac{\alpha^2}{\alpha-1} aMK + \frac{a\alpha^2}{t}. \tag{1-6}$$

Now we rewrite the inequality (1-6) as

$$|\nabla v|^2 - \alpha v_t \leq \frac{\alpha^2}{\alpha-1} aMKv + \frac{a\alpha^2 v}{t}. \tag{1-7}$$

Since $(p-1)v = pu^{p-1}$, we have $(p-1)v \to 1$ as $p \to 1$. As $p \to 1$, we have $M \to 1$,

$$|\nabla v|^2 \to \frac{|\nabla u|^2}{u^2}, \quad v_t \to \frac{u_t}{u}, \quad av \to \frac{n}{2}.$$

Consequently, (1-7) becomes Li and Yau's inequality (1-3). Therefore, for a complete noncompact Riemannian manifold $(M^n, g)$, estimate (1-6) in the result of Lu, Ni, Vázquez and Villani reduces to estimate (1-3) when $p \to 1$.

Let $\phi \in C^2(M^n)$. The Witten Laplacian associated with $\phi$ is defined by

$$\Delta_\phi = \Delta - \nabla\phi \cdot \nabla,$$

which is symmetric with respect to the $L^2(M^n)$ inner product under the weighted measure

$$d\mu = e^{-\phi}\, dv,$$

that is,

$$\int_{M^n} u \Delta_\phi v \, d\mu = -\int_{M^n} \nabla u \nabla v \, d\mu = \int_{M^n} v \Delta_\phi u \, d\mu \quad \text{for all } u, v \in C_0^\infty(M^n).$$

Following [Bakry and Émery 1985; Bakry 1994; Li 2005; Wei and Wylie 2009], we introduce the $m$-dimensional Bakry–Emery Ricci curvature associated with the Witten Laplacian by

$$\mathrm{Ric}_\phi^m = \mathrm{Ric} + \nabla^2\phi - \frac{1}{m-n} d\phi \otimes d\phi,$$

where $m \geq n$ is a constant and $m = n$ if and only if $\phi$ is a constant. Define

$$\mathrm{Ric}_\phi = \mathrm{Ric} + \nabla^2\phi.$$

Then $\mathrm{Ric}_\phi$ can be seen as the $\infty$-dimensional Bakry–Emery Ricci curvature. In this paper, we study the following equation associated with the Witten Laplacian:

$$u_t = \Delta_\phi(u^p) \tag{1-8}$$

with $p > 0$ and $p \neq 1$. For $p > 1$ and $p \in (0, 1)$, we derive an analogue of the estimates of Lu, Ni, Vázquez, and Villani and a Davies-type estimate. Moreover, for $p > 1$, we obtain a Hamilton-type estimate and an analogue of the estimates of Li and Xu. In particular, our results generalize the ones in [Huang et al. 2013].

First we consider gradient estimates of (1-8) under the assumption that the $m$-dimensional Bakry–Emery Ricci curvature is bounded from below, and obtain the following results. We set once and for all

$$\tilde{a} = \frac{m(p-1)}{m(p-1)+2}. \tag{1-9}$$

**Theorem 1.1.** *Let $(M^n, g)$ be a complete Riemannian manifold with*

$$\mathrm{Ric}_\phi^m(B_p(2R)) \geq -K,$$

*where $\mathrm{Ric}_\phi^m(B_p(2R))$ denotes the $m$-dimensional Bakry–Emery Ricci curvature on the geodesic ball $B_p(2R)$ with radius $2R$, and $K$ is a nonnegative constant. Let $u$ be a positive solution to the porous medium equation (1-8) with $p > 1$. Let $v = (p/(p-1))u^{p-1}$ and $M = (p-1)\max_{B_p(2R)\times[0,T]} v$. Then, for any $\alpha > 1$, on $B_p(R)$, we have*

$$\frac{|\nabla v|^2}{v} - \alpha\frac{v_t}{v} \leq \tilde{a}\alpha^2 M\frac{C(m)}{R^2}\left(\frac{\alpha^2}{\alpha-1}\,\frac{\tilde{a}p^2}{p-1} + 1 + \sqrt{K}R\coth(\sqrt{K}R)\right) + \frac{\alpha^2}{(\alpha-1)}\tilde{a}MK + \frac{\tilde{a}\alpha^2}{t}.$$

*Taking $R \to \infty$, we thus obtain the following estimate on $(M^n, g)$:*

$$\frac{|\nabla v|^2}{v} - \alpha\frac{v_t}{v} \leq \frac{\alpha^2}{\alpha-1}\tilde{a}MK + \frac{\tilde{a}\alpha^2}{t}. \tag{1-10}$$

**Corollary 1.2.** *Let $(M^n, g)$ be a complete Riemannian manifold with $\mathrm{Ric}_\phi^m \geq -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p > 1$. Set*

$$v = \frac{p}{p-1}u^{p-1}, \quad M = (p-1)\sup_{M^n\times[0,T]} v, \quad \widetilde{M} = \inf_{M^n\times[0,T]} v.$$

*Then, for any $x_1, x_2 \in M^n$, $0 < t_1 < t_2 < T$, $\alpha > 1$, we have*

$$v(x_1,t_1) \leq v(x_2,t_2)\left(\frac{t_2}{t_1}\right)^{\tilde{a}\alpha} \exp\left(\frac{\alpha\,\mathrm{dist}^2(x_2,x_1)}{4\widetilde{M}(t_2-t_1)} + \frac{\alpha}{\alpha-1}\tilde{a}MK(t_2-t_1)\right),$$

*where $\mathrm{dist}(x_2, x_1)$ is the distance between $x_1$ and $x_2$.*

**Theorem 1.3.** *Let $(M^n, g)$ and $K$ be as in Theorem 1.1. Let $u$ be a positive solution to the fast diffusion equation (1-8) with $p \in (1-2/m, 1)$. Set*

$$v = \frac{p}{p-1}u^{p-1}, \quad M = (1-p)\max_{B_p(2R)\times[0,T]}(-v).$$

*Then, for any* $0<\alpha<1$, *we have on* $B_p(R)$

$$-\frac{|\nabla v|^2}{v}+\alpha\frac{v_t}{v} \le \frac{(-\tilde{a})\alpha^2 M}{A(\varepsilon_1,\varepsilon_2)}\frac{C(m)}{R^2}\left(\frac{(-\tilde{a})\alpha^2 p^2}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}+1+\sqrt{K}R\coth(\sqrt{K}R)\right) + \frac{(-\tilde{a})\alpha^2 MK}{\sqrt{\varepsilon_1(1-\alpha)(1-\alpha-\tilde{a})A(\varepsilon_1,\varepsilon_2)}}+\frac{(-\tilde{a})\alpha^2}{A(\varepsilon_1,\varepsilon_2)t}, \tag{1-11}$$

*where* $\varepsilon_1,\varepsilon_2\in(0,1)$ *are positive constants satisfying*

$$A(\varepsilon_1,\varepsilon_2):=[1-\tilde{a}(1-\alpha)]-\frac{(1+\varepsilon_2)^2(1-\tilde{a})^2(1-\alpha)}{(1-\varepsilon_1)(1-\alpha-\tilde{a})}>0.$$

*Taking* $R\to\infty$ *and* $\alpha\to 1$, *we thus obtain the following estimate on* $(M^n,g)$ *with* $\mathrm{Ric}_\phi^m\ge 0$:

$$-\frac{|\nabla v|^2}{v}+\frac{v_t}{v}\le-\frac{\tilde{a}}{t}. \tag{1-12}$$

**Corollary 1.4.** *Let* $(M^n,g)$ *be a complete Riemannian manifold with* $\mathrm{Ric}_\phi^m\ge 0$. *Let* $u$ *be a positive solution to* (1-8) *with* $p\in(1-2/m,1)$. *Set*

$$v=\frac{p}{p-1}u^{p-1},\qquad M=(1-p)\sup_{M^n\times[0,T]}(-v),\qquad \widetilde{M}=\inf_{M^n\times[0,T]}(-v).$$

*Then, for any* $x_1,x_2\in M^n$ *and* $0<t_1<t_2<T$, *we have*

$$-v(x_2,t_2)\le -v(x_1,t_1)\left(\frac{t_2}{t_1}\right)^{-\tilde{a}}\exp\frac{\mathrm{dist}^2(x_2,x_1)}{4\widetilde{M}(t_2-t_1)}, \tag{1-13}$$

*where* $\mathrm{dist}(x_2,x_1)$ *is the distance between* $x_1$ *and* $x_2$.

**Remark 1.5.** Clearly, our estimate (1-10) reduces to (1-6) (see [Lu et al. 2009]) by letting $m=n$. Moreover, for $p\in(0,1)$, [Lu et al. 2009, Theorem 4.1] can be obtained from our Theorem 1.3 by taking $m=n$.

**Theorem 1.6.** *Let* $(M^n,g)$ *and* $K$ *be as in Theorem 1.1. Let* $u$ *be a positive solution to the fast diffusion equation* (1-8) *with* $p\in(1-2/m,1)$. *Set*

$$v=\frac{p}{p-1}u^{p-1},\qquad M=(1-p)\max_{B_p(2R)\times[0,T]}(-v).$$

*Then, for any* $0<\alpha<1$, *we have on* $B_p(R)$

$$-\frac{|\nabla v|^2}{v}+\alpha\frac{v_t}{v}\le\Bigg\{C(\tilde{a},\alpha)\frac{p}{\sqrt{1-p}}\sqrt{M}\frac{C}{R}+\bigg[\bigg(\frac{\alpha^2}{2(1-\alpha)}+2(1-\tilde{a})\bigg)MK+\frac{1-\alpha-\tilde{a}}{t}+(1-p)(1-\alpha-\tilde{a})M\frac{C(m)}{R^2}\big(1+\sqrt{K}R\coth(\sqrt{K}R)\big)\bigg]^{\frac{1}{2}}\Bigg\}^2.$$

*Taking $R\to\infty$, we thus obtain the following estimate on $(M^n,g)$:*

$$-\frac{|\nabla v|^2}{v}+\alpha\frac{v_t}{v}\le\bigg(\frac{\alpha^2}{2(1-\alpha)}+2(1-\tilde{a})\bigg)MK+\frac{1-\alpha-\tilde{a}}{t}. \tag{1-14}$$

**Corollary 1.7.** *Let $(M^n,g)$ be a complete Riemannian manifold with $\mathrm{Ric}_\phi^m\ge -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p\in(1-2/m,1)$. Let*

$$v=\frac{p}{p-1}u^{p-1},\qquad M=(1-p)\sup_{M^n\times[0,T]}(-v),\qquad \widetilde{M}=\inf_{M^n\times[0,T]}(-v).$$

*Then, for any $x_1,x_2\in M^n$, $0<t_1<t_2<T$, $0<\alpha<1$, we have*

$$-v(x_2,t_2)\le -v(x_1,t_1)\bigg(\frac{t_2}{t_1}\bigg)^{(1-\alpha-\tilde{a})/\alpha}\times\exp\bigg(\frac{\alpha\,\mathrm{dist}^2(x_2,x_1)}{4\widetilde{M}(t_2-t_1)}+\bigg(\frac{\alpha}{2(1-\alpha)}+\frac{2(1-\tilde{a})}{\alpha}\bigg)MK(t_2-t_1)\bigg)$$

*where $\mathrm{dist}(x_2,x_1)$ is the distance between $x_1$ and $x_2$.*

**Remark 1.8.** For complete Riemannian manifolds with $p\in(0,1)$, Corollary 4.2 of [Lu et al. 2009] shows that, if $\mathrm{Ric}\ge 0$, then

$$-\frac{|\nabla v|^2}{v}+\frac{v_t}{v}\le-\frac{a}{t}; \tag{1-15}$$

while if $\mathrm{Ric}\ge -K$ and $0<\alpha<1$, then, for any $\varepsilon>0$ satisfying

$$C(a,\alpha,\varepsilon):=1+(-a)(1-\alpha)-\frac{(1-\alpha)(1-a)^2}{(1-\alpha)-a-(1-\alpha)\varepsilon^2}>0,$$

we have

$$-\frac{|\nabla v|^2}{v}+\alpha\frac{v_t}{v}\le\frac{(-a)\alpha^2}{C(a,\alpha,\varepsilon)}\bigg(\frac{1}{t}+\frac{\sqrt{C(a,\alpha,\varepsilon)}}{(1-\alpha)\varepsilon}MK\bigg). \tag{1-16}$$

Obviously, our estimate (1-14) reduces to (1-15) when $m=n$ and $\alpha\to 1$. Moreover, (1-14) is independent of $\varepsilon$.

Denote by $R$ the scalar curvature of the metric $g$. Perelman [2002] introduced the $\mathcal{W}$-entropy functional as

$$\mathcal{W}(g, f, \tau) = \int_{M^n} \big(\tau(R + |\nabla f|^2) + f - n\big) \frac{e^{-f}}{(4\pi\tau)^{n/2}}\, dv, \tag{1-17}$$

where $\tau$ is a positive scale parameter and $f \in C^\infty(M^n)$ satisfies

$$\int_{M^n} \frac{e^{-f}}{(4\pi\tau)^{n/2}}\, dv = 1.$$

By [Perelman 2002], we know that the $\mathcal{W}$-entropy is monotone increasing under the Ricci flow, and its critical points are given by gradient shrinking solitons. Ni [2004a; 2004b] considered the $\mathcal{W}$-entropy for the linear heat equation

$$u_\tau = \Delta u \tag{1-18}$$

on complete Riemannian manifolds. More precisely, Ni [2004a] introduced the $\mathcal{W}$-entropy associated with (1-18) by

$$\mathcal{W}(g, f, \tau) = \int_{M^n} [\tau|\nabla f|^2 + f - n] \frac{e^{-f}}{(4\pi\tau)^{n/2}}\, dv, \tag{1-19}$$

where $u = \dfrac{e^{-f}}{(4\pi\tau)^{n/2}}$ is a positive solution to (1-18) and $\int_{M^n} u\, dv = 1$, and proved that

$$\frac{d}{d\tau}\mathcal{W}(g, f, \tau) = -2\int_{M^n} \tau\left(\left|\nabla^2 f - \frac{g}{2\tau}\right|^2 + \mathrm{Ric}(\nabla f, \nabla f)\right) u\, dv. \tag{1-20}$$

Thus, if the Ricci curvature is nonnegative, the $\mathcal{W}$-entropy defined by (1-19) is nonincreasing on complete Riemannian manifolds. For research on the monotonicity of $\mathcal{W}$-entropy for other geometric heat flows on Riemannian manifolds, see [Kotschwar and Ni 2009; Ecker 2007; Ni 2004a; 2004b; Lu et al. 2009]. X.-D. Li [2011; 2012; 2013] studied the $\mathcal{W}_m$-entropy associated with the Witten Laplacian to the linear heat equation

$$u_\tau = \Delta_\phi u \tag{1-21}$$

on complete Riemannian manifolds satisfying the $\mu$-bounded geometry condition. More precisely, [Li 2012] introduced the $\mathcal{W}_m$-entropy associated with (1-21) by

$$\mathcal{W}_m(g, f, \tau) = \int_{M^n} [\tau|\nabla f|^2 + f - m] \frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu, \tag{1-22}$$

where $u = \dfrac{e^{-f}}{(4\pi\tau)^{m/2}}$ is a positive solution to (1-21), and proved that if there exist two constants $m > n$ and $K \geq 0$ such that $\mathrm{Ric}_\phi^m \geq -K$, then

$$\frac{d}{d\tau}\mathcal{W}_m(g, f, \tau) = -2\int_{M^n} \tau\left(\left|\nabla^2 f - \frac{g}{2\tau}\right|^2 + \mathrm{Ric}_\phi^m(\nabla f, \nabla f)\right)u\, d\mu - \frac{2}{m-n}\int_{M^n} \tau\left(\nabla\phi\,\nabla f + \frac{m-n}{2\tau}\right)^2 u\, d\mu. \tag{1-23}$$

Thus, if $\mathrm{Ric}_\phi^m \geq 0$, then $\mathcal{W}_m(g, f, \tau)$ is nonincreasing along the heat equation (1-21). For the study of the Witten Laplacian associated with the $m$-dimensional Bakry–Emery Ricci curvature on complete Riemannian manifolds, see [Wei and Wylie 2009; Wang 2004; 1997; Qian 1998; 1997; Ni 2002; Li 2005; Fang et al. 2009; Bakry and Qian 2005; Bakry 1994; Bakry and Émery 1985]. Let $u$ be a positive solution to (1-4), and let $v = (p/(p-1))u^{p-1}$. Lu et. al. [2009] introduced

$$\mathcal{N}_p(g, u, t) = -t^a \int_{M^n} uv\, dv$$

and

$$\mathcal{W}_p(g, u, t) = \frac{d}{dt}[t\mathcal{N}_p(g, u, t)] = t^{a+1}\int_{M^n}\left(p\frac{|\nabla v|^2}{v} - \frac{a+1}{t}\right)uv\, dv, \tag{1-24}$$

where $a = \dfrac{n(p-1)}{n(p-1)+2}$. They proved that if $M^n$ is compact,

$$\begin{aligned}\frac{d}{dt}\mathcal{W}_p(g, u, t) &= -2(p-1)t^{a+1}\int_{M^n}\left(\left|\nabla^2 v + \frac{g}{[n(p-1)+2]t}\right|^2 + \mathrm{Ric}(\nabla v, \nabla v)\right)uv\, dv \\ &\quad - 2t^{a+1}\int_{M^n}\left((p-1)\Delta v + \frac{a}{t}\right)^2 uv\, dv.\end{aligned} \tag{1-25}$$

In particular, if the Ricci curvature is nonnegative, the entropy defined in (1-24) is nonincreasing on compact Riemannian manifolds when $p > 1$. For $p < 1$, using the Cauchy–Schwarz inequality, they proved from (1-25) that

$$\begin{aligned}\frac{d}{dt}\mathcal{W}_p(g, u, t) \\ \leq -2t^{a+1}\int_{M^n}\left[\frac{n(p-1)+1}{n(p-1)}\left((p-1)\Delta v + \frac{a}{t}\right)^2 + (p-1)\,\mathrm{Ric}(\nabla v, \nabla v)\right]uv\, dv.\end{aligned} \tag{1-26}$$

Clearly, if the Ricci curvature is nonnegative and $p \in (1 - 1/n, 1)$, then (1-26) shows that $(d/dt)\mathcal{W}_p(g, u, t) \leq 0$ and the entropy defined in (1-24) is nonincreasing on compact Riemannian manifolds.

Inspired by [Li 2012], in this paper we also study the $\mathcal{W}_{p,m}$-entropy for (1-8) associated with the Witten Laplacian on compact Riemannian manifolds with $p>0$ and $p\neq 1$. First we define

$$\mathcal{N}_{p,m}(g,u,t) = -t^{\tilde a}\int_{M^n} uv\,d\mu, \tag{1-27}$$

where the $\mathcal{W}_{p,m}$-entropy is defined by

$$\mathcal{W}_{p,m}(g,u,t) = \frac{d}{dt}[t\mathcal{N}_{p,m}(g,u,t)], \tag{1-28}$$

When the $m$-dimensional Bakry–Emery Ricci curvature is bounded from below, we prove the following.

**Theorem 1.9.** *Let $(M^n, g)$ be a compact Riemannian manifold. If $u$ is a positive solution to the porous medium equation (1-8) with $p>1$, then*

$$\frac{d}{dt}\mathcal{N}_{p,m}(g,u,t) = -t^{\tilde a}\int_{M^n}\left((p-1)\Delta_\phi v + \frac{\tilde a}{t}\right)uv\,d\mu, \tag{1-29}$$

*where $v=(p/(p-1))u^{p-1}$. In particular, if $\mathrm{Ric}_\phi^m \geq 0$, then $\mathcal{N}_{p,m}(g,u,t)$ is nonincreasing in $t$. Moreover,*

$$\mathcal{W}_{p,m}(g,u,t) = t^{\tilde a+1}\int_{M^n}\left(p\frac{|\nabla v|^2}{v} - \frac{\tilde a+1}{t}\right)uv\,d\mu \tag{1-30}$$

*and*

$$\begin{aligned}\frac{d}{dt}\mathcal{W}_{p,m}(g,u,t) &= -2(p-1)t^{\tilde a+1}\int_{M^n}\left(\left|\nabla^2 v + \frac{g}{[m(p-1)+2]t}\right|^2 + \mathrm{Ric}_\phi^m(\nabla v,\nabla v)\right.\\ &\qquad\left. + \frac{1}{m-n}\left|\nabla\phi\nabla v - \frac{m-n}{[m(p-1)+2]t}\right|^2\right)uv\,d\mu\\ &\quad -2t^{\tilde a+1}\int_{M^n}\left|(p-1)\Delta_\phi v + \frac{\tilde a}{t}\right|^2 uv\,d\mu.\end{aligned} \tag{1-31}$$

*In particular, if $\mathrm{Ric}_\phi^m \geq 0$, then $\mathcal{W}_{p,m}(g,u,t)$ is nonincreasing in $t$.*

**Theorem 1.10.** *Let $(M^n, g)$ be a compact Riemannian manifold. If $u$ is a positive solution to the fast diffusion equation (1-8) with $p\in(0,1)$, then*

$$\frac{d}{dt}\mathcal{N}_{p,m}(g,u,t) = -t^{\tilde a}\int_{M^n}\left((p-1)\Delta_\phi v + \frac{\tilde a}{t}\right)uv\,d\mu, \tag{1-32}$$

*where $v = \dfrac{p}{p-1}u^{p-1}$. In particular, if $\mathrm{Ric}_\phi^m \geq 0$ and $p\in(1-2/m, 1)$, then*

*$\mathcal{N}_{p,m}(g,u,t)$ is nonincreasing in $t$. Moreover, we have*

$$\mathcal{W}_{p,m}(g,u,t)=t^{\tilde{a}+1}\int_{M^n}\left(p\frac{|\nabla v|^2}{v}-\frac{\tilde{a}+1}{t}\right)uv\,d\mu, \tag{1-33}$$

*and, for any positive constant $\varepsilon\geq m-n$ and $1-\dfrac{1}{n+\varepsilon}\leq p\leq 1-\dfrac{m-n}{m\varepsilon}$,*

$$\begin{aligned}\frac{d}{dt}&\mathcal{W}_{p,m}(g,u,t)\\ &\leq 2t^{\tilde{a}+1}\int_{M^n}\left((1-p)\,\mathrm{Ric}_\phi^m(\nabla v,\nabla v)+\left(\frac{1-n(1-p)}{n(1-p)}-\frac{\varepsilon}{n}\right)\left|(p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right|^2\right.\\ &\qquad\left.+\left(\frac{m(1-p)}{n(m-n)}-\frac{1}{n\varepsilon}\right)\left|\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right|^2\right)uv\,d\mu.\end{aligned} \tag{1-34}$$

*In particular, if $\mathrm{Ric}_\phi^m\geq 0$, then $\mathcal{W}_{p,m}(g,u,t)$ is nonincreasing in $t$.*

**Remark 1.11.** If $m=n$, we see that $\phi$ is a constant. Then (1-31) becomes [Lu et al. 2009, (5.6)]. By letting $m=n$ and $\varepsilon\to 0$, (1-34) becomes (1-26), which is [Lu et al. 2009, Corollary 5.10].

**Remark 1.12.** After we submitted our paper, the referee pointed out to us [Li and Li 2013; Wang and Chen 2013; Wang et al. 2013], in which some related problems are studied. Specifically, S. Li and X.-D. Li [2013] derived the $W$-entropy formula for the Witten Laplacian on manifolds with time dependent metrics and potentials. Wang and Chen [2013] obtained Aronson–Bénilan-type estimates for the porous medium equations associated with the Witten Laplacian. Wang, Yang, and Chen [Wang et al. 2013] studied the weighted $p$-Laplacian heat equation and proved an optimal gradient estimate and the $W$-entropy monotonicity formula, which generalized the results of [Kotschwar and Ni 2009]. We note that the first version of this paper was posted on arXiv (1203.5482) on March 25 of 2012.

## 2. Proofs of Theorems 1.1 and 1.3

Let $v=(p/(p-1))u^{p-1}$. By virtue of (1-8), we have $v_t=(p-1)v\Delta_\phi v+|\nabla v|^2$, which is equivalent to

$$\frac{v_t}{v}=(p-1)\Delta_\phi v+\frac{|\nabla v|^2}{v}. \tag{2-1}$$

As in [Lu et al. 2009], we introduce the differential operator

$$\mathcal{L}=\partial_t-(p-1)v\Delta_\phi. \tag{2-2}$$

**Lemma 2.1.** *Let* $F = \frac{|\nabla v|^2}{v} - \alpha \frac{v_t}{v} - \varphi$, *where* $\alpha = \alpha(t)$ *and* $\varphi = \varphi(t)$ *are functions of* $t$. *Set*

$$L_0(F) = -\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 - 2(p-1)\operatorname{Ric}_\phi^m(\nabla v, \nabla v) + 2p\nabla v \nabla F + (1-\alpha)\Big(\frac{v_t}{v}\Big)^2 - \alpha' \frac{v_t}{v} - \varphi'.$$

(1) *If* $p > 1$, *then* $\mathscr{L}(F) \le L_0(F)$.

(2) *If* $p \in (0, 1)$, *then* $\mathscr{L}(F) \ge L_0(F)$.

*Proof.* We only give the proof for the case where $p > 1$; the other case is similar. By a direct calculation, we have

$$\mathscr{L}\Big(\frac{f}{g}\Big) = \frac{1}{g}\mathscr{L}(f) - \frac{f}{g^2}\mathscr{L}(g) + 2(p-1)v\nabla\frac{f}{g}\nabla\log g \quad \text{for all } f, g \in C^\infty(M). \tag{2-3}$$

Using (2-1), we obtain

$$\mathscr{L}(v_t) = (p-1)v_t\Delta_\phi v + 2\nabla v \nabla v_t. \tag{2-4}$$

It is well known that, for the $m$-dimensional Bakry–Emery Ricci curvature, we have the following Bochner formula (for the elementary proof, see [Ledoux 2000; Li 2005]):

$$\begin{aligned} \tfrac{1}{2}\Delta_\phi(|\nabla w|^2) &= |\nabla^2 w|^2 + \nabla w \nabla \Delta_\phi w + \operatorname{Ric}_\phi(\nabla w, \nabla w) \\ &\ge \frac{1}{n}|\Delta w|^2 + \nabla w \nabla \Delta_\phi w + \operatorname{Ric}_\phi(\nabla w, \nabla w) \\ &\ge \frac{1}{m}|\Delta_\phi w|^2 + \nabla w \nabla \Delta_\phi w + \operatorname{Ric}_\phi^m(\nabla w, \nabla w). \end{aligned} \tag{2-5}$$

It follows from $p > 1$ that

$$\begin{aligned} \mathscr{L}(|\nabla v|^2) &\le 2\nabla v \nabla v_t - 2(p-1)v\Big(\frac{1}{m}|\Delta_\phi v|^2 + \nabla v \nabla \Delta_\phi v + \operatorname{Ric}_\phi^m(\nabla v, \nabla v)\Big) \\ &= 2\nabla v \nabla[(p-1)v\Delta_\phi v + |\nabla v|^2] \\ &\qquad - 2(p-1)v\Big(\frac{1}{m}|\Delta_\phi v|^2 + \nabla v \nabla \Delta_\phi v + \operatorname{Ric}_\phi^m(\nabla v, \nabla v)\Big) \\ &= 2(p-1)|\nabla v|^2\Delta_\phi v + 2\nabla v \nabla(|\nabla v|^2) \\ &\qquad - \frac{2(p-1)}{m}v(\Delta_\phi v)^2 - 2(p-1)v\operatorname{Ric}_\phi^m(\nabla v, \nabla v). \end{aligned}$$

Applying this and (2-4) to (2-3) yields

$$\mathscr{L}\left(\frac{v_t}{v}\right)=(p-1)\frac{v_t}{v}\Delta_\phi v+\frac{2}{v}\nabla v\nabla v_t-\frac{v_t}{v}\frac{|\nabla v|^2}{v}+2(p-1)v\nabla\frac{v_t}{v}\nabla\log v \tag{2-6}$$

and

$$\begin{aligned}\mathscr{L}\left(\frac{|\nabla v|^2}{v}\right)&\le 2(p-1)\frac{|\nabla v|^2}{v}\Delta_\phi v+\frac{2}{v}\nabla v\nabla(|\nabla v|^2)\\&-\frac{2(p-1)}{m}(\Delta_\phi v)^2-2(p-1)\operatorname{Ric}_\phi^m(\nabla v,\nabla v)-\frac{|\nabla v|^4}{v^2}+2(p-1)v\nabla\frac{|\nabla v|^2}{v}\nabla\log v,\end{aligned}$$

and hence

$$\begin{aligned}\mathscr{L}(F)&=\mathscr{L}\left(\frac{|\nabla v|^2}{v}\right)-\alpha\mathscr{L}\left(\frac{v_t}{v}\right)-\alpha'\frac{v_t}{v}-\varphi'\\&\le 2(p-1)\frac{|\nabla v|^2}{v}\Delta_\phi v+\frac{2}{v}\nabla v\nabla(|\nabla v|^2)-\frac{2(p-1)}{m}(\Delta_\phi v)^2\\&\quad-2(p-1)\operatorname{Ric}_\phi^m(\nabla v,\nabla v)-\frac{|\nabla v|^4}{v^2}+2(p-1)v\nabla\frac{|\nabla v|^2}{v}\nabla\log v\\&\quad-\alpha(p-1)\frac{v_t}{v}\Delta_\phi v-\alpha\frac{2}{v}\nabla v\nabla v_t+\alpha\frac{v_t}{v}\frac{|\nabla v|^2}{v}\\&\quad-2\alpha(p-1)v\nabla\frac{v_t}{v}\nabla\log v-\alpha'\frac{v_t}{v}-\varphi'.\end{aligned}\tag{2-7}$$

Noticing that

$$2(p-1)v\nabla\frac{|\nabla v|^2}{v}\nabla\log v-2\alpha(p-1)v\nabla\frac{v_t}{v}\nabla\log v=2(p-1)\nabla v\nabla F$$

and

$$\frac{2}{v}\nabla v\nabla(|\nabla v|^2)-\alpha\frac{2}{v}\nabla v\nabla v_t=\frac{2}{v}\nabla v\nabla[(F+\varphi)v]=2(F+\varphi)\frac{|\nabla v|^2}{v}+2\nabla v\nabla F,$$

we obtain

$$\begin{aligned}2(p-1)v\left(\nabla\frac{|\nabla v|^2}{v}-\alpha\nabla\frac{v_t}{v}\right)\nabla\log v+\frac{2}{v}\nabla v\nabla(|\nabla v|^2)-\alpha\frac{2}{v}\nabla v\nabla v_t&\\=2p\nabla v\nabla F+2(F+\varphi)\frac{|\nabla v|^2}{v}&\\=2p\nabla v\nabla F+2\left(\frac{|\nabla v|^2}{v}-\alpha\frac{v_t}{v}\right)\frac{|\nabla v|^2}{v}.&\end{aligned}\tag{2-8}$$

On the other hand, using (2-1) again, we have

$$
\begin{aligned}
(2\text{-}9)\quad & 2(p-1)\frac{|\nabla v|^2}{v}\Delta_\phi v-\frac{|\nabla v|^4}{v^2}-\alpha(p-1)\frac{v_t}{v}\Delta_\phi v+\alpha\frac{v_t}{v}\frac{|\nabla v|^2}{v}\\
&=2\frac{|\nabla v|^2}{v}\left(\frac{v_t}{v}-\frac{|\nabla v|^2}{v}\right)-\frac{|\nabla v|^4}{v^2}-\alpha\frac{v_t}{v}\left(\frac{v_t}{v}-\frac{|\nabla v|^2}{v}\right)+\alpha\frac{v_t}{v}\frac{|\nabla v|^2}{v}\\
&=(2\alpha+2)\frac{v_t}{v}\frac{|\nabla v|^2}{v}-3\frac{|\nabla v|^4}{v^2}-\alpha\left(\frac{v_t}{v}\right)^2.
\end{aligned}
$$

Combining (2-8) with (2-9) gives

$$
\begin{aligned}
(2\text{-}10)\quad & 2(p-1)v\nabla\frac{|\nabla v|^2}{v}\nabla\log v-2\alpha(p-1)v\nabla\frac{v_t}{v}\nabla\log v+\frac{2}{v}\nabla v\nabla(|\nabla v|^2)\\
&-\alpha\frac{2}{v}\nabla v\nabla v_t+2(p-1)\frac{|\nabla v|^2}{v}\Delta_\phi v-\frac{|\nabla v|^4}{v^2}-\alpha(p-1)\frac{v_t}{v}\Delta_\phi v+\alpha\frac{v_t}{v}\frac{|\nabla v|^2}{v}\\
&=2p\nabla v\nabla F-\left(\frac{v_t}{v}-\frac{|\nabla v|^2}{v}\right)^2+(1-\alpha)\left(\frac{v_t}{v}\right)^2\\
&=2p\nabla v\nabla F-[(p-1)\Delta_\phi v]^2+(1-\alpha)\left(\frac{v_t}{v}\right)^2.
\end{aligned}
$$

Putting (2-10) into (2-7) yields

$$
\begin{aligned}
\mathcal{L}(F)\le{}&-\frac{2(p-1)}{m}(\Delta_\phi v)^2-2(p-1)\,\mathrm{Ric}_\phi^m(\nabla v,\nabla v)+2p\nabla v\nabla F\\
&\qquad-[(p-1)\Delta_\phi v]^2+(1-\alpha)\left(\frac{v_t}{v}\right)^2-\alpha'\frac{v_t}{v}-\varphi'\\
={}&-\frac{1}{\tilde a}[(p-1)\Delta_\phi v]^2-2(p-1)\,\mathrm{Ric}_\phi^m(\nabla v,\nabla v)+2p\nabla v\nabla F\\
&\qquad+(1-\alpha)\left(\frac{v_t}{v}\right)^2-\alpha'\frac{v_t}{v}-\varphi',
\end{aligned}
$$

which completes the proof of (1) in Lemma 2.1. □

*Proof of Theorem 1.1.* Let $\xi$ be a cut-off function such that $\xi(r)=1$ for $r\le 1$, $\xi(r)=0$ for $r\ge 2$, $0\le\xi(r)\le 1$, and

$$0\ge\xi'(r)\ge -c_1\xi^{1/2}(r),\qquad \xi''(r)\ge -c_2,$$

for positive constants $c_1$ and $c_2$. With $\rho(x)$ the distance between $x$ and $p$ in $M^n$, let

$$\psi(x)=\xi\left(\frac{\rho(x)}{R}\right).$$

Making use of an argument of Calabi [1958] (see also [Cheng and Yau 1975]), we can assume without loss of generality that the function $\psi$ is smooth in $B_p(2R)$. Then we have

$$\frac{|\nabla\psi|^2}{\psi}\le\frac{C}{R^2}. \tag{2-11}$$

By the comparison theorem with respect to the Witten Laplacian (see [Li 2005, p. 1324])

$$\Delta_\phi \rho \geq \sqrt{(m-1)K}\coth\Bigl(\sqrt{\frac{K}{m-1}}\rho\Bigr),$$

we have

$$\Delta_\phi \psi = \frac{\xi'\Delta_\phi\rho}{R} + \frac{\xi''|\nabla\rho|^2}{R^2} \geq -\frac{C(m)}{R^2}(1+\sqrt{K}R\coth(\sqrt{K}R)). \tag{2-12}$$

Define $\tilde{F} = |\nabla v|^2/v - \alpha v_t/v$, where $\alpha > 1$ is a constant. Under the assumption that $\mathrm{Ric}_\phi^m \geq -K$, Lemma 2.1(1) shows that

$$\begin{aligned} \mathscr{L}(\tilde{F}) &\leq -\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2(p-1)K|\nabla v|^2 + 2p\nabla v\nabla\tilde{F} \\ &\leq -\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2MK\frac{|\nabla v|^2}{v} + 2p\nabla v\nabla\tilde{F}. \end{aligned} \tag{2-13}$$

Set $G = t\psi\tilde{F}$. Next we will apply the maximum principle to $G$ on $B_p(2R)\times[0,T]$. Assume $G$ achieves its maximum at the point $(x_0, s) \in B_p(2R)\times[0,T]$ and assume $G(x_0, s) > 0$ (otherwise the proof is trivial), which implies $s > 0$. Then, at the point $(x_0, s)$, we have

$$\mathscr{L}(G) \geq 0, \qquad \nabla\tilde{F} = -\frac{\tilde{F}}{\psi}\nabla\psi,$$

and, by use of (2-13), we have

$$\begin{aligned} 0 &\leq \mathscr{L}(G) \\ &= s\psi\mathscr{L}(\tilde{F}) - s(p-1)v\tilde{F}\Delta_\phi\psi - 2s(p-1)v\nabla\tilde{F}\nabla\psi + \psi\tilde{F} \\ &= s\psi\mathscr{L}(\tilde{F}) - (p-1)v\frac{\Delta_\phi\psi}{\psi}G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s} \\ &\leq s\psi\Bigl(-\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2MK\frac{|\nabla v|^2}{v} + 2p\nabla v\nabla\tilde{F}\Bigr) \\ &\qquad\qquad - (p-1)v\frac{\Delta_\phi\psi}{\psi}G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s} \\ &\leq -\frac{s\psi}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2s\psi MK\frac{|\nabla v|^2}{v} + 2\frac{p}{\sqrt{p-1}}\sqrt{M}G\frac{|\nabla v|}{\sqrt{v}}\frac{|\nabla\psi|}{\psi} \\ &\qquad\qquad - (p-1)v\frac{\Delta_\phi\psi}{\psi}G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s}. \end{aligned} \tag{2-14}$$

Applying

$$[(p-1)\Delta_\phi v]^2 = \frac{1}{\alpha^2}\tilde{F}^2 + \frac{2(\alpha-1)}{\alpha^2}\tilde{F}\frac{|\nabla v|^2}{v} + \Bigl(\frac{\alpha-1}{\alpha}\Bigr)^2\frac{|\nabla v|^4}{v^2}$$

to (2-14), we obtain

$$0 \le -\frac{1}{\tilde{a}s\alpha^2}G^2 - \frac{2(\alpha-1)\psi}{\tilde{a}\alpha^2}G\frac{|\nabla v|^2}{v} - \frac{s\psi^2}{\tilde{a}}\left(\frac{\alpha-1}{\alpha}\right)^2\frac{|\nabla v|^4}{v^2} + 2s\psi^2 MK\frac{|\nabla v|^2}{v} + 2\frac{p}{\sqrt{p-1}}\sqrt{M\psi}G\frac{|\nabla v|}{\sqrt{v}}\frac{|\nabla \psi|}{\sqrt{\psi}} - (p-1)v(\Delta_\phi\psi)G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi}G + \frac{\psi G}{s}. \tag{2-15}$$

Since $-Ax^2+Bx \le \dfrac{B^2}{4A}$, we have

$$-\frac{s\psi^2}{\tilde{a}}\left(\frac{\alpha-1}{\alpha}\right)^2\frac{|\nabla v|^4}{v^2} + 2s\psi^2 MK\frac{|\nabla v|^2}{v} \le \frac{\tilde{a}\alpha^2 s\psi^2M^2K^2}{(\alpha-1)^2}$$

and

$$-\frac{2(\alpha-1)\psi}{\tilde{a}\alpha^2}G\frac{|\nabla v|^2}{v} + \frac{2p}{\sqrt{p-1}}\sqrt{M\psi}G\frac{|\nabla v|}{\sqrt{v}}\frac{|\nabla\psi|}{\sqrt{\psi}} \le \frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{|\nabla\psi|^2}{\psi}G.$$

We now set

$$P(K,R) = 1 + \sqrt{K}R\coth(\sqrt{K}R). \tag{2-16}$$

From (2-15) we obtain

$$\begin{aligned}0 &\le -\frac{1}{\tilde{a}s\alpha^2}G^2 + \frac{\tilde{a}\alpha^2 s\psi^2M^2K^2}{(\alpha-1)^2} + \frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{|\nabla\psi|^2}{\psi}G \\ &\qquad - (p-1)v(L\psi)G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi}G + \frac{\psi G}{s} \\ &\le -\frac{1}{\tilde{a}s\alpha^2}G^2 + \left(\frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s}\right)G \\ &\qquad + \frac{\tilde{a}\alpha^2 s\psi^2M^2K^2}{(\alpha-1)^2}.\end{aligned}$$

Solving this quadratic inequality for $G$ yields

$$\begin{aligned}G &\le \frac{\tilde{a}s\alpha^2}{2}\Bigg\{\frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s} \\ &\qquad + \left[\left(\frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s}\right)^2 + \frac{4\psi^2M^2K^2}{(\alpha-1)^2}\right]^{\frac{1}{2}}\Bigg\} \\ &\le \tilde{a}s\alpha^2\left\{\frac{\tilde{a}\alpha^2p^2M}{2(p-1)(\alpha-1)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s} + \frac{\psi MK}{\alpha-1}\right\}.\end{aligned}$$

Hence we have

$$\begin{aligned}G(x,T) &\le G(x_0,s)\\ &\le \tilde{a}T\alpha^2\frac{C(m)}{R^2}\left(\frac{\alpha^2}{(p-1)(\alpha-1)}\tilde{a}p^2+P(K,R)\right)M+\frac{\alpha^2}{\alpha-1}\tilde{a}TMK+\tilde{a}\alpha^2.\end{aligned}$$

This implies that, for all $x\in B_p(R)$,

$$F(x,T)\le \tilde{a}\alpha^2M\frac{C(m)}{R^2}\left(\frac{\alpha^2}{\alpha-1}\frac{\tilde{a}p^2}{p-1}+P(K,R)\right)+\frac{\alpha^2}{\alpha-1}\tilde{a}MK+\frac{\tilde{a}\alpha^2}{T}. \tag{2-17}$$

Since $T$ is arbitrary, we complete the proof of Theorem 1.1. □

*Proof of Corollary 1.2.* Along the lines of Li and Yau, we will establish a Harnack inequality from a general estimate

$$\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}-\varphi(t)\le 0. \tag{2-18}$$

Rewrite (2-18) as

$$-\frac{v_t}{v}\le\frac{1}{\alpha(t)}\left(\varphi(t)-\frac{|\nabla v|^2}{v}\right).$$

Let $f=\log v$. Then we have

$$-f_t=-\frac{v_t}{v}\le\frac{1}{\alpha(t)}\left(\varphi(t)-\frac{|\nabla v|^2}{v}\right)\le\frac{1}{\alpha(t)}(\varphi(t)-\widetilde{M}|\nabla f|^2).$$

Let $\gamma$ be a shortest geodesic joining $x_1$ and $x_2$, and set $\gamma:[t_1,t_2]\to M^n$, $\gamma(t_1)=x_1$, $\gamma(t_2)=x_2$. Define a curve $\zeta$ in $M^n\times(0,\infty)$, $\zeta:[t_1,t_2]\to M^n\times(0,\infty)$ by $\zeta(t)=(\gamma(t),t)$. Then $\zeta(t_1)=(x_1,t_1)$ and $\zeta(t_2)=(x_2,t_2)$. Set $\rho=d(x_1,x_2)$. Then $|\dot\gamma|=\rho/(t_2-t_1)$ and

$$\begin{aligned}f(x_1,t_1)-f(x_2,t_2)&=\int_{t_2}^{t_1}\frac{d}{dt}f(\zeta(t))\,dt=\int_{t_2}^{t_1}(\langle\dot\gamma,\nabla f\rangle+f_t)\,dt\\ &=\int_{t_1}^{t_2}(-\langle\dot\gamma,\nabla f\rangle+(-f_t))\,dt\\ &\le\int_{t_1}^{t_2}\left(|\dot\gamma||\nabla f|+\frac{1}{\alpha(t)}(\varphi(t)-\widetilde{M}|\nabla f|^2)\right)dt\\ &=\int_{t_1}^{t_2}\left(-\frac{\widetilde{M}}{\alpha(t)}|\nabla f|^2+|\dot\gamma||\nabla f|\right)dt+\int_{t_1}^{t_2}\frac{\varphi(t)}{\alpha(t)}\,dt\\ &\le\frac{\rho^2}{4\widetilde{M}(t_2-t_1)^2}\int_{t_1}^{t_2}\alpha(t)\,dt+\int_{t_1}^{t_2}\frac{\varphi(t)}{\alpha(t)}\,dt,\end{aligned} \tag{2-19}$$

where in the last inequality we used $-Ax^2+Bx\le\dfrac{B^2}{4A}$ and $|\dot\gamma|=\rho/(t_2-t_1)$.

Let $\alpha > 1$ be a constant and set $\varphi = \frac{\alpha^2}{(\alpha-1)}\tilde{a}MK + \frac{\tilde{a}\alpha^2}{t}$. We have from (2-19)

$$f(x_1,t_1) - f(x_2,t_2) \le \int_{t_1}^{t_2} \left( \frac{\alpha\rho^2}{4\widetilde{M}(t_2-t_1)^2} + \frac{\alpha}{\alpha-1}\tilde{a}MK + \frac{\tilde{a}\alpha}{t} \right) dt \tag{2-20}$$
$$= \frac{\alpha\rho^2}{4\widetilde{M}(t_2-t_1)} + \frac{\alpha}{\alpha-1}\tilde{a}MK(t_2-t_1) + \tilde{a}\alpha \log \frac{t_2}{t_1}.$$

Therefore, we arrive at

$$v(x_1,t_1) \le v(x_2,t_2)\left(\frac{t_2}{t_1}\right)^{\tilde{a}\alpha} \exp\left( \frac{\alpha\rho^2}{4\widetilde{M}(t_2-t_1)} + \frac{\alpha}{\alpha-1}\tilde{a}MK(t_2-t_1) \right). \qquad \square$$

*Proof of Theorem 1.3.* When $p \in (0,1)$, we have $v < 0$, and from Lemma 2.1(2)

$$\mathscr{L}(-\widetilde{F}) \le \frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2(p-1)\operatorname{Ric}_\phi^m(\nabla v, \nabla v) + 2p\nabla v \nabla(-\widetilde{F}) - (1-\alpha)\left(\frac{v_t}{v}\right)^2,$$

which implies

$$\mathscr{L}(-\widetilde{F}) \le \frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2MK\frac{|\nabla v|^2}{-v} + 2p\nabla v \nabla(-\widetilde{F}) - (1-\alpha)\left(\frac{v_t}{v}\right)^2. \tag{2-21}$$

Define $G = t\psi(-\widetilde{F})$. We'll apply the maximum principle to $G$ on $B_p(2R)\times[0,T]$. Assume $G$ achieves its maximum at the point $(x_0,s) \in B_p(2R)\times[0,T]$ and assume $G(x_0,s) > 0$ (otherwise the proof is trivial), which implies $s > 0$. Then, at the point $(x_0,s)$, we have

$$\mathscr{L}(G) \ge 0, \qquad \nabla(-\widetilde{F}) = -\frac{-\widetilde{F}}{\psi}\nabla\psi$$

and, by use of (2-21), we have

$$0 \le \mathscr{L}(G) = s\psi\mathscr{L}(-\widetilde{F}) - (p-1)v\frac{\Delta_\phi\psi}{\psi}G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s} \tag{2-22}$$
$$\le s\psi\left( \frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2MK\frac{|\nabla v|^2}{-v} + 2p\nabla v\nabla(-\widetilde{F}) \right)$$
$$- (p-1)v\frac{\Delta_\phi\psi}{\psi}G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s} - (1-\alpha)s\psi\left(\frac{v_t}{v}\right)^2$$
$$\le \frac{s\psi}{\tilde{a}}[(p-1)\Delta_\phi v]^2 + 2s\varphi MK\frac{|\nabla v|^2}{-v}$$
$$+ 2\frac{p}{\sqrt{1-p}}\sqrt{M}G\frac{|\nabla v|}{\sqrt{-v}}\frac{|\nabla\psi|}{\psi} - (p-1)v\frac{\Delta_\phi\psi}{\psi}G$$
$$+ 2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G + \frac{G}{s} - (1-\alpha)s\psi\left(\frac{v_t}{v}\right)^2.$$

Applying the equalities

$$[(p-1)\Delta_\phi v]^2=\frac{1}{\alpha^2}\tilde{F}^2+\frac{2(\alpha-1)}{\alpha^2}\tilde{F}\frac{|\nabla v|^2}{v}+\left(\frac{\alpha-1}{\alpha}\right)^2\frac{|\nabla v|^4}{v^2}$$

and

$$\left(\frac{v_t}{v}\right)^2=\frac{1}{\alpha^2}\left(-\tilde{F}+\frac{|\nabla v|^2}{v}\right)^2=\frac{1}{\alpha^2}(-\tilde{F})^2+\frac{2}{\alpha^2}(-\tilde{F})\frac{|\nabla v|^2}{v}+\frac{1}{\alpha^2}\frac{|\nabla v|^4}{v^2}$$

to (2-22), we obtain

$$\begin{aligned}(2\text{-}23)\quad 0\le \frac{1}{\tilde{a}s\alpha^2}\bigg(\big(1-\tilde{a}(1-\alpha)\big)G^2-2(1-\tilde{a})(1-\alpha)s\psi G\frac{|\nabla v|^2}{-v}\\ +s^2\psi^2(1-\alpha)(1-\alpha-\tilde{a})\frac{|\nabla v|^4}{v^2}\bigg)\\ +2s\psi^2MK\frac{|\nabla v|^2}{-v}+2\frac{p}{\sqrt{1-p}}\sqrt{M\psi}G\frac{|\nabla v|}{\sqrt{-v}}\frac{|\nabla\psi|}{\sqrt{\psi}}\\ -(p-1)v(\Delta_\phi\psi)G+2(p-1)v\frac{|\nabla\psi|^2}{\psi}G+\frac{\psi G}{s}.\end{aligned}$$

Next we employ a method similar to that in [Lu et al. 2009, Theorem 4.1]. Since $p\in(1-2/m,1)$, we have $\tilde{a}<0$. Thus we have, for any positive constants $\varepsilon_1,\varepsilon_2$,

$$2s\psi^2MK\frac{|\nabla v|^2}{-v}\le -\varepsilon_1\frac{s^2\psi^2}{\tilde{a}s\alpha^2}(1-\alpha)(1-\alpha-\tilde{a})\frac{|\nabla v|^4}{v^2}-\frac{1}{\varepsilon_1}\frac{\tilde{a}s\alpha^2(p-1)^2\psi^2M^2K^2}{(1-\alpha)(1-\alpha-\tilde{a})},$$

and

$$\begin{aligned}2\frac{p}{\sqrt{1-p}}\sqrt{M\psi}G\frac{|\nabla v|}{\sqrt{-v}}\frac{|\nabla\psi|}{\sqrt{\psi}}\\ \le -\varepsilon_2\frac{2}{\tilde{a}s\alpha^2}(1-\tilde{a})(1-\alpha)s\psi G\frac{|\nabla v|^2}{-v}-\frac{\tilde{a}\alpha^2p^2M}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}\frac{|\nabla\psi|^2}{\psi}G.\end{aligned}$$

Hence we get from (2-23) that

$$\begin{aligned}0\le -\frac{1}{\tilde{a}s\alpha^2}\bigg(-\big(1-\tilde{a}(1-\alpha)\big)G^2+2(1+\varepsilon_2)(1-\tilde{a})(1-\alpha)s\psi G\frac{|\nabla v|^2}{-v}\\ -(1-\varepsilon_1)s^2\psi^2(1-\alpha)(1-\alpha-\tilde{a})\frac{|\nabla v|^4}{v^2}\bigg)\\ -\frac{1}{\varepsilon_1}\frac{as\alpha^2\psi^2M^2K^2}{(1-\alpha)(1-\alpha-\tilde{a})}-\frac{\tilde{a}\alpha^2p^2M}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}\frac{|\nabla\psi|^2}{\psi}G\\ -(p-1)v(\Delta_\phi\psi)G+2(p-1)v\frac{|\nabla\psi|^2}{\psi}G+\frac{\psi G}{s},\end{aligned}$$

which can be rewritten as

$$0 \le \frac{1}{\tilde{a}s\alpha^2}\left(1-\tilde{a}(1-\alpha)-\frac{(1+\varepsilon_2)^2(1-\tilde{a})^2(1-\alpha)}{(1-\varepsilon_1)(1-\alpha-\tilde{a})}\right)G^2 - \frac{1}{\varepsilon_1}\frac{\tilde{a}s\alpha^2\psi^2M^2K^2}{(1-\alpha)(1-\alpha-\tilde{a})} - \frac{\tilde{a}\alpha^2p^2M}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}\frac{|\nabla\psi|^2}{\psi}G - (p-1)v(\Delta_\phi\psi)G + 2(p-1)v\frac{|\nabla\psi|^2}{\psi}G + \frac{\psi G}{s}. \tag{2-24}$$

Taking $\varepsilon_1, \varepsilon_2$ such that

$$1-\tilde{a}(1-\alpha)-\frac{(1+\varepsilon_2)^2(1-\tilde{a})^2(1-\alpha)}{(1-\varepsilon_1)(1-\alpha-\tilde{a})} =: A(\varepsilon_1,\varepsilon_2) > 0, \tag{2-25}$$

we obtain from (2-24), with $P(K,R)$ as in (2-16),

$$0 \le -\frac{1}{(-\tilde{a})s\alpha^2}A(\varepsilon_1,\varepsilon_2)G^2 + \left(\frac{(-\tilde{a})\alpha^2p^2M}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s}\right)G + \frac{(-\tilde{a})s\alpha^2\psi^2M^2K^2}{\varepsilon_1(1-\alpha)(1-\alpha-\tilde{a})}.$$

Solving this quadratic inequality for $G$ yields

$$G \le \frac{(-\tilde{a})s\alpha^2}{A(\varepsilon_1,\varepsilon_2)}\left(\frac{(-\tilde{a})\alpha^2p^2M}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)}\frac{C}{R^2} + M\frac{C(m)}{R^2}P(K,R) + \frac{\psi}{s} + \frac{\psi MK}{\sqrt{\varepsilon_1(1-\alpha)(1-\alpha-\tilde{a})}}\sqrt{A(\varepsilon_1,\varepsilon_2)}\right). \tag{2-26}$$

Hence we have

$$\begin{aligned} G(x,T) &\le G(x_0,s) \\ &\le \frac{(-\tilde{a})T\alpha^2M}{A(\varepsilon_1,\varepsilon_2)}\frac{C(m)}{R^2}\left(\frac{(-\tilde{a})\alpha^2p^2}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)} + P(K,R)\right) \\ &\quad + \frac{(-\tilde{a})T\alpha^2MK}{\sqrt{\varepsilon_1(1-\alpha)(1-\alpha-\tilde{a})A(\varepsilon_1,\varepsilon_2)}} + \frac{(-\tilde{a})\alpha^2}{A(\varepsilon_1,\varepsilon_2)}, \end{aligned} \tag{2-27}$$

and, for $x \in B_p(R)$,

$$-F(x,t) \le \frac{(-\tilde{a})\alpha^2M}{A(\varepsilon_1,\varepsilon_2)}\frac{C(m)}{R^2}\left(\frac{(-\tilde{a})\alpha^2p^2}{2\varepsilon_2(1-\tilde{a})(1-\alpha)(1-p)} + P(K,R)\right) + \frac{(-\tilde{a})\alpha^2MK}{\sqrt{\varepsilon_1(1-\alpha)(1-\alpha-\tilde{a})A(\varepsilon_1,\varepsilon_2)}} + \frac{(-\tilde{a})\alpha^2}{A(\varepsilon_1,\varepsilon_2)t}.$$

This completes the proof of Theorem 1.3. □

*Proof of Corollary 1.4.* Choosing $f=\log(-v)$ and $\varphi(t)=-\dfrac{\tilde a}{t}$, we get from (2-19)

$$f(x_2,t_2)-f(x_1,t_1)\le\int_{t_1}^{t_2}\left(\frac{\rho^2}{4\widetilde M(t_2-t_1)^2}-\frac{\tilde a}{t}\right)dt=\frac{\rho^2}{4\widetilde M(t_2-t_1)}-\tilde a\log\frac{t_2}{t_1}.\quad\square$$

## 3. Proof of Theorem 1.6

*Proof.* Define $\bar F=\dfrac{|\nabla v|^2}{v}-\alpha\dfrac{v_t}{v}$, where $\alpha\in(0,1)$ is constant. Lemma 2.1(2) shows that

$$\mathscr{L}(-\bar F)\le\frac{1}{\tilde a}[(p-1)\Delta_\phi v]^2+2MK\frac{|\nabla v|^2}{-v}+2p\nabla v\nabla(-\bar F)-(1-\alpha)\left(\frac{v_t}{v}\right)^2$$
$$=\frac{1}{\tilde a\alpha^2}\left(-\bar F-(1-\alpha)\frac{|\nabla v|^2}{-v}\right)^2+2MK\frac{|\nabla v|^2}{-v}+2p\nabla v\nabla(-\bar F)-\frac{1-\alpha}{\alpha^2}\left(-\bar F-\frac{|\nabla v|^2}{-v}\right)^2. \tag{3-1}$$

Let $G=t\psi(-\bar F)$. We apply the maximum principle to $G$ on $B_p(2R)\times[0,T]$ and assume that $G$ achieves its maximum at the point $(x_0,s)\in B_p(2R)\times[0,T]$ with $G(x_0,s)>0$ (otherwise the proof is trivial). At the point $(x_0,s)$, we have

$$\mathscr{L}(G)\ge 0,\qquad\nabla(-\bar F)=-\frac{-\bar F}{\psi}\nabla\psi,$$

and, by use of (3-1), we get

$$0\le\mathscr{L}(G)=s\psi\mathscr{L}(-\bar F)-(p-1)v\frac{\Delta_\phi\psi}{\psi}G+2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G+\frac{G}{s}$$
$$\le\frac{s\psi}{\tilde a\alpha^2}\left(-\bar F-(1-\alpha)\frac{|\nabla v|^2}{-v}\right)^2+2s\varphi MK\frac{|\nabla v|^2}{-v}$$
$$+2\frac{p}{\sqrt{1-p}}\sqrt{M}G\frac{|\nabla v|}{\sqrt{-v}}\frac{|\nabla\psi|}{\psi}-\frac{1-\alpha}{\alpha^2}s\psi\left(-\bar F-\frac{|\nabla v|^2}{-v}\right)^2$$
$$-(p-1)v\frac{\Delta_\phi\psi}{\psi}G+2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G+\frac{G}{s}.$$

Let $\dfrac{|\nabla v|^2}{-v}=\mu(-\bar F)$ at the point $(x_0,s)$. Then we have $\mu\ge0$ and

$$0\le\frac{1}{\tilde a\alpha^2 s\psi}[1-(1-\alpha)\mu]^2G^2+2\mu MKG+\frac{2\sqrt{\mu}}{\sqrt{s\psi}}\frac{p}{\sqrt{1-p}}\sqrt{M}G^{3/2}\frac{|\nabla\psi|}{\psi}$$
$$-\frac{1-\alpha}{\alpha^2}\frac{1}{s\psi}(1-\mu)^2G^2-(p-1)v\frac{\Delta_\phi\psi}{\psi}G+2(p-1)v\frac{|\nabla\psi|^2}{\psi^2}G+\frac{G}{s}. \tag{3-2}$$

Multiplying both sides of (3-2) by $s\psi/G$ yields

$$0 \le \frac{1}{\tilde{a}\alpha^2}[1-(1-\alpha)\mu]^2 G + 2\mu MKs\psi + 2\sqrt{\mu s}\frac{p}{\sqrt{1-p}}\sqrt{M}\frac{|\nabla\psi|}{\sqrt{\psi G}} - \frac{1-\alpha}{\alpha^2}(1-\mu)^2 G - (p-1)sv\Delta_\phi\psi + 2(p-1)sv\frac{|\nabla\psi|^2}{\psi} + \psi. \tag{3-3}$$

Introducing

$$\begin{aligned}\tilde{A} &= \frac{1}{-\tilde{a}\alpha^2}[1-(1-\alpha)\mu]^2 + \frac{1-\alpha}{\alpha^2}(1-\mu)^2,\\ \tilde{B} &= \sqrt{\mu s}\frac{p}{\sqrt{1-p}}\sqrt{M}\frac{|\nabla\psi|}{\sqrt{\psi}},\\ \tilde{C} &= 2\mu MKs\psi + (1-p)s(-v)\left(-\Delta_\phi\psi + 2\frac{|\nabla\psi|^2}{\psi}\right) + \psi,\end{aligned}$$

we write (3-3) as

$$0 \le -\tilde{A}G + 2\tilde{B}G^{1/2} + \tilde{C}. \tag{3-4}$$

It is easy to see that

$$\begin{aligned}\frac{1}{\tilde{A}} &= \frac{(-\tilde{a})\alpha^2}{[1-(1-\alpha)\mu]^2 + (-\tilde{a})(1-\alpha)(1-\mu)^2}\\ &= \frac{(-\tilde{a})\alpha^2}{1+(-\tilde{a})(1-\alpha)-2(1-\alpha)(1-\tilde{a})\mu+(1-\alpha)(1-\alpha-\tilde{a})\mu^2} \le 1-\alpha-\tilde{a}\end{aligned}$$

and

$$\begin{aligned}\frac{2\mu}{\tilde{A}} &= \frac{2(-\tilde{a})\alpha^2\mu}{1+(-\tilde{a})(1-\alpha)-2(1-\alpha)(1-\tilde{a})\mu+(1-\alpha)(1-\alpha-\tilde{a})\mu^2}\\ &\le \frac{(-\tilde{a})\alpha^2}{\sqrt{[1+(-\tilde{a})(1-\alpha)](1-\alpha)(1-\alpha-\tilde{a})}-(1-\alpha)(1-\tilde{a})}\\ &= \sqrt{[1/(1-\alpha)+(-\tilde{a})](1-\alpha-\tilde{a})}+(1-\tilde{a})\\ &\le \frac{\alpha^2}{2(1-\alpha)}+2(1-\tilde{a}),\end{aligned} \tag{3-5}$$

where the last inequality used that $\sqrt{xy} \le \frac{1}{2}(x+y)$. Hence there exists a constant $C(\tilde{a},\alpha)$ such that $\sqrt{\mu}/\tilde{A} \le C(\tilde{a},\alpha)$. Now, regarding (3-4) as a quadratic inequality in $\sqrt{G}$ gives

$$\sqrt{G} \le 2\tilde{B}/\tilde{A} + \sqrt{\tilde{C}/\tilde{A}},$$

and therefore

$$G^{1/2} \le C(\tilde{a},\alpha)\sqrt{sM}\frac{p}{\sqrt{1-p}}\frac{C}{R}+\left[\left(\frac{\alpha^2}{2(1-\alpha)}+2(1-\tilde{a})\right)MKs+1-\alpha-\tilde{a} + (1-p)(1-\alpha-\tilde{a})Ms\frac{C(m)}{R^2}P(K,R)\right]^{\frac{1}{2}} \tag{3-6}$$

Hence, for $x \in B_p(R)$, we have

$$\begin{aligned}&-\frac{|\nabla v|^2}{v}+\alpha\frac{v_t}{v}\\ &\quad\le\left\{C(\tilde{a},\alpha)\frac{p}{\sqrt{1-p}}\sqrt{M}\frac{C}{R}+\left[\left(\frac{\alpha^2}{2(1-\alpha)}+2(1-\tilde{a})\right)MK+\frac{1-\alpha-\tilde{a}}{t}\right.\right.\\ &\qquad\left.\left.+(1-p)(1-\alpha-\tilde{a})M\frac{C(m)}{R^2}P(K,R)\right]^{\frac{1}{2}}\right\}^2.\end{aligned} \tag{3-7}$$

This completes the proof of Theorem 1.6. □

On the other hand, under the assumption that $\mathrm{Ric}_\phi^m \ge -K$ and $p>1$, Lemma 2.1(1) shows that

$$\begin{aligned}&\mathscr{L}(F)\\ &\quad\le -\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2+2(p-1)K|\nabla v|^2+2p\nabla v\nabla F+(1-\alpha)\left(\frac{v_t}{v}\right)^2-\alpha'\frac{v_t}{v}-\varphi'\\ &\quad\le -\frac{1}{\tilde{a}}[(p-1)\Delta_\phi v]^2+2MK\frac{|\nabla v|^2}{v}+2p\nabla v\nabla F+(1-\alpha)\left(\frac{v_t}{v}\right)^2-\alpha'\frac{v_t}{v}-\varphi'.\end{aligned}$$

Following the methods in [Huang et al. 2013], we can prove the following results.

**Theorem 3.1.** *Let $(M^n, g)$ be a complete Riemannian manifold with*

$$\mathrm{Ric}_\phi^m(B_p(2R)) \ge -K,$$

*where $\mathrm{Ric}_\phi^m(B_p(2R))$ denotes the $m$-dimensional Bakry–Emery Ricci curvature on the geodesic ball $B_p(2R)$ with radius $2R$, and $K$ is a nonnegative constant. Let $u$ be a positive solution to the porous medium equation (1-8) with $p>1$. Set*

$$v=\frac{p}{p-1}u^{p-1}, \qquad M=(p-1)\max_{B_p(2R)\times[0,T]} v.$$

*Then, for any $\alpha>1$ and with $\tilde{a}$ as in (1-9), we have on $B_p(R)$*

$$\begin{aligned}&\frac{|\nabla v|^2}{v}-\alpha\frac{v_t}{v}\\ &\quad\le \tilde{a}\alpha^2\left\{\frac{\sqrt{\tilde{a}}\alpha p\sqrt{M}}{\sqrt{p-1}\sqrt{\alpha-1}}\frac{C(m)}{R}+\left(\frac{1}{t}+\frac{MK}{2(\alpha-1)}+M\frac{C(m)}{R^2}P(K,R)\right)^{\frac{1}{2}}\right\}^2.\end{aligned}$$

*Taking $R \to \infty$, we thus obtain the following estimate on $(M^n, g)$:*

$$\frac{|\nabla v|^2}{v} - \alpha \frac{v_t}{v} \le \frac{\alpha^2}{2(\alpha-1)} \tilde{a} M K + \frac{\tilde{a}\alpha^2}{t}. \tag{3-8}$$

**Corollary 3.2.** *Let $(M^n, g)$ be a complete noncompact Riemannian manifold with $\mathrm{Ric}_\phi^m \ge -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p > 1$. Set*

$$v = \frac{p}{p-1} u^{p-1}, \quad M = (p-1) \sup_{M^n \times [0,T]} v, \quad \widetilde{M} = \inf_{M^n \times [0,T]} v.$$

*Then, for any $x_1, x_2 \in M^n$, $0 < t_1 < t_2 < T$, $\alpha > 1$, we have*

$$v(x_1, t_1) \le v(x_2, t_2) \left(\frac{t_2}{t_1}\right)^{\tilde{a}\alpha} \exp\left(\frac{\alpha \operatorname{dist}^2(x_2, x_1)}{4\widetilde{M}(t_2 - t_1)} + \frac{\alpha}{2(\alpha-1)} \tilde{a} M K (t_2 - t_1)\right),$$

*where $\operatorname{dist}(x_2, x_1)$ is the distance between $x_1$ and $x_2$.*

**Theorem 3.3.** *Let $(M^n, g)$ and $K$ be as in Theorem 3.1. Let $u$ be a positive solution to the porous medium equation (1-8) with $p > 1$. Set*

$$v = \frac{p}{p-1} u^{p-1}, \quad M = (p-1) \max_{B_p(2R) \times [0,T]} v.$$

*Then, for any $\alpha > 1$, we have on $B_p(R)$*

$$\frac{|\nabla v|^2}{v} - \alpha(t) \frac{v_t}{v} \le \tilde{a}\alpha^2(t) M \frac{C(m)}{R^2} \left(\frac{p^2 \tilde{a} \alpha^2(t)}{2(p-1)(\alpha(t) - 1)} + 3 + \sqrt{K} R \coth(\sqrt{K} R)\right) + \frac{\tilde{a}\alpha^2(t)}{t},$$

*where $\alpha(t) = e^{2MKt}$. Taking $R \to \infty$, we thus obtain the following estimate on $(M^n, g)$:*

$$\frac{|\nabla v|^2}{v} - \alpha(t) \frac{v_t}{v} \le \frac{\tilde{a}\alpha^2(t)}{t}. \tag{3-9}$$

**Corollary 3.4.** *Let $(M^n, g)$ be a complete noncompact Riemannian manifold with $\mathrm{Ric}_\phi^m \ge -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p > 1$. Set*

$$v = \frac{p}{p-1} u^{p-1}, \quad M = (p-1) \sup_{M^n \times [0,T]} v, \quad \widetilde{M} = \inf_{M^n \times [0,T]} v.$$

*Then, for any $x_1, x_2 \in M^n$, $0 < t_1 < t_2 < T$, $\alpha > 1$, we have*

$$v(x_1, t_1) \le v(x_2, t_2) \exp\left\{\frac{e^{2MKt_2} - e^{2MKt_1}}{2MK} \left(\frac{\operatorname{dist}^2(x_2, x_1)}{4\widetilde{M}(t_2 - t_1)^2} + \frac{\tilde{a}}{t_1}\right)\right\},$$

*where $\operatorname{dist}(x_2, x_1)$ is the distance between $x_1$ and $x_2$.*

**Remark 3.5.** Theorems 3.1 and 3.3 reduce to Theorems 1.1 and 1.2 from [Huang et al. 2013], respectively, by letting $m=n$. In particular, the estimate (3-8) improves (1-10) on complete Riemannian manifolds.

**Theorem 3.6.** *Let $(M^n, g)$ and $K$ be as in Theorem 3.1. Let $u$ be a positive solution to the porous medium equation (1-8) with $p>1$. Let $v=(p/(p-1))u^{p-1}$ and $M=(p-1)\max_{B_p(2R)\times[0,T]} v$. Then, on $B_p(R)$, we have*

$$\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}-\varphi(t)\le \tilde{a}M\frac{C(m)}{R^2}\left(1+\sqrt{K}R\coth(\sqrt{K}R)+\frac{\tilde{a}p^2}{(p-1)\tanh(MKt)}\right),$$

*where $\alpha(t)$, $\varphi(t)$ are given by*

$$\begin{aligned}\varphi(t)&=\tilde{a}MK\big(\coth(MKt)+1\big),\\ \alpha(t)&=1+\frac{\cosh(MKt)\sinh(MKt)-MKt}{\sinh^2(MKt)}.\end{aligned}\tag{3-10}$$

*Taking $R\to\infty$, we thus obtain the following estimate on $(M^n, g)$:*

$$\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}-\varphi(t)\le 0.\tag{3-11}$$

**Corollary 3.7.** *Let $(M^n, g)$ be a complete noncompact Riemannian manifold with $\mathrm{Ric}_\phi^m\ge -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p>1$. Let $v=(p/(p-1))u^{p-1}$ and $M=(p-1)\sup_{M^n\times[0,T]} v$, $\widetilde{M}=\inf_{M^n\times[0,T]} v$. Then, for any $x_1, x_2\in M^n$, $0<t_1<t_2<T$, $\alpha>1$, we have*

$$v(x_1,t_1)\le v(x_2,t_2)A_1(t_1,t_2)\exp\left(\frac{\mathrm{dist}^2(x_2,x_1)}{4\widetilde{M}(t_2-t_1)}(1+A_2(t_1,t_2))\right),$$

*where $\mathrm{dist}(x_2,x_1)$ is the distance between $x_1$ and $x_2$ and*

$$\begin{aligned}A_1(t_1,t_2)&=\left(\frac{\exp(2MKt_2)-2MKt_2-1}{\exp(2MKt_1)-2MKt_1-1}\right)^{\tilde{a}/2},\\ A_2(t_1,t_2)&=\frac{t_2\coth(MKt_2)-t_1\coth(MKt_1)}{t_2-t_1}.\end{aligned}$$

**Theorem 3.8.** *Let $(M^n, g)$ and $K$ be as in Theorem 3.1. Let $u$ be a positive solution to the porous medium equation (1-8) with $p>1$. Let $v=(p/(p-1))u^{p-1}$ and $M=(p-1)\max_{B_p(2R)\times[0,T]} v$. Then, on $B_p(R)$, we have*

$$\begin{aligned}&\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}-\varphi(t)\\ &\qquad\le \tilde{a}\alpha^2(t)M\frac{C(m)}{R^2}\left(1+\sqrt{K}R\coth(\sqrt{K}R)+\frac{\tilde{a}p^2\alpha^2(t)}{(p-1)\tanh(MKt)}\right),\end{aligned}$$

*where*

$$\varphi(t)=\frac{\tilde a}{t}+\tilde a MK+\frac{\tilde a}{3}(MK)^2 t \quad \text{and} \quad \alpha(t)=1+\frac{2}{3}MKt. \tag{3-12}$$

*Taking $R\to\infty$, we thus obtain the following estimate on $(M^n, g)$:*

$$\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}-\varphi(t)\le 0. \tag{3-13}$$

**Corollary 3.9.** *Let $(M^n, g)$ be a complete noncompact Riemannian manifold with $\mathrm{Ric}^m_\phi \ge -K$, where $K$ is a nonnegative constant. Let $u$ be a positive solution to (1-8) with $p>1$. Set*

$$v=p/(p-1)u^{p-1},\quad M=(p-1)\sup_{M^n\times[0,T]} v,\quad \widetilde M=\inf_{M^n\times[0,T]} v.$$

*Then, for any $x_1, x_2\in M^n$, $0<t_1<t_2<T$, $\alpha>1$, we have*

$$v(x_1,t_1)\le v(x_2,t_2)\left(\frac{t_2}{t_1}\right)^{\tilde a}\left(\frac{1+\frac{2}{3}MKt_2}{1+\frac{2}{3}MKt_1}\right)^{-\tilde a/4} \times\exp\left(\frac{\mathrm{dist}^2(x_2,x_1)}{4\widetilde M(t_2-t_1)}\left(1+\tfrac{1}{3}MK(t_2+t_1)\right)+\frac{\tilde a}{2}MK(t_2-t_1)\right),$$

*where $\mathrm{dist}(x_2, x_1)$ is the distance between $x_1$ and $x_2$.*

**Remark 3.10.** Our Theorems 3.6 and 3.8 reduce to Theorems 1.3 and 1.4 from [Huang et al. 2013], respectively, by taking $m=n$. Moreover, when $t$ is small enough, $\alpha(t)$ and $\varphi(t)$ defined by (3-10) and (3-12) both satisfy $\alpha(t)\to 1$ and $\varphi(t)\le 2\tilde a MK+\tilde a/t$. Hence (3-11) and (3-13) show

$$\frac{|\nabla v|^2}{v}-\alpha(t)\frac{v_t}{v}\le 2\tilde a MK+\frac{\tilde a}{t}. \tag{3-14}$$

Clearly, for $t$ small enough, (3-14) is better than (1-10). In this sense, (3-11) and (3-13) improve (1-10) on complete Riemannian manifolds.

## 4. Proofs of Theorems 1.9 and 1.10

**Lemma 4.1.** *If $M^n$ is a compact Riemannian manifold and $u$ is a positive solution to (1-8) with $p\ne 0$, then*

$$\frac{d}{dt}\int_{M^n} uv\,d\mu=(p-1)\int_{M^n}(\Delta_\phi v)uv\,d\mu=-p\int_{M^n}|\nabla v|^2 u\,d\mu. \tag{4-1}$$

*Proof.* From (2-1), we have $(uv)_t=vu_t+uv_t=v\Delta_\phi(u^p)+(p-1)uv\Delta_\phi v+u|\nabla v|^2$. It follows from $\nabla(u^p)=u\nabla v$ that

$$\int_{M^n}[v\Delta_\phi(u^p)+u|\nabla v|^2]\,d\mu=\int_{M^n}[-\nabla v\nabla(u^p)+u|\nabla v|^2]\,d\mu=0.$$

Hence

$$\begin{aligned}\frac{d}{dt}\int_{M^n} uv\,d\mu &= \int_{M^n}(uv)_t\,d\mu = \int_{M^n}[v\Delta_\phi(u^p)+(p-1)uv\Delta_\phi v+u|\nabla v|^2]\,d\mu\\ &= (p-1)\int_{M^n}(\Delta_\phi v)uv\,d\mu = p\int_{M^n}(\Delta_\phi v)u^p\,d\mu\\ &= -p\int_{M^n}\nabla v\nabla(u^p)\,d\mu = -p\int_{M^n}|\nabla v|^2 u\,d\mu.\end{aligned}$$ □

**Lemma 4.2.** *If $M^n$ is a compact Riemannian manifold and $u$ is a positive solution to (1-8) with $p\neq 0$, then*

$$\frac{d}{dt}\int_{M^n}(\Delta_\phi v)uv\,d\mu = 2\int_{M^n}\Big[(p-1)(\Delta_\phi v)^2+|\nabla^2 v|^2+\mathrm{Ric}_\phi(\nabla v,\nabla v)\Big]uv\,d\mu.$$

*Proof.* Noticing that

$$\frac{d}{dt}\int_{M^n}(\Delta_\phi v)uv\,d\mu = \int_{M^n}[(\Delta_\phi v)_t uv+(\Delta_\phi v)(uv)_t]\,d\mu, \tag{4-2}$$

a direct calculation gives

$$\begin{aligned}&(\Delta_\phi v)_t\\ &\quad= \Delta_\phi[(p-1)v\Delta_\phi v+|\nabla v|^2]\\ &\quad= (p-1)[(\Delta_\phi v)^2+2\nabla v\nabla\Delta_\phi v+v\Delta_\phi^2 v]+\Delta_\phi|\nabla v|^2\\ &\quad= (p-1)(\Delta_\phi v)^2+2p\nabla v\nabla\Delta_\phi v+(p-1)v\Delta_\phi^2 v+2[|\nabla^2 v|^2+\mathrm{Ric}_\phi(\nabla v,\nabla v)].\end{aligned}$$

We derive from $(p-1)\nabla(uv^2)=(2p-1)uv\nabla v$ that

$$\begin{aligned}&\int_{M^n}[2p\nabla v\nabla\Delta_\phi v+(p-1)v\Delta_\phi^2 v]uv\,d\mu\\ &\qquad= \int_{M^n}2p\nabla v\nabla(\Delta_\phi v)uv\,d\mu-\int_{M^n}(p-1)\nabla(uv^2)\nabla\Delta_\phi v\,d\mu\\ &\qquad= \int_{M^n}\nabla v\nabla(\Delta_\phi v)uv\,d\mu.\end{aligned}$$

Hence

$$\begin{aligned}&\int_{M^n}(\Delta_\phi v)_t uv\,d\mu\\ &\qquad= \int_{M^n}\{(p-1)(\Delta_\phi v)^2+\nabla v\nabla\Delta_\phi v+2[|\nabla^2 v|^2+\mathrm{Ric}_\phi(\nabla v,\nabla v)]\}uv\,d\mu.\end{aligned} \tag{4-3}$$

On the other hand,

$$\begin{aligned}
(4\text{-}4)\quad \int_{M^n} \Delta_\phi v (uv)_t \, d\mu
&= \int_{M^n} \Delta_\phi v [v \Delta_\phi (u^p) + (p-1) uv \Delta_\phi v + u|\nabla v|^2] \, d\mu \\
&= \int_{M^n} [-\nabla (v \Delta_\phi v) \nabla (u^p) + (p-1) uv (\Delta_\phi v)^2 + u|\nabla v|^2 \Delta_\phi v] \, d\mu \\
&= \int_{M^n} [-\nabla (v \Delta_\phi v) u \nabla v + (p-1) uv (\Delta_\phi v)^2 + u|\nabla v|^2 \Delta_\phi v] \, d\mu \\
&= \int_{M^n} [-\nabla v \nabla \Delta_\phi v + (p-1)(\Delta_\phi v)^2] uv \, d\mu .
\end{aligned}$$

Inserting (4-3) and (4-4) into (4-2) concludes the proof of Lemma 4.2 □

*Proof of Theorems 1.9 and 1.10.* By Lemma 4.1, we have

$$\begin{aligned}
\frac{d}{dt} \mathcal{N}_{p,m}(g, u, t) &= -\tilde{a} t^{\tilde{a}-1} \int_{M^n} uv \, d\mu - (p-1) t^{\tilde{a}} \int_{M^n} (\Delta_\phi v) uv \, d\mu \\
&= -t^{\tilde{a}} \int_{M^n} \left( (p-1) \Delta_\phi v + \frac{\tilde{a}}{t} \right) uv \, d\mu .
\end{aligned}$$

We obtain (1-29) and (1-32). On the other hand, from the definition of $\mathcal{W}_{p,m}(g, u, t)$ in (1-28), we have

$$\begin{aligned}
\mathcal{W}_{p,m}(g, u, t) &= \frac{d}{dt} [t \mathcal{N}_{p,m}(g, u, t)] \\
&= \mathcal{N}_{p,m}(g, u, t) + t \frac{d}{dt} \mathcal{N}_{p,m}(g, u, t) \\
&= t^{\tilde{a}+1} \int_{M^n} \left( p \frac{|\nabla v|^2}{v} - \frac{\tilde{a}+1}{t} \right) uv \, d\mu ,
\end{aligned}$$

where Lemma 4.1 was used in the last equality. Hence we derive (1-30) and (1-33).

Notice that the estimate (1-10) also holds for compact Riemannian manifolds. Taking $K = 0$ and then letting $\alpha \to 1$ in (1-10) yields

$$(p-1) \Delta_\phi v + \frac{\tilde{a}}{t} = \frac{v_t}{v} - \frac{|\nabla v|^2}{v} + \frac{\tilde{a}}{t} \geq 0,$$

which allows us to concludes that if $\mathrm{Ric}_\phi^m \geq 0$, then $\mathcal{N}_{p,m}(g, u, t)$ is nonincreasing in $t$. When $p \in (1 - 2/m, 1)$ and $\mathrm{Ric}_\phi^m \geq 0$, we also get from (1-12) that

$$(p-1) \Delta_\phi v + \frac{\tilde{a}}{t} = \frac{v_t}{v} - \frac{|\nabla v|^2}{v} + \frac{\tilde{a}}{t} \leq 0,$$

which shows that $\mathcal{N}_{p,m}(g, u, t)$ is also nonincreasing in $t$.

Now we are in a position to prove (1-31). From (1-29), we have

$$
\begin{aligned}
&\frac{d}{dt}\Big(t\frac{d}{dt}\mathcal{N}_{p,m}(g,u,t)\Big)\\
&\quad=\frac{d}{dt}\Big(-t^{\tilde a+1}\int_{M^n}(p-1)(\Delta_\phi v)uv\,d\mu-\tilde a t^{\tilde a}\int_{M^n}uv\,d\mu\Big)\\
&\quad=\frac{d}{dt}\Big(-t^{\tilde a+1}\int_{M^n}(p-1)(\Delta_\phi v)uv\,d\mu+\tilde a\mathcal{N}_{p,m}(g,u,t)\Big)\\
&\quad=-2t^{\tilde a+1}\int_{M^n}\big((p-1)^2(\Delta_\phi v)^2+(p-1)|\nabla^2 v|^2+(p-1)\operatorname{Ric}_\phi(\nabla v,\nabla v)\big)uv\,d\mu\\
&\qquad\qquad-(\tilde a+1)t^{\tilde a}\int_{M^n}(p-1)(\Delta_\phi v)uv\,d\mu-\tilde a t^{\tilde a}\int_{M^n}\Big((p-1)\Delta_\phi v+\frac{\tilde a}{t}\Big)uv\,d\mu,
\end{aligned}
$$

where the last equality used Lemma 4.2. Hence

$$
\begin{aligned}
&\frac{d}{dt}\mathcal{W}_{p,m}(g,u,t)\\
&\quad=\frac{d}{dt}\Big(t\frac{d}{dt}\mathcal{N}_{p,m}(g,u,t)+\mathcal{N}_{p,m}(g,u,t)\Big)\\
&\quad=-2t^{\tilde a+1}\int_{M^n}\Big[(p-1)^2(\Delta_\phi v)^2+(p-1)|\nabla^2 v|^2+(p-1)\operatorname{Ric}_\phi(\nabla v,\nabla v)\Big]uv\,d\mu\\
&\qquad-(\tilde a+1)t^{\tilde a}\int_{M^n}(p-1)(\Delta_\phi v)uv\,d\mu-(\tilde a+1)t^{\tilde a}\int_{M^n}\Big((p-1)\Delta_\phi v+\frac{\tilde a}{t}\Big)uv\,d\mu\\
&\quad=-2t^{\tilde a+1}\int_{M^n}\Big((p-1)^2(\Delta_\phi v)^2+(p-1)|\nabla^2 v|^2+(p-1)\operatorname{Ric}_\phi(\nabla v,\nabla v)\\
&\qquad\qquad+(p-1)\frac{\tilde a+1}{t}\Delta_\phi v+\frac{\tilde a^2+\tilde a}{2t^2}\Big)uv\,d\mu.
\end{aligned}
\tag{4-5}
$$

Notice that

$$
\begin{aligned}
&(p-1)^2(\Delta_\phi v)^2+(p-1)\frac{\tilde a+1}{t}\Delta_\phi v+\frac{\tilde a^2+\tilde a}{2t^2}\\
&\quad=\Big|(p-1)\Delta_\phi v+\frac{m(p-1)}{[m(p-1)+2]t}\Big|^2+\frac{2(p-1)}{[m(p-1)+2]t}\Delta_\phi v+\frac{(p-1)m}{[m(p-1)+2]^2t^2},
\end{aligned}
$$

and hence

$$
\begin{aligned}
&(p-1)^2(\Delta_\phi v)^2+(p-1)\frac{\tilde a+1}{t}\Delta_\phi v+\frac{\tilde a^2+\tilde a}{2t^2}+(p-1)|\nabla^2 v|^2+\frac{p-1}{m-n}(\nabla\phi\,\nabla v)^2\\
&\quad=\Big|(p-1)\Delta_\phi v+\frac{m(p-1)}{[m(p-1)+2]t}\Big|^2+(p-1)\Big|\nabla^2 v+\frac{g}{[m(p-1)+2]t}\Big|^2\\
&\qquad+\frac{p-1}{m-n}\Big|\nabla\phi\,\nabla v-\frac{m-n}{[m(p-1)+2]t}\Big|^2.
\end{aligned}
\tag{4-6}
$$

We complete the proof of (1-31) by putting (4-6) into (4-5).

When $p \in (0,1)$, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
-(p-1)\left|\nabla^2 v + \frac{g}{[m(p-1)+2]t}\right|^2 &\geq -\frac{p-1}{n}\left|\Delta v + \frac{n}{[m(p-1)+2]t}\right|^2 \\
&= -\frac{1}{n(p-1)}\left|(p-1)\Delta_\phi v + \frac{\tilde{a}}{t}\right|^2 - \frac{p-1}{n}\left|\nabla\phi\nabla v - \frac{m-n}{[m(p-1)+2]t}\right|^2 \\
&\qquad - \frac{2}{n}\left((p-1)\Delta_\phi v + \frac{\tilde{a}}{t}\right)\left(\nabla\phi\nabla v - \frac{m-n}{[m(p-1)+2]t}\right).
\end{aligned}$$

Hence

$$\begin{aligned}
(4\text{-}7)\quad & -(p-1)\left|\nabla^2 v+\frac{g}{[m(p-1)+2]t}\right|^2 - \frac{p-1}{m-n}\left|\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right|^2 - \left|(p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right|^2 \\
&\geq \frac{1-n(1-p)}{n(1-p)}\left|(p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right|^2 + \frac{m(1-p)}{n(m-n)}\left|\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right|^2 \\
&\qquad -\frac{2}{n}\left((p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right)\left(\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right) \\
&\geq \left(\frac{1-n(1-p)}{n(1-p)}-\frac{\varepsilon}{n}\right)\left|(p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right|^2 \\
&\qquad +\left(\frac{m(1-p)}{n(m-n)}-\frac{1}{n\varepsilon}\right)\left|\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right|^2,
\end{aligned}$$

where $\varepsilon \geq m-n$ is a positive constant and satisfies

$$1-1/(n+\varepsilon) \leq p \leq 1-(m-n)/(m\varepsilon).$$

Inserting (4-7) into (1-31) gives

$$\begin{aligned}
&\frac{d}{dt}\mathcal{W}_{p,m}(g,u,t) \\
&\leq 2t^{a+1}\int_{M^n}\left((1-p)\operatorname{Ric}_\phi^m(\nabla v,\nabla v) + \left(\frac{1-n(1-p)}{n(1-p)}-\frac{\varepsilon}{n}\right)\left|(p-1)\Delta_\phi v+\frac{\tilde{a}}{t}\right|^2\right. \\
&\qquad \left. + \left(\frac{m(1-p)}{n(m-n)}-\frac{1}{n\varepsilon}\right)\left|\nabla\phi\nabla v-\frac{m-n}{[m(p-1)+2]t}\right|^2\right)uv\,d\mu.
\end{aligned}$$

This completes the proof of (1-34). □

## Acknowledgements

The authors thank Professor Xiang-Dong Li for his valuable comments and Mr. Zhijie Huang for helpful discussions. The authors also thank the referee for careful reading and many valuable suggestions to improve readability.


## References


[Aronson and Bénilan 1979] D. G. Aronson and P. Bénilan, "Régularité des solutions de l'équation des milieux poreux dans $\mathbb{R}^N$", *C. R. Acad. Sci. Paris* (*A*) **288**:2 (1979), 103–105. MR 82i:35090 Zbl 0397.35034

[Bakry 1994] D. Bakry, "L'hypercontractivité et son utilisation en théorie des semigroupes", pp. 1–114 in *Lectures on probability theory* (Saint-Flour, 1992), edited by D. Bakry et al., Lecture Notes in Math. **1581**, Springer, Berlin, 1994. MR 95m:47075 Zbl 0856.47026

[Bakry and Émery 1985] D. Bakry and M. Émery, "Diffusions hypercontractives", pp. 177–206 in *Séminaire de probabilités, XIX, 1983/84*, edited by J. Azéma and M. Yor, Lecture Notes in Math. **1123**, Springer, Berlin, 1985. MR 88j:60131 Zbl 0561.60080

[Bakry and Qian 1999] D. Bakry and Z. M. Qian, "Harnack inequalities on a manifold with positive or negative Ricci curvature", *Rev. Mat. Iberoamericana* **15**:1 (1999), 143–179. MR 2000f:58052 Zbl 0924.58096

[Bakry and Qian 2005] D. Bakry and Z. M. Qian, "Volume comparison theorems without Jacobi fields", pp. 115–122 in *Current trends in potential theory* (Bucharest, 2002/2003), edited by D. Bakry et al., Theta Ser. Adv. Math. **4**, Theta, Bucharest, 2005. MR 2007e:58048 Zbl 1212.58019

[Calabi 1958] E. Calabi, "An extension of E. Hopf's maximum principle with an application to Riemannian geometry", *Duke Math. J.* **25** (1958), 45–56. MR 19,1056e Zbl 0079.11801

[Cheng and Yau 1975] S. Y. Cheng and S.-T. Yau, "Differential equations on Riemannian manifolds and their geometric applications", *Comm. Pure Appl. Math.* **28**:3 (1975), 333–354. MR 52 #6608 Zbl 0312.53031

[Daskalopoulos and Kenig 2007] P. Daskalopoulos and C. E. Kenig, *Degenerate diffusions: initial value problems and local regularity theory*, EMS Tracts in Mathematics **1**, European Mathematical Society, Zürich, 2007. MR 2009b:35214 Zbl 1205.35002

[Davies 1989] E. B. Davies, *Heat kernels and spectral theory*, Cambridge Tracts in Mathematics **92**, Cambridge University Press, 1989. MR 90e:35123 Zbl 0699.35006

[Ecker 2007] K. Ecker, "A formula relating entropy monotonicity to Harnack inequalities", *Comm. Anal. Geom.* **15**:5 (2007), 1025–1061. MR 2009c:53088 Zbl 1167.53055

[Fang et al. 2009] F. Fang, X.-D. Li, and Z. Zhang, "Two generalizations of Cheeger–Gromoll splitting theorem via Bakry–Emery Ricci curvature", *Ann. Inst. Fourier* (*Grenoble*) **59**:2 (2009), 563–573. MR 2010b:53057 Zbl 1166.53023

[Hamilton 1993] R. S. Hamilton, "A matrix Harnack estimate for the heat equation", *Comm. Anal. Geom.* **1**:1 (1993), 113–126. MR 94g:58215 Zbl 0799.53048

[Huang et al. 2013] G. Huang, Z. Huang, and H. Li, "Gradient estimates for the porous medium equations on Riemannian manifolds", *J. Geom. Anal.* **23**:4 (2013), 1851–1875. MR 3107682 arXiv 1106.2373

[Kotschwar and Ni 2009] B. Kotschwar and L. Ni, "Local gradient estimates of $p$-harmonic functions, $1/H$-flow, and an entropy formula", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:1 (2009), 1–36. MR 2010g:53121 Zbl 1182.53060

[Ledoux 2000] M. Ledoux, "The geometry of Markov diffusion generators", *Ann. Fac. Sci. Toulouse Math.* (6) **9**:2 (2000), 305–366. MR 2002a:58045 Zbl 0980.60097

[Li 2005] X.-D. Li, "Liouville theorems for symmetric diffusion operators on complete Riemannian manifolds", *J. Math. Pures Appl.* (9) **84**:10 (2005), 1295–1361. MR 2006f:58046 Zbl 1082.58036

[Li 2011] X.-D. Li, "Perelman's $W$-entropy for the Fokker–Planck equation over complete Riemannian manifolds", *Bull. Sci. Math.* **135**:6-7 (2011), 871–882. MR 2012m:53077 Zbl 1230.82039

[Li 2012] X.-D. Li, "Perelman's entropy formula for the Witten Laplacian on Riemannian manifolds via Bakry–Emery Ricci curvature", *Math. Ann.* **353**:2 (2012), 403–437. MR 2915542 Zbl 06043348

[Li 2013] X.-D. Li, "Hamilton's Harnack inequality and the $W$-entropy formula on complete Riemannian manifolds", preprint, 2013. arXiv 1303.1242

[Li and Li 2013] S. Li and X.-D. Li, "Perelman's entropy formula for the Witten Laplacian on manifolds with time dependent metrics and potentials", preprint, 2013. arXiv 1303.6019

[Li and Xu 2011] J. Li and X. Xu, "Differential Harnack inequalities on Riemannian manifolds, I: Linear heat equation", *Adv. Math.* **226**:5 (2011), 4456–4491. MR 2012h:53079 Zbl 1226.58009

[Li and Yau 1986] P. Li and S.-T. Yau, "On the parabolic kernel of the Schrödinger operator", *Acta Math.* **156**:3-4 (1986), 153–201. MR 87f:58156 Zbl 0611.58045

[Lu et al. 2009] P. Lu, L. Ni, J. L. Vázquez, and C. Villani, "Local Aronson–Bénilan estimates and entropy formulae for porous medium and fast diffusion equations on manifolds", *J. Math. Pures Appl.* (9) **91**:1 (2009), 1–19. MR 2010j:35563 Zbl 1156.58015

[Ni 2002] L. Ni, "The Poisson equation and Hermitian–Einstein metrics on holomorphic vector bundles over complete noncompact Kähler manifolds", *Indiana Univ. Math. J.* **51**:3 (2002), 679–704. MR 2003m:32021 Zbl 1035.53032

[Ni 2004a] L. Ni, "The entropy formula for linear heat equation", *J. Geom. Anal.* **14**:1 (2004), 87–100. MR 2004m:53118a Zbl 1044.58030

[Ni 2004b] L. Ni, "Addenda to 'The entropy formula for linear heat equation'", *J. Geom. Anal.* **14**:2 (2004), 369–374. MR 2004m:53118b Zbl 1062.58028

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv math/0211159

[Qian 1997] Z. M. Qian, "Estimates for weighted volumes and applications", *Quart. J. Math. Oxford Ser.* (2) **48**:190 (1997), 235–242. MR 98e:53058 Zbl 0902.53032

[Qian 1998] Z. M. Qian, "A comparison theorem for an elliptic operator", *Potential Anal.* **8**:2 (1998), 137–142. MR 99d:58161 Zbl 0930.58012

[Vázquez 2006] J. L. Vázquez, *Smoothing and decay estimates for nonlinear diffusion equations: equations of porous medium type*, Oxford Lecture Series in Mathematics and its Applications **33**, Oxford University Press, 2006. MR 2007k:35008 Zbl 1113.35004

[Vázquez 2007] J. L. Vázquez, *The porous medium equation: mathematical theory*, Clarendon, Oxford, 2007. MR 2008e:35003 Zbl 1107.35003

[Wang 1997] F.-Y. Wang, "Logarithmic Sobolev inequalities on noncompact Riemannian manifolds", *Probab. Theory Related Fields* **109**:3 (1997), 417–424. MR 98i:58253 Zbl 0887.35012

[Wang 2004] F.-Y. Wang, "Equivalence of dimension-free Harnack inequality and curvature condition", *Integral Equations Operator Theory* **48**:4 (2004), 547–552. MR 2004m:58061 Zbl 1074.47020

[Wang and Chen 2013] Y. Wang and W. Chen, "Gradient estimates for weighted diffusion equations on smooth metric measure spaces", *J. Math.* (*Wuhan*) **33**:2 (2013), 248–258. MR 3076231

[Wang et al. 2013] Y. Wang, J. Yang, and W. Chen, "Gradient estimates and entropy formulae for weighted $p$-heat equations on smooth metric measure spaces", *Acta Math. Sci. Ser. B Engl. Ed.* **33**:4 (2013), 963–974. MR 3072133

[Wei and Wylie 2009] G. Wei and W. Wylie, "Comparison geometry for the Bakry–Emery Ricci tensor", *J. Differential Geom.* **83**:2 (2009), 377–405. MR 2011a:53064 Zbl 1189.53036

[Xu 2012] X. Xu, "Gradient estimates for $u_t = \Delta F(u)$ on manifolds and some Liouville-type theorems", *J. Differential Equations* **252**:2 (2012), 1403–1420. MR 2853544 Zbl 1235.53045

[Yau 1994] S.-T. Yau, "On the Harnack inequalities of partial differential equations", *Comm. Anal. Geom.* **2**:3 (1994), 431–450. MR 96f:58186 Zbl 0841.58059

[Yau 1995] S.-T. Yau, "Harnack inequality for non-self-adjoint evolution equations", *Math. Res. Lett.* **2**:4 (1995), 387–399. MR 96k:58211 Zbl 0884.58091

Received December 10, 2012. Revised May 13, 2013.

GUANGYUE HUANG
COLLEGE OF MATHEMATICS AND INFORMATION SCIENCE
HENAN NORMAL UNIVERSITY
XINXIANG 453007
CHINA
hgy@henannu.edu.cn

HAIZHONG LI
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING 100084
CHINA
hli@math.tsinghua.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.79

# CONTROLLED CONNECTIVITY FOR SEMIDIRECT PRODUCTS ACTING ON LOCALLY FINITE TREES

KEITH JONES

**In 2003 Bieri and Geoghegan generalized the Bieri–Neumann–Strebel invariant $\Sigma^1$ by defining $\Sigma^1(\rho)$, $\rho$ an isometric action by a finitely generated group $G$ on a proper CAT(0) space $M$. In this paper, we show how the natural and well-known connection between Bass–Serre theory and covering space theory provides a framework for the calculation of $\Sigma^1(\rho)$ when $\rho$ is a cocompact action by $G = B \rtimes A$, $A$ a finitely generated group, on a locally finite Bass–Serre tree $T$ for $A$. This framework leads to a theorem providing conditions for including an endpoint in, or excluding an endpoint from, $\Sigma^1(\rho)$. When $A$ is a finitely generated free group acting on its Cayley graph, we can restate this theorem from a more algebraic perspective, which leads to some general results on $\Sigma^1$ for such actions.**

## 1. Introduction

In [Bieri and Geoghegan 2003b], the authors begin with the following:

> Given a group $G$ and a contractible metric space $M$, consider the set $\mathrm{Hom}(G, \mathrm{Isom}(M))$ of all actions by $G$ on $M$ by isometries. Are there invariants of such actions which distinguish one from another? Are there topological properties which one such action might possess while another might not?

The tool they apply to draw distinctions between such actions is controlled $n$-connectivity, which is developed in [Bieri and Geoghegan 2003a], and which we briefly describe here. Suppose $\rho$ is an isometric action by a group $G$ having type $F_n$ on a proper CAT(0) metric space $M$. Fixing a basepoint $b \in M$, the CAT(0) boundary, $\partial M$, can be thought of as the set of geodesic rays $\tau$ emanating from $b$.[1] For an end point $e \in \partial M$ represented by a ray $\tau$, there is a nested family of subsets

*MSC2010:* primary 20E08, 20F65, 57M07; secondary 05C05, 05C25.
*Keywords:* controlled connectivity, BNS, sigma invariants, tree actions, semidirect products.

[1]For background on the topological finiteness property "type $F_n$", see [Geoghegan 2008, §7.2], and for background on CAT(0) metric spaces and their boundaries see [Bridson and Haefliger 1999, II.1 and II.8]. A metric space is *proper* if each closed metric ball is compact.

$\mathrm{HB}_k(\tau)$, $k \in \mathbb{R}$, called "horoballs" which serve as metric balls (in $M$) "at $e$."[2] This provides a sense of direction in $M$, which can be "lifted to $G$" by $\rho$ via a $G$-equivariant map from the $n$-skeleton of the universal cover of a $K(G,1)$. For a point $e \in \partial M$, if the lifts of the horoballs about $e$ are (roughly) $(n-1)$-connected, then we say $\rho$ is controlled $(n-1)$-connected over $e$. Bieri and Geoghegan show that this is independent of choice of $K(G,1)$ or equivariant map. The invariant $\Sigma^n(\rho)$ is the subset of $\partial M$ consisting of points over which $\rho$ is controlled $(n-1)$-connected. This definition generalizes the Bieri–Neumann–Strebel–Renz (BNSR) invariants $\Sigma^n(G)$, which are open subsets of the CAT(0) boundary of the vector space $G_{ab} \otimes \mathbb{R}$. A key difference between $\Sigma^1(\rho)$ and the BNSR invariant $\Sigma^1(G)$ is that $\Sigma^1(\rho)$ is in general *not* an open subset of $\partial M$.

Apart from enabling one to draw geometric distinctions between isometric actions by a group on a proper CAT(0) space, the invariant can also provide group theoretical information: if the orbits under an action $\rho$ are discrete, then the point stabilizers are finitely generated if and only if $\Sigma^1(\rho) = \partial M$ [Bieri and Geoghegan 2003a, Theorem A and Boundary Criterion].

When $M = T$ is a locally finite (simplicial) tree, the CAT(0) boundary is a metric Cantor set. Initial results in [Bieri and Geoghegan 2003a] led the authors to ask whether in this case $\Sigma^1(\rho)$ might always be one of $\varnothing$, a singleton, or the entire boundary $\partial T$. Work in [Jones 2012] establishes a class of actions for which this is the case. However, work by Ralf Lehnert in his diploma thesis demonstrates that other subsets of $\partial T$ can be realized as $\Sigma^1(\rho)$ for certain actions [Lehnert 2009]. This hints at a potentially rich world of $\Sigma^1$ invariants, which we further explore here.

**1.1.** ***Statement of results.*** We restrict our attention to $\Sigma^1$, and study only the following scenario:

**Definition 1** (actions of interest). Let $A$ be a finitely generated group with finite generating set $R$, and let $T$ be a locally finite[3] simplicial tree on which $A$ acts cocompactly and with finitely generated stabilizers. For a group $B$, suppose we have a homomorphism $\varphi : A \to \mathrm{Aut}(B)$, and let $G = B \rtimes_\varphi A$ be the resulting semidirect product. Elements of $G$ are of the form $(b, a)$, where $a \in A,\ b \in B$, and multiplication in $G$ operates under the rule

$$(b_1, a_1)(b_2, a_2) = (b_1 a_1 b_2 a_1^{-1}, a_1 a_2) = (b_1 \varphi_{a_1}(b_2), a_1 a_2).$$

[2]For background on horoballs, see [Bieri and Geoghegan 2003a, §10.1]. The convention followed there and in this paper is that as $k$ increases, we approach $e$, the reverse of the convention in [Bridson and Haefliger 1999].

[3]A simplicial tree is a proper metric space if and only if it is locally finite.

Suppose $G$ is finitely generated. Then it follows that $B$ is finitely generated as an $A$-group. By this, we mean there is a finite subset $S \subset B$ such that the set $\{\varphi_a(S) \mid a \in A\}$ generates $B$, and so $G$ is generated by $S \cup R$.

The natural projection $G \twoheadrightarrow A$ induces an action $\rho$ by $G$ on $T$ which contains the normal subgroup $B$ in its kernel. We investigate $\Sigma^1(\rho)$.

**Remark 2.** As mentioned, if the point stabilizers under $\rho$ are finitely generated, then $\Sigma^1(\rho) = \partial T$ [Bieri and Geoghegan 2003a, Theorem A and Boundary Criterion]. Moreover, since $T$ is locally finite, all point stabilizes are commensurable, so if any one is finitely generated, then all are. Thus with the assumption that the stabilizers under the $A$ action on $T$ are finitely generated, in order to obtain "interesting" invariants (those with $\Sigma^1(\rho) \neq \partial T$), one must assume that $B$ is not finitely generated, since the stabilizers under $\rho$ are simply semidirect products of $B$ with the stabilizers in $A$.

*Main result.* With the action $\rho : G \to \mathrm{Isom}(T)$ as defined above, we apply the relationship between Bass–Serre theory and covering space theory to construct a commutative diagram of $G$-equivariant cellular maps between CW-complexes:

$$\begin{tikzcd}
\tilde{X} \arrow[r, "\tilde{r}"] \arrow[d, "p"] & T \arrow[d, "\mathrm{id}"] \\
\bar{X} \arrow[r, "\bar{r}"] \arrow[d, "q"] & T \arrow[d, "\mathrm{mod}\, G"] \\
X \arrow[r, "r"] & V = T \backslash G
\end{tikzcd}$$

where $X$ is a $K(G, 1)$, $\bar{X}$ is a $K(B, 1)$, $\tilde{X}$ is a contractible universal cover, $p$ and $q$ are covering projections, and $r$, $\bar{r}$, and $\tilde{r}$ are retracts.[4] For a geodesic ray $\tau$ in $T$ and $k \in \mathbb{Z}$, consider the horoball $\mathrm{HB}_k(\tau)$.[5] For $W \subset X$ a finite subcomplex, set

$$\bar{X}_{(\tau,k,W)} = \bar{r}^{-1}(\mathrm{HB}_k(\tau)) \cap q^{-1}(W) \subset \bar{X}.$$

**Theorem 3.** *Let $e \in \partial T$ be represented by a geodesic ray $\tau$.*

(i) *If there exists a finite subcomplex $W \subset X$ such that for every $k \in \mathbb{Z}$, $\bar{X}_{(\tau,k,W)}$ is connected and the map on $\pi_1$ induced by the inclusion $\bar{X}_{(\tau,k,W)} \hookrightarrow \bar{X}$ is surjective, then $e \in \Sigma^1(\rho)$.*

(ii) *If for every $k \in \mathbb{Z}$ and every finite subcomplex $W \subset X$ such that $\bar{X}_{(\tau,k,W)}$ is connected, the induced map on $\pi_1$ is not surjective, then $e \notin \Sigma^1(\rho)$.*

[4]This is the topological construction of the Bass–Serre tree [Geoghegan 2008, §6.2; Scott and Wall 1979], discussed further in Section 2.2.

[5]A precise description of $\mathrm{HB}_k(\tau)$ is given in Equation (1-1).

*Consequences and examples.* Theorem 3 has a number of consequences in the case where $A$ is a free group and $T$ is its Cayley graph. In this case, the vertices of $T$ are the elements of $A$. Let $e$ be an endpoint of $T$ and suppose the geodesic ray $\tau$ represents $e$. For an integer $k$, let $A_k(\tau)$ be the elements of $A$ that form the vertex set of the horoball $\mathrm{HB}_k(\tau)$. To avoid confusion with the group $B$, we will use the notation $\mathrm{Ball}_r(X, p)$ to refer to the metric ball of radius $r$ in the space $X$ about the point $p$. Just as the horoball $\mathrm{HB}_k(\tau)$ can be written as the nested union of closed metric balls in $T$:

$$\mathrm{HB}_k(\tau) = \bigcup_{l \geq \max\{0,k\}} \overline{\mathrm{Ball}_{l-k}(T, \tau(l))}, \tag{1-1}$$

the set $A_k(\tau)$ can be written as a nested union of closed metric balls in the word metric on $A$:

$$A_k(\tau) = \bigcup_{l \geq \max\{0,k\}} \overline{\mathrm{Ball}_{l-k}(A, \tau(l))}. \tag{1-2}$$

We will say $B$ is finitely generated over a *subset* $A' \subseteq A$ if there is a finite subset $S \subseteq B$ such that $\{asa^{-1} \mid s \in S, a \in A'\}$ generates $B$.

In Section 4, we show that in this context Theorem 3 can be restated as follows:

**Theorem 4.** *Let $A$ be a finitely generated free group, and let $T$ be its Cayley graph with respect to a free basis. For the action $\rho$ as in Theorem 3, and for $e \in \partial T$ represented by geodesic ray $\tau$:*

(i) *If there is a finite set $S \subseteq B$ such that for each $k \in \mathbb{Z}_{\geq 0}$, $S$ generates $B$ over $A_k(\tau)$, then $e \in \Sigma^1(\rho)$.*

(ii) *If for each $k \in \mathbb{Z}_{\leq 0}$, $B$ is not finitely generated over $A_k(\tau)$, then $e \notin \Sigma^1(\rho)$.*

This is reminiscent of the invariant $\Sigma_B(A)$ of [Bieri et al. 1987] and [Bieri and Strebel 1980], but whereas $\Sigma_B(A)$ is determined by the algebraic structure of $G$, our sets $A_k(\tau)$ are given by the geometry of $T$; in particular, they are not monoids.

Since $B$ is finitely generated over $A$, we have:

**Corollary 5.** *If for each $k \in \mathbb{Z}_{\geq 0}$, $\varphi(A_k(\tau)) = \varphi(A)$, then $e \in \Sigma^1(\rho)$.*

Let $\{a_1, \ldots, a_n\}$ freely generate $A$. For a generator $a_i$, let the function $\mathrm{expsum}_{a_i}$ map a reduced word $w$ in $\{a_1, \ldots, a_n\}^{\pm}$ to the corresponding exponent sum of $a_i$ in $w$, and define the function $\mathrm{expsum}_{a_i^{-1}}$ to be $-\mathrm{expsum}_{a_i}$.

**Corollary 6.** *Let $t \in \{a_1, \ldots, a_n\}^{\pm}$. Suppose there does not exist $m \in \mathbb{Z}$ such that $B$ is finitely generated over $A - \mathrm{expsum}_t^{-1}([m, \infty))$, i.e any subset $A' \subseteq A$ must have reduced words with arbitrarily large exponent sum of $t$ in order for $B$ to be finitely generated over $A'$. Then any endpoint represented by a word eventually consisting of only $t^{-1}$ does not lie in $\Sigma^1(\rho)$.*

**Example 7.** This is a generalization of an example calculated by Ralf Lehnert, although the methods used here are different from his. Consider the semidirect product $G = B \rtimes A$, where $B = \mathbb{Z}[1/(p_1 p_2 \dots p_n)]$, where the $p_i$ are prime with $p_i \neq p_j$ for $1 \leq i, j \leq n$, and $A$ is free on $\{a_1, \dots, a_n\}$. The action is given by $a_i$ acting by multiplication by $1/p_i$. For $A' \subseteq A$, $B$ is finitely generated over $A'$ if and only if $A'$ contains reduced words with arbitrarily large exponent sum of each $a_i$. One can show that for any $k \in \mathbb{Z}$, this will always be the case for $A' = A_k(\tau)$ unless $\tau$ eventually consists of only $a_i^{-1}$ (see Lemma 24). Thus, by Corollary 6, any endpoint corresponding to an infinite word eventually consisting of $a_i^{-1}$ for some $i$ is not in $\Sigma^1$. By Theorem 4, any other endpoint is in $\Sigma^1$.

**Example 8.** Let $G = \mathbb{Z} \wr \mathbb{Z} = \oplus_{i \in \mathbb{Z}} \langle b_i \rangle \rtimes \langle t \rangle$. The action is by shifting: ${}^t b_i = b_{i+1}$. Let $T$ be the Cayley graph of $\langle t \rangle$, a simplicial line. The action $\langle t \rangle \curvearrowright T$ induces an action $G \overset{\rho}{\curvearrowright} T$. It is known from previous work that $\Sigma^1(\rho)$ is empty, as follows. Because the endpoints of the action are fixed, we can relate $\partial T$ to homomorphisms $G \twoheadrightarrow \mathbb{Z}$, and an end point lies in $\Sigma^1(\rho)$ if and only if the corresponding homomorphism represents a point of the BNSR invariant $\Sigma^1(G)$ [Bieri and Geoghegan 2003a, §10.6]. These homomorphisms do not represent points of $\Sigma^1(G)$ because they are not homomorphisms associated to HNN extension decompositions of $G$ over finitely generated base groups [Brown 1987, Proposition 3.1]. Here it follows from Theorem 4, because $B$ is not finitely generated over any proper subset of $\langle t \rangle$.

Corollary 5 can be applied to determine a nice criterion for finding endpoints of $T$ lying in $\Sigma^1(\rho)$.

**Theorem 9.** *With notation as in Theorem 4, viewing endpoints of $T$ as infinite words in the generators of $A$, $\Sigma^1(\rho)$ contains any endpoint represented by an infinite word containing infinitely many mutually distinct subwords lying in* $\ker \varphi$.

**Corollary 10.** *If $\varphi(A) \leq Aut(B)$ is abelian and $A$ has rank $n \geq 2$, then $\Sigma^1(\rho)$ is nonempty.*

For example, any endpoint represented by an infinite word containing infinitely many commutators will be contained in $\Sigma^1(\rho)$.

**Example 11.** Let $m$ and $n$ be positive integers with $m \geq n$. Let $C = \langle a_1, \dots, a_n \rangle$ and $D = \langle a_{n+1}, \dots, a_m \rangle$ be free groups, and set $A = C * D$. For a finitely generated group $K$, let $G$ be the restricted wreath product $K \operatorname{wr}_C A$, where the $A$-action on the indexing set $C$ is defined by the composition of the natural projection $\pi : A \to C$ and left multiplication. In other words, $G = B \rtimes_\varphi A$, where $B = \oplus_{\omega \in C} K_\omega$ with each $K_\omega$ a copy of $K$. The elements of $B$ are sequences $(x_\omega)$, $x_\omega \in K_\omega$, $\omega \in C$, with only finitely many $x_\omega$ nontrivial, and $C$ acts on $B$ by permuting the indices (by left multiplication on itself) while $D \leq \ker \varphi$. The projection $G \twoheadrightarrow A$ followed by the natural action by $A$ on its Cayley graph $T = \Gamma(A, \{a_1, a_2, \dots, a_m\})$ induces an action $\rho$ on $T$.

By Theorem 9, any endpoint containing infinitely many letters $a_i^{\pm}$, $n < i \le m$ will lie in $\Sigma^1(\rho)$, while Corollary 6 ensures that any endpoint eventually consisting of a single letter $a_j^{\pm}$, $1 \le j \le n$, will not lie in $\Sigma^1(\rho)$. In fact, any end point represented by a geodesic ray that eventually consists of only letters from $C$ lies outside $\Sigma^1(\rho)$, as is argued in Section 4.2. So an endpoint lies in $\Sigma^1(\rho)$ if and only if a representative geodesic ray contains infinitely many letters from $D$.

**Example 12.** One can also perform calculations in the case where $A$ is not free. For example, let $H$ be any finitely generated group and consider the group

$$G = B \rtimes_{\varphi} A, \quad \text{where } B = \prod_{i\in\mathbb{Z}} H \text{ and } A = \langle a \mid a^4\rangle * \langle b \mid b^4\rangle, \tag{1-3}$$

where $\varphi : A \to \mathrm{Aut}(B)$ consists first of the projection onto $D_\infty$ collapsing $a^2$ and $b^2$ to the identity, followed by permutation of the indices $i \in \mathbb{Z}$ given by the natural action by $D_\infty$ on $\mathbb{Z}$. Let $\rho$ be the action by $G$ on the regular 4-valent Bass–Serre tree $T_4$ corresponding to the free product structure of $A$. Notice, since this is the Bass–Serre tree corresponding to a free product, any point $e \in \partial T_4$ corresponds to a word in the normal form for the free product. One can apply Theorem 3 to calculate $\Sigma^1(\rho)$ directly to determine that a given $e \in \partial T_4$ if and only if it corresponds to an infinite normal form word containing infinitely many subwords of the form $a^2$ or $b^2$.

There is a stark similarity between this result and Theorem 9, and indeed a statement similar to Theorem 9 can be made in the case where $A$ is a free product. However, only when $A$ is a free product of *finite* groups will its corresponding Bass–Serre tree be locally finite (and hence proper); in this case the Kurosh subgroup theorem implies that $A$ has a free subgroup $A'$ of finite index. If $G = B \rtimes A$, then $G' = B \rtimes A'$ is a finite index subgroup of $G$, and the action $\rho$ by $G$ on the Bass–Serre tree corresponding to the free product decomposition of $A$ restricts to an action by $G'$ on the same tree. It follows from Theorem 12.1 of [Bieri and Geoghegan 2003a] that the invariant is the same for both actions. Hence, it is not clear that such an endeavor will add anything new to the discussion.

**1.2.** ***Defining*** $\boldsymbol{\Sigma^1}$**.** In general, there is a family of invariants $\Sigma^n$, $n \ge 0$, corresponding to the notion of controlled $(n-1)$-connectivity. The discussion below refers only to $\Sigma^1$ and controlled connectivity, but a similar discussion can be had in full generality.

We start with Bieri and Geoghegan's original definition of controlled connectivity.

**Definition 13** [Bieri and Geoghegan 2003a]. Let $\rho$ be an action by a finitely generated group $G$ on a proper CAT(0) metric space $(M, d)$. Choose a $K(G, 1)$

complex $X$ whose universal cover $\tilde{X}$ has a cocompact 1-skeleton $(\tilde{X})^{(1)}$, and a continuous $G$-map $h : (\tilde{X})^{(1)} \to M$. Given a geodesic ray $\tau$ in $M$, $\tau(\infty)$ denotes the point of $\partial M$ represented by $\tau$. For $t \in \mathbb{R}$, let $\tilde{X}_{(\tau,t)}$ denote the largest subcomplex contained in $h^{-1}(\mathrm{HB}_t(\tau))$. Then $h$ is *controlled connected over* $\tau(\infty)$ if there exists $\lambda : \mathbb{R} \to [0, \infty)$ such that for all $t \in \mathbb{R}$, any two points of $\tilde{X}_{(\tau,t)}$ can be connected by a path in $\tilde{X}_{(\tau,t-\lambda(t))}$, and $t - \lambda(t) \to \infty$ as $t \to \infty$.

The same authors also gave an "extended" definition, which we will show coincides with Definition 13 when $G$ is finitely generated.

**Definition 14** [Bieri and Geoghegan 2003b, p. 143]. Let $\rho$ be an action by a (not necessarily finitely generated!) group $G$ on a proper CAT(0) metric space $(M, d)$. Choose a nonempty free contractible $G$-CW-complex $\tilde{X}$ and a continuous $G$-map $h : \tilde{X} \to M$. Fix a geodesic ray $\tau$ in $M$. For $t \in \mathbb{R}$, define $\tilde{X}_{(\tau,t)}$ to be the largest subcomplex of $h^{-1}(\mathrm{HB}_t(\tau))$. Then $h$ is *controlled connected over* $\tau(\infty)$ if for every cocompact $G$-subspace $\widetilde{W} \subseteq \tilde{X}$, there exists a cocompact $G$-subspace $\widetilde{W}'$ containing $\widetilde{W}$ such that for all $t \in \mathbb{R}$, there exists $\lambda(t) \geq 0$ satisfying:

($*$) Any two points of $\tilde{X}_{(\tau,t)} \cap \widetilde{W}$ can be connected by a path through $\tilde{X}_{(\tau,t-\lambda(t))} \cap \widetilde{W}'$.

($**$) Any two points of $\tilde{X}_{(\tau,t+\lambda(t))} \cap \widetilde{W}$ can be connected by a path through $\tilde{X}_{(\tau,t)} \cap \widetilde{W}'$.

Both Definitions 13 and 14 are independent of choice of $G$-space $\tilde{X}$ or $G$-map $h : \tilde{X} \to M$, as is proved in [Bieri and Geoghegan 2003a; 2003b], respectively, in what the authors commonly refer to as the *invariance theorem*. For Definition 14, this is proved for the related concept of controlled connectivity over $a \in M$ [Bieri and Geoghegan 2003b, Theorem 2.3]; the proof carries over to controlled connectivity over an end point [ibid., p. 143].

The parameter $\lambda(t)$ is called a *lag*. In nice cases, $\lambda$ may be constant, or even 0. A lag is necessary for invariance, but an arbitrarily generous lag would defeat the point. In Definition 14, condition ($**$) effectively replaces the condition that $t - \lambda(t) \to \infty$ found in Definition 13.

Suppose now that $G$ is finitely generated and $h : \tilde{X} \to M$ satisfies Definition 14, but $\tilde{X}$ has noncocompact 1-skeleton. There is $h' : \tilde{X}' \to M$, where $\tilde{X}'$ has cocompact 1-skeleton, which by the invariance theorem also satisfies Definition 14. We now show that Definition 13 is satisfied by $h'|_{(\tilde{X})^{(1)}}$.

**Proposition 15.** *Let $G$ be a finitely generated group, $\tilde{X}$ a contractible free $G$-complex with cocompact 1-skeleton $(\tilde{X})^{(1)}$, and geodesic ray $\tau$ in a proper CAT(0) space $M$. A $G$-map $h : \tilde{X} \to M$ satisfies Definition 14 if and only if the restriction $h| : (\tilde{X})^{(1)} \to M$ satisfies Definition 13.*

*Proof.* If $h|$ satisfies Definition 13 over $\tau(\infty)$, then there is a lag $\lambda(t)$ satisfying $t - \lambda(t) \to \infty$ as $t \to \infty$ such that for each $t$, any two points in $(\tilde{X})^{(1)}_{(\tau,t)}$ may be joined in $(\tilde{X})^{(1)}_{(\tau,t-\lambda(t))}$. Let $\widetilde{W}$ be any cocompact $G$-subset of $\tilde{X}$. Let $Y$ be the smallest subcomplex of $\tilde{X}$ containing $\widetilde{W}$. Then $Y$ is still a cocompact $G$-set. Take $\widetilde{W}' = Y \cup (\tilde{X})^{(1)}$. Then any two points of $\tilde{X}_{(\tau,t)} \cap \widetilde{W}$ may be joined in $\tilde{X}_{(\tau,t-\lambda(t))} \cap \widetilde{W}'$ by first moving into the 1-skeleton of $\tilde{X}_{(\tau,t)} \cap Y$. We now replace $\lambda(t)$ with a lag function $\lambda'(t)$ satisfying both $(*)$ and $(**)$. For any $t$, there exists $r > t$ such that for all $s \geq r$, $s - \lambda(s) > t$. (So points of $(\tilde{X})^{(1)}_{[\tau,s)}$ can be connected through a path in $(\tilde{X})^{(1)}_{[\tau,t)}$.) Let $\lambda'(t) = \max\{\lambda(t), r - t\}$.

Now suppose $h$ satisfies Definition 14 over $\tau(\infty)$. For $\widetilde{W} = \widetilde{W}' = (\tilde{X})^{(1)}$, there is $\lambda : \mathbb{R} \to [0, \infty)$ such that by $(*)$, any two points of $\tilde{X}_{(\tau,t)} \cap (\tilde{X})^{(1)}$ may be joined through a path in $\tilde{X}_{(\tau,t-\lambda(t))} \cap (\tilde{X})^{(1)}$, since a path may be chosen which does not leave $(\tilde{X})^{(1)}$. We now find a lag $\lambda'(t)$ satisfying $t - \lambda'(t) \to \infty$. Since $\mathrm{HB}_s(\tau) \subseteq \mathrm{HB}_r(\tau)$ when $s > r$, $(**)$ says that for all $r \in \mathbb{R}$, for all $t > r + \lambda(r)$, a lag of $(t - r)$ suffices for $\mathrm{HB}_t(\tau)$. Hence, we may choose a real-valued sequence $s_1 < s_2 < \cdots$ satisfying $s_n \to \infty$ and for $t \in [s_n, s_{n+1})$ a lag of $t - n$ suffices. Define $\lambda'(t)$ by:

$$\lambda'(t) = \begin{cases} \lambda(t) & \text{if } t < s_1, \\ t - n & \text{if } s_n \leq t < s_{n+1}, \ n = 1, 2, \ldots \end{cases}$$

Then $t - \lambda'(t) = n$ when $s_n \leq t < s_{n+1}$, so $t - \lambda'(t) \to \infty$ as $t \to \infty$. □

This means that one may test for controlled connectivity of a finitely generated group in the traditional sense by applying the more general definition with a space $\tilde{X}$, even when $(\tilde{X})^{(1)}$ is not cocompact.

**Definition 16** ($\Sigma^1$). the invariance theorem ensures controlled connectivity is a property of the action $\rho$, so we define

$$\Sigma^1(\rho) = \{e \in \partial M \mid \rho \text{ is controlled connected over } e\}.$$

The action $\rho$ induces an action on $\partial M$, and under this action $\Sigma^1(\rho)$ is a $G$-invariant set.

## 2. Covering spaces and Bass–Serre theory

**2.1.** ***Some facts about covering spaces.*** The following proposition counts the number of components over a connected subset in a covering projection.

**Proposition 17** [Geoghegan 2008, Theorem 3.4.10]. *Let $(X, Z)$ be a pair of path connected CW complexes, both containing a point $z$. Let $i : (Z, z) \to (X, z)$ be the inclusion map, and let $p : (\bar{X}, \bar{z}) \to (X, z)$ be a covering projection. Let $H_1 = \operatorname{im} p_\#$ and $H_2 = \operatorname{im} i_\#$. Then the number of path components of $p^{-1}(Z)$ equals the order of the set of double cosets*

$$\{H_1 g H_2 \mid g \in \pi_1(X, z)\}.$$

*In particular, if $\bar{X} = \tilde{X}$ is the universal cover of $X$, then the number of components of $p^{-1}(Z)$ is the index of $H_2$ in $\pi_1(X, z)$.*

For us the interesting case for us will be when $Z$ has connected preimage in $\tilde{X}$. With this in mind, we will say $Z$ is $\pi_1$-*surjective* when the inclusion $(Z, z) \hookrightarrow (X, z)$ induces a surjection on $\pi_1$.

A second fact we will need is a consequence of path lifting:

**Proposition 18.** *Let $(X, Z)$ be a pair of path connected CW complexes. Let $p : \bar{X} \to X$ be a covering projection. Then each component of $p^{-1}(Z)$ surjects onto $Z$.*

**2.2.** ***Bass–Serre theory via covering spaces.*** We are concerned with cocompact actions by finitely generated groups on locally finite simplicial trees, particularly those without global fixed points. Thus all actions we consider can be understood though Bass–Serre theory [Bass 1993; Serre 1980]. There is a beautiful connection between Bass–Serre theory and covering space theory [Geoghegan 2008, §6.2; Scott and Wall 1979], which we take advantage of in order to calculate $\Sigma^1$ for actions as described by Definition 1. Here we briefly recount this topological construction of the Bass–Serre tree in the context of such actions, and in the process introduce an intermediary covering space which will be important for calculations.

Given an action $\rho$ as in Definition 1, set $V = G\backslash T$, a finite graph since $\rho$ is cocompact. Fix a base vertex $v_0$ of $V$. Choose a connected fundamental domain $F$ for $\rho$, and let $\mathcal{V}$ be the system of stabilizers for $F$. (Here a fundamental domain is not a subgraph if $V$ has loops.) Let $\bar{v}_0$ be the vertex of $F$ over $v_0$. Let $\mathbb{V} = (V, \mathcal{V}, v_0)$ be the corresponding graph of groups associated with $\rho$.

For a cell (vertex or edge) $c$ of $V$, the stabilizer $G_c \in \mathcal{V}$ is of the form $B \rtimes A_c$ (where $A_c \leq A$ is the stabilizer of $c$ under the action by $A$). Following Remark 2, we assume $G_c$ is not finitely generated. Let $R_c$ be a finite generating set for $A_c$, and let $S_c$ be an infinite generating set of $B$ which contains a finite set $S$ such that $S$ generates $B$ over $A$, as described in Definition 1. Let $X_c$ be a $K(G_c, 1)$-complex having a single 0-cell and 1-cells in correspondence with $R_c \cup S_c$ [Geoghegan 2008, Chapter 7]; this is called a "vertex (or edge) space," depending on whether $c$ is a vertex or edge. There is covering space $\bar{X}_c \twoheadrightarrow X_c$ which is a $K(B, 1)$, since $B \leq G_c$.

As in [Geoghegan 2008, Theorem 7.1.9], we assemble a $K(G, 1)$-complex $(X, x_0)$ as a total space for the graph of groups $(V, \mathcal{V}, v_0)$. This is formed as a disjoint union of the vertex spaces $X_v$, to which we attach $X_e \times I$ for each edge $e$. The attaching maps are such that the induced maps on $\pi_1$ induce inclusions $G_e \hookrightarrow G_v$ when $v$ is an endpoint of $e$. There is a retraction $r : (X, x_0) \to (V, v_0)$ collapsing $X_c$ (or $X_c \times I$ if $c$ is an edge) to $c$ for each cell $c$ of $V$. There is a covering space $q : (\bar{X}, \bar{x}_0) \to (X, x_0)$ corresponding to $B$. This, too, can be described as a total space of a graph of groups where the graph is the tree $T$ itself, and each stabilizer is isomorphic to $B$, since $T = B\backslash T$.

We then have the universal cover $p : (\tilde{X}, \tilde{x}) \twoheadrightarrow (\bar{X}, \bar{x})$. Above the map $r$ are maps $\bar{r} : (\bar{X}, \bar{x}_0) \to (T, \bar{v}_0)$ and $\tilde{r} : (\tilde{X}, \tilde{x}_0) \to (T, \bar{v}_0)$.

All maps are $G$-equivariant and continuous. We arrive at the commutative diagram given before the statement of Theorem 3.

## 3. Analysis of $\Sigma^1$ via subcomplexes of $\bar{X}$

We continue using the notation of the previous section.

**Remark 19.** Let the end point $e$ be represented by the geodesic ray $\tau$. Because $\tau$ emanates from a vertex, the horoball $\mathrm{HB}_t(\tau)$ is a subtree of $T$ if and only if $t \in \mathbb{Z}$. We are interested in $\tilde{X}_{(\tau,t)} \subset \tilde{X}$, which is by definition the largest subcomplex of $(\bar{r} \circ p)^{-1}(\mathrm{HB}_t(\tau))$; by choice of $\tau$, $X$, $\bar{r}$, and $p$, $\tilde{X}_{(\tau,t)} = (\bar{r} \circ p)^{-1}(\mathrm{HB}_t(\tau))$ exactly when $t \in \mathbb{Z}$. (There are no 0-cells of $\bar{X}$ mapped by $\bar{r}$ to the interior of an edge of $T$.) Hence, it is enough to look at horoballs of the form $\mathrm{HB}_k(\tau)$, $k \in \mathbb{Z}$. Similarly, the lag $\lambda$ can always be taken to be in $\mathbb{Z}$, so that all horoballs under consideration are subtrees of $T$.

**Definition 20.** A finite subcomplex $W$ will be called *suitable* if for each subtree $U$ of $T$, the set $\bar{r}^{-1}(U) \cap q^{-1}(W) \subset \bar{X}$ is connected. By Remark 19, it follows that if $W$ is suitable, then the set $\bar{X}_{(\tau,k,W)} = \bar{r}^{-1}(\mathrm{HB}_k(\tau)) \cap q^{-1}(W)$ is connected for any horoball $\mathrm{HB}_k(\tau)$.

**Lemma 21.** *Suppose $W$ is a connected subcomplex of $X$ such that for each vertex $v$ of $F$, $W$ contains the 1-cells of $X_v \subset X$ corresponding to $R_v$. Moreover, for each edge $e$ of $F$, let $x_e \in X_e$ be the basepoint, and suppose $W$ contains the 1-cell $\{x_e\} \times [0, 1] \in X$. Then $W$ is suitable.*

*Proof.* Let $U$ be a subtree of $T$. We show that $q^{-1}(W) \cap \bar{r}^{-1}(U)$ is connected. For a given vertex $v$ of $F$, $W$ contains loops generating $A_v$, and the image of the map $\bar{X}_v \hookrightarrow X_v$ is $B$. By Proposition 17 (with $H_1 \geq A_v$ and $H_2 = B$), $q^{-1}(W) \cap \bar{X}_v$ is connected. Hence the lemma holds if $U$ is any vertex of $T$. If $U$ contains edges, then since $W$ contains all edges of $X$ corresponding to base points of $X_e$, $e \in F$, there must be a path in $q^{-1}(W) \cap \bar{r}^{-1}(U)$ from the $\bar{r}$-preimage of any one vertex of $U$ to any other. Furthermore, the fact that there is no cell of $\bar{X}$ lying completely over the interior of an edge of $T$ ensures that there can be no components of $q^{-1}(W) \cap \bar{r}^{-1}(U)$ over the interior of an edge. □

Because each stabilizer $A_v$ is finitely generated and $V$ is finite, the following observation follows from Lemma 21.

**Observation 22.** If $W \subseteq X$ is compact, then there exists a suitable subcomplex $W' \subseteq X$ such that $W \subseteq W'$.

For convenience, we restate Theorem 3 before proving it. Recall that $\bar{X}_{(\tau,k,W)}$ denotes $\bar{r}^{-1}(\mathrm{HB}_k(\tau)) \cap q^{-1}(W) \subset \bar{X}$.

**Theorem.** *Let $e \in \partial T$ be represented by a geodesic ray $\tau$.*

(i) *If there exists a finite subcomplex $W \subset X$ such that for every $k \in \mathbb{Z}$, $\bar{X}_{(\tau,k,W)}$ is connected and the map on $\pi_1$ induced by the inclusion $\bar{X}_{(\tau,k,W)} \hookrightarrow \bar{X}$ is surjective, then $e \in \Sigma^1(\rho)$.*

(ii) *If for every $k \in \mathbb{Z}$ and every finite subcomplex $W \subset X$ such that $\bar{X}_{(\tau,k,W)}$ is connected, the induced map on $\pi_1$ is not surjective, then $e \notin \Sigma^1(\rho)$.*

*Proof.* (i) We show that Definition 14 is satisfied with lag $\lambda = 0$; in this case, conditions $(*)$ and $(**)$ are the same. Let $\tilde{L} \subseteq \tilde{X}$ be a cocompact $G$-subcomplex and set $L = q(p(\tilde{L}))$. Let $k \in \mathbb{Z}$. By Observation 22, there is a suitable subcomplex $W' \subseteq X$ with $L \cup W \subseteq W'$. Since $\bar{X}_{(\tau,k,W)}$ is $\pi_1$-surjective onto $\bar{X}$, it follows that $\bar{X}_{(\tau,k,W')}$ is as well. Because $W'$ is suitable, Proposition 17 applies to $\bar{X}_{(\tau,k,W')} \subset \bar{X}$ to ensure that $p^{-1}(q^{-1}(W')) \cap \tilde{X}_{(\tau,k)}$ is connected. Moreover this contains $L \cap \tilde{X}_{(\tau,k)}$, so condition $(*)$ is satisfied.

(ii) Let $\tilde{L}$ be a cocompact $G$-subcomplex of $\tilde{X}$, and let $\tilde{L}'$ be any cocompact $G$-subcomplex of $\tilde{X}$ containing $\tilde{L}$. We show that for any lag $k \geq 0 \in \mathbb{Z}$, there exist points of $\tilde{L} \cap \tilde{X}_{(\tau,0)}$ lying in distinct components of $\tilde{L}' \cap \tilde{X}_{(\tau,-k)}$.

Let $L = p(q(\tilde{L}))$ and $L' = p(q(\tilde{L}'))$. By Observation 22 there exists a suitable complex $W \subseteq X$ with $L' \subseteq W$. Then $\bar{X}_{(\tau,-k,W)}$ is connected, and by assumption it is not $\pi_1$-surjective. Set $\widetilde{W} = q^{-1}(p^{-1}(W))$. Then $\widetilde{W} \cap \tilde{X}_{(\tau,-k)}$ is disconnected by Proposition 17. Furthermore, Proposition 18 ensures that each of its components contains components of $\tilde{L}' \cap \tilde{X}_{(\tau,-k)}$, which in turn contain points of $\tilde{L} \cap \tilde{X}_{(\tau,-k)}$. □

## 4. $A$ a free group

Let the action $\rho$ by $G$ on $T$ be as defined in Definition 1, with the additional restriction that $A$ is a free group on the set $\{a_1, \ldots, a_n\}$ and $T$ is its Cayley graph with respect to this set. Then the vertices of $T$ are the elements of $A$. Let $X$, $q : (\bar{X}, \bar{x}_0) \to (X, \bar{x})$, $p : (\tilde{X}, \tilde{x}_0) \to (\bar{X}, \bar{x}_0)$, $r : X \to V$, and $\bar{r} : \bar{X} \to T$ be as defined in Section 2.2. The graph $V = A\backslash T$ has a unique vertex $v_0$, so the $K(G,1)$-complex $X$ can be chosen to have a unique 0-cell $x_0$, which we naturally choose as basepoint for $X$. In this case, for any cell $c$ of $V$, $X_c$ and $\bar{X}_c$ are both $K(B,1)$-complexes. In fact, we can take $\bar{X}_c = X_c = X_{v_0}$ for all $c$, since passing from $X$ to $\bar{X}$ simply "unwraps" loops in $A \subseteq G = \pi_1(X, x_0)$. Choose the base point $\bar{x}_0$ of $\bar{X}$ to be the unique 0-cell of $\bar{X}$ mapped to $1 \in A = \operatorname{vert} T$.

We uniquely represent $\partial T$ by geodesic rays $\tau$, with $\tau(0) = 1 \in A$ and $\tau(n)$ a freely reduced word on $n$ letters. Thus each geodesic ray $\tau$ corresponds to a unique infinite freely reduced word $\prod_{i \in \mathbb{Z}_{\geq 0}} c_i$.

**4.1.** ***From suitable complexes to subgroups.*** From here on, we identify $B$ with $\pi_1(\bar{X}, \bar{x}_0)$. Let $W$ be a suitable subcomplex of $X$. Since $W$ is finite, the subgroup

$$B(W) = \text{inclusion}_{\#}(q^{-1}(W) \cap \bar{r}^{-1}(1), \bar{x}_0) \leq B$$

is finitely generated. Let $S(W)$ be a finite generating set for $B(W)$. Let $T'$ be a subtree of $T$. Fix $v \in \text{vert}\, T' \subseteq A$. Then $\bar{r}^{-1}(v) \cong \bar{X}_c$ has a single 0-cell; call it $x'$. Let $B(W, T', v)$ be the image of $\pi_1\big(q^{-1}(W) \cap \bar{r}^{-1}(T'), x'\big)$ in $\pi_1(\bar{X}, x')$. Let

$$\Psi_v : \pi_1(\bar{X}, x') \to \pi_1(\bar{X}, \bar{x}_0) = B$$

be the change-of-basepoint isomorphism. Then for $g \in \pi_1(\bar{X}, x')$, $\Psi_v(g) = vgv^{-1}$.

**Lemma 23.** *The subgroup of $B$ generated by $\{usu^{-1} \mid s \in S(W), u \in T'\}$ is $\Psi_v(B(W, T', v))$.*

*Proof.* Any element $h \in B(W, T', v)$ can represented by a loop $\sigma_h$ in the 1-skeleton of $q^{-1}(W) \cap \bar{r}^{-1}(T')$ based at $x'$. Because $\bar{X}$ has no 0-cells over the interiors of edges of $T$, and because each vertex space is a copy of $X_{v_0}$ and each edge space a copy of $X_{v_0} \times [0, 1]$, the loop $\sigma_h$ may be decomposed as concatenation of subpaths $\sigma_h^0, \sigma_h^1, \ldots, \sigma_h^m$, $m \in \mathbb{N}$, where each $\sigma_h^i$, $0 \leq i \leq m$, is either a 1-cell joining one vertex space to another (a "base edge" for an edge space) or a loop contained entirely in a vertex space and corresponding to some $s \in S(W)$. Between each pair of subpaths, we may introduce a path which returns straight back to $x'$ (i.e., via 1-cells over lying over edges of $T'$ exclusively). This process rewrites $h$ as a product of conjugates of the form $v^{-1}usu^{-1}v$, $s \in S(W)$, $u \in T'$. $\square$

Combining Theorem 3 with Lemma 23, we obtain a purely algebraic condition for determining whether an endpoint lies in $\Sigma^1(\rho)$. For a geodesic ray $\tau$ corresponding to the infinite word $\prod_i c_i$ and $k \in \mathbb{Z}$, define $A_k(\tau) = \text{vert}(\text{HB}_k(\tau))$ and $w_k = \tau(k) = c_1 c_2 \ldots c_k$. Then

$$\pi_1\big(q^{-1}(W) \cap \bar{r}^{-1}(\text{HB}_k(\tau)), w_k\big) = B(W, \text{HB}_k(\tau), w_k).$$

**Theorem.** *Let $A$ be a finitely generated free group, and let $T$ be its Cayley graph with respect to a free basis. For the action $\rho$ as in Theorem 3, and for $e \in \partial T$ represented by a geodesic ray $\tau$,*

(i) *If there is a finite set $S \subseteq B$ such that for each $k \in \mathbb{Z}_{\geq 0}$, $S$ generates $B$ over $A_k(\tau)$, then $e \in \Sigma^1(\rho)$.*

(ii) *If for each $k \in \mathbb{Z}_{\leq 0}$, $B$ is not finitely generated over $A_k(\tau)$, then $e \notin \Sigma^1(\rho)$.*

*Proof.* (i) If there is such a finite set $S$, then we can choose a suitable subcomplex $W$ containing loops corresponding to $S$. For any $k \in \mathbb{Z}_{\geq 0}$, let $x'$ be the unique vertex of $\bar{r}^{-1}(w_k)$, and we have

$$B(W, \text{HB}_k(\tau), w_k) = \Psi_{w_k}^{-1}(B) = \pi_1(\bar{X}, x').$$

Thus by Theorem 3(i) we obtain $e \in \Sigma^1(\rho)$.

(ii) Given a suitable subcomplex $W$ of $X$ and $k \in \mathbb{Z}_{\le 0}$, by assumption the subgroup $\Psi(B(W, \mathrm{HB}_k(\tau), w_k))$ is a proper subgroup of $B$. Hence, $B(W, \mathrm{HB}_k(\tau), w_k)$ is a proper subgroup of $\pi_1(\bar{X}, x')$. Thus, by part (ii) of Theorem 3, $e \notin \Sigma^1(\rho)$. □

Recall that for $t \in \{a_1, \ldots, a_n\}^{\pm}$, the function $\mathrm{expsum}_t$ maps a reduced word $w$ in $\{a_1, \ldots, a_n\}^{\pm}$ to the corresponding exponent sum of $t$ in $w$. Also, recall we use the notation $\mathrm{Ball}_r(A, v)$ to refer to the $r$-ball around $v$ in $A$ (in the word metric), to avoid confusion with the subgroup $B$.

**Lemma 24.** *For an endpoint $e$ represented by the geodesic ray $\tau$, let*

$$Q_{t,k}(\tau) = \{\mathrm{expsum}_t(v) \mid v \in A_k(\tau)\} \subseteq \mathbb{Z}.$$

*Then $Q_{t,k}(\tau)$ is bounded above if and only if $\tau$ eventually consists of only $t^{-1}$. Moreover, $Q_{t,k}(\tau)$ contains every integer within its bounds.*

*Proof.* Let $\tau$ be represented by the infinite word $c_1c_2\ldots$, and fix $k \in \mathbb{Z}$. Recall that $A_k(\tau) = \bigcup_{l \ge \max\{0,k\}} \overline{\mathrm{Ball}_{l-k}(A, c_1c_2\ldots c_l)}$.

Suppose for $N \in \mathbb{Z}$, $c_i = t^{-1}$ for all $i > N$. For $j = 0, 1, 2, \ldots$, the words $g_j = c_1c_2\ldots c_{N+j}t^{N+j-k}$ all represent the same element of $A$, and $g_j$ has maximal $\mathrm{expsum}_t$ among elements of $\overline{\mathrm{Ball}_{N+j-k}(A, c_1c_2\ldots c_{N+j})}$. Since $A_k(\tau)$ is the union of these subsets, it follows that $Q_{t,k}(\tau)$ is bounded above.

On the other hand, suppose that there are infinitely many $i \in \mathbb{Z}$ such that $c_i \ne t^{-1}$. For $j \in \mathbb{Z}$, $j \ge \max\{0, k\}$, let $m(j)$ be the number of letters $c_i$ in $c_1c_2\ldots c_j$ with $c_i \ne t^{-1}$. By assumption $m(j) \to \infty$ as $j \to \infty$. Let $g_j = c_1\ldots c_jt^{j-k}$. Then

$$g_j \in \overline{\mathrm{Ball}_{j-k}(A, c_1c_2\ldots c_j)} \subseteq A_k(\tau).$$

Since $\mathrm{expsum}_t(c_1c_2\ldots c_j) \ge -(j - m(j))$,

$$\mathrm{expsum}_t(g_j) = \mathrm{expsum}_t(c_1c_2\ldots c_j) + j - k \ge m(j) - k.$$

Letting $j \to \infty$, we have that $Q_{t,k}(\tau)$ is not bounded above.

The fact that $Q_{t,k}(\tau)$ contains every integer within its bounds follows from the observation that for $v, w \in A_k(\tau)$, if

$$\mathrm{expsum}_t(v) < m < \mathrm{expsum}_t(w),$$

the path connecting $v$ to $w$ contains a vertex $u$ with $\mathrm{expsum}_t(u) = m$. □

*Proof of Corollary 6.* Let $t \in \{a_1, \ldots, a_n\}^{\pm}$. Suppose $e \in \partial T$ is represented by an infinite word eventually consisting of only $t^{-1}$, and suppose there exists no $m \in \mathbb{Z}$ such that $B$ is finitely generated over $A - \mathrm{expsum}_t^{-1}([m, \infty))$. By Lemma 24, $\{\mathrm{expsum}_t(a) \mid a \in A_k(\tau)\}$ is bounded above. Hence, $B$ cannot be finitely generated over $A_k(\tau)$, and so by Theorem 4, part (ii), $e \notin \Sigma^1(\rho)$. □

*Proof of Theorem 9.* Let $e = \tau(\infty)$, with $\tau$ corresponding to the infinite word $\prod_i c_i$. By Corollary 5, it is enough to show that $\varphi(A_k(\tau)) = \varphi(A)$ for each $k \geq 0 \in \mathbb{Z}$.

Let $w \in \mathcal{A}^*$ be a freely reduced word, and let $l$ be the reduced length of $w$. We will find $w' \in A_k(\tau)$ with $\varphi(w') = \varphi(w)$. Choose $m \in \mathbb{Z}_{\geq 0}$ large enough to ensure that the word $c_1 \dots c_m$ has $k+l$ distinct subwords in $\ker\varphi$. Call these subwords $\zeta_i$, $1 \leq i \leq k+l$, and let the remaining letters form subwords $\chi_i$, $1 \leq i \leq k+l$, so that we have the decomposition

$$c_1 \dots c_m = \chi_1\zeta_1\chi_2\zeta_2 \dots \chi_{k+l}\zeta_{k+l},$$

where each $\varphi(\zeta_i)$ is trivial, and each $\chi_i$ is possibly empty.

Now

$$\varphi(c_1c_2 \dots c_m) = \varphi(\chi_1\chi_2 \dots \chi_{k+l}),$$

and the reduced length of $\chi_1\chi_2 \dots \chi_{k+l}$ is no greater than $m - l - k$. Thus the word $\xi = c_1c_2 \dots c_m\chi_{k+l}^{-1} \dots \chi_2^{-1}\chi_1^{-1}$ is in both $\ker\varphi$ and $\overline{\mathrm{Ball}_{m-l-k}(A, c_1 \dots c_m)}$; moreover

$$\xi w \in \overline{\mathrm{Ball}_{m-k}(A, c_1 \dots c_m)} \subseteq A_k(\tau) \quad \text{and} \quad \varphi(w) = \varphi(\xi w). \qquad \square$$

**4.2. *Argument for Example 11.*** In Example 11, $G = B \rtimes_\varphi A$, where $A = C * D$ for free groups $C = \langle a_1, \dots, a_n\rangle$, $D = \langle a_{n+1}, \dots, a_m\rangle$, and $B = \bigoplus_{\omega\in C} K_\omega$ for some finitely generated group $K$. The claim is made that any endpoint of $T = \Gamma(A, \{a_1, a_2, \dots, a_m\})$ represented by a ray $\tau$ whose letters are eventually selected only from $C$ does not lie in $\Sigma^1$. Since $\Sigma^1$ is $G$-invariant, we can assume $\tau$ consists of letters entirely in $C$. Then $\pi : A \twoheadrightarrow C$ fixes each vertex of $\tau$. Moreover, it makes sense to discuss the subset $C_k(\tau) \subseteq C$.

Let $k \in \mathbb{Z}_{\leq 0}$ be given, and let $S$ be any finite subset of $B$. We will show that the set $S' = \{\varphi_a(s) \mid s \in S,\ a \in A_k(\tau)\}$ does not generate $B$. Part (ii) of Theorem 4 thereby ensures that $\tau(\infty) \notin \Sigma^1(\rho)$.

To show that $S'$ does not generate $B$, we will find an index $\psi \in C$ such that every $s \in S'$ is trivial at index $\psi$.

**Observation 25.** If $a \in A$ is in $A_k(\tau)$, then $\pi(a)$ is in $C_k(\tau)$.

*Proof.* Since $a \in A_k(\tau)$ and $k \leq 0$, there exists $l \geq 0$ such that $a \in \overline{\mathrm{Ball}_{l-k}(A, \tau(l))}$ by (1-2), so $\pi(a) \in \overline{\mathrm{Ball}_{l-k}(C, \tau(l))}$. But this is contained in $C_k(\tau)$, again by (1-2). $\square$

Define the set

$$\mathcal{I}(S) = \{\omega \in C \mid \exists\, s \in S \text{ such that } s \text{ is nontrivial at index } \omega\}.$$

Note that $\mathcal{I}(S)$ is a finite set, since $S$ is finite and each $s \in S$ is nontrivial at only finitely many indices. Define

$$\mathcal{R}(S) = \max\{\text{reduced length of } \omega \mid \omega \in \mathcal{I}(S)\}.$$

Since $\mathcal{I}(S)$ is finite, $\mathcal{R}(S)$ is a nonnegative integer representing the maximum distance (in $C$) from any index of any nontrivial component of any element of $S$ to the identity index $1 \in C$.

Since left multiplication by $c \in C$ is an isometry on $C$, it follows that the maximal distance in $C$ from any nontrivial index of any element of $\varphi_c(S)$ to $c$ is also $\mathcal{R}(S)$. Observation 25 therefore ensures that the set of nontrivial indices of elements of $S'$ is a subset of the closed $\mathcal{R}(S)$-neighborhood of $C_k(\tau)$ in $C$. In fact, this neighborhood is the set $C_{k-\mathcal{R}(S)}(\tau)$. This is a proper subset of $C$ (simply choose any geodesic ray other than $\tau$ and follow it far enough). For any $\psi \in C$ with $\psi \notin C_{k-\mathcal{R}(S)}(\tau)$, all $s \in S'$ will be trivial at index $\psi$. So $S'$ can not generate $B$.

## Acknowledgments

This paper extends results from the author's Ph.D dissertation at Binghamton University. It would not have been possible without the continuous support and encouragement of Ross Geoghegan. The author also wishes to thank Robert Bieri for his generosity with his time in discussing some of the ideas involved in this work. Additionally, the author is indebted to the anonymous referee, whose efforts have led to a more interesting and clearer paper.

## References

[Bass 1993] H. Bass, "Covering theory for graphs of groups", *J. Pure Appl. Algebra* **89**:1-2 (1993), 3–47. MR 94j:20028 Zbl 0805.57001

[Bieri and Geoghegan 2003a] R. Bieri and R. Geoghegan, "Connectivity properties of group actions on non-positively curved spaces", *Mem. Amer. Math. Soc.* **161**:765 (2003). MR 2004m:57001 Zbl 1109.20035

[Bieri and Geoghegan 2003b] R. Bieri and R. Geoghegan, "Topological properties of $SL_2$ actions on the hyperbolic plane", *Geom. Dedicata* **99** (2003), 137–166. MR 2004e:20068 Zbl 1039.20020

[Bieri and Strebel 1980] R. Bieri and R. Strebel, "Valuations and finitely presented metabelian groups", *Proc. London Math. Soc.* (3) **41**:3 (1980), 439–464. MR 81j:20080 Zbl 0448.20029

[Bieri et al. 1987] R. Bieri, W. D. Neumann, and R. Strebel, "A geometric invariant of discrete groups", *Invent. Math.* **90**:3 (1987), 451–477. MR 89b:20108 Zbl 0642.57002

[Bridson and Haefliger 1999] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften **319**, Springer, Berlin, 1999. MR 2000k:53038 Zbl 0988.53001

[Brown 1987] K. S. Brown, "Trees, valuations, and the Bieri–Neumann–Strebel invariant", *Invent. Math.* **90**:3 (1987), 479–504. MR 89e:20060 Zbl 0663.20033

[Geoghegan 2008] R. Geoghegan, *Topological methods in group theory*, Graduate Texts in Mathematics **243**, Springer, New York, 2008. MR 2008j:57002 Zbl 1141.57001

[Jones 2012] K. Jones, "Connectivity properties for actions on locally finite trees", *Pacific J. Math.* **255**:1 (2012), 143–154. MR 2923697 Zbl 06029230

[Lehnert 2009] R. Lehnert, *Kontrollierter Zusammenhang von Gruppenoperationen auf Bäumen*, Diploma thesis, Goethe Universität, Frankfurt am Main, October 2009, Available at http://tinyurl.com/Lehnert-Diplomarbeit-2009.

[Scott and Wall 1979] P. Scott and T. Wall, "Topological methods in group theory", pp. 137–203 in *Homological group theory* (Durham, 1977), edited by C. T. C. Wall, London Math. Soc. Lecture Note Ser. **36**, Cambridge University Press, 1979. MR 81m:57002 Zbl 0423.20023

[Serre 1980] J.-P. Serre, *Trees*, Springer, Berlin, 1980. MR 82c:20083 Zbl 0548.20018

Received September 20, 2012. Revised November 29, 2013.

KEITH JONES
DEPARTMENT OF MATHEMATICS, COMPUTER SCIENCE, AND STATISTICS
SUNY ONEONTA
108 RAVINE PARKWAY
ONEONTA, NY 13820
UNITED STATES

keith.jones@oneonta.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.95

# AN INDISPENSABLE CLASSIFICATION OF MONOMIAL CURVES IN $\mathbb{A}^4(\Bbbk)$

ANARGYROS KATSABEKIS AND IGNACIO OJEDA

**We give a new classification of monomial curves in $\mathbb{A}^4(\Bbbk)$. It relies on the detection of those binomials and monomials that have to appear in every system of binomial generators of the defining ideal of the monomial curve; these special binomials and monomials are called indispensable in the literature. This way to proceed has the advantage of producing a natural necessary and sufficient condition for the defining ideal of a monomial curve in $\mathbb{A}^4(\Bbbk)$ to have a unique minimal system of binomial generators. Furthermore, some other interesting results on more general classes of binomial ideals with unique minimal system of binomial generators are obtained.**

## Introduction

Let $\Bbbk[\boldsymbol{x}] := \Bbbk[x_1, \ldots, x_n]$ be the polynomial ring in $n$ variables over a field $\Bbbk$. As usual, we will denote by $\boldsymbol{x}^{\boldsymbol{u}}$ the monomial $x_1^{u_1} \cdots x_n^{u_n}$ of $\Bbbk[\boldsymbol{x}]$, with $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{N}^n$, where $\mathbb{N}$ stands for the set of non-negative integers. Recall that a pure difference binomial ideal is an ideal of $\Bbbk[\boldsymbol{x}]$ generated by differences of monic monomials. Examples of pure difference binomial ideals are the toric ideals. Indeed, let $\mathscr{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \subset \mathbb{Z}^d$ and consider the semigroup homomorphism $\pi : \Bbbk[\boldsymbol{x}] \to \Bbbk[\mathscr{A}] := \bigoplus_{\boldsymbol{a} \in \mathscr{A}} \Bbbk\, \boldsymbol{t}^{\boldsymbol{a}};\ x_i \mapsto \boldsymbol{t}^{\boldsymbol{a}_i}$. The kernel of $\pi$ is denoted by $I_{\mathscr{A}}$ and called the toric ideal of $\mathscr{A}$. Notice that the toric ideal $I_{\mathscr{A}}$ is generated by all the binomials $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$ such that $\pi(\boldsymbol{x}^{\boldsymbol{u}}) = \pi(\boldsymbol{x}^{\boldsymbol{v}})$, see, for example, [Sturmfels 1996, Lemma 4.1].

Defining ideals of monomial curves in the affine $n$-dimensional space $\mathbb{A}^n(\Bbbk)$ serve as interesting examples of toric ideals. Of particular interest is to compute and describe a minimal generating set for such an ideal. Herzog [1970] provides a minimal system of generators for the defining ideal of a monomial space curve.

Ojeda is partially supported by the project MTM2012-36917-C03-01, National Plan I+D+i and by Junta de Extremadura (FEDER funds).

*MSC2010:* primary 13F20; secondary 16W50, 13F55.

*Keywords:* binomial ideal, toric ideal, monomial curve, minimal systems of generators, indispensable monomials, indispensable binomials.

The case $n = 4$ was treated in [Bresinsky 1988], where Gröbner bases techniques were used to obtain a minimal generating set of the ideal.

A recent topic arising in algebraic statistics is to study the problem when a toric ideal has a unique minimal system of binomial generators, see [Charalambous et al. 2007; Ojeda and Vigneron-Tenorio 2010a]. To deal with this problem, Ohsugi and Hibi [2005] introduced the notion of indispensable binomials, while Aoki, Takemura and Yoshida [Aoki et al. 2008] introduced the notion of indispensable monomials. The problem was considered for the case of defining ideals of monomial curves in [García and Ojeda 2010]. Although this work offers useful information, the classification of the ideals having a unique minimal system of binomial generators remains an unsolved problem for $n \geq 4$. For monomial space curves Herzog's result provides an explicit classification of those defining ideals satisfying the above property. The aim of this work is to classify all defining ideals of monomial curves in $\mathbb{A}^4(\Bbbk)$ having a unique minimal system of generators. Our approach is inspired by the classification made by Pilar Pisón in her unpublished thesis.

The paper is organized as follows. In Section 1 we study indispensable monomials and binomials of a pure difference binomial ideal. We provide a criterion for checking whether a monomial is indispensable (Theorem 1.9) and a sufficient condition for a binomial to be indispensable (Theorem 1.10). As an application we prove that the binomial edge ideal of an undirected simple graph has a unique minimal system of binomial generators. Section 2 is devoted to special classes of binomial ideals contained in the defining ideal of a monomial curve. Corollary 2.5 underlines the significance of the critical ideal in the investigation of our problem. Theorem 2.12 and Proposition 2.13 provide necessary and sufficient conditions for a circuit to be indispensable in the toric ideal, while Corollary 2.16 will be particularly useful in the next section. In Section 3 we study defining ideals of monomial curves in $\mathbb{A}^4(\Bbbk)$. Theorem 3.6 carries out a thorough analysis of a minimal generating set of the critical ideal. This analysis is used to derive a minimal generating set for the defining ideal of the monomial curve (Theorem 3.10). As a consequence we obtain the desired classification (Theorem 3.11). Finally we prove that the defining ideal of a Gorenstein monomial curve in $\mathbb{A}^4(\Bbbk)$ has a unique minimal system of binomial generators, under the hypothesis that the ideal is not a complete intersection.

## 1. Generalities on indispensable monomials and binomials

Let $\Bbbk[\boldsymbol{x}]$ be the polynomial ring over a field $\Bbbk$. The following result is folklore, but for a lack of reference we sketch a proof.

**Theorem 1.1.** *Let $J \subset \Bbbk[\boldsymbol{x}]$ be a pure difference binomial ideal. There exist a positive integer $d$ and a vector configuration $\mathscr{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \subset \mathbb{Z}^d$ such that the toric ideal $I_{\mathscr{A}}$ is a minimal prime of $J$.*

*Proof.* By [Eisenbud and Sturmfels 1996, Corollary 2.5], $\big(J : (x_1 \cdots x_n)^\infty\big)$ is a lattice ideal. More precisely, if $\mathscr{L} = \operatorname{span}_{\mathbb{Z}}\{\boldsymbol{u} - \boldsymbol{v} \mid \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J\}$, then

$$\big(J : (x_1 \cdots x_n)^\infty\big) = \langle \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \mid \boldsymbol{u} - \boldsymbol{v} \in \mathscr{L}\rangle =: I_{\mathscr{L}}.$$

Now, by [Eisenbud and Sturmfels 1996, Corollary 2.2], the only minimal prime of $I_{\mathscr{L}}$ that is a pure difference binomial ideal is $I_{\operatorname{Sat}(\mathscr{L})} := \langle \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \mid \boldsymbol{u} - \boldsymbol{v} \in \operatorname{Sat}(\mathscr{L})\rangle$, where $\operatorname{Sat}(\mathscr{L}) := \{\boldsymbol{u} \in \mathbb{Z}^n \mid z\,\boldsymbol{u} \in \mathscr{L}$ for some $z \in \mathbb{Z}\}$. Since $\mathbb{Z}^n/\operatorname{Sat}(\mathscr{L}) \cong \mathbb{Z}^d$, for $d = n - \operatorname{rank}(\mathscr{L})$, then $\boldsymbol{e}_i + \operatorname{Sat}(\mathscr{L}) = \boldsymbol{a}_i \in \mathbb{Z}^d$, for every $i = 1, \ldots, n$, and hence the toric ideal of $\mathscr{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is equal to $I_{\operatorname{Sat}(\mathscr{L})}$; see [Sturmfels 1996, Lemma 12.2].

Finally, in order to see that $I_{\mathscr{A}}$ is a minimal prime of $J$, it suffices to note that $J \subseteq P$ implies $\big(J : (x_1 \cdots x_n)^\infty\big) \subseteq P$, for every prime ideal $P$ of $\Bbbk[\boldsymbol{x}]$. □

**Remark 1.2.** If $J = \langle \boldsymbol{x}^{\boldsymbol{u}_j} - \boldsymbol{x}^{\boldsymbol{v}_j} \mid j = 1, \ldots, s\rangle$, then $\mathscr{L} = \operatorname{span}_{\mathbb{Z}}\{\boldsymbol{u}_j - \boldsymbol{v}_j \mid j = 1, \ldots, s\}$. So, it is easy to see that, in general, $J \neq I_{\mathscr{L}}$. For example, if $J = \langle x - y, z - t, y^2 - yt\rangle$, then $I_{\mathscr{L}} = \langle x - t, y - t, z - t\rangle$.

Given a vector configuration $\mathscr{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \subset \mathbb{Z}^d$, we grade $\Bbbk[\boldsymbol{x}]$ by setting $\deg_{\mathscr{A}}(x_i) = \boldsymbol{a}_i,\ i = 1, \ldots, n$. We define the $\mathscr{A}$-degree of a monomial $\boldsymbol{x}^{\boldsymbol{u}}$ to be

$$\deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}}) = u_1\boldsymbol{a}_1 + \cdots + u_n\boldsymbol{a}_n.$$

A polynomial $f \in \Bbbk[\boldsymbol{x}]$ is $\mathscr{A}$-homogeneous if the $\mathscr{A}$-degrees of all the monomials that occur in $f$ are the same. An ideal $J \subset \Bbbk[\boldsymbol{x}]$ is $\mathscr{A}$-homogeneous if it is generated by $\mathscr{A}$-homogeneous polynomials. The toric ideal $I_{\mathscr{A}}$ is $\mathscr{A}$-homogeneous; indeed, by [Sturmfels 1996, Lemma 4.1], a binomial $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$ if and only if it is $\mathscr{A}$-homogeneous.

The proof of the following result is straightforward.

**Corollary 1.3.** *Let $J \subset \Bbbk[\boldsymbol{x}]$ be a pure difference binomial ideal and let $\mathscr{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \subset \mathbb{Z}^d$. Then $J$ is $\mathscr{A}$-homogeneous if and only if $J \subseteq I_{\mathscr{A}}$.*

Notice that the finest $\mathscr{A}$-grading on $\Bbbk[\boldsymbol{x}]$ such that a pure difference binomial ideal $J \subset \Bbbk[\boldsymbol{x}]$ is $\mathscr{A}$-homogeneous occurs when $I_{\mathscr{A}}$ is a minimal prime of $J$. Such an $\mathscr{A}$-grading does always exist by Theorem 1.1. Ideals with finest $\mathscr{A}$-grading are studied in much greater generality in [Katsabekis and Thoma 2010]. An $\mathscr{A}$-grading on $\Bbbk[\boldsymbol{x}]$ such that a pure difference binomial ideal $J \subset \Bbbk[\boldsymbol{x}]$ is $\mathscr{A}$-homogeneous is said to be positive if the quotient ring $\Bbbk[\boldsymbol{x}]/I_{\mathscr{A}}$ does not contain invertible elements or, equivalently, if the monoid $\mathbb{N}\mathscr{A}$ is free of units.

Recall (from [Sturmfels 1996, Chapter 12], for instance) that *the number of polynomials of $\mathscr{A}$-degree $\boldsymbol{b} \in \mathbb{N}\mathscr{A}$ in any minimal system of $\mathscr{A}$-homogeneous generators is* $\dim_{\Bbbk} \operatorname{Tor}_1^R(\Bbbk, \Bbbk[\mathscr{A}])_{\boldsymbol{b}}$. Thus, we say that $I_{\mathscr{A}}$ has minimal generators in degree $\boldsymbol{b}$ when $\dim_{\Bbbk} \operatorname{Tor}_1^R(\Bbbk, \Bbbk[\mathscr{A}])_{\boldsymbol{b}} \neq 0$. In this case, if $f \in I_{\mathscr{A}}$ has degree $\boldsymbol{b}$ we say that $f$ is a *minimal generator* of $I_{\mathscr{A}}$.

From now on, let $\mathcal{A} = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\} \subset \mathbb{Z}^d$ be such that the quotient ring $\Bbbk[\boldsymbol{x}]/I_{\mathcal{A}}$ does not contain invertible elements and let $J \subset \Bbbk[\boldsymbol{x}]$ be an $\mathcal{A}$-homogeneous pure difference binomial ideal.

**Definition 1.4.** A binomial $f = \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J$ is called *indispensable* in $J$ (or an *indispensable binomial* of $J$) if every system of binomial generators of $J$ contains $f$ or $-f$. A monomial $\boldsymbol{x}^{\boldsymbol{u}}$ is called *indispensable* in $J$ if every system of binomial generators of $J$ contains a binomial $f$ such that $\boldsymbol{x}^{\boldsymbol{u}}$ is a monomial of $f$.

We will write $M_J$ for the monomial ideal generated by all $\boldsymbol{x}^{\boldsymbol{u}}$ for which there exists a nonzero $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J$.

The next proposition is the natural generalization of [Charalambous et al. 2007, Proposition 3.1], but for completeness, we give a proof.

**Proposition 1.5.** *The indispensable monomials of $J$ are precisely the minimal generators of $M_J$.*

*Proof.* Let $\{f_1, \dots, f_s\}$ be a system of binomial generators of $J$. Clearly, the monomials of the $f_i$, $i = 1, \dots, s$, generate $M_J$. Let $\boldsymbol{x}^{\boldsymbol{u}}$ be a minimal generator of $M_J$. Then $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J$, for some nonzero $\boldsymbol{v} \in \mathbb{N}^n$. Now, the minimality of $\boldsymbol{x}^{\boldsymbol{u}}$ assures that $\boldsymbol{x}^{\boldsymbol{u}}$ is a monomial of $f_j$ for some $j$. Therefore every minimal generator of $M_J$ is an indispensable monomial of $J$. Conversely, let $\boldsymbol{x}^{\boldsymbol{u}}$ be an indispensable monomial of $J$. If $\boldsymbol{x}^{\boldsymbol{u}}$ is not a minimal generator of $M_J$, then there is a minimal generator $\boldsymbol{x}^{\boldsymbol{w}}$ of $M_J$ such that $\boldsymbol{x}^{\boldsymbol{u}} = \boldsymbol{x}^{\boldsymbol{w}}\boldsymbol{x}^{\boldsymbol{u}'}$ with $\boldsymbol{u}' \neq \boldsymbol{0}$. By the previous argument $\boldsymbol{x}^{\boldsymbol{w}}$ is an indispensable monomial of $J$, hence without loss of generality we may suppose that $f_k = \boldsymbol{x}^{\boldsymbol{w}} - \boldsymbol{x}^{\mathbf{z}}$ for some $k$ and $\mathbf{z} \in \mathbb{N}^n$. Thus, if $f_j = \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$, then

$$f_j' = \boldsymbol{x}^{\boldsymbol{u}'}\boldsymbol{x}^{\mathbf{z}} - \boldsymbol{x}^{\boldsymbol{v}} = f_j - \boldsymbol{x}^{\boldsymbol{u}'} f_k \in J$$

and therefore we can replace $f_j$ by $f_j'$ in $\{f_1, \dots, f_s\}$. Repeating this argument as many times as necessary, we will find a system of binomial generators of $J$ such that no element has $\boldsymbol{x}^{\boldsymbol{u}}$ as monomial, a contradiction to the fact that $\boldsymbol{x}^{\boldsymbol{u}}$ is indispensable. □

**Corollary 1.6.** *If $\boldsymbol{x}^{\boldsymbol{u}} \in M_J$ is an indispensable monomial of $I_{\mathcal{A}}$, then it is also an indispensable monomial of $J$.*

*Proof.* It suffices to note that $M_J \subseteq M_{I_{\mathcal{A}}}$ by Corollary 1.3. □

Now, we will give a combinatorial necessary and sufficient condition for a monomial $\boldsymbol{x}^{\boldsymbol{u}} \in \Bbbk[\boldsymbol{x}]$ to be indispensable in $J$.

**Definition 1.7.** Let $\boldsymbol{b} \in \mathbb{N}\mathcal{A}$. The graph $G_{\boldsymbol{b}}(J)$ has as its vertices the monomials of $M_J$ of $\mathcal{A}$-degree $\boldsymbol{b}$; two vertices $\boldsymbol{x}^{\boldsymbol{u}}$ and $\boldsymbol{x}^{\boldsymbol{v}}$ are joined by an edge if $\gcd(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{v}}) \neq 1$ and there exists a monomial $1 \neq \boldsymbol{x}^{\boldsymbol{w}}$ dividing $\gcd(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{v}})$ such that the binomial $\boldsymbol{x}^{\boldsymbol{u}-\boldsymbol{w}} - \boldsymbol{x}^{\boldsymbol{v}-\boldsymbol{w}}$ belongs to $J$.

Notice that $G_{\boldsymbol{b}}(J) = \varnothing$ exactly when $M_J$ has no element of $\mathscr{A}$-degree $\boldsymbol{b}$; in particular, $G_{\boldsymbol{b}}(J) = \varnothing$ if $\boldsymbol{b} = \boldsymbol{0}$, because $1 \notin M_J$ (otherwise, $\Bbbk[\boldsymbol{x}]/I_{\mathscr{A}}$ would contain invertible elements). Moreover, since $J \subseteq I_{\mathscr{A}}$, we have that $G_{\boldsymbol{b}}(J)$ is a subgraph of $G_{\boldsymbol{b}}(I_{\mathscr{A}})$, for all $\boldsymbol{b}$. Finally, we observe that the existence of $\boldsymbol{x}^{\boldsymbol{w}}$ as stated is trivially fulfilled for $J = I_{\mathscr{A}}$ because $\big(I_{\mathscr{A}} : (x_1 \cdots x_n)^{\infty}\big) = I_{\mathscr{A}}$, in this case, if $G_{\boldsymbol{b}}(J) \neq \varnothing$, the graph $G_{\boldsymbol{b}}(J)$ is nothing but the 1-skeleton of the simplicial complex $\nabla_{\boldsymbol{b}}$ appearing in [Ojeda and Vigneron-Tenorio 2010a]. Thus, we have the following result.

**Theorem 1.8.** *Let $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$ be a binomial of $\mathscr{A}$-degree $\boldsymbol{b}$. Then, $f$ is a minimal generator of $I_{\mathscr{A}}$ if and only if $\boldsymbol{x}^{\boldsymbol{u}}$ and $\boldsymbol{x}^{\boldsymbol{v}}$ lie in two different connected components of $G_{\boldsymbol{b}}(I_{\mathscr{A}})$, in particular, the graph is disconnected.*

*Proof.* See, for example, [Ojeda and Vigneron-Tenorio 2010b, Section 2]. □

The next theorem provides a necessary and sufficient condition for a monomial to be indispensable in $J$.

**Theorem 1.9.** *A monomial $\boldsymbol{x}^{\boldsymbol{u}}$ is indispensable in $J$ if and only if $\{\boldsymbol{x}^{\boldsymbol{u}}\}$ is connected component of $G_{\boldsymbol{b}}(J)$, where $\boldsymbol{b} = \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}})$.*

*Proof.* Suppose that $\boldsymbol{x}^{\boldsymbol{u}}$ is an indispensable monomial of $J$ and $\{\boldsymbol{x}^{\boldsymbol{u}}\}$ is not a connected component of $G_{\boldsymbol{b}}(J)$. Then, there exists $\boldsymbol{x}^{\boldsymbol{v}} \in M_J$ with $\mathscr{A}$-degree equal to $\boldsymbol{b}$ such that $\gcd(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{v}}) \neq 1$ and $\boldsymbol{x}^{\boldsymbol{u}-\boldsymbol{w}} - \boldsymbol{x}^{\boldsymbol{v}-\boldsymbol{w}} \in J$, where $1 \neq \boldsymbol{x}^{\boldsymbol{w}}$ divides $\gcd(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{v}})$. So $\boldsymbol{x}^{\boldsymbol{u}-\boldsymbol{w}} \in M_J$ and properly divides $\boldsymbol{x}^{\boldsymbol{u}}$, a contradiction to the fact that $\boldsymbol{x}^{\boldsymbol{u}}$ is a minimal generator of $M_J$ (see Proposition 1.5). Conversely, we assume that $\{\boldsymbol{x}^{\boldsymbol{u}}\}$ is connected component of $G_{\boldsymbol{b}}(J)$ with $\boldsymbol{b} = \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}})$ and that $\boldsymbol{x}^{\boldsymbol{u}}$ is not an indispensable monomial of $J$. Then, by Proposition 1.5, there exists a binomial $f = \boldsymbol{x}^{\boldsymbol{w}} - \boldsymbol{x}^{\boldsymbol{z}} \in J$, such that $\boldsymbol{x}^{\boldsymbol{w}}$ properly divides $\boldsymbol{x}^{\boldsymbol{u}}$. Let $\boldsymbol{x}^{\boldsymbol{u}} = \boldsymbol{x}^{\boldsymbol{w}}\boldsymbol{x}^{\boldsymbol{u}'}$, then $1 \neq \boldsymbol{x}^{\boldsymbol{u}'}$ divides $\gcd(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{u}'}\boldsymbol{x}^{\boldsymbol{z}})$ and hence $(\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{u}'}\boldsymbol{x}^{\boldsymbol{z}})/(\boldsymbol{x}^{\boldsymbol{u}'}) = f \in J$. Thus, $\{\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x}^{\boldsymbol{u}'}\boldsymbol{x}^{\boldsymbol{z}}\}$ is an edge of $G_{\boldsymbol{b}}(J)$, a contradiction to the fact that $\{\boldsymbol{x}^{\boldsymbol{u}}\}$ is a connected component of $G_{\boldsymbol{b}}(J)$. □

Now, we are able to give a sufficient condition for a binomial to be indispensable in $J$ by using our graphs $G_{\boldsymbol{b}}(J)$ (compare with [García and Ojeda 2010, Corollary 5]).

**Theorem 1.10.** *Given $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J$ and let $\boldsymbol{b} = \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}})$ $\big(= \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{v}})\big)$. If $G_{\boldsymbol{b}}(J) = \big\{\{\boldsymbol{x}^{\boldsymbol{u}}\}, \{\boldsymbol{x}^{\boldsymbol{v}}\}\big\}$, then $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$ is an indispensable binomial of $J$.*

*Proof.* Assume that $G_{\boldsymbol{b}}(J) = \big\{\{\boldsymbol{x}^{\boldsymbol{u}}\}, \{\boldsymbol{x}^{\boldsymbol{v}}\}\big\}$. Then, by Theorem 1.9, both $\boldsymbol{x}^{\boldsymbol{u}}$ and $\boldsymbol{x}^{\boldsymbol{v}}$ are indispensable monomials of $J$. Let $\{f_1, \ldots, f_s\}$ be a system of binomial generators of $J$. Since $\boldsymbol{x}^{\boldsymbol{u}}$ is an indispensable monomial, $f_i = \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{w}} \neq 0$, for some $i$. Thus $\deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}}) = \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{w}})$ and therefore $\boldsymbol{x}^{\boldsymbol{w}}$ is a vertex of $G_{\boldsymbol{b}}(J)$. Consequently, $\boldsymbol{w} = \boldsymbol{v}$ and we conclude that $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$ is an indispensable binomial of $J$. □

The converse of this theorem is not true in general: consider for instance the ideal $J = \langle x - y, y^2 - yt, z - t\rangle = \langle x - t, y - t, z - t\rangle \cap \langle x, y, z - t\rangle$, then $J$ is $\mathscr{A}$-homogeneous for $\mathscr{A} = \{1, 1, 1, 1\}$. Both $x - y$ and $z - t$ are indispensable binomials of $J$, while $G_{\mathbf{1}}(J) = \big\{\{x\}, \{y\}, \{z\}, \{t\}\big\}$.

**Corollary 1.11.** *If $f = \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in J$ is an indispensable binomial of $I_{\mathscr{A}}$, then $f$ is an indispensable binomial of $J$.*

*Proof.* Let $\boldsymbol{b} = \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{u}})\ \big(= \deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{v}})\big)$. By [Ojeda and Vigneron-Tenorio 2010a, Corollary 7], if $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$ is an indispensable binomial of $I_{\mathscr{A}}$, then $G_{\boldsymbol{b}}(I_{\mathscr{A}}) = \big\{\{\boldsymbol{x}^{\boldsymbol{u}}\}, \{\boldsymbol{x}^{\boldsymbol{v}}\}\big\}$. Since $\boldsymbol{x}^{\boldsymbol{u}}$ and $\boldsymbol{x}^{\boldsymbol{v}}$ are vertices of $G_{\boldsymbol{b}}(J)$ and $G_{\boldsymbol{b}}(J)$ is a subgraph of $G_{\boldsymbol{b}}(I_{\mathscr{A}})$, then $G_{\boldsymbol{b}}(J) = G_{\boldsymbol{b}}(I_{\mathscr{A}})$ and therefore, by Theorem 1.10, we conclude that $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}}$ is an indispensable binomial of $J$. $\square$

Again we have that the converse is not true; for instance, $x - y$ and $z - t$ are indispensable binomials of $J = \langle x - y, y^2 - yt, z - t\rangle$ and none of them is indispensable in the toric ideal $I_{\mathscr{A}}$.

We close this section by applying our results to show that the binomial edge ideals introduced in [Herzog et al. 2010] have unique minimal system of binomial generators.

Let $G$ be an undirected connected simple graph on the vertex set $\{1, \ldots, n\}$ and let $\Bbbk[\boldsymbol{x}, \boldsymbol{y}]$ be the polynomial ring in $2n$ variables, $x_1, \ldots, x_n, y_1, \ldots, y_n$, over $\Bbbk$.

**Definition 1.12.** The binomial edge ideal $J_G \subset \Bbbk[\boldsymbol{x}, \boldsymbol{y}]$ associated to $G$ is the ideal generated by the binomials $f_{ij} = x_i y_j - x_j y_i$, with $i < j$, such that $\{i, j\}$ is an edge of $G$.

Let $J_G \subset \Bbbk[\boldsymbol{x}, \boldsymbol{y}]$ be the binomial edge ideal associated to $G$. By definition, $J_G$ is contained in the determinantal ideal generated by the $2 \times 2$-minors of

$$\begin{pmatrix} x_1 & \ldots & x_n \\ y_1 & \ldots & y_n \end{pmatrix}.$$

This ideal is nothing but the toric ideal associated to the Lawrence lifting, $\Lambda(\mathscr{A})$, of $\mathscr{A} = \{1, \ldots, 1\}$ (see [Sturmfels 1996, Chapter 7], for instance). Thus, $J_G \subseteq I_{\Lambda(\mathscr{A})}$ and the equality holds if and only if $G$ is the complete graph on $n$ vertices. By the way, since $G$ is connected, the smallest toric ideal containing $J_G$ has codimension $n - 1$. So, the smallest toric ideal containing $J_G$ is $I_{\Lambda(\mathscr{A})}$, that is to say, $\Lambda(\mathscr{A})$ is the finest grading on $\Bbbk[\boldsymbol{x}, \boldsymbol{y}]$ such that $J_G$ is $\Lambda(\mathscr{A})$-homogeneous.

**Corollary 1.13.** *The binomial edge ideal $J_G$ has unique minimal system of binomial generators.*

*Proof.* By [Ojeda and Vigneron-Tenorio 2010a, Corollary 16], the toric ideal $I_{\Lambda(\mathscr{A})}$ is generated by its indispensable binomials, thus every $f_{ij} \in J_G$, is an indispensable binomial of $I_{\Lambda(\mathscr{A})}$. Now, by Corollary 1.11, we conclude that $J_G$ is generated by its indispensable binomials. □

The above result can be viewed as a particular case of the following general result whose proof is also straightforward consequence of [Ojeda and Vigneron-Tenorio 2010a, Corollary 16] and Corollary 1.11.

**Corollary 1.14.** *Let $\mathscr{A} = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\} \subseteq \mathbb{Z}^d$ be such that the monoid $\mathbb{N}\mathscr{A}$ is free of units. If $J \subseteq \Bbbk[\boldsymbol{x}, \boldsymbol{y}]$ is a binomial ideal generated by a subset of the minimal system of binomial generators of $I_{\Lambda(\mathscr{A})}$, then $J$ has unique minimal system of binomial generators.*

## 2. Critical binomials, circuits and primitive binomials

This section deals with binomial ideals contained in the defining ideal of a monomial curve. Special attention should be paid to the critical ideal; this is due to the fact that the ideal of a monomial space curve is equal to the critical ideal, see [Herzog 1970] (see also the definition of neat numerical semigroup in [Komeda 1982]). Throughout this section $\mathscr{A} = \{a_1, \dots, a_n\}$ is a set of relatively prime positive integers and $I_{\mathscr{A}} \subset \Bbbk[\boldsymbol{x}] = \Bbbk[x_1, \dots, x_n]$ is the defining ideal of the monomial curve $x_1 = t^{a_1}, \dots, x_n = t^{a_n}$ in the $n$-dimensional affine space over $\Bbbk$.

***Critical binomials.***

**Definition 2.1.** A binomial $x_i^{c_i} - \prod_{j \neq i} x_j^{u_{ij}} \in I_{\mathscr{A}}$ is called *critical* with respect to $x_i$ if $c_i$ is the least positive integer such that $c_i a_i \in \sum_{j \neq i} \mathbb{N} a_j$. The *critical ideal* of $\mathscr{A}$, denoted by $C_{\mathscr{A}}$, is the ideal of $\Bbbk[\boldsymbol{x}]$ generated by all the critical binomials of $I_{\mathscr{A}}$.

Observe that the critical ideal of $\mathscr{A}$ is $\mathscr{A}$-homogeneous.

**Notation 2.2.** From now on and for the rest of the paper, we will write $c_i$ for the least positive integer such that $c_i a_i \in \sum_{j \neq i} \mathbb{N} a_j$, for each $i = 1, \dots, n$.

**Proposition 2.3.** *The monomials $x_i^{c_i}$ are indispensable in $I_{\mathscr{A}}$, for every $i$. Equivalently, $\{x_i^{c_i}\}$ is a connected component of $G_b(I_{\mathscr{A}})$, where $b = c_i a_i$, for every $i$.*

*Proof.* The proof follows immediately from the minimality of $c_i$, Theorem 1.8 and Theorem 1.9. □

We now characterize the indispensable critical binomials of the toric ideal $I_{\mathscr{A}}$.

**Theorem 2.4.** *Let $f = x_i^{c_i} - \prod_{j \neq i} x_j^{u_{ij}}$ be a critical binomial of $I_{\mathscr{A}}$, then $f$ is indispensable in $I_{\mathscr{A}}$ if, and only if, $f$ is indispensable in $C_{\mathscr{A}}$.*

*Proof.* By Corollary 1.11, we have that if $f$ is indispensable in $I_{\mathscr{A}}$, then it is indispensable in $C_{\mathscr{A}}$. Conversely, assume that $f$ is indispensable in $C_{\mathscr{A}}$. Let $\{f_1, \dots, f_s\}$ be a system of binomial generators of $I_{\mathscr{A}}$ not containing $f$. Then, by Proposition 2.3, $f_l = x_i^{c_i} - \prod_{j \neq i} x_j^{v_j}$ for some $l$. So, $f_l$ is a critical binomial, that is to say, $f_l \in C_{\mathscr{A}}$. Therefore, we may replace $f$ by $f_l$ and $f - f_l \in C_{\mathscr{A}}$ in a system of binomial generators of $C_{\mathscr{A}}$, a contradiction to the fact that $f$ is indispensable in $C_{\mathscr{A}}$. □

**Corollary 2.5.** *If $I_{\mathscr{A}}$ has a unique minimal system of binomial generators, then $C_{\mathscr{A}}$ also does.*

*Proof.* The monomials $x_i^{c_i}$ are indispensable in $I_{\mathscr{A}}$, for each $i$ (see Proposition 2.3). Thus, for every $i$, there exists a unique binomial in $I_{\mathscr{A}}$ of the form $x_i^{c_i} - \prod_{j \neq i} x_j^{u_{ij}}$ and we conclude that $C_{\mathscr{A}}$ has unique minimal system of binomial generators. □

**Example 2.6.** Let $\mathscr{A} = \{4, 6, 2a+1, 2a+3\}$ where $a$ is a natural number. For $a = 0$, it is easy to see that $I_{\mathscr{A}}$ does not have a unique minimal system of binomial generators. If $a \geq 1$, then $x_4^2 - x_1^a x_2$ and $x_4^2 - x_1 x_3^2 \in C_{\mathscr{A}}$. Thus $C_{\mathscr{A}}$ is not generated by its indispensable binomials and therefore $I_{\mathscr{A}}$ does not have a unique minimal system of binomial generators.

### *Circuits.*

Recall that the support of a monomial $\boldsymbol{x}^{\boldsymbol{u}}$ is the set $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{u}}) = \{i \in \{1, \ldots, n\} \mid u_i \neq 0\}$. The support of a binomial $f = \boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$, denoted by $\operatorname{supp}(f)$, is defined as the union $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{u}}) \cup \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}})$. We say that $f$ has full support when $\operatorname{supp}(f) = \{1, \ldots, n\}$.

**Definition 2.7.** An irreducible binomial $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$ is called a *circuit* if its support is minimal with respect the inclusion.

Recall that a polynomial in $\Bbbk[\boldsymbol{x}]$ is said to be *irreducible* if it cannot be factored into the product of two (or more) non-trivial polynomials in $\Bbbk[\boldsymbol{x}]$.

**Lemma 2.8.** *Let $u_j(i) = \dfrac{a_i}{\gcd(a_i, a_j)}$, for $i \neq j$. The set of circuits in $I_{\mathscr{A}}$ is equal to*

$$\{x_i^{u_i(j)} - x_j^{u_j(i)} \mid i \neq j\}.$$

*Proof.* See [Sturmfels 1996, Chapter 4] □

The next theorem provides a class of toric ideals generated by critical binomials that, moreover, are circuits.

**Theorem 2.9.** *If $C_{\mathscr{A}} = \langle x_1^{c_1} - x_2^{c_2}, \ldots, x_{n-1}^{c_{n-1}} - x_n^{c_n} \rangle$, then $C_{\mathscr{A}} = I_{\mathscr{A}}$.*

*Proof.* From the hypothesis the binomial $x_i^{c_i} - x_{i+1}^{c_{i+1}}$ belongs to $I_{\mathscr{A}}$, for each $i \in \{1, \ldots, n-1\}$. So, every circuit of $I_{\mathscr{A}}$ is of the form $x_k^{c_k} - x_l^{c_l}$, since $\gcd(c_k, c_l) = 1$. Now, from Proposition 2.2 in [Alcántar and Villarreal 1994], the lattice $L = \ker_{\mathbb{Z}}(\mathscr{A}) = \{\boldsymbol{u} \in \mathbb{Z}^n \mid u_1 a_1 + \ldots + u_n a_n = 0\}$ is generated by $\{c_i \boldsymbol{e}_i - c_j \boldsymbol{e}_j \mid 1 \leq i \leq j \leq n\}$, where $\boldsymbol{e}_i$ is the vector with 1 in the $i$-th position and zeros elsewhere. The rank of $L$ equals $n-1$ and a lattice basis is $\{\boldsymbol{v}_i = c_i \boldsymbol{e}_i - c_{i+1} \boldsymbol{e}_{i+1} \mid 1 \leq i \leq n-1\}$. Thus $C_{\mathscr{A}}$ is a lattice basis ideal. Let $M$ be the matrix with rows $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{n-1}$, then $M$ is a mixed dominating matrix and therefore, from [Fischer and Shapiro 1996, Theorem 2.9], the equality $C_{\mathscr{A}} = I_{\mathscr{A}}$ holds. □

**Remarks 2.10.**

(1) For $n = 4$, a different proof of the above result can be found in [Bresinsky 1975].

(2) The converse of Theorem 2.9 is not true in general (see [Alcántar and Villarreal 1994], for instance).

(3) If every critical binomial of $I_{\mathscr{A}}$ is a circuit and the critical ideal has codimension $n-1$, then $c_i a_i = c_j a_j$, for every $i \neq j$. In particular, all minimal generators of $I_{\mathscr{A}}$ have the same $\mathscr{A}$-degree. This situation is explored in some detail in [García Sánchez et al. 2013] from a semigroup viewpoint.

The rest of this subsection is devoted to the investigation of necessary and sufficient conditions for a circuit to be indispensable in $I_{\mathscr{A}}$.

**Lemma 2.11.** *Let $f = x_i^{u_i(j)} - x_j^{u_j(i)} \in I_{\mathscr{A}}$ be a circuit and let $b = u_i(j)a_i$. Then there is no monomial $\boldsymbol{x}^{\boldsymbol{v}}$ in the fiber $\deg_{\mathscr{A}}^{-1}(b)$ such that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) = \{i, j\}$.*

*Proof.* Suppose to the contrary that there exists such a $\boldsymbol{v}$. Observe that $x_i^{u_i(j)} - x_j^{u_j(i)}$ is also a circuit of $I_{\{a_i/d, a_j/d\}}$, and $\boldsymbol{v} \in \deg_{\{a_i/d, a_j/d\}}^{-1}(b/d)$, with $d = \gcd(a_i, a_j)$. But $\deg_{\{a_i/d, a_j/d\}}^{-1}(b/d) = \left\{x_i^{u_i(j)}, x_j^{u_j(i)}\right\}$; see, for instance, [Rosales and García 2009, Example 8.22]. □

**Theorem 2.12.** *Let $f = x_i^{u_i(j)} - x_j^{u_j(i)} \in I_{\mathscr{A}}$ be a circuit and let $b = u_i(j)a_i$. Then, $f$ is indispensable in $I_{\mathscr{A}}$ if, and only if, $b - a_k \notin \mathbb{N}\mathscr{A}$, for every $k \neq i, j$. In particular, $u_i(j) = c_i$ and $u_j(i) = c_j$.*

*Proof.* First of all, we observe that $\deg_{\mathscr{A}}^{-1}(b) \supseteq \left\{x_i^{u_i(j)}, x_j^{u_j(i)}\right\}$ and equality holds if and only if $f$ is indispensable. So, the sufficiency condition follows. Conversely, since $b \notin \sum_{k \neq i,j} \mathbb{N}a_k$, the supports of the monomials in $\deg_{\mathscr{A}}^{-1}(b)$ are included in $\{i, j\}$ and then, by Lemma 2.11, we are done. □

From this result it follows that *if a circuit is indispensable, then it is a critical binomial.*

Let $\prec_{ij}$ be an $\mathscr{A}$-graded reverse lexicographical monomial order on $\Bbbk[\boldsymbol{x}]$ such that $x_k \prec_{ij} x_i$ and $x_k \prec_{ij} x_j$ for every $k \neq i, j$.

**Proposition 2.13.** *A circuit $f = x_i^{u_i(j)} - x_j^{u_j(i)} \in I_{\mathscr{A}}$ is indispensable in $I_{\mathscr{A}}$ if and only if it belongs to the reduced Gröbner basis of $I_{\mathscr{A}}$ with respect to $\prec_{ij}$.*

*Proof.* If $f$ is indispensable, then, by Theorem 13 of [Ojeda and Vigneron-Tenorio 2010a], it belongs to every Gröbner basis of $I_{\mathscr{A}}$. Now, suppose that $f$ belongs to the reduced Gröbner basis of $I_{\mathscr{A}}$ with respect to $\prec_{ij}$ and it is not indispensable. Since

$f$ is not indispensable, there exists a monomial $\boldsymbol{x}^{\boldsymbol{u}}$ in the fiber of $u_i(j)a_i$ different from $x_i^{u_i(j)}$ and $x_j^{u_j(i)}$. By Lemma 2.11, we have that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{u}}) \not\subset \{i, j\}$, so there is $k \in \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{u}})$ and $k \notin \{i, j\}$. Hence, both $f_i = x_i^{u_i(j)} - \boldsymbol{x}^{\boldsymbol{u}}$ and $f_j = x_j^{u_j(i)} - \boldsymbol{x}^{\boldsymbol{u}}$ belong to $I_{\mathscr{A}}$. Since the leading terms of $f_i$ and $f_j$ with respect to $\prec_{ij}$ equal to $x_i^{u_i(j)}$ and $x_j^{u_j(i)}$, respectively, we conclude that $f = x_i^{u_i(j)} - x_j^{u_j(i)} \in I_{\mathscr{A}}$ is not in the reduced Gröbner basis of $I_{\mathscr{A}}$ with respect to $\prec_{ij}$, a contradiction. □

***Primitive binomials.***

**Definition 2.14.** A binomial $\boldsymbol{x}^{\boldsymbol{u}} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$ is called *primitive* if there exists no other binomial $\boldsymbol{x}^{\boldsymbol{u}'} - \boldsymbol{x}^{\boldsymbol{v}'}$ such that $\boldsymbol{x}^{\boldsymbol{u}'}$ divides $\boldsymbol{x}^{\boldsymbol{u}}$ and $\boldsymbol{x}^{\boldsymbol{v}'}$ divides $\boldsymbol{x}^{\boldsymbol{v}}$. The set of all primitive binomials is called the Graver basis of $\mathscr{A}$ and it is denoted by $\operatorname{Gr}(\mathscr{A})$.

**Theorem 2.15.** *Let $f = x_i^{u_i} x_j^{u_j} - x_k^{u_k} x_l^{u_l} \in \operatorname{Gr}(\mathscr{A})$ be such that $u_i < c_i$, $u_j < c_j$, $u_k < c_k$ and $u_l < c_l$ with $i, j, k$ and $l$ pairwise different. Then $f$ is indispensable in $J = I_{\mathscr{A}} \cap \Bbbk[x_i, x_j, x_k, x_l]$.*

*Proof.* By [Sturmfels 1996, Proposition 4.13(a)], $J = I_{\mathscr{A}} \cap \Bbbk[x_i, x_j, x_k, x_l]$ is the toric ideal associated to $\mathscr{A}' = \{a_i, a_j, a_k, a_l\}$. Thus, without loss of generality we may assume $n = 4$, then $J = I_{\mathscr{A}}$. We prove that $G_b(I_{\mathscr{A}}) = \{x_i^{u_i} x_j^{u_j}, x_k^{u_k} x_l^{u_l}\}$, where $b = u_i a_i + u_j a_j$. Let $\boldsymbol{x}^{\boldsymbol{v}} \in \deg_{\mathscr{A}}^{-1}(b)$ be different from $x_i^{u_i} x_j^{u_j}$ and $x_k^{u_l} x_l^{u_l}$. If $u_i < v_i$, then $x_i^{u_i}(x_j^{u_j} - x_i^{v_i - u_i} x_j^{v_j} x_k^{v_k} x_l^{v_l}) \in I_{\mathscr{A}}$, thus $x_j^{u_j} - x_i^{v_i - u_i} x_j^{v_j} x_k^{v_k} x_l^{v_l} \in I_{\mathscr{A}}$ which is impossible by the minimality of $c_j$ (see Proposition 2.3). Analogously, we can prove that $u_j \geq v_j$, $u_k \geq v_k$ and $u_l \geq v_l$. Therefore $x_i^{v_i} x_j^{v_j}(x_i^{u_i - v_i} x_j^{u_j - v_j} - x_k^{v_k} x_l^{v_l}) \in I_{\mathscr{A}}$ and so $x_i^{u_i - v_i} x_j^{u_j - v_j} - x_k^{v_k} x_l^{v_l} \in I_{\mathscr{A}}$, a contradiction with the fact that $f$ is primitive. This shows that $G_b(J) = \big\{\{x_i^{u_i} x_j^{u_j}\}, \{x_k^{u_k} x_l^{u_l}\}\big\}$ and, by Theorem 1.10, we are done. □

**Corollary 2.16.** *Let $f = x_i^{u_i} x_j^{u_j} - x_k^{u_k} x_l^{u_l} \in I_{\mathscr{A}}$ be such that $u_i < c_i$, $u_j < c_j$, $u_k > 0$ and $u_l > 0$ with $i, j, k$ and $l$ pairwise different. If $x_k^{u_k} x_l^{u_l}$ is indispensable in $J = I_{\mathscr{A}} \cap \Bbbk[x_i, x_j, x_k, x_l]$, then $f$ is indispensable in $J$.*

*Proof.* Since, by Theorem 1.9, $\{x_k^{u_k} x_l^{u_l}\}$ is a connected component of $G_b(I_{\mathscr{A}})$, where $b = u_k a_k + u_l a_l$, the monomial $\boldsymbol{x}^{\boldsymbol{v}} \in \deg_{\mathscr{A}}^{-1}(b)$ in the above proof has its support in $\{i, j\}$. Thus, repeating the arguments of the proof of Theorem 2.15, we deduce that $u_i \geq v_i$ and $u_j \geq v_j$. But $x_i^{u_i} x_j^{u_j} - x_i^{v_i} x_j^{v_j} \in I_{\mathscr{A}}$, so $u_i a_i + u_j a_j = v_i a_i + v_j a_j$ which implies that $u_i = v_i$ and $u_j = v_j$. By Theorem 1.10 we have that $f$ is indispensable in $J$. □

Combining Theorem 2.15 with Corollary 1.11 we get:

**Corollary 2.17.** *Given $i, j, k$ and $l \in \{1, \ldots, n\}$ pairwise different, let $J$ be the ideal of $\Bbbk[x_i, x_j, x_k, x_l]$ generated by all Graver binomials of $I_{\mathscr{A}}$ of the form $x_i^{u_i} x_j^{u_j} - x_k^{u_k} x_l^{u_l}$ with $u_i < c_i$, $u_j < c_j$, $u_k < c_k$ and $u_l < c_l$. Then $J$ has unique minimal system of binomial generators.*

Finally we provide another class of primitive binomials that are indispensable in a toric ideal.

**Corollary 2.18.** *Let* $f = x_i^{u_i} x_j^{u_j} - x_k^{u_k} x_l^{u_l} \in \mathrm{Gr}(\mathscr{A})$ *such that* $0 < u_i < c_i$ *and* $0 < u_k < c_k$, *for* $i, j, k$ *and* $l$ *pairwise different. If* $u_i a_i + u_j a_j$ *is minimal among all Graver* $\mathscr{A}$*-degrees, then* $f$ *is indispensable in* $I_{\mathscr{A}} \cap \Bbbk[x_i, x_j, x_l, x_k]$.

*Proof.* Since $c_j a_j$ is a Graver $\mathscr{A}$-degree, we have $u_i a_i + u_j a_j \le c_j a_j$, so it follows $u_j < c_j$. Similarly, we can prove $u_l < c_l$. Therefore, by Theorem 2.15, we conclude that $f$ is indispensable in $I_{\mathscr{A}} \cap \Bbbk[x_i, x_j, x_l, x_k]$. □

It is worth to noting here that [García Sánchez et al. 2013, Theorem 6] offers a characterization of the family of affine semigroups for which $C_{\mathscr{A}} = \mathrm{Gr}(\mathscr{A})$.

## 3. Classification of monomial curves in $\mathbb{A}^4(\Bbbk)$

Let $\mathscr{A} = \{a_1, a_2, a_3, a_4\}$ be a set of relatively prime positive integers. First we will provide a minimal system of binomial generators for the critical ideal $C_{\mathscr{A}}$. This will be done by comparing the $\mathscr{A}$-degrees of the monomials $x_i^{c_i}$, for $i = 1, \ldots, 4$.

**Lemma 3.1.** *Let* $f_i = x_i^{c_i} - \prod_{j \neq i} x_j^{u_{ij}}$, $i = 1, \ldots, 4$, *be a set of critical binomials of* $I_{\mathscr{A}}$ *and let* $g_l \in I_{\mathscr{A}}$ *be a critical binomial with respect to* $x_l$, *for some* $l \in \{1, \ldots, 4\}$. *If* $f_l \neq -f_i$ *for every* $i$, *then* $g_l \in \langle f_1, f_2, f_3, f_4 \rangle$.

*Proof.* For simplicity we assume $l = 1$. Let $g_1 = x_1^{c_1} - x_2^{v_2} x_3^{v_3} x_4^{v_4} \in I_{\mathscr{A}}$ be a critical binomial. If $g_1 = f_1$, there is nothing to prove. If $g_1 \neq f_1$, without loss of generality we may assume that $u_{12} > v_2$, $u_{13} \le v_3$ and $u_{14} \le v_4$, so $g_1 - f_1 = m_1 g_2$, with $m_1 = x_2^{v_2} x_3^{u_{13}} x_4^{u_{14}}$ and $g_2 = x_2^{u_{12}-v_2} - x_3^{v_3-u_{13}} x_4^{v_4-u_{14}} \in I_{\mathscr{A}}$ (in particular $u_{12} - v_2 \ge c_2$). But $x_1^{c_1} - x_1^{u_{21}} x_2^{u_{12}-c_2} x_3^{u_{13}+u_{23}} x_4^{u_{14}+u_{24}} \in I_{\mathscr{A}}$ and also $f_1 \neq -f_2$, thus from the minimality of $c_1$ it follows that $u_{21} = 0$, that is to say, $f_2 \in \Bbbk[x_2, x_3, x_4]$. For the sake of simplicity, write $g_2 = x_2^b - x_3^c x_4^d$ with $b, c, d \in \mathbb{N}$ and $b \ge c_2$. Hence $g_2 - x_2^{b-c_2} f_2 = x_2^{b-c_2} x_3^{u_{23}} x_4^{u_{24}} - x_3^c x_4^d$. If $b - c_2 \ge c_2$, we repeat the process. After a finite number of steps, $g_2 - h_2 f_2 = x_2^{b-kc_2} x_3^{ku_{23}} x_4^{ku_{24}} - x_3^c x_4^d$ with $0 \le b - kc_2 < c_2$ and $h_2 \in \Bbbk[x_2, x_3, x_4]$. Then $(b - kc_2)a_2 + ku_{23}a_3 + ku_{24}a_4 = ca_3 + da_4$. Since $0 \le b - kc_2 < c_2$ then $x_3^{ku_{23}} x_4^{ku_{24}}$ does not divide $x_3^c x_4^d$. The case $x_3^c x_4^d$ divides $x_3^{ku_{23}} x_4^{ku_{24}}$ leads to $b = kc_2$, $c = ku_{23}$ and $d = ku_{24}$. In this setting, $g_2 = h_2 f_2$, $g_1 = f_1 + m_1 h_2 f_2$ and we are done. The remaining cases are $ku_{23} \ge c$ and $d \ge ku_{24}$, or $ku_{23} \le c$ and $d \ge ku_{24}$. Without loss of generality (by swapping variables if necessary), we may assume that $ku_{23} \le c$ and $d \le ku_{24}$. Hence $(b - kc_2)a_2 + (ku_{24} - d)a_4 = (c - ku_{23})a_3$, and consequently $c - ku_{23} \ge c_3$. We also deduce that $g_2 - h_2 f_2 = x_3^{ku_{23}} x_4^d (x_2^{b-kc_2} x_4^{ku_{24}-d} - x_3^{c-ku_{23}})$. Set $m_3 = x_3^{ku_{23}} x_4^d$ and $g_3 = x_2^{b-kc_2} x_4^{ku_{24}-d} - x_3^{c-ku_{23}}$. Since $v_3 - u_{13} - ku_{23} = c - ku_{23} \ge c_3$, we have that $v_2 \ge c_3$. Thus $x_1^{c_1} - x_1^{u_{31}} x_2^{u_{32}+v_2} x_3^{v_3-c_3} x_4^{u_{34}+v_4} \in I_{\mathscr{A}}$ and $f_1 \neq -f_3$, from the minimality of $c_1$ it follows that $u_{31} = 0$, that is to say, $f_3 \in \Bbbk[x_2, x_3, x_4]$. Analogously, by using a similar argument as before (and by swapping variables $x_2$ and $x_4$, if necessary), we obtain $h_3 \in \Bbbk[x_2, x_3, x_4]$ such that either $g_3 = h_3 f_3$ or $g_3 - h_3 f_3 = m_3 g_4$, with $m_3 = -x_2^{v'_2} x_4^{v''_4}$, $g_4 = x_4^{v'_4 - v_4 + u_{14} - v''_4} - x_2^{v''_2 - v'_2} x_3^{v''_3}$ and $v''_3 < c_3$. If $g_3 = h_3 f_3$, then $g_1 = f_1 + m_1 h_2 f_2 + m_1 m_2 h_3 f_3$ and we are done. Otherwise, since $x_1^{c_1} - x_1^{u_{41}} x_2^{v'_2 + v_2 + u_{42}} x_3^{v'_3 + u_{13} + u_{43}} x_4^{v'_4 + u_{14} - c_4} \in I_{\mathcal{A}}$ and $f_1 \neq -f_4$, the minimality of $c_1$ implies that $u_{41} = 0$, that is to say, $f_4 \in \Bbbk[x_2, x_3, x_4]$. Therefore, we have $f_2, f_3, f_4 \in \Bbbk[x_2, x_3, x_4]$. Taking into account that $I_{\mathcal{A}} \cap \Bbbk[x_2, x_3, x_4]$ is generated by $f_2$, $f_3$ and $f_4$ (see [Sturmfels 1996, Proposition 4.13(a)] and [Ojeda and Pisón Casares 2004, Theorem 2.2], for instance), we conclude that $g_2 = g_{21} f_2 + g_{23} f_3 + g_{24} f_4$ and hence $g_1 = f_1 + m_1 g_{21} f_2 + m_1 g_{23} f_3 + m_1 g_{24} f_4$, with $g_{2j} \in \Bbbk[x_2, x_3, x_4]$, $j = 1, 3, 4$. □

**Proposition 3.2.** *Let* $f_i = x_i^{c_i} - \prod_{j \neq i} x_j^{u_{ij}}$, $i = 1, \ldots, 4$, *be a set of critical binomials. If* $f_i \neq -f_j$ *for every* $i \neq j$, *then* $C_{\mathcal{A}} = \langle f_1, f_2, f_3, f_4 \rangle$.

*Proof.* The proof follows directly from Lemma 3.1. □

Observe that $f_i = -f_j$ if and only if $f_i = x_i^{c_i} - x_j^{c_j}$ and $f_j = x_j^{c_j} - x_i^{c_i}$; in particular, $f_i$ and $f_j$ are circuits. The following proposition provides an upper bound for the minimal number of generators of the critical ideal.

**Proposition 3.3.** *The minimal number of generators* $\mu(C_{\mathcal{A}})$ *of* $C_{\mathcal{A}}$ *is less than or equal to four.*

*Proof.* Let $\mathcal{F} = \{f_1, \ldots, f_4\} \subset I_{\mathcal{A}}$ be such that $f_i$ is critical with respect to $x_i$. If $f_i \neq -f_j$, for every $i \neq j$, then we are done by Proposition 3.2. Otherwise, without loss of generality we may assume $f_1 = -f_2$, that is to say, $f_1 = x_1^{c_1} - x_2^{c_2}$. Suppose that $\mathcal{F}$ is not a generating set of $C_{\mathcal{A}}$. We distinguish the following cases:

(1) $f_1$ is indispensable in $I_{\mathcal{A}}$. Then there exists a critical binomial $g \in I_{\mathcal{A}}$ with respect to at least one of the variables $x_3$ and $x_4$, say $x_4$, such that $g \neq \pm f_i$, for every $i$. By substitution of $f_4$ with $g$ in $\mathcal{F}$ we have, from Lemma 3.1, that every critical binomial with respect to $x_3$ or $x_4$ is in the ideal generated by the binomials of $\mathcal{F}$. Consequently the new set $\mathcal{F}$ generates $I_{\mathcal{A}}$.

(2) $f_1$ is not indispensable in $I_{\mathcal{A}}$. Then there exists a critical binomial $g \in I_{\mathcal{A}}$ with respect to al least one of the variables $x_1$ and $x_2$, for instance $x_2$, such that $g \neq \pm f_i$, for every $i$. We substitute $f_2$ with $g$ in $\mathcal{F}$. If $f_3 \neq -f_4$, then we have, from Proposition 3.2, that the new set $\mathcal{F}$ generates $I_{\mathcal{A}}$. Otherwise, we substitute $f_3$ with a critical binomial $h$ with respect to $x_3$ in $\mathcal{F}$ such that $h \neq \pm f_i$, for every $i$, when $f_3$ is not indispensable. So, in this case, $C_{\mathcal{A}}$ is generated by a set of four critical binomials. □

**Lemma 3.4.** *If $c_i a_i \neq c_k a_k$ and $c_i a_i \neq c_l a_l$, where $k \neq l$, then either the only critical binomial of $I_{\mathscr{A}}$ with respect to $x_i$ is $f = x_i^{c_i} - x_j^{c_j}$ or there exists a critical binomial $f \in I_{\mathscr{A}}$ with respect to $x_i$ such that* $\operatorname{supp}(f)$ *has cardinality greater than or equal to three, where $\{i, j, k, l\} = \{1, 2, 3, 4\}$.*

*Proof.* Suppose the contrary and let $f_i = x_i^{c_i} - x_j^{u_j} \in I_{\mathscr{A}}$ where $u_j > c_j$. We define $f = x_i^{c_i} - x_i^{v_i} x_j^{u_j - c_j} x_k^{v_k} x_l^{v_l} = f_i + x_j^{u_j - c_j} f_j \in I_{\mathscr{A}}$ with $f_j = x_j^{c_j} - x_i^{v_i} x_k^{v_k} x_l^{v_l} \in I_{\mathscr{A}}$. Now, from the minimality of $c_i$ it follows that $v_i = 0$, thus at least one of $v_k$ or $v_l$ is different from zero since $f_j \in I_{\mathscr{A}}$, otherwise $f - f_i = x_j^{u_j} - x_j^{u_j - c_j} \in I_{\mathscr{A}}$, and this is impossible. Therefore we conclude that $\operatorname{supp}(f)$ has cardinality greater than or equal to 3, a contradiction. The cases $f_i = x_i^{c_i} - x_k^{u_k} \in I_{\mathscr{A}}$ and $f_i = x_i^{c_i} - x_l^{u_l} \in I_{\mathscr{A}}$ are analogous, by using that $c_i a_i \neq c_k a_k$ and $c_i a_i \neq c_l a_l$, respectively. □

**Lemma 3.5.** *There is no minimal generating set of $C_{\mathscr{A}}$ of the form $\mathscr{S} = \{x_i^{c_i} - x_j^{c_j}, x_j^{c_j} - \boldsymbol{x}^{\boldsymbol{u}_j}, x_k^{c_k} - x_l^{c_l}, x_l^{c_l} - \boldsymbol{x}^{\boldsymbol{u}_l}\}$, where $\{i, j, k, l\} = \{1, 2, 3, 4\}$. In particular, if $c_i a_i = c_j a_j$ and $c_k a_k = c_l a_l$, then $\mu(C_{\mathscr{A}}) < 4$.*

*Proof.* Set $\boldsymbol{u}_j = (u_{j1}, \dots, u_{j4})$ and $\boldsymbol{u}_l = (u_{l1}, \dots, u_{l4})$. The minimality of $c_i$, $i \in \{1, 2, 3, 4\}$, forces $u_{ji} = 0 = u_{jj}$, $0 < u_{jk} < c_k$, $0 < u_{jl} < c_l$, $0 < u_{li} < c_i$, $0 < u_{lj} < c_j$, $u_{lk} = 0 = u_{ll}$.

Set $d_n = \gcd(\mathscr{A} \setminus \{a_n\})$, $n \in \{1, 2, 3, 4\}$. By [Herzog 1970, Theorem 3.10], the numerical semigroup generated by $\{a_i/d_l, a_j/d_l, a_k/d_l\}$ is symmetric and, from the proof of [Theorem 10.6,23], it is derived that $a_i/d_l = c_j c_k$, $a_j/d_l = c_i c_k$, $c_k = gcd(a_i/d_l, a_j/d_l)$ and $c_k a_k/d_l = u_{li} a_i/d_l + u_{lj} a_j/d_l$. Hence $a_i = c_j c_k d_l$, $a_j = c_i c_k d_l$ and $a_k = (u_{li} c_j + u_{lj} c_i) d_l$. Arguing analogously with $\{a_i/d_k, a_j/d_k, a_l/d_k\}$, we get $a_i = c_j c_l d_k$, $a_j = c_i c_l d_k$ and $a_l = (u_{li} c_j + u_{lj} c_i) d_k$. Thus, since $\gcd(c_i, c_j) = \gcd(c_k, c_l) = 1$, we conclude that $d_k = c_k$ and $d_l = c_l$. By considering now the symmetric semigroups $\{a_i/d_j, a_k/d_j, a_l/d_j\}$ and $\{a_j/d_i, a_k/d_i, a_l/d_i\}$, we get $a_i = (u_{jk} c_l + u_{jl} c_k) c_j$, $a_j = (u_{jk} c_l + u_{jl} c_k) c_i$, $a_k = c_i c_j c_l$ and $a_l = c_i c_j c_k$.

Putting all this together, we obtain that $u_{jk} c_l + u_{jl} c_k = c_l c_k$ which forces either $u_{jk} = 0$ or $u_{jk} \geq c_k$, and this is a contradiction in both cases. □

**Theorem 3.6.** *After permuting variables, if necessary, there exists a minimal system of binomial generators $\mathscr{S}$ of $C_{\mathscr{A}}$ of the following form:*

Case 1: *If $c_i a_i \neq c_j a_j$, for every $i \neq j$, then $\mathscr{S} = \{x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{u}_i},\ i = 1, \dots, 4\}$.*

Case 2: *If $c_1 a_1 = c_2 a_2$ and $c_3 a_3 = c_4 a_4$, then either $c_2 a_2 \neq c_3 a_3$ and*

(a) $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_3^{c_3} - x_4^{c_4}, x_4^{c_4} - \boldsymbol{x}^{\boldsymbol{u}_4}\}$ *when* $\mu(C_{\mathscr{A}}) = 3$,

(b) $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_3^{c_3} - x_4^{c_4}\}$ *when* $\mu(C_{\mathscr{A}}) = 2$,

*or $c_2 a_2 = c_3 a_3$ and*

(c) $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_2^{c_2} - x_3^{c_3}, x_3^{c_3} - x_4^{c_4}\}$.

Case 3: *If $c_1 a_1 = c_2 a_2 = c_3 a_3 \neq c_4 a_4$, then $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_2^{c_2} - x_3^{c_3}, x_4^{c_4} - \boldsymbol{x}^{\boldsymbol{u}_4}\}$.*

Case 4: *If* $c_1a_1 = c_2a_2$ *and* $c_ia_i \neq c_ja_j$ *for all* $\{i, j\} \neq \{1, 2\}$, *then*

(a) $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{u}_i} \mid i = 2, 3, 4\}$ *when* $\mu(C_{\mathscr{A}}) = 4$,

(b) $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{u}_i} \mid i = 3, 4\}$ *when* $\mu(C_{\mathscr{A}}) = 3$

*where, in each case,* $\boldsymbol{x}^{\boldsymbol{u}_i}$ *denotes an appropriate monomial whose support has cardinality greater than or equal to two.*

*Proof.* First, we observe that our assumption on the cardinality of $\boldsymbol{x}^{\boldsymbol{u}_i}$ follows from Lemma 3.4. We also notice that $C_{\mathscr{A}}$ has no minimal generating set of the form $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_2^{c_2} - \boldsymbol{x}^{u_2}, x_3^{c_3} - x_4^{c_4}, x_4^{c_4} - \boldsymbol{x}^{u_4}\}$, by Lemma 3.5.

Let $J$ be the ideal generated by $\mathscr{S}$. For the cases 1, 2(a-c), 3 and 4(a), it easily follows that $J = C_{\mathscr{A}}$ by Proposition 3.2. Indeed, in order to satisfy the hypothesis of Proposition 3.2, we may take $f_4 = x_4^{c_4} - x_1^{c_1} \in J$ and $f_3 = x_3^{c_3} - x_1^{c_1} \in J$ in the cases 2(c) and 3, respectively. The cases 2(a) and 4(b) happen when the only critical binomials of $I_{\mathscr{A}}$ with respect to $x_1$ and $x_2$ are $f_1 = x_1^{c_1} - x_2^{c_2}$ and $f_2 = -f_1$, respectively, then our claim follows from Lemma 3.1. Furthermore, the case 2(b) occurs when the only critical binomials of $I_{\mathscr{A}}$ are $\pm(x_1^{c_1} - x_2^{c_2})$ and $\pm(x_3^{c_3} - x_4^{c_4})$, so $J = C_{\mathscr{A}}$ by definition. On the other hand, since $x_i^{c_i}$ is an indispensable monomial of $I_{\mathscr{A}}$, for every $i$, by Corollary 1.6, we have that $x_i^{c_i}$ is an indispensable monomial of the ideal $J$, for every $i$. Then, we conclude that $\mathscr{S}$ is minimal in the sense that no proper subset of $\mathscr{S}$ generates $J$. □

**Example 3.7.** This example illustrates all possible cases of Theorem 3.6.

| | |
|---|---|
| Case 1: | $\mathscr{A} = \{17, 19, 21, 25\}$. |
| Case 2(a): | $\mathscr{A} = \{30, 34, 42, 51\}$. |
| Case 2(b): | $\mathscr{A} = \{39, 91, 100, 350\}$. |
| Case 2(c): | $\mathscr{A} = \{60, 132, 165, 220\}$. |
| Case 3: | $\mathscr{A} = \{12, 19, 20, 30\}$. |
| Case 4(a): | $\mathscr{A} = \{12, 13, 17, 20\}$. |
| Case 4(b): | $\mathscr{A} = \{4, 6, 11, 13\}$. |

The reader may perform the computations in detail by using the GAP package NumericalSgps ([Delgado et al. 2013]).

Since $C_{\mathscr{A}} \subseteq I_{\mathscr{A}}$, any minimal system of generators of $I_{\mathscr{A}}$ can not contain more than 4 critical binomials. This provides an affirmative answer to the question after Corollary 2 in [Bresinsky 1988]. Notice that the only cases in which $C_{\mathscr{A}}$ can have a unique minimal system of generators are 1, 2(b) and 4(b); in these cases $C_{\mathscr{A}}$ has a unique minimal system of binomial generators if and only if the monomials $\boldsymbol{x}^{\boldsymbol{u}_i}$ are indispensable.

Now we focus our attention on finding a minimal set of binomial generators of $I_{\mathscr{A}}$, that will help us to solve the classification problem. The following lemma will be useful in the proof of Proposition 3.9 and Theorem 3.10.

**Lemma 3.8.** (i) *If $f = x_i^{u_i} - \boldsymbol{x}^{\boldsymbol{v}}$ is a minimal generator of $I_{\mathscr{A}}$ that is not critical, then there exists $j \neq i$ such that* $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \{i, j\} = \varnothing$ *and $c_i a_i = c_j a_j$. Moreover, if $\boldsymbol{x}^{\boldsymbol{v}}$ is not indispensable, then $c_k a_k = c_l a_l$, with $\{i, j, k, l\} = \{1, 2, 3, 4\}$.*

(ii) *If $f = x_i^{u_i} x_j^{u_j} - \boldsymbol{x}^{\boldsymbol{v}}$ is a minimal generator of $I_{\mathscr{A}}$ with $u_i \neq 0$ and $u_j \geq c_j$, then* $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \{i, j\} = \varnothing$ *and $c_i a_i = c_j a_j$. In addition, if $\boldsymbol{x}^{\boldsymbol{v}}$ is not indispensable, then $c_k a_k = c_l a_l$, with $\{i, j, k, l\} = \{1, 2, 3, 4\}$.*

*Proof.* (i) Let $b = c_i a_i$. Since $f$ is not a critical binomial, we have that $u_i > c_i$. If $c_i a_i \neq c_j a_j$, for every $j \neq i$, then, from Lemma 3.4, there exists a critical binomial $f = x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{w}} \in I_{\mathscr{A}}$ such that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}})$ has cardinality greater than or equal to two. If $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) \neq \varnothing$, then $x_i^{u_i} \leftrightarrow x_i^{u_i - c_i} \boldsymbol{x}^{\boldsymbol{w}} \leftrightarrow \boldsymbol{x}^{\boldsymbol{v}}$ is a path in $G_b(I_{\mathscr{A}})$, a contradiction to the fact that $f$ is a minimal generator by Theorem 1.8. Hence $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) = \varnothing$. We have that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}+\boldsymbol{w}}) \subseteq \{j, k, l\}$, $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) = \varnothing$ and the cardinality of $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}})$ is at least two. This implies that $\boldsymbol{x}^{\boldsymbol{v}}$ is a power of a variable, say $\boldsymbol{x}^{\boldsymbol{v}} = x_l^{v_l}$. Observe that $v_l \geq c_l$ and as $f$ is not a critical binomial, $v_l \neq c_l$, whence $\boldsymbol{x}^{\boldsymbol{z}} = x_l^{v_l - c_l} x_i^{u_{li}} x_k^{u_{lk}} \in \deg_{\mathscr{A}}^{-1}(b)$ is a monomial such that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{z}})$ has cardinality greater than or equal to 2 and $l \in \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{z}})$. Then $x_i^{u_i} \leftrightarrow x_i^{u_i - c_i} \boldsymbol{x}^{\boldsymbol{w}} \leftrightarrow \boldsymbol{x}^{\boldsymbol{z}} \leftrightarrow \boldsymbol{x}^{\boldsymbol{v}}$ is a path in $G_b(I_{\mathscr{A}})$, a contradiction. Thus $c_i a_i = c_j a_j$, for an $j \neq i$. We have that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \cap \{i, j\} = \varnothing$; otherwise $x_i^{u_i} \leftrightarrow x_i^{u_i - c_i} x_j^{c_j} \leftrightarrow \boldsymbol{x}^{\boldsymbol{v}}$ is a path in $G_b(I_{\mathscr{A}})$, a contradiction again.

Finally, if $\boldsymbol{x}^{\boldsymbol{v}}$ is not indispensable, then, by Theorem 1.9, there exists a monomial $\boldsymbol{x}^{\boldsymbol{w}} \in \deg_{\mathscr{A}}^{-1}(b) \setminus \{\boldsymbol{x}^{\boldsymbol{v}}\}$ such that $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) \cap \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \neq \varnothing$. If $j \in \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}})$, then $x_i^{u_i} \leftrightarrow x_i^{u_i - c_i} x_j^{c_j} \leftrightarrow \boldsymbol{x}^{\boldsymbol{w}} \leftrightarrow \boldsymbol{x}^{\boldsymbol{v}}$ is a path in $G_b(I_{\mathscr{A}})$, a contradiction to the fact that $f$ is a minimal generator. Moreover $i \notin \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}})$, by the minimality of $c_i$. Thus $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) \subseteq \{k, l\}$ and also $x_k^{v_k} x_l^{v_l} - x_k^{w_k} x_l^{w_l} \in I_{\mathscr{A}}$. Suppose that $c_k a_k \neq c_l a_l$ Then $v_k a_k + v_l a_l = w_k a_k + w_l a_l$. Assume without loss of generality that $w_l \geq v_l$. We have that $(v_k - w_k) a_k = (w_l - v_l) a_l \neq 0$. Hence $v_k - w_k \geq c_k$. If $w_k \neq 0$, then $v_k > c_k$. If $w_k = 0$, $v_k a_k = (w_l - v_l) a_l$ and $v_l \neq 0$, since $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{w}}) \cap \operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) \neq \varnothing$. Thus $w_l - v_l \geq c_l$ and $w_l > c_l$. By using similar arguments as in the first part of the proof we arrive at a contradiction. Consequently $c_k a_k = c_l a_l$.

(ii) The proof is an easy adaptation of the arguments used in (i). ☐

*For the rest of this section we keep the same notation as in Theorem 3.6.*

The following result was first proved by Bresinsky [1988, Theorem 3], but our argument seems to be shorter and more appropriate in our context.

**Proposition 3.9.** *There exists a minimal system of binomial generators of $I_{\mathscr{A}}$ consisting of the union of $\mathscr{S}$ and a set of binomials in $I_{\mathscr{A}}$ with full support.*

*Proof.* By Lemma 3.8(i), if for instance $f = x_i^{u_i} - \boldsymbol{x}^{\boldsymbol{v}}$ is in a minimal generating set of $I_{\mathscr{A}}$ and it is not a critical binomial with respect to any variable, then $c_i a_i = c_j a_j$, for $j \neq i$. We replace $f$ by $g = f - x_i^{u_i - c_i}(x_i^{c_i} - x_j^{c_j}) = x_i^{u_i - c_i} x_j^{c_j} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathscr{A}}$ in the minimal generating set of $I_{\mathscr{A}}$. Moreover, either $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) = \{k, l\}$ and $\{k, l\} \cap \{i, j\} = \varnothing$, so $g$ has full support, or $\boldsymbol{x}^{\boldsymbol{v}}$ is a power of a variable, say $\boldsymbol{x}^{\boldsymbol{v}} = x_k^{v_k}$, with $v_k > c_k$. In this case, by using again Lemma 3.8(i), we replace $g$ with $h = g + x_k^{v_k - c_k}(x_k^{c_k} - x_l^{c_l}) = x_i^{u_i - c_i} x_j^{c_j} - x_k^{v_k - c_k} x_l^{c_l} \in I_{\mathscr{A}}$ with $\{k, l\} \cap \{i, j\} = \varnothing$. Hence, there exists a system of generators of $I_{\mathscr{A}}$ consisting of the union of a system of binomials generators of $C_{\mathscr{A}}$ and a set $\mathscr{S}'$ of binomials in $I_{\mathscr{A}}$ with full support. Furthermore, by Theorem 3.6, we may assume that $\mathscr{S}$ is a system of binomials generators of $C_{\mathscr{A}}$.

Now, let $f = x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{u}} \in \mathscr{S}$ and suppose that $f = \sum_{n=1}^{s} g_n f_n$ where every $f_n \in (\mathscr{S} \setminus \{f\}) \cup \mathscr{S}'$. From the minimality of $c_i$ we have that $f_n = \pm(x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{v}})$ and $|g_n| = 1$, for some $n$. Then, according to the cases in Theorem 3.6, either $\boldsymbol{x}^{\boldsymbol{u}}$ or $\boldsymbol{x}^{\boldsymbol{v}}$ is equal to $x_j^{c_j}$, for some $j \neq i$. Now in the above expression of $f$ the term $x_j^{c_j}$ should be canceled, so, from the minimality of $c_j$, we have $f_m = \pm(x_j^{c_j} - \boldsymbol{x}^{\boldsymbol{w}})$ and $|g_m| = 1$, for an $m \neq n$. Therefore, we conclude that either $\{x_i^{c_i} - x_j^{c_j}, \pm(x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{v}}), \pm(x_j^{c_j} - \boldsymbol{x}^{\boldsymbol{w}})\}$ or $\{x_i^{c_i} - \boldsymbol{x}^{\boldsymbol{u}}, \pm(x_i^{c_i} - x_j^{c_j}), \pm(x_j^{c_j} - \boldsymbol{x}^{\boldsymbol{w}})\}$ is a subset of $\mathscr{S}$. So, the only possible case is $\mathscr{S} = \{x_1^{c_1} - x_2^{c_2}, x_2^{c_2} - x_3^{c_3}, x_3^{c_3} - x_4^{c_4}\}$. Since, in this case, $I_{\mathscr{A}} = C_{\mathscr{A}}$ by Theorem 2.9, and $\mathscr{S}' = \varnothing$, we are done. $\square$

From the above proposition it follows that $I_{\mathscr{A}}$ is generic (see [Ojeda 2008], for instance) only in Case 1. The next theorem provides a minimal generating set for $I_{\mathscr{A}}$.

**Theorem 3.10.** *A minimal system of generators of $I_{\mathscr{A}}$ (up to permutation of indices) is provided by the union of $\mathscr{S}$, the set $\mathscr{I}$ of all binomials $x_{i_1}^{u_{i_1}} x_{i_2}^{u_{i_2}} - x_{i_3}^{u_{i_3}} x_{i_4}^{u_{i_4}} \in I_{\mathscr{A}}$ with $0 < u_{i_j} < c_j,\ j = 1, 2,\ u_{i_3} > 0,\ u_{i_4} > 0$ and $x_{i_3}^{u_{i_3}} x_{i_4}^{u_{i_4}}$ indispensable, and the set $\mathscr{R}$ of all binomials $x_1^{u_1} x_2^{u_2} - x_3^{u_3} x_4^{u_4} \in I_{\mathscr{A}} \setminus \mathscr{I}$ with full support and satisfying the following conditions*:

- *$u_1 \leq c_1$ and $x_3^{u_3} x_4^{u_4}$ is indispensable, in Cases* 2(a) *and* 4(b).
- *$u_1 \leq c_1$ and/or $u_3 \leq c_3$ and there is no $x_1^{v_1} x_2^{v_2} - x_3^{v_3} x_4^{v_4} \in I_{\mathscr{A}}$ with full support such that $x_1^{v_1} x_2^{v_2}$ properly divides $x_1^{u_1 + \alpha c_1} x_2^{u_2 - \alpha c_2}$ or $x_3^{v_3} x_4^{v_4}$ properly divides $x_3^{u_3 + \alpha c_3} x_4^{u_4 - \alpha u_4}$ for some $\alpha \in \mathbb{N}$, in Case* 2(b).

*Proof.* By Proposition 3.9, there exists a minimal system of binomial generators $\mathscr{S} \cup \mathscr{S}'$ of $I_{\mathscr{A}}$ such that $\mathscr{S}$ is a minimal system of generators of $C_{\mathscr{A}}$ and $\operatorname{supp}(f) = \{1, 2, 3, 4\}$, for every $f \in \mathscr{S}'$. Moreover, since all the binomials in the set $\mathscr{I}$ are indispensable by Corollary 2.16, we have $\mathscr{S}' = \mathscr{I} \cup \mathscr{R}$, where $\mathscr{R}$ is a set of binomials of $I_{\mathscr{A}}$ of the form $x_{i_1}^{u_{i_1}} x_{i_2}^{u_{i_2}} - x_{i_3}^{u_{i_3}} x_{i_4}^{u_{i_4}}$ with $u_{i_j} \neq 0$, for every $j$, and $u_{i_j} \geq c_j$ for some $j$.

Observe that if $\mathscr{R} = \varnothing$, then the set defined in the statement of the theorem coincides with $\mathscr{S} \cup \mathscr{S}'$ and therefore it is a minimal set of generators. So, we assume that $\mathscr{R} \neq \varnothing$, that is to say, there exists a minimal generator $x_1^{u_1} x_2^{u_2} - x_3^{u_3} x_4^{u_4} \in \mathscr{R}$ with $u_2 \geq c_2$ (by permuting variables if necessary). By Lemma 3.8(ii) we have

$c_1a_1 = c_2a_2$, so in Case 1 we have $\mathcal{R} = \varnothing$ and therefore we are done. Moreover, if $c_2a_2 = c_ia_i$, for an $i \in \{3,4\}$, then $x_1^{u_1}x_2^{u_2} \leftrightarrow x_1^{u_1}x_2^{u_2-c_2}x_i^{c_i} \leftrightarrow x_3^{u_3}x_4^{u_4}$ is a path in $G_b(I_{\mathcal{A}})$, where $b = u_1a_1 + u_2a_2$, a contradiction with Theorem 1.8. Therefore, we conclude that the theorem is also true in Case 2(c) and Case 3. Notice that, in Case 4(a), we can proceed similarly to reach a contradiction; indeed, since $x_2^{c_2} - \boldsymbol{x}^{\boldsymbol{v}} \in \mathcal{S}$, where $\operatorname{supp}(\boldsymbol{x}^{\boldsymbol{v}}) = \{3,4\}$, then $x_1^{c_1} - \boldsymbol{x}^{\boldsymbol{v}} \in I_{\mathcal{A}}$ and therefore $x_1^{u_1}x_2^{u_2} \leftrightarrow x_1^{u_1+c_1}x_2^{u_2-c_2} \leftrightarrow x_1^{u_1}x_2^{u_2-c_2}\boldsymbol{x}^{\boldsymbol{v}} \leftrightarrow x_3^{u_3}x_4^{u_4}$ is a path in $G_b(I_{\mathcal{A}})$, a contradiction with Theorem 1.8. Thus $\mathcal{R} = \varnothing$ in Case 4(a), too.

Suppose now that $x_1^{v_1}x_i^{v_i} - x_2^{v_2}x_j^{v_j} \in \mathcal{R}$. By Lemma 3.8(ii) again, we obtain that at least one of the equalities $c_1a_1 = c_ia_i$ and $c_2a_2 = c_ja_j$ holds. But, as we proved above, these equalities are incompatible with the condition $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4} \in \mathcal{R}$ with $u_2 \geq c_2$. Hence, all the binomials in $\mathcal{R}$ are of the form $x_1^{\bullet}x_2^{\bullet} - x_3^{\bullet}x_4^{\bullet}$ and $x_2$ arises, with exponent greater than or equal to 2, in at least one of the variables.

We distinguish the following cases:

<u>Case 2(a) or 4(b)</u>. If there exists $x_1^{v_1}x_2^{v_2} - x_3^{v_3}x_4^{v_4} \in \mathcal{R}$ such that for instance $v_4 \geq c_4$, then $c_3a_3 = c_4a_4$ by Lemma 3.8(ii). This is clearly incompatible with Cases 2(a) and 4(b), since $x_3^{v_3}x_4^{v_4} \leftrightarrow x_3^{v_3}x_4^{v_4-c_4}\boldsymbol{x}^{\boldsymbol{u}_4} \leftrightarrow x_1^{v_1}x_2^{v_2}$ is a path in $G_d(I_{\mathcal{A}})$, $d = a_1v_1 + a_2v_2$, a contradiction with Theorem 1.8. Thus the binomials in $\mathcal{R}$ are of the form $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4}$ with $u_i < c_i$, $i = 3,4$. If $x_3^{u_3}x_4^{u_4}$ is not indispensable, then there exists $x^{\boldsymbol{v}} - x_3^{v_3}x_4^{v_4} \in I_{\mathcal{A}}$ such that $0 < v_i \leq u_i$, for $i = 3,4$, with at least one inequality strict and $\operatorname{supp}(x^{\boldsymbol{v}}) \subseteq \{1,2\}$. So, $x_3^{u_3}x_4^{u_4} \leftrightarrow x_3^{u_3-v_3}x_4^{u_4-v_4}\boldsymbol{x}^{\boldsymbol{v}} \leftrightarrow x_1^{u_1}x_2^{u_2}$ is a path in $G_b(I_{\mathcal{A}})$ where $b = a_3u_3 + a_4u_4$, a contradiction with Theorem 1.8. Moreover, since $x_1^{c_1} - x_2^{c_2} \in I_{\mathcal{A}}$, we may change, if it is necessary, $\mathcal{R}$ by replacing every binomial $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4}$, where $u_1 > c_1$, with $x_1^{u_1-\alpha c_1}x_2^{u_2+\alpha c_2} - x_3^{u_3}x_4^{u_4} \in I_{\mathcal{A}}$ such that $0 < u_1 - \alpha c_1 \leq c_1$ and $u_2 + \alpha c_2 \geq c_2$. Now the new set $\mathcal{S} \cup \mathcal{I} \cup \mathcal{R}$ has the desired form. We have that

$$x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4} = (x_1^{u_1-\alpha c_1}x_2^{u_2+\alpha c_2} - x_3^{u_3}x_4^{u_4}) + x_1^{u_1-\alpha c_1}x_2^{u_2}(x_1^{\alpha c_1} - x_2^{\alpha c_2}),$$

so $\mathcal{S} \cup \mathcal{I} \cup \mathcal{R}$ is a generating set of $I_{\mathcal{A}}$. To see that this is actually minimal, by indispensability reasons, it suffices to show that if $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4} \in \mathcal{R}$ and $x_1^{v_1}x_2^{v_2} - x_3^{u_3}x_4^{u_4} \in \mathcal{S} \cup \mathcal{I} \cup \mathcal{R}$, then $x_1^{u_1}x_2^{u_2} = x_1^{v_1}x_2^{v_2}$. Otherwise $x_1^{u_1}x_2^{u_2} - x_1^{v_1}x_2^{v_2} \in I_{\mathcal{A}}$, but $0 < u_1 \leq c_1$ and $v_1 \leq c_1$. Thus $|u_1 - v_1| \leq c_1$, so $u_1 = c_1$, $v_1 = 0$ and therefore $v_2 = c_2$, since every binomial in $\mathcal{S} \cup \mathcal{I} \cup \mathcal{R}$ with cardinality less than four is critical. We have that $c_1a_1 + a_2u_2 = c_2a_2$ and also $c_1a_1 = c_2a_2$, so $u_2 = 0$ a contradiction.

<u>Case 2(b)</u>. Now, by modifying $\mathcal{R}$ as in the previous case if necessary, we have that the binomials in $\mathcal{R}$ are of the following form: $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4}$ with $0 < u_1 \leq c_1$, $u_2 \neq 0$ and/or $0 < u_3 \leq c_3$, $u_4 \neq 0$. If there exists $\alpha \in \mathbb{N}$ and $x_1^{v_1}x_2^{v_2} - x_3^{v_3}x_4^{v_4} \in I_{\mathcal{A}}$ with full support such that $x_1^{u_1+\alpha c_1}x_2^{u_2-\alpha c_2} = mx_1^{v_1}x_2^{v_2}$ (or $x_3^{u_3+\alpha c_3}x_4^{u_4-\alpha c_4} = mx_3^{v_3}x_4^{v_4}$, respectively) with $m \neq 1$, then $x_1^{u_1}x_2^{u_2} \leftrightarrow mx_3^{v_3}x_4^{v_4} \leftrightarrow x_3^{u_3}x_4^{u_4}$ (or $x_1^{u_1}x_2^{u_2} \leftrightarrow x_1^{v_1}x_2^{v_2}m \leftrightarrow$

$x_3^{u_3}x_4^{u_4}$, respectively) is a path in $G_b(I_{\mathscr{A}})$, where $b = u_1a_1 + u_2a_2$, a contradiction with Theorem 1.8. So, we conclude that all the binomials in $\mathscr{R}$ are of the desired form. Moreover, given $f = x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4} \in \mathscr{R}$ and a monomial $\boldsymbol{x}^{\boldsymbol{v}}$ with $\deg_{\mathscr{A}}(\boldsymbol{x}^{\boldsymbol{v}}) = u_1a_1 + u_2a_2$, then either $v_1 = v_2 = 0$ or $v_1 = v_3 = v_4 = 0$ and $v_2 > c_2$. Indeed, since $x_1^{u_1}x_2^{u_2} - x_1^{v_1}x_2^{v_2}x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, we have the following possibilities:

(i) $g = x_1^{u_1-v_1}x_2^{u_2-v_2} - x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, when $v_1 \le u_1$ and $v_2 < u_2$. If $g$ has full support, then $v_1 = v_2 = 0$, otherwise $f \notin \mathscr{R}$. If for instance $u_1 - v_1 = 0$, then $u_2 - v_2 \ge c_2$, because of the minimality of $c_2$. Thus, $g' = x_1^{u_1-v_1+c_1}x_2^{u_2-v_2-c_2} - x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$. If $g'$ has full support, then $v_1 = v_2 = 0$; otherwise the monomial $x_1^{u_1-v_1+c_1}x_2^{u_2-v_2-c_2}$ properly divides $x_1^{u_1+c_1}x_2^{u_2-c_2}$, that is to say, $f \notin \mathscr{R}$. If $g'$ does not have full support, say $v_3 = 0$, then $v_4 \ge c_4$ (due to the minimality of $c_4$). So, we may define $g'' = x_1^{u_1-v_1+c_1}x_2^{u_2-v_2-c_2} - x_3^{c_3}x_4^{v_4-c_4} \in I_{\mathscr{A}}$ and conclude that $v_1 = v_2 = 0$, as before.

(ii) $g = x_1^{u_1-v_1} - x_2^{v_2-u_2}x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, when $v_1 < u_1$ and $v_2 \ge u_2$. Since $0 < u_1 \le c_1$, we have that $v_1 = 0$ and also $u_1 = c_1$. Thus $v_2 - u_2 = c_2$ and $v_3 = v_4 = 0$, since $x_1^{c_1} - x_2^{c_2}$ is the only critical binomial with respect to $x_1$.

(iii) $g = x_2^{u_2-v_2} - x_1^{v_1-u_1}x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, when $v_1 \ge u_1$ and $v_2 < u_2$. Now, by the minimality of $c_2$, we have that $u_2 - v_2 \ge c_2$ and therefore $h = x_1^{c_1}x_2^{u_2-v_2-c_2} - x_1^{v_1-u_1}x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$. So, either $x_1^{c_1+u_1-v_1}x_2^{u_2-v_2-c_2} - x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, when $c_1 \ge v_1 - u_1$, or $x_2^{u_2-v_2-c_2} - x_1^{v_1-u_1-c_1}x_3^{v_3}x_4^{v_4} \in I_{\mathscr{A}}$, when $c_1 < v_1 - u_1$. In the first case we proceed as in (i), while in the other we repeat the same argument and so on. This process can not continue indefinitely, since there exists $\alpha \in \mathbb{N}$ such that $\alpha c_1 < v_1 - u_1$, and thus we are done.

From Theorem 1.8 we have that there exists a minimal generator of $\mathscr{A}$-degree $\deg_{\mathscr{A}}(f)$ for each $f \in \mathscr{R}$. Furthermore, by direct checking one can show that all the binomials in $\mathscr{I} \cup \mathscr{R}$ have a different $\mathscr{A}$-degree, and all these $\mathscr{A}$-degrees are different from both $c_1a_1$ and $c_2a_2$. Thus, we conclude that $\mathscr{S} \cup \mathscr{I} \cup \mathscr{R}$ is a minimal system of generators of $I_{\mathscr{A}}$. □

Combining Theorem 3.10 with Corollaries 2.5 and 2.16 yields the following theorem.

**Theorem 3.11.** *With the same notation as in Theorem 3.10, the ideal $I_{\mathscr{A}}$ has a unique minimal system of generators if and only if $C_{\mathscr{A}}$ has a unique minimal system of generators and $\mathscr{R} = \varnothing$.*

In [Ojeda 2008], it is shown that there exist semigroup ideals of $\Bbbk[x_1, \dots, x_4]$ with unique minimal system of binomial generators of cardinality $m$, for every $m \ge 7$.

**Example 3.12.** Let $\mathscr{A} = \{6, 8, 17, 19\}$. The critical binomial $x_1^4 - x_2^3$ of $I_{\mathscr{A}}$ is indispensable, while the critical binomial $x_4^2 - x_1x_2^4$ is not indispensable. Thus we are in Case 4(b). The binomial $x_1^2x_2^3 - x_3x_4$ belongs to $\mathcal{R}$ and therefore, from Theorem 3.11, the toric ideal $I_{\mathcal{A}}$ does not have a unique minimal system of binomial generators.

**Example 3.13.** Let $\mathcal{A} = \{25, 30, 57, 76\}$, then the minimal number of generators of $I_{\mathcal{A}}$ equals 8. The only critical binomials of $I_{\mathcal{A}}$ are $\pm(x_1^6 - x_2^5)$ and $\pm(x_3^4 - x_4^3)$, so we are in Case 2(b). The binomial $x_1^3x_2^7 - x_3x_4^3$ belongs to $\mathcal{R}$ and therefore, from Theorem 3.11, the toric ideal $I_{\mathcal{A}}$ does not have a unique minimal system of binomial generators.

Observe that $I_{\mathcal{A}}$ is a complete intersection only in cases 2(a-c), 3 and 4(b). Moreover, except from 2(b), in all the other cases $I_{\mathcal{A}} = C_{\mathcal{A}}$. In the case 2(b) a minimal system of binomial generators is $x_1^{c_1} - x_2^{c_2}$, $x_3^{c_3} - x_4^{c_4}$ and $x_1^{u_1}x_2^{u_2} - x_3^{u_3}x_4^{u_4}$ where $a_1u_1 + a_2u_2 = a_3u_3 + a_4u_4 = \operatorname{lcm}(\gcd(a_1, a_2), \gcd(a_3, a_4))$; [Delorme 1976].

It is well known that the ring $\Bbbk[\boldsymbol{x}]/I_{\mathcal{A}}$ is Gorenstein if and only if the semigroup $\mathbb{N}\mathcal{A}$ is symmetric, see [Kunz 1970]. We will prove that if $\mathbb{N}\mathcal{A}$ is symmetric and $I_{\mathcal{A}}$ is not a complete intersection, then $I_{\mathcal{A}}$ has a unique minimal system of binomial generators.

**Theorem 3.14.** *If $f_1 = x_1^{c_1} - x_3^{u_{13}}x_4^{u_{14}}$, $f_2 = x_2^{c_2} - x_1^{u_{21}}x_4^{u_{24}}$, $f_3 = x_3^{c_3} - x_1^{u_{31}}x_2^{u_{32}}$ and $f_4 = x_4^{c_4} - x_2^{u_{42}}x_3^{u_{43}}$ are critical binomials of $I_{\mathcal{A}}$ such that* $\operatorname{supp}(f_i)$ *has cardinality equal to* 3, *for every $i \in \{1, \dots, 4\}$, then $I_{\mathcal{A}}$ has a unique minimal system of binomial generators.*

*Proof.* Every exponent $u_{ij}$ of $x_j$ is strictly less than $c_j$, for each $j = 1, \dots, 4$. If for instance $u_{13} \geq c_3$, then $x_1^{c_1} - x_1^{u_{31}}x_2^{u_{32}}x_3^{u_{13}-c_3}x_4^{u_{14}} = f_1 + x_3^{u_{13}-c_3}x_4^{u_{14}}f_3 \in I_{\mathcal{A}}$ and therefore $x_1^{c_1-u_{31}} - x_2^{u_{32}}x_3^{u_{13}-c_3}x_4^{u_{14}} \in I_{\mathcal{A}}$, a contradiction to the minimality of $c_1$. By Proposition 2.3 we have that $c_ia_i \neq c_ja_j$, for every $i \neq j$. We will prove that every $f_i$ is indispensable in $C_{\mathcal{A}}$. Suppose for example that $f_1$ is not indispensable in $C_{\mathcal{A}}$, then there is a binomial $g = x_1^{c_1} - x_2^{v_2}x_3^{v_3}x_4^{v_4} \in I_{\mathcal{A}}$. So $x_3^{u_{13}}x_4^{u_{14}} - x_2^{v_2}x_3^{v_3}x_4^{v_4} \in I_{\mathcal{A}}$, and thus $v_3 < u_{13}$ and $v_4 < u_{14}$, since $u_{13} < c_3$ and $u_{14} < c_4$. We have that $x_2^{v_2} - x_3^{u_{13}-v_3}x_4^{u_{14}-v_4} \in I_{\mathcal{A}}$ and also $x_1^{c_1} - x_1^{u_{21}}x_2^{v_2-c_2}x_3^{v_3}x_4^{u_{24}+v_4} = g + x_2^{v_2-c_2}x_3^{v_3}x_4^{v_4}f_2 \in I_{\mathcal{A}}$. Therefore $x_1^{c_1-u_{21}} - x_2^{v_2-c_2}x_3^{v_3}x_4^{u_{24}+v_4} \in I_{\mathcal{A}}$, a contradiction to the minimality of $c_1$. Analogously we can prove that $f_2$, $f_3$ and $f_4$ are indispensable in $C_{\mathcal{A}}$. Thus $C_{\mathcal{A}}$ is generated by its indispensable binomials and therefore, from Theorem 3.11, the toric ideal $I_{\mathcal{A}}$ has a unique minimal system of binomial generators. □

**Corollary 3.15.** *Let $\mathbb{N}\mathcal{A}$ be a symmetric semigroup. If $I_{\mathcal{A}}$ is not a complete intersection, then it has a unique minimal system of binomial generators.*

*Proof.* From [Bresinsky 1975, Theorem 3] the toric ideal $I_{\mathcal{A}}$ has a minimal generating set consisting of five binomials, namely four critical binomials of the form defined in the above theorem and a non critical binomial. By Theorem 3.14 the toric ideal $I_{\mathcal{A}}$ is generated by its indispensable binomials. □

According to [Bresinsky 1975, Theorem 4] the integers $a_i$ are polynomials in the exponents of the binomial in a minimal generating system of $I_{\mathscr{A}}$. We can see these expressions as a system of four polynomial equations, which in light of Corollary 3.15, has a unique solution over the positive integers.

**Remark 3.16.** Theorem 6.4 of [Komeda 1982] shows that if $\mathbb{N}\mathscr{A}$ is pseudosymmetric (see [Rosales and García 2009] for a definition), then $f_1 = x_1^{c_1} - x_3 x_4^{c_4-1}$, $f_2 = x_2^{c_2} - x_1^{u_{21}} x_4$, $f_3 = x_3^{c_3} - x_1^{c_1-u_{21}-1} x_2$, $f_4 = x_4^{c_4} - x_1 x_2^{c_2-1} x_3^{c_3-1}$ and $g = x_1^{u_{21}+1} x_3^{c_3-1} - x_2 x_4^{c_4-1}$ with $c_i > 1$ for $i = 1, \ldots, 4$, and $u_{21} - 1 < c_1$, is a minimal system of generators of $I_{\mathscr{A}}$. Now, an easy check shows that $c_i a_i \neq c_j a_j$ for every $i \neq j$. The interested reader may prove that $C_{\mathscr{A}}$ has a unique minimal system of generators if and only if $u_{21} = c_1 - 2$. Thus, since $\mathscr{R} = \varnothing$, by Theorem 3.11, we conclude that $I_{\mathscr{A}}$ is generated by its indispensable binomials if and only if $c_2 n_2 \neq (c_1 - 2) n_1 + n_4$.

If the cardinality of $\mathscr{A}$ is greater than 4, the analogous of Corollary 3.15 is not true in general. In [Rosales 2001] it is shown that the semigroup generated by $\mathscr{A} = \{15, 16, 81, 82, 83, 84\}$ is symmetric. Since the monomials $x_1^{11}$, $x_3 x_6$ and $x_4 x_5$ have the same $\mathscr{A}$-degree, we conclude, by Theorem 1.8, that the ideal $I_{\mathscr{A}}$ does not have a unique minimal system of binomial generators.

## Acknowledgments

Part of this work was done during a visit of the first author to the University of Extremadura financed by the Plan Propio 2010 of the University of Extremadura. We thank the referee for helpful comments and suggestions that improved the paper.

## References

[Alcántar and Villarreal 1994] A. Alcántar and R. H. Villarreal, "Critical binomials of monomial curves", *Comm. Algebra* **22**:8 (1994), 3037–3052. MR 95c:13022 Zbl 0855.13014

[Aoki et al. 2008] S. Aoki, A. Takemura, and R. Yoshida, "Indispensable monomials of toric ideals and Markov bases", *J. Symbolic Comput.* **43**:6-7 (2008), 490–507. MR 2009c:13065 Zbl 1170.13008

[Bresinsky 1975] H. Bresinsky, "Symmetric semigroups of integers generated by 4 elements", *Manuscripta Math.* **17**:3 (1975), 205–219. MR 54 #2660 Zbl 0317.10061

[Bresinsky 1988] H. Bresinsky, "Binomial generating sets for monomial curves, with applications in $\mathbf{A}^4$", *Rend. Sem. Mat. Univ. Politec. Torino* **46**:3 (1988), 353–370. MR 92e:13004 Zbl 0738.14017

[Charalambous et al. 2007] H. Charalambous, A. Katsabekis, and A. Thoma, "Minimal systems of binomial generators and the indispensable complex of a toric ideal", *Proc. Amer. Math. Soc.* **135**:11 (2007), 3443–3451. MR 2009a:13033 Zbl 1127.13018

[Delgado et al. 2013] M. Delgado, P. A. García Sánchez, and J. Morais, "NumericalSgps: a GAP package", 2013, Available at http://cmup.fc.up.pt/cmup/mdelgado/numericalsgps. Version 0.980.

[Delorme 1976] C. Delorme, "Sous-monoïdes d'intersection complète de $N$", *Ann. Sci. École Norm. Sup.* (4) **9**:1 (1976), 145–154. MR 53 #10821 Zbl 0325.20065

[Eisenbud and Sturmfels 1996] D. Eisenbud and B. Sturmfels, "Binomial ideals", *Duke Math. J.* **84**:1 (1996), 1–45. MR 97d:13031 Zbl 0873.13021

[Fischer and Shapiro 1996] K. G. Fischer and J. Shapiro, "Mixed matrices and binomial ideals", *J. Pure Appl. Algebra* **113**:1 (1996), 39–54. MR 97h:13008 Zbl 0864.15016

[García and Ojeda 2010] P. A. García Sánchez and I. Ojeda, "Uniquely presented finitely generated commutative monoids", *Pacific J. Math.* **248**:1 (2010), 91–105. MR 2011j:20139 Zbl 1208.20052

[García Sánchez et al. 2013] P. A. García Sánchez, I. Ojeda, and J. C. Rosales, "Affine semigroups having a unique Betti element", *J. Algebra Appl.* **12**:3 (2013), Article ID #1250177. MR 3007913 Zbl 06155975

[Herzog 1970] J. Herzog, "Generators and relations of abelian semigroups and semigroup rings", *Manuscripta Math.* **3** (1970), 175–193. MR 42 #4657 Zbl 0211.33801

[Herzog et al. 2010] J. Herzog, T. Hibi, F. Hreinsdóttir, T. Kahle, and J. Rauh, "Binomial edge ideals and conditional independence statements", *Adv. in Appl. Math.* **45**:3 (2010), 317–333. MR 2011j:13041 Zbl 1196.13018

[Katsabekis and Thoma 2010] A. Katsabekis and A. Thoma, "Specializations of multigradings and the arithmetical rank of lattice ideals", *Comm. Algebra* **38**:5 (2010), 1904–1918. MR 2011e:13045 Zbl 1197.14054

[Komeda 1982] J. Komeda, "On the existence of Weierstrass points with a certain semigroup generated by 4 elements", *Tsukuba J. Math.* **6**:2 (1982), 237–270. MR 85d:14039 Zbl 0546.14011

[Kunz 1970] E. Kunz, "The value-semigroup of a one-dimensional Gorenstein ring", *Proc. Amer. Math. Soc.* **25** (1970), 748–751. MR 42 #263 Zbl 0197.31401

[Ohsugi and Hibi 2005] H. Ohsugi and T. Hibi, "Indispensable binomials of finite graphs", *J. Algebra Appl.* **4**:4 (2005), 421–434. MR 2006e:13023 Zbl 1093.13020

[Ojeda 2008] I. Ojeda, "Examples of generic lattice ideals of codimension 3", *Comm. Algebra* **36**:1 (2008), 279–287. MR 2008j:13027 Zbl 1133.13014

[Ojeda and Pisón Casares 2004] I. Ojeda and P. Pisón Casares, "On the hull resolution of an affine monomial curve", *J. Pure Appl. Algebra* **192**:1-3 (2004), 53–67. MR 2005e:13018 Zbl 1079.13007

[Ojeda and Vigneron-Tenorio 2010a] I. Ojeda and A. Vigneron-Tenorio, "Indispensable binomials in semigroup ideals", *Proc. Amer. Math. Soc.* **138**:12 (2010), 4205–4216. MR 2011i:13023 Zbl 1204.13014

[Ojeda and Vigneron-Tenorio 2010b] I. Ojeda and A. Vigneron-Tenorio, "Simplicial complexes and minimal free resolution of monomial algebras", *J. Pure Appl. Algebra* **214**:6 (2010), 850–861. MR 2011g:13032 Zbl 1195.13015

[Rosales 2001] J. C. Rosales, "Symmetric numerical semigroups with arbitrary multiplicity and embedding dimension", *Proc. Amer. Math. Soc.* **129**:8 (2001), 2197–2203. MR 2002b:20091 Zbl 0972.20036

[Rosales and García 2009] J. C. Rosales and P. A. García-Sánchez, *Numerical semigroups*, Developments in Mathematics **20**, Springer, New York, 2009. MR 2010j:20091 Zbl 1220.20047

[Sturmfels 1996] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series **8**, American Mathematical Society, Providence, RI, 1996. MR 97b:13034 Zbl 0856.13020

Received July 26, 2012. Revised October 7, 2013.

ANARGYROS KATSABEKIS
CENTRUM WISKUNDE & INFORMATICA (CWI)
POSTBUS 94079
1090 GB AMSTERDAM
THE NETHERLANDS

katsabek@aegean.gr

IGNACIO OJEDA
DEPARTAMENTO DE MATEMATICAS
UNIVERSIDAD DE EXTREMADURA
FACULTAD DE CIENCIAS
AVENIDA DE ELVAS S/N
06071 BADAJOZ
SPAIN

ojedamc@unex.es

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.117

# CONTRACTING AN AXIALLY SYMMETRIC TORUS BY ITS HARMONIC MEAN CURVATURE

CHRISTOPHER KIM

**We consider the harmonic mean curvature flow of an axially symmetric torus whose axis is a closed geodesic, where the ambient space is a hyperbolic three-manifold. Assuming the initial surface is strictly convex and its harmonic mean curvature is less than $\frac{1}{2}$, we show that the evolving surface satisfies a curvature condition comparable to that of a perfectly symmetric torus evolving under harmonic mean curvature flow. In other words, we prove that $\lambda_1 \approx e^{-t}$, $\lambda_2 \approx e^{t}$ and $\lambda_1\lambda_2 \approx 1$, where $\lambda_1$ and $\lambda_2$ are the principal curvatures of the evolving torus.**

## 1. Introduction

We consider the contraction of a convex torus embedded in a hyperbolic 3-manifold to a closed geodesic using the harmonic mean of the principal curvatures. Each point on the torus whose axis is a closed geodesic moves in the normal direction pointing to its axis with a speed equal to the harmonic mean curvature. Let $\Sigma^2 = S^1 \times S^1$ be a two-dimensional torus, $N^3$ a hyperbolic 3-manifold containing a closed geodesic and $\Phi_0 : \Sigma^2 \to N^3$ a smooth initial immersion of $\Sigma^2$ into $N^3$ centered at a closed geodesic. The evolution process is described by a one-parameter family of immersions $\Phi : \Sigma \times [0, T) \to N$ satisfying

$$\text{(HMCF)} \qquad \begin{aligned} \frac{\partial \Phi(p,t)}{\partial t} &= -F(p,t)\cdot N(p,t), \\ \Phi(p,0) &= \Phi_0(p). \end{aligned}$$

Here, $F = \lambda_1\lambda_2/(\lambda_1+\lambda_2)$ is the harmonic mean curvature of $\Sigma_t := \Phi(\Sigma, t)$ where $\lambda_1, \lambda_2$ are the principal curvatures and $N$ is the outward unit normal vector of $\Sigma_t$.

Andrews studied harmonic mean curvature flow (HMCF) of strictly convex compact hypersurfaces without boundary in Euclidean [Andrews 1994a] and Riemannian manifolds [1994b], showing that the evolving hypersurface converges to a round point in finite time. Other authors studied HMCF of hypersurfaces in Euclidean space under various curvature conditions [Caputo and Daskalopoulos

*MSC2010:* primary 53C44; secondary 35K55.
*Keywords:* harmonic mean curvature flow, hyperbolic manifold, closed geodesic.

2009; Daskalopoulos and Hamilton 2006; Daskalopoulos and Sesum 2010; Dieter 2005] and showed that the evolving hypersurface converges, when it does, to a round point. In this paper, we are interested in surfaces converging to a closed geodesic, not a point, in hyperbolic 3-manifolds by HMCF. Examples of hypersurfaces in hyperbolic manifolds converging to a totally geodesic submanifold by HMCF were constructed in [Gulliver and Xu 2009]. However, only hypersurfaces at a constant distance from totally geodesic submanifolds were considered there, so the curvature flow problem reduced to analyzing simple ODEs. This paper generalizes parts of the results of Gulliver and Xu to axially symmetric surfaces. Recently in [Andrews et al. 2013], weakly convex hypersurfaces in Euclidean space containing cylindrical regions were shown to shrink to a line segment when the hypersurface is deformed by certain curvature function. However, curvatures of the evolving surface were not analyzed in that paper.

In this paper, we will obtain curvature estimates of an axially symmetric torus contracting to a closed geodesic in hyperbolic 3-manifold by HMCF. Analyzing the principal curvatures of a torus presents a novel problem since as the torus approaches the axis we expect the small principal curvature to converge zero and the large principal curvature to approach infinity. And the product of the principal curvatures is expected to be more or less constant since it equals 1 (see (1-1)) on a perfect torus whose axis is a closed geodesic. This kind of curvature estimate is different from the estimates obtained for spherical hypersurfaces in Theorem 4.1 of [Andrews 1994b] and Theorem 5.1 of [Huisken 1984], stating that the ratio of principal curvatures are uniformly bounded. We need to estimate each principal curvature separately to show that they exhibit contrasting dynamics but the product should remain bounded throughout the evolution process.

We will consider a torus $\Sigma^2$ embedded into a hyperbolic 3-manifold $N^3$ such that it is axially symmetric about a closed geodesic $\gamma : S^1 \to N^3$. Let $r : S^1 \to [0, R]$ be a generating function defined on $\gamma$. An axially symmetric torus can be constructed by revolving the graph of the generating function about the closed geodesic.

**Theorem 1.1** (main theorem). *Let $\Sigma_0$ be an axially symmetric torus around a closed geodesic $\gamma$ in a hyperbolic 3-manifold $N$, generated by revolving a graph of $r : S^1 \to \mathbb{R}^+$ about $\gamma$. Assume $\Sigma_0$ is strictly convex and $\max_{x\in\Sigma_0} F(x) < \frac{1}{2}$ where $F(x)$ is the harmonic mean curvature at $x \in \Sigma_0$. Then, the solution of the HMCF with initial surface $\Sigma_0$ exists for all $t \in [0, \infty)$ and remains strictly convex. The evolving surface converges to the closed geodesic exponentially fast and the principal curvatures satisfy $\lambda_1 \approx e^{-t}$, $\lambda_2 \approx e^t$ and $\lambda_1\lambda_2 \approx 1$.*

**Notation.** Uniform constants are denoted by $C_i$. The same symbol $C$ might imply different constants from line to line. The approximation symbol $f \approx g$ denotes that there exist $C_1, C_2 > 0$ such that $C_1 g \le f \le C_2 g$.

**Remarks.** (1) The reason we impose the curvature condition $\max_{\Sigma_0} F < \frac{1}{2}$ is that for a perfectly symmetric torus we have $0 < F(r) < \frac{1}{2}$ for all $r \in (0, \infty)$; thus perfectly symmetric tori of any radius satisfy the condition. This can be easily seen as follows: Since the principal curvatures of a perfect torus are $\lambda_1 = \tanh r$ and $\lambda_2 = \coth r$ (by the Riccati equation, $\lambda_i' + \lambda_i^2 = 1$), the harmonic mean curvature is

$$F = \frac{1}{(\coth r)^{-1} + (\tanh r)^{-1}} = \frac{1}{\coth r + \tanh r}.$$

Thus,

$$\frac{dF}{dr} = \frac{1}{\sinh^2 r + \cosh^2 r} > 0 \quad \text{for all } r.$$

But

$$\lim_{r\to 0} \coth r = \infty, \quad \lim_{r\to 0} \tanh r = 0, \quad \lim_{r\to\infty} \coth r = \lim_{r\to\infty} \tanh r = 1.$$

Therefore

$$\lim_{r\to 0} F = 0, \quad \lim_{r\to\infty} F = \tfrac{1}{2}, \quad 0 < F(r) < \tfrac{1}{2}.$$

(2) The HMCF of a perfectly symmetric torus whose axis is a closed geodesic in a hyperbolic manifold was considered in Theorem 3 of [Gulliver and Xu 2009]. The authors showed that the radius $r(t)$ of the evolving torus satisfies

$$r(t) = \tfrac{1}{2} \sinh^{-1}(e^{-t} \sinh 2r_0) \approx e^{-t},$$

where $r_0$ is the radius of the initial torus. Since the principal curvatures of perfect torus are $\lambda_1 = \tanh r$ and $\lambda_2 = \coth r$, we obtain the asymptotic estimates of both principal curvatures:

$$\lambda_1 \approx e^{-t}, \quad \lambda_2 \approx e^{t}, \quad \lambda_1\lambda_2 = 1. \tag{1-1}$$

The main theorem of this paper shows that the principal curvatures of an axially symmetric torus contracting to a closed geodesic under HMCF retain the curvature estimates (1-1) of an evolving perfectly symmetric torus.

The paper is organized as follows. In Section 2, we derive essential geometric quantities available on axially symmetric spaces. In Section 3, we prove the short and long time existence of HMCF of axially symmetric torus and discuss the preservation of convexity of the surface. We derive the evolution equations of important geometric quantities in Section 4. In Section 5, we prove that the evolving surface remains a graph throughout the deformation process and also prove that $\lambda_2 \approx e^t$. Along the way, we obtain the optimal estimate $\lambda_1 \approx e^{-t}$ and conclude that $\lambda_1\lambda_2 \approx 1$.

## 2. Axially symmetric spaces

In this section, we will use the orthonormal frames to derive geometric quantities defined on axially symmetric surfaces. A similar computation was carried out in [Cabezas-Rivas and Miquel 2009] for general rotationally symmetric spaces. In the neighborhood of the closed geodesic, the hyperbolic metric can be expressed in Fermi coordinates as $ds^2 = dr^2 + h(r)^2 d\theta^2 + b(r)^2 dz^2$ where $r$ is the distance from the axis, $\theta$ is the angular unit of the circle perpendicular to the axis, $z$ is the position along the axis and $b(r) = \cosh r$, $h(r) = \sinh r$. We have the following orthonormal frames in $(n+1)$-dimensional rotationally symmetric space.

$$E_0 := E_r = \frac{\partial}{\partial r}, \quad E_1 := E_z = \frac{1}{b(r)}\frac{\partial}{\partial z}, \quad E_i = \frac{1}{h(r)} e_i \quad \text{for } i = 2, \dots, n,$$

where $e_i$ is an orthonormal frame of $S^{n-1}$ with the standard metric. Its dual orthonormal coframe is given by

$$\theta^r = dr, \quad \theta^z = b(r)dz, \quad \theta^i = h(r)e_i \quad \text{for } i = 2, \dots, n.$$

In these frames, the Cartan connection form $\omega_a^b$ defined by $d\theta^b = -\sum_{a=0}^{n} \omega_a^b \wedge \theta^a$ is given by

$$\omega_r^z = \frac{b'(r)}{b(r)}\theta^z, \quad \omega_r^i = \frac{h'(r)}{h(r)}\theta^i, \quad \omega_z^i = 0, \quad \omega_j^i = {}^S\omega_j^i$$

where ${}^S\omega_j^i$ represents the Cartan connection form on $S^{n-1}$. The covariant derivatives of the orthonormal frames can be computed from the equation $\overline{\nabla}_X E_a = \sum_{b=0}^n \omega_a^b(X) E_b$ and their results are given below. We denote the covariant derivative defined on the ambient manifold by $\overline{\nabla}$ and the covariant derivative on the hypersurface by $\nabla$. The symbol $'$ denotes the derivative with respect to $r$ and subscripts of $r$ mean the derivative with respect to $z$. For $i = 2, \dots, n$,

$$\begin{aligned}
&\overline{\nabla}_{E_r} E_r = 0, \quad \overline{\nabla}_{E_z} E_r = \frac{b'(r)}{b(r)} E_z, \quad \overline{\nabla}_{E_i} E_r = \frac{h'(r)}{h(r)} E_i \\
&\overline{\nabla}_{E_r} E_z = 0, \quad \overline{\nabla}_{E_z} E_z = -\frac{b'(r)}{b(r)} E_r, \quad \overline{\nabla}_{E_i} E_z = 0, \\
&\overline{\nabla}_{E_r} E_i = 0, \quad \overline{\nabla}_{E_z} E_i = 0, \quad \overline{\nabla}_{E_i} E_j = -\frac{h'(r)}{h(r)} \delta_{ij} E_r + {}^S\omega_j^k(E_i) E_k.
\end{aligned} \tag{2-1}$$

For a hypersurface constructed by revolving the graph of a generating function $r : S^1 \to \mathbb{R}^+$, the tangent vector $\sigma$ of the generating curve and the unit normal vector $N$ of the hypersurface are given by

$$\sigma = \frac{1}{\sqrt{r_z^2 + b^2}}(r_z E_r + b E_z), \quad N = \frac{1}{\sqrt{r_z^2 + b^2}}(b E_r - r_z E_z). \tag{2-2}$$

The principal curvatures of the hypersurface in the direction of $\sigma$ and $E_i$ are

$$\lambda_1 = \langle \overline{\nabla}_\sigma \sigma, N \rangle = \frac{1}{\sqrt{r_z^2+b^2}} \left( \frac{-r_{zz} b + r_z^2 b'}{r_z^2+b^2} + b' \right), \tag{2-3}$$

$$\lambda_i = \langle \overline{\nabla}_{E_i} E_i, N \rangle = \frac{b}{\sqrt{r_z^2+b^2}} \frac{h'}{h} = u \frac{h'}{h} \quad \text{for } i = 2, \dots, n, \tag{2-4}$$

respectively. Note the hypersurfaces of revolution is generated by a graph if

$$u := \langle E_r, N \rangle = \frac{b}{\sqrt{r_z^2+b^2}} \tag{2-5}$$

is greater than 0; equivalently, $v := u^{-1}$ is finite. Note that $u \le 1$ by its definition.

## 3. Short and long time existence and preserving convexity

In this section, we first prove the short time existence of HMCF of axially symmetric torus and review the long time existence and preservation of convexity proved in [Gulliver and Xu 2009]. Let $W_i^j = h_{ik} g^{kj}$ be the Weingarten map of $\Sigma_t$, where $h_{ij}$ is the second fundamental form and $g_{ij}$ is the induced metric on $\Sigma_t$. We can view the harmonic mean curvature function as $F(W_i^j) = f(\lambda(W_i^j))$, where $\lambda(W_i^j) = (\lambda_1, \lambda_2)$ is the set of eigenvalues of $W_i^j$ and $f(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 / (\lambda_1 + \lambda_2)$. Let us first discuss the short time existence of HMCF when the flow equation is cast in terms of the graph function. If we express (HMCF) in terms of the graph function using

$$\left\langle \frac{\partial \phi}{\partial t}, N \right\rangle = F,$$

we obtain

$$\frac{\partial r}{\partial t} = -\frac{r_{zz} - 2\tanh(r) r_z^2 - \sinh r \cosh r}{\tanh(r) r_{zz} - (2\tanh^2 r + 1) r_z^2 - \sinh^2 r - \cosh^2 r}, \qquad r(z, 0) = r_0(z) \tag{3-1}$$

for all $(z, t) \in S^1 \times [0, T)$. Since the initial surface is assumed to be strictly convex, from (2-3) and (2-4) we find that at $t = 0$

$$\tilde{\lambda}_1 := -r_{zz} + 2\tanh(r) r_z^2 + \sinh r \cosh r > 0. \tag{3-2}$$

We consider positive solutions

$$r > 0. \tag{3-3}$$

We define $C^\alpha(S^1)$ to be the set of standard Hölder continuous functions on $S^1$ and $C^{2+\alpha}(S^1)$ to be a space of functions $g$ on $S^1$ such that $g, g_z, g_{zz} \in C^\alpha(S^1)$. We set $Q_\tau = S^1 \times [0, \tau]$ for some $\tau > 0$ and define $C^{2+\alpha}(Q_\tau)$ to be a space of functions $g$ on $Q_\tau$ such that $g_t, g, g_z, g_{zz} \in C^\alpha(S^1)$.

**Lemma 3.1.** *Let $r_0 \in C^{2+\alpha}(S^1)$. There exists some $t_0 > 0$ such that a unique solution $r \in C^{2+\alpha}\big(S^1 \times [0, t_0]\big)$ solves (3-1).*

*Proof.* Let $M : C^{2+\alpha}(Q_\tau) \to C^\alpha(Q_\tau)$ be a fully nonlinear operator defined by

$$M(r) = r_t - F(z, t, r, r_z, r_{zz}),$$

where

$$F(z, t, r, r_z, r_{zz}) = -\frac{r_{zz} - 2\tanh r\, r_z^2 - \sinh r \cosh r}{\tanh r\, r_{zz} - (2\tanh^2 r + 1)r_z^2 - \sinh^2 r - \cosh^2 r}.$$

Consider the linearization of $M$ around a function $r \in C^{2+\alpha}(Q_\tau)$ such that $\|r - r_0\| < \delta$ for some $\delta > 0$. If we choose $\delta$ small enough, any such $r$ will satisfy conditions (3-2) and (3-3) since the initial condition $r_0 \in C^{2+\alpha}(S^1)$ satisfies those conditions. Then, the linearized equation around the function $r$, namely

$$\frac{\partial \tilde{r}}{\partial t} = DF(r)(\tilde{r}) = \alpha(r, r_z, r_{zz})\tilde{r}_{zz} + \beta(r, r_z, r_{zz})\tilde{r}_z + \gamma(r, r_z, r_{zz})\tilde{r}, \tag{3-4}$$

where

$$\alpha = \frac{-r_z^2 - \cosh^2 r}{(\tanh r\tilde{\lambda}_1 + r_z^2 + \cosh^2 r)^2},$$

$$\beta = \frac{(4\tanh^2 r - 4\tanh^2 r - 2)r_z\tilde{\lambda}_1 + 4\tanh r\, r_z(r_z^2 + \cosh^2 r)}{(\tanh r\tilde{\lambda}_1 + r_z^2 + \cosh^2 r)^2},$$

$$\gamma = \bigg[\bigg(\frac{r_{zz}}{\cosh^2 r} - \frac{2\tanh r}{\cosh^2 r} r_z^2 - 3\sinh r\cosh r + \sinh^2 r\tanh r\bigg)\tilde{\lambda}_1 + \bigg(\frac{2r_z^2}{\cosh^2 r} + \cosh^2 r + \sinh^2 r\bigg)(r_z^2 + \cosh^2 r)\bigg] \Big/ (\tanh r\tilde{\lambda}_1 + r_z^2 + \cosh^2 r)^2$$

satisfy

$$\inf_{Q_\tau} \alpha(r, r_z, r_{zz}) > \mu > 0 \quad \text{for some } \mu \text{ and } \alpha, \beta, \gamma \in C^\alpha(Q_\tau).$$

By standard theory for linear parabolic PDEs, the linearized equation (3-4) with the initial condition $\tilde{r}_0 \in C^{2+\alpha}(S^1)$ has a unique solution $\tilde{r} \in C^{2+\alpha}(Q_\tau)$. Applying the inverse function theorem for Banach spaces (see [Daskalopoulos and Hamilton 1999, Theorem 8.5]), we conclude that there exists $t_0 > 0$ such that (3-1) has a unique solution $r \in C^{2+\alpha}\big(Q_{t_0}\big)$. $\square$

**Remark.** The fully nonlinear equation (3-1) is, in fact, uniformly parabolic due to $C^1$ and $C^2$ estimates of $r$ (Corollary 5.7).

In [Gulliver and Xu 2009, Theorem 6], it is proved that the solution of (HMCF) exists for infinite time and the evolving surface remains strictly convex. We will restate the theorem dividing it into two parts: the first stating the lower bound of the harmonic mean curvature (HMC) and the second stating its upper bound. We will give the entire proof of the second part since some estimates used in the proof will be improved in Section 4 in order to obtain the asymptotically optimal upper bound for HMC.

**Theorem 3.2** [Gulliver and Xu 2009, Theorem 6]. *Let $N^3$ be a hyperbolic manifold. If the initial surface is strictly convex, then $F(x,t) \geq (\min_{M_0} F)e^{-t}$ as long as the solution of HMCF exists. In other words, the surface remains strictly convex.*

**Theorem 3.3.** *Let $N^3$ be a hyperbolic manifold. Assume that the initial hypersurface is strictly convex and $\max_{\Sigma_0} F < \frac{1}{2}$. Then, the solution of HMCF exists for infinite time and $\max_{\Sigma_t} F \leq Ce^{-t/2}$ for some constant $C$ and for all $t \in [0,\infty)$.*

**Remark.** Note that $f \leq \lambda_1 \leq 2f$ if $\lambda_i > 0$. Therefore, the theorem implies that $\max_{\Sigma_t} \lambda_1 \leq Ce^{-t/2}$, where $\lambda_1$ is the smallest principal curvature. Together with Theorem 3.2, we obtain $C_1 e^{-t} \leq F \leq C_2 e^{-t/2}$.

*Proof.* We find the upper bound for $F$ by analyzing the evolution equation of $F$. We set

$$\mathscr{L} = \frac{\partial F}{\partial h_i^j} \nabla_i \nabla^j,$$

which is an elliptic operator as long as the hypersurface is strictly convex.

$$\begin{aligned}
\frac{\partial F}{\partial t} &= \mathscr{L}(F) + F\langle \dot{F}, W^2\rangle + F\langle \dot{F}^{ij}, R_{i0j0}\rangle \\
&= \mathscr{L}(F) + \sum_i F\frac{\partial f}{\partial \lambda_i}\left(\lambda_i^2 + R_{i0i0}\right) \\
&\leq \mathscr{L}(F) + \sum_i F^3 - \sum_i F\frac{\partial f}{\partial \lambda_i} \\
&= \mathscr{L}(F) + 2F^3 - F^3\sum_i \lambda_i^{-2} \leq \mathscr{L}(F) + 2F^3 - \tfrac{1}{2}F.
\end{aligned}$$

By the maximum principle, we can solve the following ODE and obtain an upper bound for $F(x,t)$:

$$\frac{d\tilde{F}}{dt} = 2\tilde{F}^3 - \tfrac{1}{2}\tilde{F}, \qquad \tilde{F}(0) = \max_{x\in M} F(x,0).$$

The solution of the ODE is $\tilde{F}(t)^{-2} = (\tilde{F}(0)^{-2} - 4)e^t + 4$, so we have $F(x,t) \leq \tilde{F}(t)$ for all $x \in M$ as long as the solution of HMCF exists. For the proof of infinite time existence, see [Gulliver and Xu 2009, Theorem 6]. □

Since disjoint surfaces remain disjoint under HMCF by the maximum principle, given a torus whose axis is a closed geodesic, two perfect tori enclosing it from inside and outside, which are called barriers, will remain disjoint throughout the flow; thus, the radius of the evolving torus is comparable to the radii of the barriers.

**Lemma 3.4.** *Let $r$ be the generating function of an axially symmetric torus evolving by HMCF. Then there exist $C_1$ and $C_2$ such that $C_1 e^{-t} \leq r(x,t) \leq C_2 e^{-t}$ for all $x \in \Sigma_t$ as long as the solution of HMCF exists.*

## 4. Evolution equations

To show that the surface of revolution remains a graph over the closed geodesic, it is sufficient to prove that $v$ remains uniformly bounded for all time. To this end, we first derive the evolution equations of $r$ (Lemma 4.1) and $v$ (Lemma 4.3). From now on we will only consider the case $n = 2$.

**Lemma 4.1.** *The generating function satisfies following evolution equation.*

$$\left(\frac{\partial}{\partial t} - \mathscr{L}\right) r = \left(\lambda_1 \frac{\partial f}{\partial \lambda_1} - f\right) u - \frac{\partial f}{\partial \lambda_1} \frac{b'}{b} u^2 - \frac{\partial f}{\partial \lambda_2} \frac{h'}{h} (1 - u^2).$$

*Proof.* Let us compute $\dfrac{\partial r}{\partial t}$ and $\mathscr{L} r$.

$$\frac{\partial r}{\partial t} = \left\langle \frac{\partial}{\partial r}, \frac{\partial X}{\partial t} \right\rangle = -fu.$$

We choose a geodesic coordinate $\partial_1 = \sigma$, $\partial_2 = E_2$ at a fixed point such that $g_{ij} = \delta_{ij}$ and $h_{ij} = \lambda_i \delta_{ij}$ for $i, j = 1, 2$. Since $\nabla_\sigma \sigma = 0$ and $E_2(r) = 0$,

$$\mathscr{L} r = \dot{F}^{kl} \nabla_k \nabla^l r = \frac{\partial f}{\partial \lambda_1} \nabla_\sigma \nabla_\sigma r + \frac{\partial f}{\partial \lambda_2} \nabla_{E_2} \nabla_{E_2} r = \frac{\partial f}{\partial \lambda_1} \sigma \sigma(r) - \frac{\partial f}{\partial \lambda_2} (\nabla_{E_2} E_2) r.$$

Let us first compute the term $\sigma\sigma(r)$. By (2-1)–(2-5),

$$\sigma(r) = \left\langle \sigma, \frac{\partial}{\partial r} \right\rangle = \frac{r_z}{\sqrt{r_z^2 + b^2}} = -\langle N, E_z \rangle$$

and

$$\sigma\sigma(r) = -\sigma \langle N, E_z \rangle = -\langle \overline{\nabla}_\sigma N, E_z \rangle - \langle N, \overline{\nabla}_\sigma E_z \rangle$$
$$= -\langle \lambda_1 \sigma, E_z \rangle - \left\langle N, -\frac{b'}{\sqrt{r_z^2 + b^2}} E_r \right\rangle = -\lambda_1 u + \frac{b'}{b} u^2. \tag{4-1}$$

On the other hand,

$$-(\nabla_{E_2} E_2) r = -\big\langle \overline{\nabla}_{E_2} E_2, \sigma \big\rangle \sigma(r) = \frac{h'}{h} (1 - u^2). \tag{4-2}$$

Combining (4-1) and (4-2), we obtain

$$\mathscr{L} r = \frac{\partial f}{\partial \lambda_1} \left( -\lambda_1 u + \frac{b'}{b} u^2 \right) + \frac{\partial f}{\partial \lambda_2} \frac{h'}{h} (1 - u^2),$$

and this finishes the proof of the lemma. □

It is straightforward to derive the evolution equation of $\phi(r)$ for a smooth function $\phi : \mathbb{R} \to \mathbb{R}$.

**Lemma 4.2.** *The following evolution equation is satisfied by $\phi\circ r : \Sigma\times[0,\infty)\to\mathbb{R}$:*

$$\left(\frac{\partial}{\partial t}-\mathcal{L}\right)\phi(r)=\phi'\left[\left(\lambda_1\frac{\partial f}{\partial\lambda_1}-f\right)u-\frac{\partial f}{\partial\lambda_1}\frac{b'}{b}u^2-\frac{\partial f}{\partial\lambda_2}\frac{h'}{h}(1-u^2)\right]-\phi''\frac{\partial f}{\partial\lambda_1}(1-u^2).$$

*Proof.* We compute

$$\phi''\dot{F}^{kl}\nabla_k r\nabla_l r=\phi''\frac{\partial f}{\partial\lambda_1}(\sigma(r))^2=\phi''\frac{\partial f}{\partial\lambda_1}\langle E_z,N\rangle^2=\phi''\frac{\partial f}{\partial\lambda_1}(1-u^2).\qquad\square$$

**Lemma 4.3.** *The gradient function $v=u^{-1}$ satisfies the evolution equation*

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\mathcal{L}\right)v=-\frac{2}{v}\dot{f}^{kl}\nabla_k v\nabla_l v-\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)'(v-v^{-1})-\frac{\partial f}{\partial\lambda_1}v\left(v^{-1}\frac{b'}{b}-\lambda_1\right)^2\\+\left(\frac{\partial f}{\partial\lambda_2}\frac{b'h'}{bh}-\frac{\partial f}{\partial\lambda_2}\left(\frac{h'}{h}\right)'+\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)^2\right)v\\-\frac{\partial f}{\partial\lambda_2}\lambda_2\frac{b'}{b}+\left(\frac{\partial f}{\partial\lambda_2}\left(\frac{h'}{h}\right)'-\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)^2\right)v^{-1}.\end{aligned}$$

*Proof.* Let us first compute $\dot{F}^{kl}\nabla_k\nabla^l u$ by choosing the geodesic coordinate at a fixed point as before:

$$\dot{F}^{kl}\nabla_k\nabla_l u=\frac{\partial f}{\partial\lambda_1}\sigma\sigma(u)-\frac{\partial f}{\partial\lambda_2}(\nabla_{E_2}E_2)u.$$

From (2-1) and (2-2), we get $\overline{\nabla}_\sigma E_r=u\dfrac{b'}{b}E_z$. Substituting, we obtain

$$\sigma(u)=\sigma\langle E_r,N\rangle=\langle\overline{\nabla}_\sigma E_r,N\rangle+\langle E_r,\overline{\nabla}_\sigma N\rangle=\left(u\frac{b'}{b}-\lambda_1\right)\langle E_z,N\rangle.\tag{4-3}$$

As preparation for calculating $\sigma\sigma(u)$, we first observe that, by (2-2) and (4-3),

$$\sigma\left(u\frac{b'}{b}\right)=\left[\left(u\frac{b'}{b}-\lambda_1\right)\frac{b'}{b}-u\left(\frac{b'}{b}\right)'\right]\langle E_z,N\rangle.$$

From (2-1) and (2-2), we see $\overline{\nabla}_\sigma E_z=-u(b'/b)E_r$, and get

$$\sigma\langle E_z,N\rangle=\left\langle-u\frac{b'}{b}E_r,N\right\rangle+\langle E_z,\lambda_1\sigma\rangle=-u\left(u\frac{b'}{b}-\lambda_1\right).$$

Then,

$$\sigma\sigma(u)=\left[\left(u\frac{b'}{b}-\lambda_1\right)\frac{b'}{b}-u\left(\frac{b'}{b}\right)'\right](1-u^2)-\sigma(\lambda_1)\langle E_z,N\rangle-u\left(u\frac{b'}{b}-\lambda_1\right)^2,$$

where we used that $\langle E_z,N\rangle^2=1-u^2$. By (2-1) and (4-3), it is straightforward to compute

$$(\nabla_{E_2}E_2)u=\langle\overline{\nabla}_{E_2}E_2,\sigma\rangle\sigma(u)=\frac{h'}{h}\left(u\frac{b'}{b}-\lambda_1\right)(1-u^2).$$

We finally obtain

$$\dot{F}^{kl}\nabla_k\nabla_l u = \frac{\partial f}{\partial \lambda_1}\biggl(\Bigl[\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)\frac{b'}{b} - u\Bigl(\frac{b'}{b}\Bigr)'\Bigr](1-u^2) - \sigma(\lambda_1)\langle E_z, N\rangle - u\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)^2\biggr) - \frac{\partial f}{\partial \lambda_2}\frac{h'}{h}\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)(1-u^2). \tag{4-4}$$

In order to compute $\partial u/\partial t$, we will use the identities

$$\frac{\partial N}{\partial t} = \nabla F, \quad \frac{\partial E_r}{\partial t} = -F\overline{\nabla}_N E_r = -F\langle E_z, N\rangle \frac{b'}{b}E_z.$$

Then,

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{\partial}{\partial t}\langle N, E_r\rangle = \sigma(F)\langle \sigma, E_r\rangle - F(1-u^2)\frac{b'}{b} \\ &= -\frac{\partial f}{\partial \lambda_1}\sigma(\lambda_1)\langle E_z, N\rangle - \frac{\partial f}{\partial \lambda_2}\sigma(\lambda_2)\langle E_z, N\rangle - F(1-u^2)\frac{b'}{b} \\ &= -\frac{\partial f}{\partial \lambda_1}\sigma(\lambda_1)\langle E_z, N\rangle - \frac{\partial f}{\partial \lambda_2}\Bigl[\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)\frac{h'}{h} - u\Bigl(\frac{h'}{h}\Bigr)'\Bigr](1-u^2) \\ &\qquad - F(1-u^2)\frac{b'}{b},\end{aligned} \tag{4-5}$$

where we used $\langle \sigma, E_r\rangle = -\langle E_z, N\rangle$ in the third equation and in the last equation we substituted

$$\sigma(\lambda_2) = \sigma\Bigl(u\frac{h'}{h}\Bigr) = \Bigl[\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)\frac{h'}{h} - u\Bigl(\frac{h'}{h}\Bigr)'\Bigr]\langle E_z, N\rangle.$$

From (4-4) and (4-5), we derive

$$\Bigl(\frac{\partial}{\partial t} - \dot{F}^{kl}\nabla_k\nabla_l\Bigr)u = \frac{\partial f}{\partial \lambda_2}u(1-u^2)\Bigl(\frac{h'}{h}\Bigr)' - F(1-u^2)\frac{b'}{b} - \frac{\partial f}{\partial \lambda_1}\Bigl[\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)\frac{b'}{b} - u\Bigl(\frac{b'}{b}\Bigr)'\Bigr](1-u^2) + \frac{\partial f}{\partial \lambda_1}u\Bigl(u\frac{b'}{b}-\lambda_1\Bigr)^2.$$

By the definition of $v = u^{-1}$, we have $\dot{F}^{kl}\nabla_k\nabla_l u = -\dfrac{1}{v^2}\dot{F}^{kl}\nabla_k\nabla_l v + \dfrac{2}{v^3}\dot{F}^{kl}\nabla_k v\nabla_l v$ and $\partial u/\partial t = -(1/v^2)\partial v/\partial t$. Hence

$$\begin{aligned}\Bigl(\frac{\partial}{\partial t} - \dot{F}^{kl}\nabla_k\nabla_l\Bigr)v &= -\frac{2}{v}\dot{F}^{kl}\nabla_k v\nabla_l v - \frac{\partial f}{\partial \lambda_1}\Bigl(\frac{b'}{b}\Bigr)'(v-v^{-1}) \\ &\quad - \frac{\partial f}{\partial \lambda_1}v\Bigl(v^{-1}\frac{b'}{b}-\lambda_1\Bigr)^2 - \frac{\partial f}{\partial \lambda_2}\Bigl(\frac{h'}{h}\Bigr)'(v-v^{-1}) \\ &\quad + f\frac{b'}{b}(v^2-1) + \frac{\partial f}{\partial \lambda_1}\frac{b'}{b}\Bigl(v^{-1}\frac{b'}{b}-\lambda_1\Bigr)(v^2-1).\end{aligned} \tag{4-6}$$

Combining $v^2$ terms in the second line of (4-6) and applying Euler's identity

$\frac{\partial f}{\partial \lambda_1}\lambda_1 + \frac{\partial f}{\partial \lambda_2}\lambda_2 = f$ and (2-4), the $v^2$ term can be reduced to a linear term:

$$\left( f\frac{b'}{b} - \frac{\partial f}{\partial \lambda_1}\frac{b'}{b}\lambda_1 \right) v^2 = \left( f\frac{b'}{b} - \frac{b'}{b}\left( f - \frac{\partial f}{\partial \lambda_2}\lambda_2 \right) \right) v^2 = \frac{b'}{b}\frac{\partial f}{\partial \lambda_2}\lambda_2 v^2 = \frac{b'}{b}\frac{\partial f}{\partial \lambda_2}\frac{h'}{h} v.$$

We then obtain the evolution equation of $v$ as stated in the lemma. $\square$

## 5. Preserving the property of being a graph and curvature estimates

In this section, we study HMCF solutions of an axially symmetric torus centered at a closed geodesic satisfying the hypothesis of Theorem 1.1: the initial surface is strictly convex and $\max_{\Sigma_0} F < \frac{1}{2}$. Since we will prove many technical estimates, we take this opportunity to outline the overall argument. The main goal of this section is to prove that the evolving surface stays as a graph as it converges to the closed geodesic. As discussed in Section 3, this is equivalent to showing that $v = u^{-1}$ is uniformly bounded for all time (Theorem 5.5). However, we cannot prove the uniform boundedness of $v$ directly using its evolution equation, so the first step is to obtain a weak estimate: $vh \le C$ where $h(r) = \sinh r$ (Theorem 5.2). This estimate is weaker than $v < C$ since the graph function $r$, thus $\sinh r$, decays to 0 by the barrier argument in Lemma 3.4. We can then deduce by (2-4) that $\lambda_2 = h'/vh = \cosh r/(v \sinh r)$ is uniformly bounded from below. Then, together with Theorem 3.3 we can estimate the ratio of two principal curvatures: $\lambda_2/\lambda_1 \to \infty$ as $t \to \infty$ (Corollary 5.3). Equipped with this new estimate for $\lambda_2/\lambda_1$, we revisit the proof of Theorem 3.3 and obtain the optimal asymptotic upper bound of the HMC (Theorem 5.4): $\lambda_1 \approx e^{-t}$. Finally, we can prove that the gradient function $v$ is uniformly bounded (Theorem 5.5) and deduce that $\lambda_2 \approx e^t$ thanks to the formula (2-4) for $\lambda_2$ available on axially symmetric surfaces. We then conclude in Corollary 5.6 that the principal curvatures of axially symmetric torus behave like those of perfect torus evolving under HMCF as stated in (1-1).

We first consider evolution equations of $\phi(r)v$ where $\phi : \mathbb{R} \to \mathbb{R}$ is a test function to be chosen later.

**Lemma 5.1.** *The evolution equation for $\phi(r)v$ is given by*

$$\begin{aligned}\left(\frac{\partial}{\partial t} - \mathcal{L}\right)\phi v = \phi\biggl( &-\frac{\partial f}{\partial \lambda_1}\left(\frac{b'}{b}\right)'\left(v - \frac{1}{v}\right) - \frac{\partial f}{\partial \lambda_1} v\left(\frac{1}{v}\frac{b'}{b} - \lambda_1\right)^2 - \frac{\partial f}{\partial \lambda_2}\lambda_2\frac{b'}{b} \\ &+ \left[\frac{\partial f}{\partial \lambda_2}\left(\frac{h'}{h}\right)' - \frac{\partial f}{\partial \lambda_1}\left(\frac{b'}{b}\right)^2\right]\frac{1}{v}\biggr) - f\phi' + \left[-\frac{\partial f}{\partial \lambda_2}\left(\frac{h'}{h}\right)' + \frac{\partial f}{\partial \lambda_2}\frac{h'b'}{hb} + \frac{\partial f}{\partial \lambda_1}\left(\frac{b'}{b}\right)^2\right]\phi v \\ &- \frac{\partial f}{\partial \lambda_1}\left(-\lambda_1 + \frac{1}{v}\frac{b'}{b}\right)\phi' - \frac{\partial f}{\partial \lambda_2}\frac{h'}{h}\left(v - \frac{1}{v}\right)\phi' - \phi''\frac{\partial f}{\partial \lambda_1}\left(v - \frac{1}{v}\right) - \frac{2}{v}\dot{F}^{kl}\nabla_k(\phi v)\nabla_l v.\end{aligned}$$

*Proof.* Apply Lemmas 4.2 and 4.3 to

$$\left(\frac{\partial}{\partial t}-\mathcal{L}\right)\phi v=\phi\left(\frac{\partial}{\partial t}-\mathcal{L}\right)v+v\left(\frac{\partial}{\partial t}-\mathcal{L}\right)\phi-2\dot{F}^{kl}\nabla_k\phi\nabla_l v. \qquad \square$$

**Theorem 5.2.** *We have $hv\le C$ on $M\times[0,\infty)$, where $h(r)=\sinh r$.*

*Proof.* All the terms in the big parentheses straddling the first and second lines of the equation in Lemma 5.1 are nonpositive, as is the subsequent term $-f\phi'$. Substitute $\phi=h$ in that equation. Ignoring all the terms just mentioned since they are nonpositive, we obtain

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\mathcal{L}\right)hv &\le \left[-\frac{\partial f}{\partial\lambda_2}\left(\frac{h'}{h}\right)'+\frac{\partial f}{\partial\lambda_2}\frac{h'b'}{hb}+\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)^2\right]hv-\frac{\partial f}{\partial\lambda_1}\left(\frac{1}{v}\frac{b'}{b}-\lambda_1\right)h'\\ &\qquad -\frac{\partial f}{\partial\lambda_2}\frac{h'}{h}\left(v-\frac{1}{v}\right)h'-h''\frac{\partial f}{\partial\lambda_1}\left(v-\frac{1}{v}\right)-\frac{2}{v}\dot{F}^{kl}\nabla_k(hv)\nabla_l v\\ &=\left[-\frac{\partial f}{\partial\lambda_2}\left(\frac{h'}{h}\right)'+\frac{\partial f}{\partial\lambda_2}\frac{h'b'}{hb}+\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)^2-\frac{h''}{h}\frac{\partial f}{\partial\lambda_1}-\left(\frac{h'}{h}\right)^2\frac{\partial f}{\partial\lambda_2}\right]hv\\ &\qquad+\left(-hh'\frac{\partial f}{\partial\lambda_1}\frac{b'}{b}+hh''\frac{\partial f}{\partial\lambda_1}+h'^2\frac{\partial f}{\partial\lambda_2}\right)\frac{1}{hv}\\ &\qquad\qquad+h'\frac{\partial f}{\partial\lambda_1}\lambda_1-\frac{2}{v}\dot{F}^{kl}\nabla_k(hv)\nabla_l v\\ &=-\frac{hv}{\cosh^2 r}\frac{\partial f}{\partial\lambda_1}+\frac{\cosh^2 r}{hv}\frac{\partial f}{\partial\lambda_2}+\cosh r\frac{\partial f}{\partial\lambda_1}\lambda_1-\frac{2}{v}\dot{F}^{kl}\nabla_k(hv)\nabla_l v.\end{aligned}$$

There exist positive constants $C_0$, $C_1$, and $C_2$ such that

$$-\frac{1}{\cosh^2 r}\frac{\partial f}{\partial\lambda_1}\le -C_0,\quad \cosh^2 r\frac{\partial f}{\partial\lambda_2}\le C_1,\quad \cosh r\frac{\partial f}{\partial\lambda_1}\lambda_1\le C_2,$$

by Theorem 3.3, Lemma 3.4 and the fact that, if $\lambda_1,\lambda_2>0$, then

$$\frac{1}{2}\le\frac{\partial f}{\partial\lambda_1}\le 1\quad\text{and}\quad 0\le\frac{\partial f}{\partial\lambda_2}\le 1. \tag{5-1}$$

The evolution equation becomes

$$\left(\frac{\partial}{\partial t}-\mathcal{L}\right)vh\le -C_0vh+C_1(vh)^{-1}+C_2-\frac{2}{v}\dot{F}^{kl}\nabla_k(hv)\nabla_l v$$

and we can apply the maximum principle to obtain a uniform upper bound for $hv$:

$$\max_{\Sigma_t} hv\le\max\left\{\frac{1}{2C_0}\left(C_2+\sqrt{C_2^2+4C_0C_1}\right),\max_{\Sigma_0}hv\right\}. \qquad \square$$

**Corollary 5.3.** *We have $\lambda_2>C_1$ and $\lambda_2/\lambda_1\ge C_2e^{t/2}$ on $\Sigma_t$ for all $t\in[0,\infty)$.*

*Proof.* The large principal curvature $\lambda_2$ has a uniform lower bound, as can be seen by applying Theorem 5.2 to (2-4). It follows that the ratio $\lambda_2/\lambda_1$ tends to infinity at the rate $e^{t/2}$ since $\lambda_1\le Ce^{-t/2}$ from Theorem 3.3. $\square$

We will use the growth estimate of the ratio $\lambda_2/\lambda_1$ to improve the proof of Theorem 3.3 and squeeze out the optimal upper bound of the harmonic mean curvature $F$. As we shall see below, the ODE associated to the evolution equation of $F$ now has a time dependent coefficient due to the use of growth estimate $\lambda_2/\lambda_1 > Ce^{t/2}$. Therefore, we need to analyze the solution of a nonautonomous ODE in order to establish the optimal upper bound of $F$.

**Theorem 5.4.** *There exist $T > 0$ and $C_1, C_2 > 0$ such that, for all $t \geq T$,*

$$C_1 e^{-t} \leq F \leq C_2 e^{-t}.$$

*Proof.* Since Theorem 3.2 provides the lower bound, it is enough to prove the upper bound. We analyze the evolution equation of the harmonic mean curvature $F$ from Theorem 3.3 again:

$$\begin{aligned}\left(\frac{\partial}{\partial t} - \mathscr{L}\right)F &= F\langle \dot{F}, W^2\rangle + F\langle \dot{F}^{ij}, R_{i0j0}\rangle \\ &= \sum_i F\frac{\partial f}{\partial \lambda_i}(\lambda_i^2 + R_{i0i0}) \\ &= \sum_i F^3 - \sum_i F\frac{\partial f}{\partial \lambda_i} \\ &= 2F^3 - F\left(F^2\sum_i \lambda_i^{-2}\right) \\ &\leq 2F^3 - \delta(t)F,\end{aligned}$$

where

$$\delta(t) = \max\left\{\tfrac{1}{2}, 1 - Ce^{-t/2}\right\}$$

was obtained by observing that $F^2\sum_{i=1}^2 \lambda_i^{-2} = \left(\lambda_1^{-2} + \lambda_2^{-2}\right)/\left(\lambda_1^{-1} + \lambda_2^{-1}\right)^2 \geq \frac{1}{2}$ if $\lambda_i > 0$ and that

$$F^2\sum_{i=1}^2 \lambda_i^{-2} \geq 1 - 2\left(\frac{\lambda_2}{\lambda_1}\right)^{-1} \geq 1 - Ce^{-t/2}$$

due to Corollary 5.3. Then, by the maximum principle, $F(x,t) \leq \psi(t)$ for all $(x,t) \in \Sigma \times [0,\infty)$ where $\psi(t)$ is the solution of following nonautonomous ODE:

$$\frac{d\psi}{dt} = -2\psi\left(\delta(t)/2 - \psi^2\right), \quad \psi(0) = \max_{\Sigma_0} F. \tag{5-2}$$

Since we are interested in the asymptotic decay rate of the harmonic mean curvature, we will find decay rate of $\psi(t)$ for $t \in [T,\infty)$ for large $T$ by comparing the solution of (5-2) with the solutions of (5-3) and (5-4) below. Note that due to the initial condition $\max_{\Sigma_0} F < \frac{1}{2}$ it is not hard to see that $\psi(t) \to 0$ as $t \to \infty$; thus we can choose large $T$ such that $\psi(T)=\epsilon$ for any given $\epsilon>0$. Consider the ODEs

$$\frac{d\overline{\psi}}{dt}=-2\overline{\psi}\big(\delta/2-\epsilon^2\big), \tag{5-3}$$

$$\frac{d\widehat{\psi}}{dt}=-2\widehat{\psi}\big(\delta/2-\overline{\psi}^2\big), \tag{5-4}$$

on the time interval $[T,\infty)$ with conditions $\overline{\psi}(T)=\widehat{\psi}(T)=\epsilon$.

**Claim I.** $\psi\le\widehat{\psi}\le\overline{\psi}$ *for all* $t\in[T,\infty)$.

*Proof.* Since $\psi(T)=\epsilon$ and $\psi$ is nonincreasing for all $t\in[T,\infty)$, from (5-2) and (5-3)

$$\frac{d}{dt}(\log\psi-\log\overline{\psi})=2(\psi^2-\epsilon^2)\le 0.$$

Hence, $\psi\le\overline{\psi}$ on $[T,\infty)$.

Using this result, we see from (5-2) and (5-4) that

$$\frac{d}{dt}(\log\psi-\log\widehat{\psi})=2(\psi^2-\overline{\psi}^2)\le 0.$$

Hence $\psi\le\widehat{\psi}$ on $[T,\infty)$. Finally, from (5-3) and (5-4), we have

$$\frac{d}{dt}(\log\widehat{\psi}-\log\overline{\psi})=2(\overline{\psi}^2-\epsilon^2)\le 0$$

since $\overline{\psi}(T)=\epsilon$ and $\overline{\psi}$ is nonincreasing. Hence, $\widehat{\psi}\le\overline{\psi}$ on $[T,\infty)$. □

**Claim II.** $\widehat{\psi}(t)\le C_3e^{-t}$ *for all* $t\ge T$.

*Proof.* Let us find the exact solutions of (5-3) and (5-4). Noting that $\delta(t)=1-Ce^{-t/2}$ for $t\in[T,\infty)$ when $T$ is large, the solution of (5-3) is

$$\overline{\psi}(t)=\overline{\psi}(T)\exp\big[(-1+2\epsilon^2)t-2Ce^{-t/2}+C_1\big], \tag{5-5}$$

where $C_1=(1-2\epsilon^2)T+2Ce^{-T/2}$.

Next, substituting (5-5) into (5-4) and integrating in time, we obtain

$$\begin{aligned}\log\frac{\widehat{\psi}}{\widehat{\psi}(T)}&=\int_T^t(-1+Ce^{-t/2}+2\overline{\psi}^2)\,dt\\&=-t-2Ce^{-t/2}+T+2Ce^{-T/2}+2\int_T^t\overline{\psi}^2\,dt.\end{aligned}$$

But

$$\int_T^t\overline{\psi}^2\,dt=\overline{\psi}(T)^2\int_T^t\exp\big[2(-1+2\epsilon^2)t-4Ce^{-t/2}+2C_1\big]\,dt\le C_2.$$

Hence,

$$\widehat{\psi}(t)\le C_3e^{-t}.$$ □

Using Claims I and II and the maximum principle, we conclude that

$$\max_{x\in\Sigma_t} F \le \psi(t) \le \widehat{\psi}(t) \le Ce^{-t} \quad \text{for all } t \ge T. \qquad \square$$

We are now in a position to prove that $v$ is uniformly bounded.

**Theorem 5.5.** *There exists a constant $C > 0$ such that $v(x,t) \le C$ for all $(x,t) \in \Sigma \times [0,\infty)$.*

*Proof.* Define a test function $\phi(r) = e^{\mu r^{1+\alpha}}$, where $\mu$ is a positive number to be chosen and $\alpha \in (0,1)$ can be any number. Note that the asymptotic behavior $\phi \to 1$, $\phi' \to 0$, and $\phi'' \to \infty$ as $r \to 0$ becomes important when it comes to obtaining the desired estimates for the reaction terms in the evolution equation of $\phi(r)v$. In particular,

$$\phi''(r) = \mu(1+\alpha)\alpha r^{-1+\alpha}\phi + \big(\mu(1+\alpha)r^{\alpha}\big)^2\phi \ge \mu(1+\alpha)\alpha \max_{\Sigma_0} r^{-1+\alpha} \tag{5-6}$$

is useful since by choosing $\mu$ large, $\phi''$ can be made greater than any large number, but it never becomes infinite in finite time. From Lemma 5.1,

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\mathscr{L}\right)\phi v \le{}& \left[-\frac{\partial f}{\partial\lambda_2}\left(\frac{h'}{h}\right)'+\frac{\partial f}{\partial\lambda_2}\frac{h'b'}{hb}+\frac{\partial f}{\partial\lambda_1}\left(\frac{b'}{b}\right)^2\right]\phi v \\ &-\frac{\partial f}{\partial\lambda_1}\left(v^{-1}\frac{b'}{b}-\lambda_1\right)\phi'-\phi''\frac{\partial f}{\partial\lambda_1}\left(v-v^{-1}\right)-\frac{2}{v}\dot F^{kl}\nabla_k(\phi v)\nabla_l v \\ \le{}& \phi''\Bigg(\left[\frac{1}{\phi''}\frac{\partial f}{\partial\lambda_2}\frac{1}{\sinh^2 r}+\frac{1}{\phi''}\frac{\partial f}{\partial\lambda_2}+\frac{1}{\phi''}\frac{\partial f}{\partial\lambda_1}\frac{\sinh^2 r}{\cosh^2 r}-\frac{1}{\phi}\frac{\partial f}{\partial\lambda_1}\right]\phi v \\ &+\left[-\frac{\phi'\phi}{\phi''}\frac{\partial f}{\partial\lambda_1}\frac{\sinh r}{\cosh r}+\phi\frac{\partial f}{\partial\lambda_1}\right](\phi v)^{-1}+\frac{\phi'}{\phi''}\frac{\partial f}{\partial\lambda_1}\lambda_1\Bigg) \\ &-\frac{2}{v}\dot F^{kl}\nabla_k(\phi v)\nabla_l v.\end{aligned} \tag{5-7}$$

Let us first examine the coefficient of $\phi v$, in the third line of (5-7). Since $F \approx e^{-t}$ by Theorem 5.4, $\sinh r \approx e^{-t}$ by Lemma 3.4, and $\lambda_2 > C$ by Corollary 5.3, the first term is

$$\frac{\partial f}{\partial\lambda_2}\frac{1}{\sinh^2 r} = \frac{f^2\lambda_2^{-2}}{\sinh^2 r} \le C. \tag{5-8}$$

By Lemma 3.4, (5-1), (5-6), and (5-8), we see that the first three terms can be made arbitrarily small if we choose a large $\mu$. On the other hand, the last term in the third line of (5-7) is strictly negative since we can find a constant $C_0 > 0$ such that $\phi^{-1}\,\partial f/\partial\lambda_1 > C_0$; thus there is a constant $C_1 > 0$ such that

$$\frac{1}{\phi''}\frac{\partial F}{\partial\lambda_2}\frac{1}{\sinh^2 r}+\frac{1}{\phi''}\frac{\partial F}{\partial\lambda_2}+\frac{1}{\phi''}\frac{\partial F}{\partial\lambda_1}\left(\frac{\sinh r}{\cosh r}\right)^2-\frac{1}{\phi}\frac{\partial f}{\partial\lambda_1} \le -C_1.$$

Using similar argument, we see that the rest of the terms in the fourth line of (5-7) can be uniformly bounded above, so the evolution equation becomes

$$\left(\frac{\partial}{\partial t} - \mathcal{L}\right)\phi v \le \phi''\big(-C_1 \cdot \phi v + C_2(\phi v)^{-1} + C_3\big) - \frac{2}{v}\dot{F}^{kl}\nabla_k(\phi v)\nabla_l v,$$

and we can apply the maximum principle to conclude that on $\Sigma \times [0, \infty)$,

$$v \le \phi v \le \max\left\{\max_{\Sigma_0} \phi v,\ \frac{C_3 + \sqrt{C_3^2 + 4C_1C_2}}{2C_1}\right\}. \qquad \square$$

Due to the formula (2-4) for $\lambda_2$ available on axially symmetric surfaces, the uniform boundedness of $v$ implies that $\lambda_2 \approx 1/\sinh r \approx e^t$. Together with the asymptotic estimate for $\lambda_1$ from Theorem 5.4, we have shown that the principal curvatures of an axially symmetric torus evolving by HMCF have the same asymptotic curvature estimates as the perfect torus shrinking under HMCF as stated in (1-1).

**Corollary 5.6.** $\lambda_1 \approx e^{-t}$, $\lambda_2 \approx e^t$, *and* $\lambda_1\lambda_2 \approx 1$ *on* $\Sigma \times [0, \infty)$.

Note that uniform boundedness of $v$ implies that $|r_z|$ is uniformly bounded. In fact, more can be said about $|r_z|$ and $|r_{zz}|$ if we apply the results of Theorems 5.4 and 5.5 to the formula (2-3) for $\lambda_1$. Moreover, we can deduce a better estimate for $\lambda_2$.

**Corollary 5.7.** *We have* $\max_{z\in S^1} |r_{zz}| \le C_1 e^{-t}$, $\max_{z\in S^1} |r_z| \le C_2 e^{-t}$, *and*

$$\max_{z\in S^1} |v - 1| \to 0, \qquad \frac{\lambda_2}{\coth r} \to 1 \qquad \text{as } t \to \infty.$$

*Proof.* Solving for $r_{zz}$ in (2-3) , we obtain

$$r_{zz} = \frac{1}{b}\big[-\lambda_1(r_z^2 + b^2)^{3/2} + (2r_z^2 + b^2)b'\big].$$

Using that $|r_z|$ is uniformly bounded and both $\lambda_1$ and $b' = \sinh r$ decrease at the rate $e^{-t}$,

$$|r_{zz}| \le \frac{(r_z^2 + b^2)^{3/2}}{b}\lambda_1 + \frac{2r_z^2 + b^2}{b} b' \le Ce^{-t}.$$

Since $r$ is a function defined on $S^1$, the derivative $r_z$ cannot have a sign; that is, at each time $t$, there is $z_0(t)$ such that $r_z(z_0(t), t) = 0$. Then,

$$\max_{z\in S^1} |r_z(z, t)| = \max_{S^1} \big|r_z(z, t) - r_z(z_0(t), t)\big| \le \max_{S^1} \int_{z_0(t)}^{z} |r_{zz}(s, t)|\, ds \le Ce^{-t}.$$

Now, by the definition of $v$ we see that $v \to 1$ uniformly in space and time, and from the formula (2-4) for $\lambda_2$ we obtain uniform convergence $\lambda_2 \to \coth r$. $\square$

## Acknowledgements

I would like to express my gratitude to my advisor Robert Gulliver for his continuous support and encouragement throughout my Ph.D. studies.

## References

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

[Andrews 1994b] B. Andrews, "Contraction of convex hypersurfaces in Riemannian spaces", *J. Differential Geom.* **39**:2 (1994), 407–431. MR 95b:53044 Zbl 0797.53044

[Andrews et al. 2013] B. Andrews, J. McCoy, and Y. Zheng, "Contracting convex hypersurfaces by curvature", *Calc. Var. Partial Differential Equations* **47**:3-4 (2013), 611–665. MR 3070558 Zbl 06187283

[Cabezas-Rivas and Miquel 2009] E. Cabezas-Rivas and V. Miquel, "Volume-preserving mean curvature flow of revolution hypersurfaces in a rotationally symmetric space", *Math. Z.* **261**:3 (2009), 489–510. MR 2009m:53175 Zbl 1161.53053

[Caputo and Daskalopoulos 2009] M. C. Caputo and P. Daskalopoulos, "Highly degenerate harmonic mean curvature flow", *Calc. Var. Partial Differential Equations* **35**:3 (2009), 365–384. MR 2010f:35218 Zbl 1179.35174

[Daskalopoulos and Hamilton 1999] P. Daskalopoulos and R. Hamilton, "The free boundary in the Gauss curvature flow with flat sides", *J. Reine Angew. Math.* **510** (1999), 187–227. MR 2000g:53081 Zbl 0931.53031

[Daskalopoulos and Hamilton 2006] P. Daskalopoulos and R. Hamilton, "Harmonic mean curvature flow on surfaces of negative Gaussian curvature", *Comm. Anal. Geom.* **14**:5 (2006), 907–943. MR 2007i:53072 Zbl 1127.53058

[Daskalopoulos and Sesum 2010] P. Daskalopoulos and N. Sesum, "The harmonic mean curvature flow of nonconvex surfaces in $\mathbb{R}^3$", *Calc. Var. Partial Differential Equations* **37**:1-2 (2010), 187–215. MR 2010k:53096 Zbl 1189.53064

[Dieter 2005] S. Dieter, "Nonlinear degenerate curvature flows for weakly convex hypersurfaces", *Calc. Var. Partial Differential Equations* **22**:2 (2005), 229–251. MR 2005h:53117 Zbl 1076.53079

[Gulliver and Xu 2009] R. Gulliver and G. Xu, "Examples of hypersurfaces flowing by curvature in a Riemannian manifold", *Comm. Anal. Geom.* **17**:4 (2009), 701–719. MR 2011e:53102 Zbl 1197.53084

[Huisken 1984] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR 86j:53097 Zbl 0556.53001

Received May 23, 2013.

CHRISTOPHER KIM
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
127 VINCENT HALL
206 CHURCH ST. SE
MINNEAPOLIS, MN 55455
UNITED STATES

cmkim@umn.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.135

# COMPOSITION OPERATORS ON STRICTLY PSEUDOCONVEX DOMAINS WITH SMOOTH SYMBOL

HYUNGWOON KOO AND SONG-YING LI

**It is well known that the composition operator $C_\phi$ is unbounded on Hardy and Bergman spaces on the unit ball $B_n$ in $\mathbb{C}^n$ when $n > 1$ for a linear holomorphic self-map $\phi$ of $B_n$. We find a sufficient and necessary condition for a composition operator with smooth symbol to be bounded on Hardy or Bergman spaces over a bounded strictly pseudoconvex domain in $\mathbb{C}^n$. Moreover, we show that this condition is equivalent to the compactness of the composition operator from a Hardy or Bergman space into the Bergman space whose weight is $\frac{1}{4}$ bigger. We also prove that a certain jump phenomenon occurs when the composition operator is not bounded. Our results generalize known results on the unit ball to strictly pseudoconvex domains.**

## 1. Introduction

Let $D$ be a bounded strictly pseudoconvex domain in $\mathbb{C}^n$ with a smooth boundary and let $d(z)$ be the distance from $z \in D$ to $\partial D$. Let $H(D)$ be the set of all holomorphic functions on $D$. For $0 < p < \infty$ and $\alpha > -1$, the weighted Bergman space $A^p_\alpha(D)$ is the space of all $f \in H(D)$ for which

$$\|f\|^p_{A^p_\alpha} = \int_D |f(z)|^p \, dV_\alpha(z) < \infty,$$

where $dV_\alpha(z) = d(z)^\alpha \, dV(z)$ and $dV$ is the Lebesgue measure on $D$. Also, for $0 < p < \infty$, the Hardy space $H^p(D)$ is the space of all $f \in H(D)$ for which

$$\|f\|^p_{H^p} = \lim_{\epsilon \to 0} \int_{\partial D_\epsilon} |f(\zeta)|^p \, d\sigma_\epsilon(\zeta) < \infty,$$

where $\sigma_\epsilon$ is the surface measure on $\partial D_\epsilon = \{z \in D : d(z) = \epsilon\}$. It is well known

H. Koo was supported by NRF of Korea (2012R1A1A2000705). Li was partially supported by the Minjing Scholar Fund from Fujian Normal University, Fujian, China.
*MSC2010:* primary 47B33; secondary 32T15, 32A36.
*Keywords:* composition operator, strictly pseudoconvex domain, boundedness, smooth symbol.

(see [Krantz 2001]) that the admissible limit $f^*(\zeta)$ exists for almost every $\zeta \in \partial D$ when $f \in H^p(D)$ and

$$\|f\|_{H^p}^p = \int_{\partial D} |f^*(\zeta)|^p \, d\sigma_\epsilon(\zeta) < \infty,$$

where $\sigma$ is the surface area measure on $\partial D$. For notational convenience we may view $H^p(D)$ as $A^p_{-1}(D)$.

Let $\phi = (\phi_1, \ldots, \phi_n) : D \to D$ be a holomorphic self-map on $D$. Then $\phi$ induces the composition operator, $C_\phi$, defined on $H(D)$ by

$$C_\phi(f) = f \circ \phi.$$

When $D$ is the unit disk, $\Delta$, in $\mathbb{C}$, every composition operator is bounded on the weighted Bergman spaces and the Hardy spaces by Littlewood's subordination principle. On the other hand, when $D$ is the unit ball, $B_n$, in $\mathbb{C}^n$ with $n \geq 2$, it is known that not every composition is bounded on the weighted Bergman spaces or the Hardy spaces. Among the early examples of unbounded composition operators on $H^p(B_2)$, the example $\phi(z_1, z_2) = (2z_1z_2, 0)$ is due to J.H. Shapiro and the examples $\phi(z_1, z_2) = (\psi(z_1, z_2), 0)$ for $\psi$ inner were given by MacCluer [1984] and Cima, Stanton, and Wogen [Cima et al. 1984]. Other than the Carleson measure characterization there is no satisfactory criteria known for general symbols up to present time. Since a holomorphic linear map $\phi$ can not guarantee $C_\phi$ is bounded on Hardy and Bergman spaces when $n > 1$, one may concentrate on finding a good criteria for smooth holomorphic $\phi \in C^\infty(\bar{B}_n)$ so that $C_\phi$ is bounded on Hardy spaces, $H^2(B_n)$, and Bergman spaces, $A^2(B_n)$.

When $\phi$ is smooth up to the boundary, Warren Wogen [1988] found a necessary and sufficient condition for $C_\phi$ to be bounded on $H^p(B_n)$. This was generalized to $A^p_\alpha(B_n)$ in [Koo and Smith 2007], where the authors also showed what is called the jump phenomenon: if $\phi$ is smooth up to the boundary and $C_\phi$ is not bounded on $A^p_\alpha(B_n)$, then $C_\phi : A^p_\alpha(B_n) \not\to A^p_{\alpha-\epsilon}(B_n)$ for all $0 \leq \epsilon < \frac{1}{4}$. It was also proved [Koo and Park 2010] that the boundedness of $C_\phi : A^p_\alpha(B_n) \to A^p_\alpha(B_n)$ is equivalent to the compactness of $C_\phi : A^p_\alpha(B_n) \to A^p_{\alpha+1/4}(B_n)$ when $\phi$ is smooth up to the boundary. Wogen's original proof [1988] is quite long and involves various local analyses of the inducing map. Koo and Wang [2010] gave a much simpler proof of Wogen's result using certain compactness argument.

In this paper, we generalize the boundedness criteria and the jump phenomenon of composition operators with smooth symbols to bounded strictly pseudoconvex domains in $\mathbb{C}^n$. We adapt the compactness argument of [Koo and Wang 2010] in our proof. Our main theorem is the following, with $Q_\phi(\zeta)$ defined as in (3-1).

**Theorem 1.1.** *Let $0 < p < \infty$ and $\alpha \geq -1$. Let $\phi : D \to D$ be a holomorphic map with $\phi \in C^4(\bar{D})$. Then the following are equivalent.*

(1) $C_\phi : A^p_\alpha(D) \to A^p_\alpha(D)$ *is bounded.*

(2) $C_\phi : A^p_\alpha(D) \to A^p_{\alpha+1/4}(D)$ *is compact.*

(3) $Q_\phi(\zeta) < 1$ *on* $\phi^{-1}(\partial D)$.

*Moreover, if* $C_\phi : A^p_\alpha(D) \not\to A^p_\alpha(D)$, *then* $C_\phi : A^p_\alpha(D) \not\to A^p_{\alpha+\epsilon}(D)$ *for all* $0 < \epsilon < \frac{1}{4}$.

**Remark.** For $\phi(z) = (z_1 + z_2^2/2, 0) : B_2 \to B_2$, we know $C_\phi : A^p_\alpha(B_2) \to A^p_{\alpha+1/4}(B_2)$ is bounded [Koo and Smith 2007] but not compact [Koo and Park 2010].

In Section 2, we review well-known facts on strictly pseudoconvex domains $D$ and Wogen's result on the unit ball. In Section 3, we study local behavior of maps on $D$ which are smooth on $\bar{D}$, especially holomorphic self-maps of $D$. We prove our main theorem in Section 4.

Throughout the paper we use the same letter $C$ to denote various positive constants which may vary at each occurrence but do not depend on the essential parameters. Variables indicating the dependency of constants $C$ will be often specified in parentheses. For nonnegative quantities $X$ and $Y$ the notation $X \lesssim Y$ or $Y \gtrsim X$ means $X \le CY$ for some inessential constant $C$. Similarly, we write $X \approx Y$ if both $X \lesssim Y$ and $Y \lesssim X$ hold.

## 2. Background

***Strictly pseudoconvex domain.*** A $C^2$-domain $D \subset \mathbb{C}^n$ is strictly pseudoconvex if there is a defining function $r \in C^2(\mathbb{C}^n)$ such that

$$D = \{z \in \mathbb{C}^n : r(z) > 0\}$$

and there exists $C > 0$ such that

$$C|w|^2 \le -\sum_{j=1}^{n} \frac{\partial^2 r(\zeta)}{\partial \zeta_i \partial \bar{\zeta}_j} w_i \bar{w}_j \tag{2-1}$$

for all $\zeta \in \partial D$ and for all $w \in \mathbb{C}^n$. For $\epsilon > 0$, let

$$D_\epsilon = \{z \in D : r(z) > \epsilon\}.$$

For $z, w \in \bar{D}$, define a quasimetric $d(z, w)$ by

$$d(z, w) = r(z) + r(w) + \left| \sum_{j=1}^{n} \frac{\partial r(w)}{\partial w_j}(z_j - w_j) \right| + |z - w|^2. \tag{2-2}$$

For $z, w \in \bar{D}$, let

$$X(z, w) = r(w) + \sum_{j=1}^{n} \frac{\partial r(w)}{\partial w_j}(z_j - w_j) + \frac{1}{2} \sum_{j,k=1}^{n} \frac{\partial^2 r(w)}{\partial w_i \partial w_j}(z_j - w_j)(z_k - w_k).$$

Note that, by Taylor expansion of $r$ near $w$, we get

$$r(z) = -r(w) + 2\,\mathrm{Re}\,X(z,w) + \sum_{i,j=1}^{n} \frac{\partial^2 r(w)}{\partial w_i \partial \overline{w}_j}(z_i - w_i)(\bar{z}_j - \overline{w}_j) + O(|z-w|^3).$$

Thus, when $D$ is strictly pseudoconvex and $z \in \overline{D}$ is near $\eta \in \partial D$,

$$\mathrm{Re}\,X(z,\eta) \geq 0 \tag{2-3}$$

by (2-1). Moreover, it is well known from work of C. Fefferman [1974] that there exists $\delta_D > 0$ such that

$$|X(z,w)| \approx d(z,w) \tag{2-4}$$

for all $(z,w) \in R_{\delta_D}$, where

$$R_\delta = \{(z,w) \in \overline{D} \times \overline{D} : r(z) + r(w) + |z-w| < \delta\}.$$

***Carleson measures.*** For any $\zeta \in \partial D$, we can define a Carleson region centered at $\zeta$ with radius $\delta$ by

$$\mathscr{C}(\zeta,\delta) = \{z \in D : d(z,\zeta) < \delta\}.$$

A positive Borel measure $\mu$ on $\overline{D}$ is said to be a *Carleson measure* if there is a constant $M > 0$ such that, for all $\zeta \in \partial D$ and $\delta > 0$,

$$\mu(\overline{\mathscr{C}(\zeta,\delta)}) \leq M\sigma(\overline{\mathscr{C}(\zeta,\delta)} \cap \partial D),$$

and such a measure $\mu$ is said to be a *vanishing Carleson measure* if

$$\lim_{\delta\to 0} \sup_{\zeta\in\partial D} \frac{\mu(\overline{\mathscr{C}(\zeta,\delta)})}{\sigma(\overline{\mathscr{C}(\zeta,\delta)} \cap \partial D)} = 0.$$

Also, for $\alpha > -1$, a positive Borel measure $\mu$ on $D$ is said to be an $\alpha$*-Carleson measure* if there is a constant $M > 0$ such that, for all $\zeta \in \partial D$ and $\delta > 0$,

$$\mu(\mathscr{C}(\zeta,\delta)) \leq M V_\alpha(\mathscr{C}(\zeta,\delta)),$$

and such a measure $\mu$ is said to be a *vanishing* $\alpha$*-Carleson measure* if

$$\lim_{\delta\to 0} \sup_{\zeta\in\partial D} \frac{\mu(\mathscr{C}(\zeta,\delta))}{V_\alpha(\mathscr{C}(\zeta,\delta))} = 0.$$

By [Krantz and Li 1994] the $V_\alpha$-volume of $\mathscr{C}(\zeta,\delta)$ and the surface area of the intersection $\overline{\mathscr{C}(\zeta,\delta)} \cap \partial D$ are

$$V_\alpha(\mathscr{C}(\zeta,\delta)) \approx \delta^{n+1+\alpha} \quad \text{and} \quad \sigma(\overline{\mathscr{C}(\zeta,\delta)} \cap \partial D) \approx \delta^n, \tag{2-5}$$

respectively.

The next theorem follows from Hörmander's work [1967] on Carleson measures, the work on Bergman and Szegő kernels by Fefferman [1974] and Phong and Stein [1977], together with Krantz and Li's [1994; 1995a; 1995b] work on Hardy spaces and Bergman spaces.

**Theorem 2.1.** *Let $D$ be a smooth bounded strictly pseudoconvex domain in $\mathbb{C}^n$, $0 < p < \infty$ and $\alpha > -1$. Let $\mu$ be a positive Borel measure on $\bar{D}$ and $\nu$ a positive Borel measure on $D$.*

(1) *The inclusion $H^p(D) \hookrightarrow L^p(\mu)$ is continuous if and only if $\mu$ is a Carleson measure, and compact if and only if $\mu$ is a vanishing Carleson measure.*

(2) *The inclusion $A^p_\alpha(D) \hookrightarrow L^p(\nu)$ is continuous if and only if $\nu$ is an $\alpha$-Carleson measure, and compact if and only if $\mu$ is a vanishing $\alpha$-Carleson measure.*

Let $\phi : D \to D$ be a holomorphic mapping and, for a holomorphic function $f$ on $D$, let

$$C_\phi(f)(z) = f \circ \phi(z).$$

Since $D$ is bounded, $\phi$ has admissible limit $\phi^*(\zeta)$ almost everywhere in $\partial D$. So, when $\xi \in \partial D$, we define $\phi(\xi) =: \phi^*(\xi)$. Let $\sigma \circ \phi^{-1}$ and $V_\alpha \circ \phi^{-1}$ be the measures on $\bar{D}$ and $D$ defined by

$$\sigma \circ \phi^{-1}(E) = \int_{\phi^{*-1}(E)} d\sigma(\zeta)$$

for all $E \subset \bar{D}$ and

$$V_\alpha \circ \phi^{-1}(E) = \int_{\phi^{-1}(E)} dV_\alpha(z)$$

for all $E \subset D$, respectively. Then, by a change of variables, we have

$$\int_{\partial D} |C_\phi f(\zeta)|^p \, d\sigma(\zeta) = \int_{\bar{D}} |f(z)|^p \, d\sigma \circ \phi^{-1}(z)$$

and

$$\int_D |C_\phi f(z)|^p \, dV_\alpha(z) = \int_D |f(z)|^p \, dV_\alpha \circ \phi^{-1}(z).$$

Therefore, as a corollary of Theorem 2.1 we have the following characterization.

**Corollary 2.2.** *Let $0 < p < \infty$, $\alpha, \beta > -1$, and $\phi : D \to D$ be a holomorphic mapping.*

(1) *$C_\phi : H^p(D) \to H^p(D)$ is bounded if and only if $\sigma \circ \phi^{-1}$ is a Carleson measure, and compact if and only if $\sigma \circ \phi^{-1}$ is a vanishing Carleson measure.*

(2) *$C_\phi : H^p(D) \to A^p_\alpha(D)$ is bounded if and only if $V_\alpha \circ \phi^{-1}$ is a Carleson measure, and compact if and only if $V_\alpha \circ \phi^{-1}$ is a vanishing Carleson measure.*

(3) *$C_\phi : A^p_\alpha(D) \to A^p_\beta(D)$ bounded if and only if $V_\beta \circ \phi^{-1}$ is an $\alpha$-Carleson measure, and compact if and only if $V_\beta \circ \phi^{-1}$ is a vanishing $\alpha$-Carleson measure.*

***Wogen's theorem.*** Let $\phi : B_n \to B_n$ be holomorphic and $\phi \in C^4(\bar{B}_n)$. Then Wogen proved [1988] the following characterization for $C_\phi$ to be bounded in $H^2(B_n)$, which was generalized by Koo and Smith to $A^p_\alpha(B_n)$ [2007], and by Koo and Park to holomorphic Sobolev spaces [2010]. For $z, \zeta \in \mathbb{C}^n$ and a smooth function $g$, let

$$\mathfrak{D}_\zeta g(z) = \sum_{j=1}^{n} \zeta_j \frac{\partial g}{\partial z_j}(z) \quad \text{and} \quad \mathfrak{D}_{\bar{\zeta}} g(z) = \sum_{j=1}^{n} \bar{\zeta}_j \frac{\partial g}{\partial \bar{z}_j}(z). \tag{2-6}$$

For $z, w \in \mathscr{C}^n$, let $\langle z, w\rangle$ be the Hermitian inner product defined by

$$\langle z, w\rangle = \sum_{j=1}^{n} z_j \bar{w}_j.$$

**Theorem 2.3.** *Let $\phi : B_n \to B_n$ be holomorphic and $\phi \in C^4(\bar{B}_n)$. Let $0 < p < \infty$, $\alpha \geq -1$. For $\eta \in \partial B_n$, let $H_\eta(z) = \langle \phi(z), \eta\rangle$. Then $C_\phi : A^p_\alpha(B_n) \to A^p_\alpha(B_n)$ is bounded if and only if*

$$|\mathfrak{D}_{\tau\tau} H_\eta(\zeta)| < \mathfrak{D}_\zeta H_\eta(\zeta)$$

*for all $\zeta, \eta, \tau \in \partial B_n$ such that*

$$\zeta \in \phi^{-1}(\partial B_n), \quad \eta = \phi(\zeta), \quad \langle \zeta, \tau\rangle = 0.$$

Koo and Smith [2007] proved that the following jump phenomenon occurs when $C_\phi$ is not bounded.

**Theorem 2.4.** *Let $\phi : B_n \to B_n$ be holomorphic and $\phi \in C^4(\bar{B}_n)$. Let $0 < p < \infty$, $\alpha \geq -1$. If $C_\phi$ is not bounded on $A^p_\alpha(B_n)$, then $C_\phi : A^p_\alpha(B_n) \not\to A^p_{\alpha+\epsilon}(B_n)$ for all $0 \leq \epsilon < \frac{1}{4}$.*

The following was proved for the critical index $\epsilon = \frac{1}{4}$ [Koo and Park 2010].

**Theorem 2.5.** *Let $\phi : B_n \to B_n$ be holomorphic and $\phi \in C^4(\bar{B}_n)$. Let $0 < p < \infty$ and $\alpha \geq -1$. Then $C_\phi : A^p_\alpha(B_n) \to A^p_\alpha(B_n)$ is bounded if and only if $C_\phi : A^p_\alpha(B_n) \to A^p_{\alpha+1/4}(B_n)$ is compact.*

## 3. Local estimates of smooth holomorphic maps on $D$

Throughout this section we assume that $\phi : D \to D$ is a holomorphic mapping with $\phi \in C^4(\bar{D})$ where $D$ is a bounded strictly pseudoconvex domain with a smooth boundary. For $z \in \mathbb{C}^n$, we use the following notation:

$$z = (z_1, z_2, \dots, z_n) = (z_1, z') = (z_1, z_2, z''), \quad z_j = x_j + iy_j \ (1 \leq j \leq n).$$

For $w$ near $\partial D$, let

$$\nu(w) = |\partial r(w)|^{-1} \partial r(w),$$

where

$$\partial r(z)=\left(\frac{\partial r(z)}{\partial z_1},\ldots,\frac{\partial r(z)}{\partial z_n}\right).$$

For $\eta\in\partial D$, let

$$\phi_\eta(z)=X(\phi(z),\eta)$$

and let

$$Q_\phi(\zeta,\eta)=\sup_\tau\left\{\left|\frac{\mathfrak{D}^2_{\tau\tau}\phi_\eta(\zeta)}{\mathfrak{D}_{\nu(\zeta)}\phi_\eta(\zeta)}-\frac{\mathfrak{D}^2_{\tau\tau}r(\zeta)}{|\partial r(\zeta)|}\right|\cdot\frac{|\partial r(\zeta)|}{|\mathfrak{D}^2_{\tau\bar\tau}r(\zeta)|}:\langle\tau,\nu(\zeta)\rangle=0\right\}.$$

If $\eta=\phi(\zeta)$, we let

$$Q_\phi(\zeta)=Q_\phi(\zeta,\phi(\zeta)).\tag{3-1}$$

For $D=B_n$, it is easy to check that $\phi_\eta=2H_\eta-2$ and the condition on Theorem 2.3 is equivalent to $Q_\phi(\zeta)<1$ for all $\zeta\in\phi^{-1}(\partial D)$.

**Proposition 3.1.** *Let $\zeta\in\partial D$ and $\eta=\phi(\zeta)\in\partial D$. Then*

(1) $\mathfrak{D}_{\nu(\zeta)}\phi_\eta(\zeta)>0$,

(2) $\mathfrak{D}_\tau\phi_\eta(\zeta)=0$ *for all $\tau$ with* $\langle\nu(\zeta),\tau\rangle=0$,

(3) $Q_\phi(\zeta)\le 1$.

*Proof.* Let $\zeta,\eta\in\partial D$, and $\langle\nu(\zeta),\tau\rangle=0$. Without loss of generality, we may choose local coordinates near $(\zeta,\eta)\in\partial D\times\partial D\subset\mathbb{C}^{2n}$ such that

$$\zeta=\eta=(0,\ldots,0),\quad \nu(\zeta)=\nu(\eta)=(1,0,\ldots,0),\quad \tau=(0,1,0,\ldots,0).$$

For $1\le i,j\le n$, let

$$r_i=\frac{\partial r(\zeta)}{\partial z_i},\quad r_{ij}=\frac{\partial^2 r(\zeta)}{\partial z_i\partial z_j},\quad r_{i\bar j}=\frac{\partial^2 r(\zeta)}{\partial z_i\partial\bar z_j},$$

and let

$$a_i=\frac{\partial r(\eta)}{\partial z_i},\quad a_{ij}=\frac{\partial^2 r(\eta)}{\partial z_i\partial z_j}.$$

Also, for $1\le i,j,\ell\le n$, let

$$b_i^\ell=\frac{\partial\phi_\ell(\zeta)}{\partial z_i},\quad b_{ij}^\ell=\frac{\partial^2\phi_\ell(\zeta)}{\partial z_i\partial z_j}.$$

From the definition of $X$, we have

$$\begin{aligned}\phi_\eta(z)&=:X(\phi(z),\eta)\\&=\sum_{j=1}^n\frac{\partial r(\eta)}{\partial\eta_j}(\phi_j(z)-\eta_j)+\frac12\sum_{i,j=1}^n\frac{\partial^2 r(\eta)}{\partial\eta_i\partial\eta_j}(\phi_i(z)-\eta_i)(\phi_j(z)-\eta_j),\end{aligned}$$

and thus

$$\phi_\eta(z) = a_1\phi_1(z) + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\phi_i(z)\phi_j(z). \tag{3-2}$$

Since the harmonic function $\operatorname{Re}\phi_1$ takes a minimum at $\zeta$ and $\nu(\zeta)$ is the inward normal vector at $\zeta \in \partial D$, by Hopf's lemma, we have

$$b_1^1 = \frac{\partial\phi_1(\zeta)}{\partial\zeta_1} = \frac{\partial \operatorname{Re}\phi_1}{\partial x_1}(\zeta) > 0. \tag{3-3}$$

Since $\nu(\zeta) = (1, 0, \dots, 0)$, for $z$ near $\zeta$

$$r(z) = 2r_1x_1 + O(|z|^2) \quad (r_1 > 0).$$

Therefore, there are $\epsilon, \delta > 0$ such that

$$z = (x_1, z') \in D \quad \text{if } 0 < x_1 \le \delta \quad \text{and} \quad |z'|^2 = \epsilon|x_1|.$$

Then, for all $(x_1, z')$ with $0 < x_1 \le \delta$ and $|z'|^2 = \epsilon|z_1|$, we have

$$0 \le \operatorname{Re}\phi_1(x_1, z') = \operatorname{Re}\Bigl(b_1^1x_1 + \sum_{j=2}^{n} b_j^1 z_j\Bigr) + O(|z|^2).$$

From this, we can easily deduce that

$$b_j^1 = \frac{\partial\phi_1(\zeta)}{\partial\zeta_j} = 0 \quad (2 \le j \le n). \tag{3-4}$$

Then, from (3-2), (3-3), and (3-4), we have

$$\begin{aligned}\phi_\eta(z) &= a_1\Bigl(b_1^1z_1 + \frac{1}{2}\sum_{i,j=1}^{n} b_{ij}^1 z_iz_j\Bigr) + \frac{1}{2}\sum_{k,\ell=1}^{n}\Bigl(\sum_{i,j=1}^{n} a_{ij}b_k^ib_\ell^j\Bigr)z_kz_\ell + O(|z|^3)\\ &= a_1b_1^1\Bigl[z_1 + \frac{1}{2a_1b_1^1}\sum_{i,j=1}^{n}\Bigl[a_1b_{ij}^1 + \sum_{k,\ell=1}^{n} a_{k\ell}b_i^kb_j^\ell\Bigr]z_iz_j\Bigr] + O(|z|^3).\end{aligned}$$

From this we easily conclude (1) and (2).

For (3), let

$$c_{ij} = \frac{r_1}{2a_1b_1^1}\Bigl[a_1b_{ij}^1 + \sum_{k,\ell=1}^{n} a_{k\ell}b_i^kb_j^\ell\Bigr] - \frac{r_{ij}}{2}. \tag{3-5}$$

Then we get

$$\phi_\eta(z) = \frac{a_1 b_1^1}{r_1}\Big[r_1 z_1 + \frac{1}{2}\sum_{i,j=1}^n r_{ij} z_i z_j + \frac{1}{2}\sum_{i,j=1}^n r_{i\bar{j}} z_i \bar{z}_j\Big] + \frac{a_1 b_1^1}{r_1}\Big[\sum_{i,j=1}^n c_{ij} z_i z_j - \frac{1}{2}\sum_{i,j=1}^n r_{i\bar{j}} z_i \bar{z}_j\Big] + O(|z|^3). \tag{3-6}$$

Note that, for $z$ near $\zeta$,

$$r(z) = 2\,\mathrm{Re}\Big(r_1 z_1 + \frac{1}{2}\sum_{i,j=1}^n r_{ij} z_i z_j + \frac{1}{2}\sum_{i,j=1}^n r_{i\bar{j}} z_i \bar{z}_j\Big) + O(|z|^3).$$

Now consider a point $(s, te^{i\theta}, 0'')$ near $\zeta$, with $s, t \geq 0$. (Here and below, $0''$ stands for the origin in $\mathscr{C}^{n-2}$; see start of Section 3.) We have

$$r(s, te^{i\theta}, 0'') = 2r_1 s + (\mathrm{Re}(r_{22}e^{2i\theta}) + r_{2\bar{2}})t^2 + O(s^2 + st + t^3),$$

and thus

$$r(s, te^{i\theta}, 0'') \approx t^{5/2} \quad \text{if } s = t^{5/2} - \frac{1}{2r_1}\big(\mathrm{Re}(r_{22}e^{2i\theta}) + r_{2\bar{2}}\big)t^2. \tag{3-7}$$

Then, with $z := (s, te^{i\theta}, 0'')$, by (2-3) and (3-6), we have

$$\begin{aligned} 0 \leq \mathrm{Re}\,\phi_\eta(z) &= \frac{a_1 b_1^1}{2r_1} r(z) + \frac{a_1 b_1^1}{r_1}\mathrm{Re}\big(c_{22}t^2 e^{2i\theta} - \tfrac{1}{2} r_{2\bar{2}} t^2\big) + O(t^3) \\ &= \frac{a_1 b_1^1}{r_1}\mathrm{Re}\big(c_{22}e^{2i\theta} - \tfrac{1}{2} r_{2\bar{2}}\big)t^2 + O(t^{5/2}) \end{aligned}$$

for all $\theta$. Thus

$$\mathrm{Re}\big(c_{22}e^{2i\theta} - \tfrac{1}{2} r_{2\bar{2}}\big) \geq 0, \qquad \theta \in [0, 2\pi].$$

This implies

$$|c_{22}| \leq -\frac{r_{2\bar{2}}}{2}.$$

Since $\nu(\zeta) = (1, 0, \ldots, 0)$ and $\tau = (0, 1, 0, \ldots, 0)$, by (3-6) we have

$$c_{22} = r_1 \frac{1}{2}\frac{\partial^2 \phi_\eta(\zeta)}{\partial\zeta_2\partial\zeta_2}\Big(\frac{\partial\phi_\eta(\zeta)}{\partial\zeta_1}\Big)^{-1} - \frac{r_{22}}{2} = \frac{|\partial r(\zeta)|}{2}\Big(\frac{\mathscr{D}^2_{\tau\tau}\phi_\eta(\zeta)}{\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta)} - \frac{\mathscr{D}^2_{\tau\tau} r(\zeta)}{|\partial r(\zeta)|}\Big).$$

Therefore, we have

$$\frac{|\partial r(\zeta)|}{2}\left|\frac{\mathscr{D}^2_{\tau\tau}\phi_\eta(\zeta)}{\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta)} - \frac{\mathscr{D}^2_{\tau\tau} r(\zeta)}{|\partial r(\zeta)|}\right| = |c_{22}| \leq -\frac{1}{2}\frac{\partial^2 r(\zeta)}{\partial z_2 \partial\bar{z}_2} = -\tfrac{1}{2}\mathscr{D}^2_{\tau\bar{\tau}} r(\zeta). \qquad \square$$

The following lemma is the key local estimate for the proof of (3) $\Longrightarrow$ (1) of Theorem 1.1. First we introduce some notation. For $\delta > 0$, let

$$\begin{aligned}
V_\delta &= \{\xi \in \partial D : |X(\xi, \zeta)| < \delta \text{ for some } \zeta \in \phi^{-1}(\partial D)\},\\
W_\delta &= \{\eta \in \partial D : |X(\eta, \phi(\zeta))| < \delta \text{ for some } \zeta \in \phi^{-1}(\partial D)\},\\
K &= \{(\zeta, \phi(\zeta)) \in \partial D \times \partial D : \zeta \in \phi^{-1}(\partial D)\},\\
K_\delta &= \{(z, \eta) \in \bar{D} \times \partial D : |X(z, \zeta)| + |X(\phi(\zeta), \eta)| < \delta, \zeta \in \phi^{-1}(\partial D)\}.
\end{aligned}$$

**Lemma 3.2.** *Suppose $Q_\phi(\xi) < 1$ on $\phi^{-1}(\partial D)$. Then there are $\delta > 0$ and $C > 1$ such that, for all $(z, \eta) \in K_\delta$,*

$$\frac{1}{C}\big(|X(\phi(\zeta), \eta)| + |X(z, \zeta)|\big) \le |X(\phi(z), \eta)| \le C\big(|X(\phi(\zeta), \eta)| + |X(z, \zeta)|\big), \tag{3-8}$$

*where the point $\zeta \in \partial D$ is defined by the relation*

$$\min\{|X(\phi(w), \eta)| : w \in \overline{O_z}\} = |X(\phi(\zeta), \eta)|$$

*and $O_z$ is the connected component of $\phi^{-1}(\mathscr{C}(\eta, \delta))$ containing $z$.*

*Proof.* Since $\phi \in C^2(\bar{D})$, there are $\epsilon, \delta > 0$ such that $Q_\phi(z, \eta) \le 1 - \epsilon$ for all $(z, \eta) \in K_\delta$. Fix $(z, \eta) \in K_\delta$ and let $\zeta$ be any point such that

$$\min\{|X(\phi(w), \eta)| : w \in\} = |X(\phi(\zeta), \eta)|.$$

Note that $\zeta \in \partial D$, since $\phi_\eta(w) = X(\phi(w), \eta)$ is an open map as a holomorphic function on $D$. Without loss of generality, we may choose local coordinates near $(\zeta, \eta) \in \partial D \times \partial D \subset \mathbb{C}^{2n}$ as in the proof of Proposition 3.1 so that

$$\zeta = \eta = (0, \dots, 0), \qquad \nu(\zeta) = \nu(\eta) = (1, 0, \dots, 0).$$

Then, by Taylor expansion of $\phi_\eta$ at $\zeta$, we have

$$\phi_\eta(z) = \phi_\eta(\zeta) + \sum_{j=1}^{n} a_j z_j + \frac{1}{2}\sum_{i,j=2}^{n} a_{ij}\, z_i z_j + O(|z_1|^2 + |z_1||z'| + |z'|^3).$$

By Proposition 3.1(1), we have $\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta) > 0$ when $\eta = \phi(\zeta)$. Therefore, by shrinking $\delta$ if necessary, we may assume that $\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta) \ne 0$ for all $(\zeta, \eta) \in K_\delta$, and thus

$$a_1 = \frac{\partial \phi_\eta}{\partial z_1}(\zeta) = \mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta) \ne 0.$$

Since $\zeta$ is the local minimum point of $|\phi_\eta|$, by Taylor expansion of $\phi_\eta(z)$ at $\zeta$ with $z = (s, te^{i\theta}, 0'')$ as in (3-7), we see that

$$a_j = \frac{\partial \phi_\eta}{\partial z_j}(\zeta) = 0 \quad \text{ if } j \ge 2.$$

Thus we have

$$\phi_\eta(z)=\phi_\eta(\zeta)+a_1z_1+\frac{1}{2}\sum_{i,j=2}^{n}a_{ij}\,z_iz_j+O(|z_1|^2+|z_1||z'|+|z'|^3). \tag{3-9}$$

Note that by assumption we have $Q_\phi(\zeta,\eta)\le 1-\epsilon$, since $(\zeta,\eta)\in K_\delta$. Define $F$ and $G$ on $\mathbb{C}^{n-1}$ by

$$F(z')=\frac{1}{2}\sum_{i,j=2}^{n}\left(\frac{a_{ij}}{a_1}-\frac{r_{ij}}{r_1}\right)z_iz_j,\qquad G(z')=-(1-\epsilon)\sum_{i,j=2}^{n}\frac{r_{i\bar{j}}}{r_1}\,z_i\bar{z}_j.$$

Then the condition $Q_\phi(\zeta,\eta)\le 1-\epsilon$ implies $|\mathcal{D}_{\tau'\tau'}F|\le\mathcal{D}_{\tau'\bar{\tau}'}G$ for all $\tau'\in\mathbb{C}^{n-1}$. But straightforward calculations show that

$$\mathcal{D}_{\tau'\tau'}F(z')=2F(\tau'),\qquad \mathcal{D}_{\tau'\bar{\tau}'}G(z')=G(\tau').$$

Therefore, we have

$$\left|\sum_{i,j=2}^{n}\left(\frac{a_{ij}}{a_1}-\frac{r_{ij}}{r_1}\right)z_iz_j\right|\le-(1-\epsilon)\sum_{i,j=2}^{n}\frac{r_{i\bar{j}}}{r_1}\,z_i\bar{z}_j.$$

Since $D$ is strictly pseudoconvex, from this inequality together with (2-1), we have

$$-\sum_{i,j=2}^{n}\frac{r_{i\bar{j}}}{r_1}\,z_i\bar{z}_j-\left|\sum_{i,j=2}^{n}\left(\frac{a_{ij}}{a_1}-\frac{r_{ij}}{r_1}\right)z_iz_j\right|\ge\epsilon C|z'|^2.$$

Therefore, by (3-9) we have

$$\begin{aligned}
&|\mathrm{Re}(\phi_\eta(z)-\phi_\eta(\zeta))|\\
&\ge|a_1|\,\mathrm{Re}\left(z_1+\frac{1}{2}\sum_{i,j=2}^{n}\frac{r_{ij}}{r_1}z_iz_j+\frac{1}{2}\sum_{i,j=2}^{n}\frac{r_{i\bar{j}}}{r_1}\,z_i\bar{z}_j\right)\\
&\qquad-|a_1|\left(\frac{1}{2}\sum_{i,j=2}^{n}\frac{r_{i\bar{j}}}{r_1}z_i\bar{z}_j+\frac{1}{2}\left|\sum_{i,j=2}^{n}\left(\frac{a_{ij}}{a_1}-\frac{r_{ij}}{r_1}\right)z_iz_j\right|\right)+O(|z_1|^2+|z_1||z'|+|z'|^3)\\
&\ge\frac{|a_1|}{2r_1}r(z)+|a_1|\frac{\epsilon C|z'|^2}{2}+O(|z_1|^2+|z_1||z'|+|z'|^3).
\end{aligned}$$

Since $|\phi_\eta(z)-\phi_\eta(\zeta)|\lesssim|\phi_\eta(z)-\phi_\eta(\zeta)|+|\mathrm{Re}(\phi_\eta(z)-\phi_\eta(\zeta))|$, by (3-9) we then have

$$|\phi_\eta(z)-\phi_\eta(\zeta)|\gtrsim\left|a_1z_1+\frac{1}{2}\sum_{i,j=2}^{n}a_{ij}z_iz_j\right|+|z'|^2+O(|z_1|^2+|z_1||z'|+|z'|^3).$$

Since $|a+b|+c > |a|/M + (Mc-|b|)/M$ for any $M \geq 1$, we see that there is $C > 0$ such that

$$|\phi_\eta(z) - \phi_\eta(\zeta)| \geq C(|z_1| + |z'|^2) + O(|z_1|^2 + |z_1||z'| + |z'|^3). \tag{3-10}$$

Note that by (2-4) we have

$$\begin{aligned} |X(z,\zeta)| &\approx d(z,\zeta) \\ &= r(z) + r_1|z_1| + |z'|^2 \\ &\approx |z_1| + |z'|^2 + O(|z_1|^2 + |z_1||z'| + |z'|^3). \end{aligned}$$

Therefore, from (3-10), there exist $C > 1$ (by shrinking $\delta > 0$ if necessary) such that

$$|X(\phi(z),\eta) - X(\phi(\zeta),\eta)| \geq \frac{1}{C}|X(z,\zeta)|, \quad |z| < \delta.$$

Note that if $|X(\phi(\zeta),\eta)| < \dfrac{1}{2C}|X(z,\zeta)|$, the triangular inequality yields

$$|X(\phi(z),\eta)| \gtrsim [|X(\phi(\zeta),\eta)| + |X(z,\zeta)|], \quad |z| < \delta.$$

This inequality also holds when

$$|X(\phi(\zeta),\eta)| \geq \frac{1}{2C}|X(z,\zeta)|,$$

since $|X(\phi(z),\eta)|$ has a minimum at $\zeta$. The constants involved depend continuously on $\eta$ throughout the calculations, and thus, by shrinking $\delta > 0$ again if necessary, there are $C > 0$ and $\delta > 0$ such that

$$|X(\phi(z),\eta)| \geq C[|X(\phi(\zeta),\eta)| + |X(z,\zeta)|] \tag{3-11}$$

for all $(z,\eta) \in K_\delta$.

Since

$$|X(z,\zeta)| \approx |z_1| + |z'|^2 + O(|z_1|^2 + |z_1||z'| + |z'|^3),$$

the converse inequality follows from (3-9). □

We use the same notation as in the proof of Proposition 3.1, and let

$$r_{222} = \frac{\partial^3 r(\zeta)}{\partial z_2^3}, \quad r_{22\bar{2}} = \frac{\partial^3 r(\zeta)}{\partial z_2^2 \partial \bar{z}_2}.$$

We use the following lemma to prove the jump phenomenon when $C_\phi$ is not bounded on $A^p_\alpha(D)$.

**Lemma 3.3.** *Let* $\zeta = (0, \dots, 0) \in \partial D$ *with*

$$\nu(\zeta) = (1, 0, \dots, 0),$$

*and let $R$ be a holomorphic polynomial*

$$R(z_1, z_2) = r_1 z_1 + (r_{12} + r_{1\bar{2}}) z_1 z_2 + \frac{(r_{22} + r_{2\bar{2}})}{2} z_2^2 + \frac{(r_{222} + 3r_{22\bar{2}})}{6} z_2^3. \tag{3-12}$$

*Let $a \in \mathbb{C}$, $b \in \mathbb{R}$, and*

$$g(z) = (1 + a z_2) R(z_1, z_2) + i b z_2^3 + O(|z_1|^2 + |z_2|^4 + |z''|^2).$$

*Then, for $\alpha \geq -1$, there is $C > 0$ such that, for all $\delta > 0$,*

$$V_{\alpha+1/4}(\{z \in D : |g(z)| \leq \delta\}) \geq C \delta^{n+\alpha+1}.$$

*Proof.* It suffices to prove for $\delta > 0$ small, and hence we assume $\delta > 0$ is sufficiently small. For the rest of proof we assume

$$z' = (z_2, z'') \in A_\delta := \{(z_2, z'') \in \mathbb{C}^{n-1} : x_2^4 + y_2^2 + |z''|^2 \leq \delta\}. \tag{3-13}$$

From the fact that $\nu(\zeta) = (1, 0, \dots, 0)$, there are constants $p_j \in \mathbb{R}$ for $1 \leq j \leq 5$ such that

$$\begin{aligned} r(z_1, z_2, z'') = r_1 x_1 + p_1 x_1 x_2 + p_2 y_1 x_2 + p_3 x_2^2 + p_4 x_2^3 + p_5 x_2 y_2 \\ + O(x_1^2 + y_1^2 + y_2^2 + x_2^4 + |z''|^2). \end{aligned} \tag{3-14}$$

Also, there are $q_j \in \mathbb{R}$ for $1 \leq j \leq 5$ such that

$$\begin{aligned} \operatorname{Im}[R(z_1 + i y_1, z_2) + i b z_2^3] = r_1 y_1 + q_1 y_1 x_2 + q_1 x_1 x_2 + q_3 x_2 y_2 + q_4 x_2^2 + q_5 x_2^3 \\ + O(x_1^2 + y_1^2 + y_2^2 + x_2^4), \end{aligned} \tag{3-15}$$

since $|z_1||y_2| + |x_2^2 y_2| = O(x_1^2 + y_1^2 + y_2^2 + x_2^4)$.

Taking $\delta > 0$ sufficiently small if necessary, we may assume $r_1 + p_1 x_2 \geq r_1/2$ and $r_1 + q_1 x_2 \geq r_1/2$. Let $(u, v) = (u(z_2), v(z_2)) \in \mathbb{R}^2$ be the solution of the equations

$$\begin{aligned} 0 &= (r_1 + p_1 x_2) u + p_2 x_2 v + p_3 x_2^2 + p_4 x_2^3 + p_5 x_2 y_2, \\ 0 &= (r_1 + q_1 x_2) v + q_2 x_2 u + q_3 x_2 y_2 + q_4 x_2^2 + q_5 x_2^3. \end{aligned}$$

Since $z' \in A_\delta$, the solution $(u, v)$ always exists and satisfies

$$|u| + |v| \lesssim \delta^{1/2}.$$

Hence, by (3-14) and (3-15), we have

$$r(u + i v, z_2, z'') = O(\delta), \quad \operatorname{Im}[R(u + i v, z_2) + i b z_2^3] = O(\delta). \tag{3-16}$$

By (2-1) we have $r_{2\bar{2}} \in \mathbb{R}$, and thus

$$\operatorname{Re}[r_{2\bar{2}} z_2 (z_2 - \bar{z}_2)] = -2 r_{2\bar{2}} y_2^2.$$

Therefore,

$$
\begin{aligned}
&2\,\mathrm{Re}[R(z_1,z_2)]\\
&\qquad = r(z_1,z_2,0'') + 2\,\mathrm{Re}[r_{1\bar 2} z_1(z_2-\bar z_2)] + \mathrm{Re}[r_{2\bar 2} z_2(z_2-\bar z_2)]\\
&\qquad\qquad\qquad\qquad\qquad\qquad + \mathrm{Re}[r_{22\bar 2} z_2^2(z_2-\bar z_2)] + O(|z_1|^2+|z_2|^4)\\
&\qquad = r(z_1,z_2,0'') - 4y_2\,\mathrm{Im}[r_{1\bar 2} z_1] - 2r_{2\bar 2} y_2^2 - 2y_2\,\mathrm{Re}[r_{22\bar 2} z_2^2] + O(|z_1|^2+|z_2|^4)\\
&\qquad = r(z_1,z_2,0'') + O(|z_1|^2+|z_1y_2|+y_2^2+|y_2|\,|z_2|^2+|z_2|^4)\\
&\qquad = r(z_1,z_2,0'') + O(x_1^2+y_1^2+y_2^2+x_2^4).
\end{aligned}
$$

Therefore, from (3-16) we have

$$2\,\mathrm{Re}[R(u+iv,z_2)] = O(\delta),$$

and thus, from the second equation of (3-16), we have

$$|R(u+iv,z_2)| \approx |\mathrm{Re}[R(u+iv,z_2)]| + |\mathrm{Im}[R(u+iv,z_2)]| = O(\delta).$$

From these estimates we then have

$$
\begin{aligned}
|g(u+iv,z')| &\lesssim |\mathrm{Re}[R(u+iv,z_2)]| + |z_2|\,|R(u+iv,z_2)|\\
&\qquad + |\mathrm{Im}[R(u+iv,z_2)+ibz_2^3]| + O(|u+iv|^2+x_2^4+y_2^2+|z''|^2)\\
&= O(\delta).
\end{aligned}
$$

Since $\partial g(\zeta)/\partial z_1 = r_1$, by taking $\delta$ sufficiently small if necessary, we have

$$z_1 = u(z_2)+iv(z_2)+O(\delta) \implies |g(z)| \lesssim \delta. \tag{3-17}$$

Let

$$B_\delta^C(z_2) := \{z_1 : u(z_2)+C\delta \le x_1 \le u(z_2)+2C\delta,\ v(z_2)\le y_1 \le v(z_2)+\delta\}$$

and

$$\Lambda_\delta^C = \{z : z' \in A_\delta,\ z_1 \in B_\delta^C(z_2)\}.$$

Then, by (3-14), there is $C>0$ such that, for all $z\in\Lambda_\delta^C$, we have

$$r(z)\approx\delta,$$

and from (3-17), for all $z\in\Lambda_\delta^C$, we have

$$|g(z_1,z_2,z'')| \lesssim \delta.$$

Therefore, there are constants $c, C > 0$ such that

$$V_{\alpha+1/4}(\{z\in D : |g(z)|\le\delta\}) \ge V_{\alpha+1/4}(\Lambda_{c\delta}^C) \gtrsim \delta^{\alpha+1/4} V(\Lambda_{c\delta}^C).$$

Since $B_\delta^C(z_2)$ is a rectangle with area $C\delta^2$ for a fixed $z_2$, from the definition of $A_\delta$ in (3-13) we have

$$V_{\alpha+1/4}(\{z \in D : |g(z)| \le \delta\}) \gtrsim \delta^{\alpha+1/4} V(\Lambda_{c\delta}^C) \approx \delta^{\alpha+n+1}.$$

The proof is complete, since the constants suppressed in the inequalities throughout our calculations are independent of $\delta$. □

## 4. Proof of Theorem 1.1

First, we prove the last statement, the jump phenomenon, assuming the equivalence of (1), (2), and (3).

Let $0 < \epsilon < \frac{1}{4}$ and suppose

$$C_\phi : A_\alpha^p(D) \to A_{\alpha+\epsilon}^p(D)$$

is bounded. Then

$$C_\phi : A_\alpha^p(D) \to A_{\alpha+1/4}^p(D)$$

is compact, since the inclusion the map $I : A_{\alpha+\epsilon}^p(D) \hookrightarrow A_{\alpha+1/4}^p(D)$ is compact. Thus, from the equivalence of (1) and (2) we conclude the boundedness of

$$C_\phi : A_\alpha^p(D) \to A_\alpha^p(D).$$

To prove the equivalence of (1), (2), and (3), note that (1) $\Longrightarrow$ (2) is trivial since the inclusion map $I : A_\alpha^p(D) \hookrightarrow A_{\alpha+1/4}^p(D)$ is compact. Thus, it suffices to show that (2) $\Longrightarrow$ (3) and (3) $\Longrightarrow$ (1). First (3) $\Longrightarrow$ (1) follows from the following theorem.

**Theorem 4.1.** *Let $0 < p < \infty$ and $\alpha \ge -1$. Let $\phi : D \to D$ be a holomorphic map with $\phi \in C^4(\bar{D})$. If $Q_\phi(\zeta) < 1$ on $\phi^{-1}(\partial D)$, then $C_\phi$ is bounded on $A_\alpha^p(D)$.*

*Proof.* Let $\mu = \sigma \circ \phi^{-1}$ and $\mu_\alpha = V_\alpha \circ \phi^{-1}$ for $\alpha > -1$. By Corollary 2.2, it suffices to show that there exist $\delta_0 > 0$ and $M > 0$ such that, for all $\eta \in \partial D$ and $0 < \delta < \delta_0$,

$$\mu(\overline{C(\eta, \delta)}) \le M\delta^n \tag{4-1}$$

and

$$\mu_\alpha(C(\eta, \delta)) \le M\delta^{n+1+\alpha}. \tag{4-2}$$

We may assume $\delta > 0$ is sufficiently small, since, otherwise, (4-1) and (4-2) hold trivially. Note that $\phi(D) \cap \partial D = \varnothing$ since $\phi$ is a holomorphic self-map of $D$. Thus $\phi(\bar{D}) \cap [\partial D \setminus V] = \varnothing$ for any neighborhood $V \subset \partial D$ of $\partial D \cap \phi(\partial D)$. By (2-4), with $W_\delta$ as defined right before Lemma 3.2, it suffices to show that there are constants $\delta_1 > 0$ and $\delta_2 > 0$ such that (4-1) and (4-2) hold for all $\delta < \delta_1$ and $\eta \in W_{\delta_2}$. Choose $\delta_1$ and $\delta_2$ small so that Lemma 3.2 holds with $\delta = \delta_0 := (\delta_1 + \delta_2)$, and let $C > 1$ be the corresponding constant in Lemma 3.2.

For $\eta \in W_{\delta_2}$, let $O_j$ be any component of $\phi^{-1}(\mathscr{C}(\eta, \delta_0))$ which also intersects with $\phi^{-1}(\mathscr{C}(\eta, \delta_0/2C))$. Let $\zeta_j \in \overline{O_j}$ be a point such that

$$\min\{|X(\phi(w), \eta)| : w \in \overline{O_j}\} = |X(\phi(\zeta_j), \eta)|.$$

Since $|X(\phi(\zeta_j), \eta)| \le \delta_0/2C$, by (3-8) we have

$$\phi(\mathscr{C}(\zeta_j, \delta_0/2C)) \subset \mathscr{C}(\eta, \delta_0).$$

Therefore, $\mathscr{C}(\zeta_j, \delta_0/2C) \subset O_j$, since $O_j$ is a component which contains $\zeta_j$. This implies that the number of components $O_j$ has an upper bound $M < \infty$ independent of $\eta$, since

$$M\delta_0^{n+1+\alpha} \approx \sum_{j=1}^{M} V_\alpha(\mathscr{C}(\zeta_j, \delta_0/2C)) \le V_\alpha(\phi^{-1}(\mathscr{C}(\eta, \delta_0))) \lesssim 1.$$

Now fix such a component $O_j$ as above. Then, by Lemma 3.2,

$$O_j \cap \phi^{-1}(\mathscr{C}(\eta, \delta)) \subset \mathscr{C}(\zeta_j, C\delta)$$

for all $\delta < \delta_0$.

Then, (4-1) and (4-2) follows immediately since the number of components has a uniform upper bound $M$. □

Next, (2) $\Longrightarrow$ (3) follows from the following theorem together with the Carleson measure criteria, Corollary 2.2.

**Theorem 4.2.** *Let $\phi : D \to D$ be a holomorphic map with $\phi \in C^4(\bar{D})$. Suppose $\zeta, \eta = \phi(\zeta) \in \partial D$ and $Q_\phi(\zeta) = 1$. Then there is $C > 0$ such that, for all $\delta > 0$,*

$$V_{\alpha+1/4}(\phi^{-1}(\mathscr{C}(\eta, \delta))) \ge C\, V_\alpha(\mathscr{C}(\eta, \delta))$$

*and*

$$V_{-3/4} \circ \phi^{-1}(\overline{\mathscr{C}(\eta, \delta)}) \ge C\sigma(\overline{\mathscr{C}(\eta, \delta)} \cap \partial D).$$

*Proof.* For $z \in \mathbb{C}^n$, let $z = (z_1, \ldots, z_n) = (z_1, z') = (z_1, z_2, z'')$. Near $(\zeta, \eta) \in \partial D \times \partial D$, we choose the same coordinates as in the proof of Proposition 3.1 so that

$$\zeta = \eta = (0, \ldots, 0), \qquad \nu(\zeta) = \nu(\eta) = (1, 0, \ldots, 0).$$

By change of coordinates in $z'$ variables if necessary, we may assume $Q_\phi(\zeta) = 1$ for $\tau = (0, 1, 0, \ldots, 0)$, that is,

$$\left| \frac{\mathscr{D}^2_{\tau\tau}\phi_\eta(\zeta)}{\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta)} - \frac{\mathscr{D}^2_{\tau\tau} r(\zeta)}{|\partial r(\zeta)|} \right| \cdot \frac{|\partial r(\zeta)|}{|\mathscr{D}^2_{\tau\bar{\tau}} r(\zeta)|} = 1 \quad (\tau = (0, 1, 0, \ldots, 0)).$$

Since this relation is invariant under rotation in the $z_2$ variable, we may assume

$$\frac{\mathscr{D}^2_{\tau\tau}\phi_\eta(\zeta)}{\mathscr{D}_{\nu(\zeta)}\phi_\eta(\zeta)} - \frac{r_{22}}{r_1} = \frac{r_{2\bar{2}}}{r_1}$$

By (1) and (2) of Proposition 3.1, we have

$$\phi_\eta(z) = a_1 z_1 + \sum_{j=2}^{n} a_{2j} z_2 z_j + a_{32} z_2^3 + O(|z_1|^2 + |z_2|^4 + |z''|^2) \tag{4-3}$$

with $a_1 > 0$. Therefore, the condition $Q_\phi(\zeta) = 1$ is equivalent to

$$\frac{2a_{22}}{a_1} - \frac{r_{22}}{r_1} = \frac{r_{2\bar{2}}}{r_1}. \tag{4-4}$$

Let $R(z_1, z_2)$ be as in (3-12). Then, by (4-3) and (4-4), we get

$$\begin{aligned}\phi_\eta(z) = \frac{a_1}{r_1}(1 + Az_2)R(z_1, z_2) + Bz_2^3 \\ + \sum_{j=3}^{n} a_{2j} z_2 z_j + O\biggl(|z_1|^2 + |z_2|^4 + |z_1|\,|z_3|^2 + \sum_{j=4}^{n} |z_j|^2\biggr),\end{aligned}$$

where

$$A = \frac{a_{12}}{a_1} - \frac{(r_{22} + r_{2\bar{2}})a_{12}}{2r_1}, \quad B = a_{32} - \frac{(r_{222} + 3r_{22\bar{2}})a_1}{6r_1} - A\frac{(r_{22} + r_{2\bar{2}})a_1}{2r_1}.$$

Then, by Lemma 3.3, to complete the proof it suffices to show that

$$\operatorname{Re} B = 0, \quad a_{2j} = 0 \quad (j = 3, \ldots, n).$$

Since $\nu(\zeta) = (1, 0, \ldots, 0)$, for $(s, t) \in \mathbb{R}^2$ we have

$$r(s, t, te^{i\theta}, 0, \ldots, 0) = 2r_1 s + O(s^2 + t^2).$$

Thus, for each $\theta, t \in \mathbb{R}$, there is $s \in \mathbb{R}$ with $|s| \lesssim t^2$ such that $\operatorname{Re}[R(s, t)] = r(s, t, te^{i\theta}, 0, \ldots, 0) = 0$.

Since $\operatorname{Re} \phi_\eta(s, t, te^{i\theta}, 0, \ldots, 0) \ge 0$ by (2-3), we get

$$\begin{aligned} 0 &\le \operatorname{Re} \phi_\eta(s, t, te^{i\theta}, 0, \ldots, 0) \\ &= \operatorname{Re}\biggl[\frac{a_1}{r_1}(1 + At)R(s, t) + Bt^3 + a_{23}t^2 e^{i\theta}\biggr] + O(s^2 + t^4) \\ &= \operatorname{Re}\biggl[\frac{a_1}{r_1} At R(s, t) + Bt^3 + a_{23}t^2 e^{i\theta}\biggr] + O(s^2 + t^4) \\ &= \operatorname{Re}[Bt^3 + a_{23}t^2 e^{i\theta}] + O(s^2 + t^4) \end{aligned}$$

for all $\theta$. This implies $a_{23} = 0$, and, with the same argument, we get

$$a_{2j} = 0 \quad (j = 3, \ldots, n).$$

Also, note that $r(s, \pm t, 0'') = 2r_1 s + O(s^2 + t^2)$ which implies that for each $\pm t$

there is $s = s(\pm t)$ such that $r(s, \pm t, 0'') = 0$ with $|s(\pm t)| \lesssim t^2$. Then, by (2-3), with $s = s(\pm t)$ we have

$$0 \le \operatorname{Re}\phi_\eta(s, \pm t, 0'') = \frac{a_1}{r_1}\operatorname{Re}[R(s, \pm t)] \pm t^3 \operatorname{Re} B + O(t\,|\operatorname{Im}[R(s, \pm t)]| + t^4)$$
$$= \pm t^3 \operatorname{Re} B + O(t^4).$$

Therefore, we get $\operatorname{Re} B = 0$ and the proof is complete. □

## Acknowledgement

Part of this research was performed during the first author's visit to University of California at Irvine. He thanks the Mathematics Department of University California at Irvine for its hospitality and support.

## References

[Cima et al. 1984] J. A. Cima, C. S. Stanton, and W. R. Wogen, "On boundedness of composition operators on $H^2(B_2)$", *Proc. Amer. Math. Soc.* **91**:2 (1984), 217–222. MR 85j:47030 Zbl 0546.47015

[Fefferman 1974] C. Fefferman, "The Bergman kernel and biholomorphic mappings of pseudoconvex domains", *Invent. Math.* **26** (1974), 1–65. MR 50 #2562 Zbl 0289.32012

[Hörmander 1967] L. Hörmander, "$L^p$ estimates for (pluri-) subharmonic functions", *Math. Scand.* **20** (1967), 65–78. MR 38 #2323 Zbl 0156.12201

[Koo and Park 2010] H. Koo and I. Park, "Composition operators on holomorphic Sobolev spaces in $B_n$", *J. Math. Anal. Appl.* **369**:1 (2010), 232–244. MR 2011e:47045 Zbl 1191.47031

[Koo and Smith 2007] H. Koo and W. Smith, "Composition operators induced by smooth self-maps of the unit ball in $\mathscr{C}^N$", *J. Math. Anal. Appl.* **329**:1 (2007), 617–633. MR 2008e:47059 Zbl 1115.32004

[Koo and Wang 2010] H. Koo and M. Wang, "Revisit to a theorem of Wogen", pp. 355–363 in *Topics in operator theory, 1: Operators, matrices and analytic functions*, edited by J. A. Ball et al., Oper. Theory Adv. Appl. **202**, Birkhäuser, Basel, 2010. MR 2012a:47057 Zbl 1221.47044

[Krantz 2001] S. G. Krantz, *Function theory of several complex variables*, AMS Chelsea, Providence, RI, 2001. MR 2002e:32001 Zbl 1087.32001

[Krantz and Li 1994] S. G. Krantz and S.-Y. Li, "A note on Hardy spaces and functions of bounded mean oscillation on domains in $\mathscr{C}^n$", *Michigan Math. J.* **41**:1 (1994), 51–71. MR 95f:32008 Zbl 0802.32013

[Krantz and Li 1995a] S. G. Krantz and S.-Y. Li, "Duality theorems for Hardy and Bergman spaces on convex domains of finite type in $\mathscr{C}^n$", *Ann. Inst. Fourier* (*Grenoble*) **45**:5 (1995), 1305–1327. MR 96m:32002 Zbl 0835.32004

[Krantz and Li 1995b] S. G. Krantz and S.-Y. Li, "On decomposition theorems for Hardy spaces on domains in $\mathscr{C}^n$ and applications", *J. Fourier Anal. Appl.* **2**:1 (1995), 65–107. MR 96m:32003 Zbl 0886.32003

[MacCluer 1984] B. D. MacCluer, "Spectra of compact composition operators on $H^p(B_N)$", *Analysis* **4**:1-2 (1984), 87–103. MR 86e:47038 Zbl 0582.32009

[Phong and Stein 1977] D. H. Phong and E. M. Stein, "Estimates for the Bergman and Szegö projections on strongly pseudo-convex domains", *Duke Math. J.* **44**:3 (1977), 695–704. MR 56 #8916 Zbl 0392.32014

[Wogen 1988] W. R. Wogen, "The smooth mappings which preserve the Hardy space $H^2(B_n)$", pp. 249–263 in *Contributions to operator theory and its applications* (Mesa, AZ, 1987), edited by I. Gohberg et al., Oper. Theory Adv. Appl. **35**, Birkhäuser, Basel, 1988. MR 90k:32018 Zbl 0685.46029

Received September 26, 2012. Revised May 19, 2013.

HYUNGWOON KOO
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEOUL 136-713
SOUTH KOREA
koohw@korea.ac.kr

SONG-YING LI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE, CA 92697
UNITED STATES
sli@math.uci.edu

and

SCHOOL OF MATHEMATICS AND COMPUTER SCIENCES
FUJIAN NORMAL UNIVERSITY
FUJIAN
CHINA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.155

# THE ALEXANDROV PROBLEM IN A QUOTIENT SPACE OF $\mathbb{H}^2 \times \mathbb{R}$

ANA MENEZES

**We prove an Alexandrov-type theorem for a quotient space of $\mathbb{H}^2 \times \mathbb{R}$. More precisely, we classify the compact embedded surfaces with constant mean curvature in the quotient of $\mathbb{H}^2 \times \mathbb{R}$ by a subgroup of isometries generated by a horizontal translation along horocycles of $\mathbb{H}^2$ and a vertical translation. We also construct some examples of periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$ and we prove a multivalued Rado theorem for small perturbations of the helicoid in $\mathbb{H}^2 \times \mathbb{R}$.**

## 1. Introduction

Alexandrov [1962] proved that the only compact embedded constant mean curvature hypersurface in $\mathbb{R}^n$, $\mathbb{H}^n$ and $\mathbb{S}^n_+$ is the round sphere. Since then, many people have proved Alexandrov-type theorems in other spaces.

For instance, W. T. Hsiang and W. Y. Hsiang [1989] showed that a compact embedded constant mean curvature surface in $\mathbb{H}^2 \times \mathbb{R}$ or in $\mathbb{S}^2_+ \times \mathbb{R}$ is a rotational sphere. They used the Alexandrov reflection method with vertical planes in order to prove that for any horizontal direction, there is a vertical plane of symmetry of the surface orthogonal to that direction.

To apply the Alexandrov reflection method we need to start with a vertical plane orthogonal to a given direction that does not intersect the surface, and in $\mathbb{S}^2 \times \mathbb{R}$ this fact is guaranteed by the hypothesis that the surface is contained in the product of a hemisphere with the real line. We remark that in $\mathbb{S}^2 \times \mathbb{R}$, we know that there are embedded rotational constant mean curvature tori, but the Alexandrov problem is not completely solved in $\mathbb{S}^2 \times \mathbb{R}$. In other simply connected homogeneous spaces with 4-dimensional isometry groups ($\mathrm{Nil}_3$, $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$, some Berger spheres), we do not know if the solutions to the Alexandrov problem are spheres.

In $\mathrm{Sol}_3$, Rosenberg proved that an embedded compact constant mean curvature surface is a sphere [Daniel and Mira 2013].

Recently, Mazet, Rodríguez and Rosenberg [Mazet et al. 2011b] considered the quotient of $\mathbb{H}^2 \times \mathbb{R}$ by a discrete group of isometries of $\mathbb{H}^2 \times \mathbb{R}$ generated by

*MSC2010:* 53A10, 53C42.
*Keywords:* constant mean curvature surface, periodic surface, Alexandrov reflection.

a horizontal translation along a geodesic of $\mathbb{H}^2$ and a vertical translation. They classified the compact embedded constant mean curvature surfaces in the quotient space. Moreover, they constructed examples of periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$, where by periodic we mean a surface which is invariant by a nontrivial discrete group of isometries of $\mathbb{H}^2 \times \mathbb{R}$.

In this paper we also consider periodic surfaces in $\mathbb{H}^2 \times \mathbb{R}$. The discrete groups of isometries of $\mathbb{H}^2 \times \mathbb{R}$ we consider are generated by a horizontal translation $\psi$ along horocycles $c(s)$ of $\mathbb{H}^2$ and/or a vertical translation $T(h)$ for some $h > 0$. In the case the group is the $\mathbb{Z}^2$ subgroup generated by $\psi$ and $T(h)$, the quotient space $\mathcal{M} = \mathbb{H}^2 \times \mathbb{R}/[\psi, T(h)]$ is diffeomorphic to $\mathbb{T}^2 \times \mathbb{R}$, where $\mathbb{T}^2$ is the 2-torus. Moreover, $\mathcal{M}$ is foliated by the family of tori $\mathbb{T}(s) = c(s) \times \mathbb{R}/[\psi, T(h)]$ which are intrinsically flat and have constant mean curvature $\frac{1}{2}$. We prove an Alexandrov-type theorem in this quotient space $\mathcal{M}$.

Moreover, in the last part of this paper, we consider a multivalued Rado theorem for small perturbations of the helicoid. Rado's theorem (see [Radó 1930]) is one of the fundamental results of minimal surface theory. It is connected to the famous Plateau problem, and states that if $\Omega \subset \mathbb{R}^2$ is a convex subset and $\Gamma \subset \mathbb{R}^3$ is a simple closed curve which is graphical over $\partial\Omega$, then any compact minimal surface $\Sigma \subset \mathbb{R}^3$ with $\partial\Sigma = \Gamma$ must be a disk which is graphical over $\Omega$, and then unique, by the maximum principle. Dean and Tinaglia [2005] proved a generalization of Rado's theorem. They showed that for a minimal surface of any genus whose boundary is almost graphical in some sense, the minimal surface must be graphical once we move sufficiently far from the boundary. In our work, we consider this problem for minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$ whose boundary is a small perturbation of the boundary of a helicoid, and we prove that the solution to the Plateau problem is the only compact minimal disk with that boundary (see Theorem 2).

This paper is organized as follows. In Section 2, we introduce some notation. In Section 3, we classify the compact embedded constant mean curvature surfaces in the space $\mathcal{M}$, that is, we prove an Alexandrov-type theorem for doubly periodic $H$-surfaces (see Theorem 1). In Section 4, we construct some examples of periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$. In Section 5, we prove a multivalued Rado theorem for small perturbations of the helicoid (see Theorem 2).

## 2. Preliminaries

Throughout this paper, the Poincaré disk model is used for the hyperbolic plane; that is,

$$\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$$

with the hyperbolic metric

$$g_{-1} = \frac{4}{(1-x^2-y^2)^2} g_0,$$

where $g_0$ is the Euclidean metric in $\mathbb{R}^2$. In this model, the asymptotic boundary $\partial_\infty \mathbb{H}^2$ of $\mathbb{H}^2$ is identified with the unit circle. Consequently, any point in the closed unit disk is viewed as either a point in $\mathbb{H}^2$ or a point in $\partial_\infty \mathbb{H}^2$. We denote by $\mathbf{0}$ the origin of $\mathbb{H}^2$.

In $\mathbb{H}^2$ we consider $\gamma_0, \gamma_1$ the geodesic lines $\{x = 0\}, \{y = 0\}$, respectively. For $j = 0, 1$, we denote by $Y_j$ the Killing vector field whose flow $(\phi_l)_{l \in (-1,1)}$ is given by hyperbolic translation along $\gamma_j$ with $\phi_l(\mathbf{0}) = (l \sin(\pi j), l \cos(\pi j))$, and with $(\sin(\pi j), \cos(\pi j))$ as attractive point at infinity. We call $(\phi_l)_{l \in (-1,1)}$ the flow of $Y_j$ even though the family $(\phi_l)_{l \in (-1,1)}$ is not parameterized at the right speed.

We denote by $\pi : \mathbb{H}^2 \times \mathbb{R} \to \mathbb{H}^2$ the vertical projection and we write $t$ for the height coordinate in $\mathbb{H}^2 \times \mathbb{R}$. In what follows, we will often identify the hyperbolic plane $\mathbb{H}^2$ with the horizontal slice $\{t = 0\}$ of $\mathbb{H}^2 \times \mathbb{R}$. The vector fields $Y_j$, $j = 0, 1$, and their flows naturally extend to horizontal vector fields and their flows in $\mathbb{H}^2 \times \mathbb{R}$.

Consider any geodesic $\gamma$ tending to the point at infinity $p_0 \in \partial_\infty \mathbb{H}^2$, parametrized by arc length. Let $c(s)$ denote the horocycle in $\mathbb{H}^2$ tangent to $\partial_\infty \mathbb{H}^2$ at $p_0$ that intersects $\gamma$ at $\gamma(s)$. Given two points $p, q \in c(s)$, we denote by $\psi : \mathbb{H}^2 \times \mathbb{R} \to \mathbb{H}^2 \times \mathbb{R}$ the parabolic translation along $c(s)$ such that $\psi(p) = q$.

We write $\overline{pq}$ to denote the geodesic arc between the two points $p, q$ of $\mathbb{H}^2 \times \mathbb{R}$.

## 3. The Alexandrov problem for doubly periodic constant mean curvature surfaces

Take two points $p, q$ in a horocycle $c(s)$, and let $\psi$ be the parabolic translation along $c(s)$ such that $\psi(p) = q$. We have $\psi(c(s)) = c(s)$ for all $s$. Consider the $\mathbb{Z}^2$ subgroup $G$ of isometries of $\mathbb{H}^2 \times \mathbb{R}$ generated by $\psi$ and a vertical translation $T(h)$, for some positive $h$. We denote by $\mathcal{M}$ the quotient of $\mathbb{H}^2 \times \mathbb{R}$ by $G$. The manifold $\mathcal{M}$ is diffeomorphic but not isometric to $\mathbb{T}^2 \times \mathbb{R}$ and is foliated by the family of tori $\mathbb{T}(s) = (c(s) \times \mathbb{R})/G$, $s \in \mathbb{R}$, which are intrinsically flat and have constant mean curvature $\frac{1}{2}$. Thus the tori $\mathbb{T}(s)$ are examples of compact embedded constant mean curvature surfaces in $\mathcal{M}$.

We have the following answer to the Alexandrov problem in $\mathcal{M}$.

**Theorem 1.** *Let $\Sigma \subset \mathcal{M}$ be a compact immersed surface with constant mean curvature $H$. Then $H \geq \frac{1}{2}$. Moreover*:

(1) *If $H = \frac{1}{2}$, then $\Sigma$ is a torus $\mathbb{T}(s)$, for some $s$.*

(2) *If $H > \frac{1}{2}$ and $\Sigma$ is embedded, then $\Sigma$ is either the quotient of a rotational sphere, or the quotient of a vertical unduloid* (*in particular, a vertical cylinder over a circle*).

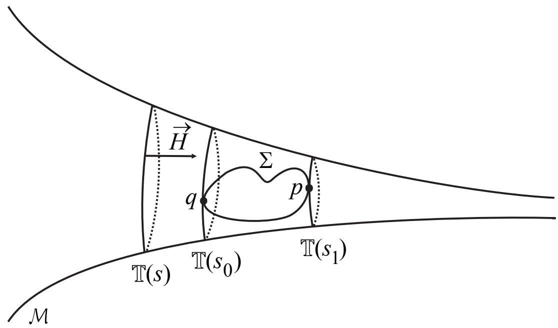


**Figure 1.** $\Sigma \subset \mathcal{M}$.

*Proof.* Let $\Sigma$ be a compact immersed surface in $\mathcal{M}$ with constant mean curvature $H$. As $\Sigma$ is compact, there exist $s_0 \leq s_1 \in \mathbb{R}$ such that $\Sigma$ is between $\mathbb{T}(s_0)$ and $\mathbb{T}(s_1)$, and it is tangent to $\mathbb{T}(s_0), \mathbb{T}(s_1)$ at points $q$, $p$, respectively, as illustrated in Figure 1.

For $s < s_0$, the torus $\mathbb{T}(s)$ does not intersect $\Sigma$, and $\Sigma$ stays in the mean convex region bounded by $\mathbb{T}(s)$.

By comparison at $q$, we conclude that $H \geq \frac{1}{2}$. If $H = \frac{1}{2}$, then by the maximum principle, $\Sigma$ is the torus $\mathbb{T}(s_0)$, and we have proved the first part of the theorem.

To prove the last part, suppose $\Sigma$ is embedded and consider the quotient space $\widetilde{\mathcal{M}} = \mathbb{H}^2 \times \mathbb{R}/[T(h)]$, which is diffeomorphic to $\mathbb{H}^2 \times \mathbb{S}^1$. Take a connected component $\widetilde{\Sigma}$ of the lift of $\Sigma$ to $\widetilde{\mathcal{M}}$, and denote by $\tilde{c}(s)$ the surface $c(s) \times \mathbb{S}^1$. Observe that $\tilde{c}(s)$ is the lift of $\mathbb{T}(s)$ to $\widetilde{\mathcal{M}}$. Moreover, let us consider two points $\tilde{p}, \tilde{q} \in \widetilde{\Sigma}$ whose projections in $\mathcal{M}$ are the points $p, q$, respectively.

It is easy to prove that $\widetilde{\Sigma}$ separates $\widetilde{\mathcal{M}}$. In fact, suppose by contradiction this is not true, then we can consider a geodesic arc $\alpha : (-\epsilon, \epsilon) \to \widetilde{\mathcal{M}}$ such that $\alpha(0) \in \widetilde{\Sigma}$, $\alpha'(0) \in T\widetilde{\Sigma}^{\perp}$ and we can join the points $\alpha(-\epsilon), \alpha(\epsilon)$ by a curve that does not intersect $\widetilde{\Sigma}$, hence we obtain a Jordan curve, which we still call $\alpha$, whose intersection number with $\widetilde{\Sigma}$ is $1 \mod 2$. Notice that the distance between $\widetilde{\Sigma}$ and $\tilde{c}(s_0)$ is bounded. Since we can homotope $\alpha$ so it is arbitrarily far from $\tilde{c}(s_0)$, we conclude that a translate of $\alpha$ does not intersect $\widetilde{\Sigma}$, contradicting the fact that the intersection number of $\alpha$ and $\widetilde{\Sigma}$ is $1 \mod 2$. Thus $\widetilde{\Sigma}$ does separate $\widetilde{\mathcal{M}}$.

Let us call $A$ the mean convex component of $\widetilde{\mathcal{M}} \setminus \widetilde{\Sigma}$ with boundary $\widetilde{\Sigma}$ and $B$ the other component. Hence $\widetilde{\mathcal{M}} \setminus \widetilde{\Sigma} = A \cup B$.

Let $\gamma$ be a geodesic in $\mathbb{H}^2$ that limits to $p_0 = \gamma(+\infty) \in \partial_\infty \mathbb{H}^2$ (the point where the horocycles $c(s)$ are centered) and let us assume that $\gamma$ intersects $\widetilde{\Sigma}$ in *at least* two points.

Consider the family $(l_t)_{t \in \mathbb{R}}$ of geodesics in $\mathbb{H}^2$ orthogonal to $\gamma$ and denote by $P(t)$ the totally geodesic vertical annulus $l_t \times \mathbb{S}^1$ of $\widetilde{\mathcal{M}} = \mathbb{H}^2 \times \mathbb{S}^1$ (see Figure 2). Since $\widetilde{\Sigma}$ is a lift of the compact surface $\Sigma$, it stays in the region between $\tilde{c}(s_0)$

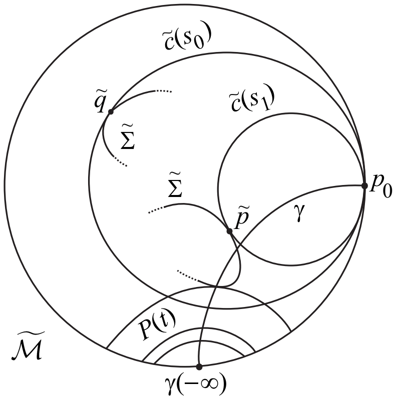


**Figure 2.** The family of totally geodesic annuli $P(t)$.

and $\tilde{c}(s_1)$, and the distance from any point of $\tilde{\Sigma}$ to $\tilde{c}(s_0)$ and to $\tilde{c}(s_1)$ is uniformly bounded.

By our choice of $\gamma$, the ends of each $P(t)$ are outside the region bounded by $\tilde{c}(s)$, hence $P(t) \cap \tilde{\Sigma}$ is compact for all $t$. Moreover, for $t$ close to $-\infty$, $P(t)$ is contained in $B$ and $P(t) \cap \tilde{\Sigma}$ is empty. Then start with $t$ close to $-\infty$ and let $t$ increase until a first contact point between $\tilde{\Sigma}$ and some vertical annulus, say $P(t_0)$. In particular, we know that the mean curvature vector of $\tilde{\Sigma}$ does not point into $\bigcup_{t \leq t_0} P(t)$.

Continuing to increase $t$ and starting the Alexandrov reflection procedure for $\tilde{\Sigma}$ and the family of vertical totally geodesic annuli $P(t)$, we get a first contact point between the reflected part of $\tilde{\Sigma}$ and $\tilde{\Sigma}$, for some $t_1 \in \mathbb{R}$. Observe that this first contact point occurs because we are assuming that the geodesic $\gamma$ intersects $\tilde{\Sigma}$ in at least two points.

Then $\tilde{\Sigma}$ is symmetric with respect to $P(t_1)$. As $\tilde{\Sigma} \cap \big(\bigcup_{t_0 \leq t \leq t_1} P(t)\big)$ is compact, $\tilde{\Sigma}$ is compact. Hence, given any horizontal geodesic $\alpha$ we can apply the Alexandrov procedure with the family of totally geodesic vertical annuli $Q(t) = \tilde{l}_t \times \mathbb{S}^1$, where $(\tilde{l}_t)_{t \in \mathbb{R}}$ is the family of horizontal geodesics orthogonal to $\alpha$, and we obtain a symmetry plane for $\tilde{\Sigma}$.

Hence we have shown that if some geodesic that limits to $p_0$ intersects $\tilde{\Sigma}$ in two or more points, then $\tilde{\Sigma}$ lifts to a rotational cylindrically bounded surface $\bar{\Sigma}$ in $\mathbb{H}^2 \times \mathbb{R}$. If $\bar{\Sigma}$ is not compact then $\bar{\Sigma}$ is a vertical unduloid, and if $\bar{\Sigma}$ is compact we know by the theorem of Hsiang and Hsiang [1989] $\bar{\Sigma}$ is a rotational sphere. Therefore, we have proved that in this case $\Sigma \subset \mathcal{M}$ is either the quotient of a rotational sphere or the quotient of a vertical unduloid.

Now to finish the proof let us assume that every geodesic that limits to $p_0$ intersects $\tilde{\Sigma}$ in *at most* one point. In particular, the geodesic $\beta$ that limits to $p_0$ and

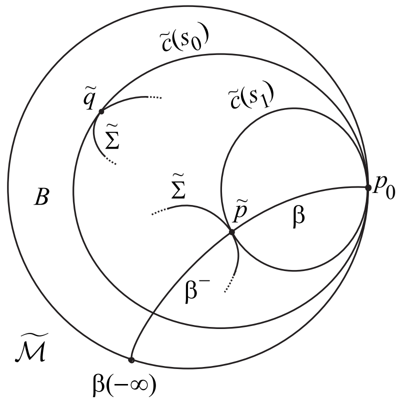


**Figure 3.** Geodesic $\beta$.

passes through $\tilde{p} \in \tilde{c}(s_1)$ intersects $\tilde{\Sigma}$ only at $\tilde{p}$. Write $\beta^-$ to denote the arc of $\beta$ between $\beta(-\infty)$ and $\tilde{p}$ (see Figure 3).

As $\beta \cap \tilde{\Sigma} = \{\tilde{p}\}$, we have $\beta^- \cap \tilde{\Sigma} = \varnothing$ and then $\beta^- \subset B$, since $\tilde{\Sigma}$ separates $\widetilde{\mathcal{M}}$.

Hence at the point $\tilde{p} \in \tilde{\Sigma} \cap \tilde{c}(s_1)$, the mean curvature vectors of $\tilde{\Sigma}$ and $\tilde{c}(s_1)$ point to the mean convex side of $\tilde{c}(s_1)$ and $\tilde{\Sigma}$ lies on the mean concave side of $\tilde{c}(s_1)$, then by comparison we get $H \leq \frac{1}{2}$. But we already know that $H \geq \frac{1}{2}$. Hence $H = \frac{1}{2}$ and $\tilde{\Sigma} = \tilde{c}(s_1)$, by the maximum principle. Therefore, in this case we conclude $\Sigma = \mathbb{T}(s_1)$. □

**Remark.** Note that a vertical unduloid, contained in a cylinder $D \times \mathbb{R}$ and invariant by a vertical translation $T(l)$ in $\mathbb{H}^2 \times \mathbb{R}$, passes to the quotient $\mathcal{M} = \mathbb{H}^2 \times \mathbb{R}/[\psi, T(h)]$ as an embedded surface if the quotient of $D$ is embedded and the number $l$ is a multiple of $h$. Analogously, a rotational sphere of height $l$ contained in a cylinder $D \times \mathbb{R}$ in $\mathbb{H}^2 \times \mathbb{R}$ passes to the quotient as an embedded surface if $l < h$ and the quotient of $D$ is embedded in $\mathcal{M}$.

## 4. Periodic minimal surfaces

In this section we are interested in constructing some new examples of periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$ invariant by a subgroup of isometries, which is either isomorphic to $\mathbb{Z}^2$, or generated by a vertical translation, or generated by a screw motion. In fact, we only consider subgroups generated by a parabolic translation $\psi$ along a horocycle and/or a vertical translation $T(h)$, for some $h > 0$.

Periodic minimal surfaces in $\mathbb{R}^3$ have received great attention since Riemann, Schwarz, Scherk (and many others) studied them. They also appear in the natural sciences. Meeks and Rosenberg [1993] proved that a periodic properly embedded minimal surface of finite topology (in $\mathbb{R}^3/G$, where $G$ is a nontrivial discrete group of isometries acting properly discontinuously on $\mathbb{R}^3$) has finite total curvature and the ends are asymptotic to standard ends (planar, catenoidal, or helicoidal). In [Hauswirth and Menezes 2013], we consider the same study for periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$. The first step is to understand what are the possible models for the ends in the quotient. This is one reason to construct examples.

**4.1.** ***Doubly periodic minimal surface.*** In $\mathbb{H}^2$ consider two geodesics $\alpha, \beta$ that limit to the same point at infinity, say $\alpha(-\infty) = p_0 = \beta(-\infty)$. Denote $B = \alpha(+\infty)$ and $D = \beta(+\infty)$. Take a geodesic $\gamma$ contained in the region bounded by $\alpha$ and $\beta$ that limits to the same point $p_0$ at infinity. Parametrize these geodesics so that $\alpha(t) \to B$, $\beta(t) \to D$ and $\gamma(t) \to p_0$ when $t \to +\infty$.

Fix $h > \pi$ and consider the following Jordan curve:

$$\Gamma_t = \overline{(\alpha(t),0)\,(\gamma(t),0)} \cup \overline{(\alpha(t),0)\,(\alpha(t),h)} \cup \overline{(\beta(t),0)\,(\gamma(t),0)} \cup \overline{(\beta(t),0)\,(\beta(t),h)} \cup \overline{(\alpha(t),h)\,(\gamma(t),h)} \cup \overline{(\beta(t),h)\,(\gamma(t),h)},$$

as illustrated in Figure 4.

Consider a least area embedded minimal disk $\Sigma_t$ with boundary $\Gamma_t$. Let $Y$ be the Killing field whose flow $(\phi_l)_{l\in\mathbb{R}}$ is given by translation along the geodesic $\gamma$. Notice that $\Gamma_t$ is transversal to the Killing field $Y$. Hence given any geodesic $\bar{\gamma}$ orthogonal to $\gamma$, we can use the Alexandrov reflection technique with the foliation of $\mathbb{H}^2 \times \mathbb{R}$ by the vertical planes $(\phi_l(\bar{\gamma}))_{l\in\mathbb{R}}$ to show that $\Sigma_t$ is a $Y$-Killing graph. In particular, $\Sigma_t$ is stable and unique (see [Nelli and Rosenberg 2006, Lemma 2.1]). This gives uniform curvature estimates for $\Sigma_{t_0}$ for points far from the boundary (see [Rosenberg et al. 2010, Main Theorem]). Rotating $\Sigma_t$ by angle $\pi$ around the geodesic arc $\overline{(\alpha(t),0)\,(\gamma(t),0)}$ gives a minimal surface that extends $\Sigma_t$, has int $\overline{(\alpha(t),0)\,(\gamma(t),0)}$ in its interior, and is still a $Y$-Killing graph. Thus we get uniform curvature estimates for $\Sigma_t$ in a neighborhood of $\overline{(\alpha(t),0)\,(\gamma(t),0)}$. This is also true for the three other horizontal geodesic arcs in $\Gamma_t$.

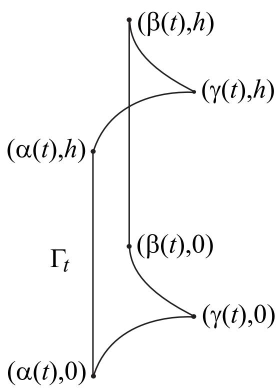


**Figure 4.** Curve $\Gamma_t$.

Observe that for any $t$, $\Sigma_t$ stays in the halfspace determined by $\overline{BD} \times \mathbb{R}$ that contains $\Gamma_t$, by the maximum principle.

As $h > \pi$, we can use as a barrier the minimal surface $S_h \subset \mathbb{H}^2 \times (0, h)$ which is a vertical bigraph with respect to the horizontal slice $\{t = h/2\}$. The surface $S_h$ is invariant by translations along the horizontal geodesic $\gamma_0 = \{x = 0\}$ and its asymptotic boundary is $(\tau \times \{0\}) \cup \overline{(0, 1, 0)\,(0, 1, h)} \cup (\tau \times \{h\}) \cup \overline{(0, -1, 0)\,(0, -1, h)}$, where $\tau = \partial_\infty \mathbb{H}^2 \cap \{x > 0\}$. For more details about the surface $S_h$, see [Mazet et al. 2011a; 2011b; Sá Earp 2008].

For $l$ sufficiently large, the translated surface $\phi_l(S_h)$ does not intersect $\Sigma_t$; hence the surface $\Sigma_t$ is contained between $\phi_l(S_h)$ and $\overline{BD} \times \mathbb{R}$.

Notice that when $t \to +\infty$, $\Gamma_t$ converges to $\Gamma$, where

$$\Gamma = (\alpha \times \{0\}) \cup (\beta \times \{0\}) \cup (\alpha \times \{h\}) \cup (\beta \times \{h\}) \cup \overline{(D, 0)\,(D, h)} \cup \overline{(B, 0)\,(B, h)}.$$

Therefore, as we have uniform curvature estimates and barriers at infinity, there exists a subsequence of $\Sigma_t$ that converges to a minimal surface $\Sigma$, where $\Sigma$ lies in the region of $\mathbb{H}^2 \times [0, h]$ bounded by $\alpha \times \mathbb{R}$, $\beta \times \mathbb{R}$, $\overline{BD} \times \mathbb{R}$ and $\phi_l(S_h)$, and with boundary $\partial\Sigma = \Gamma$.

Hence the surface obtained by reflection in all horizontal boundary geodesics of $\Sigma$ is invariant by $\psi^2$ and $T(2h)$, where $\psi$ is the horizontal translation along horocycles that sends $\alpha$ to $\beta$. Moreover, this surface in the quotient space $\mathbb{H}^2 \times \mathbb{R}/[\psi^2, T(2h)]$ is topologically a sphere minus four points. Two ends are asymptotic to vertical planes and two are asymptotic to horizontal planes (cusps), all of them with finite total curvature.

**Proposition 1.** *There exists a doubly periodic minimal surface* (*invariant by horizontal translations along a horocycle and by a vertical translation*) *such that*, *in the quotient space*, *this surface is topologically a sphere minus four points*, *with two ends asymptotic to vertical planes and two asymptotic to horizontal planes*, *all of them with finite total curvature.*

**4.2.** ***Vertically periodic minimal surfaces.*** Take $\alpha$ any geodesic in $\mathbb{H}^2 \times \{0\}$. For $h > \pi$, consider the vertical segment $\alpha(-\infty) \times [0, 2h]$, and a point $p \in \partial_\infty \mathbb{H}^2$, $p \neq \alpha(-\infty), \alpha(+\infty)$. For some small $\epsilon > 0$, consider the asymptotic vertical segment joining $(p, \epsilon)$ and $(p, h + \epsilon)$. Now, connect $(p, \epsilon)$ to $(\alpha(-\infty), 0)$ and $(p, h+\epsilon)$ to $(\alpha(-\infty), 2h)$ by curves in $\partial_\infty \mathbb{H}^2 \times \mathbb{R}$, whose tangent vectors are never horizontal or vertical, and so that the resulting curve $\Gamma$ is differentiable. Also, consider the horizontal geodesic $\beta$ connecting $p$ to $\alpha(+\infty)$.

Parametrize $\alpha$ by arc length, and consider $\gamma$ a geodesic orthogonal to $\alpha$ passing through $\alpha(0)$. Let us denote by $d(t)$ the equidistant curve to $\gamma$ at a distance $|t|$ that intersects $\alpha$ at $\alpha(t)$. For each $t$ consider a curve $\Gamma_t$ contained in the plane $d(t) \times \mathbb{R}$ with endpoints $(\alpha(t), 0)$ and $(\alpha(t), 2h)$ such that $\Gamma_t$ is contained in the

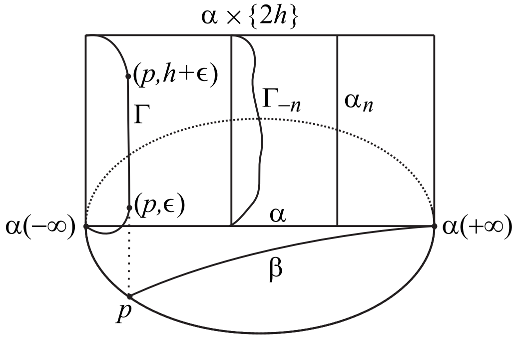


**Figure 5.** Curves $\Gamma_{-n}$ and $\Gamma$.

region $R$ bounded by $\alpha \times \mathbb{R}$, $\beta \times \mathbb{R}$, $\mathbb{H}^2 \times \{0\}$ and $\mathbb{H}^2 \times \{2h\}$ with the properties that its tangent vectors do not point in the horizontal direction and $\Gamma_t$ converges to $\Gamma$ when $t \to -\infty$. In particular, $\Gamma_t$ is transversal to the Killing field $Y$ whose flow $(\phi_l)_{l \in \mathbb{R}}$ is given by translation along the geodesic $\gamma$.

Write $\alpha_t$ to denote the vertical segment $\alpha(t) \times [0, 2h]$ (see Figure 5).

For each $n$, let $\Sigma_n$ be the solution to the Plateau problem with boundary

$$\Gamma_{-n} \cup (\alpha([-n, n]) \times \{0\}) \cup (\alpha([-n, n]) \times \{2h\}) \cup \alpha_n.$$

By our choice of the curves $\Gamma_t$, the boundary $\partial\Sigma_n$ is transverse to the Killing field $Y$. Using the foliation of $\mathbb{H}^2 \times \mathbb{R}$ by the vertical planes $\phi_l(\alpha)$, $l \in \mathbb{R}$, the Alexandrov reflection technique shows that $\Sigma_n$ is a $Y$-Killing graph. In particular, it is unique and stable [Nelli and Rosenberg 2006], and we have uniform curvature estimates far from the boundary [Rosenberg et al. 2010]. When we apply the rotation by angle $\pi$ around $\alpha \times \{0\}$ to the minimal surface $\Sigma_n$, we get another minimal surface which extends $\Sigma_n$, is still a $Y$-Killing graph and has $\mathrm{int}(\alpha([-n, n]) \times \{0\})$ in its interior. Hence we obtain uniform curvature estimates for $\Sigma_n$ in a neighborhood of $\alpha([-n, n]) \times \{0\}$. This is also true for $\alpha([-n, n]) \times \{2h\}$ and $\alpha_n$.

Observe that $\Sigma_n$ is contained in the region $R$, for all $n$.

By our choice of $\Gamma$, for each $q \in \Gamma$, we can consider two translations of the minimal surfaces $S_h$ (considered in the last section) that pass through $q$ so that one of them has asymptotic boundary under $\Gamma$, the other one has asymptotic boundary above $\Gamma$ and their intersection with $\Gamma$ is just the point $q$ considered or is the whole vertical segment $\overline{(p, \epsilon)\,(p, h + \epsilon)}$. Hence, the envelope of the union of all these translated surfaces $S_h$ forms a barrier to $\Sigma_n$, for all $n$.

Then, as we have uniform curvature estimates and barriers at infinity, we conclude that there exists a subsequence of $\Sigma_n$ that converges to a minimal surface $\Sigma$ with

$$(\alpha(+\infty) \times [0, 2h]) \cup \Gamma = \partial_\infty \Sigma,$$

and then

$$\partial\Sigma = \Gamma \cup (\alpha \times \{0\}) \cup (\alpha \times \{2h\}) \cup (\alpha(+\infty) \times [0, 2h]).$$

Therefore, the surface obtained by reflection in all horizontal boundary geodesics of $\Sigma$ is a vertically periodic minimal surface invariant by $T(4h)$. In the quotient space this minimal surface has two ends; one is asymptotic to a vertical plane and has finite total curvature, while the other one is topologically an annular end and has infinite total curvature.

**Proposition 2.** *There exists a singly periodic minimal surface* (*invariant by a vertical translation*) *such that*, *in the quotient space*, *this surface has two ends*; *one end is asymptotic to a vertical plane and has finite total curvature*, *while the other one is topologically an annular end and has infinite total curvature.*

**4.3.** ***Periodic minimal surfaces invariant by screw motion.*** Now we construct some examples of periodic minimal surfaces invariant by a screw motion, that is, invariant by a subgroup of isometries generated by the composition of a horizontal translation with a vertical translation.

Consider two geodesics $\alpha, \beta$ in $\mathbb{H}^2$ that limit to the same point at infinity, say $\alpha(+\infty) = p_0 = \beta(+\infty)$. For $h > \pi$, consider a smooth curve $\Gamma$ contained in the asymptotic boundary of $\mathbb{H}^2 \times \mathbb{R}$, connecting $(\alpha(-\infty), 2h)$ to $(\beta(-\infty), 0)$ and such that its tangent vectors are never horizontal or vertical. Also, take a point $p \in \partial_\infty \mathbb{H}^2$ in the halfspace determined by $\beta \times \mathbb{R}$ that does not contain $\alpha$.

For some small $\epsilon > 0$, consider the asymptotic vertical segment joining $(p, \epsilon)$ and $(p, h + \epsilon)$. Now, connect $(p, \epsilon)$ to $(p_0, 0)$ and $(p, h + \epsilon)$ to $(p_0, 2h)$ by curves in $\partial_\infty \mathbb{H}^2 \times \mathbb{R}$ whose tangent vectors are never horizontal or vertical, and such that the resulting curve $\widehat{\Gamma}$ is differentiable.

Parametrize $\alpha$ by arc length, and consider $\gamma$ a geodesic orthogonal to $\alpha$ passing through $\alpha(0)$. Let us denote by $d(t)$ the equidistant curve to $\gamma$ at a distance $|t|$ that intersects $\alpha$ at $\alpha(t)$. For each $t, s$ consider two curves $\widehat{\Gamma}_t$ and $\Gamma_s$ contained in the plane $d(t) \times \mathbb{R}$ and $d(s) \times \mathbb{R}$, respectively, with the properties that their tangent vectors are never horizontal, $\widehat{\Gamma}_t$ joins $(\alpha(t), 2h)$ to $(\beta(t), 0)$, $\Gamma_s$ joins $(\alpha(s), 2h)$ to $(\beta(s), 0)$, $\widehat{\Gamma}_t$ converges to $\widehat{\Gamma}$ when $t \to +\infty$, $\Gamma_s$ converges to $\Gamma$ when $s \to -\infty$, and both curves are contained in the region $R$ bounded by $\alpha \times \mathbb{R}$, $\theta \times \mathbb{R}$, $\mathbb{H}^2 \times \{0\}$ and $\mathbb{H}^2 \times \{2h\}$, where $\theta$ is the geodesic with endpoints $p$ and $\beta(-\infty)$ (see Figure 6).

For each $n$, let $\Sigma_n$ be the solution to the Plateau problem with boundary

$$\Gamma_{-n} \cup (\alpha([-n, n]) \times \{2h\}) \cup \widehat{\Gamma}_n \cup (\beta([-n, n]) \times \{0\}).$$

The surface $\Sigma_n$ is contained in the region $R$. As in the previous section, we can show that $\Sigma_n$ is a Killing graph, then it is stable, unique and we have uniform curvature estimates far from the boundary. Rotating $\Sigma_n$ by angle $\pi$ around the geodesic

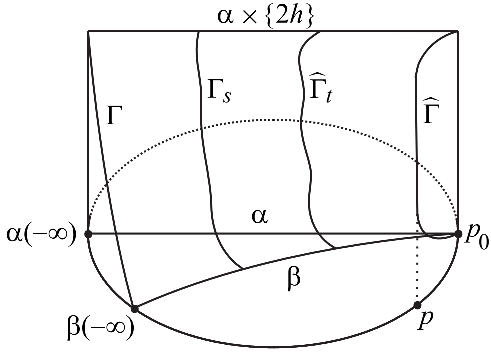


**Figure 6.** Curves $\widehat{\Gamma}_t, \Gamma_s, \widehat{\Gamma}$ and $\Gamma$.

$\alpha \times \{2h\}$ we get a minimal surface which extends $\Sigma_n$, is still a Killing graph, and has $\operatorname{int}(\alpha([-n,n]) \times \{2h\})$ in its interior. Hence we get uniform curvature estimates for $\Sigma_n$ in a neighborhood of $\alpha([-n,n]) \times \{2h\}$. This is also true for $\beta([-n,n]) \times \{0\}$. Thus when $n \to +\infty$, there exists a subsequence of $\Sigma_n$ that converges to a minimal surface $\Sigma$ with $\Gamma \cup \widehat{\Gamma} \subset \partial_\infty \Sigma_n$. Using the same argument as before with suitable translations of the surface $S_h$ as barriers, we conclude that in fact $\partial_\infty \Sigma = \Gamma \cup \widehat{\Gamma}$, and then

$$\partial \Sigma = \Gamma \cup (\alpha \times \{2h\}) \cup (\beta \times \{0\}) \cup \widehat{\Gamma}.$$

The surface obtained by reflection in all horizontal boundary geodesics of $\Sigma$ is a minimal surface invariant by $\psi^2 \circ T(4h)$, where $\psi$ is the horizontal translation along horocycles that sends $\alpha$ to $\beta$. There are two annular embedded ends in the quotient, each of infinite total curvature.

**Proposition 3.** *There exists a minimal surface invariant by a screw motion such that*, *in the quotient space*, *this minimal surface has two annular embedded ends*, *each one of infinite total curvature.*

Now we will construct another interesting example of a periodic minimal surface invariant by a screw motion.

Denote by $\gamma_0, \gamma_1$ the geodesic lines $\{x = 0\}, \{y = 0\}$ in $\mathbb{H}^2$, respectively. Let $c$ be a horocycle orthogonal to $\gamma_1$, and consider $p, q \in c$ equidistant points to $\gamma_1$. Take $\alpha, \beta$ geodesics which limit to $p_0 = (1,0) = \gamma_1(+\infty)$ and pass through $p, q$, respectively. Fix $\epsilon > 0$ and $h > \pi$. Define the points

$$A = \alpha(-t_0),$$
$$C = \alpha(t_0),$$
$$B = \beta(-t_0),$$
$$D = \beta(t_0),$$

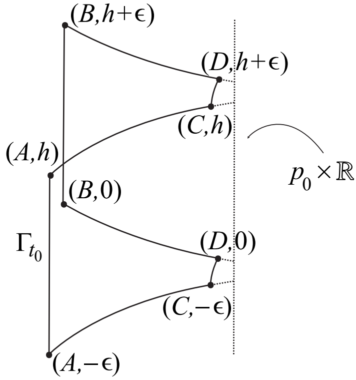


**Figure 7.** Curve $\Gamma_{t_0}$.

and let us consider the following Jordan curve (see Figure 7):

$$\begin{aligned}\Gamma_{t_0} = (\alpha([-t_0,t_0])\times\{-\epsilon\}) &\cup \overline{(C,-\epsilon)\,(D,0)}\\ &\cup(\beta([-t_0,t_0])\times\{0\})\cup(\alpha([-t_0,t_0])\times\{h\})\cup\overline{(C,h)\,(D,h+\epsilon)}\\ &\cup(\beta([-t_0,t_0])\times\{h+\epsilon\})\cup\overline{(A,-\epsilon)\,(A,h)}\cup\overline{(B,0)\,(B,h+\epsilon)}.\end{aligned}$$

We consider a least area embedded minimal disk $\Sigma_{t_0}$ with boundary $\Gamma_{t_0}$.

Denote by $Y_1$ the Killing vector field whose flow $(\phi_l)_{l\in(-1,1)}$ gives the hyperbolic translation along $\gamma_1$ with $\phi_l(0)=(l,0)$ and $p_0$ as attractive point at infinity. As $\Gamma_{t_0}$ is transversal to the Killing field $Y_1$, we can prove, using the Alexandrov reflection procedure, that $\Sigma_{t_0}$ is a $Y_1$-Killing graph with convex boundary, in particular, $\Sigma_{t_0}$ is stable and unique [Nelli and Rosenberg 2006]. This yields uniform curvature estimates far from the boundary [Rosenberg et al. 2010]. Rotating $\Sigma_{t_0}$ by angle $\pi$ around the geodesic arc $\alpha([-t_0,t_0])\times\{-\epsilon\}$ gives a minimal surface that extends $\Sigma_{t_0}$, has $\mathrm{int}(\alpha([-t_0,t_0])\times\{-\epsilon\})$ in its interior, and is still a $Y_1$-Killing graph. Thus we get uniform curvature estimates for $\Sigma_{t_0}$ in a neighborhood of $\alpha([-t_0,t_0])\times\{-\epsilon\}$. This is also true for the three other horizontal geodesic arcs in $\Gamma_{t_0}$.

Write $F=\alpha(-\infty)$, $G=\beta(-\infty)$. Observe that, by the maximum principle, for any $t_0$, $\Sigma_{t_0}$ stays in the halfspace determined by $\overline{FG}\times\mathbb{R}$ that contains $\Gamma_{t_0}$.

Since $h>\pi$, we can consider the minimal surface $S_h$ (considered in Section 4.1) as a barrier. For $l$ close to 1, the translated surface $\phi_l(S_h)$ does not intersect $\Sigma_{t_0}$.

The surface $\Sigma_{t_0}$ is contained between $\phi_l(S_h)$ and $\overline{FG}\times\mathbb{R}$. When $t_0\to+\infty$, $\Gamma_{t_0}$ converges to $\Gamma$, where

$$\begin{aligned}\Gamma = (\alpha\times\{-\epsilon\})\cup\overline{(p_0,-\epsilon)\,(p_0,0)}&\cup(\beta\times\{0\})\cup(\alpha\times\{h\})\\ \cup\overline{(p_0,h)\,(p_0,h+\epsilon)}\cup(\beta\times\{h+\epsilon\})&\cup\overline{(F,-\epsilon)\,(F,h)}\cup\overline{(G,0)\,(G,h+\epsilon)}.\end{aligned}$$

Using the maximum principle, we can prove that $\Sigma_t$ is contained between $\phi_l(S_h)$ and $\overline{FG} \times \mathbb{R}$, for all $t > t_0$. Therefore, there exists a subsequence of the surfaces $\Sigma_t$ that converges to a minimal surface $\Sigma$, where $\Sigma$ lies in the region between $\mathbb{H}^2 \times \{-\epsilon\}$ and $\mathbb{H}^2 \times \{h+\epsilon\}$ bounded by $\alpha \times \mathbb{R}$, $\beta \times \mathbb{R}$, $\overline{FG} \times \mathbb{R}$ and $\phi_l(S_h)$, and has boundary $\partial\Sigma = \Gamma$.

Hence the surface obtained by reflection in all horizontal boundary geodesics of $\Sigma$ is invariant by $\psi^2 \circ T(2(h+\epsilon))$, where $\psi$ is the horizontal translation along horocycles that sends $\alpha$ to $\beta$. Moreover, this surface in the quotient space has two vertical ends and two helicoidal ends, each one of finite total curvature.

**Proposition 4.** *There exists a minimal surface invariant by a screw motion such that, in the quotient space, this minimal surface has four ends: two vertical ends and two helicoidal ends, all of them with finite total curvature.*

## 5. A multivalued Rado theorem

The aim of this section is to prove a multivalued Rado theorem for small perturbations of the helicoid. Recall that Rado's theorem says that minimal surfaces over a convex domain with graphical boundaries must be disks which are themselves graphical. We will prove that for certain small perturbations of the boundary of a (compact) helicoid there exists only one compact minimal disk with that boundary. By a compact helicoid we mean the intersection of a helicoid with certain compact regions in $\mathbb{H}^2 \times \mathbb{R}$. The idea here originated in [Hardt and Rosenberg 1990]. We will apply this multivalued Rado theorem to construct an embedded minimal surface in $\mathbb{H}^2 \times \mathbb{R}$ whose boundary is a small perturbation of the boundary of a complete helicoid.

Consider $Y$ the Killing field whose flow $(\phi_\theta)_{\theta \in [0,2\pi)}$ is given by rotations around the $z$-axis. For some $0 < c < 1$, let

$$D = \{(x, y) \in \mathbb{H}^2; x^2 + y^2 \le c\}.$$

Take a helix $h$ of constant pitch contained in a solid cylinder $D \times [0, d]$, so that the vertical projection of $h$ over $\mathbb{H}^2 \times \{0\}$ is $\partial D$, and the endpoints of $h$ are in the same vertical line. Let us denote by $\Gamma$ the Jordan curve which is the union of $h$, the two horizontal geodesic arcs joining the endpoints of $h$ to the $z$-axis, and the part of the $z$-axis. Call $\mathcal{H}$ the compact part of the helicoid that has $\Gamma$ as its boundary. We know that $\mathcal{H}$ is a minimal surface transversal to the Killing field $Y$ at the interior points. Take $\theta < \pi/4$, and consider $\mathcal{H}_1 = \phi_{-\theta}(\mathcal{H})$ and $\mathcal{H}_2 = \phi_\theta(\mathcal{H})$. Hence $\mathcal{H}_1, \mathcal{H}_2$ are two compact helicoids with boundary $\partial\mathcal{H}_1 = \phi_{-\theta}(\Gamma)$, $\partial\mathcal{H}_2 = \phi_\theta(\Gamma)$.

Consider $h_0$ a small smooth perturbation of the helix $h$ with fixed endpoints such that $h_0$ is transversal to $Y$ and $h_0$ is contained in the region between $\phi_{-\theta}(h)$ and $\phi_\theta(h)$ in $\partial D \times [0, d]$. Call $\Gamma_0$ the Jordan curve which is the union of $h_0$, the

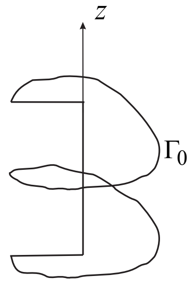


**Figure 8.** Curve $\Gamma_0$.

two horizontal geodesic arcs and a part of the $z$-axis, hence $\Gamma_0 = (\Gamma \setminus h) \cup h_0$ (see Figure 8).

Denote by $R$ the convex region bounded by $\mathcal{H}_1$ and $\mathcal{H}_2$ in the solid cylinder $D \times [0, d]$. The Jordan curve $\Gamma_0$ is contained in the simply connected region $R$ which has mean convex boundary. Then we can consider the solution to the Plateau problem in this region $R$, and we get a compact minimal disk $H$ contained in $R$ with boundary $\partial H = \Gamma_0$.

**Proposition 5.** *Under the assumptions above, $H$ is transversal to the Killing field $Y$ at the interior points. Moreover, the family $(\phi_\theta(H))_{\theta \in [0,2\pi)}$ foliates $D \times [0, d] \setminus \{z\text{-axis}\}$.*

*Proof.* As $H$ is a disk, we already know that each integral curve of $Y$ intersects $H$ in at least one point.

Observe that $\phi_{\pi/2}(R) \cap R \setminus \{z\text{-axis}\} = \varnothing$ and, in particular, $\phi_{\pi/2}(H) \cap H \setminus \{z\text{-axis}\} = \varnothing$. Moreover, notice that the tangent plane of $\phi_{\pi/2}(H)$ never coincides with the tangent plane of $H$ along the $z$-axis; at each point of the $z$-axis the surfaces are in disjoint sectors. So as one decreases $t$ from $\pi/2$ to 0, the surfaces $\phi_t(H)$ and $H$ have only the $z$-axis in common and they are never tangent along the $z$-axis. More precisely, as $t$ decreases and $t > 0$, there can not be a first interior point of contact between the two surfaces by the maximum principle. Also there can not be a point on the $z$-axis which is a first point of tangency of the two surfaces for $t > 0$, by the boundary maximum principle. Thus the surfaces $\phi_t(H)$ and $H$ have only the $z$-axis in common for $0 < t \leq \pi/2$. The same argument works for $-\pi/2 \leq t < 0$. Thus each integral curve of $Y$ intersects $H$ in exactly one point.

Denote by $R_2$ the region in $R$ bounded by $H$ and $\mathcal{H}_2$, and denote by $N$ the unit normal vector field of $H$ pointing toward $R_2$. As each integral curve of $Y$ intersects $H$ in exactly one point, we have $\langle N, Y \rangle \geq 0$ on $H$. As $\langle N, Y \rangle$ is a Jacobi function on the minimal surface $H$, we conclude that necessarily $\langle N, Y \rangle > 0$ in $\operatorname{int} H$. Therefore, $H$ is transversal to the Killing field $Y$ at the interior points, and the surfaces $\phi_t(H)$ foliate $D \times [0, d] \setminus \{z\text{-axis}\}$ for $t \in [0, 2\pi)$. $\square$

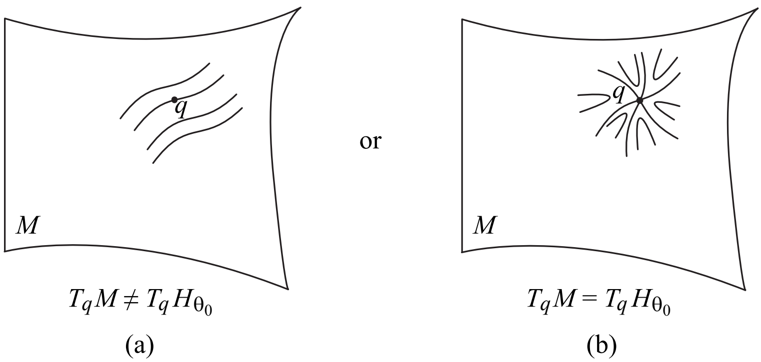


**Figure 9.** $q \in \operatorname{int} M$.

**Theorem 2** (multivalued Rado theorem). *Under the assumptions above*, $H$ *is the unique compact minimal disk with boundary* $\Gamma_0$.

*Proof.* Set $\Gamma_\theta = \phi_\theta(\Gamma_0)$ and $H_\theta = \phi_\theta(H)$, so $H_\theta$ is a minimal disk with $\partial H_\theta = \Gamma_\theta$. By Proposition 5, the family $(H_\theta)_{\theta \in [0,2\pi)}$ gives a foliation of the region $D \times [0, d] \setminus \{z\text{-axis}\}$.

Let $M \neq H$ be another compact minimal disk with boundary $\Gamma_0$. We will analyze the intersection between $M$ and each $H_\theta$.

First, observe that $M \subset D \times [0, d]$ by the maximum principle, and $M \cap H_\theta \neq \varnothing$ for all $\theta$.

Fix $\theta_0$. Given $q \in H_{\theta_0} \cap M$, then either $q \in \operatorname{int} M$ or $q \in \Gamma_0 = \partial M$.

Suppose $q \in \operatorname{int} M$.

If the intersection is transversal at $q$, then in a neighborhood of $q$ we have that $H_{\theta_0} \cap M$ is a simple curve passing through $q$. If we let $\theta_0$ vary a little, we see in $M$ a foliation as in part (a) of Figure 9.

On the other hand, if $M$ is tangent to $H_{\theta_0}$ at $q$, as the intersection of any two minimal surfaces is locally given by an $n$-prong singularity, that is, $2n$ embedded arcs that meet at equal angles (see [Hoffman and Meeks 1989, Claim 1 of Lemma 4]), then in a neighborhood of $q$ we have that $H_{\theta_0} \cap M$ consists of $2n$ curves passing through $q$ and making equal angles at $q$. If we let $\theta_0$ vary a little, we see in $M$ a foliation as in part (b) of Figure 9.

Now suppose $q \in \Gamma_0$.

If $q \in \Gamma_0 \cap \{z\text{-axis}\}$, to understand the trace of $H_{\theta_0}$ on $M$ in a neighborhood of $q$ we proceed as follows. Rotation by angle $\pi$ of $\mathbb{H}^2 \times \mathbb{R}$ about the $z$-axis extends $M$ smoothly to a minimal surface $\widetilde{M}$ that has $q$ as an interior point. Each $H_\theta$ also extends by this rotation (giving a helicoid $\widetilde{H}_\theta$). So in a neighborhood of $q$, we understand the intersection of $\widetilde{M}$ and $\widetilde{H}_{\theta_0}$. The surfaces $\widetilde{\mathcal{M}}$ and $\widetilde{H}_{\theta_0}$ are either transverse or tangent at $q$ as in Figure 9. Then when we restrict to $M \cap H_{\theta_0}$ and

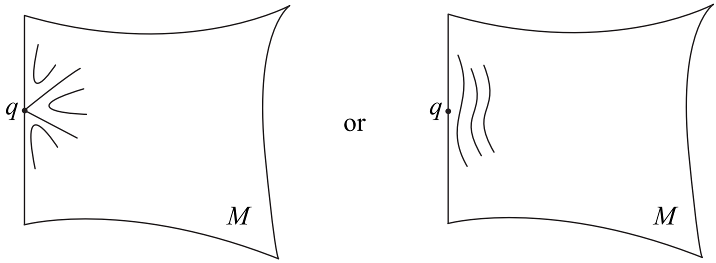


**Figure 10.** $q \in \Gamma_0 \cap \{z\text{-axis}\}$.

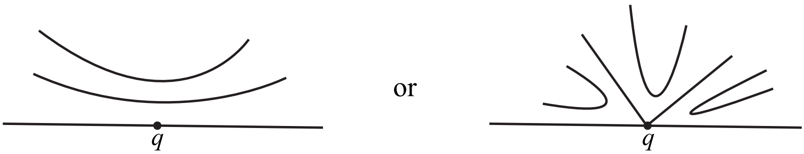


**Figure 11.** $q \in \Gamma_0 \setminus \{z\text{-axis}\}$.

let $\theta_0$ vary slightly, we see that the trace of $H_{\theta_0}$ on $M$ near $q$ is as in Figure 10, since the segment on the $z$-axis through $q$ is in $M \cap H_{\theta_0}$.

On the other hand, if $q \in \Gamma_0 \setminus \{z\text{-axis}\}$ then $\theta_0 = 0$, since $\Gamma_\theta \cap \Gamma_0 \setminus \{z\text{-axis}\} = \varnothing$ for any $\theta \neq 0$. Note that we cannot have $M \cap H$ homeomorphic to a semicircle in a neighborhood of $q$, since this would imply that $M$ is on one side of $H$ at $q$ and this contradicts the boundary maximum principle. Thus when we let $\theta_0 = 0$ vary a little, we have two possible foliations for $M$ in a neighborhood of $q$ as indicated in Figure 11.

Now consider two copies of $M$ and glue them together along the boundary.

Since $M$ is a disk, when we glue these two copies of $M$ we obtain a sphere with a foliation whose singularities have negative index by the analysis above. But this is impossible. Therefore, there is no minimal disk with boundary $\Gamma_0$ besides $H$. $\square$

**Remark.** This proof clearly works to prove Theorem 2 for slightly perturbed helicoids in $\mathbb{R}^3$.

Now let us construct an example of a complete embedded minimal surface in $\mathbb{H}^2 \times \mathbb{R}$ whose asymptotic boundary is a small perturbation of the asymptotic boundary of a complete helicoid.

Consider the (compact) helix $\beta(u) = (\cos u, \sin u, 2u)$ for $u \in [0, 4\pi]$. Notice that $\beta$ is a multigraph over $\partial_\infty \mathbb{H}^2$. Take $\theta < \pi/4$ and consider a small perturbation $\alpha(u)$ of $\beta(u)$ in $\partial_\infty \mathbb{H}^2 \times \mathbb{R}$ contained between $\phi_{-\theta}(\beta)$ and $\phi_\theta(\beta)$ such that $\alpha$ is transversal to $\partial_t$ and $\mathbb{H}^2 \times \{\tau\}$ for any $\tau \in [0, 8\pi]$, $\alpha(0) = \beta(0)$, $\alpha(4\pi) = \beta(4\pi)$ and so that the vertical distance between $\alpha(s)$ and $\alpha(s + 2\pi)$ is bigger than $\pi$ for any $s \in (0, 2\pi)$.

Now for $t \in [0, 1]$, consider the curves $\alpha_t(u) = (1-t)(0, 0, u) + t\alpha(u)$, $u \in [0, 4\pi]$. Call $\Gamma_t$ (respectively $\Gamma_1$) the Jordan curve which is the union of $\alpha_t$ (respectively $\alpha$), the two horizontal geodesics joining the endpoints of $\alpha_t$ (respectively $\alpha$) to the $z$-axis, and the part of the $z$-axis between $z = 0$ and $z = 8\pi$. Note that when $t$ goes to 1, the curves $\Gamma_t$ converge to the curve $\Gamma_1$. Denote by $H_t$ the minimal disk with boundary $\Gamma_t$. By Theorem 2, $H_t$ is stable and unique. In particular, we have uniform curvature estimates for points far from the boundary. As before, using rotation by angle $\pi$ around horizontal geodesics, we can prove that there is uniform curvature estimates for $H_t$ in a neighborhood of the two horizontal geodesic arcs of $\Gamma_t$.

As in the previous section, the envelope of the union of the translated surfaces $S_h$, $h > \pi$, forms a barrier to the sequence $H_t$, hence we conclude that there exists a subsequence of $H_t$ that converges to a minimal surface $H_1$ with boundary $\partial H_1 = \Gamma_1$. Rotation by angle $\pi$ of $\mathbb{H}^2 \times \mathbb{R}$ around the $z$-axis extends $H_1$ smoothly to a minimal surface which has two horizontal (straight) geodesics in its boundary. Thus the surface obtained by reflection in all horizontal boundary geodesics of $H_1$ is a minimal surface whose asymptotic boundary is a small perturbation of the asymptotic boundary of the complete helicoid in $\mathbb{H}^2 \times \mathbb{R}$ which has $\beta$ contained in its asymptotic boundary.

## Acknowledgements

This work is part of the author's Ph.D. thesis at the Instituto Nacional de Matemática Pura e Aplicada (IMPA). She would like to express her sincere gratitude to her advisor Professor Harold Rosenberg for his constant encouragement and guidance throughout the preparation of this work. The author was financially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil, and IMPA.

## References

[Alexandrov 1962] A. D. Alexandrov, "A characteristic property of spheres", *Ann. Mat. Pura Appl.* (4) **58** (1962), 303–315. MR 26 #722 Zbl 0107.15603

[Daniel and Mira 2013] B. Daniel and P. Mira, "Existence and uniqueness of constant mean curvature spheres in $\mathrm{Sol}_3$", *J. Reine Angew. Math.* 685 (2013), 1–32.

[Dean and Tinaglia 2005] B. Dean and G. Tinaglia, "A generalization of Rado's theorem for almost graphical boundaries", *Math. Z.* **251**:4 (2005), 849–858. MR 2007c:53009 Zbl 1085.53009

[Hardt and Rosenberg 1990] R. Hardt and H. Rosenberg, "Open book structures and unicity of minimal submanifolds", *Ann. Inst. Fourier* (*Grenoble*) **40**:3 (1990), 701–708. MR 92e:53009 Zbl 0702.53039

[Hauswirth and Menezes 2013] L. Hauswirth and A. Menezes, "On doubly periodic minimal surfaces in $\mathbb{H}^2 \times \mathbb{R}$ with finite total curvature in the quotient space", preprint, 2013. arXiv 1305.4813

[Hoffman and Meeks 1989] D. Hoffman and W. H. Meeks, III, "The asymptotic behavior of properly embedded minimal surfaces of finite topology", *J. Amer. Math. Soc.* **2**:4 (1989), 667–682. MR 90f:53010 Zbl 0683.53005

[Hsiang and Hsiang 1989] W.-T. Hsiang and W.-Y. Hsiang, "On the uniqueness of isoperimetric solutions and imbedded soap bubbles in non-compact symmetric spaces, I", *Invent. Math.* **98**:1 (1989), 39–58. MR 90h:53078 Zbl 0682.53057

[Mazet et al. 2011a] L. Mazet, M. M. Rodríguez, and H. Rosenberg, "The Dirichlet problem for the minimal surface equation, with possible infinite boundary data, over domains in a Riemannian surface", *Proc. Lond. Math. Soc.* (3) **102**:6 (2011), 985–1023. MR 2012f:53013 Zbl 1235.53007

[Mazet et al. 2011b] L. Mazet, M. M. Rodríguez, and H. Rosenberg, "Periodic constant mean curvature surfaces in $\mathbb{H}^2 \times \mathbb{R}$", preprint, 2011. To appear in Asian J. Math. arXiv 1106.5900

[Meeks and Rosenberg 1993] W. H. Meeks, III and H. Rosenberg, "The geometry of periodic minimal surfaces", *Comment. Math. Helv.* **68**:4 (1993), 538–578. MR 95a:53011 Zbl 0807.53049

[Nelli and Rosenberg 2006] B. Nelli and H. Rosenberg, "Simply connected constant mean curvature surfaces in $\mathbb{H}^2 \times \mathbb{R}$", *Michigan Math. J.* **54**:3 (2006), 537–543. MR 2008f:53007 Zbl 1152.53307

[Radó 1930] T. Radó, "Some remarks on the problem of plateau", *Proc. Natl. Acad. Sci. USA* **16**:3 (1930), 242–248. JFM 56.0437.01

[Rosenberg et al. 2010] H. Rosenberg, R. Souam, and E. Toubiana, "General curvature estimates for stable $H$-surfaces in 3-manifolds and applications", *J. Differential Geom.* **84**:3 (2010), 623–648. MR 2011g:53015 Zbl 1198.53062

[Sá Earp 2008] R. Sá Earp, "Parabolic and hyperbolic screw motion surfaces in $\mathbb{H}^2 \times \mathbb{R}$", *J. Aust. Math. Soc.* **85**:1 (2008), 113–143. MR 2010d:53067 Zbl 1178.53060

Received June 4, 2012. Revised September 9, 2013.

ANA MENEZES
INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA
ESTRADA DONA CASTORINA, 110
22460-320 RIO DE JANEIRO
BRAZIL
anamaria@impa.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.173

# TWISTED QUANTUM DRINFELD HECKE ALGEBRAS

DEEPAK NAIDU

**We generalize quantum Drinfeld Hecke algebras by incorporating a 2-cocycle on the associated finite group. We identify these algebras as specializations of deformations of twisted skew group algebras, giving an explicit connection to Hochschild cohomology. We classify these algebras for diagonal actions, as well as for the symmetric groups with their natural representations. Our results show that the parameter spaces for the symmetric groups in the twisted setting is smaller than in the untwisted setting.**

## 1. Introduction

Drinfeld Hecke algebras were defined by V. Drinfeld [1986]. They arise as symplectic reflection algebras in the work of P. Etingof and V. Ginzburg [2002], as braided Cherednik algebras in the work of Y. Bazlov and A. Berenstein [2009], and as graded versions of affine Hecke algebras in the work of G. Lusztig [1989]. They arise in diverse areas, such as representation theory, combinatorics, and orbifold theory, and they were used by I. Gordon [2003] to prove a version of the $n!$ conjecture for Weyl groups.

In this paper, we consider quantum and twisted analogs of Drinfeld Hecke algebras by incorporating quantum parameters as well as a 2-cocycle on the associated finite group. We simultaneously generalize twisted Drinfeld Hecke algebras and quantum Drinfeld Hecke algebras. The former was studied by S. Witherspoon [2007], and the latter was studied in [Levandovskyy and Shepler 2011] and [Naidu and Witherspoon 2011]. T. Chmutova [2005] generalized symplectic reflection algebras by incorporating a 2-cocycle on the associated finite group, and showed that such a 2-cocycle arises naturally for nonfaithful representations. Such a 2-cocycle also arises in orbifold theory, where they are known as discrete torsion [Adem and Ruan 2003; Căldăraru et al. 2004; Vafa and Witten 1995].

Let $V$ be a complex vector space with basis $v_1, v_2, \dots, v_n$, and $\boldsymbol{q} := (q_{ij})_{1\le i,j\le n}$, a tuple of nonzero scalars for which $q_{ii} = 1$ and $q_{ji} = q_{ij}^{-1}$ for all $i, j$. Let $S_{\boldsymbol{q}}(V)$

*MSC2010:* 16E40, 16S35.

*Keywords:* Hochschild cohomology, deformations, skew group algebras, graded Hecke algebras, symplectic reflection algebras.

denote the *quantum symmetric algebra*

$$S_{\boldsymbol{q}}(V) := \mathbb{C}\langle v_1, \ldots, v_n \mid v_i v_j = q_{ij} v_j v_i \text{ for all } 1 \le i, j \le n\rangle.$$

Let $G$ be a finite group acting linearly on $V$, and $\alpha : G \times G \to \mathbb{C}^\times$, a normalized 2-cocycle on $G$. Let $\kappa : V \times V \to \mathbb{C}^\alpha G$ be a bilinear map for which $\kappa(v_i, v_j) = -q_{ij}\kappa(v_j, v_i)$ for all $1 \le i, j \le n$. Let $T(V)$ be the tensor algebra on $V$, and define

$$\mathcal{H}_{\boldsymbol{q},\kappa,\alpha} := T(V)\#_\alpha G/(v_i v_j - q_{ij} v_j v_i - \kappa(v_i, v_j) \mid 1 \le i, j \le n),$$

the quotient of the twisted skew group algebra $T(V)\#_\alpha G$ by the ideal generated by all elements of the form specified. Suppose that the action of $G$ on $V$ induces an action of $G$ on $S_{\boldsymbol{q}}(V)$ by automorphisms, so we may form the twisted skew group algebra $S_{\boldsymbol{q}}(V)\#_\alpha G$. Assigning each $v_i$ degree one and each group element degree zero makes $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ a filtered algebra, and makes $S_{\boldsymbol{q}}(V)\#_\alpha G$ a graded algebra. We call $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ a *twisted quantum Drinfeld Hecke algebra* (over $\mathbb{C}$) if it satisfies the Poincaré–Birkhoff–Witt condition: the associated graded algebra $\operatorname{gr}\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is isomorphic, as a graded algebra, to $S_{\boldsymbol{q}}(V)\#_\alpha G$. The space of all maps $\kappa : V \times V \to \mathbb{C}^\alpha G$ for which $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra will be referred to as the *parameter space*.

***Main results and organization.*** In Section 2, we use G. Bergman's Diamond Lemma [1978] to give necessary and sufficient conditions for the algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ to be a twisted quantum Drinfeld Hecke algebra.

In Section 3, we identify the twisted quantum Drinfeld Hecke algebras $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ as specializations of particular types of deformations of the twisted skew group algebras $S_{\boldsymbol{q}}(V)\#_\alpha G$.

Section 4 develops the homological algebra needed for the sections that follow. Specifically, this section is concerned with the computation of the degree two Hochschild cohomology of $S_{\boldsymbol{q}}(V)\#_\alpha G$.

In Section 5, we establish a one-to-one correspondence between the subspace of *constant* Hochschild 2-cocycles (see Section 3 for definition) contained in $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$ and twisted quantum Drinfeld Hecke algebras associated to the quadruple $(G, V, \boldsymbol{q}, \alpha)$. As a consequence, we show that every constant Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$ lifts to a deformation of $S_{\boldsymbol{q}}(V)\#_\alpha G$.

In Section 6, we consider diagonal actions of $G$ on a chosen basis for $V$, and, using results from [Naidu et al. 2011], we classify the corresponding twisted quantum Drinfeld Hecke algebras.

In Section 7, we consider the symmetric groups $S_n$, $n \ge 5$, with their natural representations, with the unique nontrivial quantum parameters $q_{ij} = -1$, $i \ne j$, and with a cohomologically nontrivial 2-cocycle on $S_n$, which is unique up to coboundary. We classify the corresponding twisted quantum Drinfeld Hecke algebras. Our results show that the parameter space in the twisted setting is smaller than in the untwisted setting.

Throughout the paper, let $G$ denote a finite group acting linearly on a complex vector space $V$ with basis $v_1, v_2, \ldots, v_n$. Let $\boldsymbol{q} := (q_{ij})_{1\le i,j\le n}$ denote a tuple of nonzero scalars for which $q_{ii} = 1$ and $q_{ji} = q_{ij}^{-1}$ for all $i, j$. We work over the complex numbers $\mathbb{C}$, and all tensor products are taken over $\mathbb{C}$ unless otherwise indicated.

## 2. Necessary and sufficient conditions

In this section, we use Bergman's Diamond Lemma [1978] to give necessary and sufficient conditions for the algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ (defined in the introduction and recalled below) to be a twisted quantum Drinfeld Hecke algebra. First, we recall the notion of a twisted skew group algebra. Let $G$ be a finite group, and let $\alpha : G \times G \to \mathbb{C}^\times$ be a normalized 2-cocycle on $G$, that is,

$$\alpha(g_1, g_2)\alpha(g_1g_2, g_3) = \alpha(g_2, g_3)\alpha(g_1, g_2g_3) \quad \text{and} \quad \alpha(g, 1) = 1 = \alpha(1, g)$$

for all $g, g_1, g_2, g_3 \in G$. Let $A$ be an algebra on which $G$ acts by automorphisms. The *twisted skew group algebra* $A\#_\alpha G$ is defined as follows. As a vector space, $A\#_\alpha G$ is $A \otimes \mathbb{C}G$. Multiplication on $A\#_\alpha G$ is determined by

$$(a \otimes g)(b \otimes h) := \alpha(g, h)a({}^g b) \otimes gh$$

for all $a, b \in A$ and all $g, h \in G$, where a left superscript denotes the action of the group element. The 2-cocycle condition on $\alpha$ ensures that $A\#_\alpha G$ is an associative algebra. Note that $A$ is a subalgebra of $A\#_\alpha G$ via the isomorphism $A \xrightarrow{\sim} A \otimes 1$, and the twisted group algebra $\mathbb{C}^\alpha G$ is a subalgebra of $A\#_\alpha G$ via the isomorphism $\mathbb{C}^\alpha G \xrightarrow{\sim} 1 \otimes \mathbb{C}^\alpha G$. The image of a group element $g$ in the twisted group algebra $\mathbb{C}^\alpha G$ is denoted by $t_g$. To shorten notation, we write the element $a \otimes g$ of $A\#_\alpha G$ by $at_g$. Since $\alpha$ is assumed to be normalized, $t_1$ is the multiplicative identity for $A\#_\alpha G$. For all $g \in G$, we have

$$(t_g)^{-1} = \alpha^{-1}(g, g^{-1})t_{g^{-1}} = \alpha^{-1}(g^{-1}, g)t_{g^{-1}}.$$

Suppose $G$ acts linearly on a complex vector space $V$ with basis $v_1, v_2, \ldots, v_n$, and let $\boldsymbol{q} := (q_{ij})_{1\le i,j\le n}$ denote a tuple of nonzero scalars for which $q_{ii} = 1$ and $q_{ji} = q_{ij}^{-1}$ for all $i, j$. For each group element $g \in G$, let $g_k^i$ denote the scalar determined by the equation

$${}^g v_i = \sum_{k=1}^{n} g_k^i v_k,$$

and define the *quantum* $(i, j, k, l)$*-minor determinant* of $g$ as

$$\det_{ijkl}(g) := g_l^j g_k^i - q_{ji} g_l^i g_k^j.$$

The following lemma is used in the proof of Theorem 2.2.

**Lemma 2.1.** *Suppose that the action of $G$ on $V$ extends to an action on $S_{\boldsymbol{q}}(V)$ by automorphisms, and let $g \in G$. We have*:

(i) $q_{lk} \det_{ijkl}(g) = -\det_{ijlk}(g)$ *for all* $i, j, k, l$.

(ii) *For each* $i, j$, *if* $q_{ij} \neq 1$, *then* $g^i_k g^j_k = 0$ *for all* $k$.

*Proof.* For a proof of (i), see [Levandovskyy and Shepler 2011, Lemma 3.2]. Part (ii) follows from the assumption that $G$ acts on $S_{\boldsymbol{q}}(V)$ by automorphisms and that $q_{ij} \neq 1$: we have ${}^g v_i {}^g v_j = q_{ij} {}^g v_j {}^g v_i$, and so $(\sum_{k=1}^n g^i_k v_k)(\sum_{l=1}^n g^j_l v_l) = q_{ij}(\sum_{k=1}^n g^j_k v_k)(\sum_{l=1}^n g^i_l v_l)$. Equating coefficients of $v_k^2$ yields $g^i_k g^j_k = q_{ij} g^i_k g^j_k$, and since $q_{ij} \neq 1$, we get $g^i_k g^j_k = 0$. $\square$

Let $\kappa : V \times V \to \mathbb{C}^\alpha G$ be a bilinear map for which $\kappa(v_i, v_j) = -q_{ij}\kappa(v_j, v_i)$ for all $1 \leq i, j \leq n$. For each $g \in G$, let $\kappa_g : V \times V \to \mathbb{C}$ be the function determined by the condition

$$\kappa(v, w) = \sum_{g \in G} \kappa_g(v, w) t_g \quad \text{for all } v, w \in V.$$

The condition $\kappa(v_i, v_j) = -q_{ij}\kappa(v_j, v_i)$ implies that $\kappa_g(v_i, v_j) = -q_{ij}\kappa_g(v_j, v_i)$ for all $g \in G$.

Recall that the algebra

$$\mathcal{H}_{\boldsymbol{q},\kappa,\alpha} := T(V)\#_\alpha G/(v_i v_j - q_{ij} v_j v_i - \kappa(v_i, v_j) \mid 1 \leq i, j \leq n)$$

is called a twisted quantum Drinfeld Hecke algebra if it satisfies the Poincaré–Birkhoff–Witt condition: $\operatorname{gr} \mathcal{H}_{\boldsymbol{q},\kappa,\alpha} \cong S_{\boldsymbol{q}}(V)\#_\alpha G$, as graded algebras. This is equivalent to the condition that the set $\{v_1^{m_1} v_2^{m_2} \cdots v_n^{m_n} t_g \mid m_i \geq 0, g \in G\}$ is a $\mathbb{C}$-basis for $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$.

In the proof of the next theorem, we assume familiarity with, and will freely use, terminology from [Bergman 1978] (for example, “reduction system”).

**Theorem 2.2.** *The algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra if and only if the following conditions hold.*

(1) *For all* $g, h \in G$ *and* $1 \leq i < j \leq n$,

$$\frac{\alpha(h, g)}{\alpha(hgh^{-1}, h)} \kappa_g(v_j, v_i) = \sum_{k<l} \det_{ijkl}(h) \kappa_{hgh^{-1}}(v_l, v_k).$$

(2) *For all* $g \in G$ *and* $1 \leq i < j < k \leq n$,

$$\kappa_g(v_k, v_j)({}^g v_i - q_{ji} q_{ki} v_i) + \kappa_g(v_k, v_i)(q_{kj} v_j - q_{ji} {}^g v_j) + \kappa_g(v_j, v_i)(q_{kj} q_{ki} {}^g v_k - v_k) = 0.$$

*Proof.* We begin by expressing the algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ as a quotient of a free associative $\mathbb{C}$-algebra. Let $X = \{v_1, v_2, \ldots, v_n\} \cup \{t_g \mid g \in G\}$, and let $\mathbb{C}\langle X\rangle$ be the free associative $\mathbb{C}$-algebra generated by $X$. Consider the reduction system

$$S=\{(t_g v_i, {}^g v_i t_g), (t_g t_h, \alpha(g,h)t_{gh}), (v_j v_i, q_{ji} v_i v_j+\kappa(v_j, v_i)) \mid g,h\in G, 1\le i<j\le n\}$$

for $\mathbb{C}\langle X\rangle$. Let $I$ be the ideal of $\mathbb{C}\langle X\rangle$ generated by the elements

$$t_g v_i - {}^g v_i t_g, \quad t_g t_h - \alpha(g,h)t_{gh}, \quad v_j v_i - q_{ji} v_i v_j - \kappa(v_j, v_i), \quad g,h\in G, 1\le i<j\le n.$$

In what follows, we use the Diamond Lemma [Bergman 1978] to show that the set

$$\{v_1^{m_1} v_2^{m_2} \cdots v_n^{m_n} t_g \mid m_i \ge 0, g \in G\}$$

is a $\mathbb{C}$-basis for $\mathbb{C}\langle X\rangle/I$ if and only if the two conditions in the statement of the theorem hold.

Define a partial order $\le$ on the free semigroup $\langle X\rangle$ as follow: First, we declare that $v_1 < v_2 < \cdots < v_n < g$ for all $g \in G$, and then we set $A < B$ if

(i) $A$ is of smaller length than $B$, or

(ii) $A$ and $B$ have the same length but $A$ is less than $B$ relative to the lexicographical order.

Then $\le$ is a semigroup partial order on $\langle X\rangle$, compatible with the reduction system $S$, and having the descending chain condition. Thus the hypothesis of the Diamond Lemma holds.

Observe that the set $\langle X\rangle_{\text{irr}}$ of irreducible elements of $\langle X\rangle$ is precisely the alleged $\mathbb{C}$-basis for $\mathbb{C}\langle X\rangle/I$. That is,

$$\langle X\rangle_{\text{irr}} = \{v_1^{m_1} v_2^{m_2} \cdots v_n^{m_n} t_g \mid m_i \ge 0, g \in G\}.$$

In what follows, we show that all ambiguities of $S$ are resolvable if and only if the two conditions in the statement of the theorem hold. The theorem will then follow by the Diamond Lemma. There are no inclusion ambiguities, but there exist overlap ambiguities, and these correspond to the monomials

$$t_g t_h t_k, \quad t_g t_h v_i, \quad t_h v_j v_i, \quad v_k v_j v_i, \quad \text{where } 1 \le i < j < k \le n,\ g, h \in G.$$

Associativity of the multiplication in the twisted group algebra $\mathbb{C}^\alpha G$ implies that the ambiguities corresponding to the monomials $t_g t_h t_k$ are resolvable. The equality ${}^{gh}v_i = {}^g({}^h v_i)$ implies that the ambiguities corresponding to the monomials $t_g t_h v_i$ are resolvable. Next, we show that the ambiguities corresponding to the monomials $t_h v_j v_i$ are resolvable if and only if condition (1) in the statement of the theorem holds. Applying a reduction to the factor $v_j v_i$ in $t_h v_j v_i$, we get

$$q_{ji} t_h v_i v_j + t_h \kappa(v_j, v_i).$$

Applying a reduction to the factor $t_h v_i$ and then to the resulting factor $t_h v_j$ gives

$$\begin{aligned}
&q_{ji}\,{}^h v_i\,{}^h v_j t_h + t_h\kappa(v_j, v_i)\\
&\qquad = q_{ji}\Big(\sum_{l=1}^{n} h^i_l v_l\Big)\Big(\sum_{k=1}^{n} h^j_k v_k\Big) t_h + t_h\kappa(v_j, v_i)\\
&\qquad = q_{ji}\sum_{l<k} h^i_l h^j_k v_l v_k t_h + q_{ji}\sum_{k<l} h^i_l h^j_k v_l v_k t_h + q_{ji}\sum_{k=1}^{n} h^i_k h^j_k v_k^2 t_h + t_h\kappa(v_j, v_i).
\end{aligned}$$

Applying a reduction to the factor $v_l v_k$ in the second summation above yields

$$\begin{aligned}
&q_{ji}\sum_{l<k} h^i_l h^j_k v_l v_k t_h + q_{ji}\sum_{k<l} h^i_l h^j_k q_{lk} v_k v_l t_h\\
&\qquad\qquad + q_{ji}\sum_{k<l} h^i_l h^j_k \kappa(v_l, v_k) t_h + \; q_{ji}\sum_{k=1}^{n} h^i_k h^j_k v_k^2 t_h + t_h\kappa(v_j, v_i).
\end{aligned}$$

Combining the first two summations, expanding $\kappa(v_l, v_k)$ and $\kappa(v_j, v_i)$, and then applying reductions to each term in $\kappa(v_l, v_k)t_h$ and to each term in $t_h\kappa(v_j, v_i)$ gives

$$\begin{aligned}
&q_{ji}\sum_{k<l}\big(h^i_k h^j_l + q_{lk} h^i_l h^j_k\big) v_k v_l t_h + q_{ji}\sum_{k=1}^{n} h^i_k h^j_k v_k^2 t_h\\
&\qquad\qquad + q_{ji}\sum_{g\in G}\Big(\alpha(g, h)\sum_{k<l} h^i_l h^j_k \kappa_g(v_l, v_k)\Big) t_{gh} + \sum_{g\in G}\alpha(h, g)\kappa_g(v_j, v_i) t_{hg}\\
&\qquad = q_{ji}\sum_{k<l}(h^i_k h^j_l + q_{lk} h^i_l h^j_k) v_k v_l t_h + q_{ji}\sum_{k=1}^{n} h^i_k h^j_k v_k^2 t_h\\
&\qquad\quad + \sum_{g\in G}\Big(\alpha(hgh^{-1}, h) q_{ji}\sum_{k<l} h^i_l h^j_k \kappa_{hgh^{-1}}(v_l, v_k) + \alpha(h, g)\kappa_g(v_j, v_i)\Big) t_{hg}.
\end{aligned}$$

Next, we apply to $t_h v_j v_i$ a reduction different from the one in the computation above: Applying a reduction to the factor $t_h v_j$ in $t_h v_j v_i$, and then to the resulting factor $t_h v_i$, we get

$$\begin{aligned}
{}^h v_j\,{}^h v_i t_h &= \Big(\sum_{l=1}^{n} h^j_l v_l\Big)\Big(\sum_{k=1}^{n} h^i_k v_k\Big) t_h\\
&= \sum_{l<k} h^j_l h^i_k v_l v_k t_h + \sum_{k<l} h^j_l h^i_k v_l v_k t_h + \sum_{k=1}^{n} h^j_k h^i_k v_k^2 t_h.
\end{aligned}$$

Applying a reduction to the factor $v_l v_k$ in the second summation above yields

$$\sum_{l<k} h^j_l h^i_k v_l v_k t_h + \sum_{k<l} q_{lk} h^j_l h^i_k v_k v_l t_h + \sum_{k<l} h^j_l h^i_k \kappa(v_l, v_k) t_h + \sum_{k=1}^{n} h^j_k h^i_k v_k^2 t_h.$$

Combining the first two summations, expanding $\kappa(v_l, v_k)$, and then applying a reduction to each term in $\kappa(v_l, v_k)t_h$ gives

$$\begin{aligned}\sum_{k<l}(h_k^j h_l^i + q_{lk} h_l^j h_k^i) v_k v_l t_h + \sum_{k=1}^{n} h_k^j h_k^i v_k^2 t_h + \sum_{g\in G}\Big(\alpha(g,h)\sum_{k<l} h_l^j h_k^i \kappa_g(v_l, v_k)\Big) t_{gh} \\ = \sum_{k<l}(h_k^j h_l^i + q_{lk} h_l^j h_k^i) v_k v_l t_h + \sum_{k=1}^{n} h_k^j h_k^i v_k^2 t_h \\ + \sum_{g\in G}\Big(\alpha(hgh^{-1},h)\sum_{k<l} h_l^j h_k^i \kappa_{hgh^{-1}}(v_l, v_k)\Big) t_{hg}.\end{aligned}$$

By equating coefficients, we see that the final expressions in the previous two computations are equal if and only if

(a) $q_{ji} h_k^i h_l^j + q_{ji} q_{lk} h_l^i h_k^j = h_k^j h_l^i + q_{lk} h_l^j h_k^i$ for all $k < l$,

(b) $q_{ji} h_k^i h_k^j = h_k^i h_k^j$ for all $k$, and

(c) for all $g \in G$, we have

$$\begin{aligned}\alpha(hgh^{-1},h) q_{ji}\sum_{k<l} h_l^i h_k^j \kappa_{hgh^{-1}}(v_l, v_k) + \alpha(h,g)\kappa_g(v_j, v_i) \\ = \alpha(hgh^{-1},h)\sum_{k<l} h_l^j h_k^i \kappa_{hgh^{-1}}(v_l, v_k).\end{aligned}$$

Conditions (a) and (b) follow from Lemma 2.1(i) and (ii), respectively. The equation in (c) is equivalent to condition (1) in the statement of the theorem.

Lastly, we show that the ambiguities corresponding to the monomials $v_k v_j v_i$ are resolvable if and only if condition (2) in the statement of the theorem holds. Applying a reduction to the factor $v_k v_j$ in $v_k v_j v_i$, we get

$$q_{kj} v_j v_k v_i + \kappa(v_k, v_j) v_i.$$

Applying a reduction to the factor $v_k v_i$ gives

$$q_{kj} q_{ki} v_j v_i v_k + q_{kj} v_j \kappa(v_k, v_i) + \kappa(v_k, v_j) v_i.$$

Applying a reduction to the factor $v_j v_i$ yields

$$q_{kj} q_{ki} q_{ji} v_i v_j v_k + q_{kj} q_{ki} \kappa(v_j, v_i) v_k + q_{kj} v_j \kappa(v_k, v_i) + \kappa(v_k, v_j) v_i.$$

Expanding $\kappa(v_j, v_i)$, $\kappa(v_k, v_i)$, and $\kappa(v_k, v_j)$, applying reductions to each term in $\kappa(v_j, v_i)v_k$ and to each term in $\kappa(v_k, v_j)v_i$, and then rearranging gives

$$q_{kj} q_{ki} q_{ji} v_i v_j v_k + \sum_{g\in G}(\kappa_g(v_k, v_j)\,{}^g v_i + q_{kj}\kappa_g(v_k, v_i) v_j + q_{kj} q_{ki} \kappa_g(v_j, v_i)\,{}^g v_k) t_g.$$

Next, we apply to $v_k v_j v_i$ a reduction different from the one in the computation above: Applying a reduction to the factor $v_j v_i$ in $v_k v_j v_i$, we get

$$q_{ji} v_k v_i v_j + v_k \kappa(v_j, v_i).$$

Applying a reduction to the factor $v_k v_i$ gives

$$q_{ji} q_{ki} v_i v_k v_j + q_{ji} \kappa(v_k, v_i) v_j + v_k \kappa(v_j, v_i).$$

Applying a reduction to the factor $v_k v_j$ yields

$$q_{ji} q_{ki} q_{kj} v_i v_j v_k + q_{ji} q_{ki} v_i \kappa(v_k, v_j) + q_{ji} \kappa(v_k, v_i) v_j + v_k \kappa(v_j, v_i).$$

Expanding $\kappa(v_k, v_j)$, $\kappa(v_k, v_i)$, and $\kappa(v_j, v_i)$, and then applying reductions to each term in $\kappa(v_k, v_i) v_j$ gives

$$q_{ji} q_{ki} q_{kj} v_i v_j v_k + \sum_{g \in G} (q_{ji} q_{ki} \kappa_g(v_k, v_j) v_i + q_{ji} \kappa_g(v_k, v_i)^g v_j + \kappa_g(v_j, v_i) v_k) t_g.$$

The final expressions in the two computations above are equal if and only if condition (2) in the statement of the theorem holds. This finishes the proof. □

## 3. Deformations

The primary goal of this section is to show that the twisted quantum Drinfeld Hecke algebras $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ are isomorphic to specializations of particular types of deformations of the twisted skew group algebras $S_{\boldsymbol{q}}(V) \#_\alpha G$.

Let $\hbar$ denote an indeterminate. Recall that, for a $\mathbb{C}$-algebra $A$, a *deformation of $A$ over* $\mathbb{C}[\hbar]$ is an associative $\mathbb{C}[\hbar]$-algebra whose underlying vector space is $A[\hbar] = \mathbb{C}[\hbar] \otimes A$, and which reduces modulo $\hbar$ to the original algebra $A$. Thus the multiplication $\mu$ on $A[\hbar]$ is determined by

$$\mu(a, b) = \mu_0(a, b) + \mu_1(a, b)\hbar + \mu_2(a, b)\hbar^2 + \cdots$$

for all $a, b \in A$, where $\mu_0(a, b)$ is the product in $A$, the $\mu_i : A \times A \to A$ are $\mathbb{C}$-bilinear maps extended to be bilinear over $\mathbb{C}[\hbar]$, and for each pair $(a, b)$ the sum above is finite. A consequence of associativity of $\mu$ is that $\mu_1$ is a *Hochschild 2-cocycle*, that is,

$$a\mu_1(b, c) + \mu_1(a, bc) = \mu_1(ab, c) + \mu_1(a, b)c \tag{3.1}$$

for all $a, b, c \in A$.

In order to see that the twisted quantum Drinfeld Hecke algebras $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ may be realized as specializations of deformations of $S_{\boldsymbol{q}}(V) \#_\alpha G$, we define the algebra

$$\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar} := (T(V) \#_\alpha G)[\hbar] / (v_i v_j - q_{ij} v_j v_i - \kappa(v_i, v_j)\hbar \mid 1 \le i, j \le n).$$

Assigning $\hbar$ degree zero, each $v_i$ degree one, and each $t_g$ ($g \in G$) degree zero, we see that $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a filtered algebra, and that $(S_{\boldsymbol{q}}(V) \#_\alpha G)[\hbar]$ is a graded algebra. We call the algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ a *twisted quantum Drinfeld Hecke algebra over* $\mathbb{C}[\hbar]$ if $\operatorname{gr} \mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar} \cong (S_{\boldsymbol{q}}(V) \#_\alpha G)[\hbar]$, as graded algebras. Specializing a twisted quantum

Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$ to $\hbar = 1$ yields the twisted quantum Drinfeld Hecke algebra over $\mathbb{C}$, as defined earlier.

In the following theorem, by the *degree* of $\mu_i$ we mean its degree as a function between graded algebras.

**Theorem 3.2.** *Every twisted quantum Drinfeld Hecke algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ over $\mathbb{C}[\hbar]$ is isomorphic to some deformation $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$ with* $\deg \mu_i = -2i$ *for all $i \geq 1$.*

*Proof.* Suppose that $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$. Consider the natural projection $T(V)\#_\alpha G \to S_{\boldsymbol{q}}(V)\#_\alpha G$, and let $s : S_{\boldsymbol{q}}(V)\#_\alpha G \to T(V)\#_\alpha G$ be the $\mathbb{C}$-linear section determined by the ordering $v_1, v_2, \ldots, v_n$ of the basis of $V$. For example, $s(v_2 v_1^2 t_g) = q_{21}^2 v_1^2 v_2 t_g$.

Extend $s$ to a $\mathbb{C}[\hbar]$-linear map $\tilde{s} : (S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar] \to (T(V)\#_\alpha G)[\hbar]$, and let $p$ denote the natural projection from $(T(V)\#_\alpha G)[\hbar]$ to $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$. Since $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$, the composition $f := p \circ \tilde{s}$ is an isomorphism of $\mathbb{C}[\hbar]$-modules.

Next, define a $\mathbb{C}[\hbar]$-bilinear multiplication $\mu$ on $(S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar]$ by

$$\mu := f^{-1} \circ \mathrm{mult} \circ (f \times f),$$

where mult is the multiplication map in $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$. Since $\mu$ is $\mathbb{C}[\hbar]$-bilinear, it must necessarily be a power series

$$\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar + \cdots,$$

where the $\mu_i$ are $\mathbb{C}$-bilinear maps from $(S_{\boldsymbol{q}}(V)\#_\alpha G) \times (S_{\boldsymbol{q}}(V)\#_\alpha G)$ to $S_{\boldsymbol{q}}(V)\#_\alpha G$. Note that, by definition of $f$, the map $\mu_0$ is precisely the multiplication map in $S_{\boldsymbol{q}}(V)\#_\alpha G$, and so $\mu$ is a deformation $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$. By definition, the map $f$ is an isomorphism between the $\mathbb{C}[\hbar]$-algebras $(S_{\boldsymbol{q}}(V)\#_\alpha G[\hbar], \mu)$ and $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$, proving that $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is isomorphic to a deformation of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$.

Finally we prove the degree condition on the $\mu_i$. Given elements

$$a = v_1^{\beta_1} v_2^{\beta_2} \cdots v_n^{\beta_n} t_g \quad \text{and} \quad b = v_1^{\gamma_1} v_2^{\gamma_2} \cdots v_n^{\gamma_n} t_h$$

in $S_{\boldsymbol{q}}(V)\#_\alpha G$, to find $\mu_1(a,b), \mu_2(a,b), \ldots$, we must put the product $f(a)f(b) \in \mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ in the normal form by repeatedly applying the relations defining $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$. Induction on the degree $\sum_{k=1}^n \beta_k + \gamma_k$ of $ab$ implies that $\deg \mu_i = -2i$ for all $i \geq 1$, as claimed. □

**Lemma 3.3.** *The algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}$ if and only if $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$.*

*Proof.* The proof given for $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ in Theorem 2.2 generalizes for $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ by extending scalars to $\mathbb{C}[\hbar]$. That is, $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$ if and only if the two conditions in Theorem 2.2 hold. □

**Corollary 3.4.** *Every twisted quantum Drinfeld Hecke algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is isomorphic to a specialization of a deformation $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$ with* $\deg \mu_i = -2i$ *for all* $i \geq 1$.

A Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$ is said to be *constant* if it is of degree $-2$ as a function between graded algebras. In the next section, it is shown that such 2-cocycles correspond to certain *constant* polynomials, justifying the choice of terminology.

**Proposition 3.5.** *Let $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ be a twisted quantum Drinfeld Hecke algebra. The map $\kappa : V \times V \to \mathbb{C}^\alpha G$ is equal to the* quantum skew-symmetrization *of some constant Hochschild 2-cocycle $\mu_1$ on $S_{\boldsymbol{q}}(V)\#_\alpha G$, that is,*

$$\kappa(v_i, v_j) = \mu_1(v_i, v_j) - q_{ij}\mu_1(v_j, v_i)$$

*for all $i$, $j$.*

*Proof.* By Lemma 3.3, $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$. By Theorem 3.2, associated to $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a deformation $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$ with $\deg \mu_i = -2i$ for all $i \geq 1$. Note that $\mu_1$ is a constant Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$. We claim that $\kappa$ is equal to the quantum skew-symmetrization of $\mu_1$.

Let $f$ be the map defined in the proof of Theorem 3.2. For any two monomials $a, b \in S_{\boldsymbol{q}}(V)\#_\alpha G$, the value of $\mu_1(a, b)$ is determined by writing the product $f(a)f(b) \in \mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ in the normal form by repeatedly applying the relations defining $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$. If $i \leq j$, the product $f(v_i)f(v_j) = v_iv_j$ is already in the desired form, so $\mu_1(v_i, v_j) = 0$. If $i > j$, we write $v_iv_j \longrightarrow q_{ij}v_jv_i + \kappa(v_i, v_j)\hbar$, and so $\kappa(v_i, v_j) = \mu_1(v_i, v_j)$. If $i \leq j$, we have $\kappa(v_i, v_j) = -q_{ij}\kappa(v_j, v_i) = -q_{ij}\mu_1(v_j, v_i)$. Thus $\kappa(v_i, v_j) = \mu_1(v_i, v_j) - q_{ij}\mu_1(v_j, v_i)$ for all $i$, $j$. □

The proof of the following theorem is a generalization of [Naidu and Witherspoon 2011, Theorem 2.2]; see also [Witherspoon 2007, Theorem 3.2].

**Theorem 3.6.** *Every deformation $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$ with* $\deg \mu_i = -2i$ *for all $i \geq 1$ is isomorphic to some twisted quantum Drinfeld Hecke algebra over* $\mathbb{C}[\hbar]$.

*Proof.* Suppose that $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ is a deformation of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$ with $\deg \mu_i = -2i$ for all $i \geq 1$. In what follows, we identity $T(V)\#_\alpha G$ with the free associative $\mathbb{C}$-algebra generated by the set $\{v_1, v_2, \ldots, v_n\} \cup \{t_g \mid g \in G\}$ subject to the relations $t_gv_i = {}^g v_i t_g$ and $t_gt_h = \alpha(g, h)t_{gh}$ for all $i \in \{1, 2, \ldots, n\}$ and all $g, h \in G$. Define a map $\phi : (T(V)\#_\alpha G)[\hbar] \to (S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar]$ as follows. First, set $\phi(v_i) = v_i$ and $\phi(t_g) = t_g$ for all $i \in \{1, 2, \ldots, n\}$ and all $g \in G$. Since $\deg \mu_k = -2k$ for all $k \geq 1$, we have

$$\mu_k(\mathbb{C}^\alpha G, \mathbb{C}^\alpha G) = \mu_k(\mathbb{C}^\alpha G, V) = \mu_k(V, \mathbb{C}^\alpha G) = 0$$

for all $k \geq 1$. This implies that the relations $t_g v_i = {}^g v_i t_g$ and $t_g t_h = \alpha(g,h) t_{gh}$ hold in the algebra $((S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar],\ \mu)$, and so we obtain a $\mathbb{C}$-algebra homomorphism on $T(V)\#_\alpha G$, which extends to a $\mathbb{C}[\hbar]$-algebra homomorphism $\phi$ from $(T(V)\#_\alpha G)[\hbar]$ to $(S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar]$, where the algebra structure on the latter is given by $\mu$.

Next, we show that $\phi$ is surjective. It is enough to show that each monomial $v_{i_1} \cdots v_{i_m} t_g$ is in the image of $\phi$. The proof is by induction on the degree of the monomial. Suppose that all monomials of degree less than $m$ are in the image of $\phi$. In particular, $\phi(X) = v_{i_2} \cdots v_{i_m} g$ for some $X \in (T(V)\#_\alpha G)[\hbar]$. Then

$$\begin{aligned}\phi(v_{i_1} X) &= \mu(v_{i_1}, \phi(X)) \\ &= \mu(v_{i_1}, v_{i_2} \cdots v_{i_m} t_g) \\ &= v_{i_1} \cdots v_{i_m} t_g + \mu_1(v_{i_1}, v_{i_2} \cdots v_{i_m} t_g)\hbar + \mu_2(v_{i_1}, v_{i_2} \cdots v_{i_m} t_g)\hbar^2 + \cdots .\end{aligned}$$

Since $\deg(\mu_k) = -2k$, by the induction hypothesis, each $\mu_k(v_{i_1}, v_{i_2} \cdots v_{i_m} t_g)$ is in the image of $\phi$. Therefore, $v_{i_1} \cdots v_{i_m} t_g$ is in the image of $\phi$, and it follows that $\phi$ is surjective.

Finally, we determine the kernel of $\phi$. Since $\deg(\mu_1) = -2$, we can define a bilinear map $\kappa : V \times V \to \mathbb{C}^\alpha G$ by setting $\kappa(v_i, v_j) := \mu_1(v_i, v_j) - q_{ij}\mu_1(v_j, v_i)$ for all $i, j$. Let $I$ denote the ideal in $(T(V)\#_\alpha G)[\hbar]$ generated by the elements

$$v_i v_j - q_{ij} v_j v_i - \kappa(v_i, v_j)\hbar.$$

Since $\mu_k(v_i, v_j) = 0$ for all $k \geq 2$, we have

$$\begin{aligned}\phi(v_i v_j) &= \mu(v_i, v_j) = v_i v_j + \mu_1(v_i, v_j)\hbar, \\ \phi(v_j v_i) &= \mu(v_j, v_i) = v_j v_i + \mu_1(v_j, v_i)\hbar,\end{aligned}$$

and so $I$ is contained in the kernel of $\phi$. The form of the relations and surjectivity of $\phi$ imply that the kernel of $\phi$ is precisely $I$, and it follows that the deformation $((S_{\boldsymbol{q}}(V)\#_\alpha G)[\hbar],\ \mu)$ is isomorphic to the twisted quantum Drinfeld Hecke algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ over $\mathbb{C}[\hbar]$. □

## 4. Computing $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$

Let $A$ be an algebra on which the finite group $G$ acts by automorphisms, and let $\alpha$ be a 2-cocycle on $G$. This section is concerned with the computation of the Hochschild cohomology $\mathrm{HH}^\bullet(A\#_\alpha G)$ of the twisted skew group algebra $A\#_\alpha G$. We are particularly interested in degree two cohomology in the case when $A$ is the quantum symmetric algebra $S_{\boldsymbol{q}}(V)$. The results of this section are used in the sections that follow.

Recall that the Hochschild cohomology of an algebra $R$ is

$$\mathrm{HH}^\bullet(R) := \mathrm{Ext}^\bullet_{R^e}(R, R),$$

where the enveloping algebra $R^e := R \otimes R^{\mathrm{op}}$ acts on $R$ by left and right multiplication. When $R$ is a twisted skew group algebra $A\#_\alpha G$ in a characteristic not dividing the order of the finite group $G$, by [Ştefan 1995, Corollary 3.4], there is an action of $G$ on $\mathrm{HH}^\bullet(A, A\#_\alpha G) = \mathrm{Ext}^\bullet_{A^e}(A, A\#_\alpha G)$ for which $\mathrm{HH}^\bullet(A\#_\alpha G)$ is isomorphic to $\mathrm{HH}^\bullet(A, A\#_\alpha G)^G$, the space of elements of $\mathrm{HH}^\bullet(A, A\#_\alpha G)$ that are invariant under the action of $G$. Thus, one can compute $\mathrm{HH}^\bullet(A\#_\alpha G)$ by first computing $\mathrm{HH}^\bullet(A, A\#_\alpha G)$ and then determining the space of $G$-invariant elements. When $A$ is the quantum symmetric algebra $S_{\boldsymbol{q}}(V)$, we compute $\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#_\alpha G)$ using the quantum Koszul resolution, recalled below.

The *quantum exterior algebra* $\bigwedge_{\boldsymbol{q}}(V)$ associated to the tuple $\boldsymbol{q} = (q_{ij})$ is

$$\textstyle\bigwedge_{\boldsymbol{q}}(V) := \mathbb{C}\langle v_1, \ldots, v_n \mid v_i v_j = -q_{ij} v_j v_i \text{ for all } 1 \le i, j \le n\rangle.$$

Since we are working in characteristic 0, the defining relations imply in particular that $v_i^2 = 0$ for each $v_i$ in $\bigwedge_{\boldsymbol{q}}(V)$. This algebra has a basis given by all $v_{i_1} \cdots v_{i_m}$ $(0 \le m \le n,\ 1 \le i_1 < \cdots < i_m \le n)$; we write such a basis element as $v_{i_1} \wedge \cdots \wedge v_{i_m}$ by analogy with the ordinary exterior algebra.

By [Wambst 1993, Proposition 4.1(c)], the following is a free $S_{\boldsymbol{q}}(V)^e$-resolution of $S_{\boldsymbol{q}}(V)$:

$$\cdots \to S_{\boldsymbol{q}}(V)^e \otimes \textstyle\bigwedge^2_{\boldsymbol{q}}(V) \xrightarrow{d_2} S_{\boldsymbol{q}}(V)^e \otimes \bigwedge^1_{\boldsymbol{q}}(V) \xrightarrow{d_1} S_{\boldsymbol{q}}(V)^e \xrightarrow{\text{mult}} S_{\boldsymbol{q}}(V) \to 0, \tag{4.1}$$

that is, for $1 \le m \le n$, the degree $m$ term is $S_{\boldsymbol{q}}(V)^e \otimes \bigwedge^m_{\boldsymbol{q}}(V)$; the differential $d_m$ is defined by

$$\begin{aligned}&d_m(1^{\otimes 2} \otimes v_{j_1} \wedge \cdots \wedge v_{j_m})\\ &= \sum_{i=1}^{m} (-1)^{i+1} \Bigg[ \bigg( \prod_{s=1}^{i} q_{j_s, j_i} \bigg) v_{j_i} \otimes 1 - \bigg( \prod_{s=i}^{m} q_{j_i, j_s} \bigg) \otimes v_{j_i} \Bigg] \otimes v_{j_1} \wedge \cdots \wedge \hat{v}_{j_i} \wedge \cdots \wedge v_{j_m}\end{aligned}$$

whenever $1 \le j_1 < \cdots < j_m \le n$, and mult denotes the multiplication map. The complex (4.1) is a quantum version of the usual Koszul resolution for a polynomial ring.

Suppose that the action of $G$ on $V$ induces an action on $\bigwedge_{\boldsymbol{q}}(V)$. Thus, there is an action of $G$ on the quantum Koszul complex (4.1), that is, an action of $G$ on each $S_{\boldsymbol{q}}(V)^e \otimes \bigwedge^i_{\boldsymbol{q}}(V)$ that commutes with the differentials.

Assume that $\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V)\#_\alpha G)$ has been computed using the quantum Koszul resolution. So, elements of $\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V)\#_\alpha G)$ are given as $G$-invariant elements of $\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#_\alpha G)$. For our purposes, we need to find representatives for elements in $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$ that are given as maps from $(S_{\boldsymbol{q}}(V)\#_\alpha G) \otimes (S_{\boldsymbol{q}}(V)\#_\alpha G)$ to $S_{\boldsymbol{q}}(V)\#_\alpha G$ satisfying the 2-cocycle condition (3.1). To this end, we consider chain maps between the quantum Koszul resolution (4.1) and the bar resolution of $A$:

$$\begin{tikzcd} \cdots \arrow[r] & S_{\boldsymbol{q}}(V)^{\otimes 4} \arrow[r, "\delta_2"] \arrow[d, shift right, "\Psi_2"'] & S_{\boldsymbol{q}}(V)^{\otimes 3} \arrow[r, "\delta_1"] \arrow[d, shift right, "\Psi_1"'] & S_{\boldsymbol{q}}(V)^e \arrow[r, "\text{mult}"] \arrow[d, leftrightarrow, "="'] & S_{\boldsymbol{q}}(V) \arrow[r] \arrow[d, leftrightarrow, "="'] & 0 \\ \cdots \arrow[r] & S_{\boldsymbol{q}}(V)^e \otimes \bigwedge_{\boldsymbol{q}}^2 V \arrow[r, "d_2"] \arrow[u, shift right, "\Phi_2"'] & S_{\boldsymbol{q}}(V)^e \otimes \bigwedge_{\boldsymbol{q}}^1 V \arrow[r, "d_1"] \arrow[u, shift right, "\Phi_1"'] & S_{\boldsymbol{q}}(V)^e \arrow[r, "\text{mult}"] & S_{\boldsymbol{q}}(V) \arrow[r] & 0. \end{tikzcd}$$

Here the differentials $\delta_i$ in the bar resolution are defined as

$$\delta_i(a_0 \otimes \cdots \otimes a_{i+1}) = \sum_{j=0}^{i} (-1)^j a_0 \otimes \cdots \otimes a_j a_{j+1} \otimes \cdots \otimes a_{i+1}$$

for all $a_0, \ldots, a_{i+1} \in A$. We will only need to know the values of $\Psi_2$ on elements of the form $1 \otimes v_i \otimes v_j \otimes 1$, and these can be chosen to be

$$\Psi_2(1 \otimes v_i \otimes v_j \otimes 1) = \begin{cases} 1 \otimes 1 \otimes v_i \wedge v_j & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \tag{4.2}$$

Chain maps $\Phi_i$ are defined in [Naidu et al. 2011], and more generally in [Wambst 1993], that embed the quantum Koszul resolution as a subcomplex of the bar resolution. We will only need $\Phi_2$, and this is defined by

$$\Phi_m(1 \otimes 1 \otimes v_i \wedge \wedge v_j) = 1 \otimes v_i \otimes v_j \otimes 1 - q_{ij} \otimes v_j \otimes v_i \otimes 1 \tag{4.3}$$

for all $1 \leq i, j \leq n$.

We define the Reynold's operator, or averaging map, which ensures $G$-invariance of the image, compensating for the possibility that $\Psi_2$ may not preserve the action of $G$:

$$\mathcal{R}_2 : \operatorname{Hom}_{\mathbb{C}}(S_{\boldsymbol{q}}(V)^{\otimes 2}, S_{\boldsymbol{q}}(V) \#_\alpha G) \to \operatorname{Hom}_{\mathbb{C}}(S_{\boldsymbol{q}}(V)^{\otimes 2}, S_{\boldsymbol{q}}(V) \#_\alpha G)^G,$$
$$\mathcal{R}_2(\gamma) := \frac{1}{|G|} \sum_{g \in G} {}^g\gamma.$$

The following map tells how to extend a function defined on $S_{\boldsymbol{q}}(V)^{\otimes 2}$ to a function defined on $(S_{\boldsymbol{q}}(V)\#_\alpha G)^{\otimes 2}$ [Căldăraru et al. 2004]:

$$\Theta_2^* : \operatorname{Hom}_{\mathbb{C}}(S_{\boldsymbol{q}}(V)^{\otimes 2}, S_{\boldsymbol{q}}(V) \#_\alpha G)^G \to \operatorname{Hom}_{\mathbb{C}}((S_{\boldsymbol{q}}(V) \#_\alpha G)^{\otimes 2}, S_{\boldsymbol{q}}(V) \#_\alpha G),$$
$$\Theta_2^*(\kappa)(a_1 t_{g_1} \otimes a_2 t_{g_2}) := \alpha(g_1, g_2) \kappa(a_1 \otimes {}^{g_1}a_2) t_{g_1 g_2}.$$

The theorem below is from [Căldăraru et al. 2004]; see also [Shepler and Witherspoon 2012].

**Theorem 4.4.** *Suppose that the action of $G$ on $V$ extends to an action on $\bigwedge_{\boldsymbol{q}}(V)$ by automorphisms. The map*

$$\Theta_2^* \mathcal{R}_2 \Psi_2^* : \operatorname{Hom}_{\mathbb{C}}\big(\textstyle\bigwedge_{\boldsymbol{q}}^2(V), S_{\boldsymbol{q}}(V) \#_\alpha G\big) \to \operatorname{Hom}_{\mathbb{C}}(S_{\boldsymbol{q}}(V)^{\otimes 2}, S_{\boldsymbol{q}}(V) \#_\alpha G)$$

*induces an isomorphism*

$$\mathrm{HH}^2(S_{\boldsymbol{q}}(V),\, S_{\boldsymbol{q}}(V)\#_\alpha G)^G \xrightarrow{\sim} \mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G).$$

*Moreover*, $\Theta_2^*\mathcal{R}_2\Psi_2^*$ *maps* $\mathrm{HH}^2(S_{\boldsymbol{q}}(V),\, S_{\boldsymbol{q}}(V)\#_\alpha G)$ *onto* $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$.

Next, we will introduce some notation and give some formulas that are useful in the sections that follow. For each $g \in G$, the space $S_{\boldsymbol{q}}(V)t_g \subseteq S_{\boldsymbol{q}}(V)\#_\alpha G$ is a (left) $S_{\boldsymbol{q}}(V)^e$-module via the action

$$(a \otimes b) \cdot (ct_g) := act_g b = ac({}^g b)t_g$$

for all $a, b, c \in S_{\boldsymbol{q}}(V)$ and all $g \in G$. Note that $\mathrm{HH}^2(S_{\boldsymbol{q}}(V),\, S_{\boldsymbol{q}}(V)\#_\alpha G)$ is isomorphic to the direct sum $\bigoplus_{g\in G} \mathrm{HH}^2(S_{\boldsymbol{q}}(V),\, S_{\boldsymbol{q}}(V)t_g)$.

We wish to express the formula for the differentials $d_m$ in the quantum Koszul resolution (4.1) in a more convenient form. To this end, let $\mathbb{N}^n$ denote the set of all $n$-tuples of elements from $\mathbb{N}$. The *length* of $\gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{N}^n$, denoted $|\gamma|$, is the sum $\sum_{i=1}^n \gamma_i$. For each $\gamma \in \mathbb{N}^n$, define $v^\gamma := v_1^{\gamma_1} v_2^{\gamma_2} \cdots v_n^{\gamma_n}$. For each $i \in \{1, \dots, n\}$, define $[i] \in \mathbb{N}^n$ by $[i]_j = \delta_{i,j}$, for all $j \in \{1, \dots, n\}$. For each $\beta = (\beta_1, \dots, \beta_n) \in \{0, 1\}^n$, let $v^{\wedge\beta}$ denote the vector $v_{j_1} \wedge \cdots \wedge v_{j_m} \in \bigwedge_{\boldsymbol{q}}^m(V)$ determined by the conditions $m = |\beta|$, $\beta_{j_k} = 1$ for all $k \in \{1, \dots, m\}$, and $j_1 < \ldots < j_m$. For each $\beta \in \{0, 1\}^n$ with $|\beta| = m$, we have

$$d_m(1^{\otimes 2} \otimes v^{\wedge\beta}) = \sum_{i=1}^n \delta_{\beta_i,1}(-1)^{\sum_{s=1}^{i-1}\beta_s} \left[ \left( \prod_{s=1}^i q_{s,i}^{\beta_s} \right) v_i \otimes 1 - \left( \prod_{s=i}^n q_{i,s}^{\beta_s} \right) \otimes v_i \right] \otimes v^{\wedge(\beta-[i])}.$$

Removing the term $S_{\boldsymbol{q}}(V)$ from the quantum Koszul resolution (4.1), applying the functor $\mathrm{Hom}_{S_{\boldsymbol{q}}(V)^e}(\,\cdot\,,\; S_{\boldsymbol{q}}(V)t_g)$, and then identifying

$$\mathrm{Hom}_{S_{\boldsymbol{q}}(V)^e}\big(S_{\boldsymbol{q}}(V)^e \otimes {\textstyle\bigwedge_{\boldsymbol{q}}^\bullet}(V),\, S_{\boldsymbol{q}}(V)t_g\big) \cong \mathrm{Hom}_{\mathbb{C}}\big({\textstyle\bigwedge_{\boldsymbol{q}}^\bullet}(V),\, S_{\boldsymbol{q}}(V)t_g\big)$$

with $S_{\boldsymbol{q}}(V)t_g \otimes \bigwedge_{\boldsymbol{q}^{-1}}^\bullet(V^*)$, we obtain the complex

$$0 \to S_{\boldsymbol{q}}(V)t_g \xrightarrow{d_1^*} S_{\boldsymbol{q}}(V)t_g \otimes {\textstyle\bigwedge_{\boldsymbol{q}^{-1}}^1}(V^*) \xrightarrow{d_2^*} S_{\boldsymbol{q}}(V)t_g \otimes {\textstyle\bigwedge_{\boldsymbol{q}^{-1}}^2}(V^*) \to \cdots. \tag{4.5}$$

For all $a \in S_{\boldsymbol{q}}(V)$ and all $\beta \in \{0, 1\}^n$ with $|\beta| = m - 1$, the differential $d_m^*$ sends the element $at_g \otimes (v^*)^{\wedge\beta}$ to

$$\sum_{i=1}^n \delta_{\beta_i,0}(-1)^{\sum_{s=1}^{i}\beta_s} \left[ \left( \left( \prod_{s=1}^i q_{s,i}^{\beta_s} \right) v_i a - \left( \prod_{s=i}^n q_{i,s}^{\beta_s} \right) a({}^g v_i) \right) t_g \right] \otimes (v^*)^{\wedge(\beta+[i])}. \tag{4.6}$$

For later use, we record the following formula. Let $\eta \in (S_{\boldsymbol{q}}(V)\#_\alpha G) \otimes \bigwedge_{\boldsymbol{q}^{-1}}^2(V^*)$. Then

$$[\Theta_2^*\mathcal{R}_2\Psi_2^*(\eta)](v_i \otimes v_j) = \frac{1}{|G|} \sum_{g\in G} {}^g\big(\eta(\Psi_2(1 \otimes {}^{g^{-1}}v_i \otimes {}^{g^{-1}}v_j \otimes 1))\big). \tag{4.7}$$

The elements of $\big((S_{\boldsymbol{q}}(V)\#_\alpha G)\otimes \bigwedge^2_{\boldsymbol{q}^{-1}}(V^*)\big)^G$ that correspond to *constant* Hochschild 2-cocycles, i.e., those of degree $-2$ as maps from $(S_{\boldsymbol{q}}(V)\#_\alpha G)\otimes(S_{\boldsymbol{q}}(V)\#_\alpha G)$ to $S_{\boldsymbol{q}}(V)\#_\alpha G$, are precisely those in $\big(\mathbb{C}^\alpha G\otimes \bigwedge^2_{\boldsymbol{q}^{-1}}(V^*)\big)^G$, due to the form of the chain map $\Psi_2$. Note that the intersection of the image of $d_2^*$ with $\mathbb{C}^\alpha G\otimes \bigwedge^2_{\boldsymbol{q}^{-1}}(V^*)$ is 0. Applying our earlier formula, letting $\beta=[j]+[k]$,

$$(4.8)\quad d_3^*(t_g\otimes v_j^*\wedge v_k^*) = \sum_{i\notin\{j,k\}} (-1)^{\sum_{s=1}^{i}\beta_s}\Bigg[\bigg(\Big(\prod_{s=1}^{i} q_{s,i}^{\beta_s}\Big)v_i-\Big(\prod_{s=i}^{n} q_{i,s}^{\beta_s}\Big){}^g v_i\bigg)t_g\Bigg]\otimes(v^*)^{\wedge(\beta+[i])}.$$

## 5. Constant Hochschild 2-cocycles

In this section, we establish the following bijection:

$$\left\{\begin{array}{c}\text{constant Hochschild}\\ \text{2-cocycles on } S_{\boldsymbol{q}}(V)\#_\alpha G\end{array}\right\}\longleftrightarrow\left\{\begin{array}{c}\text{twisted quantum Drinfeld}\\ \text{Hecke algebras } \mathcal{H}_{\boldsymbol{q},\kappa,\alpha}\end{array}\right\}.$$

We also show that every constant Hochschild 2-cocycles on $S_{\boldsymbol{q}}(V)\#_\alpha G$ lifts to a deformation of $S_{\boldsymbol{q}}(V)\#_\alpha G$.

We use the following two lemmas shortly.

**Lemma 5.1.** *The action of $G$ on $V$ extends to an action on $\bigwedge_{\boldsymbol{q}}(V)$ by automorphisms if, and only if, for all $g\in G$, $i\neq j$, and $k<l$,*

$$(1-q_{ij}q_{lk})g_k^i g_l^j+(q_{ij}-q_{lk})g_l^i g_k^j=0.$$

*Proof.* See [Naidu and Witherspoon 2011, Lemma 4.2]. □

**Lemma 5.2.** *Suppose that the action of $G$ on $V$ extends to an action, by automorphisms, on $S_{\boldsymbol{q}}(V)$ and on $\bigwedge_{\boldsymbol{q}}(V)$. Then, for all $g\in G$ and all $i,j,k,l$ ($i<j$, $k<l$), if $g_l^i g_k^j\neq 0$, then $q_{lk}=q_{ij}$, and if $g_k^i g_l^j\neq 0$, then $q_{lk}=q_{ij}^{-1}$.*

*Proof.* See [Naidu and Witherspoon 2011, Lemma 4.3]. □

Proposition 3.5 showed that every twisted quantum Drinfeld Hecke algebra arises from the quantum skew-symmetrization of a constant Hochschild 2-cocycle. The following theorem shows that the converse is also true. The proof of the following theorem involves the maps $\Theta_2^*$, $\mathcal{R}_2$, $\Psi_2^*$, and $d_3^*$ defined in Section 4.

**Theorem 5.3.** *Suppose that the action of $G$ on $V$ extends to an action, by automorphisms, on $S_{\boldsymbol{q}}(V)$ and on $\bigwedge_{\boldsymbol{q}}(V)$. Let $\alpha$ be a normalized 2-cocycle on $G$, let $\mu_1$ be a constant Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$, and let $\kappa: V\times V\to\mathbb{C}^\alpha G$ be the quantum skew-symmetrization of $\mu_1$. Then $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra.*

*Proof.* We show that the map $\kappa$ satisfies the conditions of Theorem 2.2. Let $\eta$ be a $G$-invariant element of

$$\operatorname{Hom}_{\mathbb{C}}\big(\textstyle\bigwedge_{\boldsymbol{q}}^2(V),\, S_{\boldsymbol{q}}(V)\#_\alpha G\big) \cong (S_{\boldsymbol{q}}(V)\#_\alpha G)\otimes \textstyle\bigwedge_{\boldsymbol{q}^{-1}}^2(V^*)$$

such that $[\Theta_2^*\mathcal{R}_2\Psi_2^*](\eta) = \mu_1$. Since $\mu_1$ is a *constant* Hochschild 2-cocycle, the image of $\eta$ as a map from $\bigwedge_{\boldsymbol{q}}^2(V)$ to $S_{\boldsymbol{q}}(V)\#_\alpha G$ is contained in $\mathbb{C}^\alpha G$, or, equivalently, $\eta$ belongs to $(\mathbb{C}^\alpha G)\otimes \bigwedge_{\boldsymbol{q}^{-1}}^2(V^*)$.

For all $1 \le k, l \le n$, we have $[\Psi_2^*(\eta)](v_k \otimes v_l - q_{kl} v_l \otimes v_k) = \eta(v_k \wedge v_l)$. This equality and the $G$-invariance of $\eta$ imply that $\kappa(v_i, v_j) = \eta(v_i \wedge v_j)$ for all $1 \le i, j \le n$. Indeed, we have

$$\begin{aligned}
\kappa(v_i, v_j) &= [\Theta_2^*\mathcal{R}_2\Psi_2^*(\eta)](v_i\otimes v_j - q_{ij}v_j\otimes v_i)\\
&= \frac{1}{|G|}\sum_{g\in G}\Theta_2^*({}^g(\Psi_2^*(\eta)))(v_i\otimes v_j - q_{ij}v_j\otimes v_i)\\
&= \frac{1}{|G|}\sum_{g\in G}{}^g((\Psi_2^*(\eta))^{g^{-1}}(v_i\otimes v_j - q_{ij}v_j\otimes v_i))\\
&= \frac{1}{|G|}\sum_{g\in G}{}^g\bigg((\Psi_2^*(\eta))\bigg(\sum_{k,l}(g^{-1})^i_k(g^{-1})^j_l\,(v_k\otimes v_l - q_{ij}v_l\otimes v_k)\bigg)\bigg)\\
&= \frac{1}{|G|}\sum_{g\in G}{}^g\bigg(\sum_{k,l}(g^{-1})^i_k(g^{-1})^j_l\,(\Psi_2^*(\eta))(v_k\otimes v_l - q_{ij}v_l\otimes v_k)\bigg)\\
&= \frac{1}{|G|}\sum_{g\in G}{}^g\bigg(\sum_{k,l}(g^{-1})^i_k(g^{-1})^j_l\,\eta(v_k\wedge v_l)\bigg)\\
&= \frac{1}{|G|}\sum_{g\in G}{}^g(\eta({}^{g^{-1}}(v_i\wedge v_j)))\\
&= \frac{1}{|G|}\sum_{g\in G}({}^g\eta)(v_i\wedge v_j)\\
&= \eta(v_i\wedge v_j).
\end{aligned}$$

Next, write

$$\eta = \sum_{g\in G}\sum_{1\le r<s\le n}\eta^g_{rs}t_g\otimes v_r^*\wedge v_s^* \in \mathbb{C}^\alpha G\otimes \textstyle\bigwedge_{\boldsymbol{q}^{-1}}^2(V^*) \subseteq (S_{\boldsymbol{q}}(V)\#_\alpha G)\otimes \textstyle\bigwedge_{\boldsymbol{q}^{-1}}^2(V^*).$$

The calculation above implies that $\kappa_g(v_i, v_j) = \eta^g_{ij}$ for all $i < j$ and all $g \in G$. Since $\eta$ is a Hochschild 2-cocycle, we have $d_3^*(\eta) = 0$. Using (4.8), we see that, for all $1 \le i < j < k \le n$, we must have

$$\sum_{g\in G}(\eta^g_{jk}v_it_g - \eta^g_{jk}q_{ij}q_{ik}{}^gv_it_g - \eta^g_{ik}q_{ij}v_jt_g + \eta^g_{ik}q_{jk}{}^gv_jt_g + \eta^g_{ij}q_{ik}q_{jk}v_kt_g - \eta^g_{ij}{}^gv_kt_g) = 0.$$

Equivalently,

$$-\eta^g_{jk}(q_{ij}q_{ik}\,{}^g v_i - v_i) - \eta^g_{ik}(q_{ij}v_j - q_{jk}\,{}^g v_j) - \eta^g_{ij}({}^g v_k - q_{ik}q_{jk}v_k) = 0$$

for all $1 \le i < j < k \le n$ and all $g \in G$.

Multiplying both sides by $q_{ji}q_{ki}q_{kj}$ yields

$$-q_{jk}\eta^g_{jk}({}^g v_i - q_{ji}q_{ki}v_i) - q_{ki}\eta^g_{ik}(q_{kj}v_j - q_{ji}\,{}^g v_j) - q_{ji}\eta^g_{ij}(q_{kj}q_{ki}\,{}^g v_k - v_k) = 0.$$

Now substituting $\kappa_g(v_k, v_j)$, $\kappa_g(v_k, v_i)$, and $\kappa_g(v_j, v_i)$ for $-q_{jk}\eta^g_{jk}$, $-q_{ki}\eta^g_{ik}$, and $-q_{ji}\eta^g_{ij}$, respectively, we obtain

$$\kappa_g(v_k, v_j)({}^g v_i - q_{ji}q_{ki}v_i) + \kappa_g(v_k, v_i)(q_{kj}v_j - q_{ji}\,{}^g v_j) + \kappa_g(v_j, v_i)(q_{kj}q_{ki}\,{}^g v_k - v_k) = 0,$$

which is condition (2) of Theorem 2.2.

Next, we show that $\kappa$ also satisfies condition (1) of Theorem 2.2. Since $\eta$ is $G$-invariant, we have $\eta({}^h v_i \wedge {}^h v_j) = {}^h(\eta(v_i \wedge v_j))$ for all $i, j$ and all $h \in G$. We have

$$\begin{aligned}
\eta({}^h v_i \wedge {}^h v_j) &= \sum_{k,l} h^i_k h^j_l \eta(v_k \wedge v_l) \\
&= \sum_{k<l} h^i_k h^j_l \eta(v_k \wedge v_l) - \sum_{k<l} q_{lk} h^i_l h^j_k \eta(v_k \wedge v_l) \\
&= \sum_{k<l,\ g\in G} (h^i_k h^j_l - q_{lk} h^i_l h^j_k)\eta^g_{kl} t_g,
\end{aligned}$$

and for all $i < j$, we have

$$\begin{aligned}
{}^h(\eta(v_i \wedge v_j)) &= {}^h\Big(\sum_{g\in G} \eta^g_{ij} t_g\Big) \\
&= \sum_{g\in G} \eta^g_{ij} t_h t_g (t_h)^{-1} \\
&= \sum_{g\in G} \frac{\alpha(h,g)\alpha(hg,h^{-1})}{\alpha(h^{-1},h)} \eta^g_{ij} t_{hgh^{-1}} \\
&= \sum_{g\in G} \frac{\alpha(h,g)}{\alpha(hgh^{-1},h)} \eta^g_{ij} t_{hgh^{-1}}.
\end{aligned}$$

Equating the coefficients of $t_{hgh^{-1}}$, we find that, for all $i < j$ and all $h, g \in G$, we have

$$\frac{\alpha(h,g)}{\alpha(hgh^{-1},h)} \eta^g_{ij} = \sum_{k<l} (h^i_k h^j_l - q_{lk} h^i_l h^j_k)\eta^{hgh^{-1}}_{kl}.$$

Substituting $\kappa_g(v_i, v_j)$ and $\kappa_{hgh^{-1}}(v_k, v_l)$ for $\eta_{ij}^g$ and $\eta_{kl}^{hgh^{-1}}$, respectively, and then multiplying both sides by $-q_{ji}$ yields

$$\frac{\alpha(h,g)}{\alpha(hgh^{-1},h)}\kappa_g(v_j, v_i) = \sum_{k<l}(q_{ji}q_{lk}h_l^i h_k^j - q_{ji}h_k^i h_l^j)\kappa_{hgh^{-1}}(v_k, v_l).$$

Substituting $-q_{kl}\kappa_{hgh^{-1}}(v_l, v_k)$ for $\kappa_{hgh^{-1}}(v_k, v_l)$, and then using Lemma 5.2, we obtain

$$\frac{\alpha(h,g)}{\alpha(hgh^{-1},h)}\kappa_g(v_j, v_i) = \sum_{k<l}\det{}_{ijkl}(h)\kappa_{hgh^{-1}}(v_l, v_k),$$

which is condition (1) of Theorem 2.2. □

The proof of the following theorem involves the map $\Phi_2^*$ defined in Section 4.

**Theorem 5.4.** *Let $\alpha$ be a normalized 2-cocycle on $G$. Suppose that the action of $G$ on $V$ extends to an action, by automorphisms, on $S_{\boldsymbol{q}}(V)$ and on $\bigwedge_{\boldsymbol{q}}(V)$. The assignment*

$$\mu_1 \mapsto \mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$$

*where $\kappa$ is the quantum skew-symmetrization of $\mu_1$ is a bijection from the space of equivalence classes of constant Hochschild 2-cocycles on $S_{\boldsymbol{q}}(V)\#_\alpha G$ to the space of twisted quantum Drinfeld Hecke algebras associated to the quadruple $(G, V, \boldsymbol{q}, \alpha)$.*

*Proof.* Proposition 3.5 showed that the assignment specified in the statement of the theorem is surjective. To see that the assignment is also injective, let $\mu_1$ and $\mu_1'$ be constant Hochschild 2-cocycles on $S_{\boldsymbol{q}}(V)\#_\alpha G$ such that their quantum skew-symmetrizations are equal. We have

$$\begin{aligned}[\Phi_2^*(\mu_1)](1\otimes 1\otimes v_i\otimes v_j) &= \mu_1(v_i, v_j) - q_{ij}\mu_1(v_j, v_i)\\ &= \mu_1'(v_i, v_j) - q_{ij}\mu_1'(v_j, v_i)\\ &= [\Phi_2^*(\mu_1')](1\otimes 1\otimes v_i\otimes v_j),\end{aligned}$$

so $\Phi_2^*(\mu_1) = \Phi_2^*(\mu_1')$, and it follows that $\mu_1$ and $\mu_1'$ are cohomologous. □

**Theorem 5.5.** *Let $\alpha$ be a normalized 2-cocycle on $G$. Suppose that the action of $G$ on $V$ extends to an action, by automorphisms, on $S_{\boldsymbol{q}}(V)$ and on $\bigwedge_{\boldsymbol{q}}(V)$. Each constant Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$ lifts to a deformation of $S_{\boldsymbol{q}}(V)\#_\alpha G$ over $\mathbb{C}[\hbar]$.*

*Proof.* Let $\mu_1'$ be a constant Hochschild 2-cocycle on $S_{\boldsymbol{q}}(V)\#_\alpha G$. By Theorem 5.3, $\mu_1'$ gives rise to a twisted quantum Drinfeld Hecke algebra $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$, where $\kappa$ is the quantum skew-symmetrization of $\mu_1'$. By Lemma 3.3, $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a twisted quantum Drinfeld Hecke algebra over $\mathbb{C}[\hbar]$. By Theorem 3.2, associated to $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha,\hbar}$ is a deformation $\mu = \mu_0 + \mu_1\hbar + \mu_2\hbar^2 + \cdots$ of $S_{\boldsymbol{q}}(V)\#_\alpha G$. The proof of Proposition 3.5 shows that $\kappa$ is the quantum skew-symmetrization of $\mu_1$, and it follows from Theorem 5.4 that $\mu_1'$ is cohomologous to $\mu_1$. □

## 6. Diagonal actions

As before, let $G$ be a finite group acting linearly on a vector space $V$ with basis $v_1, \dots, v_n$. Assume that $v_1, \dots, v_n$ are common eigenvectors for $G$. In this case, the Hochschild cohomology $\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#G)$ was computed in [Naidu et al. 2011]. Let $\alpha$ be a normalized 2-cocycle on $G$. In this section, we use results from [Naidu et al. 2011] to give an explicit description of the subspace of $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$ consisting of constant Hochschild 2-cocycles. As a consequence, we obtain a classification of twisted quantum Drinfeld Hecke algebras associated to the quadruple $(G, V, \boldsymbol{q}, \alpha)$.

Let $\lambda_{g,i} \in \mathbb{C}$ be the scalars for which ${}^g v_i = \lambda_{g,i} v_i$ for all $g \in G$ and all $i \in \{1, \dots, n\}$. For each $g \in G$, define

$$C_g := \left\{ \gamma \in (\mathbb{N} \cup \{-1\})^n \mid \text{ for each } i \in \{1, \dots, n\}, \prod_{s=1}^{n} q_{is}^{\gamma_s} = \lambda_{g,i} \text{ or } \gamma_i = -1 \right\}. \tag{6.1}$$

**Theorem 6.2** [Naidu et al. 2011]. *If $G$ acts diagonally on $V$, then*

$$\mathrm{HH}^\bullet(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#G)$$

*is isomorphic to the graded vector subspace of $(S_{\boldsymbol{q}}(V)\#_G) \otimes \bigwedge_{\boldsymbol{q}^{-1}}(V^*)$ given by*

$$\mathrm{HH}^m(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#G) \cong \bigoplus_{g \in G} \bigoplus_{\substack{\beta \in \{0,1\}^n \\ |\beta| = m}} \bigoplus_{\substack{\tau \in \mathbb{N}^n \\ \tau - \beta \in C_g}} \mathrm{span}_{\mathbb{C}}\{(v^\tau t_g) \otimes (v^*)^{\wedge \beta}\},$$

*for all $m \in \mathbb{N}$.*

**Corollary 6.3.** *The constant Hochschild 2-cocycles representing elements in the cohomology $\mathrm{HH}^2(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#G)$ form a vector space having as a basis the set of all*

$$t_g \otimes v_r^* \wedge v_s^*,$$

*where $r < s$ and $g \in G$ satisfy $q_{rr'} q_{sr'} = \lambda_{g,r'}$ for all $r' \notin \{r, s\}$.*

Note that the $S_{\boldsymbol{q}}(V)$-bimodule structure of $S_{\boldsymbol{q}}(V)\#_\alpha G$ does not depend on the 2-cocycle $\alpha$, and so $\mathrm{HH}^2(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#_\alpha G) = \mathrm{HH}^2(S_{\boldsymbol{q}}(V), S_{\boldsymbol{q}}(V)\#G)$.

Let $\mathcal{R}$ denote a complete set of representatives of conjugacy classes in $G$, let $C_G(a)$ denote the centralizer of $a$ in $G$, and let $[G/C_G(a)]$ denote a complete set of representatives of left cosets of $C_G(a)$ in $G$. In the following theorem, the notation $\delta_{i,j}$ is the Kronecker delta.

**Theorem 6.4.** *The constant Hochschild 2-cocycles representing elements in the cohomology $\mathrm{HH}^2(S_{\boldsymbol{q}}(V)\#_\alpha G)$ form a vector space having as a basis the set of all*

$$\sum_{g \in [G/C_G(a)]} \frac{\alpha(g, a)}{\alpha(gag^{-1}, g)} \lambda_{g,r}^{-1} \lambda_{g,s}^{-1} t_{gag^{-1}} \otimes v_r^* \wedge v_s^*,$$

*where* $r < s$ *and* $a \in \mathcal{R}$ *satisfy* $q_{rr'}q_{sr'} = \lambda_{a,r'}$ *for all* $r' \notin \{r, s\}$, *and* $\lambda_{h,r}\lambda_{h,s} = \alpha(h, a)/\alpha(a, h)$ *for all* $h \in C_G(a)$.

*Proof.* We show that the space of $G$-invariant elements of the vector space given in Corollary 6.3 is precisely the vector space stated in the theorem. The stated result then follows from Theorem 4.4.

First, we show that the scalar $(\alpha(g, a)/\alpha(gag^{-1}, g))\lambda_{g,r}^{-1}\lambda_{g,s}^{-1}$ is independent of choice of representative $g$ of a coset of $C_G(a)$ under the assumption that $\lambda_{h,r}\lambda_{h,s} = \alpha(h, a)/\alpha(a, h)$ for all $h \in C_G(a)$. Suppose that $gag^{-1} = g'ag'^{-1}$. Then $g' = gh$ for some $h \in C_G(a)$, and we have

$$\frac{\alpha(g', a)}{\alpha(g'ag'^{-1}, g')}\lambda_{g',r}^{-1}\lambda_{g',s}^{-1} = \frac{\alpha(gh, a)}{\alpha(gag^{-1}, gh)}\lambda_{g,r}^{-1}\lambda_{g,s}^{-1}\lambda_{h,r}^{-1}\lambda_{h,s}^{-1}.$$

Substituting $\lambda_{h,r}\lambda_{h,s} = \alpha(h, a)/\alpha(a, h)$ yields

$$\frac{\alpha(gh, a)\alpha(a, h)}{\alpha(gag^{-1}, gh)\alpha(h, a)}\lambda_{g,r}^{-1}\lambda_{g,s}^{-1}.$$

Applying the 2-cocycle condition of $\alpha$ to the triple $(g, h, a)$ gives

$$\alpha(gh, a)/\alpha(h, a) = \alpha(g, ha)/\alpha(g, h).$$

Making this substitution in the expression above yields

$$\frac{\alpha(g, ha)\alpha(a, h)}{\alpha(gag^{-1}, gh)\alpha(g, h)}\lambda_{g,r}^{-1}\lambda_{g,s}^{-1}.$$

Applying the 2-cocycle condition of $\alpha$ to the triple $(g, a, h)$ gives $\alpha(g, ha)\alpha(a, h) = \alpha(ga, h)\alpha(g, a)$. Making this substitution in the expression above yields

$$\frac{\alpha(ga, h)\alpha(g, a)}{\alpha(gag^{-1}, gh)\alpha(g, h)}\lambda_{g,r}^{-1}\lambda_{g,s}^{-1}.$$

Finally, applying the 2-cocycle condition of $\alpha$ to the triple $(gag^{-1}, g, h)$ gives

$$\frac{\alpha(ga, h)}{\alpha(gag^{-1}, gh)\alpha(g, h)} = \frac{1}{\alpha(gag^{-1}, g)}.$$

Making this substitution in the expression above yields

$$\frac{\alpha(g, a)}{\alpha(gag^{-1}, g)}\lambda_{g,r}^{-1}\lambda_{g,s}^{-1},$$

proving that the scalar above is independent of choice of representative $g$ of a coset of $C_G(a)$ under the assumption that $\lambda_{h,r}\lambda_{h,s} = \alpha(h, a)/\alpha(a, h)$ for all $h \in C_G(a)$. Thus, each of the alleged basis element is well defined, and is evidently $G$-invariant.

Conversely, let $\eta = \sum_{a\in G} \sum \eta^a_{rs} t_a \otimes v^*_r \wedge v^*_s$, where $\eta^a_{rs}$ are scalars and the second sum runs over all $r < s$ that satisfy $q_{rr'}q_{sr'} = \lambda_{a,r'}$ for all $r' \notin \{r, s\}$. We have

$$^g\eta = \sum_{a\in G} \eta^a_{rs} t_g t_a (t_g)^{-1} \otimes {}^g(v^*_r) \wedge {}^g(v^*_s) = \sum_{a\in G} \frac{\alpha(g, a)}{\alpha(gag^{-1}, g)} \lambda^{-1}_{g,r} \lambda^{-1}_{g,s} \eta^a_{rs} t_{gag^{-1}} \otimes v^*_r \wedge v^*_s.$$

Assume that $\eta$ is $G$-invariant. Then

$$\eta^{gag^{-1}}_{rs} = \frac{\alpha(g, a)}{\alpha(gag^{-1}, g)} \lambda^{-1}_{g,r} \lambda^{-1}_{g,s} \eta^a_{rs},$$

for all $g \in G$. Letting $g = h \in C_G(a)$ yields

$$\lambda_{h,r}\lambda_{h,s} = \frac{\alpha(h, a)}{\alpha(a, h)},$$

showing that $\eta$ is in the span of the alleged basis elements. The stated result now follows from Theorem 4.4. $\square$

The proof of the following theorem involves the maps $\Theta^*_2$, $\mathcal{R}_2$, and $\Psi^*_2$ defined in Section 4.

**Theorem 6.5.** *The maps $\kappa : V \times V \to \mathbb{C}^\alpha G$ for which $\mathcal{H}_{\boldsymbol{q},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra form a vector space with basis consisting of maps*

$$\begin{aligned} f_{r,s,a} &: V \times V \to \mathbb{C}^\alpha G, \\ &(v_i, v_j) \mapsto (\delta_{i,r}\delta_{j,s} - q_{sr}\delta_{i,s}\delta_{j,r}) \sum_{g\in[G/C_G(a)]} \frac{\alpha(g, a)}{\alpha(gag^{-1}, g)} \lambda^{-1}_{g,r} \lambda^{-1}_{g,s}\, t_{gag^{-1}}, \end{aligned}$$

*where $r < s$ and $a \in \mathcal{R}$ satisfy $q_{rr'}q_{sr'} = \lambda_{a,r'}$ for all $r' \notin \{r, s\}$ and $\lambda_{h,r}\lambda_{h,s} = \alpha(h, a)/\alpha(a, h)$ for all $h \in C_G(a)$.*

*Proof.* Let $\eta = \sum_{g\in[G/C_G(a)]} (\alpha(g, a)/\alpha(gag^{-1}, g)) \lambda^{-1}_{g,r} \lambda^{-1}_{g,s} t_{gag^{-1}} \otimes v^*_r \wedge v^*_s$, where $r < s$ and $a \in \mathcal{R}$ satisfy the conditions specified in Theorem 6.4. In the proof of Theorem 5.3 we saw that $[\Theta^*_2 \mathcal{R}_2 \Psi^*_2(\eta)](v_i \otimes v_j - q_{ij} v_j \otimes v_i) = \eta(v_i \wedge v_j)$, and the latter is equal to

$$(\delta_{i,r}\delta_{j,s} - q_{sr}\delta_{i,s}\delta_{j,r}) \sum_{g\in[G/C_G(a)]} \frac{\alpha(g, a)}{\alpha(gag^{-1}, g)} \lambda^{-1}_{g,r} \lambda^{-1}_{g,s}\, t_{gag^{-1}}.$$

The stated result now follows from Theorems 4.4 and 5.4. $\square$

## 7. Symmetric groups: natural representations

In this section, we classify twisted quantum Drinfeld Hecke algebras for the symmetric groups $S_n$, $n \geq 4$, acting naturally on a vector space of dimension $n$.

Consider the natural action of $S_n$ on a vector space $V$ with ordered basis $v_1, \dots, v_n$. Let $\boldsymbol{q} := (q_{ij})_{1\leq i,j\leq n}$ denote a tuple of nonzero scalars for which $q_{ii} = 1$ and $q_{ji} = q^{-1}_{ij}$ for all $i, j$. The action of $S_n$ extends to an action on the quantum symmetric algebra $S_{\boldsymbol{q}}(V)$ by automorphisms if and only if either $q_{ij} = 1$ for all $i$, $j$ or $q_{ij} = -1$ for all $i \neq j$. The tuple corresponding to the former will be denoted by $\mathbf{1}$, and the tuple corresponding to the latter by $-\mathbf{1}$. The action of $S_n$ on $V$ extends to an action on the quantum exterior algebra $\bigwedge_{-\mathbf{1}}$ by automorphisms. Note that the algebra $\bigwedge_{-\mathbf{1}}$ is commutative.

The Schur multiplier $\mathrm{H}^2(S_n, \mathbb{C}^\times)$ of the symmetric group $S_n$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ for all $n \geq 4$ [Schur 2001]. Let $\alpha$ be a 2-cocycle on $S_n$, and let $[\alpha]$ denote the image of $\alpha$ in $\mathrm{H}^2(S_n, \mathbb{C}^\times)$. A classification of twisted quantum Drinfeld Hecke algebras for $S_n$, acting naturally on a vector space of dimension $n$, is given in [Ram and Shepler 2003] for $[\alpha] = 1$ and $\boldsymbol{q} = \mathbf{1}$, in [Wambst 1993] for $[\alpha] \neq 1$ and $\boldsymbol{q} = \mathbf{1}$, and in [Naidu and Witherspoon 2011] for $[\alpha] = 1$ and $\boldsymbol{q} = -\mathbf{1}$. The goal of this section is to address the remaining case: $[\alpha] \neq 1$ and $\boldsymbol{q} = -\mathbf{1}$.

Next, we recall a Schur covering group of $S_n$, which we use to obtain a cohomologically nontrivial 2-cocycle on $S_n$. Let $T_n$ be the group with generators $t_1, \ldots, t_{n-1}, z$ and relations

$$\begin{aligned}
z^2 &= 1, \\
t_r^2 &= 1 && \text{for } 1 \leq r \leq n-1, \\
t_r t_s &= t_s t_r z && \text{for } |r-s| > 1 \text{ and } 1 \leq r, s \leq n-1, \\
t_r t_{r+1} t_r &= t_{r+1} t_r t_{r+1} && \text{for } 1 \leq r \leq n-2, \\
z t_r &= t_r z && \text{for } 1 \leq r \leq n-1.
\end{aligned}$$

The group $T_n$ is a central extension of $S_n$ by $\langle z \rangle$:

$$1 \to \langle z \rangle \to T_n \xrightarrow{p} S_n \to 1,$$

where the surjection $p$ sends $z$ to 1 and sends $t_r$ to the transposition $(r\, r+1)$. The group $T_n$ is a Schur covering group of $S_n$ [Schur 2001].

We define certain distinguished elements of $T_n$: For every $r, s \in \{1, \ldots, n\}$, $r \neq s$, denote by $[rs]$ the element of $T_n$ defined recursively as follows:

$$\begin{aligned}
[r\, r+1] &:= t_r, \\
[rs] &:= t_r [r+1\, s] t_r z && \text{if } r < s-1, \\
[rs] &:= [sr] z && \text{if } r > s.
\end{aligned}$$

Note that $p([rs]) = (rs)$.

Next, we define a section $u : S_n \to T_n$ of the surjection $p : T_n \to S_n$ by $u(\sigma) = u_\sigma$. If $\sigma \in S_n$ is the $k$-cycle $(a_1, \ldots, a_k)$, where $a_1, \ldots, a_k \in \{1, \ldots, n\}$ and $a_1$ is the smallest element of the set $\{a_1, \ldots, a_k\}$, define

$$u_\sigma := [a_1 a_k][a_1 a_{k-1}] \cdots [a_1 a_2].$$

If $\sigma \in S_n$ is the product $(a_1, \dots, a_k)(b_1, \dots, b_l) \cdots$ of disjoint cycles, where $a_1$ is the smallest element of the set $\{a_1, \dots, a_k\}$, $b_1$ is the smallest element of the set $\{b_1, \dots, b_l\}$, and so on, and $a_1 < b_1 < \cdots$, define

$$u_\sigma := u_{(a_1,\dots,a_k)} u_{(b_1,\dots,b_l)} \cdots .$$

It is evident that $u : S_n \to T_n$ is a section, that is, $pu = \mathrm{id}_{S_n}$.

Consider any irreducible representation of the group $T_n$. Since the element $z$ is central and has order two, it must necessarily act on this representation as multiplication by either 1 or $-1$. Assume the latter. In this case, we obtain a cohomologically nontrivial (normalized) 2-cocyle $\alpha : S_n \times S_n \to \mathbb{C}^\times$ defined by

$$\alpha(\sigma, \tau) := \begin{cases} 1 & \text{if } u_\sigma u_\tau u_{\sigma\tau}^{-1} = 1, \\ -1 & \text{if } u_\sigma u_\tau u_{\sigma\tau}^{-1} = z, \end{cases} \tag{7.1}$$

for all $\sigma, \tau \in S_n$.

Our goal is to classify twisted quantum Drinfeld Hecke algebras associated to the quadruple $(S_n, V, -\mathbf{1}, \alpha)$, where $V$ is the natural representation of $S_n$ and $-\mathbf{1}$ is the tuple defined earlier in this section. To this end, in what follows, we establish several lemmas that aid in accomplishing our goal.

Since the subgroup $\langle z \rangle$ of $T_n$ is central, there is an action of $S_n$ on $T_n$ induced by conjugation. If $\sigma$ belongs to $S_n$ and $\nu$ belongs to $T_n$, we denote by $\sigma \triangleright \nu$ the result of $\sigma$ acting upon $\nu$. We have $\sigma \triangleright \nu = \hat{\sigma} \nu (\hat{\sigma})^{-1}$, where $\hat{\sigma}$ is any element in the set $p^{-1}(\sigma)$.

For each $\sigma \in S_n$, let $\epsilon(\sigma)$ denote the signature of $\sigma$:

$$\epsilon(\sigma) = \begin{cases} 0 & \text{if } \sigma \text{ is an even permutation,} \\ 1 & \text{if } \sigma \text{ is an odd permutation.} \end{cases}$$

The following result from [Vendramin 2012] will be put to use shortly.

**Lemma 7.2.** *For all distinct $r, s \in \{1, \dots, n\}$ and all $\sigma \in S_n$, we have*

$$\sigma \triangleright [rs] = [\sigma(r)\sigma(s)] z^{\epsilon(\sigma)}.$$

For later use, we record two lemmas.

**Lemma 7.3.** *For all distinct $r, r', s, s' \in \{1, \dots, n\}$, we have*

$$[rs][sr']z = [rr'][rs] = [r's][rr']z.$$

*Proof.* We have $[rs]^{-1}[rr'][rs] = (rs)^{-1} \triangleright [rr'] = (rs) \triangleright [rr']$, and, by Lemma 7.2, the last expression equals $[sr']z$, proving the first equality. The second equality is proved similarly. ◻

**Lemma 7.4.** *For all distinct $r, r', s, s' \in \{1, \dots, n\}$, we have*

$$[rs][r's'] = [r's'][rs]z.$$

*Proof.* We have $[rs][r's'][rs]^{-1} = (rs) \triangleright [r's']$, and, by Lemma 7.2, the last expression equals $[r's']z$. $\square$

For all distinct $r, s, r', s' \in \{1, \ldots, n\}$, let $d(r, s, r', s')$ denote the number of inequalities

$$\min\{r, s\} > \min\{r', s'\}, \quad r > s, \quad r' > s'$$

that hold. For all distinct $r, s, r', s' \in \{1, \ldots, n\}$ and all $\sigma \in S_n$, define

$$d_\sigma(r, s, r', s') := d(\sigma(r), \sigma(s), \sigma(r'), \sigma(s')).$$

For later use, we record the following obvious result.

**Lemma 7.5.** *For all distinct $r, s, r', s' \in \{1, \ldots, n\}$, we have*

$$|d(r, s, r', s') - d(r, s, s', r')| = 1 = |d(r, s, r', s') - d(s, r, r', s')|.$$

We need the following lemma, which is a generalization of [Vendramin 2012, Lemma 3.7].

**Lemma 7.6.** *Let $\sigma$ be any element of $S_n$.*

(a) *For all $r, s \in \{1, \ldots, n\}$ with $r < s$, we have*

$$\sigma \triangleright u_{(rs)} = \begin{cases} u_{\sigma(rs)\sigma^{-1}} z^{\epsilon(\sigma)} & \text{if } \sigma(r) < \sigma(s), \\ u_{\sigma(rs)\sigma^{-1}} z^{\epsilon(\sigma)+1} & \text{if } \sigma(r) > \sigma(s). \end{cases}$$

(b) *For all distinct $r, s, r', s' \in \{1, \ldots, n\}$ with $r < s$, $r' < s'$, and $r < r'$, we have*

$$\sigma \triangleright u_{(rs)(r's')} = u_{\sigma(rs)(r's')\sigma^{-1}} z^{d_\sigma(r,s,r',s')}.$$

(c) *For all distinct $r, s, r' \in \{1, \ldots, n\}$ with $r < s$ and $r < r'$, we have*

$$\sigma \triangleright u_{(rsr')} = u_{\sigma(rsr')\sigma^{-1}}.$$

*Proof.* (a) By Lemma 7.2, $\sigma \triangleright u_{(rs)} = \sigma \triangleright [rs] = [\sigma(r)\sigma(s)]z^{\epsilon(\sigma)}$. If $\sigma(r) < \sigma(s)$, then

$$[\sigma(r)\sigma(s)]z^{\epsilon(\sigma)} = u_{(\sigma(r)\sigma(s))} z^{\epsilon(\sigma)} = u_{\sigma(rs)\sigma^{-1}} z^{\epsilon(\sigma)}.$$

If $\sigma(r) > \sigma(s)$, then

$$[\sigma(r)\sigma(s)]z^{\epsilon(\sigma)} = [\sigma(s)\sigma(r)]z^{\epsilon(\sigma)+1} = u_{(\sigma(s)\sigma(r))} z^{\epsilon(\sigma)+1} = u_{\sigma(rs)\sigma^{-1}} z^{\epsilon(\sigma)+1}.$$

(b) Again, by Lemma 7.2,

$$\begin{aligned} \sigma \triangleright u_{(rs)(r's')} = \sigma \triangleright [rs][r's'] &= (\sigma \triangleright [rs])(\sigma \triangleright [r's']) \\ &= [\sigma(r)\sigma(s)]z^{\epsilon(\sigma)}[\sigma(r')\sigma(s')]z^{\epsilon(\sigma)} \\ &= [\sigma(r)\sigma(s)][\sigma(r')\sigma(s')]. \end{aligned}$$

If $\min\{\sigma(r), \sigma(s)\} > \min\{\sigma(r'), \sigma(s')\}$, then, using Lemma 7.4, we rewrite the product above as $[\sigma(r')\sigma(s')][\sigma(r)\sigma(s)]z$. If $\sigma(r) > \sigma(s)$, we replace $[\sigma(r)\sigma(s)]$ by

$[\sigma(s)\sigma(r)]z$. Similarly, if $\sigma(r') > \sigma(s')$, we replace $[\sigma(r')\sigma(s')]$ by $[\sigma(s')\sigma(r')]z$. Since the element $z$ has order two, the stated result follows. For example, suppose that $d_\sigma(r, s, r', s') = 3$. Then $\sigma(r) > \sigma(s)$, $\sigma(r') > \sigma(s')$, and $\sigma(s) > \sigma(s')$, and in this case we write

$$\begin{aligned}[\sigma(r)\sigma(s)][\sigma(r')\sigma(s')] = [\sigma(r')\sigma(s')][\sigma(r)\sigma(s)]z &= [\sigma(s')\sigma(r')]z[\sigma(s)\sigma(r)]zz \\ &= [\sigma(s')\sigma(r')][\sigma(s)\sigma(r)]z \\ &= u_{(\sigma(s')\sigma(r'))(\sigma(s)\sigma(r))}z \\ &= u_{\sigma(rs)(r's')\sigma^{-1}}z.\end{aligned}$$

(c) Again, by Lemma 7.2,

$$\begin{aligned}\sigma \triangleright u_{(rsr')} = \sigma \triangleright [rr'][rs] = (\sigma \triangleright [rr'])(\sigma \triangleright [rs]) &= [\sigma(r)\sigma(r')]z^{\epsilon(\sigma)}[\sigma(r)\sigma(s)]z^{\epsilon(\sigma)} \\ &= [\sigma(r)\sigma(r')][\sigma(r)\sigma(s)].\end{aligned}$$

*Case* ($c_1$). $\sigma(r) < \sigma(r')$ and $\sigma(r) < \sigma(s)$. In this case,

$$[\sigma(r)\sigma(r')][\sigma(r)\sigma(s)] = u_{(\sigma(r)\sigma(s)\sigma(r'))} = u_{\sigma(rsr')\sigma^{-1}}.$$

*Case* ($c_2$). Either $\sigma(s) < \sigma(r) < \sigma(r')$ or $\sigma(s) < \sigma(r') < \sigma(r)$. Using the first equality of Lemma 7.3,

$$\begin{aligned}[\sigma(r)\sigma(r')][\sigma(r)\sigma(s)] = [\sigma(r)\sigma(s)][\sigma(s)\sigma(r')]z &= [\sigma(s)\sigma(r)]z[\sigma(s)\sigma(r')]z \\ &= u_{(\sigma(s)\sigma(r')\sigma(r))} \\ &= u_{\sigma(rsr')\sigma^{-1}}.\end{aligned}$$

*Case* ($c_3$). Either $\sigma(r') < \sigma(r) < \sigma(s)$ or $\sigma(r') < \sigma(s) < \sigma(r)$. Using the second equality of Lemma 7.3,

$$\begin{aligned}[\sigma(r)\sigma(r')][\sigma(r)\sigma(s)] &= [\sigma(r')\sigma(s)][\sigma(r)\sigma(r')]z \\ &= [\sigma(r')\sigma(s)][\sigma(r')\sigma(r)]zz \\ &= u_{(\sigma(r')\sigma(r)\sigma(s))} \\ &= u_{\sigma(rsr')\sigma^{-1}}.\end{aligned}$$

□

We now turn our attention to the Hochschild cohomology of $S_{-\mathbf{1}}(V)\#_\alpha S_n$.

**Theorem 7.7** [Naidu and Witherspoon 2011, Theorem 6.8]. *Assume that* $n \geq 4$. *The constant Hochschild* 2*-cocycles representing elements in* $\mathrm{HH}^2(S_{-\mathbf{1}}(V), S_{-\mathbf{1}}(V)\#S_n)$ *form a vector subspace of* $(S_{-\mathbf{1}}(V)\#G) \otimes \bigwedge_{-\mathbf{1}}(V^*)$ *having as a basis the set of all*

$$\begin{aligned}
\eta_1 &= t_1 \otimes v_r^* \wedge v_s^* && (r < s),\\
\eta_2 &= t_{(rs)} \otimes v_r^* \wedge v_s^* && (r < s),\\
\eta_3 &= t_{(rs)} \otimes (v_r^* \wedge v_{r'}^* + v_s^* \wedge v_{r'}^*) && (r < s),\\
\eta_4 &= t_{(rs)(r's')} \otimes (v_r^* \wedge v_{r'}^* + v_r^* \wedge v_{s'}^* + v_s^* \wedge v_{r'}^* + v_s^* \wedge v_{s'}^*) && (r < s,\, r' < s',\, r < r'),\\
\eta_5 &= t_{(rsr')} \otimes (v_r^* \wedge v_s^* + v_s^* \wedge v_{r'}^* + v_r^* \wedge v_{r'}^*) && (r < s,\, r < r').
\end{aligned}$$

Note that the $S_{-\mathbf{1}}(V)$-bimodule structure of $S_{-\mathbf{1}}(V)\#_\alpha G$ does not depend on the 2-cocycle $\alpha$, and so $\mathrm{HH}^2(S_{-\mathbf{1}}(V), S_{-\mathbf{1}}(V)\#_\alpha G) = \mathrm{HH}^2(S_{-\mathbf{1}}(V), S_{-\mathbf{1}}(V)\#G)$.

The lemma below involves the maps $\Theta_2^*$, $\mathcal{R}_2$, and $\Psi_2^*$ defined in Section 4. Recall that the image of an element $\sigma \in S_n$ in the twisted group algebra $\mathbb{C}^\alpha S_n$ is denoted by $t_\sigma$. Also, recall the definition of the 2-cocycle $\alpha$ given in (7.1).

**Lemma 7.8.** *For all $i \neq j$,*

$$[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_a)](v_i \otimes v_j) = \begin{cases} \dfrac{1}{n(n-1)} t_1 & \text{if } a = 1,\\ 0 & \text{if } a = 2,\\ 0 & \text{if } a = 3 \text{ and } n \geq 5,\\ 0 & \text{if } a = 4,\\ \dfrac{1}{n(n-1)(n-2)} \sum_{k \neq i,j} (2t_{(ijk)} + t_{(ikj)}) & \text{if } a = 5, \end{cases}$$

*Proof.* Using (4.7),

$$\begin{aligned}
[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_1)](v_i \otimes v_j) &= \frac{1}{n!} \sum_{\sigma \in S_n} {}^\sigma(\eta_1(\Psi_2(1 \otimes v_{\sigma^{-1}(i)} \otimes v_{\sigma^{-1}(j)} \otimes 1)))\\
&= \frac{1}{n!} \sum_{\substack{\sigma \in S_n\\ \sigma^{-1}(i) < \sigma^{-1}(j)}} {}^\sigma(\eta_1(1 \otimes 1 \otimes v_{\sigma^{-1}(i)} \otimes v_{\sigma^{-1}(j)}))\\
&= \frac{1}{n!} \sum_{\substack{\sigma \in S_n\\ \sigma(r)=i,\, \sigma(s)=j}} {}^\sigma(t_1)\\
&= \frac{1}{n(n-1)} t_1.
\end{aligned}$$

Similarly,

$$[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_2)](v_i \otimes v_j) = \frac{1}{n!} \sum_{\substack{\sigma \in S_n\\ \sigma(r)=i,\, \sigma(s)=j}} {}^\sigma(t_{(rs)}).$$

Applying the conjugation action in $\mathbb{C}^\alpha G$, we get

$$\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j}}\frac{\alpha(\sigma,(rs))}{\alpha(\sigma(rs)\sigma^{-1},\sigma)}t_{\sigma(rs)\sigma^{-1}}=\Bigg(\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j}}\frac{\alpha(\sigma,(rs))}{\alpha((ij),\sigma)}\Bigg)t_{(ij)}.$$

The scalar $\alpha(\sigma,(rs))/\alpha((ij),\sigma)$ in the summation above is determined by the following element of $T_n$:

$$u_\sigma u_{(rs)}u_{\sigma(rs)}^{-1}u_{(ij)\sigma}u_\sigma^{-1}u_{(ij)}^{-1}=u_\sigma u_{(rs)}u_\sigma^{-1}u_{\sigma(rs)\sigma^{-1}}^{-1}.$$

By Lemma 7.6(a),

$$u_\sigma u_{(rs)}u_\sigma^{-1}u_{\sigma(rs)\sigma^{-1}}^{-1}=\begin{cases}z^{\epsilon(\sigma)} & \text{if } i<j,\\ z^{\epsilon(\sigma)+1} & \text{if } i>j.\end{cases}$$

Since we assume $n$ is greater than or equal to 4, the set $\{\sigma\in S_n\mid\sigma(r)=i,\ \sigma(s)=j\}$ contains an equal number of odd and even permutations, and so

$$\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j}}\frac{\alpha(\sigma,(rs))}{\alpha((ij),\sigma)}=0,$$

proving that $[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_2)](v_i\otimes v_j)=0$.

Next, we consider the $a=3$ case. In addition to the stated assumption $r<s$, assume further that $r<r'$ and $s<r'$. The other cases can be handled similarly. We have

$$[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_3)](v_i\otimes v_j)=\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}}{}^{\sigma}(t_{(rs)})+\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}}{}^{\sigma}(t_{(rs)}).$$

Applying the conjugation action in $\mathbb{C}^\alpha G$, we get

$$\begin{aligned}\frac{1}{n!}&\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}}\frac{\alpha(\sigma,(rs))}{\alpha((i\sigma(s)),\sigma)}t_{(i\sigma(s))}+\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}}\frac{\alpha(\sigma,(rs))}{\alpha((\sigma(r)i),\sigma)}t_{(\sigma(r)i)}\\ &=\frac{1}{n!}\sum_{k\neq i,j}\Bigg(\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j,\,\sigma(s)=k}}\frac{\alpha(\sigma,(rs))}{\alpha((ik),\sigma)}+\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j,\,\sigma(r)=k}}\frac{\alpha(\sigma,(rs))}{\alpha((ik),\sigma)}\Bigg)t_{(ik)}.\end{aligned}$$

The scalar $\alpha(\sigma,(rs))/\alpha((ik),\sigma)$ in the first of the two inner summations above is determined by the element $u_\sigma u_{(rs)}u_\sigma^{-1}u_{\sigma(rs)\sigma^{-1}}^{-1}$ of $T_n$. Again, by Lemma 7.6(a),

$$u_\sigma u_{(rs)}u_\sigma^{-1}u_{\sigma(rs)\sigma^{-1}}^{-1}=\begin{cases}z^{\epsilon(\sigma)} & \text{if } i<k,\\ z^{\epsilon(\sigma)+1} & \text{if } i>k.\end{cases}$$

Since $n$ is assumed to be greater than or equal to 5, the set $\{\sigma\in S_n\mid\sigma(r)=i,\ \sigma(r')=j,\ \sigma(s)=k\}$ contains an equal number of odd and even permutations, and so

$$\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j,\,\sigma(s)=k}}\frac{\alpha(\sigma,(rs))}{\alpha((ik),\sigma)}=0.$$

Similarly,

$$\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j,\,\sigma(r)=k}} \frac{\alpha(\sigma,(rs))}{\alpha((ik),\sigma)}=0,$$

and it follows that $[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_3)](v_i\otimes v_j)=0$.

For the $a=4$ case, in addition to the stated assumptions $r<s$, $r'<s'$, $r<r'$, assume further that $r<s'$, $s<r'$, and $s<s'$. The other cases can be handled similarly. We have

$$\begin{aligned}
&[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_4)](v_i\otimes v_j)\\
&\qquad=\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}} {}^{\sigma}(t_{(rs)(r's')})+\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s')=j}} {}^{\sigma}(t_{(rs)(r's')})\\
&\qquad\quad+\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}} {}^{\sigma}(t_{(rs)(r's')})+\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(s')=j}} {}^{\sigma}(t_{(rs)(r's')}).
\end{aligned}$$

Applying the conjugation action in $\mathbb{C}^\alpha G$, we get

$$\begin{aligned}
&\frac{1}{n!}\Biggl(\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((i\sigma(s))(j\sigma(s')),\sigma)}t_{(i\sigma(s))(j\sigma(s'))}\\
&\qquad+\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s')=j}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((i\sigma(s))(\sigma(r')j),\sigma)}t_{(i\sigma(s))(\sigma(r')j)}\\
&\qquad+\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((\sigma(r)i)(j\sigma(s')),\sigma)}t_{(\sigma(r)i)(j\sigma(s'))}\\
&\qquad\qquad+\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(s')=j}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((\sigma(r)i)(\sigma(r')j),\sigma)}t_{(\sigma(r)i)(\sigma(r')j)}\Biggr)\\
&=\frac{1}{n!}\sum_{k,l\notin\{i,j\}}\Biggl(\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j\\ \sigma(s)=k,\,\sigma(s')=l}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((ik)(jl),\sigma)}+\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s')=j\\ \sigma(s)=k,\,\sigma(r')=l}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((ik)(jl),\sigma)}\\
&\qquad+\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j\\ \sigma(r)=k,\,\sigma(s')=l}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((ik)(jl),\sigma)}+\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(s')=j\\ \sigma(r)=k,\,\sigma(r')=l}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((ik)(jl),\sigma)}\Biggr)t_{(ik)(jl)}.
\end{aligned}$$

The scalar $\alpha(\sigma,(rs)(r's'))/\alpha((ik)(jl),\sigma)$ in the first of the four inner summations above is determined by the element $u_\sigma u_{(rs)(r's')}u_\sigma^{-1}u^{-1}_{\sigma(rs)(r's')\sigma^{-1}}$ of $T_n$. By Lemma 7.6(b),

$$u_\sigma u_{(rs)(r's)} u_\sigma^{-1} u^{-1}_{\sigma(rs)(r's')\sigma^{-1}} = z^{d_\sigma(r,s,r's')} = z^{d(i,k,j,l)}.$$

Thus,

$$\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j\\ \sigma(s)=k,\,\sigma(s')=l}} \frac{\alpha(\sigma,(rs)(r's'))}{\alpha((ik)(jl),\sigma)} = (n-4)!(-1)^{d(i,k,j,l)}.$$

Similarly, the second, third, and fourth summations are equal to $(n-1)!$ times $(-1)^{d(i,k,l,j)}$, $(-1)^{d(k,i,j,l)}$, and $(-1)^{d(k,i,l,j)}$, respectively. From Lemma 7.5 it follows that the sum of the four summations above is equal to zero, and so $[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_4)](v_i\otimes v_j)=0$.

Finally, for the $a=5$ case, in addition to the stated assumptions $r<s$, $r<r'$, assume further that $s<r'$. Again, the other case can be handled similarly. We have

$$\begin{aligned}
&[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_5)](v_i\otimes v_j)\\
&= \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j}} {}^\sigma(t_{(rsr')}) + \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}} {}^\sigma(t_{(rsr')}) + \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}} {}^\sigma(t_{(rsr')})\\
&= \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j}} \frac{\alpha(\sigma,(rsr'))}{\alpha((ij\sigma(r')),\sigma)} t_{(ij\sigma(r'))} + \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j}} \frac{\alpha(\sigma,(rsr'))}{\alpha((\sigma(r)ij),\sigma)} t_{(\sigma(r)ij)}\\
&\qquad + \frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j}} \frac{\alpha(\sigma,(rsr'))}{\alpha((i\sigma(s)j),\sigma)} t_{(i\sigma(s)j)}\\
&= \frac{1}{n!}\sum_{k,l\notin\{i,j\}}\Bigg[\Bigg(\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j,\,\sigma(r')=k}} \frac{\alpha(\sigma,(rsr'))}{\alpha((ijk),\sigma)} + \sum_{\substack{\sigma\in S_n\\ \sigma(s)=i,\,\sigma(r')=j,\,\sigma(r)=k}} \frac{\alpha(\sigma,(rsr'))}{\alpha((ijk),\sigma)}\Bigg) t_{(ijk)}\\
&\qquad + \sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(r')=j,\,\sigma(s)=k}} \frac{\alpha(\sigma,(rsr'))}{\alpha((ikj),\sigma)} t_{(ikj)}\Bigg].
\end{aligned}$$

The scalar $\alpha(\sigma,(rsr'))/\alpha((ijk),\sigma)$ in the first of the three inner summations above is determined by the element $u_\sigma u_{(rsr')} u_\sigma^{-1} u^{-1}_{\sigma(rsr')\sigma^{-1}}$ of $T_n$. By Lemma 7.6(c), $u_\sigma u_{(rsr')} u_\sigma^{-1} u^{-1}_{\sigma(rsr')\sigma^{-1}} = 1$. Thus,

$$\sum_{\substack{\sigma\in S_n\\ \sigma(r)=i,\,\sigma(s)=j\sigma(r')=k}} \frac{\alpha(\sigma,(rsr')}{\alpha((ijk),\sigma)} = (n-3)!\,.$$

Similarly, the second and third summations are also equal to $(n-3)!$. It follows that

$$[(\Theta_2^*\mathcal{R}_2\Psi_2^*)(\eta_5)](v_i\otimes v_j) = \frac{1}{n(n-1)(n-2)}\sum_{k\neq i,j}(2t_{(ijk)}+t_{(ikj)}). \qquad \square$$

Combining Theorems 7.7, 4.4, 5.4, and Lemma 7.8 establishes the following.

**Theorem 7.9.** *Assume that $n \geq 5$. The maps $\kappa : V \times V \to \mathbb{C}^\alpha S_n$ for which $\mathcal{H}_{-\mathbf{1},\kappa,\alpha}$ is a twisted quantum Drinfeld Hecke algebra form a two-dimensional vector space with basis consisting of bilinear maps $\kappa_1 : V \times V \to \mathbb{C}^\alpha S_n$ and $\kappa_2 : V \times V \to \mathbb{C}^\alpha S_n$ determined by*

$$\kappa_1(v_i, v_j) = t_1 \quad \text{and} \quad \kappa_2(v_i, v_j) = \sum_{k \neq i,j} (t_{(ijk)} + t_{(ikj)}),$$

*for all $i \neq j$.*

## References

[Adem and Ruan 2003] A. Adem and Y. Ruan, "Twisted orbifold $K$-theory", *Comm. Math. Phys.* **237**:3 (2003), 533–556. MR 2004e:19004 Zbl 1051.57022

[Bazlov and Berenstein 2009] Y. Bazlov and A. Berenstein, "Noncommutative Dunkl operators and braided Cherednik algebras", *Selecta Math.* (*N.S.*) **14**:3-4 (2009), 325–372. MR 2010k:16044 Zbl 1220.16027

[Bergman 1978] G. M. Bergman, "The diamond lemma for ring theory", *Adv. in Math.* **29**:2 (1978), 178–218. MR 81b:16001 Zbl 0326.16019

[Căldăraru et al. 2004] A. Căldăraru, A. Giaquinto, and S. Witherspoon, "Algebraic deformations arising from orbifolds with discrete torsion", *J. Pure Appl. Algebra* **187**:1-3 (2004), 51–70. MR 2005c:16013 Zbl 1055.16010

[Chmutova 2005] T. Chmutova, "Twisted symplectic reflection algebras", preprint, 2005. arXiv math/0505653

[Drinfeld 1986] V. G. Drinfeld, "Degenerate affine Hecke algebras and Yangians", *Funktsional. Anal. i Prilozhen.* **20**:1 (1986), 69–70. In Russian; translated in *Funct. Anal. Appl.* **20** (1986), 58–60. MR 87m:22044 Zbl 0599.20049

[Etingof and Ginzburg 2002] P. Etingof and V. Ginzburg, "Symplectic reflection algebras, Calogero–Moser space, and deformed Harish-Chandra homomorphism", *Invent. Math.* **147**:2 (2002), 243–348. MR 2003b:16021 Zbl 1061.16032

[Gordon 2003] I. Gordon, "On the quotient ring by diagonal invariants", *Invent. Math.* **153**:3 (2003), 503–518. MR 2004f:20075 Zbl 1039.20019

[Levandovskyy and Shepler 2011] V. Levandovskyy and A. Shepler, "Quantum Drinfeld Hecke algebras", preprint, 2011. arXiv 1111.4975

[Lusztig 1989] G. Lusztig, "Affine Hecke algebras and their graded version", *J. Amer. Math. Soc.* **2**:3 (1989), 599–635. MR 90e:16049 Zbl 0715.22020

[Naidu and Witherspoon 2011] D. Naidu and S. Witherspoon, "Hochschild cohomology and quantum Drinfeld Hecke algebras", preprint, 2011. arXiv 1111.5243

[Naidu et al. 2011] D. Naidu, P. Shroff, and S. Witherspoon, "Hochschild cohomology of group extensions of quantum symmetric algebras", *Proc. Amer. Math. Soc.* **139**:5 (2011), 1553–1567. MR 2012b:16026 Zbl 1259.16011

[Ram and Shepler 2003] A. Ram and A. V. Shepler, "Classification of graded Hecke algebras for complex reflection groups", *Comment. Math. Helv.* **78**:2 (2003), 308–334. MR 2004d:20007 Zbl 1063.20005

[Schur 2001] J. Schur, "On the representation of the symmetric and alternating groups by fractional linear substitutions", *Internat. J. Theoret. Phys.* **40**:1 (2001), 413–458. MR 2003a:20016 Zbl 0969.20002

[Shepler and Witherspoon 2012] A. V. Shepler and S. Witherspoon, "Group actions on algebras and the graded Lie structure of Hochschild cohomology", *J. Algebra* **351** (2012), 350–381. MR 2862214 Zbl 06046981

[Ştefan 1995] D. Ştefan, "Hochschild cohomology on Hopf Galois extensions", *J. Pure Appl. Algebra* **103**:2 (1995), 221–233. MR 96h:16013 Zbl 0838.16008

[Vafa and Witten 1995] C. Vafa and E. Witten, "On orbifolds with discrete torsion", *J. Geom. Phys.* **15**:3 (1995), 189–214. MR 95m:81190 Zbl 0816.53053

[Vendramin 2012] L. Vendramin, "Nichols algebras associated to the transpositions of the symmetric group are twist-equivalent", *Proc. Amer. Math. Soc.* **140**:11 (2012), 3715–3723. MR 2944712

[Wambst 1993] M. Wambst, "Complexes de Koszul quantiques", *Ann. Inst. Fourier* (*Grenoble*) **43**:4 (1993), 1089–1156. MR 95a:17023 Zbl 0810.16010

[Witherspoon 2007] S. Witherspoon, "Twisted graded Hecke algebras", *J. Algebra* **317**:1 (2007), 30–42. MR 2009a:20009 Zbl 1139.20005

Received October 24, 2012.

DEEPAK NAIDU
DEPARTMENT OF MATHEMATICAL SCIENCES
NORTHERN ILLINOIS UNIVERSITY
DEKALB, ILLINOIS 60115
UNITED STATES
dnaidu@math.niu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.205

# $L^p$ HARMONIC 1-FORMS AND FIRST EIGENVALUE OF A STABLE MINIMAL HYPERSURFACE

KEOMKYO SEO

**We estimate the bottom of the spectrum of the Laplace operator on a stable minimal hypersurface in a negatively curved manifold. We also derive various vanishing theorems for $L^p$ harmonic 1-forms on minimal hypersurfaces in terms of the bottom of the spectrum of the Laplace operator. As consequences, the corresponding Liouville type theorems for harmonic functions with finite $L^p$ energy on minimal hypersurfaces in a Riemannian manifold are obtained.**

## 1. Introduction

Hodge theory plays an important role in the topology of compact Riemannian manifolds. Unfortunately, the Hodge theory does not work anymore in noncompact manifolds. However, the $L^2$-Hodge theory works well in noncompact cases [Anderson 1988; Dodziuk 1982]. In this direction, there are various results for $L^2$ harmonic 1-forms on stable minimal hypersurfaces. Recall that a minimal hypersurface in a Riemannian manifold is called *stable* if the second variation of its volume is always nonnegative for any normal variation with compact support. More precisely, an $n$-dimensional minimal hypersurface $M$ in a Riemannian manifold $N$ is called *stable* if it holds that, for any compactly supported Lipschitz function $f$ on $M$,

$$\int_M |\nabla f|^2 - (|A|^2 + \overline{\mathrm{Ric}}(\nu, \nu)) f^2 \, dv \geq 0,$$

where $\nu$ is the unit normal vector of $M$, $\overline{\mathrm{Ric}}(\nu, \nu)$ denotes the Ricci curvature of $N$ in the $\nu$ direction, $|A|^2$ is the square length of the second fundamental form $A$, and $dv$ is the volume form for the induced metric on $M$.

Using the nonexistence of $L^2$ harmonic 1-forms, Palmer [1991] proved that if there exists a codimension-one cycle on a complete minimal hypersurface $M$ in Euclidean space, which does not separate $M$, $M$ is unstable. Using Bochner's

This research was supported by the Sookmyung Women's University Research Grants (1-1303-0116).
*MSC2010:* primary 53C42; secondary 58C40.
*Keywords:* minimal hypersurface, stability, first eigenvalue, $L^p$ harmonic 1-form, Liouville type theorem.

vanishing technique, Miyaoka [1993] showed that a complete noncompact stable minimal hypersurface in a nonnegatively curved manifold has no nontrivial $L^2$ harmonic 1-forms. Pigola, Rigoli, and Setti [Pigola et al. 2005] gave general Liouville type results and the corresponding vanishing theorems on the $L^2$ cohomology of stable minimal hypersurfaces. Refer to [Carron 2002; Pigola et al. 2008] for a survey in this area. While the $L^2$ theory is quite well understood, in the case $p \neq 2$, the $L^p$ theory is less developed. See [Scott 1995] for general $L^p$ theory of differential forms on a manifold.

The purpose of this paper is twofold. Firstly, we estimate the smallest spectral value of the Laplace operator on a complete noncompact stable minimal hypersurface in a Riemannian manifold under the assumption on $L^p$ norm of the second fundamental form. Secondly, we obtain various vanishing theorems for $L^p$ harmonic 1-forms on minimal hypersurfaces.

Let $M$ be a complete noncompact Riemannian manifold and let $\Omega$ be a compact domain in $M$. Let $\lambda_1(\Omega) > 0$ denote the first eigenvalue of the Dirichlet boundary value problem

$$\begin{cases} \Delta f + \lambda f = 0 & \text{in } \Omega, \\ f = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\Delta$ denotes the Laplace operator on $M$. Then the first eigenvalue $\lambda_1(M)$ is defined by

$$\lambda_1(M) = \inf_{\Omega} \lambda_1(\Omega),$$

where the infimum is taken over all compact domains in $M$. Cheung and Leung [2001] gave the first eigenvalue estimate for an $n$-dimensional complete noncompact submanifold $M$ with the norm of its mean curvature vector bounded in the hyperbolic space. In particular, they proved that if $M$ is minimal, the first eigenvalue $\lambda_1(M)$ satisfies

$$\tfrac{1}{4}(n-1)^2 \leq \lambda_1(M).$$

Note that this inequality is sharp because equality holds if $M$ is totally geodesic [McKean 1970]. This result was extended to an $n$-dimensional complete noncompact submanifold with the norm of its mean curvature vector bounded in a complete simply connected Riemannian manifold with sectional curvature bounded above by a negative constant. More precisely, we have the following theorem.

**Theorem** [Bessa and Montenegro 2003; Seo 2012]. *Let $N$ be an $n$-dimensional complete simply connected Riemannian manifold with sectional curvature $K_N$ satisfying $K_N \leq -a^2 < 0$ for a positive constant $a > 0$. Let $M$ be a an $m$-dimensional complete noncompact submanifold with bounded mean curvature vector $H$ in $N$ satisfying $|H| \leq b < (m-1)a$. Then*

$$\tfrac{1}{4}[(m-1)a - b]^2 \leq \lambda_1(M). \tag{1}$$

On the other hand, Candel [2007] obtained an upper bound for the bottom of the spectrum of a complete simply connected stable minimal surface in 3-dimensional hyperbolic space. With finite $L^2$ norm of the second fundamental form, one may estimate an upper bound for the bottom of the spectrum of a stable minimal hypersurface in a Riemannian manifold with pinched negative sectional curvature [Dung and Seo 2012; Seo 2011]. In Section 2, we estimate the bottom of the spectrum of the Laplace operator on stable minimal hypersurfaces under the assumption on the $L^p$ norm of the second fundamental form. Indeed, we prove the following.

**Theorem.** *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature satisfying $K_1 \le K_N \le K_2$, where $K_1, K_2$ are constants and $K_1 \le K_2 < 0$. Let $M$ be a complete stable non-totally geodesic minimal hypersurface in $N$. Assume that, for $1-\sqrt{2/n} < p < 1+\sqrt{2/n}$,*

$$\lim_{R\to\infty} R^{-2}\int_{B(R)} |A|^{2p} = 0,$$

*where $B(R)$ is a geodesic ball of radius $R$ on $M$. If $|\nabla K|^2 = \sum_{i,j,k,l,m} K^2_{ijkl;m} \le K_3^2|A|^2$ for some constant $K_3 \ge 0$, we have*

$$-K_2\frac{(n-1)^2}{4} \le \lambda_1(M) \le \frac{np^2(2K_3 - n(K_1+K_2))}{2-n(p-1)^2}.$$

The author [2010] proved that if $M$ is an $n$-dimensional complete stable minimal hypersurface in hyperbolic space with $\lambda_1(M) > (2n-1)(n-1)$, there is no nontrivial $L^2$ harmonic 1-form on $M$. This result was generalized [Dung and Seo 2012] to a complete stable minimal hypersurface in a Riemannian manifold with sectional curvature bounded below by a nonpositive constant. In Section 3, we prove an extended result for $L^p$ harmonic 1-forms on a complete noncompact stable minimal hypersurface as follows.

**Theorem.** *Let $N$ be an $(n+1)$-dimensional complete Riemannian manifold with sectional curvature satisfying that $K \le K_N$ where $K \le 0$ is a constant. Let $M$ be a complete noncompact stable minimal hypersurface in $N$. Assume that, for $0 < p < n/(n-1)+\sqrt{2n}$,*

$$\lambda_1(M) > \frac{-2n(n-1)^2p^2K}{2n-[(n-1)p-n]^2}.$$

*Then there is no nontrivial $L^{2p}$ harmonic 1-form on $M$.*

Yau [1976] proved that there are no nonconstant $L^p$ harmonic functions on a complete Riemannian manifold for $1 < p < \infty$. Li and Schoen [1984] proved that Yau's result is still true for $L^p$ harmonic functions on a complete manifold of nonnegative Ricci curvature when $0 < p < \infty$. In the case of harmonic forms, Greene and Wu [1974; 1981] announced nonexistence of nontrivial $L^p$ harmonic forms ($1 \le p < \infty$) on complete Riemannian and Kählerian manifolds of nonnegative curvature. See also [Colding and Minicozzi 1996; 1997; 1998; Li and Tam 1987; 1992] for Liouville type theorems for harmonic functions on a complete Riemannian manifold. The Liouville property holds also for harmonic functions on minimal hypersurfaces in a Riemannian manifold. For instance, Schoen and Yau proved the Liouville type theorem on minimal hypersurfaces as follows.

**Theorem** [Schoen and Yau 1976]. *Let $M$ be a complete noncompact stable minimal hypersurface in a Riemannian manifold with nonnegative sectional curvature. If $f$ is a harmonic function on $M$ with finite $L^2$ energy, $f$ is constant.*

Recall that a function $f$ on a Riemannian manifold $M$ has *finite $L^p$ energy* if $|\nabla f| \in L^p(M)$. As an application of our theorem, we immediately obtain the following, which is a generalization of Schoen and Yau's result (see Corollary 3.10).

**Theorem.** *Let $M$ be a complete noncompact stable minimal hypersurface in a Riemannian manifold with nonnegative sectional curvature with $\lambda_1(M) > 0$. Then there is no nontrivial harmonic function on $M$ with finite $L^p$ energy for $0 < p < n/(n-1) + \sqrt{2n}$.*

For $n \ge 3$, it is well known [Cao et al. 1997] that an $n$-dimensional complete stable minimal hypersurface $M$ in Euclidean space cannot have more than one end. This topological result was generalized to minimal hypersurfaces with finite index in Euclidean space and stable minimal hypersurfaces in a nonnegatively curved manifold by Li and Wang [2002; 2004]. If we assume that $M$ has sufficiently small total scalar curvature instead of assuming that $M$ is stable, we can also have the same conclusion [Ni 2001; Seo 2008]. See also [Pigola and Veronelli 2012] for more general results related with $L^p$ norm of the second fundamental form. In the same spirit, Yun [2002] proved that if $M \subset \mathbb{R}^{n+1}$ is a complete minimal hypersurface with sufficiently small total scalar curvature, there is no nontrivial $L^2$ harmonic 1-form on $M$. Yun's result was generalized [Dung and Seo 2012] to a complete noncompact stable minimal hypersurface in a complete Riemannian manifold with sectional curvature bounded below by a nonpositive constant. The corresponding vanishing theorems for $L^p$ harmonic 1-forms are obtained in Section 4.

One crucial step in the proofs of our theorems is to obtain an inequality of Simons' type for $|\phi|^p$ rather than $|\phi|$, where $\phi$ is a geometric quantity which we want to analyze. This kind of inequalities has been used in [Deng 2008; Fu 2012; Shen and Zhu 2005]. Equipped with this Simons' type inequality, we extend the original Bochner technique to our cases.

## 2. An estimate for the bottom of the spectrum of the Laplace operator

Let $M$ be an $n$-dimensional manifold immersed in an $(n+1)$-dimensional Riemannian manifold $N$. We choose a local vector field of orthonormal frames $e_1, \dots, e_{n+1}$ in $N$ such that the vectors $e_1, \dots, e_n$ are tangent to $M$ and the vector $e_{n+1}$ is normal to $M$. With respect to this frame field of $N$, let $K_{ijkl}$ be a curvature tensor of $N$. We denote by $K_{ijkl;m}$ the covariant derivative of $K_{ijkl}$. In this section, we follow the notation of [Schoen et al. 1975].

**Theorem 2.1.** *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature satisfying $K_1 \le K_N \le K_2$, where $K_1$, $K_2$ are constants and $K_1 \le K_2 < 0$. Let $M$ be a complete stable non-totally geodesic minimal hypersurface in $N$. Assume that, for $1-\sqrt{2/n} < p < 1+\sqrt{2/n}$,*

$$\lim_{R\to\infty} R^{-2} \int_{B(R)} |A|^{2p} = 0,$$

*where $B(R)$ is a geodesic ball of radius $R$ on $M$. If $|\nabla K|^2 = \sum_{i,j,k,l,m} K^2_{ijkl;m} \le K_3^2 |A|^2$ for some constant $K_3 \ge 0$, we have*

$$-K_2 \frac{(n-1)^2}{4} \le \lambda_1(M) \le \frac{np^2(2K_3 - n(K_1+K_2))}{2-n(p-1)^2}.$$

*Proof.* As mentioned in the introduction, one sees that the lower bound of $\lambda_1(M)$ is given as $-K_2(n-1)^2/4$ from inequality (1) [Bessa and Montenegro 2003; Seo 2012]. Namely, the first eigenvalue of an $n$-dimensional minimal hypersurface in a complete simply connected Riemannian manifold with sectional curvature bounded above by a negative constant $K_2$ is bounded below by $-K_2(n-1)^2/4$. Therefore, in the rest of the proof, we shall find the upper bound of the first eigenvalue $\lambda_1(M)$.

By [Schoen et al. 1975, (1.22), (1.27)], we have

$$|A|\Delta|A| + 2K_3|A|^2 - n(2K_2 - K_1)|A|^2 + |A|^4 \ge \sum h^2_{ijk} - \big|\nabla|A|\big|^2$$

at all points where $|A| \ne 0$. Because $K_2 - K_1 \ge 0$, this inequality implies

$$\begin{aligned} |A|\Delta|A| + 2K_3|A|^2 - nK_2|A|^2 + |A|^4 &\ge \sum h^2_{ijk} - \big|\nabla|A|\big|^2 \\ &= |\nabla A|^2 - \big|\nabla|A|\big|^2. \end{aligned}$$

Applying the Kato-type inequality

$$|\nabla A|^2 - \big|\nabla|A|\big|^2 \ge \frac{2}{n}\big|\nabla|A|\big|^2,$$

due to Y. L. Xin [2005], we get

$$|A|\Delta|A| + (2K_3 - nK_2)|A|^2 + |A|^4 \ge \frac{2}{n}\big|\nabla|A|\big|^2. \tag{2}$$

For a positive number $p > 0$, we have

$$\begin{aligned}|A|^p\Delta|A|^p &= |A|^p\mathrm{div}(\nabla|A|^p)\\ &= |A|^p\mathrm{div}(p\,|A|^{p-1}\nabla|A|)\\ &= p(p-1)|A|^{2p-2}\big|\nabla|A|\big|^2 + p|A|^{2p-1}\Delta|A|\\ &= \frac{p-1}{p}\big|\nabla|A|^p\big|^2 + p|A|^{2p-2}|A|\Delta|A|.\end{aligned}$$

It follows from inequality (2) that

$$\begin{aligned}&|A|^p\Delta|A|^p\\ &\quad\geq \frac{p-1}{p}\big|\nabla|A|^p\big|^2 + \frac{2p}{n}|A|^{2p-2}\big|\nabla|A|\big|^2 - p|A|^{2p+2} - p(2K_3 - nK_2)|A|^{2p}\\ &\quad= \frac{p-1}{p}\big|\nabla|A|^p\big|^2 + \frac{2}{np}\big|\nabla|A|^p\big|^2 - p|A|^{2p+2} - p(2K_3 - nK_2)|A|^{2p}.\end{aligned}$$

Thus

$$|A|^p\Delta|A|^p + p(2K_3 - nK_2)|A|^{2p} + p|A|^{2p+2} \geq \left(1 - \frac{n-2}{np}\right)\big|\nabla|A|^p\big|^2.$$

Choose a Lipschitz function $f$ with compact support in a geodesic ball $B(R)$ of radius $R$ centered at a point $x \in M$. Multiplying both sides by $f^2$ and integrating over $B(R)$, we obtain

$$\begin{aligned}\int_{B(R)} f^2|A|^p\Delta|A|^p + p(2K_3 - nK_2)\int_{B(R)} f^2|A|^{2p} + p\int_{B(R)} f^2|A|^{2p+2}&\\ \geq \left(1 - \frac{n-2}{np}\right)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2&.\end{aligned}$$

The divergence theorem yields

$$\begin{aligned}&\int_{B(R)} f^2|A|^p\Delta|A|^p\\ &\quad= \int_{B(R)} \mathrm{div}(f^2|A|^p\nabla|A|^p) - \int_{B(R)} f^2\big|\nabla|A|^p\big|^2 - 2\int_{B(R)} f|A|^p\langle\nabla f, \nabla|A|^p\rangle\\ &\quad= -\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 - 2\int_{B(R)} f|A|^p\langle\nabla f, \nabla|A|^p\rangle.\end{aligned}$$

Therefore

$$\left(1-\frac{n-2}{np}\right)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 \le p(2K_3-nK_2)\int_{B(R)} f^2|A|^{2p}+p\int_{B(R)} f^2|A|^{2p+2} - \int_{B(R)} f^2\big|\nabla|A|^p\big|^2 - 2\int_{B(R)} f|A|^p\langle\nabla f,\nabla|A|^p\rangle. \tag{3}$$

The stability of $M$ implies that

$$\int_M |\nabla f|^2-(|A|^2+\overline{\mathrm{Ric}}(e_{n+1}))f^2\ge 0 \tag{4}$$

for any compactly supported Lipschitz function $f$ on $M$. From our assumption on the sectional curvature of $N$, we see that

$$nK_1\le \overline{\mathrm{Ric}}(e_{n+1})=R_{n+1,1,n+1,1}+\cdots+R_{n+1,n,n+1,n}\le nK_2.$$

Hence the stability inequality (4) gives

$$\int_M |\nabla f|^2-(|A|^2+nK_1)f^2\ge 0 \tag{5}$$

for any compactly supported Lipschitz function $f$ on $M$. Choose a Lipschitz function $f$ with compact support in a geodesic ball $B(R)\subset M$, as before. Replacing $f$ by $|A|^p f$ in inequality (5), we have

$$\int_M \big|\nabla(|A|^p f)\big|^2-(|A|^{2p+2}f^2+nK_1|A|^{2p}f^2)\ge 0.$$

Thus

$$\int_{B(R)}\big|\nabla|A|^p\big|^2 f^2+\int_{B(R)}|\nabla f|^2|A|^{2p}+2\int_{B(R)} f|A|^p\langle\nabla f,\nabla|A|^p\rangle \ge \int_{B(R)}|A|^{2p+2}f^2+nK_1\int_{B(R)}|A|^{2p}f^2. \tag{6}$$

Combining the inequalities (3) and (6), we get

$$\left(1-\frac{n-2}{np}\right)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 \le p(2K_3-nK_1-nK_2)\int_{B(R)} f^2|A|^{2p}+(p-1)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 + p\int_{B(R)}|\nabla f|^2|A|^{2p}+2(p-1)\int_{B(R)} f|A|^p\langle\nabla f,\nabla|A|^p\rangle. \tag{7}$$

On the other hand, from the definition of $\lambda_1(M)$ and the domain monotonicity of eigenvalues, it follows that

$$\lambda_1(M) \le \lambda_1(B(R)) \le \frac{\int_{B(R)} |\nabla f|^2}{\int_{B(R)} f^2} \tag{8}$$

for any compactly supported nonconstant Lipschitz function $f$ on $M$. Substituting $|A|^p f$ for $f$ in inequality (8), we see that

$$\begin{aligned} \lambda_1(M) \int_{B(R)} |A|^{2p} f^2 &\le \int_{B(R)} \big|\nabla(|A|^p f)\big|^2 \\ &= \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 + \int_{B(R)} |A|^{2p} |\nabla f|^2 + 2\int_{B(R)} f |A|^p \langle \nabla f, \nabla |A|^p\rangle. \end{aligned} \tag{9}$$

Plugging inequality (9) into (7), we have

$$\begin{aligned} &\left(1 - \frac{n-2}{np}\right) \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 \\ &\le \frac{p}{\lambda_1(M)} (2K_3 - nK_1 - nK_2) \biggl( \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 \\ &\qquad\qquad + |\nabla f|^2 |A|^{2p} + 2\int_{B(R)} f |A|^p \langle \nabla f, \nabla |A|^p\rangle \biggr) \\ &+ (p-1) \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 + p \int_{B(R)} |\nabla f|^2 |A|^{2p} + 2(p-1) \int_{B(R)} f |A|^p \langle \nabla f, \nabla |A|^p\rangle. \end{aligned}$$

Thus

$$\begin{aligned} \left(1 - \frac{n-2}{np}\right) \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 &\le \left( \frac{p}{\lambda_1(M)} (2K_3 - nK_1 - nK_2) + p - 1 \right) \int_{B(R)} f^2 \big|\nabla |A|^p\big|^2 \\ &\quad + \left( \frac{p}{\lambda_1(M)} (2K_3 - nK_1 - nK_2) + p \right) \int_{B(R)} |\nabla f|^2 |A|^{2p} \\ &\quad + 2\left( \frac{p}{\lambda_1(M)} (2K_3 - nK_1 - nK_2) + p - 1 \right) \int_{B(R)} f |A|^p \langle \nabla f, \nabla |A|^p\rangle. \end{aligned} \tag{10}$$

Note that Young's inequality yields

$$2\int_{B(R)} f |A|^p \langle \nabla f, \nabla |A|^p\rangle \le \varepsilon \int_{B(R)} |\nabla f|^2 |A|^{2p} + \frac{1}{\varepsilon} \int_{B(R)} f^2 |\nabla |A|^p|^2 \tag{11}$$

for any $\varepsilon > 0$. From inequalities (10) and (11), it follows that

$$\begin{aligned}\left(1-\frac{n-2}{np}\right)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 &\le \left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p-1\right)\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 \\ &\quad + \left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p\right)\int_{B(R)}|\nabla f|^2|A|^{2p} \\ + \left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p-1\right)&\left(\varepsilon\int_{B(R)}|\nabla f|^2|A|^{2p}+\frac{1}{\varepsilon}\int_{B(R)} f^2|\nabla|A|^p|^2\right),\end{aligned}$$

which yields that

$$\begin{aligned}\left[1-\frac{n-2}{np}-\left(1+\frac{1}{\varepsilon}\right)\left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p-1\right)\right]\int_{B(R)} f^2\big|\nabla|A|^p\big|^2 \\ \le\left[(1+\varepsilon)\left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p\right)-\varepsilon\right]\int_{B(R)}|\nabla f|^2|A|^{2p}.\end{aligned}$$

For a contradiction, we suppose that

$$\lambda_1(M) > \frac{p(2K_3-nK_1-nK_2)}{1-(n-2)/np-(p-1)} = \frac{np^2(2K_3-n(K_1+K_2))}{2-n(p-1)^2}.$$

Note the assumption that $1-\sqrt{2/n} < p < 1+\sqrt{2/n}$ is equivalent to

$$2-n(p-1)^2 > 0.$$

Choose a sufficiently large $\varepsilon > 0$ satisfying

$$\left[1-\frac{n-2}{np}-\left(1+\frac{1}{\varepsilon}\right)\left(\frac{p}{\lambda_1(M)}(2K_3-nK_1-nK_2)+p-1\right)\right] > 0.$$

Since $|\nabla f| \le 1/R$ by our choice of $f$, one can conclude that, by letting $R\to\infty$,

$$\int_M |\nabla|A|^p|^2 = 0,$$

where we used the growth condition on $\int_{B(R)}|A|^{2p}$. Thus we see that $|A|$ is constant. Since the volume of $M$ is infinite [Wei 2003], we get $|A|\equiv 0$. This implies that $M$ is totally geodesic, which is impossible by our assumption. Therefore we obtain the upper bound of $\lambda_1(M)$:

$$\lambda_1(M) \le \frac{np^2(2K_3-n(K_1+K_2))}{2-n(p-1)^2}. \qquad \square$$

Dung and the author [2012] gave an estimate of the bottom of the spectrum for the Laplace operator on a complete noncompact stable minimal hypersurface $M$ in a complete simply connected Riemannian manifold with pinched negative sectional curvature under the assumption on $L^2$-norm of the second fundamental form $A$ of $M$. In Theorem 2.1, if we take $p = 1$, we get the following.

**Corollary 2.2** [Dung and Seo 2012]. *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature satisfying $K_1 \le K_N \le K_2$, where $K_1$, $K_2$ are constants and $K_1 \le K_2 < 0$. Let $M$ be a complete stable non-totally geodesic minimal hypersurface in $N$. Assume that*

$$\lim_{R\to\infty} R^{-2} \int_{B(R)} |A|^2 = 0,$$

*where $B(R)$ is a geodesic ball of radius $R$ on $M$. If $|\nabla K|^2 = \sum_{i,j,k,l,m} K^2_{ijkl;m} \le K_3^2 |A|^2$ for some constant $K_3 > 0$, we have*

$$-K_2 \frac{(n-1)^2}{4} \le \lambda_1(M) \le \frac{(2K_3 - n(K_1 + K_2))n}{2}.$$

In particular, if $N$ is the $(n+1)$-dimensional hyperbolic space $\mathbb{H}^{n+1}$, one sees that $K_1 = K_2 = -1$, and hence $|\nabla K|^2 = 0$, that is, $K_3 = 0$. Moreover, it follows from McKean's result [1970] that the first eigenvalue $\lambda_1(M)$ of any complete totally geodesic hypersurface $M \subset \mathbb{H}^{n+1}$ satisfies $\lambda_1(M) = (n-1)^2/4$. Therefore we have the following consequence which is an extension of the result in [Seo 2011].

**Corollary 2.3.** *Let $M$ be a complete stable minimal hypersurface in $\mathbb{H}^{n+1}$ with $\int_M |A|^{2p}\, dv < \infty$ for $1 - \sqrt{2/n} < p < 1 + \sqrt{2/n}$. Then we have*

$$-K_2 \frac{(n-1)^2}{4} \le \lambda_1(M) \le \frac{2n^2 p^2}{2 - n(p-1)^2}.$$

As another application of Theorem 2.1, we have the following when $n < 8$.

**Corollary 2.4.** *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature satisfying $K_1 \le K_N \le K_2$, where $K_1$, $K_2$ are constants and $K_1 \le K_2 < 0$ for $n < 8$. Let $M$ be a complete stable non-totally geodesic minimal hypersurface in $N$. For $p = 1, 2, 3$, if $\int_M |A|^p < \infty$, we have*

$$-K_2 \frac{(n-1)^2}{4} \le \lambda_1(M) \le \frac{np^2(2K_3 - n(K_1 + K_2))}{2 - n(p-1)^2}.$$

*Proof.* Since $\sqrt{2/n} > 1/2$ when $n < 8$, the conclusion can be derived from Theorem 2.1. $\square$

## 3. Vanishing theorems on minimal hypersurfaces with $\lambda_1(M)$ bounded below

Before we prove the vanishing theorems for $L^p$ harmonic 1-forms on complete minimal hypersurface, we begin with some useful facts.

**Lemma 3.1** [Leung 1992]. *Let $M$ be an $n$-dimensional complete immersed minimal hypersurface in a Riemannian manifold $N$. If all the sectional curvatures of $N$ are bounded below by a constant $K$,*

$$\mathrm{Ric} \geq (n-1)K - \frac{n-1}{n}|A|^2.$$

**Lemma 3.2** [Wang 2001]. *Let $\omega$ be a harmonic 1-form on an $n$-dimensional Riemannian manifold $M$. Then*

$$|\nabla\omega|^2 - \big|\nabla|\omega|\big|^2 \geq \frac{1}{n-1}\big|\nabla|\omega|\big|^2. \tag{12}$$

We also need the following well-known Sobolev inequality on a Riemannian manifold.

**Lemma 3.3** [Hoffman and Spruck 1974]. *Let $M^n$ be a complete immersed minimal submanifold in a nonpositively curved manifold $N^{n+p}$, $n \geq 3$. Then, for any $\phi \in W_0^{1,2}(M)$, we have*

$$\left(\int_M |\phi|^{2n/(n-2)}\,dv\right)^{(n-2)/n} \leq C_s \int_M |\nabla\phi|^2\,dv, \tag{13}$$

*where $C_s$ is the Sobolev constant which depends only on $n \geq 3$.*

A complete Riemannian manifold $M$ is called *nonparabolic* if it admits a nonconstant positive superharmonic function. Otherwise, $M$ is said to be *parabolic*. The following sufficient condition for parabolicity is well known.

**Theorem** [Grigoryan 1983; 1985; Karp 1982; Varopoulos 1983]. *Let $M$ be a complete Riemannian manifold. If, for any point $p \in M$ and a geodesic ball $B_p(r)$,*

$$\int_1^\infty \frac{r}{\mathrm{Vol}(B_p(r))}\,dr = \infty,$$

*$M$ is parabolic.*

It immediately follows from this result that if $M$ is nonparabolic,

$$\int_1^\infty \frac{r}{\mathrm{Vol}(B_p(r))}\,dr < \infty,$$

and hence $M$ has infinite volume. Moreover, if $\lambda_1(M) > 0$, $M$ is nonparabolic [Grigoryan 1999]. Therefore one can conclude the following.

**Proposition 3.4.** *Let $M$ be an $n$-dimensional complete noncompact Riemannian manifold with $\lambda_1(M) > 0$. Then* $\mathrm{Vol}(M) = \infty$.

Note that, in the case of submanifolds, Cheung and Leung [1998] proved that the volume $\mathrm{Vol}(B_p(r))$ of every complete noncompact submanifold $M$ in the Euclidean or hyperbolic space grows at least as a linear function of $r$ under the assumption that the mean curvature vector $H$ of $M$ is bounded in absolute value.

We are now ready to state and prove vanishing theorems for $L^p$ harmonic 1-forms on a complete noncompact stable minimal hypersurface.

**Theorem 3.5.** *Let $N$ be an $(n+1)$-dimensional complete Riemannian manifold with sectional curvature satisfying $K \le K_N$ where $K \le 0$ is a constant. Let $M$ be a complete noncompact stable minimal hypersurface in $N$. Assume that, for $0 < p < n/(n-1) + \sqrt{2n}$,*

$$\lambda_1(M) > \frac{-2n(n-1)^2 p^2 K}{2n - [(n-1)p - n]^2}.$$

*Then there is no nontrivial $L^{2p}$ harmonic 1-form on $M$.*

*Proof.* We consider two cases: $K < 0$ and $K = 0$.

*Case* 1: $K < 0$. Let $\omega$ be an $L^{2p}$ harmonic 1-form on $M$, that is,

$$\Delta\omega = 0 \quad \text{and} \quad \int_M |\omega|^{2p}\, dv < \infty.$$

In an abuse of notation, we refer to both a harmonic 1-form and its dual harmonic vector field by $\omega$. Bochner's formula yields

$$\Delta|\omega|^2 = 2(|\nabla\omega|^2 + \mathrm{Ric}(\omega, \omega)).$$

Moreover,

$$\Delta|\omega|^2 = 2(|\omega|\Delta|\omega| + \big|\nabla|\omega|\big|^2).$$

Applying Lemma 3.1 and the Kato-type inequality (12), we see that

$$|\omega|\Delta|\omega| + \frac{n-1}{n}|A|^2|\omega|^2 - (n-1)K|\omega|^2 \ge \frac{1}{n-1}\big|\nabla|\omega|\big|^2. \tag{14}$$

For any positive number $p$, we have

$$\begin{aligned}
|\omega|^p \Delta|\omega|^p &= |\omega|^p \mathrm{div}(\nabla|\omega|^p) \\
&= |\omega|^p \mathrm{div}(p\,|\omega|^{p-1}\nabla|\omega|) \\
&= p(p-1)|\omega|^{2p-2}\big|\nabla|\omega|\big|^2 + p|\omega|^{2p-1}\Delta|\omega| \\
&= \frac{p-1}{p}\big|\nabla|\omega|^p\big|^2 + p|\omega|^{2p-2}|\omega|\Delta|\omega|.
\end{aligned}$$

Plugging inequality (14) into the above equality, we have

$$|\omega|^p\Delta|\omega|^p + p(n-1)\left(\frac{|A|^2}{n} - K\right)|\omega|^{2p} \geq \left(1 - \frac{1}{p} + \frac{1}{p(n-1)}\right)\left|\nabla|\omega|^p\right|^2.$$

Choose a Lipschitz function $f$ with compact support in a geodesic ball $B(R)$ of radius $R$ centered at $p \in M$. Multiplying both side by $f^2$ and integrating over $B(R)$, we obtain

$$\begin{aligned}&\left(1 - \frac{1}{p} + \frac{1}{p(n-1)}\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \\ &\quad\leq \int_{B(R)} f^2|\omega|^p\Delta|\omega|^p + \frac{p(n-1)}{n}\int_{B(R)} f^2|A|^2|\omega|^{2p} - p(n-1)K\int_{B(R)} f^2|\omega|^{2p}.\end{aligned}$$

The divergence theorem gives

$$\int_{B(R)} f^2|\omega|^p\Delta|\omega|^p = -\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 - 2\int_{B(R)} f|\omega|^p\langle\nabla f, \nabla|\omega|^p\rangle.$$

Thus

$$\begin{aligned}(15)\quad &\left(1 - \frac{1}{p} + \frac{1}{p(n-1)}\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \\ &\quad\leq \frac{p(n-1)}{n}\int_{B(R)} f^2|A|^2|\omega|^{2p} - p(n-1)K\int_{B(R)} f^2|\omega|^{2p} \\ &\qquad - \int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 - 2\int_{B(R)} f|\omega|^p\langle\nabla f, \nabla|\omega|^p\rangle.\end{aligned}$$

Since $M$ is stable,

$$\int_M |\nabla f|^2 - (|A|^2 + \overline{\mathrm{Ric}}(e_{n+1}))f^2 \geq 0$$

for any compactly supported Lipschitz function $f$ on $M$. From the assumption on the sectional curvature of $N$, it follows that

$$\int_M |\nabla f|^2 - (|A|^2 + nK)f^2 \geq 0$$

for any compactly supported Lipschitz function $f$ on $M$. Replacing $f$ by $|\omega|^p f$, we have

$$\begin{aligned}(16)\quad &\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 + \int_{B(R)} |\nabla f|^2|\omega|^{2p} + 2\int_{B(R)} f|\omega|^p\langle\nabla f, \nabla|\omega|^p\rangle \\ &\qquad\geq \int_{B(R)} f^2|A|^2|\omega|^{2p} + nK\int_{B(R)} f^2|\omega|^{2p}.\end{aligned}$$

Combining the inequalities (15) and (16) gives

$$\begin{aligned}\left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)&\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\ &\leq \frac{p(n-1)}{n}\bigg[\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\int_{B(R)}|\nabla f|^2|\omega|^{2p}\\ &\qquad +2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle - nK\int_{B(R)} f^2|\omega|^{2p}\bigg]\\ &\quad -p(n-1)K\int_{B(R)} f^2|\omega|^{2p}-\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2-2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}$$

Hence

$$\begin{aligned}\left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)&\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\ &\leq \left(\frac{p(n-1)}{n}-1\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\frac{p(n-1)}{n}\int_{B(R)}|\nabla f|^2|\omega|^{2p}\\ &\quad -2p(n-1)K\int_{B(R)} f^2|\omega|^{2p}+2\left(\frac{p(n-1)}{n}-1\right)\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}\tag{17}$$

Moreover, using the definition of the bottom of the spectrum, we see that

$$\begin{aligned}\lambda_1(M)&\int_{B(R)}|\omega|^{2p}f^2\\ &\leq \int_{B(R)}\big|\nabla(|\omega|^p f)\big|^2\\ &= \int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\int_{B(R)}|\omega|^{2p}|\nabla f|^2+2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}\tag{18}$$

From inequalities (17) and (18), it follows that

$$\begin{aligned}\left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)&\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\ &\leq \left(\frac{p(n-1)}{n}-1-\frac{2p(n-1)K}{\lambda_1(M)}\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\ &\quad +\left(\frac{p(n-1)}{n}-\frac{2p(n-1)K}{\lambda_1(M)}\right)\int_{B(R)}|\nabla f|^2|\omega|^{2p}\\ &\quad +2\left(\frac{p(n-1)}{n}-1-\frac{2p(n-1)K}{\lambda_1(M)}\right)\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}$$

Applying Young's inequality, we have

$$2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle \le \varepsilon\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\frac{1}{\varepsilon}\int_{B(R)}|\nabla f|^2|\omega|^{2p}$$

for any $\varepsilon>0$. Thus

$$\begin{aligned}\left[2-\frac{1}{p}+\frac{1}{p(n-1)}+\frac{2p(n-1)K}{\lambda_1(M)}-\frac{p(n-1)}{n}-\varepsilon\left(\frac{p(n-1)}{n}-1-\frac{2p(n-1)K}{\lambda_1(M)}\right)\right]&\\ \times\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2&\\ \le\left[\frac{p(n-1)}{n}-\frac{2p(n-1)K}{\lambda_1(M)}+\frac{1}{\varepsilon}\left(\frac{p(n-1)}{n}-1-\frac{2p(n-1)K}{\lambda_1(M)}\right)\right]\int_{B(R)}|\nabla f|^2|\omega|^{2p}.&\end{aligned}$$

Since

$$\lambda_1(M)>\frac{-2p(n-1)K}{2-1/p+1/(p(n-1))-p(n-1)/n}=\frac{-2n(n-1)^2p^2K}{2n-[(n-1)p-n]^2}$$

by the hypothesis, one can choose a sufficiently small $\varepsilon>0$ satisfying that

$$\left[2-\frac{1}{p}+\frac{1}{p(n-1)}+\frac{2p(n-1)K}{\lambda_1(M)}-\frac{p(n-1)}{n}-\varepsilon\left(\frac{p(n-1)}{n}-1-\frac{2p(n-1)K}{\lambda_1(M)}\right)\right]>0.$$

Note that $\int_M|\omega|^{2p}<\infty$, since $\omega$ is an $L^{2p}$ harmonic 1-form on $M$. Letting $R$ tend to infinity, we obtain

$$\int_M\big|\nabla|\omega|^p\big|^2=0,$$

which implies that $\big|\nabla|\omega|\big|\equiv 0$. Hence $|\omega|\equiv$ constant. From Proposition 3.4, it follows that $|\omega|\equiv 0$.

*Case* 2: $K=0$. Using the inequality (17) and Young's inequality, we obtain

$$\begin{aligned}\left[2-\frac{1}{p}+\frac{1}{p(n-1)}-\frac{p(n-1)}{n}-\varepsilon\left(\frac{p(n-1)}{n}-1\right)\right]\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2&\\ \le\left[\frac{p(n-1)}{n}+\frac{1}{\varepsilon}\left(\frac{p(n-1)}{n}-1\right)\right]\int_{B(R)}|\nabla f|^2|\omega|^{2p}.&\end{aligned}$$

Since $0<p<n/(n-1)+\sqrt{2n}$, one may choose a sufficiently small $\varepsilon>0$ satisfying

$$2-\frac{1}{p}+\frac{1}{p(n-1)}-\frac{p(n-1)}{n}-\varepsilon\left(\frac{p(n-1)}{n}-1\right)>0.$$

Letting $R$ tend to infinity gives

$$\int_{B(R)}\big|\nabla|\omega|^p\big|^2=0,$$

which implies that $|\omega| \equiv$ constant. From the assumption that $\lambda_1(M) > 0$ and Proposition 3.4, it follows that $|\omega| \equiv 0$. □

As a consequence of Theorem 3.5, given a complete noncompact stable minimal hypersurface in a nonnegatively curved Riemannian manifold, one has the following result.

**Corollary 3.6.** *Let $N$ be an $(n+1)$-dimensional complete nonnegatively curved Riemannian manifold. Let $M$ be a complete noncompact stable minimal hypersurface in $N$ with $\lambda_1(M) > 0$. If $n \leq 11$, there is no nontrivial $L^p$ harmonic $1$-form on $M$ for any $0 < p \leq n$.*

*Proof.* For $n \leq 11$, the inequality $2(n/(n-1) + \sqrt{2n}) \geq n$ holds. □

**Corollary 3.7.** *Let $N$ be an $(n+1)$-dimensional complete nonnegatively curved Riemannian manifold. Let $M$ be a complete noncompact stable minimal hypersurface in $N$ with $\lambda_1(M) > 0$. If $n \leq 11$, there is no nontrivial $L^2$ harmonic $1$-form on $M$.*

In the case of $L^2$ harmonic 1-forms, Theorem 3.5 gives a generalization of [Dung and Seo 2012] as follows.

**Corollary 3.8.** *Let $N$ be an $(n+1)$-dimensional complete Riemannian manifold with sectional curvature satisfying $K \leq K_N$ where $K < 0$ is a constant. Let $M$ be a complete noncompact stable minimal hypersurface in $N$. Assume that*

$$\lambda_1(M) > \frac{-2n(n-1)^2 K}{2n-1}.$$

*Then there are no nontrivial $L^2$ harmonic $1$-forms on $M$.*

In particular, if $N$ is $(n+1)$-dimensional hyperbolic space $\mathbb{H}^{n+1}$, Corollary 3.8 improves the previous result of [Seo 2010]. Related to this result, Cavalcante, Mirandola, and Vitório [Cavalcante et al. 2012] obtained the vanishing theorem for $L^2$ harmonic 1-forms on complete noncompact submanifolds in a Cartan–Hadamard manifold.

Palmer [1991] showed that if there exists a codimension-one cycle in a complete minimal hypersurface $M$ in $\mathbb{R}^{n+1}$ which does not separate $M$, $M$ is unstable. We obtain a generalization of Palmer's result as follows.

**Corollary 3.9.** *Let $N$ be an $(n+1)$-dimensional complete Riemannian manifold with sectional curvature satisfying $K \leq K_N$ where $K \leq 0$ is a constant. Let $M$ be a complete noncompact minimal hypersurface in $N$. Assume that*

$$\lambda_1(M) > \frac{-2n(n-1)^2 K}{2n-1}.$$

*Suppose that there exists a codimension-one cycle in $M$ which does not separate $M$. Then $M$ cannot be stable.*

*Proof.* Suppose that $M$ is stable in $N$. From [Dodziuk 1982], there exists a nontrivial $L^2$ harmonic 1-form on $M$, which is a contradiction to Corollary 3.8. □

Let $M$ be a complete Riemannian manifold and let $f$ be a harmonic function on $M$ with finite $L^p$ energy. Then the total differential $df$ is obviously an $L^p$ harmonic 1-form on $M$. As another application of Theorem 3.5, we prove the following Liouville type theorem for harmonic functions with finite $L^p$ energy on a complete noncompact stable minimal hypersurface, which is a generalization of Schoen and Yau's result [1976], as mentioned in the introduction.

**Corollary 3.10.** *Let $N$ be an $(n+1)$-dimensional complete Riemannian manifold with sectional curvature satisfying $K \le K_N$ where $K \le 0$ is a constant. Let $M$ be a complete noncompact stable minimal hypersurface in $N$. Assume that, for $0 < p < n/(n-1) + \sqrt{2n}$,*

$$\lambda_1(M) > \frac{-2n(n-1)^2 p^2 K}{2n - [(n-1)p - n]^2}.$$

*Then there is no nontrivial harmonic function on $M$ with finite $L^p$ energy.*

So far, we have assumed that $\lambda_1(M) > 0$ for a complete noncompact stable minimal hypersurface $M$ in a nonnegatively curved Riemannian manifold. However, we do not know whether the assumption that $\lambda_1(M) > 0$ is necessary or not. It would be interesting to remove the condition in these results.

## 4. Vanishing theorems on minimal hypersurfaces with small $L^n$ or $L^\infty$ norm of the second fundamental form

In the following, we prove a vanishing theorem for $L^p$ harmonic 1-forms on a complete stable minimal hypersurface $M$, assuming that $M$ has sufficiently small total scalar curvature instead of assuming that $M$ is stable.

**Theorem 4.1.** *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature $K_N$ satisfying that $K_1 \le K_N \le K_2 < 0$, where $K_1$, $K_2$ are constants and $n \ge 3$. Let $M$ be a complete minimal hypersurface in $N$. Assume that $K := K_2/K_1$ satisfies*

$$K > \frac{4(n-2)}{(n-1)^2}.$$

*For*

$$\frac{(n-1)K}{4}-\frac{1}{2}\sqrt{\frac{(n-1)^2K^2}{4}-(n-2)K} < p < \frac{(n-1)K}{4}+\frac{1}{2}\sqrt{\frac{(n-1)^2K^2}{4}-(n-2)K},$$

*assume that*

$$\left(\int_M |A|^n\right)^{2/n} < \frac{n(2p(n-1)-n+2-4p^2K)}{p^2(n-1)^2C_s},$$

*where $C_s$ is the Sobolev constant in [Hoffman and Spruck 1974]. Then there are no nontrivial $L^{2p}$ harmonic 1-forms on $M$.*

*Proof.* A similar argument as in the proof of Theorem 3.5 shows

$$|\omega|^p\Delta|\omega|^p + p(n-1)\left(\frac{|A|^2}{n}-K_1\right)|\omega|^{2p} \geq \left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)\big|\nabla|\omega|^p\big|^2$$

for any Lipschitz function $f$ with compact support in a geodesic ball $B(R)$ of radius $R$ centered at a point $p \in M$. Multiplying both sides by $f^2$, integrating over $B(R)$, and applying the divergence theorem, we see that

$$\begin{aligned}(19)\quad \left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2 &\leq \frac{p(n-1)}{n}\int_{B(R)} f^2|A|^2|\omega|^{2p} - p(n-1)K_1\int_{B(R)} f^2|\omega|^{2p} \\ &\quad - \int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2 - 2\int_{B(R)} f|\omega|^p\langle\nabla f, \nabla|\omega|^p\rangle.\end{aligned}$$

On the other hand, the Sobolev inequality (13) implies that

$$\begin{aligned}\int_{B(R)} f^2|A|^2|\omega|^{2p} &\leq \left(\int_M |A|^n\right)^{2/n}\left(\int_M (|\omega|^p f)^{(2n)/n-2}\right)^{(n-2)/n} \\ &\leq C_s\left(\int_M |A|^n\right)^{2/n}\int_M \big|\nabla(|\omega|^p f)\big|^2 \\ &\leq C_s\left(\int_M |A|^n\right)^{2/n}\left(\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2 + \int_{B(R)} |\nabla f|^2|\omega|^{2p} \right. \\ &\qquad \left. + 2\int_{B(R)} f|\omega|^p\langle\nabla f, \nabla|\omega|^p\rangle\right).\end{aligned}$$

Plugging this inequality into (19) gives

$$
\begin{aligned}
(20)\quad \left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2
&\le \frac{p(n-1)C_s}{n}\left(\int_M |A|^n\right)^{2/n}\int_{B(R)}|\nabla f|^2|\omega|^{2p}\\
&\quad+\left(\frac{p(n-1)C_s}{n}\left(\int_M |A|^n\right)^{2/n}-1\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\
&\quad+2\left(\frac{p(n-1)C_s}{n}\left(\int_M |A|^n\right)^{2/n}-1\right)\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle\\
&\quad-p(n-1)K_1\int_{B(R)} f^2|\omega|^{2p}.
\end{aligned}
$$

An estimate (1) for the bottom of the spectrum yields

$$
-\frac{K_2(n-1)^2}{4}\le\lambda_1(M)\le\frac{\int_{B(R)}\big|\nabla(|\omega|^p f)\big|^2}{\int_{B(R)}(|\omega|^p f)^2},
$$

which gives

$$
\begin{aligned}
(21)\quad \int_{B(R)}(|\omega|^p f)^2
&\le -\frac{4}{K_2(n-1)^2}\left(\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\int_{B(R)} u|\nabla f|^2|\omega|^{2p}\right.\\
&\qquad\left.+2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle\right).
\end{aligned}
$$

Thus, from inequalities (20) and (21), it follows that

$$
\begin{aligned}
&\left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2\\
&\le B\int_{B(R)}|\nabla f|^2|\omega|^{2p}+(B-1)\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+2(B-1)\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle,
\end{aligned}
$$

where

$$
B=\frac{p(n-1)C_s}{n}\left(\int_M |A|^n\right)^{2/n}+\frac{4p}{(n-1)}\frac{1}{K}.
$$

Applying Young's inequality

$$
2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle\le\varepsilon\int_{B(R)} f^2\big|\nabla|\omega|^p\big|^2+\frac{1}{\varepsilon}\int_{B(R)}|\nabla f|^2|\omega|^{2p}
$$

for any $\varepsilon > 0$, we see that

$$\left(2-\frac{1}{p}+\frac{1}{p(n-1)}-B-\varepsilon(B-1)\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \le \left(B+\frac{1}{\varepsilon}(B-1)\right)\int_{B(R)} |\nabla f|^2|\omega|^{2p}.$$

From the assumption on the total curvature of $M$, one can make

$$\left(2-\frac{1}{p}+\frac{1}{p(n-1)}-B-\varepsilon(B-1)\right) > 0$$

by choosing a sufficiently small $\varepsilon > 0$. Letting $R \to \infty$ and using that $\omega$ is an $L^{2p}$ harmonic 1-form, we conclude that

$$\int_M \left|\nabla|\omega|^p\right|^2 = 0.$$

The same argument as before shows that $|\omega| \equiv 0$. □

**Corollary 4.2.** *Let $M$ be a complete minimal hypersurface in $\mathbb{H}^{n+1}$ satisfying*

$$\left(\int_M |A|^n\right)^{2/n} < \frac{n(-4p^2+2p(n-1)-n+2)}{p^2(n-1)^2 C_s}$$

*for $1/2 < p < n/2-1$. Then there are no nontrivial $L^{2p}$ harmonic 1-forms on $M$.*

**Corollary 4.3.** *Under the same conditions as in Theorem 4.1, there is no nontrivial harmonic function on $M$ with finite $L^p$ energy.*

When the $L^\infty$ norm of the second fundamental form of a complete minimal hypersurface is bounded, the following vanishing theorem holds.

**Theorem 4.4.** *Let $N$ be an $(n+1)$-dimensional complete simply connected Riemannian manifold with sectional curvature $K_N$ satisfying $K_1 \le K_N \le K_2 < 0$, where $K_1, K_2$ are constants and $n \ge 3$. Let $M$ be a complete noncompact minimal hypersurface in $N$. Assume that $K := K_2/K_1 > 4(n-2)/(n-1)^2$ and the second fundamental form $A$ satisfies*

$$|A|^2 \le C < \frac{4p^2K_1-(2p(n-1)-n+2)K_2}{4p^2}$$

*for*

$$\frac{(n-1)K}{4}-\frac{1}{2}\sqrt{\frac{(n-1)^2K^2}{4}-(n-2)K} < p < \frac{(n-1)K}{4}+\frac{1}{2}\sqrt{\frac{(n-1)^2K^2}{4}-(n-2)K}.$$

*Then there are no nontrivial $L^{2p}$ harmonic 1-forms on $M$.*

*Proof.* A similar argument as before shows

$$\begin{aligned}\left(1-\frac{1}{p}+\frac{1}{p(n-1)}\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \\ \le \frac{p(n-1)}{n}\int_{B(R)} f^2|A|^2|\omega|^{2p} - p(n-1)K_1\int_{B(R)} f^2|\omega|^{2p} \\ -\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 - 2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}$$

Since $|A|^2 \le C$,

$$\begin{aligned}\left(2-\frac{1}{p}+\frac{1}{p(n-1)}\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \\ \le \left(\frac{p(n-1)C}{n} - p(n-1)K_1\right)\int_{B(R)} f^2|\omega|^{2p} - 2\int_{B(R)} f|\omega|^p\langle\nabla f,\nabla|\omega|^p\rangle.\end{aligned}$$

Using an estimate for the bottom of the spectrum and Young's inequality again, we have

$$\begin{aligned}\left(2-\frac{1}{p}+\frac{1}{p(n-1)}-D-\varepsilon(D-1)\right)\int_{B(R)} f^2\left|\nabla|\omega|^p\right|^2 \\ \le \left(D+\frac{1}{\varepsilon}(D-1)\right)\int_{B(R)}|\nabla f|^2|\omega|^{2p},\end{aligned}$$

where

$$D = \frac{-4}{(n-1)^2K_2}\left(\frac{p(n-1)C}{n} - p(n-1)K_1\right).$$

Since

$$C < \frac{4p^2K_1 - (2p(n-1)-n+2)K_2}{4p^2},$$

by our assumption, we may choose a sufficiently small $\varepsilon > 0$ satisfying

$$\left(2-\frac{1}{p}+\frac{1}{p(n-1)}-D-\varepsilon(D-1)\right) > 0.$$

Thus we get

$$\int_{B(R)}\left|\nabla|\omega|^p\right|^2 = 0$$

by letting $R$ tend to infinity. Hence $\omega \equiv 0$. □

**Corollary 4.5.** *Let $M$ be a complete minimal hypersurface in $\mathbb{H}^{n+1}$ with the second fundamental form $A$ satisfying*

$$|A|^2 \le C < \frac{-4p^2 + 2p(n-1) - n + 2}{4p^2}$$

*for $1/2 < p < n/2 - 1$. Then there are no nontrivial $L^{2p}$ harmonic 1-forms on $M$.*

**Corollary 4.6.** *Under the same conditions as in Theorem 4.4, there is no nontrivial harmonic function on $M$ with finite $L^p$ energy.*

We remark that there are lots of examples of minimal hypersurfaces with finite $L^n$ or $L^\infty$ norm of the second fundamental form in $\mathbb{H}^{n+1}$ [do Carmo and Dajczer 1983; Mori 1981; Ripoll 1989; Seo 2011].

## References

[Anderson 1988] M. T. Anderson, "$L^2$ harmonic forms on complete Riemannian manifolds", pp. 1–19 in *Geometry and analysis on manifolds* (Katata/Kyoto, 1987), edited by T. Sunada, Lecture Notes in Math. **1339**, Springer, Berlin, 1988. MR 89j:58004 Zbl 0652.53030

[Bessa and Montenegro 2003] G. P. Bessa and J. F. Montenegro, "Eigenvalue estimates for submanifolds with locally bounded mean curvature", *Ann. Global Anal. Geom.* **24**:3 (2003), 279–290. MR 2004f:53068 Zbl 1060.53063

[Candel 2007] A. Candel, "Eigenvalue estimates for minimal surfaces in hyperbolic space", *Trans. Amer. Math. Soc.* **359**:8 (2007), 3567–3575. MR 2007m:53076 Zbl 1115.53005

[Cao et al. 1997] H.-D. Cao, Y. Shen, and S. Zhu, "The structure of stable minimal hypersurfaces in $\mathbf{R}^{n+1}$", *Math. Res. Lett.* **4**:5 (1997), 637–644. MR 99a:53037 Zbl 0906.53004

[do Carmo and Dajczer 1983] M. do Carmo and M. Dajczer, "Rotation hypersurfaces in spaces of constant curvature", *Trans. Amer. Math. Soc.* **277**:2 (1983), 685–709. MR 85b:53055 Zbl 0518.53059

[Carron 2002] G. Carron, "$L^2$ harmonic forms on non-compact Riemannian manifolds", pp. 49–59 in *Surveys in analysis and operator theory* (Canberra, 2001), edited by A. Hassell, Proc. Centre Math. Appl. Austral. Nat. Univ. **40**, Austral. Nat. Univ., Canberra, 2002. MR 2003j:58001 Zbl 1038.58023

[Cavalcante et al. 2012] M. Cavalcante, H. Mirandola, and F. Vitorio, "$L^2$ harmonic 1-forms on submanifolds with finite total curvature", preprint, 2012. arXiv 1201.5392

[Cheung and Leung 1998] L.-F. Cheung and P.-F. Leung, "The mean curvature and volume growth of complete noncompact submanifolds", *Differential Geom. Appl.* **8**:3 (1998), 251–256. MR 99k:53111 Zbl 0942.53037

[Cheung and Leung 2001] L.-F. Cheung and P.-F. Leung, "Eigenvalue estimates for submanifolds with bounded mean curvature in the hyperbolic space", *Math. Z.* **236**:3 (2001), 525–530. MR 2002c:53094 Zbl 0990.53029

[Colding and Minicozzi 1996] T. H. Colding and W. P. Minicozzi, II, "Generalized Liouville properties of manifolds", *Math. Res. Lett.* **3**:6 (1996), 723–729. MR 97h:53039 Zbl 0884.58032

[Colding and Minicozzi 1997] T. H. Colding and W. P. Minicozzi, II, "Harmonic functions on manifolds", *Ann. of Math.* (2) **146**:3 (1997), 725–747. MR 98m:53052 Zbl 0928.53030

[Colding and Minicozzi 1998] T. H. Colding and W. P. Minicozzi, II, "Liouville theorems for harmonic sections and applications", *Comm. Pure Appl. Math.* **51**:2 (1998), 113–138. MR 98m:53053 Zbl 0928.58022

[Deng 2008] Q. Deng, "Complete hypersurfaces with constant mean curvature and finite $L^p$ norm curvature in Euclidean spaces", *Arch. Math.* (*Basel*) **90**:4 (2008), 360–373. MR 2009i:53054 Zbl 1137.53017

[Dodziuk 1982] J. Dodziuk, "$L^2$−harmonic forms on complete manifolds.", in *Semin. differential geometry*, edited by S.-T. Yao, Ann. Math. Stud. **102**, 1982. Zbl 0484.53033

[Dung and Seo 2012] N. T. Dung and K. Seo, "Stable minimal hypersurfaces in a Riemannian manifold with pinched negative sectional curvature", *Ann. Global Anal. Geom.* **41**:4 (2012), 447–460. MR 2891296 Zbl 1242.53073

[Fu 2012] H.-P. Fu, "Bernstein type theorems for complete submanifolds in space forms", *Math. Nachr.* **285**:2-3 (2012), 236–244. MR 2012m:53135 Zbl 06012479

[Greene and Wu 1974] R. E. Greene and H. Wu, "Integrals of subharmonic functions on manifolds of nonnegative curvature", *Invent. Math.* **27** (1974), 265–298. MR 52 #3605 Zbl 0342.31003

[Greene and Wu 1981] R. E. Greene and H. Wu, "Harmonic forms on noncompact Riemannian and Kähler manifolds", *Michigan Math. J.* **28**:1 (1981), 63–81. MR 82e:58005

[Grigoryan 1983] A. A. Grigoryan, "Existence of the Green function on a manifold", *Uspekhi Mat. Nauk* **38**:1(229) (1983), 161–162. In Russian; translated in *Russian Math. Surveys* **38**:1 (1983), 190–191. MR 84i:58128

[Grigoryan 1985] A. A. Grigoryan, "The existence of positive fundamental solutions of the Laplace equation on Riemannian manifolds", *Mat. Sb.* (*N.S.*) **128(170)**:3 (1985), 354–363, 446. MR 87d:58140

[Grigoryan 1999] A. Grigoryan, "Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds", *Bull. Amer. Math. Soc.* (*N.S.*) **36**:2 (1999), 135–249. MR 99k:58195

[Hoffman and Spruck 1974] D. Hoffman and J. Spruck, "Sobolev and isoperimetric inequalities for Riemannian submanifolds", *Comm. Pure Appl. Math.* **27** (1974), 715–727. MR 51 #1676 Zbl 0295.53025

[Karp 1982] L. Karp, "Subharmonic functions, harmonic mappings and isometric immersions.", in *Semin. differential geometry*, edited by S.-T. Yau, Ann. Math. Stud. **102**, 1982. Zbl 0487.53046

[Leung 1992] P. F. Leung, "An estimate on the Ricci curvature of a submanifold and some applications", *Proc. Amer. Math. Soc.* **114**:4 (1992), 1051–1061. MR 92g:53052 Zbl 0753.53003

[Li and Schoen 1984] P. Li and R. Schoen, "$L^p$ and mean value properties of subharmonic functions on Riemannian manifolds", *Acta Math.* **153**:3-4 (1984), 279–301. MR 86j:58147 Zbl 0556.31005

[Li and Tam 1987] P. Li and L.-F. Tam, "Positive harmonic functions on complete manifolds with nonnegative curvature outside a compact set", *Ann. of Math.* (2) **125**:1 (1987), 171–207. MR 88m:58039

[Li and Tam 1992] P. Li and L.-F. Tam, "Harmonic functions and the structure of complete manifolds", *J. Differential Geom.* **35**:2 (1992), 359–383. MR 93b:53033 Zbl 0768.53018

[Li and Wang 2002] P. Li and J. Wang, "Minimal hypersurfaces with finite index", *Math. Res. Lett.* **9**:1 (2002), 95–103. MR 2003b:53066 Zbl 1019.53025

[Li and Wang 2004] P. Li and J. Wang, "Stable minimal hypersurfaces in a nonnegatively curved manifold", *J. Reine Angew. Math.* **566** (2004), 215–230. MR 2005e:53093 Zbl 1050.53049

[McKean 1970] H. P. McKean, "An upper bound to the spectrum of $\Delta$ on a manifold of negative curvature", *J. Differential Geometry* **4** (1970), 359–366. MR 42 #1009 Zbl 0197.18003

[Miyaoka 1993] R. Miyaoka, "$L^2$ harmonic 1-forms on a complete stable minimal hypersurface", pp. 289–293 in *Geometry and global analysis* (Sendai, 1993), edited by T. Kotake et al., Tohoku Univ., Sendai, 1993. MR 96g:53102 Zbl 0912.53042

[Mori 1981] H. Mori, "Minimal surfaces of revolution in $H^3$ and their global stability", *Indiana Univ. Math. J.* **30**:5 (1981), 787–794. MR 82k:53082 Zbl 0589.53007

[Ni 2001] L. Ni, "Gap theorems for minimal submanifolds in $\mathbf{R}^{n+1}$", *Comm. Anal. Geom.* **9**:3 (2001), 641–656. MR 2002m:53097 Zbl 1020.53041

[Palmer 1991] B. Palmer, "Stability of minimal hypersurfaces", *Comment. Math. Helv.* **66**:2 (1991), 185–188. MR 92m:58023 Zbl 0736.53054

[Pigola and Veronelli 2012] S. Pigola and G. Veronelli, "Remarks on $L^p$-vanishing results in geometric analysis", *Internat. J. Math.* **23**:1 (2012), 1250008, 18. MR 2888937 Zbl 1252.53072

[Pigola et al. 2005] S. Pigola, M. Rigoli, and A. G. Setti, "Vanishing theorems on Riemannian manifolds, and geometric applications", *J. Funct. Anal.* **229**:2 (2005), 424–461. MR 2006k:53055 Zbl 1087.58022

[Pigola et al. 2008] S. Pigola, M. Rigoli, and A. G. Setti, *Vanishing and finiteness results in geometric analysis: a generalization of the Bochner technique*, Progress in Mathematics **266**, Birkhäuser, Basel, 2008. MR 2009m:58001 Zbl 1150.53001

[Ripoll 1989] J. B. Ripoll, "Helicoidal minimal surfaces in hyperbolic space", *Nagoya Math. J.* **114** (1989), 65–75. MR 91a:53015 Zbl 0699.53067

[Schoen and Yau 1976] R. Schoen and S. T. Yau, "Harmonic maps and the topology of stable hypersurfaces and manifolds with non-negative Ricci curvature", *Comment. Math. Helv.* **51**:3 (1976), 333–341. MR 55 #11302 Zbl 0361.53040

[Schoen et al. 1975] R. Schoen, L. Simon, and S. T. Yau, "Curvature estimates for minimal hypersurfaces", *Acta Math.* **134**:3-4 (1975), 275–288. MR 54 #11243 Zbl 0323.53039

[Scott 1995] C. Scott, "$L^p$ theory of differential forms on manifolds", *Trans. Amer. Math. Soc.* **347**:6 (1995), 2075–2096. MR 95i:58009 Zbl 0849.58002

[Seo 2008] K. Seo, "Minimal submanifolds with small total scalar curvature in Euclidean space", *Kodai Math. J.* **31**:1 (2008), 113–119. MR 2009f:53095 Zbl 1147.53313

[Seo 2010] K. Seo, "$L^2$ harmonic 1-forms on minimal submanifolds in hyperbolic space", *J. Math. Anal. Appl.* **371**:2 (2010), 546–551. MR 2011j:58054 Zbl 1195.53087

[Seo 2011] K. Seo, "Stable minimal hypersurfaces in the hyperbolic space", *J. Korean Math. Soc.* **48**:2 (2011), 253–266. MR 2012c:53096 Zbl 1211.53080

[Seo 2012] K. Seo, "Isoperimetric inequalities for submanifolds with bounded mean curvature", *Monatsh. Math.* **166**:3-4 (2012), 525–542. MR 2925153 Zbl 1245.58009

[Shen and Zhu 2005] Y. B. Shen and X. H. Zhu, "On complete hypersurfaces with constant mean curvature and finite $L^p$-norm curvature in $\mathbf{R}^{n+1}$", *Acta Math. Sin.* (*Engl. Ser.*) **21**:3 (2005), 631–642. MR 2006f:53088 Zbl 1087.53054

[Varopoulos 1983] N. T. Varopoulos, "Potential theory and diffusion on Riemannian manifolds", pp. 821–837 in *Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I, II* (Chicago, 1981), edited by W. Beckner et al., Wadsworth, Belmont, CA, 1983. MR 85a:58103 Zbl 0558.31009

[Wang 2001] X. Wang, "On conformally compact Einstein manifolds", *Math. Res. Lett.* **8**:5-6 (2001), 671–688. MR 2003d:53075 Zbl 1053.53030

[Wei 2003] S. W. Wei, "The structure of complete minimal submanifolds in complete manifolds of nonpositive curvature", *Houston J. Math.* **29**:3 (2003), 675–689. MR 2004f:53075 Zbl 1130.53307

[Xin 2005] Y. L. Xin, "Berstein type theorems without graphic condition", *Asian J. Math.* **9**:1 (2005), 31–44. MR 2006b:53010

[Yau 1976] S. T. Yau, "Some function-theoretic properties of complete Riemannian manifold and their applications to geometry", *Indiana Univ. Math. J.* **25**:7 (1976), 659–670. MR 54 #5502 Zbl 0335.53041

[Yun 2002] G. Yun, "Total scalar curvature and $L^2$ harmonic 1-forms on a minimal hypersurface in Euclidean space", *Geom. Dedicata* **89** (2002), 135–141. MR 2003a:53091

Received October 9, 2012.

KEOMKYO SEO
DEPARTMENT OF MATHEMATICS
SOOKMYUNG WOMEN'S UNIVERSITY
HYOCHANGWONGIL 52
SEOUL 140-742
SOUTH KOREA
kseo@sookmyung.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.231

# RECONSTRUCTION FROM KOSZUL HOMOLOGY AND APPLICATIONS TO MODULE AND DERIVED CATEGORIES

RYO TAKAHASHI

**Let $R$ be a commutative noetherian ring and $M$ a finitely generated $R$-module. In this paper, we reconstruct $M$ from its Koszul homology with respect to a suitable sequence of elements of $R$ by taking direct summands, syzygies and extensions, and count the number of those operations. Using this result, we consider generation and classification of certain subcategories of the category of finitely generated $R$-modules, its bounded derived category and the singularity category of $R$.**

## 1. Introduction

For the past five decades, a lot of classification theorems of subcategories of abelian categories and triangulated categories have been given in ring theory, representation theory, algebraic geometry and algebraic topology; see, for instance, [Balmer 2002; 2005; Benson et al. 2011; Dao and Takahashi 2014; Friedlander and Pevtsova 2007; Gabriel 1962; Hopkins and Smith 1998; Hovey 2001; Krause 2008; Krause and Stevenson 2013; Neeman 1992; Stevenson 2014; Takahashi 2010; 2013; Thomason 1997]. Reconstruction of an object from its *support* in the spectrum of a suitable commutative ring plays a crucial role in the proofs of those theorems.

The notion of dimension for triangulated categories was introduced by Bondal and Van den Bergh [2003] and by Rouquier [2008]; analogues for abelian categories were introduced by Dao and Takahashi [2011; 2012a]. They essentially indicate the number of *extensions* necessary to build all objects out of a single object. There are many related studies; for example, see [Aihara and Takahashi 2011; Avramov et al. 2010a; Ballard et al. 2012; Bergh et al. 2010; Burke et al. 2012; Christensen 1998; Krause and Kussin 2006; Dao and Takahashi 2012b; Oppermann 2009; Orlov 2009b; Rouquier 2006; Schoutens 2003; Takahashi 2009].

Takahashi was partially supported by JSPS Grant-in-Aid for Young Scientists (B) 22740008 and by JSPS Postdoctoral Fellowships for Research Abroad.

*MSC2010:* 18E30, 18E35, 13C60, 13D09.

*Keywords:* Koszul complex, Koszul homology, module category, derived category, singularity category, resolving subcategory, thick subcategory.

In this paper, we study reconstructing a given module from its Koszul homology and counting the number of necessary operations. Our main result is the following theorem.

**Theorem 1.1.** *Let $R$ be a commutative noetherian ring, and let $M$ be a finitely generated $R$-module. Let $\boldsymbol{x} = x_1, \ldots, x_n$ be a sequence of elements of $R$ such that $M$ is locally free on $\mathsf{D}(\boldsymbol{x})$. Then there exists a positive integer $k$ such that the Koszul complex $\mathsf{K}(\boldsymbol{x}^k, M)$ is equivalent to a complex of finitely generated $R$-modules*

$$(0 \to N \to P_{n-1} \to \cdots \to P_0 \to 0),$$

*where $P_0, \ldots, P_{n-1}$ are projective and $M$ is a direct summand of $N$. In particular, $M$ can be built out of the Koszul homologies $\mathsf{H}_0(\boldsymbol{x}^k, M), \ldots, \mathsf{H}_n(\boldsymbol{x}^k, M)$ by taking $n$ syzygies, $n$ extensions and 1 direct summand.*

Note that since the free locus of a finitely generated $R$-module is an open subset of $\operatorname{Spec} R$ in the Zariski topology, there exist many such sequences $\boldsymbol{x}$ that satisfy the assumption of the theorem. We shall prove a more general result in Theorem 3.1.

Theorem 1.1 has a lot of applications. To state some of them, we fix notation. Let $\operatorname{mod} R$ be the category of finitely generated $R$-modules and $\mathbf{D}_{\mathrm{b}}(R)$ the bounded derived category of $\operatorname{mod} R$. We denote by $\mathbf{D}_{\mathrm{sg}}(R)$ the *singularity category* of $R$. This category has been introduced and studied by Buchweitz [1986] in connection with Cohen–Macaulay modules over Gorenstein rings. In recent years, it has been investigated by Orlov [2004; 2006; 2009a; 2011; 2012] in relation to the homological mirror symmetry conjecture.

Let $\mathsf{S}(R)$ be the set of prime ideals $\mathfrak{p}$ of $R$ such that $R_{\mathfrak{p}}$ is not a field, and denote by $\operatorname{Sing} R$ the singular locus of $R$. Applying Theorem 1.1, we can prove the following result on classification of subcategories.

**Corollary 1.2.** *Let $R$ be a commutative noetherian ring.*

(1) *There is a one-to-one correspondence between*:

(a) *the specialization-closed subsets of* $\mathsf{S}(R)$,

(b) *the resolving subcategories of* $\operatorname{mod} R$ *generated by a Serre subcategory of* $\operatorname{mod} R$.

(2) *There are one-to-one correspondences among*:

(a) *the specialization-closed subsets of* $\operatorname{Sing} R$,

(b) *the thick subcategories of* $\mathbf{D}_{\mathrm{b}}(R)$ *generated by $R$ and a Serre subcategory of* $\operatorname{mod} R$,

(c) *the thick subcategories of* $\mathbf{D}_{\mathrm{sg}}(R)$ *generated by a Serre subcategory of* $\operatorname{mod} R$.

When $R$ is local, let $\mathsf{mod}^\circ(R)$ (respectively, $\mathbf{D}_{\mathrm{b}}^\circ(R)$, $\mathbf{D}_{\mathrm{sg}}^\circ(R)$) be the full subcategories of $\mathsf{mod}\, R$ (respectively, $\mathbf{D}_{\mathrm{b}}(R)$, $\mathbf{D}_{\mathrm{sg}}(R)$) consisting of modules (respectively, complexes) that are locally free (respectively, perfect, zero) on the punctured spectrum of $R$. Applying Theorem 1.1, we can prove the following result on generation of subcategories.

**Corollary 1.3.** *Let $R$ be a commutative noetherian local ring of Krull dimension $d$ with residue field $k$.*

(1) *Every object in* $\mathsf{mod}^\circ(R)$ *is built out of a module of finite length by taking $d$ extensions in* $\mathsf{mod}\, R$, *up to finite direct sums, direct summands and syzygies.*

(2) *Every object in $\mathbf{D}_{\mathrm{sg}}^\circ(R)$ is built out of a module of finite length by taking $d$ extensions in $\mathbf{D}_{\mathrm{sg}}(R)$, up to finite direct sums, direct summands and shifts.*

*In particular, one has that* $\mathsf{mod}^\circ(R)$ *is generated by $k$ as a resolving subcategory of* $\mathsf{mod}\, R$, *that $\mathbf{D}_{\mathrm{b}}^\circ(R)$ is generated by $R$ and $k$ as a thick subcategory of $\mathbf{D}_{\mathrm{b}}(R)$, and that $\mathbf{D}_{\mathrm{sg}}^\circ(R)$ is generated by $k$ as a thick subcategory of $\mathbf{D}_{\mathrm{sg}}(R)$.*

Corollary 1.3 yields variants of results shown by Schoutens [2003] and Takahashi [2009; 2010]. It also recovers a result on isolated singularities given by Keller–Murfet–Van den Bergh [2011]. Furthermore, utilizing it, one can show the following result.

**Corollary 1.4.** *Let $R$ be a commutative noetherian ring. The following are equivalent for a resolving subcategory $\mathcal{X}$ of* $\mathsf{mod}\, R$:

(1) *$\mathcal{X}$ is generated by a Serre subcategory of* $\mathsf{mod}\, R$.

(2) *$\mathcal{X}$ is closed under tensor products and transposes.*

*Hence there is a one-to-one correspondence between the specialization-closed subsets of $\mathsf{S}(R)$ and the resolving subcategories of* $\mathsf{mod}\, R$ *closed under tensor products and transposes.*

The last assertion of this corollary greatly improves the main result of [Takahashi 2013]. Indeed, it removes the superfluous assumptions that $R$ is local and that $R$ is Cohen–Macaulay.

The organization of this paper is as follows. In the next Section 2 we prepare some fundamental notions. In Section 3 we state and prove the most general result in this paper, which includes Theorem 1.1. In the final Section 4 we apply the results shown in the preceding section to find out the structure of certain subcategories, and give several results including Corollaries 1.2, 1.3 and 1.4.

## 2. Basic definitions

This section is devoted to stating the definitions and basic properties of notions which we will *freely* use in the later sections. We begin with our convention.

**Convention 2.1.** Throughout the present paper, let $R$ be a commutative noetherian ring with identity. We assume that all $R$-modules are finitely generated, that all $R$-complexes are homologically bounded, and that all subcategories of categories are full.

In what follows, $\mathcal{T}$ and $\mathcal{A}$ denote a triangulated category and an abelian category with enough projective objects, respectively.

**Definition 2.2.** (1) For a subcategory $\mathcal{X}$ of an additive category $\mathcal{C}$, the *additive closure* $\mathsf{add}_{\mathcal{C}}\,\mathcal{X}$ of $\mathcal{X}$ is defined to be the smallest subcategory of $\mathcal{C}$ containing $\mathcal{X}$ and closed under finite direct sums and direct summands.

(2) A *Serre subcategory* of $\mathcal{A}$ is defined to be a subcategory of $\mathcal{A}$ closed under subobjects, quotients and extensions.

(3) A *thick subcategory* of $\mathcal{T}$ is by definition a triangulated subcategory of $\mathcal{T}$ closed under direct summands. The *thick closure* of a subcategory $\mathcal{X}$ of $\mathcal{T}$ is defined as the smallest thick subcategory of $\mathcal{T}$ containing $\mathcal{X}$, and denoted by $\mathsf{thick}_{\mathcal{T}}\,\mathcal{X}$ or simply by $\mathsf{thick}\,\mathcal{X}$. When $\mathcal{X}$ consists of a single object $M$, we denote it by $\mathsf{thick}_{\mathcal{T}}\,M$ or $\mathsf{thick}\,M$.

(4) We denote by $\mathsf{proj}\,\mathcal{A}$ the subcategory of $\mathcal{A}$ consisting of projective objects.

(5) Let $P = (\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \to 0)$ be a projective resolution of $M \in \mathcal{A}$. Then for each $n > 0$ we define the *$n$-th syzygy* $\Omega^n M$ of $M$ (with respect to $P$) as the image of $d_n$. This is uniquely determined up to projective summands.

(6) We define a *resolving subcategory* of $\mathcal{A}$ as a subcategory of $\mathcal{A}$ containing $\mathsf{proj}\,\mathcal{A}$ and closed under direct summands, extensions and syzygies. The *resolving closure* of a subcategory $\mathcal{X}$ of $\mathcal{A}$ is by definition the smallest resolving subcategory of $\mathcal{A}$ containing $\mathcal{X}$, and denoted by $\mathsf{res}_{\mathcal{A}}\,\mathcal{X}$ or simply by $\mathsf{res}\,\mathcal{X}$. When $\mathcal{X}$ consists of a single object $M$, we denote it by $\mathsf{res}_{\mathcal{A}}\,M$ or $\mathsf{res}\,M$.

(7) Let $X, Y$ be complexes of objects of $\mathcal{A}$.

(a) A homomorphism $f : X \to Y$ of complexes is called a *quasiisomorphism* if the induced map $\mathsf{H}_i(f) : \mathsf{H}_i(X) \to \mathsf{H}_i(Y)$ on the $i$-th homologies is an isomorphism for all integers $i$.

(b) We say that $X$ is *equivalent* to $Y$ if there exists a sequence $X^0, X^1, \dots, X^n$ of complexes such that $X^0 = X$, $X^n = Y$, and there is a quasiisomorphism between $X^i$ and $X^{i+1}$ for all $0 \le i \le n-1$. Then we write $X \simeq Y$.

**Remark 2.3.** (1) A Serre subcategory is defined for an arbitrary abelian category.

(2) A resolving subcategory is usually defined as a subcategory containing the projective objects and closed under direct summands, extensions and kernels of epimorphisms. This definition and ours are equivalent.

(3) Let $\mathscr{X}$ be a resolving subcategory of $\mathscr{A}$. Let $M$ be an object of $\mathscr{X}$ and $n > 0$ an integer. The $n$-th syzygy of $M$ with respect to *some* projective resolution of $M$ is in $\mathscr{X}$ if and only if the $n$-th syzygy of $M$ with respect to *every* projective resolution of $M$ is in $\mathscr{X}$.

We recall the notions of balls in $\mathscr{T}$ and $\mathscr{A}$ introduced in [Bondal and Van den Bergh 2003; Dao and Takahashi 2011; Rouquier 2008].

**Definition 2.4.** (1a) For a subcategory $\mathscr{X}$ of $\mathscr{T}$ we denote by $\langle\mathscr{X}\rangle$ the smallest subcategory of $\mathscr{T}$ containing $\mathscr{X}$ that is closed under finite direct sums, direct summands and shifts; in symbols, $\langle\mathscr{X}\rangle = \mathsf{add}_{\mathscr{T}}\{X[i] \mid i \in \mathbb{Z},\ X \in \mathscr{X}\}$. When $\mathscr{X}$ consists of a single object $M$, we simply denote it by $\langle M\rangle$.

(1b) For subcategories $\mathscr{X}, \mathscr{Y}$ of $\mathscr{T}$ we denote by $\mathscr{X}*\mathscr{Y}$ the subcategory of $\mathscr{T}$ consisting of objects $M$ which fits into an exact triangle $X \to M \to Y \rightsquigarrow$ in $\mathscr{T}$ with $X \in \mathscr{X}$ and $Y \in \mathscr{Y}$. We set $\mathscr{X} \diamond \mathscr{Y} = \langle\langle\mathscr{X}\rangle * \langle\mathscr{Y}\rangle\rangle$.

(1c) Let $\mathscr{C}$ be a subcategory of $\mathscr{T}$. We define the *ball of radius $r$ centered at* $\mathscr{C}$ as

$$\langle\mathscr{C}\rangle_r = \begin{cases} \langle\mathscr{C}\rangle & (r = 1), \\ \langle\mathscr{C}\rangle_{r-1} \diamond \mathscr{C} = \langle\langle\mathscr{C}\rangle_{r-1} * \langle\mathscr{C}\rangle\rangle & (r \geq 2). \end{cases}$$

If $\mathscr{C}$ consists of a single object $M$, then we simply denote it by $\langle M\rangle_r$. We write $\langle\mathscr{C}\rangle_r^{\mathscr{T}}$ when we should specify that $\mathscr{T}$ is the ground category where the ball is defined.

(2a) For a subcategory $\mathscr{X}$ of $\mathscr{A}$ we denote by $[\mathscr{X}]$ the smallest subcategory of $\mathscr{A}$ containing $\mathsf{proj}\,\mathscr{A}$ and $\mathscr{X}$ that is closed under finite direct sums, direct summands and syzygies, that is, $[\mathscr{X}] = \mathsf{add}_{\mathscr{A}}(\mathsf{proj}\,\mathscr{A} \cup \{\, \Omega^i X \mid i \geq 0,\ X \in \mathscr{X}\,\})$. When $\mathscr{X}$ consists of a single object $M$, we simply denote it by $[M]$.

(2b) For subcategories $\mathscr{X}, \mathscr{Y}$ of $\mathscr{A}$ we denote by $\mathscr{X}\circ\mathscr{Y}$ the subcategory of $\mathscr{A}$ consisting of objects $M$ which fits into an exact sequence $0 \to X \to M \to Y \to 0$ in $\mathscr{A}$ with $X \in \mathscr{X}$ and $Y \in \mathscr{Y}$. We set $\mathscr{X} \bullet \mathscr{Y} = [[\mathscr{X}] \circ [\mathscr{Y}]]$.

(2c) Let $\mathscr{C}$ be a subcategory of $\mathscr{A}$. We define the *ball of radius $r$ centered at* $\mathscr{C}$ as

$$[\mathscr{C}]_r = \begin{cases} [\mathscr{C}] & (r = 1), \\ [\mathscr{C}]_{r-1} \bullet \mathscr{C} = [[\mathscr{C}]_{r-1} \circ [\mathscr{C}]] & (r \geq 2). \end{cases}$$

If $\mathscr{C}$ consists of a single object $M$, then we simply denote it by $[M]_r$. We write $[\mathscr{C}]_r^{\mathscr{A}}$ when we should specify that $\mathscr{A}$ is the ground category where the ball is defined.

**Remark 2.5** [Bondal and Van den Bergh 2003; Dao and Takahashi 2011; Rouquier 2008]. (1) Let $\mathscr{X}, \mathscr{Y}, \mathscr{Z}, \mathscr{C}$ be subcategories of $\mathscr{T}$.

(a) An object $M \in \mathscr{T}$ belongs to $\mathscr{X} \diamond \mathscr{Y}$ if and only if there is an exact triangle $X \to Z \to Y \rightsquigarrow$ with $X \in \langle\mathscr{X}\rangle$, $Y \in \langle\mathscr{Y}\rangle$, and $M$ a direct summand of $Z$.

(b) One has $(\mathscr{X} \diamond \mathscr{Y}) \diamond \mathscr{Z} = \mathscr{X} \diamond (\mathscr{Y} \diamond \mathscr{Z})$ and $\langle\mathscr{C}\rangle_a \diamond \langle\mathscr{C}\rangle_b = \langle\mathscr{C}\rangle_{a+b}$ for all $a, b > 0$.

(2) Let $\mathscr{X}, \mathscr{Y}, \mathscr{Z}, \mathscr{C}$ be subcategories of $\mathscr{A}$.

(a) An object $M \in \mathscr{A}$ belongs to $\mathscr{X} \bullet \mathscr{Y}$ if and only if there is an exact sequence $0 \to X \to Z \to Y \to 0$ with $X \in [\mathscr{X}]$ and $Y \in [\mathscr{Y}]$ such that $M$ is a direct summand of $Z$.

(b) One has $(\mathscr{X} \bullet \mathscr{Y}) \bullet \mathscr{Z} = \mathscr{X} \bullet (\mathscr{Y} \bullet \mathscr{Z})$ and $[\mathscr{C}]_a \bullet [\mathscr{C}]_b = [\mathscr{C}]_{a+b}$ for all $a, b > 0$.

**Definition 2.6.** An $R$-complex is called *perfect* if it is a bounded complex of projective $R$-modules. The *singularity category* $\mathbf{D}_{\mathrm{sg}}(R)$ of $R$ is defined as the Verdier quotient of $\mathbf{D}_{\mathrm{b}}(R)$ by the perfect complexes. For the definition of a Verdier quotient, we refer to [Neeman 2001, Remark 2.1.9]. Whenever we discuss the singularity category $\mathbf{D}_{\mathrm{sg}}(R)$, we identify each object or subcategory of $\operatorname{mod} R$ with its image in $\mathbf{D}_{\mathrm{sg}}(R)$ by the composition of the canonical functors $\operatorname{mod} R \to \mathbf{D}_{\mathrm{b}}(R) \to \mathbf{D}_{\mathrm{sg}}(R)$.

**Remark 2.7** [Dao and Takahashi 2012b, Lemma 2.4]. (1) For all $X \in \mathbf{D}_{\mathrm{b}}(R)$ there exists an exact triangle $P \to X \to M[n] \rightsquigarrow$ in $\mathbf{D}_{\mathrm{b}}(R)$ such that $P$ is a perfect complex, $M$ is a module and $n$ is an integer. In particular, $X \cong M[n]$ in $\mathbf{D}_{\mathrm{sg}}(R)$.

(2) For every $M \in \operatorname{mod} R$ and every $n \geq 0$ there is an isomorphism $M \cong \Omega^n M[n]$ in $\mathbf{D}_{\mathrm{sg}}(R)$. Hence, for a subcategory $\mathscr{C}$ of $\operatorname{mod} R$ and an integer $k > 0$, each module in $[\mathscr{C}]_k^{\operatorname{mod} R}$ belongs to $\langle \mathscr{C} \rangle_k^{\mathbf{D}_{\mathrm{sg}}(R)}$.

We introduce subcategories which will be investigated in Section 4.

**Definition 2.8.** Let $\Phi$ be a subset of $\operatorname{Spec} R$. Set $\Phi^{\mathrm{c}} = \operatorname{Spec} R \setminus \Phi$. We denote by $\mathsf{e}^{\Phi}(R)$ (respectively, $\operatorname{mod}^{\Phi}(R)$) the subcategory of $\operatorname{mod} R$ consisting of $R$-modules $M$ such that $M_{\mathfrak{p}} = 0$ (respectively, $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free) for all $\mathfrak{p} \in \Phi^{\mathrm{c}}$. Also, $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$ (respectively, $\mathbf{D}_{\mathrm{sg}}^{\Phi}(R)$) denotes the subcategory of $\mathbf{D}_{\mathrm{b}}(R)$ (respectively, $\mathbf{D}_{\mathrm{sg}}(R)$) consisting of $R$-complexes $X$ such that $X_{\mathfrak{p}}$ isomorphic to a perfect $R_{\mathfrak{p}}$-complex in $\mathbf{D}_{\mathrm{b}}(R_{\mathfrak{p}})$ (respectively, $X_{\mathfrak{p}} \cong 0$ in $\mathbf{D}_{\mathrm{sg}}(R_{\mathfrak{p}})$) for all $\mathfrak{p} \in \Phi^{\mathrm{c}}$. We have that $\mathsf{e}^{\Phi}(R)$ is a Serre subcategory of $\operatorname{mod} R$, that $\operatorname{mod}^{\Phi}(R)$ is a resolving subcategory of $\operatorname{mod} R$, and that $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$, $\mathbf{D}_{\mathrm{sg}}^{\Phi}(R)$ are thick subcategories of $\mathbf{D}_{\mathrm{b}}(R)$, $\mathbf{D}_{\mathrm{sg}}(R)$ respectively.

**Definition 2.9.** (1) For an $R$-module $M$ we denote by $\mathsf{NF}(M)$ the *nonfree locus* of $M$, that is, the set of prime ideals $\mathfrak{p}$ of $R$ such that the $R_{\mathfrak{p}}$-module $M_{\mathfrak{p}}$ is nonfree. As is well-known, $\mathsf{NF}(M)$ is a closed subset of $\operatorname{Spec} R$ in the Zariski topology.

(2) For an $R$-complex $M$ we denote by $\mathsf{IPD}(M)$ the *infinite projective dimension locus* of $M$, that is, the set of prime ideals $\mathfrak{p}$ of $R$ such that the $R_{\mathfrak{p}}$-complex $M_{\mathfrak{p}}$ has infinite projective dimension.

(3) For a subcategory $\mathscr{X}$ of $\operatorname{mod} R$ we set $\operatorname{Supp} \mathscr{X} = \bigcup_{M \in \mathscr{X}} \operatorname{Supp} M$ and $\mathsf{NF}(\mathscr{X}) = \bigcup_{M \in \mathscr{X}} \mathsf{NF}(M)$.

(4) For a subcategory $\mathscr{X}$ of $\mathbf{D}_{\mathrm{b}}(R)$ we set $\mathsf{IPD}(\mathscr{X}) = \bigcup_{M \in \mathscr{X}} \mathsf{IPD}(M)$.

(5) For a subcategory $\mathscr{X}$ of $\mathbf{D}_{\mathrm{sg}}(R)$ we set $\operatorname{Supp}_{\mathrm{sg}}(\mathscr{X}) = \bigcup_{M \in \mathscr{X}} \mathsf{IPD}(M)$.

**Definition 2.10.** (1) Let $M$ be an $R$-module.

(a) Let $\boldsymbol{x}$ be a sequence of elements of $R$. Then $\mathsf{K}(\boldsymbol{x}, M)$ denotes the *Koszul complex* of $M$ with respect to $\boldsymbol{x}$. We call $\mathsf{H}_i(\boldsymbol{x}, M) := \mathsf{H}_i(\mathsf{K}(\boldsymbol{x}, M))$ the $i$-th *Koszul homology* ($i \in \mathbb{Z}$) and $\mathsf{H}(\boldsymbol{x}, M) := \bigoplus_{i\in\mathbb{Z}} \mathsf{H}_i(\boldsymbol{x}, M)$ the *Koszul homology* of $M$ with respect to $\boldsymbol{x}$.

(b) Let $P_1 \xrightarrow{d} P_0 \to M \to 0$ be a projective presentation of $M$. Then the cokernel of the $R$-dual map of $d$ is called the *transpose* of $M$ and denoted by $\mathsf{Tr}\, M$. This is uniquely determined up to projective summands.

(2) A subset $\Phi$ of $\mathsf{Spec}\, R$ is called *specialization-closed* if $\mathsf{V}(\mathfrak{p}) \subseteq \Phi$ for all $\mathfrak{p} \in \Phi$. This is nothing but a union of closed subsets of $\mathsf{Spec}\, R$ in the Zariski topology.

(3) We denote by $\mathsf{Sing}\, R$ the *singular locus* of $R$, namely, the set of prime ideals $\mathfrak{p}$ of $R$ such that $R_{\mathfrak{p}}$ is not a regular local ring.

(4) A local ring $R$ with maximal ideal $\mathfrak{m}$ is called an *isolated singularity* if $\mathsf{Sing}\, R \subseteq \{\mathfrak{m}\}$.

## 3. Reconstruction from Koszul homology

In this section, we consider reconstructing a given module from its Koszul homology by taking direct summands, extensions and syzygies. We start by stating and proving the most general result in this paper; actually, almost all of the other results given in this paper are deduced from this.

**Theorem 3.1.** *Let $M$ be an $R$-module. Let $\boldsymbol{x} = x_1, \dots, x_n$ be a sequence of elements of $R$ such that $x_p \operatorname{Ext}_R^q(M, \Omega^r M) = 0$ for all $1 \le p \le n$ and $1 \le q, r \le p$. Let $P$ be a projective resolution of $M$. Then $\mathsf{K}(\boldsymbol{x}, M)$ is equivalent to a complex*

$$X = (0 \to X_n \to X_{n-1} \to \cdots \to X_1 \to X_0 \to 0)$$

*such that $X_i = \bigoplus_{j=0}^{i} P_j^{\oplus\binom{n}{i-j}}$ for each $0 \le i \le n-1$ and $X_n = \bigoplus_{j=0}^{n} (\Omega^j M)^{\oplus\binom{n}{j}}$.*

*Proof.* We prove the theorem by induction on $n$. Let us first consider the case where $n = 1$. Multiplication by $x_1$ makes a pullback diagram:

$$\begin{tikzcd}
\sigma: & 0 \arrow[r] & \Omega M \arrow[r] & P_0 \arrow[r] & M \arrow[r] & 0 \\
x_1\sigma: & 0 \arrow[r] & \Omega M \arrow[r] \arrow[u, equal] & N \arrow[r] \arrow[u] & M \arrow[r] \arrow[u, "x_1"] & 0.
\end{tikzcd}$$

Since $x_1 \operatorname{Ext}_R^1(M, \Omega M) = 0$, we see that the exact sequence $x_1\sigma$ splits and get an isomorphism $N \cong \Omega M \oplus M$. Thus we obtain a short exact sequence of complexes

$$0 \to W \to X \to \mathsf{K}(x_1, M) \to 0,$$

where $W=(0\to\Omega M\xrightarrow{=}\Omega M\to 0)$ and $X=(0\to\Omega M\oplus M\to P_0\to 0)$. As $W$ is acyclic, $\mathsf{K}(x_1,M)$ is equivalent to $X$.

Next we assume $n\geq 2$. The induction hypothesis implies that $\mathsf{K}(x_1,\dots,x_{n-1},M)$ is equivalent to a complex

$$Y=(0\to Y_{n-1}\xrightarrow{f}Y_{n-2}\to\cdots\to Y_1\to Y_0\to 0)$$

with $Y_i=\bigoplus_{j=0}^{i}P_j^{\oplus\binom{n-1}{i-j}}$ for $0\leq i\leq n-2$ and $Y_{n-1}=\bigoplus_{j=0}^{n-1}(\Omega^jM)^{\oplus\binom{n-1}{j}}$. In general, taking a tensor product with a perfect complex preserves equivalence of complexes (cf. [Christensen 2000, A.4.1]). Hence we have

$$\begin{aligned}
&\mathsf{K}(\boldsymbol{x},M)\\
&\quad=\mathsf{K}(x_1,\dots,x_{n-1},M)\otimes_R\mathsf{K}(x_n,R)\simeq Y\otimes_R\mathsf{K}(x_n,R)\\
&\quad=(0\to Y_{n-1}\xrightarrow{g}Y_{n-1}\oplus Y_{n-2}\xrightarrow{d_{n-1}}Y_{n-2}\oplus Y_{n-3}\xrightarrow{d_{n-2}}\cdots\xrightarrow{d_2}Y_1\oplus Y_0\xrightarrow{d_1}Y_0\to 0)\\
&\quad=:Z,
\end{aligned}$$

where $g=\binom{(-1)^{n-1}x_n}{f}$. Note that there is an exact sequence $0\to\Omega Y_{n-1}\to Q\xrightarrow{\pi}Y_{n-1}\to 0$ with $Q=\bigoplus_{j=0}^{n-1}P_j^{\oplus\binom{n-1}{j}}$. Consider the pullback diagram

$$\begin{array}{lccccccccc}
\tau: & 0 & \longrightarrow & \Omega Y_{n-1} & \longrightarrow & Q\oplus Y_{n-2} & \xrightarrow{h} & Y_{n-1}\oplus Y_{n-2} & \longrightarrow & 0\\
 & & & \| & & \uparrow & & g\uparrow & & \\
g^*(\tau): & 0 & \longrightarrow & \Omega Y_{n-1} & \longrightarrow & L & \longrightarrow & Y_{n-1} & \longrightarrow & 0,
\end{array}$$

where $h=\begin{pmatrix}\pi&0\\0&1\end{pmatrix}$ and $g^*=\mathsf{Ext}_R^1(g,\Omega Y_{n-1})$. As $Y_{n-2}$ is projective, the map $g^*$ can be identified with the multiplication map

$$\mathsf{Ext}_R^1(Y_{n-1},\Omega Y_{n-1})\xrightarrow{(-1)^{n-1}x_n}\mathsf{Ext}_R^1(Y_{n-1},\Omega Y_{n-1}).$$

There are isomorphisms

$$\begin{aligned}
\mathsf{Ext}_R^1(Y_{n-1},\Omega Y_{n-1})&\cong\bigoplus_{j,k=0}^{n-1}\mathsf{Ext}_R^1(\Omega^jM,\Omega(\Omega^kM))^{\oplus\left(\binom{n-1}{j}+\binom{n-1}{k}\right)}\\
&\cong\bigoplus_{j,k=0}^{n-1}\mathsf{Ext}_R^{j+1}(M,\Omega^{k+1}M)^{\oplus\left(\binom{n-1}{j}+\binom{n-1}{k}\right)},
\end{aligned}$$

and hence $x_n$ annihilates $\mathsf{Ext}_R^1(Y_{n-1},\Omega Y_{n-1})$. Therefore $g^*(\tau)$ is a split exact sequence, and we obtain a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \Omega Y_{n-1} & \longrightarrow & Q\oplus Y_{n-2} & \xrightarrow{h} & Y_{n-1}\oplus Y_{n-2} & \longrightarrow & 0\\
 & & \| & & l\uparrow & & g\uparrow & & \\
0 & \longrightarrow & \Omega Y_{n-1} & \longrightarrow & \Omega Y_{n-1}\oplus Y_{n-1} & \longrightarrow & Y_{n-1} & \longrightarrow & 0,
\end{array}$$

with exact rows. We observe that the complex $Z$ is equivalent to the complex

$$X = (0 \to \Omega Y_{n-1} \oplus Y_{n-1} \xrightarrow{l} Q \oplus Y_{n-2} \xrightarrow{d_{n-1}h} Y_{n-2} \oplus Y_{n-3} \xrightarrow{d_{n-2}} \cdots \xrightarrow{d_2} Y_1 \oplus Y_0 \xrightarrow{d_1} Y_0 \to 0).$$

There are equalities

$$\Omega Y_{n-1} \oplus Y_{n-1} = \bigoplus_{j=0}^{n} (\Omega^j M)^{\oplus \binom{n}{j}},$$
$$Q \oplus Y_{n-2} = \bigoplus_{j=0}^{n-1} P_j^{\oplus \binom{n}{(n-1)-j}},$$
$$Y_i \oplus Y_{i-1} = \bigoplus_{j=0}^{i} P_j^{\oplus \binom{n}{i-j}}$$

for $1 \le i \le n-2$ and $Y_0 = P_0$. Thus we are done. □

Using Theorem 3.1, we obtain the following corollary.

**Corollary 3.2.** *Let $M$ and $\boldsymbol{x}$ be as in Theorem 3.1.*

(1) *If $\boldsymbol{x}$ is a regular sequence on $M$, then $\Omega^n(M/\boldsymbol{x}M) \cong \bigoplus_{k=0}^{n} (\Omega^k M)^{\oplus \binom{n}{k}}$ in* $\mathsf{mod}\, R$.

(2) *For each $1 \le i \le n$ there exists an exact sequence of $R$-modules*

$$0 \to \mathsf{H}_i(\boldsymbol{x}, M) \to E_i \to \Omega E_{i-1} \to 0$$

*with $E_0 = \mathsf{H}_0(\boldsymbol{x}, M)$ such that $M$ is a direct summand of $E_n$. Hence $M$ is built out of $\mathsf{H}_0(\boldsymbol{x}, M), \dots, \mathsf{H}_n(\boldsymbol{x}, M)$ by taking $n$ syzygies, $n$ extensions and $1$ direct summand. In particular, $M$ belongs to the ball $[\mathsf{H}(\boldsymbol{x}, M)]_{n+1}^{\mathsf{mod}\, R}$.*

(3) *There is an exact triangle*

$$F \to \mathsf{K}(\boldsymbol{x}, M) \to \bigoplus_{k=0}^{n} (\Omega^k M)^{\oplus \binom{n}{k}}[n] \rightsquigarrow$$

*in $\mathbf{D}_{\mathrm{b}}(R)$, where $F = (0 \to F_{n-1} \to \cdots \to F_0 \to 0)$ is a perfect complex.*

(4) *The module $M$ belongs to the ball $\langle R \oplus \mathsf{K}(\boldsymbol{x}, M) \rangle_{n+1}^{\mathbf{D}_{\mathrm{b}}(R)}$.*

(5) *One has $\mathsf{K}(\boldsymbol{x}, M) \cong \bigoplus_{k=0}^{n} M^{\oplus \binom{n}{k}}[k]$ in $\mathbf{D}_{\mathrm{sg}}(R)$. In particular, $M$ is a direct summand of $\mathsf{K}(\boldsymbol{x}, M)$ in $\mathbf{D}_{\mathrm{sg}}(R)$.*

*Proof.* We use the notation of Theorem 3.1 and its assertion.

(1) Since $\boldsymbol{x}$ is regular on $M$, we have an equivalence $\mathsf{K}(\boldsymbol{x}, M) \simeq M/\boldsymbol{x}M$. There is an exact sequence

$$0 \to X_n \to X_{n-1} \to \cdots \to X_0 \to M/\boldsymbol{x}M \to 0$$

of $R$-modules. As $X_n = \bigoplus_{j=0}^{n} (\Omega^j M)^{\oplus \binom{n}{j}}$ and $X_i$ is projective for all $0 \le i \le n-1$, the module $X_n$ is the $n$-th syzygy of $M/\boldsymbol{x}M$ as an $R$-module.

(2) For each $0 \le i \le n$ take a truncation $X^i = (0 \to X_n \to \cdots \to X_{i+1} \to X_i \to 0)$ of $X$ with $(X^i)_j = X_{i+j}$ for $0 \le j \le n$. Then there is a short exact sequence

$$0 \to X_{i-1} \to X^{i-1} \to X^i[1] \to 0$$

of complexes for each $1 \le i \le n$. The long exact sequence in homology gives an exact sequence $0 \to \mathsf{H}_1(X^{i-1}) \to \mathsf{H}_0(X^i) \to X_{i-1} \to \mathsf{H}_0(X^{i-1}) \to 0$ of modules. As $X_{i-1}$ is projective, we have an exact sequence

$$0 \to \mathsf{H}_1(X^{i-1}) \to \mathsf{H}_0(X^i) \to \Omega\mathsf{H}_0(X^{i-1}) \to 0$$

for all $1 \le i \le n$. Notice $\mathsf{H}_1(X^{i-1}) = \mathsf{H}_i(\boldsymbol{x}, M)$, $\mathsf{H}_0(X^0) = \mathsf{H}_0(\boldsymbol{x}, M)$ and $\mathsf{H}_0(X^n) = X_n$. Setting $E_i = \mathsf{H}_0(X^i)$ for $0 \le i \le n$, we obtain desired exact sequences.

(3) Truncating the complex $X$ provides such an exact triangle.

(4) Decomposing $F$ into short exact sequences of complexes, we observe that $F$ is in $\langle R \rangle_n^{\mathbf{D}_{\mathrm{b}}(R)}$. As $M$ is a direct summand of $\bigoplus_{k=0}^n (\Omega^k M)^{\oplus \binom{n}{k}}$, the assertion follows from (3).

(5) By (3) we have an isomorphism $\mathsf{K}(\boldsymbol{x}, M) \cong \bigoplus_{k=0}^n (\Omega^k M)^{\oplus \binom{n}{k}}[n]$ in $\mathbf{D}_{\mathrm{sg}}(R)$. Since $M \cong \Omega^k M[k]$ in $\mathbf{D}_{\mathrm{sg}}(R)$, we are done. □

**Remark 3.3.** (1) Corollary 3.2(1) is a refinement of [Takahashi 2010, Proposition 2.2], which shows the same conclusion under the additional assumption that $\boldsymbol{x}$ is a regular sequence on $R$ annihilating more Ext modules.

(2) Corollary 3.2(5) can also be shown by using the proof of [Dao and Takahashi 2012b, Proposition 2.3]. It also implies that $M$ belongs to $\langle R \oplus \mathsf{K}(\boldsymbol{x}, M) \rangle_m^{\mathbf{D}_{\mathrm{b}}(R)}$ for *some* integer $m > 0$. However, it cannot determine how big/small $m$ is, while Corollary 3.2(4) can.

We are interested in existence of a sequence $\boldsymbol{x}$ as in Theorem 3.1. The lemma below guarantees that such a sequence always exists. Moreover, one can make such a sequence as a power of an arbitrary sequence whose defining closed subset covers the nonfree locus.

**Lemma 3.4.** *Let $M$ be an $R$-module. Let $\boldsymbol{x} = x_1, \ldots, x_n$ be a sequence of elements of $R$ with $\mathsf{NF}(M) \subseteq \mathsf{V}(\boldsymbol{x})$. Then there exists an integer $k > 0$ such that the sequence $\boldsymbol{x}^k = x_1^k, \ldots, x_n^k$ annihilates $\mathsf{Ext}_R^i(M, N)$ for all $i > 0$ and all $N \in \mathsf{mod}\, R$.*

*Proof.* Let $I$ be an ideal of $R$ with $\mathsf{NF}(M) = \mathsf{V}(I)$. Then by [Dao and Takahashi 2012a, Remark 5.2(1)] there exists an integer $p > 0$ such that $I^p \mathsf{Ext}_R^i(M, N) = 0$ for all $i > 0$ and all $N \in \mathsf{mod}\, R$. By assumption, we have $(\boldsymbol{x}^q) \subseteq I$ for some $q > 0$. Setting $k = pq$ completes the proof. □

Combining Theorem 3.1, Corollary 3.2(2) and Lemma 3.4, we immediately obtain the following result, which includes Theorem 1.1.

**Corollary 3.5.** *Let $M$ be an $R$-module. Let $\boldsymbol{x} = x_1, \dots, x_n$ be a sequence of elements of $R$ with $\mathsf{NF}(M) \subseteq \mathsf{V}(\boldsymbol{x})$. Then there exists an integer $k > 0$ such that $\mathsf{K}(\boldsymbol{x}^k, M)$ is equivalent to a complex*

$$(0 \to N \to P_{n-1} \to \cdots \to P_0 \to 0),$$

*where each $P_i$ is projective and $M$ is a direct summand of $N$. Hence, $M$ is built out of $\mathsf{H}_0(\boldsymbol{x}^k, M), \dots, \mathsf{H}_n(\boldsymbol{x}^k, M)$ by taking $n$ syzygies, $n$ extensions and 1 direct summand. In particular, $M$ is in $[\mathsf{H}(\boldsymbol{x}^k, M)]_{n+1}^{\mathsf{mod}\, R}$.*

## 4. Generation of subcategories

In this section, we apply our results obtained in the previous section to investigate generation of subcategories. To be precise, for a subset $\Phi$ of $\mathsf{Spec}\, R$ we analyze the structure of the subcategories $\mathsf{mod}^{\Phi}(R)$, $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$ and $\mathbf{D}_{\mathrm{sg}}^{\Phi}(R)$. We also consider classification of these subcategories.

First of all, we want to make a generator of $\mathsf{mod}^{\Phi}(R)$ as a resolving subcategory of $\mathsf{mod}\, R$ and generators of $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$, $\mathbf{D}_{\mathrm{sg}}^{\Phi}(R)$ as thick subcategories of $\mathbf{D}_{\mathrm{b}}(R)$, $\mathbf{D}_{\mathrm{sg}}(R)$. In fact, $\mathsf{e}^{\Phi}(R)$ gives generators of these three subcategories:

**Theorem 4.1.** *Let $\Phi$ be a subset of $\mathsf{Spec}\, R$. Then one has equalities*

$$\mathsf{mod}^{\Phi}(R) = \mathsf{res}_{\mathsf{mod}\, R}(\mathsf{e}^{\Phi}(R)), \tag{1}$$

$$\mathbf{D}_{\mathrm{b}}^{\Phi}(R) = \mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}(\{R\} \cup \mathsf{e}^{\Phi}(R)), \tag{2}$$

$$\mathbf{D}_{\mathrm{sg}}^{\Phi}(R) = \mathsf{thick}_{\mathbf{D}_{\mathrm{sg}}(R)}(\mathsf{e}^{\Phi}(R)). \tag{3}$$

*Proof.* (1) It is obvious that $\mathsf{e}^{\Phi}(R)$ is contained in $\mathsf{mod}^{\Phi}(R)$, and hence so is its resolving closure. To show the opposite inclusion, let $M$ be an object of $\mathsf{mod}^{\Phi}(R)$. Then by definition $\mathsf{NF}(M)$ is contained in $\Phi$. It is seen from Corollary 3.5 that there is a sequence $\boldsymbol{x} = x_1, \dots, x_n$ of elements of $R$ with $\mathsf{NF}(M) = \mathsf{V}(\boldsymbol{x})$ such that $M$ belongs to $\mathsf{res}_{\mathsf{mod}\, R}\, \mathsf{H}(\boldsymbol{x}, M)$. Since $\mathsf{H}(\boldsymbol{x}, M)$ is annihilated by $\boldsymbol{x}$, we have

$$\mathsf{Supp}\, \mathsf{H}(\boldsymbol{x}, M) \subseteq \mathsf{V}(\boldsymbol{x}) = \mathsf{NF}(M) \subseteq \Phi,$$

which shows $\mathsf{H}(\boldsymbol{x}, M) \in \mathsf{e}^{\Phi}(R)$. Consequently, $M$ is in $\mathsf{res}_{\mathsf{mod}\, R}(\mathsf{e}^{\Phi}(R))$.

(2) Clearly, $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$ contains $R$ and $\mathsf{e}^{\Phi}(R)$, and the thick closure of $\{R\} \cup \mathsf{e}^{\Phi}(R)$. Let $X$ be an object of $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$. Then there is an exact triangle

$$P \to X \to M[n] \rightsquigarrow$$

in $\mathbf{D}_{\mathrm{b}}(R)$ such that $P$ is a perfect $R$-complex, $M$ is an $R$-module and $n$ is an integer. We use the *large restricted flat dimension* of $M$, namely

$$\mathsf{Rfd}_R\, M = \sup_{\mathfrak{p} \in \mathsf{Spec}\, R} \{\mathsf{depth}\, R_{\mathfrak{p}} - \mathsf{depth}_{R_{\mathfrak{p}}} M_{\mathfrak{p}}\}$$

By [Avramov et al. 2010b, Theorem 1.1] this is finite. Put $r = \mathsf{Rfd}_R\, M$. Let $\mathfrak{p}$ be a prime ideal in $\Phi^{\mathrm{c}}$. Localizing the above exact triangle at $\mathfrak{p}$, we see that the $R_{\mathfrak{p}}$-module $M_{\mathfrak{p}}$ has finite projective dimension. Hence

$$\mathsf{pd}_{R_{\mathfrak{p}}} M_{\mathfrak{p}} = \mathsf{depth}\, R_{\mathfrak{p}} - \mathsf{depth}_{R_{\mathfrak{p}}} M_{\mathfrak{p}} \le r.$$

Setting $N = \Omega^r M$, we note that $N$ belongs to $\mathsf{mod}^{\Phi}(R)$, hence to $\mathsf{res}_{\mathsf{mod}\, R}(\mathsf{e}^{\Phi}(R))$ by (1). Therefore $N$ is in $\mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}(\{R\} \cup \mathsf{e}^{\Phi}(R))$, and so is $M$. As $P \in \mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}\, R$, the object $X$ belongs to $\mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}(\{R\} \cup \mathsf{e}^{\Phi}(R))$ by the above exact triangle.

(3) This equality is obtained by using (2). □

One can describe the structure of $\mathsf{e}^{\Phi}(R)$ in more detail, which makes more visible representations of $\mathsf{mod}^{\Phi}(R)$, $\mathbf{D}_{\mathrm{b}}^{\Phi}(R)$ and $\mathbf{D}_{\mathrm{sg}}^{\Phi}(R)$.

**Corollary 4.2.** *Let $\Phi$ be a subset of $\mathsf{Spec}\, R$. Then $\mathsf{e}^{\Phi}(R)$ is the smallest subcategory of $\mathsf{mod}\, R$ containing $R/\mathfrak{p}$ for all $\mathfrak{p} \in \Phi^{\mathrm{sp}}$ and closed under extensions. Here $\Phi^{\mathrm{sp}}$ denotes the largest specialization-closed subset of $\mathsf{Spec}\, R$ contained in $\Phi$. Hence*

$$\begin{aligned}
\mathsf{mod}^{\Phi}(R) &= \mathsf{res}_{\mathsf{mod}\, R}\{\, R/\mathfrak{p} \mid \mathfrak{p} \in \Phi^{\mathrm{sp}} \,\},\\
\mathbf{D}_{\mathrm{b}}^{\Phi}(R) &= \mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}\{\, R,\, R/\mathfrak{p} \mid \mathfrak{p} \in \Phi^{\mathrm{sp}} \,\},\\
\mathbf{D}_{\mathrm{sg}}^{\Phi}(R) &= \mathsf{thick}_{\mathbf{D}_{\mathrm{sg}}(R)}\{\, R/\mathfrak{p} \mid \mathfrak{p} \in \Phi^{\mathrm{sp}} \,\}.
\end{aligned}$$

*Proof.* The last assertion follows from Theorem 4.1.

We claim that $\Phi^{\mathrm{sp}} = \mathsf{Supp}(\mathsf{e}^{\Phi}(R))$ holds. Indeed, it is evident that $\mathsf{Supp}(\mathsf{e}^{\Phi}(R))$ is a specialization-closed subset of $\mathsf{Spec}\, R$ contained in $\Phi$. Let $\Psi$ be a specialization-closed subset of $\mathsf{Spec}\, R$ contained in $\Phi$. Then we have $\mathsf{e}^{\Psi}(R) \subseteq \mathsf{e}^{\Phi}(R)$, and hence $\Psi = \mathsf{Supp}(\mathsf{e}^{\Psi}(R)) \subseteq \mathsf{Supp}(\mathsf{e}^{\Phi}(R))$. Thus the claim holds.

Let $\mathscr{X}$ be the smallest subcategory of $\mathsf{mod}\, R$ containing $R/\mathfrak{p}$ for all $\mathfrak{p} \in \Phi^{\mathrm{sp}}$ and closed under extensions. First, let $\mathfrak{p}$ be a prime ideal in $\Phi^{\mathrm{sp}}$. As $\Phi^{\mathrm{sp}}$ is specialization-closed, we have $\mathsf{Supp}(R/\mathfrak{p}) = \mathsf{V}(\mathfrak{p}) \subseteq \Phi^{\mathrm{sp}} \subseteq \Phi$, whence $R/\mathfrak{p}$ belongs to $\mathsf{e}^{\Phi}(R)$. Since $\mathsf{e}^{\Phi}(R)$ is closed under extensions, $\mathsf{e}^{\Phi}(R)$ contains $\mathscr{X}$. Next, let $M$ be a module in $\mathsf{e}^{\Phi}(R)$. Take a filtration

$$M = M_0 \supsetneq M_1 \supsetneq \cdots \supsetneq M_n = 0$$

of submodules of $M$ such that $M_{i-1}/M_i \cong R/\mathfrak{p}_i$ with $\mathfrak{p}_i \in \mathsf{Spec}\, R$ for each $1 \le i \le n$. Then $\mathfrak{p}_i$ is in $\mathsf{Supp}\, M$, and so in $\mathsf{Supp}(\mathsf{e}^{\Phi}(R))$. By the claim, we have $\mathfrak{p}_i \in \Phi^{\mathrm{sp}}$ for all $1 \le i \le n$. Decomposing the above filtration into short exact sequences, we see that $M$ is in $\mathscr{X}$. Therefore $\mathscr{X}$ contains $\mathsf{e}^{\Phi}(R)$, and the proof is completed. □

The next result, which includes part of Corollary 1.3, follows immediately from Corollary 4.2. Note that the objects of $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$ are the $R$-modules that are locally free on the punctured spectrum of $R$.

**Corollary 4.3.** (1) $\mathbf{D}_{\mathrm{b}}(R) = \mathsf{thick}\{\, R,\, R/\mathfrak{p} \mid \mathfrak{p} \in \mathsf{Sing}\, R \,\}$.

(2) $\mathbf{D}_{\mathrm{sg}}(R) = \mathsf{thick}\{\, R/\mathfrak{p} \mid \mathfrak{p} \in \mathsf{Sing}\, R \,\}$.

(3) *If $R$ is a local ring with maximal ideal $\mathfrak{m}$ and residue field $k$, then Then* $\mathsf{mod}^{\{\mathfrak{m}\}}(R) = \mathsf{res}(k)$, $\mathbf{D}_{\mathrm{b}}^{\{\mathfrak{m}\}}(R) = \mathsf{thick}(R \oplus k)$ *and* $\mathbf{D}_{\mathrm{sg}}^{\{\mathfrak{m}\}}(R) = \mathsf{thick}(k)$.

**Remark 4.4.** The equalities in (1) and (2) can also be shown using Theorem VI.8 of [Schoutens 2003], while similar results to (3) have been obtained in Theorem 2.4 of [Takahashi 2010] as well as by H. Abe and O. Iyama (work in progress).

As a common consequence of the two assertions of Corollary 4.3, one can recover [Keller et al. 2011, Proposition A.2]:

**Corollary 4.5.** *Let $R$ be an isolated singularity with residue field $k$. Then* $\mathbf{D}_{\mathrm{b}}(R) = \mathsf{thick}(R \oplus k)$ *and* $\mathbf{D}_{\mathrm{sg}}(R) = \mathsf{thick}(k)$.

Next, we make a closer investigation on the inner structure of subcategories. In fact, we can refine the assertions as to $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$ and $\mathbf{D}_{\mathrm{sg}}^{\{\mathfrak{m}\}}(R)$ in Corollary 4.3(3) in terms of balls in the abelian category $\mathsf{mod}\, R$ and the triangulated category $\mathbf{D}_{\mathrm{sg}}(R)$. Denote by $\mathsf{fl}(R)$ the subcategory of $\mathsf{mod}\, R$ consisting of modules of finite length. The following theorem holds, which is the main part of Corollary 1.3.

**Theorem 4.6.** *Let $R$ be a $d$-dimensional local ring with maximal ideal $\mathfrak{m}$. Then there are equalities*

$$\mathsf{mod}^{\{\mathfrak{m}\}}(R) = [\mathsf{fl}(R)]_{d+1}^{\mathsf{mod}\, R} \quad \textit{and} \quad \mathbf{D}_{\mathrm{sg}}^{\{\mathfrak{m}\}}(R) = \langle \mathsf{fl}(R) \rangle_{d+1}^{\mathbf{D}_{\mathrm{sg}}(R)}.$$

*Proof.* (1) Let us show the first equality. It clearly holds when $d = 0$, so we assume $d > 0$. Let $M$ be an $R$-module in $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$. Take any system of parameters $\boldsymbol{x} = x_1, \dots, x_d$ of $R$. As $M$ is in $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$, we have $\mathsf{NF}(M) \subseteq \{\mathfrak{m}\} = \mathsf{V}(\boldsymbol{x})$. Corollary 3.5 implies that $M$ belongs to $[\mathsf{H}(\boldsymbol{x}^k, M)]_{d+1}$ for some $k > 0$. Since the $R$-module $\mathsf{H}(\boldsymbol{x}^k, M)$ is annihilated by the $\mathfrak{m}$-primary ideal $(\boldsymbol{x}^k)$, it has finite length. Thus we obtain $M \in [\mathsf{fl}(R)]_{d+1}$, and the first equality follows.

(2) We prove the second equality. Let $X$ be an $R$-complex in $\mathbf{D}_{\mathrm{sg}}^{\{\mathfrak{m}\}}(R)$. Note that $X \cong \Omega^d M[n]$ in $\mathbf{D}_{\mathrm{sg}}(R)$ for some $R$-module $M$ and some integer $n$. By the Auslander–Buchsbaum formula, we see that $\Omega^d M$ belongs to $\mathsf{mod}^{\{\mathfrak{m}\}}(R) = [\mathsf{fl}(R)]_{d+1}$. Now the second equality follows from the first one. $\square$

Here is an immediate consequence of Theorem 4.6.

**Corollary 4.7.** *If $R$ is a $d$-dimensional isolated singularity*, $\mathbf{D}_{\mathrm{sg}}(R) = \langle \mathsf{fl}(R) \rangle_{d+1}$.

**Remark 4.8.** (1) Rewording the second equality in Theorem 4.6 by the terminology introduced in [Aihara et al. 2014], one has the following inequality:

$$\mathsf{fl}(R)\text{-tri.dim}\, \mathbf{D}_{\mathrm{sg}}^{\{\mathfrak{m}\}}(R) \le \dim R.$$

(2) Theorem A in [Takahashi 2009] constructs *some* object in $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$ from *every* object in $\mathsf{mod}\, R$ and counts the number of necessary operations (containing syzygies). In contrast to this, Theorem 4.6 constructs *every* object in $\mathsf{mod}^{\{\mathfrak{m}\}}(R)$ from *some* object in $\mathsf{fl}(R)$ and counts the number of necessary operations.

(3) Similar equalities to the first equality in Theorem 4.6 are given for $\mathsf{mod}\, R$ in [Schoutens 2003, Theorem VI.8] and [Burke et al. 2012, Theorem 2], but these are different from ours in respect of how to count operations. The biggest difference is that neither of those two results counts the number of necessary extensions.

(4) In the case where $R$ is Cohen–Macaulay, Corollary 4.7 also follows from [Aihara et al. 2014, 4.5.1], because every maximal Cohen–Macaulay $R$-module is a direct summand of the $d$-th syzygy of some module of finite length by [Takahashi 2010, Proposition 2.2].

Finally, we are interested in classifying resolving and thick subcategories by using $\mathsf{mod}^{\Phi}(R)$, $\mathbf{D}^{\Phi}_{\mathrm{b}}(R)$ and $\mathbf{D}^{\Phi}_{\mathrm{sg}}(R)$. For this purpose, we prepare a lemma:

**Lemma 4.9.** (1) *The assignments* $\mathscr{X} \mapsto \mathsf{Supp}\,\mathscr{X}$ *and* $\Phi \mapsto \mathsf{e}^{\Phi}(R)$ *make a one-to-one correspondence between the Serre subcategories of* $\mathsf{mod}\, R$ *and the specialization-closed subsets of* $\mathsf{Spec}\, R$.

(2) *Let* $\Phi$ *be a specialization-closed subset of* $\mathsf{Spec}\, R$. *Then* $\mathsf{NF}(\mathsf{mod}^{\Phi}(R)) = \Phi \cap \mathsf{S}(R)$ *and* $\mathsf{IPD}(\mathbf{D}^{\Phi}_{\mathrm{b}}(R)) = \mathsf{Supp}_{\mathrm{sg}}(\mathbf{D}^{\Phi}_{\mathrm{sg}}(R)) = \Phi \cap \mathsf{Sing}\, R$.

*Proof.* (1) This is Gabriel's classification theorem [1962] for Serre subcategories.

(2) Let $\mathfrak{p} \in \Phi$. Then $\mathsf{IPD}(R/\mathfrak{p}) \subseteq \mathsf{NF}(R/\mathfrak{p}) \subseteq \mathsf{V}(\mathfrak{p}) \subseteq \Phi$. Hence $R/\mathfrak{p}$ belongs to $\mathsf{mod}^{\Phi}(R)$, $\mathbf{D}^{\Phi}_{\mathrm{b}}(R)$ and $\mathbf{D}^{\Phi}_{\mathrm{sg}}(R)$. If $\mathfrak{p} \in \mathsf{S}(R)$ (respectively, $\mathsf{Sing}\, R$), then $\mathfrak{p} \in \mathsf{NF}(R/\mathfrak{p})$ (respectively, $\mathsf{IPD}(R/\mathfrak{p})$). The assertion now follows. $\square$

We can obtain the following theorem, which includes Corollary 1.2.

**Theorem 4.10.** (1) *The assignment* $\Phi \mapsto \mathsf{mod}^{\Phi}(R)$ *is a bijection from the set of specialization-closed subsets of* $\mathsf{Spec}\, R$ *contained in* $\mathsf{S}(R)$ *to the set of resolving closures* $\mathsf{res}_{\mathsf{mod}\, R}\, \mathscr{X}$, *where* $\mathscr{X}$ *runs through the Serre subcategories of* $\mathsf{mod}\, R$.

(2) *The assignment* $\Phi \mapsto \mathbf{D}^{\Phi}_{\mathrm{b}}(R)$ *is a bijection from the set of specialization-closed subsets of* $\mathsf{Spec}\, R$ *contained in* $\mathsf{Sing}\, R$ *to the set of thick closures* $\mathsf{thick}_{\mathbf{D}_{\mathrm{b}}(R)}(\{R\} \cup \mathscr{X})$, *where* $\mathscr{X}$ *runs through the Serre subcategories of* $\mathsf{mod}\, R$.

(3) *The assignment* $\Phi \mapsto \mathbf{D}^{\Phi}_{\mathrm{sg}}(R)$ *is a bijection from the set of specialization-closed subsets of* $\mathsf{Spec}\, R$ *contained in* $\mathsf{Sing}\, R$ *to the set of thick closures* $\mathsf{thick}_{\mathbf{D}_{\mathrm{sg}}(R)}\, \mathscr{X}$, *where* $\mathscr{X}$ *runs through the Serre subcategories of* $\mathsf{mod}\, R$.

*Proof.* In view of Theorem 4.1, the three assignments make well-defined maps, and they are injective by Lemma 4.9(2). Thus it only remains to show that they are surjective.

(1) Let $\mathscr{X}$ be a Serre subcategory of $\operatorname{mod} R$. According to Lemma 4.9(1), we have $\mathscr{X} = \mathsf{e}^Z(R)$ for some specialization-closed subset $Z$ of $\operatorname{Spec} R$. Putting $\Phi = Z \cap \mathsf{S}(R)$, we easily see that $\Phi$ is a specialization-closed subset of $\operatorname{Spec} R$ which is contained in $\mathsf{S}(R)$ and satisfies $\operatorname{mod}^Z(R) = \operatorname{mod}^\Phi(R)$. Theorem 4.1 implies $\operatorname{res}_{\operatorname{mod} R} \mathscr{X} = \operatorname{mod}^\Phi(R)$.

(2), (3) We use the proof of (1). Set $\Psi = Z \cap \operatorname{Sing} R$. Then $\Psi$ is a specialization-closed subset of $\operatorname{Spec} R$ contained in $\operatorname{Sing} R$ such that the equalities $\mathbf{D}^Z_{\mathrm{b}}(R) = \mathbf{D}^\Psi_{\mathrm{b}}(R)$ and $\mathbf{D}^Z_{\mathrm{sg}}(R) = \mathbf{D}^\Psi_{\mathrm{sg}}(R)$ hold. Hence the surjectivity of the map follows from Theorem 4.1. □

The next statement subsumes Corollary 1.4 and also some earlier results: namely, (1) and the equivalence of (b)–(d) in (2) are proved in [Takahashi 2013, Theorem 1.1 and Proposition 4.6] under the assumption that $R$ is a Cohen–Macaulay local ring. Our results show that this assumption is superfluous.

**Corollary 4.11.** (1) *The assignments $\Phi \mapsto \operatorname{mod}^\Phi(R)$ and $\mathscr{X} \mapsto \mathsf{NF}(\mathscr{X})$ gives mutually inverse bijections between*

(a) *the specialization-closed subsets of $\operatorname{Spec} R$ contained in $\mathsf{S}(R)$, and*

(b) *the resolving subcategories of $\operatorname{mod} R$ closed under tensor products and transposes.*

(2) *Let $\mathscr{X}$ be a resolving subcategory of $\operatorname{mod} R$. Then the following are equivalent*:

(a) *$\mathscr{X}$ is the resolving closure of a Serre subcategory of $\operatorname{mod} R$.*

(b) *$\mathscr{X}$ is closed under tensor products and transposes.*

(c) *$R/\mathfrak{p}$ belongs to $\mathscr{X}$ for all $\mathfrak{p} \in \mathsf{NF}(\mathscr{X})$.*

(d) *For all $\mathfrak{p} \in \mathsf{NF}(\mathscr{X})$ there exists $M \in \mathscr{X}$ such that $\kappa(\mathfrak{p})$ is a direct summand of $M_\mathfrak{p}$.*

*Proof.* Recall that we have proved in Corollary 4.3(3) that if $R$ is a local ring with maximal ideal $\mathfrak{m}$ and residue field $k$, then the equality $\operatorname{mod}^{\{\mathfrak{m}\}}(R) = \operatorname{res}_{\operatorname{mod} R}(k)$ holds. Hence, in view of [Dao and Takahashi 2014, Lemma 3.2], we see that all the ten assertions in [Takahashi 2013, Lemma 2.5] hold without the assumption that $R$ is Cohen–Macaulay. Therefore, it is observed from [Dao and Takahashi 2014, Proposition 3.3] and the proof of [Takahashi 2013, Proposition 3.1] that one can remove from [Takahashi 2013, Proposition 3.1] the two assumptions that $R$ is local and that $R$ is Cohen–Macaulay. Thus, the proof of [Takahashi 2013, Theorem 3.3] actually proves that the statement [Takahashi 2013, Theorem 3.3] holds without the assumption that $R$ is a Cohen–Macaulay local ring. Since [Takahashi 2013, Lemma 4.5] (respectively, [Takahashi 2013, Lemma 4.4]) is still valid for an arbitrary commutative noetherian ring (respectively, local ring) $R$, so are [Takahashi 2013, Proposition 4.6 and Theorem 4.7]. Now our Theorem 4.10 completes the proof. □

## Acknowledgments

The author is grateful to Osamu Iyama for giving him helpful suggestions and letting him know about their results. He thanks Takuma Aihara and Xiao-Wu Chen for useful comments, and the referee for reading the paper carefully.

## References

[Aihara and Takahashi 2011] T. Aihara and R. Takahashi, "Generators and dimensions of derived categories", preprint, 2011. arXiv 1106.0205

[Aihara et al. 2014] T. Aihara, T. Araya, O. Iyama, R. Takahashi, and M. Yoshiwaki, "Dimensions of triangulated categories with respect to subcategories", *J. Algebra* **399** (2014), 205–219. MR 3144584

[Avramov et al. 2010a] L. L. Avramov, R.-O. Buchweitz, S. B. Iyengar, and C. Miller, "Homology of perfect complexes", *Adv. Math.* **223**:5 (2010), 1731–1781. MR 2011k:13014 Zbl 1186.13006

[Avramov et al. 2010b] L. L. Avramov, S. B. Iyengar, and J. Lipman, "Reflexivity and rigidity for complexes, I: Commutative rings", *Algebra Number Theory* **4**:1 (2010), 47–86. MR 2011k:13016 Zbl 1194.13017

[Ballard et al. 2012] M. Ballard, D. Favero, and L. Katzarkov, "Orlov spectra: bounds and gaps", *Invent. Math.* **189**:2 (2012), 359–430. MR 2947547 Zbl 1266.14013

[Balmer 2002] P. Balmer, "Presheaves of triangulated categories and reconstruction of schemes", *Math. Ann.* **324**:3 (2002), 557–580. MR 2003j:18016 Zbl 1011.18007

[Balmer 2005] P. Balmer, "The spectrum of prime ideals in tensor triangulated categories", *J. Reine Angew. Math.* **588** (2005), 149–168. MR 2007b:18012 Zbl 1080.18007

[Benson et al. 2011] D. J. Benson, S. B. Iyengar, and H. Krause, "Stratifying modular representations of finite groups", *Ann. of Math.* (2) **174**:3 (2011), 1643–1684. MR 2846489 Zbl 1261.20057

[Bergh et al. 2010] P. A. Bergh, S. B. Iyengar, H. Krause, and S. Oppermann, "Dimensions of triangulated categories via Koszul objects", *Math. Z.* **265**:4 (2010), 849–864. MR 2011f:18016 Zbl 1263.18006

[Bondal and Van den Bergh 2003] A. Bondal and M. Van den Bergh, "Generators and representability of functors in commutative and noncommutative geometry", *Mosc. Math. J.* **3**:1 (2003), 1–36. MR 2004h:18009 Zbl 1135.18302

[Buchweitz 1986] R.-O. Buchweitz, "Maximal Cohen–Macaulay modules and Tate-cohomology over Gorenstein rings", preprint, University of Hannover, 1986, http://hdl.handle.net/1807/16682.

[Burke et al. 2012] J. Burke, L. W. Christensen, and R. Takahashi, "Building modules from the singular locus", preprint, 2012. To appear in *Math. Scand.* arXiv 1210.0055

[Christensen 1998] J. D. Christensen, "Ideals in triangulated categories: phantoms, ghosts and skeleta", *Adv. Math.* **136**:2 (1998), 284–339. MR 99g:18007 Zbl 0928.55010

[Christensen 2000] L. W. Christensen, *Gorenstein dimensions*, Lecture Notes in Math. **1747**, Springer, Berlin, 2000. MR 2002e:13032 Zbl 0965.13010

[Dao and Takahashi 2011] H. Dao and R. Takahashi, "The radius of a subcategory of modules", preprint, 2011. To appear in *Algebra Number Theory*. arXiv 1111.2902

[Dao and Takahashi 2012a] H. Dao and R. Takahashi, "The dimension of a subcategory of modules", preprint, 2012. arXiv 1203.1955

[Dao and Takahashi 2012b] H. Dao and R. Takahashi, "Upper bounds for dimensions of singularity categories", preprint, 2012. arXiv 1203.1683

[Dao and Takahashi 2014] H. Dao and R. Takahashi, "Classification of resolving subcategories and grade consistent functions", *Int. Math. Res. Not* (2014). arXiv 1202.5605

[Friedlander and Pevtsova 2007] E. M. Friedlander and J. Pevtsova, "Π-supports for modules for finite group schemes", *Duke Math. J.* **139**:2 (2007), 317–368. MR 2008g:14081 Zbl 1128.20031

[Gabriel 1962] P. Gabriel, "Des catégories abéliennes", *Bull. Soc. Math. France* **90** (1962), 323–448. MR 38 #1144 Zbl 0201.35602

[Hopkins and Smith 1998] M. J. Hopkins and J. H. Smith, "Nilpotence and stable homotopy theory, II", *Ann. of Math.* (2) **148**:1 (1998), 1–49. MR 99h:55009 Zbl 0927.55015

[Hovey 2001] M. Hovey, "Classifying subcategories of modules", *Trans. Amer. Math. Soc.* **353**:8 (2001), 3181–3191. MR 2002i:13007 Zbl 0981.13006

[Keller et al. 2011] B. Keller, D. Murfet, and M. Van den Bergh, "On two examples by Iyama and Yoshino", *Compos. Math.* **147**:2 (2011), 591–612. MR 2012c:13029 Zbl 1264.13016

[Krause 2008] H. Krause, "Thick subcategories of modules over commutative Noetherian rings (with an appendix by Srikanth Iyengar)", *Math. Ann.* **340**:4 (2008), 733–747. MR 2008m:13017 Zbl 1143.13012

[Krause and Kussin 2006] H. Krause and D. Kussin, "Rouquier's theorem on representation dimension", pp. 95–103 in *Trends in representation theory of algebras and related topics* (Querétaro, 2004), edited by J. A. de la Peña and R. Bautista, Contemp. Math. **406**, Amer. Math. Soc., Providence, RI, 2006. MR 2008c:16010 Zbl 1107.16013

[Krause and Stevenson 2013] H. Krause and G. Stevenson, "A note on thick subcategories of stable derived categories", *Nagoya Math. J.* **212** (2013), 87–96. arXiv 1111.2220

[Neeman 1992] A. Neeman, "The chromatic tower for $D(R)$", *Topology* **31**:3 (1992), 519–532. MR 93h:18018 Zbl 0793.18008

[Neeman 2001] A. Neeman, *Triangulated categories*, Ann. of Math. Stud. **148**, Princeton University Press, 2001. MR 2001k:18010 Zbl 0974.18008

[Oppermann 2009] S. Oppermann, "Lower bounds for Auslander's representation dimension", *Duke Math. J.* **148**:2 (2009), 211–249. MR 2010i:16018 Zbl 1173.16007

[Orlov 2004] D. O. Orlov, "Триангулированные категории особенностей и D-браны в моделях Ландау–Гинзбурга", *Tr. Mat. Inst. Steklova* **246** (2004), 240–262. Translated as "Triangulated categories of singularities and D-branes in Landau–Ginzburg models", *Proc. Steklov Inst. Math.* **246**:3 (2004), 227–248. MR 2006i:81173 Zbl 1101.81093 arXiv math/0302304

[Orlov 2006] D. O. Orlov, "Триангулированные категории особенностей и эквивалентности между моделями Ландау–Гинзбурга", *Mat. Sb.* **197**:12 (2006), 117–132. Translated as "Triangulated categories of singularities, and equivalences between Landau–Ginzburg models", *Sb. Math.* **197**:12 (2006), 1827–1840. MR 2009g:14013 Zbl 1161.14301

[Orlov 2009a] D. O. Orlov, "Derived categories of coherent sheaves and triangulated categories of singularities", pp. 503–531 in *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin*, vol. II, edited by Y. Tschinkel and Y. G. Zarhin, Progr. Math. **270**, Birkhäuser, Boston, 2009. MR 2011c:14050 Zbl 1200.18007

[Orlov 2009b] D. O. Orlov, "Remarks on generators and dimensions of triangulated categories", *Mosc. Math. J.* **9**:1 (2009), 143–149. MR 2011a:14031 Zbl 1197.18004

[Orlov 2011] D. O. Orlov, "Formal completions and idempotent completions of triangulated categories of singularities", *Adv. Math.* **226**:1 (2011), 206–217. MR 2012c:14035 Zbl 1216.18012

[Orlov 2012] D. O. Orlov, "Matrix factorizations for nonaffine LG-models", *Math. Ann.* **353**:1 (2012), 95–108. MR 2910782 Zbl 1243.81178

[Rouquier 2006] R. Rouquier, "Representation dimension of exterior algebras", *Invent. Math.* **165**:2 (2006), 357–367. MR 2007f:16031 Zbl 1101.18006

[Rouquier 2008] R. Rouquier, "Dimensions of triangulated categories", *J. K-Theory* **1**:2 (2008), 193–256. MR 2009i:18008 Zbl 1165.18008

[Schoutens 2003] H. Schoutens, "Projective dimension and the singular locus", *Comm. Algebra* **31**:1 (2003), 217–239. MR 2005e:13020 Zbl 1014.13003

[Stevenson 2014] G. Stevenson, "Subcategories of singularity categories via tensor actions", *Compos. Math* (2014). arXiv 1105.4698

[Takahashi 2009] R. Takahashi, "Modules in resolving subcategories which are free on the punctured spectrum", *Pacific J. Math.* **241**:2 (2009), 347–367. MR 2010b:13027 Zbl 1172.13005

[Takahashi 2010] R. Takahashi, "Classifying thick subcategories of the stable category of Cohen–Macaulay modules", *Adv. Math.* **225**:4 (2010), 2076–2116. MR 2011h:13014 Zbl 1202.13009

[Takahashi 2013] R. Takahashi, "Classifying resolving subcategories over a Cohen–Macaulay local ring", *Math. Z.* **273**:1-2 (2013), 569–587. MR 3010176 Zbl 1267.13024

[Thomason 1997] R. W. Thomason, "The classification of triangulated subcategories", *Compos. Math.* **105**:1 (1997), 1–27. MR 98b:18017 Zbl 0873.18003

Received June 12, 2013.

RYO TAKAHASHI
GRADUATE SCHOOL OF MATHEMATICS
NAGOYA UNIVERSITY
FUROCHO, CHIKUSAKU
NAGOYA 464-8602
JAPAN
takahashi@math.nagoya-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.268.249

# A VIRTUAL KAWASAKI–RIEMANN–ROCH FORMULA

VALENTIN TONITA

**Kawasaki's formula is a tool to compute holomorphic Euler characteristics of vector bundles on a compact orbifold $\mathscr{X}$. Let $\mathscr{X}$ be an orbispace with perfect obstruction theory which admits an embedding in a smooth orbifold. One can then construct the virtual structure sheaf and the virtual fundamental class of $\mathscr{X}$. In this paper we prove that Kawasaki's formula "behaves well" with working "virtually" on $\mathscr{X}$ in the following sense: if we replace the structure sheaves, tangent and normal bundles in the formula by their virtual counterparts then Kawasaki's formula stays true. Our motivation comes from studying the quantum $K$-theory of a complex manifold $X$ (Givental and Tonita, 2014), with the formula applied to Kontsevich moduli spaces of genus-0 stable maps to $X$.**

## 1. Introduction

Given a manifold $\mathscr{X}$ and a vector bundle $V$ on $\mathscr{X}$, then the Hirzebruch–Riemann–Roch formula states that

$$\chi(\mathscr{X}, V) = \int_{\mathscr{X}} \mathrm{ch}(V) T d(T_{\mathscr{X}}).$$

Kawasaki [1979] generalized this formula to the case when $\mathscr{X}$ is an orbifold. He reduces the computation of Euler characteristics on $\mathscr{X}$ to the computation of certain cohomological integrals on *the inertia orbifold* $I\mathscr{X}$:

$$\chi(\mathscr{X}, V) = \sum_{\mu} \frac{1}{m_\mu} \int_{\mathscr{X}_\mu} T d(T_{\mathscr{X}_\mu}) \, \mathrm{ch}\Big(\frac{\mathrm{Tr}(V)}{\mathrm{Tr}(\Lambda^\bullet N_\mu^*)}\Big). \tag{1}$$

We explain below the ingredients in the formula:

$I\mathscr{X}$ is defined as follows: around any point $p \in \mathscr{X}$ there is a local chart $(\widetilde{U}_p, G_p)$ such that locally $\mathscr{X}$ is represented as the quotient of $\widetilde{U}_p$ by $G_p$. Consider the set of conjugacy classes $(1) = (h_p^1), (h_p^2), \ldots, (h_p^{n_p})$ in $G_p$. Define

$$I\mathscr{X} := \{(p, (h_p^i)) \mid i = 1, 2, \ldots, n_p\}.$$

*MSC2010:* 19L10.
*Keywords:* Gromov–Witten theory, Riemann–Roch type formulae.

Pick an element $h_p^i$ in each conjugacy class. Then a local chart on $I\mathscr{X}$ is given by

$$\coprod_{i=1}^{n_p} \widetilde{U}_p^{(h_p^i)}/Z_{G_p}(h_p^i),$$

where $Z_{G_p}(h_p^i)$ is the centralizer of $h_p^i$ in $G_p$. Denote by $\mathscr{X}_\mu$ the connected components of the inertia orbifold (we'll often refer to them as Kawasaki strata). The multiplicity $m_\mu$ associated to each $\mathscr{X}_\mu$ is given by

$$m_\mu := \left|\ker(Z_{G_p}(g) \to \mathrm{Aut}(\widetilde{U}_p^g))\right|.$$

For a vector bundle $V$ we will denote by $V^*$ the dual bundle to $V$. The restriction of $V$ to $\mathscr{X}_\mu$ decomposes in characters of the $g$ action. Let $E_r^{(l)}$ be the subbundle of the restriction of $E$ to $\mathscr{X}_\mu$ on which $g$ acts with eigenvalue $e^{2\pi il/r}$. Then the trace $\mathrm{Tr}(V)$ is defined to be the orbibundle whose fiber over the point $(p, (g))$ of $\mathscr{X}_\mu$ is

$$\mathrm{Tr}(V) := \sum_l e^{\frac{2\pi il}{r}} E_r^{(l)}.$$

Finally, $\Lambda^\bullet N_\mu^*$ is the $K$-theoretic Euler class of the normal bundle $N_\mu$ of $\mathscr{X}_\mu$ in $\mathscr{X}$. $\mathrm{Tr}(\Lambda^\bullet N_\mu^*)$ is invertible because the symmetry $g$ acts with eigenvalues different from 1 on the normal bundle to the fixed point locus. We call the terms corresponding to the identity component in the formula *fake Euler characteristics*:

$$\chi^f(\mathscr{X}, V) = \int_{\mathscr{X}} \mathrm{ch}(V) T d(T_{\mathscr{X}}).$$

In the case where $\mathscr{X}$ is a global quotient, formula (1) is the Lefschetz fixed point formula.

Now let $\mathscr{X}$ be a compact, complex orbispace (Deligne–Mumford stack) with a perfect obstruction theory $E^{-1} \to E^0$. This is used to define the intrinsic normal cone, which is embedded in $E_1$ — the dual bundle to $E^{-1}$ (see [Li and Tian 1998; Behrend and Fantechi 1997]). The virtual structure sheaf $\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}$ was defined in [Lee 2004] as the $K$-theoretic pullback by the zero section of the structure sheaf of this cone. Let $I\mathscr{X} = \coprod_\mu \mathscr{X}_\mu$ be the inertia orbifold of $\mathscr{X}$. We denote by $i_\mu$ the inclusion of a stratum $\mathscr{X}_\mu$ in $\mathscr{X}$. For a bundle $V$ on $\mathscr{X}$, we write $i_\mu^* V = V_\mu^f \oplus V_\mu^m$ for its decomposition as the direct sum of the fixed part and the moving part under the action of the symmetry associated to $\mathscr{X}_\mu$. To avoid ugly notation we will often simply write $V^m$, $V^f$. The virtual normal bundle to $\mathscr{X}_\mu$ in $\mathscr{X}$ is defined as $[E_0^m] - [E_1^m]$. We will in addition assume that $\mathscr{X}$ admits an embedding $j$ in a smooth compact orbifold $\mathscr{Y}$. This is always true for the moduli spaces of genus-0 stable maps $X_{0,n,d}$ because an embedding $X \hookrightarrow \mathbb{P}^N$ induces an embedding $X_{0,n,d} \hookrightarrow (\mathbb{P}^N)_{0,n,d}$.

**Theorem 1.1.** *Denote by $N_\mu^{\mathrm{vir}}$ the virtual normal bundle of $\mathscr{X}_\mu$ in $\mathscr{X}$. Then*

$$\chi(\mathscr{X}, j^*(V) \otimes \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}) = \sum_\mu \frac{1}{m_\mu} \chi^f\left(\mathscr{X}_\mu, \frac{\mathrm{Tr}(V_\mu \otimes \mathbb{O}_{\mathscr{X}_\mu}^{\mathrm{vir}})}{\mathrm{Tr}(\Lambda^\bullet(N_\mu^{\mathrm{vir}})^*)}\right). \tag{2}$$

**Remark 1.2.** A perfect obstruction theory $E^{-1} \to E^0$ on $\mathscr{X}$ induces canonically a perfect obstruction theory on $\mathscr{X}_\mu$ by taking the fixed part of the complex $E_\mu^{-1,f} \to E_\mu^{0,f}$. The proof is the same as that of Proposition 1 in [Graber and Pandharipande 1999]. This is then used to define the sheaf $\mathbb{O}_{\mathscr{X}_\mu}^{\mathrm{vir}}$.

**Remark 1.3.** It is proved in [Fantechi and Göttsche 2010] that if $\mathscr{X}$ is a scheme, the Grothendieck–Riemann–Roch theorem is compatible with virtual fundamental classes and virtual fundamental sheaves, that is,

$$\chi^f(\mathscr{X}, V \otimes \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}) = \int_{[\mathscr{X}]} \mathrm{ch}(V \otimes \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}) \cdot T\,d(T^{\mathrm{vir}}),$$

where $[\mathscr{X}]$ is the virtual fundamental class of $\mathscr{X}$ and $T^{\mathrm{vir}}$ is its virtual tangent bundle. Their arguments carry over to the case when $\mathscr{X}$ is a stack.

**Remark 1.4.** The bundles $V$ to which we apply Theorem 1.1 in [Givental and Tonita 2014] are (sums and products of) cotangent line bundles $L_i$ and evaluation classes $\mathrm{ev}_i^*(a_i)$ (where $a_i$ are $K$-theoretic classes on the target). They are pullbacks of the corresponding bundles on $(\mathbb{P}^N)_{0,n,d}$.

## 2. Proof of Theorem 1.1

Before proving Theorem 1.1 we recall a couple of background facts and lemmata on $K$-theory which we will use.

Let $K_0(X)$ be the Grothendieck group of coherent sheaves on $X$. Given a map $f : X \to Y$, the $K$-theoretic pullback $f^*(\mathscr{F}) : K_0(Y) \to K_0(X)$ is defined as the alternating sum of derived functors $\mathrm{Tor}^i_{\mathbb{O}_Y}(\mathscr{F}, \mathbb{O}_X)$, provided that the sum is finite. This is always true for instance if $f$ is flat or if it is a regular embedding.

For any fiber square

$$\begin{array}{ccc} V' & \longrightarrow & V \\ \downarrow & & \downarrow \\ B' & \xrightarrow{i} & B \end{array}$$

with $i$ a regular embedding one can define $K$-theoretic refined Gysin homomorphisms $i^! : K_0(V) \to K_0(V')$ (see [Lee 2004]). One way to define the map $i^!$ is the following: The class $i_*(\mathbb{O}_{B'}) \in K^0(B)$ has a finite resolution of vector bundles, which is exact off $B'$. We pull it back to $V$ and then cap (i.e., tensor product) with classes in $K_0(V)$, to get a class on $K_0(V)$ with homology supported on $V'$, which we can regard as an element of $K_0(V')$, because there is a canonical isomorphism between complexes on $V$ with homology supported on $V'$ and $K_0(V')$.

In the following two lemmata, $X$, $Y$, $Y'$ are assumed DM stacks. We will use the following result:

**Lemma 2.1.** *Consider the diagram*:

$$\begin{tikzcd}
\iota^* C_{X/Y} \arrow[r] \arrow[d] & C_{X/Y} \arrow[d] \\
X' \arrow[r, "\iota"] \arrow[d] & X \arrow[d, "j"'] \\
Y' \arrow[r, "i"] & Y
\end{tikzcd}$$

*with $i$ a regular embedding and $j$ an embedding, $C_{X/Y}$ is the normal cone of $X$ in $Y$ and both squares are fiber diagrams. Then*

$$i^![\mathcal{O}_{C_{X/Y}}] = [\mathcal{O}_{C_{X'/Y'}}] \in K_0(\iota^* C_{X/Y}). \tag{3}$$

This is stated and proved in [Lee 2004, Lemma 2]. The proof is based on a more general statement (Lemma 1 of [Lee 2004]), which has been worked out in [Kresch 1999] on the level of Chow rings. Since $K$-theoretic statements are stronger, we give below the key ingredient which allows one to carry over Kresch's proof to $K$-theory:

**Lemma 2.2.** *Let $f : X \to Y$ be a closed embedding and let $g : Y \to \mathbb{P}^1$ be a surjection such that $g \circ f$ is flat. Denote by $X_0$ and $Y_0$ the fibers over* 0 *of $g \circ f$ and $g$, respectively. Moreover, assume that the restriction of $f$ to $X \setminus X_0$ is an isomorphism. Then if $i$ is the inclusion of* $\{0\}$ *in $\mathbb{P}^1$, we have $i^!(\mathcal{O}_Y) = \mathcal{O}_{X_0} \in K_0(Y_0)$.*

*Proof.* The skyscraper sheaves at all points of $\mathbb{P}^1$ represent the same element in $K_0(\mathbb{P}^1)$, hence if we pull back a resolution of any point $P \in \mathbb{P}^1$ by $g$ we get the same elements of $K_0(Y)$. On the other hand since $f$ is an isomorphism above $\mathbb{P}^1 \setminus \{0\}$, pulling back by $g$ of the structure sheaf of a point $P \neq 0$ is the same as pulling back by $g \circ f$ followed by $f_*$. By what we said above we can replace $P$ with 0. Now from the flatness of $g \circ f$ above 0 the pullback of the structure sheaf of 0 by $g \circ f$ is the structure sheaf of the fiber $X_0$. The result then follows from the definition of $i^!$. □

**Remark 2.3.** Lemma 2.2 allows one to show Lemma 2.1: intermediately one shows, following [Kresch 1999] (notation is as in Lemma 2.1), that $[\mathcal{O}_{C_1}] = [\mathcal{O}_{C_2}]$ in $K_0(C_{X'}Y \times_Y C_X Y)$, where $C_1 := C_{i^* C_X Y}(C_X Y)$ and $C_2 := C_{j^* C_{Y'} Y}(C_{Y'} Y)$.

We now go on to prove Theorem 1.1. We have

$$\chi(\mathscr{X}, j^*V \otimes \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}) = \chi(\mathscr{Y}, V \otimes j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}).$$

Kawasaki's formula applied to the sheaf $V \otimes j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}$ on $\mathscr{Y}$ gives

$$\chi(\mathscr{Y}, V \otimes j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}) = \sum_{\mu} \frac{1}{m_\mu} \chi^f\left(\mathscr{Y}_\mu, \frac{\mathrm{Tr}(V_\mu \otimes i_\mu^* j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}})}{\mathrm{Tr}(\Lambda^\bullet N_\mu^*)}\right). \tag{4}$$

From the fiber diagram

$$\begin{array}{ccc} \mathscr{X}_\mu & \xrightarrow{i'_\mu} & \mathscr{X} \\ {\scriptstyle j'}\downarrow & & \downarrow{\scriptstyle j} \\ \mathscr{Y}_\mu & \xrightarrow{i_\mu} & \mathscr{Y} \end{array}$$

and Theorem 6.2 in [Fulton 1998] (where this is proved for Chow rings) we have $i_\mu^* j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}} = j'_* i_\mu^! \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}$. Plugging this in (4) gives

$$\chi^f\left(\mathscr{Y}_\mu, \frac{\mathrm{Tr}(V_\mu \otimes i_\mu^* j_*\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}})}{\mathrm{Tr}(\Lambda^\bullet N_\mu^*)}\right) = \chi^f\left(\mathscr{Y}_\mu, \frac{\mathrm{Tr}(V_\mu \otimes j'_* i_\mu^! \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}})}{\mathrm{Tr}(\Lambda^\bullet N_\mu^*)}\right). \tag{5}$$

Let $G_\mu$ be the cyclic group generated by one element of the conjugacy class associated to $\mathscr{X}_\mu$. Then we will show that

$$\mathrm{Tr}\left(\frac{i_\mu^! \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}}{\Lambda^\bullet(N_\mu^*)}\right) = \mathrm{Tr}\left(\frac{\mathbb{O}_{\mathscr{X}_\mu}^{\mathrm{vir}}}{\Lambda^\bullet(N_\mu^{\mathrm{vir}})^*}\right) \tag{6}$$

in the $G_\mu$-equivariant $K$-ring of $\mathscr{X}_\mu$. This is essentially the computation of Section 3 in [Graber and Pandharipande 1999] carried out in $\mathbb{C}^*$-equivariant $K$-theory. Relation (6) then follows by embedding the group $G_\mu$ in the torus and specializing the value of the variable $t$ in the ground ring of $\mathbb{C}^*$-equivariant $K$-theory to a $|G_\mu|$-root of unity.

If we define a cone $D := C_{\mathscr{X}/\mathscr{Y}} \times_{\mathscr{X}} E_0$, then this is a $T_{\mathscr{Y}}$ cone (see [Behrend and Fantechi 1997]). The virtual normal cone $D^{\mathrm{vir}}$ is defined as $D/T_{\mathscr{Y}}$ and $\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}$ is the pullback by the zero section of the structure sheaf of $D^{\mathrm{vir}}$. Alternatively there is a fiber diagram

$$\begin{array}{ccc} T_{\mathscr{Y}} & \longrightarrow & D \\ \downarrow & & \downarrow \\ \mathscr{X} & \xrightarrow{0_{E_1}} & E_1 \end{array}$$

where the bottom map is the zero section of $E_1$. Then one can define $\mathbb{O}_{\mathscr{X}}^{\mathrm{vir}}$ as $0_{T_{\mathscr{Y}}}^* 0_{E_1}^! [\mathbb{O}_D]$. We'll prove formula (6) following closely the calculation in [Graber and Pandharipande 1999]. First, by definition of $\mathbb{O}^{\mathrm{vir}}_{\mathcal{X}}$ and by commutativity of Gysin maps, we have

$$i_\mu^! \mathbb{O}^{\mathrm{vir}}_{\mathcal{X}} = i_\mu^! 0^*_{T_{\mathcal{Y}}} 0^!_{E_1}[\mathbb{O}_D] = 0^*_{T_{\mathcal{Y}}} 0^!_{E_1} i_\mu^! [\mathbb{O}_D]. \tag{7}$$

We pull back relation (3) to $(i'_\mu)^* D = (i'_\mu)^*(C_{\mathcal{X}/\mathcal{Y}} \times E_0)$ to get

$$i_\mu^! [\mathbb{O}_D] = [\mathbb{O}_{D_\mu} \times (E_0^m)^*]. \tag{8}$$

In the equality above we have used the fact that $D_\mu = C_{\mathcal{X}_\mu/\mathcal{Y}_\mu} \times E_0^f$ and we identified the sheaf of sections of the bundle $E_0^m$ with the dual bundle $(E_0^m)^*$. Plugging (8) in (7) we get

$$i_\mu^! \mathbb{O}^{\mathrm{vir}}_{\mathcal{X}} = 0^*_{T_{\mathcal{Y}}} 0^!_{E_1}[\mathbb{O}_{D_\mu} \times (E_0^m)^*]. \tag{9}$$

Notice that the action of $T_{\mathcal{Y}_\mu}$ leaves $D_\mu \times (E_0^m)^*$ invariant (it acts trivially on $(E_0^m)^*$). Now we can write $0^*_{T_{\mathcal{Y}}} = 0^*_{T_{\mathcal{Y}^f_\mu}} \times 0^*_{T_{\mathcal{Y}^m_\mu}}$ and since $D_\mu^{\mathrm{vir}} = D_\mu / T_{\mathcal{Y}_\mu}$ we rewrite (9) as

$$i_\mu^! \mathbb{O}^{\mathrm{vir}}_{\mathcal{X}} = 0^*_{T_{\mathcal{Y}^m_\mu}} 0^!_{E_1}[\mathbb{O}_{D_\mu^{\mathrm{vir}}} \times (E_0^m)^*]. \tag{10}$$

The proof of Lemma 1 in [Graber and Pandharipande 1999] works in our set-up as well: it uses excess intersection formula which holds in $K$-theory. It shows that the following relation holds in the $\mathbb{C}^*$-equivariant $K$-ring of $\mathcal{X}_\mu$:

$$0^*_{T\mathcal{Y}^m_\mu} 0^!_{E_1}[\mathbb{O}_{D_\mu^{\mathrm{vir}}} \times (E_0^m)^*] = 0^*_{E_0^m}\big(0^!_{E_1}[\mathbb{O}_{D_\mu^{\mathrm{vir}}} \times (E_0^m)^*]\big) \cdot \frac{\Lambda^\bullet(T_{\mathcal{Y}^m})^*}{\Lambda^\bullet(E_0^m)^*}. \tag{11}$$

The class $0^!_{E_1}[\mathbb{O}_{D_\mu^{\mathrm{vir}}} \times E_0^m]$ lives in the $\mathbb{C}^*$-equivariant $K$-ring of $E_0^m$. The class doesn't depend on the bundle map $E_0^m \to E_1^m$ so we can assume this map to be 0. Then by excess intersection formula and the definition of $\mathbb{O}^{\mathrm{vir}}_{\mathcal{X}_\mu}$ we get

$$0^*_{E_0^m}\big(0^!_{E_1}[\mathbb{O}_{D_\mu^{\mathrm{vir}}} \times (E_0^m)^*]\big) = \mathbb{O}^{\mathrm{vir}}_{\mathcal{X}_\mu} \cdot \Lambda^\bullet(E_1^m)^*. \tag{12}$$

Formula (12) holds because $D_\mu^{\mathrm{vir}} \times (E_0^m) \subset E_1^f \times E_0^m$ and $0^!_{E_1}$ acts as $0^!_{E_1^f} \times 0^!_{E_1^m}$ on factors. $0^!_{E_1^f}[\mathbb{O}_{D_\mu^{\mathrm{vir}}}] = \mathbb{O}^{\mathrm{vir}}_{\mathcal{X}_\mu}$ by definition of $\mathbb{O}^{\mathrm{vir}}_{\mathcal{X}_\mu}$. By excess intersection formula applied to the fiber square

$$\begin{tikzcd} E_0^m \arrow[r] \arrow[d, "\pi"'] & E_0^m \arrow[d] \\ \mathcal{X}_\mu \arrow[r, "0_{E_1^m}"] & E_1^m \end{tikzcd}$$

we have $0^*_{E_0^m} 0^!_{E_1^m}[(E_0^m)^*] = 0^*_{E_0^m} \pi^* \Lambda^\bullet(E_1^m)^* = \Lambda^\bullet(E_1^m)^*$. Plugging formula (12) in (11) (note that $N_\mu = T_{\mathcal{Y}^m_\mu}$ and $N_\mu^{\mathrm{vir}} = [E_0^m] - [E_1^m]$) and taking traces proves (6).

We now plug (6) in (5) and then pull back to $\mathscr{X}_\mu$ to get

$$\chi^f\left(\mathscr{Y}_\mu, \frac{\operatorname{Tr}(V_\mu \otimes j_* i_\mu^* \mathbb{O}_{\mathscr{X}}^{\mathrm{vir}})}{\operatorname{Tr}(\Lambda^\bullet N_\mu^*)}\right) = \chi^f\left(\mathscr{Y}_\mu, \operatorname{Tr}(V_\mu) \otimes j'_* \frac{\operatorname{Tr}(\mathbb{O}_{\mathscr{X}_\mu}^{\mathrm{vir}})}{\operatorname{Tr}(\Lambda^\bullet (N_\mu^{\mathrm{vir}})^*)}\right)$$
$$= \chi^f\left(\mathscr{X}_\mu, \frac{\operatorname{Tr}(V_\mu \otimes \mathbb{O}_{\mathscr{X}_\mu}^{\mathrm{vir}})}{\operatorname{Tr}(\Lambda^\bullet (N_\mu^{\mathrm{vir}})^*)}\right). \qquad \square$$

## Acknowledgements

I would like to thank Alexander Givental for suggesting the problem and for useful discussions. Thanks are also due to Yuan-Pin Lee who patiently answered my questions on the material in his work [Lee 2004] and to Hsian-Hua Tseng who read a preliminary draft of the paper.

## References

[Behrend and Fantechi 1997] K. Behrend and B. Fantechi, "The intrinsic normal cone", *Invent. Math.* **128**:1 (1997), 45–88. MR 98e:14022 Zbl 0909.14006

[Fantechi and Göttsche 2010] B. Fantechi and L. Göttsche, "Riemann–Roch theorems and elliptic genus for virtually smooth schemes", *Geom. Topol.* **14**:1 (2010), 83–115. MR 2011a:14016 Zbl 1194.14017

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer, Berlin, 1998. MR 99d:14003 Zbl 0885.14002

[Givental and Tonita 2014] A. Givental and V. Tonita, "The Hirzebruch–Riemann–Roch theorem in true genus-0 quantum K-theory", pp. 43–90 in *Symplectic, Poisson and noncommutative geometry*, Mathematical Sciences Research Publications **62**, Cambridge Univ. Press, New York, 2014. arXiv 1106.3136

[Graber and Pandharipande 1999] T. Graber and R. Pandharipande, "Localization of virtual classes", *Invent. Math.* **135**:2 (1999), 487–518. MR 2000h:14005 Zbl 0953.14035

[Kawasaki 1979] T. Kawasaki, "The Riemann–Roch theorem for complex $V$-manifolds", *Osaka J. Math.* **16**:1 (1979), 151–159. MR 80f:58042 Zbl 0405.32010

[Kresch 1999] A. Kresch, "Canonical rational equivalence of intersections of divisors", *Invent. Math.* **136**:3 (1999), 483–496. MR 2000d:14005 Zbl 0923.14003

[Lee 2004] Y.-P. Lee, "Quantum $K$-theory, I: Foundations", *Duke Math. J.* **121**:3 (2004), 389–424. MR 2005f:14107 Zbl 1051.14064

[Li and Tian 1998] J. Li and G. Tian, "Virtual moduli cycles and Gromov–Witten invariants of algebraic varieties", *J. Amer. Math. Soc.* **11**:1 (1998), 119–174. MR 99d:14011 Zbl 0912.14004

Received October 19, 2012. Revised September 23, 2013.

VALENTIN TONITA
KAVLI INSTITUTE FOR THE PHYSICS AND MATHEMATICS OF THE UNIVERSE (KAVLI IPMU)
5-1-5 KASHIWANOHA
KASHIWA 277-8583
JAPAN
valentin.tonita@ipmu.jp

# Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

# PACIFIC JOURNAL OF MATHEMATICS

Volume 268 No. 1 March 2014

---