Pacific Journal of Mathematics

Volume 251 No. 2 June 2011

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

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

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

The subscription price for 2011 is US $420/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

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

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

TWO KAZDAN–WARNER-TYPE IDENTITIES FOR THE RENORMALIZED VOLUME COEFFICIENTS AND THE GAUSS–BONNET CURVATURES OF A RIEMANNIAN METRIC

BIN GUO, ZHENG-CHAO HAN AND HAIZHONG LI

We prove two Kazdan–Warner-type identities involving the renormalized volume coefficients $v^{(2k)}$ of a Riemannian manifold (M^n, g), the Gauss–Bonnet curvature G_{2r}, and a conformal Killing vector field on (M^n, g). In the case when the Riemannian manifold is locally conformally flat, we find

$$v^{(2k)} = (-2)^{-k}\sigma_k \quad \text{and} \quad G_{2r}(g) = \frac{4^r(n-r)!r!}{(n-2r)!}\sigma_r$$

and our results reduce to earlier ones established by Viaclovsky in 2000 and the second author in 2006.

1. Introduction

Theorem A [Viaclovsky 2000b; Han 2006a]. *Let (M, g) be a compact Riemannian manifold of dimension $n \geq 3$, let $\sigma_k(g^{-1} \circ A_g)$ be the σ_k curvature of g, and let X be a conformal Killing vector field on (M, g). When $k \geq 3$, assume also that (M, g) is locally conformally flat. Then*

$$\int_M \langle X, \nabla\sigma_k(g^{-1} \circ A_g)\rangle dv_g = 0. \tag{1-1}$$

Recall that on an n-dimensional Riemannian manifold (M, g) with $n \geq 3$, the full Riemannian curvature tensor Rm decomposes as

$$Rm = W_g \oplus (A_g \odot g),$$

where W_g denotes the Weyl tensor of g,

$$A_g = \frac{1}{n-2}\Big(\mathrm{Ric}_g - \frac{R_g}{2(n-1)}g\Big)$$

denotes the Schouten tensor, and $\odot$ is the Kulkarni–Nomizu wedge product. Under a conformal change of metrics $g_w = e^{2w}g$, where w is a smooth function over the

Haizhong Li is supported by NSFC grant number 10971110.
MSC2000: primary 53C20; secondary 53A30.
Keywords: renormalized volume coefficients, $v^{(2k)}$ curvature, conformal transformation, locally conformally flat, σ_k curvature, Gauss–Bonnet curvatures, Kazdan–Warner.

manifold, the Weyl curvature changes pointwise as $W_{g_w} = e^{2w} W_g$. Thus, essential information about the Riemannian curvature tensor under a conformal change of metrics is reflected by the change in the Schouten tensor. One often tries to study the Schouten tensor through the elementary symmetric functions $\sigma_k(g^{-1} \circ A_g)$ (which we later denote as $\sigma_k(g)$) of the eigenvalues of the Schouten tensor, called the σ_k curvatures of g, by studying how they deform under conformal change of metrics.

Question. For all $k \geq 1$, can we generalize Theorem A without the condition that (M, g) is locally conformally flat?

In this note, we show the answer is yes. The renormalized volume coefficients $v^{(2k)}(g)$ of a Riemannian metric g, were introduced in the physics literature in the late 1990s in the context of AdS/CFT correspondence — see [Graham 2009] for a mathematical discussion — and were shown in [Graham and Juhl 2007] to be equal to $\sigma_k(g^{-1} A_g)$, up to a scaling constant, when (M, g) is locally conformally flat. In fact, in the normalization we are going to adopt,

$$v^{(2)}(g) = -\tfrac{1}{2}\sigma_1(g) \quad \text{and} \quad v^{(4)}(g) = \tfrac{1}{4}\sigma_2(g). \tag{1-2}$$

For $k = 3$, Graham and Juhl [2007, page 5] have also listed the formula

$$v^{(6)}(g) = -\frac{1}{8}\Big(\sigma_3(g) + \frac{1}{3(n-4)}(A_g)^{ij}(B_g)_{ij}\Big), \tag{1-3}$$

where

$$(B_g)_{ij} := \frac{1}{n-3}\nabla^k\nabla^l W_{likj} + \frac{1}{n-2}R^{kl} W_{likj}$$

is the *Bach* tensor of the metric. Just as $\int_M \sigma_k(g^{-1} \circ A_g)\, dv_g$ is conformally invariant when $2k = n$ and (M, g) is locally conformally flat, Graham [2009] showed that $\int_M v^{(2k)}(g)\, dv_g$ is also conformally invariant on a general manifold when $2k = n$. Chang and Fang [2008] showed that, for $n \neq 2k$, the Euler–Lagrange equations for the functional $\int_M v^{(2k)}(g)\, dv_g$ under conformal variations subject to the constraint $\mathrm{Vol}_g(M) = 1$ satisfies $v^{(2k)}(g) = \text{const}$, which is a generalized characterization for the curvatures $\sigma_k(g^{-1} \circ A_g)$ when (M, g) is locally conformally flat, as given by Viaclovsky [2000a].

In this note, we will first show that the curvatures $v^{(2k)}(g)$ will play the role of $\sigma_k(g^{-1} \circ A_g)$ in (1-1) for a general manifold. Graham [2009] also gives an explicit expression of $v^{(8)}(g)$, but the explicit expression of $v^{(2k)}(g)$ for general k is not known because they are algebraically complicated; see [Graham 2009, page 1958]. Thus the study of the $v^{(2k)}(g)$ curvatures involves significant challenges not shared by that of $\sigma_k(g)$: First, $v^{(2k)}(g)$ for $k \geq 3$ depends on derivatives of curvature of g; in fact, these depend on derivatives of curvatures of order up to $2k-4$. Second, the $v^{(2k)}(g)$ are defined in [Graham 2009] via an indirect, highly nonlinear inductive algorithm. Despite these difficulties, we can use some properties of these $v^{(2k)}(g)$ curvatures to prove the following.

Theorem 1.1. *Let (M, g) be a compact Riemannian manifold of dimension $n \geq 3$, and let X be a conformal Killing vector field on (M^n, g). For $k \geq 1$, we have*

$$\int_M \langle X, \nabla v^{(2k)}(g)\rangle dv_g = 0. \tag{1-4}$$

Remark 1.2. From (1-2), we know that Theorem 1.1 is equivalent to Theorem A when $k = 1, 2$, or when (M^n, g) is locally conformally flat for $k \geq 3$.

One main reason for interest in identities such as (1-1) and (1-4) is that they play crucial roles in analyzing potentially blowing up conformal metrics with a prescribed curvature function, with $v^{(2k)}(g)$ prescribed in this case. Although little is known about this problem at this stage, Theorem 1.1 establishes one ingredient for attacking this problem.

Our second result involves the Gauss–Bonnet curvatures G_{2r} for $2r \leq n$, introduced by H. Weyl in 1939 and defined by

$$G_{2r}(g) = \delta^{j_1 j_2 \cdots j_{2r-1} j_{2r}}_{i_1 i_2 \cdots i_{2r-1} i_{2r}} R^{i_1 i_2}{}_{j_1 j_2} \cdots R^{i_{2r-1} i_{2r}}{}_{j_{2r-1} j_{2r}},$$

where $\delta^{j_1 j_2 \cdots j_{2r-1} j_{2r}}_{i_1 i_2 \cdots i_{2r-1} i_{2r}}$ is the generalized Kronecker symbol; see also [Labbi 2008]. Note that $G_2 = 2R$, with R the scalar curvature.

Theorem 1.3. *Let (M^n, g) be a compact Riemannian manifold, and let X be a conformal Killing vector field. Then for the Gauss–Bonnet curvatures defined above, we have*

$$\int_M \langle X, G_{2r}(g)\rangle dv_g = 0.$$

Remark 1.4. When (M, g) is locally conformally flat, we see that the Gauss curvature satisfies

$$G_{2r}(g) = \frac{4^r (n-r)! r!}{(n-2r)!} \sigma_r,$$

so Theorem 1.3 reduces to Theorem A.

Remark 1.5. M. Labbi [2008] proved that the first variation of the functional $\int_M G_{2r} dv_g$ within metrics with constant volume gave the so-called generalized Einstein metric, and this functional has the variational property for $2r < n$ and is a topological invariant for $2r = n$. In fact, if $n = 2r$, this functional is the Gauss–Bonnet integrand up to a constant [Chern 1944].

In the next section, we first provide a general proof for Theorem 1.1 by adapting an ingredient in a preprint version [Han 2006b] of [Han 2006a], and using of a variation formula for $v^{(2k)}(g)$ established in [Graham 2009] and [Chang and Fang 2008]. Because of the explicit expression for $v^{(6)}(g)$ and potential applications to other related problems in low dimensions, we provide in Section 3 a self-contained proof for Theorem 1.1 in the case $k=3$. We prove Theorem 1.3 in Section 4.

2. Proof of Theorem 1.1

We will need the following variation formula for $v^{(2k)}(g)$; see [Graham 2009].

Proposition 2.1. *Under the conformal transformation $g_t = e^{2t\eta}g$, the variation of $v^{(2k)}(g_t)$ is given by*

$$\frac{\partial}{\partial t}\Big|_{t=0} v^{(2k)}(g_t) = -2k\eta v^{(2k)} + \nabla_i(L^{ij}_{(k)}\eta_j), \tag{2-1}$$

where $L^{ij}_{(k)}$ is defined as in [Graham 2009] *by*

$$L^{ij}_{(k)} = -\sum_{l=1}^{k}\frac{1}{l!}v^{(2k-2l)}(g)\partial_\rho^{l-1}g^{ij}(\rho)\big|_{\rho=0},$$

with $g_{ij}(\rho)$ denoting the extension of g such that

$$g_+ = \frac{(d\rho)^2 - 2\rho g(\rho)}{4\rho^2}$$

is an asymptotic solution to $\mathrm{Ric}(g_+) = -ng_+$ *near* $\rho = 0$.

An integral version of (2-1) first appeared in [Chang and Fang 2008]:

$$\int_M \Big(\frac{\partial}{\partial t}\Big|_{t=0}(v^{(2k)}(g_t)) + 2k\eta v^{(2k)}(g)\Big)dv_g = 0.$$

Proof of Theorem 1.1 in the case $n \neq 2k$. Let X be a conformal vector field on M. Let ϕ_t denote the local one-parameter family of conformal diffeomorphisms of (M, g) generated by X. Thus for some smooth function ω_t on M, we have

$$\phi_t^*(g) = e^{2\omega_t}g =: g_t.$$

We have the properties

$$\phi_t^* v^{(2k)}(g) = v^{(2k)}(\phi_t^* g) = v^{(2k)}(e^{2\omega_t}g), \tag{2-2}$$

$$\dot\omega := \frac{d}{dt}\Big|_{t=0}\omega_t = \frac{\mathrm{div}X}{n}, \tag{2-3}$$

$$\frac{\partial}{\partial t}\Big|_{t=0}(g_t^{-1}\circ A(g_t)) = -\nabla^2\dot\omega - 2\dot\omega g^{-1}\circ A(g), \tag{2-4}$$

$$\frac{\partial}{\partial t}\Big|_{t=0}\mathrm{div}_{g_t} X = nX\eta = n\langle X, \nabla\eta\rangle. \tag{2-5}$$

Using (2-2), (2-3), and (2-1), we have

$$\begin{aligned}\langle X, \nabla v^{(2k)}(g)\rangle &= \frac{\partial}{\partial t}\Big|_{t=0}(v^{(2k)}(g_t))\\ &= -2k\dot{\omega}v^{(2k)} + \nabla_i(L^{ij}_{(k)}\dot{\omega}_j)\\ &= -\frac{2k}{n}(\operatorname{div} X)v^{(2k)} + \nabla_i(L^{ij}_{(k)}\dot{\omega}_j)\\ &= -\frac{2k}{n}\operatorname{div}(v^{(2k)}X) + \frac{2k}{n}\langle X, \nabla v^{(2k)}(g)\rangle + \frac{1}{n}\nabla_i(L^{ij}_{(k)}(\operatorname{div} X)_j),\end{aligned}$$

from which it follows that

$$\Big(1-\frac{2k}{n}\Big)\langle X, \nabla v^{(2k)}(g)\rangle = -\frac{2k}{n}\operatorname{div}(v^{(2k)}X) + \frac{1}{n}\nabla_i(L^{ij}_{(k)}(\operatorname{div} X)_j). \tag{2-6}$$

Theorem 1.1 in the case $2k \neq n$ follows directly by integrating (2-6) over M. □

Proof of Theorem 1.1 in the case $2k = n$. As in [Han 2006b], we will prove that for any conformal metric $g_1 = e^{2\eta}g$ of g,

$$\int_M \langle X, v^{(2k)}(g_1)\rangle dv_{g_1} = \int_M \langle X, v^{(2k)}(g)\rangle dv_g = -\int_M \operatorname{div}_g X v^{(2k)}(g) dv_g,$$

that is, $\int_M \langle X, v^{(2k)}(g)\rangle dv_g$ is independent of the particular choice of metric in the conformal class. We only have to prove that

$$\frac{\partial}{\partial t}\Big|_{t=0}\int_M \operatorname{div}_{g_t} X v^{(2k)}(g_t) dv_{g_t} = 0 \quad \text{for } g_t = e^{2t\eta}g. \tag{2-7}$$

We prove (2-7) by direct computations using Proposition 2.1. Indeed,

$$\begin{aligned}&\frac{\partial}{\partial t}\Big|_{t=0}\int_M \operatorname{div}_{g_t} X v^{(2k)}(g_t) dv_{g_t}\\ &\quad= \int_M \big(n\langle X, \nabla\eta\rangle v^{(2k)} + \operatorname{div} X(-2k\eta v^{(2k)} + \nabla_i(L^{ij}_{(k)}\eta_j)) + n\eta \operatorname{div} X v^{(2k)}\big) dv_g\\ &\quad= \int_M \big(n\langle X, \nabla\eta\rangle v^{(2k)} + \operatorname{div} X\nabla_i(L^{ij}_{(k)}\eta_j)\big) dv_g\\ &\quad= \int_M \big(\langle n v^{(2k)}X, \nabla\eta\rangle - L^{ij}_{(k)}(\operatorname{div} X)_i\eta_j\big) dv_g\\ &\quad= \int_M \big(-\operatorname{div}(n v^{(2k)}X) + \nabla_j(L^{ij}_{(k)}(\operatorname{div} X)_i)\big)\eta dv_g = 0\end{aligned}$$

in the case $n = 2k$ by (2-6).

The remaining argument is an adaptation of an argument of Bourguignon and Ezin [1987]: either the connected component of the identity of the conformal group $C_0(M, g)$ is compact, and then there is a metric $\hat{g}$ conformal to g admitting $C_0(M, g)$ as a group of isometries, from which it follows that $\operatorname{div}_{\hat{g}} X \equiv 0$ and therefore (1-4) holds; or, $C_0(M, g)$ is noncompact, and then by a theorem of

Obata and Ferrand, (M, g) is conformal to the standard sphere, in which case we can pick the canonical metric to compute the integral on the left hand side of (1-4) and conclude that it is zero. $\square$

3. A self-contained proof of Theorem 1.1 in the case $k = 3$

We aim to give a direct, self-contained derivation for a more explicit version of (2-1); namely, under conformal change of metric $g_t = e^{2t\eta} g$,

$$\frac{\partial}{\partial t}\Big|_{t=0} v^{(6)}(g_t) = -6v^{(6)}(g)\eta + \nabla^j\left(\left(\frac{T^{(2)}_{ij}(g)}{8} + \frac{B_{ij}(g)}{24(n-4)}\right)\nabla^i\eta\right), \tag{3-1}$$

where $T^{(2)}_{ij}(g)$ is the Newton tensor associated with A_g, as defined in Reilly [1977]:

Definition. For an integer $k \geq 0$, the k-th Newton tensor is

$$T^{(k)}_{ij} = \frac{1}{k!}\sum \delta^{j_1\cdots j_k j}_{i_1\cdots i_k i} A_{i_1 j_1}\cdots A_{i_k j_k},$$

where $\delta^{j_1\cdots j_k j}_{i_1\cdots i_k i}$ is the generalized Kronecker symbol.

With (3-1) we can repeat the proof in the last section to prove Theorem 1.1 in the case $k = 3$.

First we recall the transformation laws for the tensors B_{ij} and A_{ij} under conformal change of metric $g_t = e^{2t\eta} g$ — see [Chang and Fang 2008]:

$$\begin{aligned}
A_{ij}(g_t) &= A_{ij} - t\nabla^2_{ij}\eta + t^2\nabla_i\eta\nabla_j\eta - \tfrac{1}{2}t^2|\nabla\eta|^2_g g_{ij},\\
B_{ij}(g_t) &= e^{-2t\eta}\big(B_{ij} + (n-4)t(C_{ijk} + C_{jik})\nabla^k\eta + (n-4)t^2 W_{ikjl}\nabla^k\eta\nabla^l\eta\big),
\end{aligned}$$

where $C_{ijk} := A_{ij,k} - A_{ik,j}$ are the components of the *Cotton* tensor, with $A_{ij,k}$ the components of the covariant derivative of the Schouten tensor A_{ij}.

Thus

$$\begin{aligned}
\frac{\partial}{\partial t}\Big|_{t=0} A^{ij}(g_t) &= -\nabla^{ij}\eta - 4A^{ij}(g)\eta,\\
\frac{\partial}{\partial t}\Big|_{t=0} B_{ij}(g_t) &= (n-4)(C_{ijk} + C_{jik})\nabla^k\eta - 2\eta B_{ij}.
\end{aligned}$$

Proposition 3.1 [Viaclovsky 2000a; Han 2006b; Hu and Li 2004]. *We have*

(i) $k\sigma_k(g) = \sum_{i,j} T^{(k-1)}_{ij} A_{ij}$

(ii) $\sum_i T^{(k)}_{ii} = (n-k)\sigma_k(g)$.

(iii) $\sum_l \nabla^l W_{lijk} = -(n-3)C_{ijk}$.

Using the relation between $v^{(6)}$ and $\sigma_3(g)$, and with $A^{ij}B_{ij}$ as in (1-3), we find

$$\begin{aligned}
&-8\frac{\partial}{\partial t}\Big|_{t=0} v^{(6)}(g_t)\\
&= T^{(2)}_{ij}(g)\big(-\nabla^{ij}\eta-2\eta A^{ij}(g)\big)\\
&\qquad+\frac{1}{3(n-4)}\big(-B_{ij}(g)\nabla^{ij}\eta+(n-4)A^{ij}(g)(C_{ijk}+C_{jik})\nabla^k\eta-6\eta A^{ij}B_{ij}\big)\\
&= -6\Big(\sigma_3(g)+\frac{1}{3(n-4)}A^{ij}B_{ij}\Big)\eta-\Big(T^{(2)}_{ij}(g)+\frac{B_{ij}(g)}{3(n-4)}\Big)\nabla^{ij}\eta+\tfrac{2}{3}A^{ij}(g)C_{ijk}\nabla^k\eta\\
&= 48v^{(6)}(g)\eta-\nabla^j\Big(\Big(T^{(2)}_{ij}(g)+\frac{B_{ij}(g)}{3(n-4)}\Big)\nabla^i\eta\Big)\\
&\qquad+\Big(\sum_j\Big(T^{(2)}_{ij,j}(g)+\frac{B_{ij,j}(g)}{3(n-4)}\Big)+\tfrac{2}{3}A^{kl}C_{kli}\Big)\nabla^i\eta,
\end{aligned}$$

where we used (1-3) and Proposition 3.1(i). The following lemma implies that

$$\sum_j\Big(T^{(2)}_{ij,j}(g)+\frac{B_{ij,j}(g)}{3(n-4)}\Big)+\tfrac{2}{3}A^{kl}C_{kli}=0,$$

thus establishing (3-1).

Lemma 3.2. (i) $\displaystyle\sum_j T^{(2)}_{ij,j}=-A^{pq}C_{pqi}$.

(ii) $\displaystyle\sum_j B_{ij,j}=(n-4)A^{kl}C_{kli}$.

Proof of (i). In normal coordinates, we have

$$\sum_j T^{(2)}_{ij,j}=\sum\Big(\frac{1}{2!}\sum\delta^{j_1j_2j}_{i_1i_2i}A_{i_1j_1}A_{i_2j_2}\Big)_j=\sum\delta^{j_1j_2j}_{i_1i_2i}A_{i_1j_1}A_{i_2j_2,j}=-A^{pq}C_{pqi},$$

where we used

$$\delta^{j_1j_2j}_{i_1i_2i}=\begin{vmatrix}\delta_{i_1j_1}&\delta_{i_1j_2}&\delta_{i_1j}\\ \delta_{i_2j_1}&\delta_{i_2j_2}&\delta_{i_2j}\\ \delta_{ij_1}&\delta_{ij_2}&\delta_{ij}\end{vmatrix}$$

and $\sum_i A_{ii,j}=\sum_i A_{ij,i}$, itself a consequence of the second Bianchi identity. □

Proof of (ii). First, using Proposition 3.1(iii) and substituting R_{ij} in terms of A_{ij} in the definition of the Bach tensor B_{ij}, we obtain

$$\begin{aligned}
B_{ij}&=-\sum_k C_{ikj,k}+\sum_{k,l}A_{kl}W_{likj}\\
&=-\sum_k(A_{ik,jk}-A_{ij,kk})+\sum_{k,l}A_{kl}W_{likj}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\sum_j B_{ij,j}\\
&= -\sum_{j,k}(A_{ik,jkj} - A_{ij,kkj}) + \sum_{k,l,j}(A_{kl,j}W_{likj} + A_{kl}W_{likj,j})\\
&= -\sum_{j,k}(A_{ik,jkj} - A_{ik,jjk}) + \sum_{k,l,j} A_{kl,j}W_{likj} - (n-3)\sum_{k,l} A_{kl}C_{kil}\\
&= -\sum_{j,k,m}(A_{ik,m}R_{mjkj} + A_{im,j}R_{mkkj} + A_{mk,j}R_{mikj})\\
&\quad + \sum_{k,l,j} A_{kl,j}W_{likj} + (n-3)\sum_{k,l} A_{kl}C_{kli}\\
&= \sum_{j,k,m}(-A_{mk,j}R_{mikj} + A_{km,j}W_{mikj}) + (n-3)\sum_{k,l} A_{kl}C_{kil}\\
&= \sum_{j,k,m} A_{mk,j}(-A_{mk}g_{ij} + A_{mj}g_{ik} - g_{mk}A_{ij} + g_{mj}A_{ik}) + (n-3)\sum_{k,l} A_{kl}C_{kli}\\
&= \sum_{m,k}(-A_{mk,i}A_{mk} + A_{mi,k}A_{mk} - A_{mk,j}g_{mk}A_{ij} + A_{mj,k}g_{mk}A_{ij})\\
&\quad + (n-3)\sum_{k,l} A_{kl}C_{kli}\\
&= \sum_{m,k} A_{mk}(A_{mi,k} - A_{mk,i}) + (n-3)\sum_{k,l} A_{kl}C_{kli}\\
&= \sum_{m,k} A_{mk}C_{mik} + (n-3)\sum_{k,l} A_{kl}C_{kli}\\
&= (n-4)\sum_{k,l} A_{kl}C_{kli},
\end{aligned}$$

where we have used

$$R_{mikj} = W_{mikj} + A_{mk}g_{ij} - A_{mj}g_{ik} + g_{mk}A_{ij} - g_{mj}A_{ik}. \qquad \square$$

Proof of Theorem 1.1 in the special case $k = 3$. We use the notation of Section 2. Let ϕ_t be the local one-parameter family of conformal diffeomorphisms of (M, g) generated by X. For $g_t = \phi_t^*(g) = e^{2\omega_t}g$, similarly to (3-1), we have

$$\begin{aligned}
\langle X, v^{(6)}\rangle &= \frac{\partial}{\partial t}\Big|_{t=0} v^{(6)}(g_t)\\
&= -6v^{(6)}(g)\dot\omega + \sum_{i,j}\nabla^j\left(\left(\frac{T_{ij}^{(2)}(g)}{8} + \frac{B_{ij}(g)}{24(n-4)}\right)\nabla^i\dot\omega\right),
\end{aligned} \tag{3-2}$$

if $n \neq 2k$. Then integrating (3-2) we can get Theorem 1.1.

If $n = 2k$, then by use of (3-1) and (3-2), we can prove that $\int_M \langle X, v^{(6)}(g)\rangle dv_g$ is independent of the particular choice of the metric within the conformal class. The remainder of the proof repeats verbatim that of Section 2. □

4. Proof of Theorem 1.3

In this section, we will prove Theorem 1.3 using a method similar to the one used in Section 2. Let (M^n, g) be a compact Riemannian manifold, and denote by R_{ijkl} the Riemann curvature tensor in local coordinates. Define a tensor P_r by

$$P_r{}_i^j = \delta^{jj_1j_2\cdots j_{2r-1}j_{2r}}_{ii_1i_2\cdots i_{2r-1}i_{2r}} R^{i_1i_2}{}_{j_1j_2} \cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \quad \text{for } 2r \leq n,$$

where $\delta^{jj_1j_2\cdots j_{2r-1}j_{2r}}_{ii_1i_2\cdots i_{2r-1}i_{2r}}$ is the generalized Kronecker symbol.

Lemma 4.1. *The tensor P_r is divergence free, that is,*

$$P_r{}_{i,j}^j = 0 \quad \textit{for any } i.$$

This property was present in [Labbi 2008] and [Lovelock 1971], although with different notation and formalism. Since we define the tensor P_r explicitly as above, and the property of P_r in Lemma 4.1 is a direct consequence of the Bianchi identity, we include a proof here.

Proof. We have

$$\begin{aligned}
P_r{}_{i,j}^j &= r\delta^{jj_1j_2\ldots j_{2r-1}j_{2r}}_{ii_1i_2\ldots i_{2r-1}i_{2r}} R^{i_1i_2}{}_{j_1j_2,j} \cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \\
&= -r\delta^{jj_1j_2\cdots j_{2r-1}j_{2r}}_{ii_1i_2\cdots i_{2r-1}i_{2r}} R^{i_1i_2}{}_{j_2j,j_1} \cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \\
&\quad - r\delta^{jj_1j_2\cdots j_{2r-1}j_{2r}}_{ii_1i_2\cdots i_{2r-1}i_{2r}} R^{i_1i_2}{}_{jj_1,j_2} \cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \\
&= -2r\delta^{jj_1j_2\cdots j_{2r-1}j_{2r}}_{ii_1i_2\cdots i_{2r-1}i_{2r}} R^{i_1i_2}{}_{j_1j_2,j} \cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \\
&= -2P_r{}_{i,j}^j,
\end{aligned}$$

where we have used the second Bianchi identity. It then follows that $P_r{}_{i,j}^j = 0$. □

Lemma 4.2. *The generalized Kronecker symbol satisfies*

$$\sum_{i,j=1}^n \delta^i_j \delta^{jj_1\ldots j_r}_{ii_1\ldots i_r} = (n-r)\delta^{j_1\ldots j_r}_{i_1\ldots i_r} \quad \textit{for any } 1 \leq i_1, \ldots, j_r \leq n \textit{ and } r \leq n.$$

The proof follows by a direct calculation from the definition.

Let X be a conformal vector field, and denote by ϕ_t the one-parameter subgroup of diffeomorphisms generated by X. Then there exists a family of functions ω_t such that $g_t=\phi_t^* g=e^{2\omega_t}g$. We have (2-3), $\omega_0=0$, and

$$G_{2r}(g_t)=\phi_t^* G_{2r}(g). \tag{4-1}$$

Under the conformal change of metric $g_t=e^{2\omega_t}g$, we have the formula (see for example [Chow et al. 2006])

$$R^{ij}{}_{kl}(g_t)=e^{-2\omega_t}(R^{ij}{}_{kl}-(\alpha\odot g)^{ij}{}_{kl}), \tag{4-2}$$

where we denote $\alpha_{ij}=(\omega_t)_{ij}-(\omega_t)_i(\omega_t)_j+\frac{1}{2}|\nabla\omega_t|^2 g_{ij}$ for convenience (note that $(\omega_t)_{ij}$ is the covariant derivative with respect to the fixed metric g) and $\odot$ is the Kulkarni–Nomizu product, defined by

$$(\alpha\odot g)_{ijkl}=\alpha_{ik}g_{jl}+\alpha_{jl}g_{ik}-\alpha_{il}g_{jk}-\alpha_{jk}g_{il}.$$

From (4-2) we see that

$$\begin{aligned} G_{2r}(g_t)=e^{-2r\omega_t}&\delta^{j_1 j_2\dots j_{2r-1}j_{2r}}_{i_1 i_2\dots i_{2r-1}i_{2r}} \\ &\cdot\big(R^{i_1 i_2}{}_{j_1 j_2}-(\alpha\odot g)^{i_1 i_2}{}_{j_1 j_2}\big)\cdots\big(R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}}-(\alpha\odot g)^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}}\big).\end{aligned} \tag{4-3}$$

Taking derivative with respect to t on both sides of (4-1) and using (4-3), we see by using (2-3) that

$$\begin{aligned} &\langle X, G_{2r}(g)\rangle \\ &=\frac{\partial}{\partial t}\Big|_{t=0}G_{2r}(g_t) \\ &=-2r\dot\omega G_{2r}(g)-r\delta^{j_1 j_2\cdots j_{2r-1}j_{2r}}_{i_1 i_2\cdots i_{2r-1}i_{2r}}\left(\frac{\partial\alpha}{\partial t}\Big|_{t=0}\odot g\right)^{i_1 i_2}{}_{j_1 j_2}R^{i_3 i_4}{}_{j_3 j_4}\cdots R^{i_{2r-1}i_{2r}}{}_{j_{2r-1}j_{2r}} \\ &=-2r\dot\omega G_{2r}(g)-4r(n-2r+1)P_{r-1}{}_i^j\,\dot\omega^i_j \\ &=-2r\frac{\operatorname{div} X}{n}G_{2r}(g)-\frac{4r(n-2r+1)}{n}P_{r-1}{}_i^j(\operatorname{div} X)^i{}_j \\ &=-2r\frac{\operatorname{div} X}{n}G_{2r}(g)-\frac{4r(n-2r+1)}{n}\nabla_j\big(P_{r-1}{}_i^j(\operatorname{div} X)^i\big).\end{aligned} \tag{4-4}$$

where we have used Lemma 4.2 in the third equality and Lemma 4.1 in the last. Integrating (4-4) over M and using the divergence theorem, we see that

$$\int_M\langle X, G_{2r}(g)\rangle dv=-2r\int_M\frac{\operatorname{div} X}{n}G_{2r}(g)dv=\frac{2r}{n}\int_M\langle X, G_{2r}(g)\rangle dv, \tag{4-5}$$

Hence, if $n>2r$, it follows from (4-5) that $\int_M\langle X, G_{2r}(g)\rangle dv=0$. If $n=2r$, we follow ideas in Section 2, that is, we need to prove that the integral

$$\int_M G_{2r}(g)\operatorname{div}_g X dv_g,$$

is independent of a particular choice of metric within a conformal class. Let $g_1 = e^{2\eta} g (\eta \in C^\infty(M))$ be any metric in the conformal class $[g]$. Considering a family of metrics $g_t = e^{2t\eta} g$ connecting g and g_1, we need to prove that

$$\frac{\partial}{\partial t}\Big|_{t=0} \int_M G_{2r}(g_t) \operatorname{div}_{g_t} X dv_{g_t} = 0.$$

By a direct computation, we have

$$\begin{aligned}
&\frac{\partial}{\partial t}\Big|_{t=0} \int_M G_{2r}(g_t) \operatorname{div}_{g_t} X dv_{g_t} \\
&\quad = \int_M \Big(\frac{\partial}{\partial t}\Big|_{t=0} G_{2r}(g_t) \operatorname{div} X + G_{2r}(g) \frac{\partial}{\partial t}\Big|_{t=0} \operatorname{div}_{g_t} X + n\eta G_{2r}(g) \operatorname{div} X \Big) dv_g \\
&\quad = \int_M \big(-2r\eta G_{2r}(g) \operatorname{div} X - 4r(n-2r+1) P_{r-1}{}_i^j \eta^i{}_j \operatorname{div} X \\
&\qquad\qquad\qquad\qquad + nG_{2r}(g)\langle \nabla\eta, X\rangle + nG_{2r}(g) \operatorname{div} X\eta \big) dv_g \\
&\quad = \int_M \big(-2r\eta G_{2r}(g) \operatorname{div} X - 4\eta r(n-2r+1) P_{r-1}{}_i^j (\operatorname{div} X)^i_j \\
&\qquad\qquad\qquad\qquad - n\eta \langle \nabla G_{2r}(g), X\rangle \big) dv_g \\
&\quad = 0,
\end{aligned}$$

where we have used (2-5) in the second equality, the divergence theorem in the third and (4-4) in the last. The remainder of the proof follows the idea of [Bourguignon and Ezin 1987] as in Section 2. Hence we complete the proof of Theorem 1.3.

References

[Bourguignon and Ezin 1987] J.-P. Bourguignon and J.-P. Ezin, "Scalar curvature functions in a conformal class of metrics and conformal transformations", *Trans. Amer. Math. Soc.* **301**:2 (1987), 723–736. MR 88e:53054 Zbl 0622.53023

[Chang and Fang 2008] S.-Y. A. Chang and H. Fang, "A class of variational functionals in conformal geometry", *Int. Math. Res. Not.* **2008**:7 (2008), Art. ID rnn008. MR 2009h:53072 Zbl 1154.53019

[Chern 1944] S.-s. Chern, "A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds", *Ann. of Math.* (2) **45** (1944), 747–752. MR 6,106a Zbl 0060.38103

[Chow et al. 2006] B. Chow, P. Lu, and L. Ni, *Hamilton's Ricci flow*, Graduate Studies in Mathematics **77**, American Mathematical Society, Providence, RI, 2006. MR 2008a:53068 Zbl 1118.53001

[Graham 2009] C. R. Graham, "Extended obstruction tensors and renormalized volume coefficients", *Adv. Math.* **220**:6 (2009), 1956–1985. MR 2010e:53060 Zbl 1161.53062

[Graham and Juhl 2007] C. R. Graham and A. Juhl, "Holographic formula for Q-curvature", *Adv. Math.* **216**:2 (2007), 841–853. MR 2009a:53062 Zbl 1147.53030

[Han 2006a] Z.-C. Han, "A Kazdan–Warner type identity for the σ_k curvature", *C. R. Math. Acad. Sci. Paris* **342**:7 (2006), 475–478. MR 2006j:53045 Zbl 1099.53028

[Han 2006b] Z.-C. Han, "A Kazdan–Warner type identity for the σ_k curvature", preprint, 2006, available at http://www.math.rutgers.edu/~zchan/current-preprint/KW.pdf.

[Hu and Li 2004] Z. Hu and H. Li, "A new variational characterization of n-dimensional space forms", *Trans. Amer. Math. Soc.* **356**:8 (2004), 3005–3023. MR 2005d:53058 Zbl 1058.53029

[Labbi 2008] M.-L. Labbi, "Variational properties of the Gauss–Bonnet curvatures", *Calc. Var. Partial Differential Equations* **32**:2 (2008), 175–189. MR 2009a:58013 Zbl 1139.58009

[Lovelock 1971] D. Lovelock, "The Einstein tensor and its generalizations", *J. Mathematical Phys.* **12** (1971), 498–501. MR 43 #1588 Zbl 0213.48801

[Reilly 1977] R. C. Reilly, "Applications of the Hessian operator in a Riemannian manifold", *Indiana Univ. Math. J.* **26**:3 (1977), 459–472. MR 57 #13799 Zbl 0391.53019

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

[Viaclovsky 2000b] J. A. Viaclovsky, "Some fully nonlinear equations in conformal geometry", pp. 425–433 in *Differential equations and mathematical physics* (Birmingham, AL, 1999), edited by R. Weikard and G. Weinstein, AMS/IP Stud. Adv. Math. **16**, Amer. Math. Soc., Providence, RI, 2000. MR 2001i:53057 Zbl 1161.53346

Received May 18, 2010.

BIN GUO
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING 100084
CHINA
Current address:
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NJ 08854
UNITED STATES

bguo@math.rutgers.edu

ZHENG-CHAO HAN
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NJ 08854
UNITED STATES

zchan@math.rutgers.edu

HAIZHONG LI
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING 100084
CHINA

hli@math.tsinghua.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

GONALITY OF A GENERAL ACM CURVE IN $\mathbb{P}^3$

ROBIN HARTSHORNE AND ENRICO SCHLESINGER

Let C be an ACM (projectively normal) nonsingular curve in $\mathbb{P}^3_{\mathbb{C}}$ not contained in a plane, and suppose C is general in its Hilbert scheme — this is irreducible once the postulation is fixed. Answering a question posed by Peskine, we show the gonality of C is $d - l$, where d is the degree of the curve and l is the maximum order of a multisecant line of C. Furthermore $l = 4$ except for two series of cases, in which the postulation of C forces every surface of minimum degree containing C to contain a line as well. We compute the value of l in terms of the postulation of C in these exceptional cases. We also show the Clifford index of C is equal to $\mathrm{gon}(C) - 2$.

1. Introduction

Let C be a nonsingular projective curve over an algebraically closed field $\mathbb{K}$. The *gonality* of C, written $\mathrm{gon}(C)$, is the minimum degree of a surjective morphism $C \to \mathbb{P}^1$, or equivalently the minimum positive integer k such that there exists a g^1_k on C.

For curves of genus $g \geq 1$ the gonality varies between 2, the value it takes on hyperelliptic curves, and $\left[\frac{1}{2}(g+3)\right]$, which by Brill–Noether theory is the gonality of a general curve of genus g. It may be regarded as the most fundamental invariant of the algebraic structure of C after the genus, providing a stratification of the moduli space of curves of genus g.

When a curve is embedded in some projective space, it is natural to wonder whether the gonality may be related to extrinsic properties of the curve. Here is a classical result in this direction, already known to Noether [Ciliberto 1984; Hartshorne 1986]:

Theorem 1.1. *A smooth curve $C \subset \mathbb{P}^2$ of degree $d \geq 3$ has gonality* $\mathrm{gon}(C) = d - 1$, *and any morphism $C \to \mathbb{P}^1$ of degree $d-1$ is obtained by projecting C from one of its points.*

Hartshorne was partially supported by Gnsaga - Programma Professori visitatori. Schlesinger was partially supported by MIUR PRIN 2005 *Spazi di moduli e teoria di Lie* and PRIN 2007: *Moduli, strutture geometriche e loro applicazioni*.

MSC2000: 14H50, 14H51.

Keywords: gonality, Clifford index, ACM space curves, multisecant lines.

See [Hartshorne 2002] for a proof and references. It is a simple exercise to prove the statement using the method of [Lazarsfeld 1986], which associates a vector bundle on $\mathbb{P}^2$ to a basepoint-free pencil on C. It is this method that we will exploit in the proof of our result.

One may ask a similar question for a curve $C \subset \mathbb{P}^3$. If L is a line in $\mathbb{P}^3$, projection from L induces a morphism $\pi_L : C \to \mathbb{P}^1$, whose degree is the degree of C minus the number of points of intersection of C and L. Thus the morphisms π_L of minimal degree are those corresponding to maximal order multisecant lines. We define

$$l = l(C) = \max\{\deg(C \cap L) : L \text{ a line in } \mathbb{P}^3\}$$

By analogy with the plane curves case one might wonder whether

$$\operatorname{gon}(C) = \deg(C) - l(C) \tag{1-1}$$

for a curve in $\mathbb{P}^3$, in which case following the terminology of [Hartshorne 2002] we say the gonality of C is computed by multisecants. Of course, this is usually not the case. For example, a general curve of genus g has gonality $\left[\frac{1}{2}(g+3)\right]$ and can be embedded in $\mathbb{P}^3$ as a nonspecial linearly normal curve of degree $g+3$. Since the Grassmannian of lines in $\mathbb{P}^3$ has dimension 4, and the set of lines meeting C is a codimension one subvariety, one expects $l(C)$ to be 4, and so

$$\deg(C) - l(C) = g - 1 > \left[\tfrac{1}{2}(g+3)\right] = \operatorname{gon}(C).$$

See [Hartshorne 2002, Examples 2.8 and 2.9] for specific counterexamples.

On the other hand, if the embedding of C in $\mathbb{P}^3$ is very special, one may hope the gonality of C is computed by multisecants. In this vein Peskine raised the question:

Question 1.2. *If C is a smooth ACM curve in $\mathbb{P}^3$, is its gonality computed by multisecants*?

Here ACM means arithmetically Cohen–Macaulay: a curve in $\mathbb{P}^3$ is ACM if the natural maps $H^0(\mathbb{P}^3, \mathcal{O}(n)) \to H^0(C, \mathcal{O}_C(n))$ are surjective for every $n \geq 0$.

Some special cases have been treated in the literature. Early results about uniqueness of the linear series $|\mathcal{O}_C(1)|$ for complete intersections and other ACM curves are in [Ciliberto and Lazarsfeld 1984]. Basili [1996] has proven that the gonality of a smooth complete intersection is indeed computed by multisecants. Ellia and Franco [2001] showed that the maximum order l of a multisecant to a general complete intersection of type (a, b) is 4 if $a \geq b \geq 4$ as one expects. Lazarsfeld [1997, 4.12] finds lower bounds for the gonality of a complete intersection curve in $\mathbb{P}^n$.

Results from [Martens 1996] and [Ballico 1997] show that the gonality of a smooth curve $C \subset \mathbb{P}^3$ on a smooth quadric surface is computed by multisecants.

In [Hartshorne 2002] it is shown that if a smooth curve $C \subset \mathbb{P}^3$ is ACM, lies on a smooth cubic surface X, and is general in its linear system on X, then its gonality is computed by multisecants. Farkas [2001] has shown that smooth ACM curves $C \subset \mathbb{P}^3$ lying on certain smooth quartic surfaces that do not contain rational or elliptic curves have gonality computed by multisecants.

In this paper, we show that, with the exception of very few cases we cannot decide, the gonality of a *general* ACM curve is indeed computed by multisecants. We have to make sense of the expression general ACM curve. To obtain an irreducible parameter space for ACM curves one needs to fix the *Hilbert function*, that is, the sequence of integers $h^0(\mathcal{O}_C(n))$. This is more conveniently expressed by its second difference or h-vector:

$$h_C(n) = h^0(\mathcal{O}_C(n)) - 2h^0(\mathcal{O}_C(n-1)) + h^0(\mathcal{O}_C(n-2)).$$

which has the advantage of being finitely supported while still nonnegative. We will denote by $A(h)$ the Hilbert scheme parametrizing ACM curves in $\mathbb{P}^3$ with h-vector h. By a theorem of Ellingsrud (see Remark 6.4), the Hilbert scheme $A(h)$ is *smooth and irreducible*. Thus by a general ACM curve we will mean a curve in a Zariski open nonempty subset of $A(h)$. We believe it is reasonable to assume that C is general in the statement of our theorem, because it might happen that a special ACM curve had a low degree pencil unrelated to the line bundle $\mathcal{O}_C(1)$.

Theorem 1.3. *Assume $\mathbb{K} = \mathbb{C}$ is the field of complex numbers. Let $C \subset \mathbb{P}^3$ be a nonplanar smooth ACM curve. If C is general in the Hilbert scheme $A(h_C)$, then*

$$\operatorname{gon}(C) = d - l,$$

where $d = \deg(C)$ and $l = l(C)$ is the maximum order of a multisecant line to C, except perhaps when the degree d, the genus g and the least degree s of a surface containing C form one of the following triples: (15, 26, 5), (16, 30, 5), (21, 50, 6), (22, 55, 6), (23, 60, 6), (28, 85, 7), (29, 91, 7), (36, 133, 8).

For curves C contained in a quadric or a cubic surface, the statement follows from the references cited above. So our contribution is for curves not lying on a cubic surface.

We can also determine the integer $l(C)$ in terms of the h-vector of C. Most of the time $l(C) = 4$, with two families of exceptions. These exceptional cases arise because the h-vector forces surfaces of minimal degree containing C to contain a line as well; this line is then a multisecant of order higher than expected. If s as above denotes the *least degree* of a surface containing C, we let

$$t = \min\{n : h^0(\mathcal{I}_C(n)) - h^0(\mathcal{O}_{\mathbb{P}^3}(n-s)) > 0\},$$

so that (s,t) is the smallest type of a complete intersection containing C. We denote by e the *index of speciality* of C: $e = \max\{n : h^1\mathcal{O}_C(n) > 0\}$.

The value of $l(C)$ is given as follows:

Theorem 1.4. *Let $C \subset \mathbb{P}^3_{\mathbb{C}}$ be a general smooth ACM curve with $s \geq 4$. Let $l = l(C)$ denote the maximum order of a multisecant line to C. Then $l = 4$, unless*

- *the h-vector of C satisfies $h(e+1) = 3$ and $h(e+2) = 2$, in which case $l = e+3$ and C has a unique $(e+3)$-secant line, or*
- *$t > s+3$ and the h-vector of C satisfies $h(t) = s-2$ and $h(t+1) = s-3$, but not $h(e+1) = 3$, $h(e+2) = 2$, in which case $l = t-s+1$ and C has a unique $(t-s+1)$-secant line.*

Nollet [1998] has found a sharp bound for the maximal order $l = l(C)$ of a multisecant line in terms of the h-vector of C, valid for *any* irreducible ACM curve. If C is not a complete intersection, the bound is the largest integer n for which $h_C(n-1) - h_C(n) > 1$. Since this number is at least s, we see that $l(C)$ and the gonality of C *vary* in the family $A(h)$, provided $s \geq 5$, and the gonality of the general curve is $d - 4$ (in fact the argument of Theorem 4.1 shows that $l(C)$ varies in the linear system $|C|$ on a smooth surface X of degree $s \geq 5$). On the other hand, in the special case $h(e+1) = 3$ and $h(e+2) = 2$, then Nollet's bound is precisely $e+3$, so that $l(C)$ is constant in $A(h)$.

Finally, in most cases we can prove that *every* pencil computing the gonality of C arises from a maximum order multisecant: the finite list of exceptions is given in Theorem 9.1. In particular, C has a finite number of pencils of minimal degree, and therefore its Clifford index is $\mathrm{Cliff}(C) = \mathrm{gon}(C) - 2 = d - l(C) - 2$.

It would be interesting to investigate linear series g^r_k on general ACM curves also for $r \geq 2$. For results in this direction we refer to [Lopez and Pirola 1995].

Outline of proof and structure of the paper. Since the conclusions of our result are semicontinuous on the Hilbert scheme $A(h)$, it suffices to show the existence of a *single* curve C for which the result holds. Let C be a smooth ACM curve in $\mathbb{P}^3$ with given h-vector h, not lying on any surface of degree at most 3. In Section 3 we review the classical result that for every smooth space curve D of degree at least 10 there exists a line L that is at least a 4-secant line of D. Thus $\mathrm{gon}(C) \leq d-4$. Next, if C is general in $A(h)$, it is contained in a *smooth* surface X of degree s. We prove in Corollary 4.2 that, if C is general in its linear system on X and L is an l-secant line of C with $l \geq 5$, then L is contained in X. In fact, we prove a slightly more general result, which gives explicit conditions for a space curve not to have 5-secant lines:

Theorem 4.1. *Let $C \subset \mathbb{P}^3_{\mathbb{K}}$ be a curve contained in an irreducible surface X of degree s. Suppose C is a Cartier divisor on X and*

$$H^0(\mathbb{P}^3, \mathcal{I}_C(s-2)) = 0, \quad H^1(\mathbb{P}^3, \mathcal{I}_C(m)) = 0 \text{ for } m = s-2, s-3, s-4.$$

If C is general in its linear system on X, then $\deg(C \cap L) \leq 4$ *for every line L not contained in X, and C has only finitely many* 4*-secant lines not contained in X.*

In particular, if X does not contain a line, then C does not have an l-secant line for any $l \geq 5$.

At this point to prove our main theorem we need to show that every pencil of minimal degree arises from a multisecant line. The proof uses the technique from [Lazarsfeld 1986], which associates to a basepoint-free pencil on C a vector bundle $\mathscr{E}$ on the surface X, as explained in Section 5. In Section 6 we review enough liaison theory for ACM curves to be able to show that the bundle $\mathscr{E}$ is Bogomolov unstable. Thus it has a destabilizing divisor $A \in \operatorname{Pic}(X)$, whose degree $x = A.H$ satisfies stringent numerical restrictions in terms of the intersection numbers A^2, $A.C$ and C^2.

To use these constraints effectively we need to control the Picard group of X. The hypothesis that the ground field is $\mathbb{C}$ allows us to apply the Noether–Lefschetz type theorem of [Lopez 1991, II.3.1] or the more recent [Brevik and Nollet 2008] to conclude that, if C is general in $A(h)$ and X is very general among surfaces of minimal degree containing C, then $\operatorname{Pic}(X)$ is freely generated by H and the irreducible components of a curve Γ that is general among curves minimally linked to C. Such a Γ is a general ACM curve, but it may not be irreducible. Thus we are led to establish a structure theorem for general ACM curves. Section 7 is devoted to the proof of this result. It generalizes Gruson–Peskine's theorem [1978], according to which the general ACM curve in $A(h)$ is smooth and irreducible if h is of decreasing type ("has no gaps"):

Theorem 7.21. *Let $A(h)$ denote the Hilbert scheme parametrizing ACM curves in $\mathbb{P}^3_{\mathbb{K}}$ with h-vector h. If Γ is general in $A(h)$, then*

$$\Gamma = D_1 \cup D_2 \cup \cdots \cup D_r,$$

where $r-1$ is the number of Gruson–Peskine gaps of h, and the D_i are distinct smooth irreducible ACM curves whose h-vectors are determined by the gap decomposition of h as explained in Section 7. Furthermore, for every $1 \leq i_1 < i_2 < \cdots < i_h \leq r$, the curve

$$D_{i_1} \cup D_{i_2} \cup \cdots \cup D_{i_h}$$

is still ACM.

Thus we can write the destabilizing divisor as $A = aH + \sum a_i D_i$. In the proof of the main Theorem 9.1, using the fact that the curves D_i and their unions are ACM, together with the numerical constraints on $x = A.H$ we show $-s-1 \le x < 0$. We then play this inequality against the bounds of Corollary 8.9, which are essentially upper bounds for the genus of an ACM curve lying on X in terms of the degree of the curve and of degree of X. In fact, these bounds are a refinement of the bounds for the genus of an ACM curve proven in [Gruson and Peskine 1978] (see Remark 8.8). The end result is that there are only two possibilities for A: either $-A = H$ (the plane section) or $-A = H - L$ for some line L on X.

Corollary 5.7 shows that in case $A = -H$ the pencil arises from a multisecant line not contained in X, while in case $A = L - H$ the pencil arises from L. This shows pencils of minimal degree on C all arise from multisecant lines, thus completing the proof of the theorem.

2. Notation and terminology

A linear system of degree k and projective dimension r on C is denoted with the symbol g^r_k, and a g^1_k is called a pencil. The *gonality* of C, written $\mathrm{gon}(C)$, is the least positive integer k such that there exists a g^1_k on C. Since a pencil of least degree is automatically basepoint-free, the gonality of C is the least degree of a surjective morphism $C \to \mathbb{P}^1$. One can further notice that a g^1_k with $k = \mathrm{gon}(C)$ is complete, so that $h^0(C, \mathscr{O}_C(Z)) = 2$ for every divisor Z in the pencil.

Definition 2.1. Assume $C \subset \mathbb{P}^3$ is a nonplanar curve. Given a line L, let $\pi_L : C \to \mathbb{P}^1$ be obtained projecting C from L, and let $\mathfrak{X}(L)$ denote the g^1_k corresponding to π_L. Note that $\mathfrak{X}(L)$ is obtained from the pencil cut out on C by planes through L removing its base locus, which coincides with the scheme theoretic intersection $C \cap L$. In particular,

$$\deg(\pi_L) = \deg \mathfrak{X}(L) = \deg(C) - \deg(C.L)$$

and $\mathfrak{X}(L)$ is complete if $\deg(C.L) \ge 2$. We say that a g^1_k on C arises from a multisecant if it is of the form $\mathfrak{X}(L)$ for some line L. We say the gonality of C can be *computed by multisecants* if there exists a line L such that $\mathfrak{X}(L)$ has degree $\mathrm{gon}(C)$.

3. Existence of 4-secant lines

The following statement is classical and well known, but it seems hard to find a reference.

Proposition 3.1. *Let C be a smooth irreducible curve of degree $d \ge 10$ in $\mathbb{P}^3$. Then C has an l-secant line L with $l \ge 4$. In particular, the gonality of C is at most $d-4$.*

Proof. The statement is clear if $\deg(C) \ge 4$ and C is contained in a plane or $\deg(C) \ge 7$ and C is contained in a quadric surface. If C is not contained in a quadric surface, we will show the Cayley number of 4-secants

$$\mathscr{C}(d,g)=\frac{(d-2)(d-3)^2(d-4)}{12}-\frac{g(d^2-7d+13-g)}{2}$$

is positive. The existence of L then follows from intersection theory as explained in [Le Barz 1987] or in [Arbarello et al. 1985]. For fixed $d \geq 7$, the number $\mathscr{C}(d,g)$ is a decreasing function of g, because the partial derivative with respect to g is

$$g-\frac{d^2-7d+13}{2},$$

which is negative because $g \leq d^2/4-d+1$ when C is not contained in a plane.

But C is not even contained in a quadric surface; thus its genus is bounded above by $\frac{1}{6}d(d-3)+1$, and

$$\mathscr{C}(d,g)\geq \mathscr{C}\big(d,\tfrac{1}{6}d(d-3)+1\big)=\frac{d(d-3)(d-6)(d-9)}{72},$$

which is positive for $d \geq 10$. □

Remark 3.2. The result is sharp, because a smooth complete intersection of two cubic surfaces has degree 9 and no 4-secant line.

4. Nonexistence of 5-secant lines

Theorem 4.1. *Let $C \subset \mathbb{P}^3$ be a curve contained in an irreducible surface X of degree s. Suppose C is a Cartier divisor on X and*

$$H^0(\mathbb{P}^3,\mathscr{I}_C(s-2))=0,\qquad H^1(\mathbb{P}^3,\mathscr{I}_C(m))=0 \text{ for } m=s-2,s-3,s-4.$$

If C is general in its linear system on X, then $\deg(C.L)\leq 4$ *for every line L not contained in X, and C has only finitely many* 4*-secant lines not contained in X.*

In particular, if X does not contain a line, then C does not have an l-secant line for any $l \geq 5$.

Proof. The statement is obvious if $s \leq 3$, so assume $s \geq 4$. The hypotheses imply $h^1\mathbb{O}(D)=0$ for $D=C, C-H, C-2H$ because, by Serre duality,

$$h^1(\mathbb{P}^3,\mathscr{I}_C(m))=h^1(X,\mathbb{O}_X(mH-C))=h^1(X,\mathbb{O}_X(C+(s-4-m)H)).$$

Similarly, $H^2(\mathbb{O}_X(C-nH))$ is dual to

$$H^0(\mathbb{O}_X((s-4+n)H-C))=H^0(X,\mathscr{I}_{C,X}(s-4+n)),$$

which by assumption is zero for $n \leq 2$. Thus we see that $h^0\mathbb{O}_X(D)=\chi\mathbb{O}_X(D)$ for $D=C, C-H, C-2H$.

Let L be a line not contained in X, and let V be the scheme theoretic intersection of X and L. Then V has degree s, and there is an exact sequence

$$0 \to \mathcal{O}_X(-2H) \to \mathcal{O}_X(-H)^{\oplus 2} \to \mathcal{I}_{V,X} \to 0.$$

Twisting by $\mathcal{O}_X(C)$ and taking cohomology we see that

$$h^0(\mathcal{I}_V(C)) = 2h^0(\mathcal{O}_X(C-H)) - h^0(\mathcal{O}_X(C-2H)).$$

Therefore

$$\begin{aligned} h^0(\mathcal{O}_X(C)) - h^0(\mathcal{I}_V(C)) &= h^0(\mathcal{O}_X(C)) - 2h^0(\mathcal{O}_X(C-H)) + h^0(\mathcal{O}_X(C-2H)) \\ &= \chi(\mathcal{O}_X(C)) - 2\chi(\mathcal{O}_X(C-H)) + \chi(\mathcal{O}_X(C-2H)) = s. \end{aligned}$$

This shows that the points of V impose independent conditions on the linear system $|C|$. It follows that the family of curves in $|C|$ meeting L in a scheme of length $l \le s$ has codimension l in $|C|$. This implies the statement because L varies in a four-dimensional family. □

Corollary 4.2. *Let $C \subset \mathbb{P}^3$ be an ACM curve. Suppose that C is contained in a smooth surface $X \subset \mathbb{P}^3$ of degree $s = s_C$, and that C is general in its linear system on X. Then* $\deg(C.L) \le 4$ *for any line L not contained in X.*

In particular, if X does not contain a line, then C does not have an l-secant line for any $l \ge 5$.

Proof. The statement follows from Theorem 4.1 because C is ACM precisely when $H^1(\mathbb{P}^3, \mathcal{I}_C(m)) = 0$ for every m. □

5. Gonality of curves on a smooth surface: Lazarsfeld's method

In this section we explain a construction due to Lazarsfeld [1986; 1997] that will be crucial in proving that every pencil of minimal degree on a general ACM curve arises from a multisecant.

When a curve C is contained in a smooth surface X, we associate a rank two vector bundle on X to a basepoint-free g^1_k on C as follows. The basepoint-free g^1_k is determined by a degree k line bundle $\mathcal{O}_C(Z)$ on C, and a surjective map of $\mathcal{O}_C$-modules

$$\beta : \mathcal{O}_C^{\oplus 2} \to \mathcal{O}_C(Z).$$

(Note that, since $k \ge 1$, the map $H^0(\beta) : H^0(\mathcal{O}_C^{\oplus 2}) \to H^0(\mathcal{O}_C(Z))$ is injective.)

Definition 5.1. Suppose C is an integral curve on the smooth projective surface X, and $\mathfrak{X}$ is a basepoint-free pencil on C defined by $\beta : \mathcal{O}_C^{\oplus 2} \to \mathcal{O}_C(Z)$. Let

$$\alpha : \mathcal{O}_X^{\oplus 2} \to \mathcal{O}_C(Z)$$

denote the map obtained composing β with the natural surjection $\mathcal{O}_X^{\oplus 2} \to \mathcal{O}_C^{\oplus 2}$. Then the kernel $\mathcal{E}$ of α is called the *bundle associated* to the pencil $\mathfrak{X}$.

Proposition 5.2. *Let $\mathscr{E}$ be the bundle associated to a pencil of degree k on C as in the previous definition. Then*

(a) *$\mathscr{E}$ is a rank two vector bundle on X.*

(b) $H^0(\mathscr{E}) = 0$.

(c) $c_1(\mathscr{E}) = \mathscr{O}_X(-C)$ *and* $c_2(\mathscr{E}) = \deg(Z)$, *so that*

$$\Delta(\mathscr{E}) \stackrel{\text{def}}{=} c_1^2(\mathscr{E}) - 4c_2(\mathscr{E}) = C^2 - 4k.$$

(Here we consider the first Chern class as an element of $A^1(X) \cong \mathrm{Pic}(X)$, while we view the c_1^2 and c_2 as integers, via the degree map for zero cycles.)

Proof. By definition of $\mathscr{E}$ there is an exact sequence:

$$0 \to \mathscr{E} \to \mathscr{O}_X^{\oplus 2} \to \mathscr{O}_C(Z) \to 0$$

Since $\mathscr{O}_C$ has rank zero and projective dimension 1 as an $\mathscr{O}_X$-module, $\mathscr{E}$ is a rank two vector bundle on X, whose Chern classes can be computed from the above sequence. If $H^0(\mathscr{E})$ were not zero, then $H^0(\alpha) : H^0(\mathscr{O}_C^{\oplus 2}) \to H^0(\mathscr{O}_C(Z))$ would not be injective, so α would induce a surjective map $\mathscr{O}_C \to \mathscr{O}_C(Z)$, contradicting $\deg Z = k \geq 1$. □

We recall the definition of Bogomolov instability for rank two vector bundles on a surface, and Bogomolov's theorem which gives a numerical condition for instability.

Definition 5.3. Let $\mathscr{E}$ be a rank two vector bundle on X. One says that $\mathscr{E}$ is *Bogomolov unstable* if there exist a finite subscheme $W \subset X$ (possibly empty) and divisors A and B on X sitting in an exact sequence

$$0 \to \mathscr{O}_X(A) \to \mathscr{E} \to \mathscr{I}_W \otimes \mathscr{O}_X(B) \to 0. \tag{5-1}$$

where $(A - B)^2 > 0$ and $(A - B).H > 0$ for some (hence every) ample divisor H. We say A is a *destabilizing divisor* of $\mathscr{E}$. It is unique up to linear equivalence.

Theorem 5.4 ([Bogomolov 1978]; compare [Huybrechts and Lehn 1997, 7.3.3] and [Lazarsfeld 1997, 4.2]). *Suppose the ground field $\mathbb{K}$ has characteristic zero. Let $\mathscr{E}$ be a rank two vector bundle on the smooth projective surface X, and let $\Delta(\mathscr{E}) = c_1(\mathscr{E})^2 - 4c_2(\mathscr{E})$.*

If $\Delta(\mathscr{E}) > 0$, then $\mathscr{E}$ is Bogomolov unstable.

Following Lazarsfeld's approach, we will show in Section 6 that the bundle associated to a pencil computing the gonality of a smooth ACM curve satisfies $\Delta(\mathscr{E}) > 0$, hence it is Bogomolov unstable, and there is a destabilizing divisor A. To work effectively we will need the following technical result that will be useful in two ways. First it immediately implies that, when $-A = H$ (plane section) or

$-A = H - L$ (plane section minus a line), the given pencil arises from a multisecant; later on the inequalities $A^2 \geq 0$ and $A.H < 0$ will be used to exclude all other possibilities for A.

Proposition 5.5. *Suppose X is a smooth projective surface, C is an integral curve on X, and $|Z|$ is a complete basepoint-free pencil on C. Let $\mathscr{E}$ be the rank 2 bundle on X associated to $|Z|$. Suppose there is an exact sequence*

$$0 \to \mathcal{O}_X(A) \xrightarrow{h} \mathscr{E} \to \mathscr{I}_W \otimes \mathcal{O}_X(B) \to 0 \tag{5-2}$$

with W zero-dimensional and B not effective. Then the linear system $|-A|$ on X contains two effective curves D_1 and D_2 with the following properties:

(a) *D_1 and D_2 meet properly in a 0-dimensional scheme V containing W.*

(b) *D_1 and D_2 meet C properly, and, if R is the base locus of the pencil cut out on C by $C.D_1$ and $C.D_2$, then*

$$\mathcal{O}_C(Z) \cong \mathcal{O}_X(-A) \otimes \mathcal{O}_C(-R);$$

that is, the pencil $|Z|$ is obtained by first restricting D_1 and D_2 to C and then removing the base locus R.

(c) *R is the residual scheme to W in V, that is, there is an exact sequence*

$$0 \to \mathcal{O}_W \to \mathcal{O}_V \to \mathcal{O}_R \to 0.$$

In particular $h^0 \mathscr{I}_W(-A) \geq 2$, $A.H < 0$ for every ample divisor H, and $A^2 \geq 0$.

Remark 5.6. The proposition applies if $\mathscr{E}$ is Bogomolov unstable with destabilizing sequence (5-2). Indeed in this case, if H is an ample divisor on X, then $(A - B).H > 0$. Since $c_1(\mathscr{E}) = A + B = -C$ in $\mathrm{Pic}(X)$, we compute

$$-2B.H = (A - B).H + C.H > 0.$$

Therefore B is not effective.

Proof of Proposition 5.5. Dualizing $0 \to \mathscr{E} \to \mathcal{O}_X^{\oplus 2} \to \mathcal{O}_C(Z) \to 0$ we obtain an exact sequence

$$0 \to \mathcal{O}_X^{\oplus 2} \to \mathscr{E}(C) \to \mathcal{O}_C(C - Z) \to 0.$$

We now look at the composite map $g : \mathcal{O}_X^{\oplus 2} \to \mathscr{E}(C) \to \mathscr{I}_W(-A)$.

This map is nonzero, otherwise $\mathcal{O}_X^{\oplus 2}$ would map injectively into the kernel of $\mathscr{E}(C) \to \mathscr{I}_W(-A)$, which is $\mathcal{O}_X(C + A)$, absurd. Hence the image of g has rank one, and has the form $\mathscr{I}_Y(-A)$ for some proper subscheme $Y \subset X$ containing W. Then $\mathscr{I}_Y = \mathscr{I}_V(-D)$ where D is the divisorial part of Y, and V is zero dimensional. We obtain an exact sequence

$$0 \to \mathrm{Ker}(g) \to \mathcal{O}_X^{\oplus 2} \to \mathscr{I}_V(-A - D) \to 0.$$

It follows $\operatorname{Ker}(g) = \mathcal{O}_X(A+D)$ and $-A-D$ is effective. A diagram chase shows there is an exact sequence

$$0 \to \mathcal{O}_X(A+D) \to \mathcal{O}_X(C+A) \to \mathcal{O}_C(C-Z)$$

from which we see there is an effective curve C_0 linearly equivalent to $C-D$ contained in C. Since C is irreducible, this implies either $D = C$ or $D = 0$.

Now $-A-D$ is effective, so, if we had $D = C$, then $B = -A-C$ would be effective, contradicting the hypotheses. Hence the only possibility is $D = 0$.

Putting everything together we obtain a commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{O}_X(A) \arrow[r, "g^\vee"] \arrow[d, "h"] & \mathcal{O}_X^{\oplus 2} \arrow[r, "(s_1,s_2)"] \arrow[d, equal] & \mathcal{I}_V(-A) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathcal{E} \arrow[r] & \mathcal{O}_X^{\oplus 2} \arrow[r] & \mathcal{O}_C(Z) \arrow[r] & 0
\end{tikzcd}$$

Now let D_1 and D_2 the divisors defined by the sections s_1 and s_2 of $\mathcal{O}_X(-A)$. The first row of the diagram shows D_1 and D_2 meet properly in the zero dimensional scheme V, which contains W by construction. The two sections remain independent in $H^0(\mathcal{O}_C(Z))$ because $H^0(\mathcal{E}) = 0$. Hence D_1 and D_2 meet C properly, and $D_1.C$ and $D_2.C$ span a pencil on C.

By the snake lemma, the kernel of the vertical map $\mathcal{I}_V(-A) \to \mathcal{O}_C(Z)$ is $\mathcal{I}_W(B) = \mathcal{I}_W(-A-C)$, hence a diagram chase produces an exact sequence

$$0 \to \mathcal{O}_C(Z) \to \mathcal{O}_X(-A) \otimes \mathcal{O}_C \to \mathcal{O}_V/\mathcal{O}_W \to 0$$

which proves the rest of the statement. □

Corollary 5.7. *Assume $X \subset \mathbb{P}^3$ is a smooth surface with plane section H, containing a smooth irreducible curve C. Suppose C is not contained in a plane. Let $|Z|$ be a complete basepoint-free pencil on C, and let $\mathcal{E}$ be the bundle on X associated to $|Z|$.*

(a) *If there is an exact sequence*

$$0 \to \mathcal{O}_X(A) \to \mathcal{E} \to \mathcal{I}_W(B) \to 0$$

with W zero dimensional and $A+H$ effective, then there is a line L such that $|Z| = \mathfrak{L}(L)$ is the pencil cut out on C by planes through L. Furthermore, if X does not contain L, then $A = -H$ and W is the residual scheme to $C \cap L$ in $X \cap L$, while, if X contains L, then $A = L - H$ and W is empty.

(b) *Assume C is linearly normal and $|Z|$ is the pencil cut out on C by planes through a line L meeting C in a scheme of length at least 2. Then there exists an exact sequence as above with $A = -H$ if X does not contain L and $A = L - H$ if X contains L.*

Proof. (a) The divisor B is not effective; otherwise

$$B+(A+H)=(-A-C)+(A+H)=H-C$$

would be effective, which contradicts the assumption that C is not contained in plane.

Thus we may apply Proposition 5.5 to the given exact sequence to conclude the linear system $|-A|$ contains a pencil. By assumption $P=A+H$ is effective, and therefore in order that $|-A|=|H-P|$ may contain a pencil it is necessary that P be empty or a line.

If P is empty, by 5.5 the are two plane sections $D_1=H_1\cap X$ and $D_2=H_2\cap X$ of X meeting in a zero dimensional scheme V, hence the line $L=H_1\cap H_2$ is not contained in X. Proposition 5.5b shows $|Z|$ is obtained removing from the pencil spanned by $C\cap H_1$ and $C\cap H_2$ its base locus $C\cap L$, that is, $|Z|=\mathfrak{X}(L)$, and Proposition 5.5c shows W is the residual scheme to $C\cap L$ in $X\cap L$.

Finally, if P is a line, then D_1 and D_2 belong to $|H-P|$, hence their intersection $V=D_1\cap D_2$ is empty. It follows from Proposition 5.5 that $|Z|=\mathfrak{X}(P)$ and that and W is empty.

(b) By the definition of $\mathscr{E}$ there is an exact sequence

$$0\to\mathscr{E}\to\mathscr{O}_X^{\oplus 2}\to\mathscr{O}_C(Z)\to 0.$$

Comparing this sequence with

$$0\to\mathscr{O}_C\to\mathscr{O}_C(Z)\to\mathscr{O}_Z\to 0,$$

we obtain

$$0\to\mathscr{O}_X(-C)\to\mathscr{E}\to\mathscr{I}_{Z,X}\to 0.$$

Now twist by H and take cohomology to get a long exact sequence

$$0\to H^0(\mathscr{O}_X(H-C))\to H^0(\mathscr{E}(H))\to H^0(\mathscr{I}_{Z,X}(H))\to H^1(\mathscr{O}_X(H-C)).$$

Since Z is contained in a plane, $h^0(\mathscr{I}_{Z,X}(H))>0$, while $H^1(\mathscr{O}_X(H-C))=H^1(\mathscr{I}_C(H))=0$ because C is linearly normal. Hence $\mathscr{E}(H)$ has a section, and after removing torsion in the cokernel if necessary we find an exact sequence:

$$0\to\mathscr{O}_X(P-H)\to\mathscr{E}\to\mathscr{I}_W(H-P-C)\to 0,$$

with W zero dimensional and P effective. Now (b) follows from (a). $\square$

6. ACM curves

In this section we show that, if C is an ACM curve of degree d having a pencil of minimal degree $k\le d-4$ on a smooth surface of degree $s=s_C$, then the bundle $\mathscr{E}$ associated to the given pencil satisfies $\Delta(\mathscr{E})>0$ (except for a small list of cases given in Proposition 6.10); hence, if the ground field has characteristic zero, it is Bogomolov unstable. The proof is based on the structure of the biliaison class of ACM curves which we now briefly recall. We also include some information about the minimal link Γ of a curve C, which we will need later.

Given a curve C in $\mathbb{P}^3$ its fundamental numerical invariants are, besides its degree d_C and its arithmetic genus $g(C) = 1 - \chi(\mathcal{O}_C)$:

- its index of speciality $e(C) = \max\{n : h^1\mathcal{O}_C(n) > 0\}$;
- the minimal degree s_C of a surface containing C;
- the integer $t_C = \min\{n : h^0(\mathcal{I}_C(n)) - h^0(\mathcal{O}_{\mathbb{P}^3}(n - s_C)) > 0)\}$. If C is integral or more generally if C lies on an integral surface of degree s_C, the integer t_C is the smallest n such that C is contained in a complete intersection of two surfaces of degree s_C and n.

When C is ACM, all its basic numerical invariants can be computed from the Hilbert function. It is convenient to express the Hilbert function through its second difference function, the so called *h-vector* h_C of C — see [Migliore 1998, §1.4] — because h_C is a finitely supported function. Thus one defines

$$h_C(n) = h^0(\mathcal{O}_C(n)) - 2h^0(\mathcal{O}_C(n-1)) + h^0(\mathcal{O}_C(n-2)).$$

If $s = s(C)$ and $e = e(C)$, the function h_C satisfies

$$\text{(6-1)} \qquad \begin{cases} h(n) = n+1 & \text{if } 0 \le n \le s-1, \\ h(n) \ge h(n+1) & \text{if } n \ge s-1, \\ h(e+2) > 0 \quad \text{and} \quad h(n) = 0 & \text{for } n \ge e+3. \end{cases}$$

Thus we may write h as

$$h_C = \{1, 2, \dots, s, h_C(s), \dots, h_C(e+2)\}.$$

with $s = h_C(s-1) \ge h_C(s) \ge h_C(s+1) \ge \cdots \ge h_C(e+2)$.

We say that a finitely supported function $h : \mathbb{N} \to \mathbb{N}$ is an *h-vector* if it satisfies (6-1) for some $s \ge 1$. Every h-vector arises as the h-vector of an ACM curve in $\mathbb{P}^3$; see [Martin-Deschamps and Perrin 1990, Theorem V.1.3, p. 111] and Remark 7.7 below. It will be convenient to allow the identically zero function among h-vectors, and think of it as the h-vector of the empty curve. In terms of the h-vector, the fundamental invariants of C are:

Proposition 6.1. *For an ACM curve C in $\mathbb{P}^3$, with h-vector h_C, we have*

(1) $d_C = \sum h_C(n)$,

(2) $g(C) = 1 + \sum (n-1) h_C(n)$,

(3) $e(C) + 2 = \max\{n : h_C(n) > 0\}$,

(4) $s_C = \min\{n \geq 0 : h_C(n) < n+1\}$, *and*

(5) $t_C = \min\{n \geq 0 : h_C(n-1) > h_C(n)\}$.

Consistently with these formulas, for the empty curve we define $s = 0$, $d = 0$, $g = 1$, $e = -\infty$.

Remark 6.2. If C is an ACM curve with $s_C = s$, then

$$d_C = \sum h_C(n) \geq \sum_{n=0}^{s-1} (n+1) = \tfrac{1}{2}s(s+1).$$

The h-vectors of integral curves have a special form:

Definition 6.3 [Maggioni and Ragusa 1988]. An h-vector is of *decreasing type* if $h(a) > h(a+1)$ implies that for each $n \geq a$ either $h(n) > h(n+1)$ or $h(n) = 0$.

Remark 6.4. By a result from [Ellingsrud 1975] (see also [Martin-Deschamps and Perrin 1990, p. 5; corollaire 1.2 on p. 134; §1.7, p. 139]), the Hilbert scheme $A(h)$ of ACM curves in $\mathbb{P}^3$ with a given h-vector is smooth and irreducible, even when h is not of decreasing type.

Gruson and Peskine [1978] (see also [Maggioni and Ragusa 1988] and [Nollet 1998]) showed that, if C is an integral ACM curve, then h_C is of decreasing type, and conversely, if h is an h-vector of decreasing type, then there exists a smooth irreducible ACM curve C with $h_C = h$. Thus an h-vector h is of decreasing type if and only if the general curve C in $A(h)$ is smooth and irreducible.

If C is not irreducible, it may happen that every pair of surfaces X_1 and X_2 containing C of minimal degrees s_C and t_C have a common component. Nollet [1998, Proposition 1.5] generalized the result of Gruson and Peskine by showing that if C is contained in a complete intersection of type (s_C, t_C), then h_C is of decreasing type. We partially reproduce his argument here:

Lemma 6.5. (i) *Suppose an ACM curve D is contained in a complete intersection Y of type (s_D, t_D), and let Γ be the curve and linked to D by Y. Then*

$$e(\Gamma) + 3 < s_D.$$

(ii) *Let Γ be an ACM curve, and suppose $a \leq b$ are integers such that $a \geq e(\Gamma)+3$ and $b \geq e(\Gamma) + 4$. Then the h-vector of a curve D linked to Γ by a complete intersection of type (a, b) is of decreasing type. If $a \geq e(\Gamma) + 4$, then $s_D = a$ and $t_D = b$. If $a = e(\Gamma) + 3$, then $s_D = a$ and $t_D = b - 1$.*

Proof. If Γ and D are linked by a complete intersection Y of type (a, b), we have, by [Migliore 1998, 5.2.19],

$$h_\Gamma(n) = h_Y(n) - h_D(a+b-2-n) = h_Y(a+b-2-n) - h_D(a+b-2-n).$$

Suppose first $a = s_D$ and $b = t_D$. Then

$$h_\Gamma(s_D - 1) = h_Y(t_Y - 1) - h_D(t_D - 1) = s_Y - s_D = 0.$$

Therefore $e(\Gamma) + 3 \le s_D - 1$.

Next suppose $b \ge a \ge e(\Gamma) + 4$. Then $s_D \le a$ because $D \subseteq Y$, and

$$h_D(b-1) = h_Y(a-1) - h_\Gamma(a-1) = h_Y(a-1) = a$$

while

$$h_D(b) = h_Y(a-2) - h_\Gamma(a-2) \le h_Y(a-2) = a-1$$

hence $s_D = a$ and $t_D = b$.

If $a = e(\Gamma)+3$ and $b \ge e(\Gamma)+4$, then a similar calculation shows $h_D(b-2) = a$ and $h_D(b-1) < a$, so that $s_D = a$ and $t_D = b-1$.

It remains to show h_D is of decreasing type. Let $u = s(\Gamma)$. Then $u \le e(\Gamma)+3 \le a$ and $h_\Gamma(n) = h_Y(n) = n+1$ for $n \le u-1$; hence $h_D(n) = 0$ for $n \ge a+b-1-u$.

Since $h_\Gamma(n) \ge h_\Gamma(n+1)$ for $n \ge u-1$, we see that for $b-1 \le m \le a+b-2-u$

$$\begin{aligned} h_D(m) - h_D(m+1) &= h_Y(m) - h_Y(m+1) - h_\Gamma(a+b-2-m) + h_\Gamma(a+b-1-m) \\ &= 1 - \partial h_\Gamma(a+b-1-m) \ge 1, \end{aligned}$$

which shows that h_D is of decreasing type. □

Fix a smooth surface $X \subset \mathbb{P}^3$ of degree s. Two curves C and D on X are said to be *biliaison equivalent* if C is linearly equivalent to $D + nH$ for some integer n.

Definition 6.6. A curve C on a surface X is *minimal on* X if $C - H$ is not effective.

Proposition 6.7. *A curve C is minimal on a smooth surface X if and only if*

$$e(C) + 3 < \deg(X).$$

Proof. To say C is minimal is equivalent to saying $h^0(\mathcal{O}_X(C-H)) = 0$. By duality on X this is the same as $h^2(\mathcal{I}_C(s-3)) = 0$, where $s = \deg(X)$. On the other hand, $h^2(\mathcal{I}_C(s-3)) = h^1(\mathcal{O}_C(s-3))$, so the condition says $s - 3 > e(C)$, or equivalently, $e(C) + 3 < s$. □

Definition 6.8. We say that an h-vector is *s-minimal* if the corresponding curve satisfies $e + 3 < s$. We say that an h-vector is *s-basic* if it is the h-vector of an integral curve C satisfying $s_C = t_C = s$. Thus the s-basic h-vectors are those h-vectors of decreasing type that begin with a string

$$\{1, 2, \ldots, s-1, s, m\}$$

with $m = h(s) \le s-1$.

Table 1 on the next page lists s-basic h-vectors for $s = 4$ and $s = 5$.

	d	g	h-vector	$C^2-4(d-4)$	$C^2-4(d-5)$	λ ($\Gamma = tH-C$)	$q(\lambda)$
$s=4$	10	11	1, 2, 3, 4	−4	0	1, 2, 3	20
	11	14	1, 2, 3, 4, 1	−2	2	2, 3	17
	12	17	1, 2, 3, 4, 2	0	4	1, 3	16
	13	20	1, 2, 3, 4, 3	2	6	1, 2	17
	13	21	1, 2, 3, 4, 2, 1	4	8	3	9
	14	24	1, 2, 3, 4, 3, 1	6	10	2	12
	15	28	1, 2, 3, 4, 3, 2	10	14	1	9
	16	33	1, 2, 3, 4, 3, 2, 1	16	20	∅	0
$s=5$	15	26	1, 2, 3, 4, 5	−9	−5	1, 2, 3, 4	50
	16	30	1, 2, 3, 4, 5, 1	−6	−2	2, 3, 4	46
	17	34	1, 2, 3, 4, 5, 2	−3	1	1, 3, 4	44
	18	38	1, 2, 3, 4, 5, 3	0	4	1, 2, 4	44
	18	39	1, 2, 3, 4, 5, 2, 1	2	6	3, 4	34
	19	42	1, 2, 3, 4, 5, 4	3	7	1, 2, 3	46
	19	43	1, 2, 3, 4, 5, 3, 1	5	9	2, 4	36
	20	47	1, 2, 3, 4, 5, 4, 1	8	12	2, 3	40
	20	48	1, 2, 3, 4, 5, 3, 2	10	14	1, 4	30
	21	52	1, 2, 3, 4, 5, 4, 2	13	17	1, 3	36
	21	54	1, 2, 3, 4, 5, 3, 2, 1	17	21	4	16
	22	57	1, 2, 3, 4, 5, 4, 3	18	22	1, 2	34
	22	58	1, 2, 3, 4, 5, 4, 2, 1	20	24	3	24
	23	63	1, 2, 3, 4, 5, 4, 3, 1	25	29	2	24
	24	69	1, 2, 3, 4, 5, 4, 3, 2	32	36	1	16
	25	76	1, 2, 3, 4, 5, 4, 3, 2, 1	41	45	∅	0
$s=6$	21	50	1, 2, 3, 4, 5, 6	−12	−8	1, 2, 3, 4, 5	105
	22	55	1, 2, 3, 4, 5, 6, 1	−8	−4	2, 3, 4, 5	100
	23	60	1, 2, 3, 4, 5, 6, 2	−4	0	1, 3, 4, 5	97
	24	65	1, 2, 3, 4, 5, 6, 3	0	4	1, 2, 4, 5	96
$s=7$	28	85	1, 2, 3, 4, 5, 6, 7	−12	−8	1, 2, 3, 4, 5, 6	196
	29	91	1, 2, 3, 4, 5, 6, 7, 1	−7	−3	2, 3, 4, 5, 6	190
	30	97	1, 2, 3, 4, 5, 6, 7, 2	−2	2	1, 3, 4, 5, 6	186
$s=8$	35	130	1, 2, 3, 4, 5, 6, 7, 4, 3	29	33	1, 2, 5, 6	154
	36	133	1, 2, 3, 4, 5, 6, 7, 8	−8	−4	1, 2, 3, 4, 5, 6, 7	336
	37	140	1, 2, 3, 4, 5, 6, 7, 8, 1	−2	2	2, 3, 4, 5, 6, 7	329
	45	196	1, 2, 3, 4, 5, 6, 7, 8, 9	1	5	1, 2, 3, 4, 5, 6, 7, 8	540

Table 1. s-basic h-vectors and s-minimal biliaison types.

Proposition 6.9. *Suppose C is an ACM curve contained in a smooth surface X of degree s_C. Let $s = s_C$, $t = t_C$ and $e = e(C)$. Then $e+3 \geq t \geq s$ and*

(a) *h_C is of decreasing type*;

(b) *if $\Gamma \in |tH - C|$, then $e(\Gamma)+3 < s$ and Γ is minimal on X*;

(c) *$C - mH$ is effective if and only if $m \leq e+4-s$*;

(d) *if $C_1 \in |C-(t-s)H|$, h_{C_1} is s-basic*;

(e) *if $C_2 \in |C-(t-s+1)H|$, h_{C_2} is of decreasing type.*

There is a one to one correspondence $h_\Gamma \mapsto h_{C_1}$ mapping s-minimal h-vectors to s-basic h-vectors.

Proof. Since C is ACM, the ideal sheaf $\mathscr{I}_{C,\mathbb{P}^3}$ is $(e+3)$-regular, hence $e+3 \geq t$. By definition of t, we have $t \geq s$, and C is contained in a surface F of degree t that does not contain X. Therefore C is contained in the complete intersection $X \cap F$ of type (s,t). Let $\Gamma \in |tH - C|$ be the curve linked to C by $X \cap F$: then $e(\Gamma_0)+3 < s$ and Γ is minimal (by either Lemma 6.5 or by definition of t).

Each of the curves C, C_1, C_2 is linked to a curve in the linear system $|\Gamma|$ by a complete intersection of type (s,t), (s,s), or $(s-1,s)$, respectively. By Lemma 6.5 the h-vectors of C, C_1 and C_2 are of decreasing type, and h_{C_1} is s-basic. □

There is a unique 1-basic h-vector, namely $h_0 = \{1\}$, the h-vector of a line. Every $(s-1)$-basic h-vector gives rise to two s-basic h vectors by performing a type A or type B transformation, defined as follows: (1) A type $A = A_s$ transformation consists of inserting an s to an $(s-1)$-basic h-vector $h = \{1, 2, \ldots, s-1, m, \ldots\}$ to transform it into the s-basic vector $h' = \{1, 2, \ldots, s-1, s, m \ldots\}$. Geometrically, if h is the h-vector of a curve C on a surface X of degree s, h' is the h-vector of the effective divisor $C+H$ on X. (2) A type B transformation consists of inserting a string $s, s-1$ to an $(s-1)$-basic h-vector $h = \{1, 2, \ldots, s-1, m, \ldots\}$ to transform it into the s-basic vector $h'' = \{1, 2, \ldots, s-1, s, s-1, m \ldots\}$. Geometrically, this operation breaks into two steps: suppose h is the h-vector of a curve C on a surface X_1 of degree $s-1$. Let $C_1 = C+H$ be obtained by adding to C a plane section of X_1, then pick a surface X_2 of degree s containing C_1, and finally let $C_2 = C_1 + H$ be obtained by adding to C_1 a plane section of X_2. Then h'' is the h-vector of C_2.

Conversely, any s-basic h-vector with $m = h(s) \leq s-2$ arises from a type A transformation of an $(s-1)$-basic h-vector, while any s-basic h-vector with $m = h(s) = s-1$ arises from a type B transformation of an $(s-1)$-basic h-vector. In particular, the number of s-basic h-vectors is 2^{s-1} (see Table 1).

Proposition 6.10. *Let C be an integral ACM curve in $\mathbb{P}^3$ with $s_C \geq 4$. Suppose C is contained in a smooth surface X of degree $s = s(C)$. Suppose C has a basepoint-free pencil of degree k, and let $\mathscr{E}$ be the bundle on X associated to such a pencil.*

(a) *If* $k \le d-5$, *then* $\Delta(\mathscr{E}) > 0$ *unless*

- $s = 4$ *and* $(d, g) = (10, 11)$, *or*
- $s = 5$ *and* $(d, g) = (15, 26),\ (16, 30)$, *or*
- $s = 6$ *and* $(d, g) = (21, 50),\ (22, 55),\ (23, 60)$, *or*
- $s = 7$ *and* $(d, g) = (28, 85),\ (29, 91)$, *or*
- $s = 8$ *and* $(d, g) = (36, 133)$.

(b) *If* $k = d-4$, *then* $\Delta(\mathscr{E}) > 0$ *unless*

- $s = 4$ *and* $(d, g) = (10, 11), (11, 14),\ (12, 17)$, *or*
- $s = 5$ *and* $(d, g) = (15, 26),\ (16, 30),\ (17, 34),\ (18, 38)$, *or*
- $s = 6$ *and* $(d, g) = (21, 50),\ (22, 55),\ (23, 60),\ (24, 65)$, *or*
- $s = 7$ *and* $(d, g) = (28, 85),\ (29, 91), (30, 97)$, *or*
- $s = 8$ *and* $(d, g) = (36, 133),\ (37, 140)$.

Proof. We can compute $\Delta(\mathscr{E})$ in terms of $d = d_C$ and $g = g(C)$:

$$\Delta(\mathscr{E}) = C^2 - 4k = 2g - 2 - (s-4)d - 4k = \delta_s(d, g) + 4(d-k),$$

where we have set $\delta_s(C) = \delta_s(d, g) = 2g - 2 - ds$. One can easily verify the following facts:

(1) Let $C \subseteq X_s$ be a curve on a surface X of degree s in $\mathbb{P}^3$, and consider the divisor $C + H$ on X_s. Then

$$\delta_s(C + H) - \delta_s(C) = 2d - 3s.$$

In particular, if $d \ge \frac{1}{2}s(s+1)$ and $s \ge 3$, $\delta_s(C + H) > \delta_s(C)$.

(2) Suppose $C \subseteq X_{s+1}$ is a curve on a surface X of degree $s+1$ in $\mathbb{P}^3$, and consider the divisor $C + H$ on X_{s+1} Then

$$\delta_{s+1}(C + H) - \delta_s(C) = d - 3(s+1).$$

In particular, if $d \ge \frac{1}{2}s(s+1)$ and $s \ge 6$, $\delta_{s+1}(C + H) \ge \delta_s(C)$, and the inequality is strict unless $s = 6$ and $d = 21$.

To prove the proposition, we have seen that $\Delta(\mathscr{E})$ can be computed in terms of d, g, s, k, which depend only on the h-vector and the choice of s, k. Therefore, using the two remarks (1), (2) just made and using biliaisons on each surface to reduce to s-basic h-vectors, and using the transformations of type A and B mentioned before the statement, it would be sufficient to prove that $\Delta > 0$ for all s-basic h-vectors with $s = 4$. Unfortunately this is not so, as $\Delta \le 0$ for the first three 4-basic h-vectors (see Table 1). Still the two remarks show that Δ becomes positive using the transformations of type A and B, with the only exceptions listed in the statement. Table 1 displays all h-vectors for which $\Delta \le 0$ for $k = d-4$ and $k = d-5$. $\square$

7. General ACM curves

We now generalize the results of [Gruson and Peskine 1978] by giving a description of a general ACM curve C with a given h-vector h, even when h is not of decreasing type. We show (Theorem 7.21) that C is a union of smooth ACM subcurves whose h-vectors are determined by that of C. The basic step is Proposition 7.18, which is a special case of [Davis 1985, Corollary 4.2], and says that C is the union of two ACM subcurves whenever h_C is not of decreasing type. As a corollary we show the existence of multisecant lines for ACM curves with h-vector of special types.

Definition 7.1. Let C_0 and C be two curves in $\mathbb{P}^3$.

(a) Following [Martin-Deschamps and Perrin 1990] we say that C is obtained by an *elementary biliaison* of height h from C_0 if there exists a surface X in $\mathbb{P}^3$ containing C_0 and C so that $\mathcal{I}_{C,X} \cong \mathcal{I}_{C_0,X}(-h)$. In the language of generalized divisors [Hartshorne 1994] this means C is linearly equivalent to $C_0 + hH$ on X, where H denotes the plane section.

(b) As a particular case, we say C is obtained by a *trivial* biliaison of height h if $\mathcal{I}_{C,X} = \mathcal{I}_{C_0,X}\mathcal{I}_{Y,X}$ where Y is a complete intersection of X and a surface of degree h. If Y meets C_0 properly, this means C is the union of C_0 and Y.

(c) By a *special biliaison of degree k* we mean an elementary biliaison of height one $C \sim C_0 + H$ on a surface of degree $k \geq e(C_0) + 4$. The condition $k \geq e(C_0) + 4$ guarantees $s_C = s_{C_0} + 1$ and $k = e(C) + 3$ by [Martin-Deschamps and Perrin 1990, p. 68].

Proposition 7.2 (Lazarsfeld–Rao property). *Suppose C is an ACM curve with index of speciality e. Then C can be obtained by a special biliaison of degree $k = e+3$ from some ACM curve C_0 satisfying $s_{C_0} = s_C - 1$.*

Proof. One knows — see for example [Strano 2004] — that an ACM curve C with index of speciality e can be obtained by an elementary biliaison of height 1 on a surface X of degree $e + 3$ from an ACM curve C_0 satisfying

$$s_{C_0} = s_C - 1 \quad \text{and} \quad e(C_0) < e(C).$$

Since $\deg(X) = e + 3 \geq e(C_0) + 4$, this is a special biliaison. □

Remark 7.3. When $s_C = 1$, the curve C_0 above is the empty curve, which is therefore convenient to allow among ACM curves.

Corollary 7.4. *Let C be an ACM curve. Then there exist positive integers $k_1 < k_2 < \cdots < k_u$ such that C is obtained from the empty curve by a chain of u special biliaisons of degrees $k_1, \ldots, k_u$. The sequence $\lambda_C = (k_1, k_2, \ldots, k_u)$ is uniquely* *determined by C, and we will call it the biliaison type of C. Morever, we have*

$$d_C = \sum_{i=1}^{u} k_i, \quad g(C) = 1 + \frac{1}{2}\sum_{i=1}^{u} k_i(k_i-3) + \sum_{i=1}^{u}(s_C - i)k_i,$$
$$s_C = u, \quad t_C - s_C + 1 = k_1, \quad e(C) + 3 = k_u.$$

Example 7.5. If $C \subset \mathbb{P}^3$ is ACM , then $d_C \geq \frac{1}{2}s_C(s_C+1)$, with equality if and only if $\lambda_C = (1, 2, 3, \ldots, s_C - 1, s_C)$.

Remark 7.6. The biliaison type λ_C was introduced from a different point of view in [Green 1998], and it essentially the same thing as the numerical character $\{n_j\}$ of [Gruson and Peskine 1978]: the precise relationship, if $s = s_C$, is

$$n_j - j = k_{s-j} \quad \text{for } j = 0, \ldots, s-1.$$

The biliaison type (hence the numerical character) is equivalent to the h-vector of C. Indeed, h_C can be recovered from λ_C because one knows how h_C vector varies in an elementary biliaison, while λ_C can be computed out of h_C via the formula

$$k_i = \#\{n : h_C(n) \geq s_C + 1 - i\}.$$

One can visualize h_C and λ_C as follows. In the first quadrant of the (x, y) plane, draw a dot at (n, p) if n and p are integers satisfying $1 \leq p \leq h(n)$. Then $h(n)$ is the number of dots on the vertical line $x = n$, while k_i is the number of dots on the horizontal line $y = s - i + 1$. In particular, $k_1 = t_C - s_C + 1$ is the number of dots on the top horizontal line $y = s$, and $k_s = e(C) + 3$ is the number of dots on the bottom line $y = 1$.

Remark 7.7. The statement that every h-vector arises as the h-vector of an ACM curve in $\mathbb{P}^3$ is equivalent to the statement that every finite, strictly increasing sequence of positive integers $\lambda = (k_1, \ldots, k_u)$ occurs as λ_C for some ACM curve $C \subset \mathbb{P}^3$. We can see this by induction on u. When $u = 1$, $\lambda = (k)$ is the biliaison type of a plane curve of degree k. If $u > 1$, by induction there is an ACM curve C_0 with $\lambda_{C_0} = (k_1, \ldots, k_{u-1})$. Now $s_{C_0} \leq e(C_0) + 3 = k_{u-1} < k_u$. Therefore we can find a surface X of degree k_u containing C_0, and construct C from C_0 by a biliaison of height one on X. Since $e(C_0) + 3 < k_u$, the biliaison is special, hence λ_C equals the given λ. A refined version of this construction is in Theorem 7.21.

Definition 7.8. A sequence $\lambda = (k_1, k_2, \ldots, k_u)$ *has a gap* at i if $k_{i+1} - k_i \geq 3$.

For example, the sequence λ_C of Figure 1 has a gap at $i = 2$.

Davis [1985] shows that a gap in λ_C forces C to break in the union of two ACM subcurves. We now give a more geometric proof of this result. For this we need some preliminary remarks. While in general the union C of two ACM curves B and D can fail to be ACM, it is certainly ACM if I_D/I_C is isomorphic to R_B up to

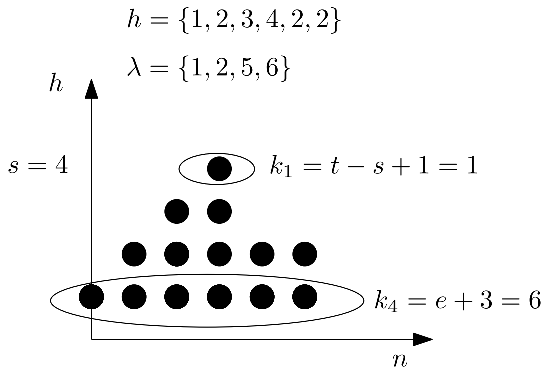


Figure 1. Biliaison type and h-vector.

a twist. This condition is satisfied when C is obtained from B by a trivial biliaison, and also when C is obtained from B by a chain of elementary biliaisons "trivial on B" (Lemma 7.16 below). Here are some preliminary examples.

Example 7.9. If C is obtained from a curve B by a trivial biliaison of height h on a surface X, "adding" to C the complete intersection Y of X with a surface of degree h, then

$$I_Y/I_C \cong \frac{I_Y/I_X}{I_C/I_X} \cong \frac{H^0_*(\mathcal{I}_{Y,X})}{H^0_*(\mathcal{I}_{C,X})} \cong \frac{H^0_*(\mathcal{O}_X(-h))}{H^0_*(\mathcal{I}_{B,X}(-h))} \cong R_B(-h)$$

Example 7.10. Let $D \subset \mathbb{P}^3$ be a curve, and L a line not contained in D. Set $C = D \cup L$, and let f be the degree of the scheme theoretic intersection $D \cap L$. Then $\mathcal{I}_{D,C} \cong \mathcal{I}_{D\cap L,L} \cong \mathcal{O}_L(-f)$. If D is ACM, it follows that $C = D \cup L$ is ACM if and only if $I_D/I_C \cong R_L(-f)$.

By the same argument, if B and D are two ACM curves meeting properly and $\mathcal{I}_{B\cap D,B} \cong \mathcal{O}_B(-f)$, then $C = B \cup D$ is ACM if and only if $I_D/I_C \cong R_B(-f)$.

From another point of view, suppose B and D are two ACM curves contained in a *smooth* surface X, and let $C = B + D$. Then

$$\mathcal{O}_B(-D) \stackrel{\text{def}}{=} \mathcal{O}_X(-D) \otimes \mathcal{O}_B \cong \mathcal{I}_{D,C}.$$

If $\mathcal{O}_B(-D) \cong \mathcal{O}_B(-f)$, then C is ACM if and only if $I_D/I_C \cong R_B(-f)$.

The condition $I_D/I_C \cong R_B(-f)$ implies that C is obtained by a "generalized liaison addition" of B and D in the sense of [Geramita and Migliore 1994]. The following proposition is essentially a special case of Theorem 1.3 of that reference.

Proposition 7.11. *Suppose that C contains two subcurves B and D, and that for some integer f there is an isomorphism of R_C-modules*:

$$I_D/I_C \cong R_B(-f). \tag{7-1}$$

(a) *There is a surface S of degree f containing D but not C, and the curve D is the scheme theoretic intersection of C and S. In particular, $f \geq s_D$.*

(b) *The degrees and genera of B, C and D are related by the formulas*

$$d_C = d_B + d_D, \quad g(C) = g(B) + g(D) + f d_B - 1.$$

If B and D have no common component, then C is the scheme-theoretic union of B and D, $\mathcal{I}_{B\cap D,B} \cong \mathcal{O}_B(-f)$, and $B.D = f d_B$.

If C is contained in a smooth surface X, then $C = B + D$ on X, and $\mathcal{O}_X(D) \otimes \mathcal{O}_B \cong \mathcal{O}_B(f)$. In particular, $B.D = f d_B$.

(c) *Suppose D is ACM. Then B is ACM if and only if C is ACM, in which case*

$$h_C(n) = h_B(n-f) + h_D(n)$$

(d) *Suppose B, C and D are ACM and $f = s_D$. If* $\max\{\lambda_B\} < \min\{\lambda_D\}$ *then*

$$\lambda_C = \lambda_B \cup \lambda_D.$$

Proof. The hypothesis $I_D/I_C \cong R_B(-f)$ is equivalent to there being a form $F \in H^0(\mathbb{P}^3, \mathcal{O}(f))$ such that the sequence

$$0 \to I_B/I_C(-f) \to R_C(-f) \xrightarrow{F} R_C \to R_D \to 0$$

is exact. In particular, $I_D = I_C + I_S$ where S is the surface of equation $F = 0$, hence D is the scheme theoretic union of C and S. Sheafifying the exact sequence

$$0 \to I_B(-f) \to I_C \to I_D/(F) \to 0$$

we obtain another exact sequence

$$0 \to H^1_*(\mathcal{I}_B)(-f) \to H^1_*(\mathcal{I}_C) \to H^1_*(\mathcal{I}_D).$$

It follows that, if D is ACM, then $H^1_*(\mathcal{I}_B)(-f) \cong H^1_*(\mathcal{I}_C)$, and B is ACM if and only if C is ACM.

If B and D are ACM, the relation between the h-vectors follows immediately from the exact sequence $0 \to R_B(-f) \to R_C \to R_D \to 0$.

The relation between the degrees and genera follows computing the Euler characteristics of the two sides of $\mathcal{I}_{D,C} \cong \mathcal{O}_B(-f)$.

Suppose B and D have no common components. The kernel of the natural surjective map

$$\mathcal{O}_B(-f) \cong \mathcal{I}_{D,C} \to \mathcal{I}_{B\cap D,B}$$

is supported on D and is a subsheaf of $\mathcal{O}_B$. Since B is locally Cohen–Macaulay and has no component in common with D, the kernel is zero, hence $\mathcal{O}_B(-f) \cong \mathcal{I}_{B\cap D,B}$.

Suppose C is contained in a smooth surface X. Since $D \subseteq C$, there is an effective divisor A on X such that $C = A + D$. Then

$$\mathcal{O}_B(-f) \cong \mathcal{I}_{D,C} \cong \mathcal{O}_X(-D) \otimes \mathcal{O}_A$$

from which we deduce $A = B$ and $\mathcal{O}_B(f) \cong \mathcal{O}_X(D) \otimes \mathcal{O}_B$, hence $B.D = f d_B$.

We deduce (d) from (c). By assumption

$$e(B) + 3 = \max\{\lambda_B\} < \min\{\lambda_D\} = t_D - s_D + 1.$$

On the other hand, $h_D(n) = s_D$ if and only if $s_D - 1 \le n \le t_D - 1$, and $h_B(n - s_D)$ is nonzero if and only if $s_D \le n \le s_D + e(B) + 2$. Since $t_D > s_D + e(B) + 2$, we see $h_D(n) = s_D$ whenever $h_B(n - s_D)$ is nonzero (h_B so to speak sits on the top of h_D, as in Figure 1). Now it follows from $h_C(n) = h_B(n - f) + h_D(n)$ that $\lambda_C = \lambda_B \cup \lambda_D$. □

Example 7.12. Figure 1 on page 289 shows the h-vector of a curve which is the union of a twisted cubic curve B and a divisor D of type (6, 5) on a smooth quadric surface. The biliaison types are $\lambda_B = \{1, 2\}$ and $\lambda_D = \{5, 6\}$.

Definition 7.13. Suppose $D_0 \subseteq C_0$ are curves in $\mathbb{P}^3$ contained in a surface X, and D is obtained from D_0 by an elementary biliaison of height h on X. The biliaison is defined by an injective morphism $v : \mathcal{I}_{D_0,X}(-h) \to \mathcal{O}_X$ whose image is $\mathcal{I}_{D,X}$. Then the image of the restriction of v to $\mathcal{I}_{C_0,X}(-h)$, is the ideal $\mathcal{I}_{C,X}$ of a curve $C \subset X$, obtained by biliaison from C_0. In this case, we say that the biliaison from C_0 to C is *induced* by the given biliaison from D_0 to D. Note that C contains D.

Remark 7.14. When D_0 is empty, a biliaison induced from D_0 is the same thing as a trivial biliaison. Indeed, in this case v is multiplication by a local equation of the complete intersection D in $\mathcal{O}_X$, and v maps $\mathcal{I}_{C_0,X}(-h)$ onto $\mathcal{I}_{C_0,X}\mathcal{I}_{D,X}$.

Remark 7.15. For an elementary biliaison from C_0 to C to be induced by a biliaison of D_0 it is enough that the corresponding morphism $u : \mathcal{I}_{C_0,X}(-h) \to \mathcal{O}_X$ lift to a morphism $\hat{u} : \mathcal{I}_{D_0,X}(-h) \to \mathcal{O}_X$. Indeed, $\hat{u}$ is automatically injective because its kernel $\mathcal{K}$ is isomorphic to a subsheaf of $\mathcal{I}_{D_0,C_0}(-h) \subseteq \mathcal{O}_{C_0}(-h)$, and at the same time is a subsheaf of $\mathcal{O}_X(-h)$; since $\mathcal{O}_X$ and $\mathcal{O}_{C_0}$ have no common associated points, we must have $\mathcal{K} = 0$.

Lemma 7.16. *Suppose C_0 contains B and D_0, and $I_{D_0}/I_{C_0} \cong R_B(-f)$. Suppose C is obtained by an elementary biliaison from C_0 induced by an elementary biliaison of height h from D_0 to D on a surface X. Then C contains D and B, and*

$$I_D/I_C \cong R_B(-f-h).$$

Proof. Since the biliaison from C_0 to C is induced by that from D_0 to D, C contains D, and

$$I_D/I_C \cong \frac{I_{D_0}/I_X(-h)}{I_{C_0}/I_X(-h)} \cong R_B(-f-h)$$

In particular, $R_B(-h-f)$ is an R_C-module, therefore $B \subseteq C$. ☐

Lemma 7.17. *Suppose C_0 contains B and D_0, and $I_{D_0}/I_{C_0} \cong R_B(-s_{D_0})$. If k is an integer such that*

$$k \geq \max(s_{D_0}+e(B)+6,\, e(C_0)+4),$$

then any height-one biliaison from C_0 to C on a surface of degree k is induced by a biliaison from D_0 to a curve D such that

$$I_D/I_C \cong R_B(-s_D)$$

Proof. The lemma generalizes [Martin-Deschamps and Perrin 1990, Remark 2.7c, p. 65], which treats the case $C_0 = B$ and $D_0 = \varnothing$. The statement in this case becomes: if $k \geq e(C_0)+6$, then every height-one elementary biliaison from C_0 to C on a surface of degree k is trivial.

To prove the statement, let X be the degree k surface on which the biliaison from C_0 to C is defined, and apply $\mathrm{Hom}_{\mathcal{O}_X}(\,\cdot\,,\mathcal{O}_X)$ to the exact sequence

$$0 \to \mathcal{I}_{C_0,X}(-1) \to \mathcal{I}_{D_0,X}(-1) \to \mathcal{O}_B(-s_{D_0}-1) \to 0$$

to see that $u : \mathcal{I}_{C_0,X}(-1) \to \mathcal{O}_X$ lifts to $\hat{u} : \mathcal{I}_{D_0,X}(-1) \to \mathcal{O}_X$ if and only if the image of u in $\mathrm{Ext}^1_{\mathcal{O}_X}(\mathcal{O}_B(-s_{D_0}-1), \mathcal{O}_X)$ vanishes. Now by Serre duality on X the latter Ext group is dual to

$$H^1(X, \mathcal{O}_B(k-s_{D_0}-5))$$

which is zero because $k \geq s_{D_0}+e(B)+6$. Thus u lifts to give a height-one biliaison from D_0 to a curve D inducing the biliaison from C_0 to C. By Lemma 7.16 above $I_D/I_C \cong R_B(-s_{D_0}-1)$. Finally, since $k \geq s_{D_0}+1$, we have $s_D = s_{D_0}+1$. ☐

The following proposition is a special case of [Davis 1985, Corollary 4.2].

Proposition 7.18. *Suppose the biliaison type $\lambda_C = (k_1, k_2, \ldots, k_s)$ of an ACM curve C has a gap at j. Then C contains ACM curves B and D such that*

$$\lambda_B = (k_1, k_2, \ldots, k_j), \quad \lambda_D = (k_{j+1}, k_{j+2}, \ldots, k_s), \quad \textit{and} \quad I_D/I_C \cong R_B(-s_D).$$

Furthermore, (B, D) is the unique pair of ACM curves with the above properties.

Proof. Note that $s = s_C$. Suppose first $j = s-1$, that is, $k_s \geq k_{s-1}+3$. Since $k_s = e(C)+3$, by Proposition 7.2 C is obtained by a special biliaison on a surface

X of degree k_s from an ACM curve B. By definition of biliaison type, $\lambda_B = (k_1, k_2, \ldots, k_{s-1})$. As $k_{s-1} = e(B) + 3$, we see

$$k_s \geq k_{s-1} + 3 = e(B) + 6.$$

By Lemma 7.17 the biliaison is trivial, so C contains a plane section D of X, and $I_D/I_C \cong R_B(-1)$. Since $\lambda_D = (\deg(X)) = (k_s)$, the statement holds when $j = s-1$

We now suppose $j < s-1$ and proceed by induction on $s-j$. By Proposition 7.2 C is obtained by a special biliaison on a surface X of degree k_s from an ACM curve C_0 whose biliaison type is $\lambda_0 := \lambda_{C_0} = (k_1, k_2, \ldots, k_{s-1})$. Thus λ_0 has a gap at j, and $s_{C_0} = s-1$, hence by induction C_0 contains ACM curves B and D_0 such that $\lambda_B = (k_1, k_2, \ldots, k_j)$, $\lambda_{D_0} = (k_{j+1}, k_{j+2}, \ldots, k_{s-1})$, and $I_{D_0}/I_{C_0} \cong R_B(-s_{D_0})$.

In particular, $s_{D_0} = s - j - 1$, so that

$$k_s \geq k_{j+1} + s - j - 1 \geq k_j + 3 + s_{D_0} = e(B) + 6 + s_{D_0}.$$

Since $k_s = e(C) + 3 \geq e(C_0) + 4$, by Lemma 7.17 the biliaison from C_0 to C is induced by a biliaison from D_0 to a curve D, and $I_D/I_C \cong R_B(-s_D)$. Finally, since D is obtained from D_0 by a special biliaison, D is ACM and $\lambda_D = \lambda_{D_0} \cup (k_s) = (k_{j+1}, k_{j+2}, \ldots, k_s)$.

It remains to prove uniqueness. Note that $s_D = s - j$ is determined by C, hence so is t_D because

$$t_D - s_d + 1 = \min(\lambda_D) = k_{j+1}.$$

By assumption $e(B) + 3 = k_j \leq k_{j+1} - 3 = t_D - s_D - 2$, hence from the exact sequence

$$0 \to \omega_D(m) \to \omega_C(m) \to \omega_B(s_D + m) \to 0$$

we see

$$H^0(\omega_D(m)) = H^0(\omega_C(m)) \quad \text{for every } m \leq 3 - t_D.$$

We will show that $\Omega_D = H^0_*(\omega_D)$ is generated over the polynomial ring $R = H^0_*(\mathbb{P}^3)$ by its elements of degree at most $3 - t_D$. Taking this for granted for the moment, it follows that Ω_D is the submodule of Ω_C generated by

$$\bigoplus_{m \leq 3 - t_D} H^0(\omega_C(m));$$

hence it is determined by C. But I_D is the annihilator of Ω_D, because R_D is Cohen–Macaulay with canonical module Ω_D, hence D is determined by C.

Since $t_D - s_D + 1 = k_{j+1} > 1$, the curve D is contained in a unique surface S of degree s_D, and therefore B is also determined, being the residual curve to $D = C \cap S$ in C.

To finish, we need to show $\Omega_D = H^0_*(\omega_D)$ is generated by its sections of degree at most $3 - t_D$. For this we choose a complete intersection Y of type (s_D, u)

containing D and let E be the curve linked to D by Y. As $\Omega_D \cong I_E/I_Y(-e_Y)$ and I_E is generated by its elements of degree at most $e(E)+3$, it is enough to show $e(Y)-t_D \geq e(E)$.

From $\omega_E(-e(Y)) \cong \mathcal{I}_D/\mathcal{I}_Y$ and $h^0(\mathcal{I}_D(t_D-1)) = h^0(\mathcal{I}_Y(t_D-1))$, we see that $h^0(\omega(t_D-1-e(Y))) = 0$; that is, $t_D - e(Y) \leq -e(E)$, as desired. □

Corollary 7.19. *Let $C \subset \mathbb{P}^3$ be an irreducible, reduced ACM curve that is contained in a smooth surface X of degree $s = s_C$. Let $t = t_C$ and $e = e(C)$.*

(a) *If $h_C(e+1) = 3$, $h_C(e+2) = 2$, then C has a unique $(e+3)$-secant line L, and every surface of degree at most $e+2$ containing C contains L as well.*

(b) *If $h_C(t) = s-2$, $h_C(t+1) = s-3$ (so that $s \geq 3$), then X contains a line L that is a $(t-s+1)$-secant of C.*

Remark 7.20. As a partial converse, we will see in the proof of Theorem 9.1 that, if, for every smooth C in the Hilbert scheme $A(h)$, the general surface of degree s containing C contains a line, then the h-vector of C satisfies either (a) or (b).

Proof of Corollary 7.19. Since X is smooth, by definition of t there is surface X_t of degree t containing C but not X. Thus C is contained in the complete intersection $Y = X \cap X_t$. Let Γ the curve linked to C by Y. Then on X

$$C \sim tH - \Gamma$$

where H denotes a plane section of X, and $\sim$ stands for linear equivalence. By [Migliore 1998, Corollary 5.2.19],

$$h_\Gamma(n) = h_Y(s+t-2-n) - h_C(s+t-2-n).$$

<u>Case A: $h(e+1) = 3$ and $h(e+2) = 2$.</u> The formula above implies

$$s_\Gamma = \min\{s, s+t-4-e\}.$$

But $t \leq e+3$ because $h_C(e+3) = 0$, hence $s_\Gamma = s+t-4-e$. The conditions on h_C then translate as follows:

$$h_\Gamma(s_\Gamma) = h_\Gamma(s_\Gamma+1) = s_\Gamma - 1.$$

If $s_\Gamma = 1$, this implies $\Gamma = L$ is a line. If $s_\Gamma \geq 2$, then the condition on h_Γ is equivalent to $\lambda_\Gamma = (1, k_2, \dots)$, with $k_2 \geq 4$ because $h_\Gamma(n) \geq s_\Gamma - 1$ at least for $n = s_\Gamma - 2, s_\Gamma - 1, s_\Gamma, s_\Gamma + 1$. By Proposition 7.18 Γ contains a line L and an ACM curve D with $I_D/I_\Gamma \cong R_L(1-s_\Gamma)$. We can treat the two cases simultaneously if we take D to be the empty curve when $s_\Gamma = 1$.

By Proposition 7.11, $\Gamma = L + D$ on X, and $L.D = s_\Gamma - 1$. Thus

$$C.L = (tH - L - D).L = t + s - 2 - s_\Gamma + 1 = s + t - s_\Gamma - 1 = e + 3.$$

In particular, every surface of degree at most $e+2$ containing C contains L as well. On the other hand, $C+L$ is an ACM curve, because it is linearly equivalent to $D+tH$. Therefore

$$I_C/I_{C+L} \cong R_L(-C.L) = R_L(-e-3).$$

It follows that $h_{C\cup L}(n)$ and $h_C(n)$ differ only for $n=e+3$, where their value is 1 and 0 respectively. In particular, $h_{C\cup L}(e+2) = h_C(e+2) = 2$ and $h_{C\cup L}(e+3) = 1$, so that by [Nollet 1998, Proposition 1.5] the homogeneous ideal of $C\cup L$ is generated by its forms of degree at most $e+2$, hence by the forms in I_C of degree at most $e+2$.

Suppose now M is an $(e+3)$-secant line of C. Then the homogeneous ideals of C and $C\cup M$ coincide in degrees at most $e+2$. It follows that the ideal of $C\cup L$ is contained in that of $C\cup M$, hence $C\cup L = C\cup M$ and $L = M$. Therefore L is the unique $(e+3)$-secant of C.

Case B: $h_C(t) = s-2$ and $h_C(t+1) = s-3$. Then $h_\Gamma(s-3) = h_\Gamma(s-2) = 1$ and $h_\Gamma(s-1) = 0$. This implies either $\lambda_\Gamma = (s-1)$, or $\lambda_\Gamma = (\dots, k_{u-1}, s-1)$ with $s-1-k_{u-1} \geq 3$. By Proposition 7.18, Γ contains a plane curve P of degree $s-1$ and an ACM curve B (possibly empty) such that $I_P/I_\Gamma \cong R_B(-1)$.

By Proposition 7.11, $\Gamma = B+P$ on X, and $B.P = d_B$. Let L be the line residual to P in the intersection of X with the plane of P. Then $B.L = B.H - B.P = 0$; hence

$$C.L = (tH - B - P).L = ((t-1)H - B + L).L = t-1+2-s = t-s+1. \quad \square$$

Given any sequence $\lambda = (k_1, k_2, , \dots, k_u)$ with $r-1$ gaps (for any $r \geq 1$), we can decompose λ uniquely as

$$\lambda = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_r, \tag{7-2}$$

where each λ_i has no gaps and, if a_i and b_i denote respectively the minimum and the maximum integer in λ_i, we have $a_{i+1} - b_i \geq 3$. We call (7-2) the *gap decomposition* of λ.

Theorem 7.21. *Let $A(\lambda)$ denote the Hilbert scheme parametrizing ACM curves having biliaison type λ. If C is general in $A(\lambda)$, then C is reduced and for every $f \geq e(C)+3$, there exists a smooth surface F of degree f containing C.*

Let $\lambda = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_r$ be the gap decomposition of λ. Then:

(a) *Every ACM curve $C \in A(\lambda)$ contains ACM subcurves D_i, $i = 1, 2, \dots, r$, such that $\lambda_{D_i} = \lambda_i$.*

(b) *If C is general in $A(\lambda)$, we have*

$$C = D_1 \cup D_2 \cup \cdots \cup D_r,$$

where the D_i are distinct smooth irreducible ACM curves satisfying $\lambda_{D_i} = \lambda_i$; for every $1 \le i_1 < i_2 < \cdots < i_h \le r$, the curve

$$D_{i_1} \cup D_{i_2} \cup \cdots \cup D_{i_h}$$

is ACM and has biliaison type $\lambda_{i_1} \cup \lambda_{i_2} \cup \cdots \cup \lambda_{i_h}$.

Remark 7.22. The D_i in Theorem 7.21 (for $i \ge 2$) are not necessarily general in $A(\lambda_i)$: this is because they are forced to lie on surfaces containing D_j for $j < i$.

Proof of Theorem 7.21. Recall that by a theorem of Ellingsrud $A(\lambda)$ is irreducible (see Remark 6.4). By Proposition 7.18 and induction on the number of gaps we see that for each i, $1 \le i \le r$, there are ACM curves C_i and D_i with the following properties:

(1) $C_r = C$ and $C_1 = D_1$.

(2) If $2 \le i \le r$, C_i contains C_{i-1} and D_i, and $I_{D_i}/I_{C_i} = R_{C_{i-1}}(-s_{D_i})$.

(3) $\lambda_{D_i} = \lambda_i$ for every $1 \le i \le r$.

(4) $\lambda_{C_i} = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_i$ for every $1 \le i \le r$.

We claim that for every $1 \le i_1 < i_2 < \cdots < i_h \le r$ there are ACM curves $E_{i_1,i_2,\ldots,i_h} \subseteq C_{i_h}$ such that

(1) if $h = 1$, $E_i = D_i$, and, if $h = r$, $E_{1,2,\ldots,r} = C$;

(2) if $2 \le h \le r$, $E_{i_1,i_2,\ldots,i_h}$ contains $E_{i_1,i_2,\ldots,i_{h-1}}$ and D_{i_h}, and

$$I_{D_{i_h}}/E_{i_1,i_2,\ldots,i_h} = R_{E_{i_1,i_2,\ldots,i_{h-1}}}(-s_{D_{i_h}});$$

(3) $\lambda_{E_{i_1,i_2,\ldots,i_h}} = \lambda_{i_1} \cup \lambda_{i_2} \cup \cdots \cup \lambda_{i_h}$.

We prove the statement by induction on h. When $h = 1$ there is nothing to prove. Suppose $h > 1$. By the induction hypothesis, there is a curve $A = E_{i_1,i_2,\ldots,i_{h-1}} \subseteq C_{i_{h-1}}$ with the properties above. Let $B = C_{i_h - 1}$. By Lemma 7.23 below there exists a curve $C_0 \subseteq C_{i_h}$ containing B and D_{i_h} such that $I_{D_{i_h}}/I_{C_0} \cong R_A(-s_{D_{i_h}})$. Since A and D_{i_h} are ACM, it follows from Proposition 7.11 that C_0 is ACM as well. We define $E_{i_1,i_2,\ldots,i_h}$ to be C_0. Then $E_{i_1,i_2,\ldots,i_h}$ has the required properties (the formula for the biliaison type follows from part (d) of the same proposition).

To see the components D_i of a generic C are smooth, we follow the original proof of [Gruson and Peskine 1978, 2.5]. More precisely we show that, if

$$\lambda = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_r$$

is the gap decomposition of $\lambda = (k_1, \ldots, k_s)$, there exists an ACM curve C with $\lambda_C = \lambda$ satisfying the following properties:

(1) C is contained in a smooth surface for every $f \ge k_s = e(C) + 3$.

(2) $C = D_1 \cup D_2 \cup \cdots \cup D_r$, where the D_i are smooth irreducible ACM curves satisfying $\lambda_{D_i} = \lambda_i$; in particular, C is reduced.

(3) $\omega_{D_r}(-e(D_r))$ has a section whose scheme of zeros is smooth (contains no multiple points).

We prove this statement by induction on s as in [Gruson and Peskine 1978, 2.5]. For $s = 1$, the statement is about plane curves and is well known (note that $e(C) + 3 = d_C$ for a plane curve C).

Assume now the statement is true for λ, fix a curve C with the properties above, and consider $\lambda^+ = \lambda \cup \{k_{s+1}\}$. We have two cases to consider:

Case 1: $k_{s+1} \leq k_s + 3$. In this case λ^+ has a gap at s, and its gap decomposition is $\lambda^+ = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_r \cup \{k_{s+1}\}$.

By assumption, $k_{s+1} \geq k_s + 3 = e(C) + 6$; thus there exists a smooth surface X of degree k_{s+1} containing C. Let D_{r+1} be a general plane section of X, and let $C^+ = C \cup D_{r+1}$. Then D_{r+1} is smooth with $\lambda = (k_{s+1})$, thus C^+ satisfies (2) with respect to λ^+. It also satisfies (3) because $\omega_{D_{r+1}}(-e(D_{r+1})) \cong \mathcal{O}_{D_{r+1}}$. By construction C^+ lies on the smooth surface X of degree $k_{s+1} = e(C^+) + 3$. The fact that C^+ is contained in a smooth surface of degree f, for every $f > e(C^+)+3$, follows now from the fact that $\mathcal{I}_{C^+}(e(C^+)+3)$ is generated by its global sections; see, for example, [Peskine and Szpiro 1974] and [Nollet 1998, Corollary 2.9]. Thus C^+ also satisfies (1), and we are done in case 1.

Case 2: $k_{s+1} = k_s + 1$ or $k_s + 2$. In this case the gap decomposition of λ^+ is

$$\lambda^+ = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_{r-1} \cup \lambda_r^+$$

where $\lambda_r^+ = \lambda_r \cup \{k_{s+1}\}$.

We can still find a smooth surface X of degree k_{s+1} containing C because $k_{s+1} > e(C) + 3$. In particular, X contains D_r. The proof of [Gruson and Peskine 1978, 2.5] shows that the general curve D_r^+ in the linear system $D_r + H$ on X is smooth with $\lambda_{D_r^+} = \lambda_r^+$, and that $\omega_{D_r^+}(-e(D_r^+))$ has a section whose scheme of zeros is smooth. Thus

$$C^+ = D_1 \cup D_2 \cup \cdots \cup D_r^+$$

has the required properties (note that $e(C^+) + 3 = k_{s+1} = \deg(X)$). □

Lemma 7.23. *Suppose $C \subset \mathbb{P}^3$ is a curve, with subcurves B, D such that*

$$I_D/I_C \overset{\beta}{\cong} R_B(-f).$$

If A is a subcurve of B, there exists a unique curve C_0 with the following properties:

(1) *C_0 is contained in C.*

(2) *C_0 contains A and D, and there is an isomorphism $I_D/I_{C_0} \overset{\alpha}{\cong} R_A(-f)$ which makes commutative the diagram*

$$\begin{array}{ccc} I_D/I_C & \xrightarrow[\cong]{\beta} & R_B(-f) \\ \downarrow & & \downarrow \\ I_D/I_{C_0} & \xrightarrow[\cong]{\alpha} & R_A(-f) \end{array}$$

where the vertical arrows are induced by the inclusions $C_0 \subseteq C$ and $A \subseteq B$.

If A and D have no common components, then $C_0 = A \cup D$.

Proof. The inclusion

$$I_A/I_B(-f) \hookrightarrow R_B(-f) \overset{\beta^{-1}}{\cong} I_D/I_C \hookrightarrow R_C$$

defines an ideal J in R_C. Uniqueness is clear, because if such a C_0 exists, we must have $I_{C_0}/I_C = J$. To show existence, let I be the inverse image of J in the polynomial ring $R = H^0_*(\mathcal{O}_{\mathbb{P}^3})$, so that $I/I_C \cong I_A/I_B(-f)$. The given isomorphism $I_D/I_C \overset{\beta}{\cong} R_B(-f)$ induces $I_D/I \overset{\alpha}{\cong} R_A(-f)$, hence an exact sequence

$$0 \to R_A(-f) \to R/I \to R_D \to 0.$$

From this exact sequence we see that R/I has depth at least one, hence I is the saturated ideal of a subscheme $C_0 \subset C$.

By construction I_{C_0}/I_C and $I_A/I_B(-f)$ are isomorphic, so that the given isomorphism $I_D/I_C \overset{\beta}{\cong} R_B(-f)$ induces another, $I_D/I_{C_0} \overset{\alpha}{\cong} R_A(-f)$, with the desired properties. Finally, we can check C_0 is a locally Cohen–Macaulay curve looking at the exact sequence

$$0 \to \mathcal{O}_A(-f) \to \mathcal{O}_{C_0} \to \mathcal{O}_D \to 0.$$

If A and D have no common components, then C_0 contains the union $A \cup D$. Since both C_0 and $A \cup D$ are locally Cohen–Macaulay curves of degree $d_A + d_D$, they must be equal. $\square$

8. Bounds on the quadratic form $\phi(D, D)$

Let $X \subset \mathbb{P}^3$ be a smooth surface of degree $s \geq 2$. We will make use of the bilinear form on $\operatorname{Pic}(X)$:

$$\phi(D, E) = (D.H)\,(E.H) - s\,(D.E) = \det \begin{bmatrix} D.H & H^2 \\ D.E & E.H \end{bmatrix}.$$

This is essentially the positive definite product on $\mathrm{Pic}(X)/\mathbb{Z}H$ induced by the intersection product: by the algebraic Hodge index theorem, $\phi(D, D) \geq 0$ for any divisor D on X, and $\phi(D, D) = 0$ if and only if D is numerically (hence linearly) equivalent to a multiple of H.

In the proof of our main theorem it will be crucial to be able to bound $\phi(D, D)$ from below in terms of the degree d_D when D is an ACM curve on X. Note that if D is a curve on X, then

$$\phi(D, D) = d_D^2 + s(s-4)d_D - 2s(g(D)-1) \tag{8-1}$$

Thus, if we fix the degree d_D and s, then knowing $\phi(D, D)$ is the same as knowing the genus $g(D)$, and bounding $\phi(D, D)$ from below is the same as bounding $g(D)$ from above. In fact, the bounds of this section can be seen as a refinement of the bounds on the genus of an ACM curve of [Gruson and Peskine 1978]; see Remark 8.8. The form $\phi(D, D)$ has the advantage of being invariant if we replace D with $mH - D$ or $D + nH$, that is, it is invariant under liaison and biliaison on X. Thus one can compute $\phi(D, D)$ assuming D is a minimal curve on X.

To compute these bounds we note that, by (8-1), the form $\phi(D, D)$ for an ACM curve D depends only on the h-vector (or the biliaison type λ) of D and on s. Since it is enough to consider only minimal curves on X, and there only finitely many possible biliaison types λ of minimal curves for each s, our proof will proceed by a careful analysis of these λ.

We call a biliaison type λ *s-minimal* if it corresponds to a minimal ACM curve on a smooth surface X of degree s. Since minimal is equivalent to $e + 3 < s$ by Proposition 6.7, the s-minimal types λ are just those increasing sequences of positive integers $\lambda = (k_1, k_2, \ldots, k_u)$ satisfying $k_u < s$. There are 2^{s-1} such possible sequences (including the empty one), and by Proposition 6.9 the corresponding curves are linked by a complete intersection (s, s) to curves with s-basic h-vectors. For any such λ, we let d, g, e be the corresponding invariants of the associated curve Γ, and we define

$$q(\lambda) = \phi(\Gamma, \Gamma) = d^2 + s(s-4)d - 2s(g-1). \tag{8-2}$$

Then one verifies the formula

$$q(\lambda) = \sum_{i=1}^{u} k_i(s-1)(s-k_i) \; - \; 2 \sum_{1 \leq i < j \leq u} k_i(s-k_j). \tag{8-3}$$

Table 1 on page 284 lists all the s-basic h-vectors and associated s-minimal biliaison types λ for $s = 4, 5$ and a few for $s = 6, 7, 8, 9$, together with the values q takes on them.

Definition 8.1. Suppose $\lambda = (k_1, k_2, \dots, k_u)$ is s-minimal. Then we define the s-*dual* λ' of λ to be

$$\lambda' = (s - k_u, s - k_{u-1}, \dots, s - k_1)$$

if $\lambda \neq \varnothing$. If $\lambda = \varnothing$, then $\lambda' = \varnothing$. Note that, if λ is the biliaison type of an ACM curve Γ, then λ' is the biliaison type of a curve linked to Γ by a complete intersection of two surfaces of degree $s_\Gamma = u_\lambda$ and s (see Section 6).

Proposition 8.2. *The invariants of* λ' *are* $u_{\lambda'} = u_\lambda$, $d_{\lambda'} = u_\lambda s - d_\lambda$, $q(\lambda') = q(\lambda)$.

Proof. The first two equalities are obvious. The equality $q(\lambda') = q(\lambda)$ follows from (8-3), or can be deduced from the invariance of $\phi(D, D)$ under liaison on X. □

We say that $\lambda_1 = (k_1, k_2, \dots, k_u)$ *precedes* $\lambda_2 = (l_1, l_2, \dots, l_v)$ and write $\lambda_1 < \lambda_2$ if $k_u < l_1$. In this case, if λ_2 is s-minimal, then

$$\lambda_1 \cup \lambda_2 = (k_1, k_2, \dots, k_u, l_1, \dots, l_v)$$

is also s-minimal. Note that $(\lambda \cup \mu)' = \mu' \cup \lambda'$.

Example 8.3. A plane curve of degree $k < s$ on a surface X of degree $s \geq 2$ is minimal. The corresponding λ sequence is $\lambda = (k)$, and $q(\lambda) = k(s-1)(s-k)$.

More generally if λ is the biliaison type of a complete intersection of two surfaces of degrees $a \leq b < s$ then $q(\lambda) = ab(s-a)(s-b)$.

Example 8.4. Let $\lambda = (1, 2, \dots, k-1, k)$ with $k < s$. Then $d_\lambda = \frac{1}{2}k(k+1)$ and

$$q(\lambda) = d_\lambda \left(s^2 - \tfrac{2}{3}s(2k+1) + d_\lambda\right)$$

The first statement of Proposition 8.5 below determines, once $q((k))$ is known, the function $q(\lambda)$ by induction on the number u_λ of elements of λ.

Proposition 8.5. *Suppose* $\lambda < \mu$ *are* s*-minimal.*

(a) $q(\lambda \cup \mu) = q(\lambda) + q(\mu) - 2d_\lambda d_{\mu'}$.

(b) *If* $\lambda < (k)$ *and* $(k+1) < \mu$, *then*

$$q(\lambda \cup (k+1) \cup \mu) - q(\lambda \cup (k) \cup \mu) = (s-1)(s-1-2k) - 2(d_{\mu'} - d_\lambda).$$

(c) *Suppose* β *is another* s*-minimal biliaison type*, *and* h, k *are two integers such that* $\lambda < (h-1)$, $(h) < \beta < (k)$, *and* $(k+1) < \mu$. *Let* $\delta = \lambda \cup (h) \cup \beta \cup (k) \cup \mu$ *and* $\epsilon = \lambda \cup (h-1) \cup \beta \cup (k+1) \cup \mu$. *Then*

$$q(\delta) - q(\epsilon) = 2s(k - h - u_\beta) \geq 2s > 0.$$

We next show that $q(\lambda)$ increases if one inserts a new integer in a sequence λ.

Corollary 8.6. *Let $(k_1, k_2, \ldots, k_u)$ be s-minimal.*

(a) *If $k_u < k < s$, then*

$$q(k_1, k_2, \ldots, k_u, k) \geq q(k_1, k_2, \ldots, k_u) + k(s-k)^2$$

In particular, $q(\lambda) \geq (s-1)^2$ unless $\lambda = \varnothing$.

(b) *If $k_i < k < k_{i+1}$, then*

$$q(k_1, k_2, \ldots, k_i, k, k_{i+1}, \ldots, k_u) \geq q(k_1, k_2, \ldots, k_r) + k(s-k).$$

Proof. Let $\lambda = (k_1, k_2, \ldots, k_u)$ By Proposition 8.5 we have

$$q(\lambda \cup (k)) = q(\lambda) + q(k) - 2d_\lambda(s-k) = q(\lambda) + (s-k)\,(k(s-1) - 2d_\lambda)\,.$$

Thus the first claim follows from

$$d_\lambda = \sum_1^r k_i \leq \frac{1}{2}k(k-1). \tag{8-4}$$

For the second claim, set $\lambda = (k_1, k_2, \ldots, k_i)$ and $\mu = (k_{i+1}, k_{i+2}, \ldots, k_u)$. Using Proposition 8.5 we compute

$$q(\lambda \cup ((k) \cup \mu)) - q(\lambda \cup \mu) = q((k) \cup \mu) - q(\mu) + 2d_\lambda(d_{\mu'} - d_{((k)\cup\mu)'})$$

Now $d_{\mu'} - d_{((k)\cup\mu)'} = -(s-k)$, while by duality and the first claim

$$q((k) \cup \mu) - q(\mu) = q(\mu' \cup (s-k)) - q(\mu') \geq (s-k)k^2.$$

Hence

$$\begin{aligned} q(\lambda \cup ((k) \cup \mu)) - q(\lambda \cup \mu) &\geq (s-k)k^2 - 2d_\lambda(s-k) \\ &= (s-k)(k^2 - 2d_\lambda) \geq k(s-k), \end{aligned}$$

where the last inequality follows from (8-4). □

We now prove a lower bound for $q(\lambda)$ in terms of the residue class of d_λ modulo s.

Proposition 8.7. *Let λ be s-minimal, of degree d congruent to f modulo s, with $0 \leq f < s$. Then*

(a) *If $u_\lambda = 2$, so that $\lambda = (h, k)$ with $h + k \equiv f \pmod{s}$, then*

$$q(\lambda) = \begin{cases} f(s-1)(s-f) + 2h(k-1)s & \text{if } h+k < s, \\ f(s-1)(s-f) + 2(s-k)(s-h-1)s & \text{if } h+k \geq s. \end{cases}$$

(b) *If $u_\lambda \geq 3$ and $s \geq 5$, we have*

$$q(\lambda) \geq 2s + m(f, s),$$

where $m(f, s)$ denotes the minimum of $q(\mu)$ as μ varies among s-minimal *biliaison types satisfying* $u_\mu = 2$ *and* $d_\mu \equiv f$ *or* $d_\mu \equiv s-f \pmod{s}$. *In fact,*

$$m(f,s) = \begin{cases} f(s-1)(s-f) + 2s(f-2) & \text{if } 3 \le f \le s-f \text{ or } f = s-2, s-1, \\ f(s-1)(s-f) + 2s(s-f-2) & \text{if } 3 \le s-f \le f \text{ or } f = 0, 1, 2. \end{cases}$$

This minimum is attained by $\lambda = (1, f-1)$ *and* $\lambda' = (s-f+1, s-1)$ *when* $3 \le f \le s-f$ *or if* $f = s-2, s-1$, *and by* $\lambda = (1, s-f-1)$ *and* $\lambda' = (f+1, s-1)$ *when* $3 \le s-f \le f$ *or* $f = 0, 1, 2$.

Proof. Part (a) is a simple computation. To prove part (b), note that the role of f and $s-f$ is symmetric, reflecting the fact that $q(\lambda) = q(\lambda')$. Thus we can replace λ with λ' whenever convenient. If $\lambda = (k_1, k_2, \ldots, k_r)$ and there are two indices $i < j$ such that $k_i - 1 > k_{i-1}$ and $k_j + 1 < k_{j+1}$, we replace k_i by $k_i - 1$ and k_j by $k_j + 1$ to obtain a new increasing sequence λ_1 with the same degree as λ, hence the same f. Then $q(\lambda) \ge q(\lambda_1) + 2s$ by Proposition 8.5(c). When $u_\lambda = 2$, it follows that the minimum $m(f,s)$ is attained by sequences of the form $(1,k)$ or $(h, s-1)$, as in the statement. When $u_\lambda \ge 3$, iterating the procedure above and passing to the dual word if necessary, we may assume that λ is one of the following sequences:

$$\begin{array}{cl} (1, 2, \ldots, h) & 3 \le h < s \\ (1, 2, \ldots, h, s-m, s-(m-1), \ldots, s-1) & 1 \le m \le h,\ 2 \le h \le s-m-2 \\ (1, 2, \ldots, h, k) & 2 \le h \le k-2 \\ (1, 2, \ldots, h, k, s-m, s-(m-1), \ldots, s-1) & m \le h,\ 1 \le h \le k-2,\ k \le s-m-2 \end{array}$$

If $\lambda = (1, 2, \ldots, s-1)$, we replace it with $(2, \ldots, s-2)$, since

$$q(1, 2, \ldots, s-1) > q(2, \ldots, s-2)$$

If $h \ge 2$, we define

$$\mu = (2, \ldots, h-1, h+1, \ldots)$$

to be the sequence obtained removing 1 and h from λ and adding $h+1$. If $h = 1$, then $\lambda = (1, k, s-1)$ with $3 \le k \le s-3$, in which case we define $\mu = (k+1, s-1)$.

Then $d_\mu = d_\lambda$, $u_\mu = u_\lambda - 1$, hence we will be done by induction on u_λ if we show $q(\lambda) \ge q(\mu) + 2s$. By Proposition 8.5(a) we can assume $\lambda = (1, 2, \ldots, h)$ and $\mu = (2, \ldots, h-1, h+1)$. Then one computes $q(\lambda) - q(\mu) = 2s$. ☐

Remark 8.8. One can show that the bound $q(\lambda) \ge f(s-1)(s-f)$ is equivalent to the bound in [Gruson and Peskine 1978] for the genus of an ACM curve of degree $d > s(s-1)$ not lying on a surface degree $s-1$. They also show that curves of maximal genus are linked to plane curves: in our notation this means $u_\lambda = 1$ if $q(\lambda)$ attains its minimal value $f(s-1)(s-f)$.

Corollary 8.9. *Let λ be s-minimal of degree d congruent to f modulo s, with $0 \le f < s$. If $u_\lambda \ge 2$, then*

$$q(\lambda) \ge \begin{cases} 2s(s-2) & \text{if } f = 0, \\ 3s^2 - 8s+1 & \text{if } f = 1 \text{ or } f = s-1, \\ 2s^2 - 4s + 4 & \text{if } f \notin \{0, 1, s-1\}. \end{cases}$$

Proof. We may assume $s \ge 5$ because the cases $s = 3, 4$ are easily checked; see Table 1. If $f = 0$, 1 or $s-1$, the statement follows immediately from Proposition 8.7. If $f \ne 0, 1, s-1$, again by the smae proposition we have

$$q(\lambda) \ge q(f) + 2s \ge q(2) + 2s = 2s^2 - 4s + 4. \qquad \square$$

Corollary 8.10. *Suppose $s \ge 5$ and let λ be s-minimal. Suppose $q(\lambda) \le (s+1)^2$. Then one of the following occurs*:

(1) $\lambda = \varnothing$ *and* $q(\lambda) = 0$.

(2) $\lambda = (1)$ *or* $\lambda = (s-1)$, *and* $q(\lambda) = (s-1)^2$.

(3) $5 \le s \le 7$ *and* $\lambda = (2)$ *or* $\lambda = (s-2)$, *so that* $q(\lambda) = 2(s-1)(s-2)$.

(4) $s = 6$ *and* $\lambda = (3)$, *so that* $q(\lambda) = 3(s-1)(s-3) = 45$.

(5) $s = 5$ *or* 6 *and* $\lambda = (1, s-1)$.

(6) $s = 5$ *and* $\lambda = (1, 3)$ *or* $\lambda = (2, 4)$, *in which case* $q(\lambda) = 36 = (s+1)^2$.

(7) $s = 5$ *and* $\lambda = (1, 2)$ *or* $\lambda = (3, 4)$, *in which case* $q(\lambda) = 34$.

Furthermore, if $q(\lambda) \le (s-1)^2$, *then either* (1) *or* (2) *occurs. If* $(s-1)^2 < q(\lambda) \le s^2$, *then either* $s = 4$ *and* $\lambda = (2)$ *or* $(1, 3)$, *or* $s = 5$ *and* $\lambda = (2)$ *or* (3).

Proof. Suppose first $\lambda = (f)$. Then $q(\lambda) = f(s-1)(s-f)$. One checks this is bigger than $(s+1)^2$ except in the cases listed in the statement.

Suppose now $u_\lambda \ge 2$. If $f = 0$, then $q(\lambda) \ge 2s(s-2)$ by Corollary 8.9, and this is bigger than $(s+1)^2$ unless $s \le 6$. When $s = 5$ or 6, one checks by hand the only possibility is $\lambda = (1, s-1)$.

If $f = 1$ or $s-1$, the lower bound for $q(\lambda)$ is

$$3s^2 - 8s+1,$$

which is bigger than $(s+1)^2$ unless $s \le 5$. When $s = 5$, one finds the two sequences $\lambda = (1, 3)$ or $\lambda = (2, 4)$.

If $f \ne 0, 1, s-1$, then $q(\lambda) \ge 2s^2 - 4s + 4$ which is bigger than $(s+1)^2$ unless $s \le 5$. When $s = 5$, one finds the two sequences $\lambda = (1, 2)$ or $\lambda = (3, 4)$ for which $q(\lambda) = 34$. $\square$

9. Gonality of a general ACM curve

In this section we give the proof of our main result.

Theorem 9.1. *Assume $\mathbb{K}$ has characteristic zero. Let $C \subset \mathbb{P}^3_{\mathbb{K}}$ be an irreducible, nonsingular ACM curve with h-vector h, and let $s = s_C$, $t = t_C$, $e = e(C)$ and $g = g(C)$. Assume that $s \geq 4$ and that (s, d, g) is not one of the following*: (4, 10, 11), (5, 15, 26), (5, 16, 30), (6, 21, 50), (6, 22, 55), (6, 23, 60), (7, 28, 85), (7, 29, 91), (8, 36, 133).

Suppose there is a smooth surface X of degree s containing C with the following properties:

(1) *The linear system $|tH - C|$ on X contains a reduced curve Γ, such that the irreducible components $D_1, \dots D_r$ are ACM curves, and*

$$\lambda_\Gamma = \lambda_{D_1} \cup \lambda_{D_2} \cup \cdots \cup \lambda_{D_r}$$

is the gap decomposition of λ_Γ.

(2) *The Picard group of X is* $\mathrm{Pic}(X) = \mathbb{Z}[H] \oplus \mathbb{Z}[D_1] \oplus \cdots \oplus \mathbb{Z}[D_r]$.

(3) *C is general in its linear system on X.*

Then

$$\mathrm{gon}(C) = d - l,$$

where $l = l(C)$ is the maximum order of a multisecant of C. Furthermore, with the possible exception of the values of (s, d, g) listed in Proposition 6.10(b), C has finitely many g^1_{d-l}; hence its Clifford index is

$$\mathrm{Cliff}(C) = \mathrm{gon}(C) - 2 = d - l - 2.$$

More precisely:

(a) *If $h(e+1) = 3$, $h(e+2) = 2$, then the gonality of C is $d-e-3$ and there is unique pencil of minimal degree, arising from the unique $(e+3)$-secant line of C* (compare Corollary 7.19).

(b) *if $h(t) = s-2$, $h(t+1) = s-3$, $t > s+3$, but the condition of case (a) above does not occur, then the gonality of C is $d - (t-s+1)$, and there is unique pencil of minimal degree, arising from the unique $(t-s+1)$-secant line of C.*

(c) *if neither case (a) nor (b) above occurs, then the gonality of C is $d-4$, and every g^1_{d-4} on C arises from a 4-secant line, unless either*

   (1) *(s, d, g) is in the list of Proposition 6.10(b), or*

   (2) *$s = 4$, $C \in |C_0 + bH|$ where $b \geq 2$ and C_0 has degree 4 and arithmetic genus 1; in this case $|\mathcal{O}_C(b)|$ is the unique g^1_{d-4} that does not arise from a 4-secant.*

Finally, if C has a complete basepoint-free pencil of degree $k < d-4$, then the pencil arises either from an $(e+3)$-secant line or from a $(t-s+1)$-secant line.

Remark 9.2. The conditions on h in (a) and (b) are not satisfied in any of the cases listed in Proposition 6.10(b).

Proof of Theorem 9.1. The gonality of C is at most $d-4$ by Proposition 3.1.

Suppose $\mathcal{L}$ is a complete basepoint-free pencil of degree k on C, and assume $k \leq d-4$, unless we are in one of the cases listed in Proposition 6.10(b), for which we assume $k \leq d-5$. We will classify these pencils as follows. By the same proposition the bundle $\mathcal{E}$ associated to $\mathcal{L}$ on X satisfies $\Delta(\mathcal{E}) > 0$, and then by Bogomolov's result (Theorem 5.4) it follows that $\mathcal{E}$ is Bogomolov unstable. Let $\mathcal{O}_X(A)$ be the line bundle that destabilizes $\mathcal{E}$. We will show that only the following cases can occur:

(1) for any h-vector, we can have $A = -H$; then by Corollary 5.7 the pencil $\mathcal{L}$ arises from a multisecant line L that is not contained in X. Corollary 4.2 shows that $k = \deg \mathcal{L} = d-4$ and that there is a finite set of such pencils.

(2) when $h(e+1) = 3$ and $h(e+2) = 2$, then C has a unique $(e+3)$-secant line L, and $\mathcal{L} = \mathcal{L}(L)$. In this case $L \subset X$ and $A = L - H$.

(3) if $t > s+3$, $h(t) = s-2$, $h(t+1) = s-3$, then C has a unique $(t-s+1)$-secant line L, and $\mathcal{L} = \mathcal{L}(L)$. In this case $L \subset X$ and $A = L - H$.

(4) $s = 4$, $C \in |C_0 + bH|$ where $b \geq 2$ and C_0 has degree 4 and arithmetic genus 1. In this case $\mathcal{L} = |\mathcal{O}_C(b)|$ and $A = -C_0$. In particular, $\deg \mathcal{L} = d-4$ and $\mathcal{L}$ does not arise from a multisecant.

The statement of the theorem clearly follows from this classification. For the Clifford index, we use the fact, proved in [Coppens and Martens 1991], that $\mathrm{Cliff}(C) = \mathrm{gon}(C) - 2$ when C has a finite number of pencils of minimal degree.

We now proceed to classify the possible basepoint-free complete pencils $\mathcal{L}$ of degree at most $d-4$. Let A be the divisor that destabilizes the bundle $\mathcal{E}$ associated to $\mathcal{L}$. Recall that A sits in an exact sequence

$$0 \to \mathcal{O}_X(A) \to \mathcal{E} \to \mathcal{I}_{W,X}(B) \to 0$$

where W is zero-dimensional and $(A-B).H > 0$. From the exact sequence we see $A - B = 2A + C$ and

$$(2A+C)^2 = (A-B)^2 \geq \Delta(\mathcal{E}) = C^2 - 4k.$$

By Proposition 5.5 we also have $(-A).H > 0$ and $A^2 \geq 0$.

To be able to work effectively with the above inequalities, we write $x = A.H$ for the degree of A, and consider the bilinear form on $\mathrm{Pic}(X)$

$$\phi(D, E) = (D.H)\,(E.H) - s\,(D.E) = \det\begin{bmatrix} D.H & H^2 \\ D.E & E.H \end{bmatrix}.$$

We then obtain the following numerical constraints on x:

$$-d < 2x < 0, \quad x^2 \geq \phi(A, A), \quad x^2 + dx + ks \geq \phi(A, A+C), \tag{9-1}$$

the last two inequalities being equivalent to $A^2 \geq 0$ and $(2A + C)^2 \geq C^2 - 4k$ respectively.

In $\mathrm{Pic}(X)$ we can write $A = \sum a_i D_i + cH$ with $a_i \in \mathbb{Z}$, $c \in \mathbb{Z}$. We wish to show

$$\phi(A, A+C) \geq 0.$$

We first prove $\phi(D_i, D_j) < 0$. Let $\lambda_\Gamma = \lambda_1 \cup \lambda_2 \cup \cdots \cup \lambda_r$ be the gap decomposition of λ_Γ, so that $\lambda_{D_i} = \lambda_i$. If $i < j$, $D_i + D_j$ is ACM with $\lambda_{D_i+D_j} = \lambda_i \cup \lambda_j$ by Theorem 7.21. Since $\phi(D, D) = q(\lambda_D)$ for an ACM curve D with $s_D < s$, by Proposition 8.5

$$\phi(D_i, D_j) = -d_{\lambda_i} d_{\lambda'_j} < 0 \tag{9-2}$$

(note that the formula $\phi(D_i, D_j) = -d_{\lambda_i} d_{\lambda'_j}$ is correct only for $i < j$).

To simplify notation we let $q_i = \phi(D_i, D_i)$ and $b_i = -\sum_{j \neq i} \phi(D_i, D_j)$. We claim that $q_i > 2b_i$ for every i. To prove this let $E_i = \sum_{j \neq i} D_j$. Then

$$\begin{aligned} \phi(\Gamma, \Gamma) &= \phi(D_i + E_i, D_i + E_i) = \phi(D_i, D_i) + \phi(E_i, E_i) + 2\phi(D_i, E_i) \\ &= \phi(E_i, E_i) + q_i - 2b_i; \end{aligned}$$

thus it is enough to show $\phi(\Gamma, \Gamma) > \phi(E_i, E_i)$, that is, $q(\lambda_\Gamma) > q(\lambda_{E_i})$. The latter inequality holds by Corollary 8.6; hence $q_i > 2b_i$.

We now compute

$$\begin{aligned} \phi(A, A) &= \sum_i a_i^2 \phi(D_i, D_i) + 2 \sum_{i<j} a_i a_j \phi(D_i, D_j) \\ &= \sum_i a_i^2 (q_i - b_i) - \sum_i a_i^2 \sum_{j \neq i} \phi(D_i, D_j) + 2 \sum_{i<j} a_i a_j \phi(D_i, D_j) \\ &= \sum_i a_i^2 (q_i - b_i) - \sum_{i<j} (a_i - a_j)^2 \phi(D_i, D_j), \end{aligned}$$

$$\begin{aligned} \phi(A, C) &= \phi\Bigl(\sum_i a_i D_i, t_C H - \sum_j D_j\Bigr) = \phi\Bigl(\sum_i a_i D_i, -\sum_j D_j\Bigr) \\ &= -\sum_{i,j} a_i \phi(D_i, D_j) = -\sum_i a_i (q_i - b_i). \end{aligned}$$

Therefore

$$\phi(A,A)=\sum_i a_i^2(q_i-b_i)-\sum_{i<j}(a_i-a_j)^2\phi(D_i,D_j), \tag{9-3}$$

$$\phi(A,C)=-\sum_i a_i(q_i-b_i), \tag{9-4}$$

$$\phi(A,A+C)=\sum_i(a_i^2-a_i)(q_i-b_i)-\sum_{i<j}(a_i-a_j)^2\phi(D_i,D_j). \tag{9-5}$$

The last equality implies $\phi(A,A+C)\ge 0$ because the a_i are integers, $q_i>2b_i\ge b_i$ and $\phi(D_i,D_j)<0$.

We now show that $\phi(A,A+C)\ge 0$ implies $x\ge -s-1$.

By hypothesis $k\le d-4$; therefore

$$x^2+dx+(d-4)s\ge x^2+dx+ks\ge\phi(A,A+C)\ge 0.$$

Let δ be the discriminant of the equation $x^2+dx+(d-4)s=0$:

$$\delta=d^2-4sd+16s=(d-2s)^2-4s(s-4).$$

Let $y=d-2s$. Since C is ACM and $s=s_C$, we have $d\ge\frac{1}{2}s(s+1)$ by Remark 6.2, hence

$$y-2=d-2s-2\ge\tfrac{1}{2}(s^2-3s-4)\ge\tfrac{1}{2}(s^2-4s).$$

In fact, we can have equality only if $s=4$ and $d=10$, while the hypotheses of the theorem when $s=4$ require d to be at least 11. Thus $y-2>\frac{1}{2}s(s-4)$ and

$$\delta=y^2-4s(s-4)>y^2-8y+16=(y-4)^2.$$

Thus δ is positive, and the equation has two real roots, one smaller than $-d/2$, the other one, say $\bar{x}$, larger than $-d/2$. Since $-d/2<x<0$, we conclude $x\ge\bar{x}$. Furthermore, unless $s=4$ and $d=11$, we have $y-4\ge 0$ under the hypotheses of the theorem, hence

$$\bar{x}=-\frac{d}{2}+\frac{1}{2}\sqrt{\delta}>-\frac{d}{2}+\frac{1}{2}\sqrt{y^2-8y+16}=-\frac{d}{2}+\frac{1}{2}(y-4)=-s-2.$$

The inequality $\bar{x}>-6$ holds also in case $s=4$ and $d=11$. Thus $x\ge -s-1$. Then from $x^2\ge\phi(A,A)$ we see that

$$(s+1)^2\ge\phi(A,A).$$

If all the a_i are zero, then $A=cH$ (this is the case if C is a complete intersection of X and another surface). Since $-s-1\le x=\deg A<0$, we must have $A=-H$.

If not all the a_i are zero, let $1\le i_1<\cdots<i_h\le r$ be the indices for which $a_i\ne 0$. Formula (9-3) holds with this new set of indices, and shows that, if all the

coefficients a_i are nonzero, then $\phi(A, A)$ attains its minimum when all the a_i are equal to 1. Thus

$$\phi(A, A) \geq \phi(D, D),$$

where $D = D_{i_1} + \cdots + D_{i_h}$ is the support of A.

Now D is ACM with biliaison type $\lambda_D = \lambda_{i_1} \cup \cdots \cup \lambda_{i_h}$ by Theorem 7.21. If λ_D is not one of the special cases listed in Corollary 8.10, then

$$\phi(D, D) = q(\lambda_D) > (s+1)^2,$$

contradicting $(s+1)^2 \geq \phi(A, A)$.

Suppose now λ_D is one of the special cases listed in Corollary 8.10. We still have $\phi(A, A) \geq (s-1)^2$ because λ_D is not empty. Before examining the various cases, let us remark that, if only one of the a_i is nonzero, so that

$$A = aD + cH$$

with D irreducible and $a \neq 0$, then either $a = 1$ or $a = -1$. This follows from

$$a^2 = \frac{\phi(A, A)}{\phi(D, D)} \leq \frac{(s+1)^2}{(s-1)^2} < 4.$$

Also note that D is irreducible precisely when λ_D has no gaps, that is, in all cases of Corollary 8.10 except when $s = 5$ or 6 and $\lambda = (1, s-1)$.

To complete the list of Corollary 8.10, observe from Table 1 that for $s = 4$ there are 7 possibilities for λ_D, because $\lambda \neq \varnothing$ and $u_\lambda < 4$, namely

$$(1),\ (2),\ (3),\ (1, 2),\ (1, 3),\ (2, 3),\ (1, 2, 3).$$

Case 1: $\lambda_D \neq (1)$, $\lambda_D \neq (s-1)$, and, when $s = 5$ or 6, $\lambda_D \neq (1, s-1)$.

Then $\phi(D, D) > (s-1)^2$ and λ_D has no gaps by Corollary 8.10. Thus D is irreducible, $A = aD + cH$ with $a = \pm 1$ and

$$(s+1)^2 \geq x^2 \geq \phi(A, A) = a^2\phi(D, D) > (s-1)^2.$$

Hence $x = -s-1$ or $x = -s$.

Case 1a: $a = 1$, $x = -s-1$. In this case $d_D \equiv x \equiv -1 \pmod{s}$, and by Corollary 8.10 we must have $s \leq 5$. Furthermore by the last inequality in (9-1)

$$x^2 + dx + (d-4)s \geq 0,$$

that is

$$s^2 + 2s+1 - sd - d + (d-4)s \geq 0$$

so $d \leq s^2 - 2s+1$. This gives $d \leq 9$ if $s = 4$, and $d \leq 16$ if $s = 5$, while $d \geq \frac{1}{2}s(s+1)$ because C is an ACM curve $s_C = s$. Thus we must have $s = 5$, and examining the list in Corollary 8.10 we find $\lambda_D = (1, 3)$ is the only possibility. Then, for

$\Gamma = tH - C$, we know λ_Γ contains $\lambda_D = (1,3)$ in its gap decomposition and $u_{\lambda_\Gamma} < 5$. This forces $\lambda_\Gamma = \lambda_D$, hence $D = \Gamma$ and therefore

$$d = st - \deg(\Gamma) \geq 25 - 4 = 21$$

a contradiction, so this case does not occur.

Case 1b: $a = 1$, $x = -s$. In this case $d_D \equiv x \equiv 0 \pmod{s}$ and $s^2 = x^2 \geq q(\lambda)$. By Corollary 8.10 the only possibility is $s = 4$ and $\lambda_D = (1,3)$, which forces $D = \Gamma = tH - C$. Furthermore, we must have $\operatorname{gon}(C) = k = d - 4$ for the inequality $x^2 + dx + ks \geq \phi(A, A + C)$ of (9-1) to hold.

Since $x = -4 = \deg(D + cH)$, we see $c = -2$. Now pick an effective divisor $C_0 \in |-A| = |2H - D|$. Then C_0 is ACM with biliaison type $(1,3)$, thus C_0 is up to a deformation with constant cohomology an elliptic quartic. By construction $C \in |C_0 + bH|$ with $b = t - 2 \geq 2$. (Note that $b = 2$ gives $(d, g) = (12, 17)$, which is in the list of Proposition 6.10(b).) For $b \geq 2$ the restriction of $|C_0|$ to C is $|\mathcal{O}_C(b)|$, and is a g^1_{d-4} on C that does not arise from a multisecant.

Case 1c: $a = -1$, $x = -s-1$ or $-s$. In this case $A = -D + cH$, hence, if $D = D_i$,

$$\phi(A, A) + \phi(A, C) = 2\phi(D_i, D_i) + \sum_{j \neq i} \phi(-D_i, -D_j) = 2q_i - b_i \geq \tfrac{3}{2} q_i > \tfrac{3}{2}(s-1)^2.$$

Therefore

$$x^2 + dx + (d-4)s \geq \tfrac{3}{2}(s-1)^2,$$

which contradicts both $x = -s-1$ and $x = -s$, so this case does not occur.

Case 2: $\lambda_D = (1)$, so that D is a line $L \subset X$, and $A = cH + aL$ with $a = \pm 1$. In this case either $\Gamma = L$ and $\lambda_\Gamma = (1)$, or λ_Γ has a gap at the beginning:

$$\lambda_\Gamma = (1, 4, \dots)$$

In both cases $L = D_1$ is unique. The proof of Corollary 7.19 shows that the h-vector of C satisfies $h_C(e+1) = 3$ and $h_C(e+2) = 2$, and that $C.L = e + 3$. Thus in any case

$$\deg(Z) = \operatorname{gon}(C) \leq d - e - 3.$$

We wish to show that $A = L - H$ and $Z = \mathfrak{X}(L)$.

Recall that the degree x of A must satisfy the inequalities $-s-1 < x < 0$ and

$$x^2 \geq a^2 \phi(L, L) = (s-1)^2.$$

We also know $x = cs + a$ with $a = \pm 1$. Therefore $c = -1$ and either $A = -H - L$ or $A = -H + L$.

Suppose first $A = -H - L$. Since $\deg(X) = s \geq 4$,

$$H^0 \mathcal{O}_X(H + L) \cong H^0 \mathcal{O}_X(H)$$

thus every curve B in the linear system $|-A| = |H+L|$ contains the line L. This contradicts Proposition 5.5, according to which we can find two effective divisors in $|-A|$ meeting properly. So $A = -H-L$ is impossible. Therefore $A = -H+L$, and $Z = \mathfrak{X}(L)$ by Corollary 5.7.

Case 3: $\lambda_D = (s-1)$, so that $D = H-L$ is a plane curve of degree $s-1$, residual to a line L in a plane section of X. Furthermore, $A = cH + aD = (c+a)H - aL$ with $a = \pm 1$.

In this case $D = D_r$, thus L is unique, and either $\Gamma = D_r$ or λ_Γ has a gap at the end. The proof of Corollary 7.19 shows that the h-vector of C satisfies $h_C(t) = s-2$, $h_C(t+1) = s-3$ and that L is a $(t-s+1)$-secant line for C. An argument analogous to the one of the previous case shows $A = -H+L$, so that $\mathfrak{X} = \mathfrak{X}(L)$.

Case 4: $\lambda_D = (1, s-1)$ with $s = 5$ or 6, hence $A = cH + a_1L_1 + a_2P$ where L_1 is a line, P is a plane curve of degree $s-1$, and a_1 and a_2 are nonzero. Note that $\phi(L_1, P) = -1$, therefore

$$\begin{aligned}\phi(A, A) &= (a_1^2 + a_2^2)(s-1)^2 - 2a_1a_2\\ &= (a_1^2 + a_2^2)(s^2-2s) + (a_1-a_2)^2 \geq 2(s^2-2s) > s^2.\end{aligned}$$

On the other hand, $(s+1)^2 \geq x^2 \geq \phi(A, A)$. Therefore we must have $x = -s-1$ and $a_1^2 + a_2^2 < 3$, that is, a_1 and a_2 can only be 1 or -1.

Then

$$-s-1 = x = cs + a_1 + a_2(s-1),$$

from which we see $-1 \equiv a_1 - a_2 \pmod{s}$. This is impossible because $a_1 = \pm 1$ and $a_2 = \pm 1$.

This complete the list of possible cases, and proves the classification of complete basepoint-free pencils $\mathfrak{X}$ of degree at most $d-4$, hence the theorem □

Remark 9.3. In the first of the cases excluded in the theorem, namely $s = 4$ and $(d, g) = (10, 11)$, we can prove $\operatorname{gon}(C) = 6 = d-4$ by the method of [Hartshorne 2002].

Theorem 9.4. *Assume the ground field is the complex numbers. Then the conclusions of Theorem 9.1 hold for the general ACM curve C in $A(h)$.*

Proof. Since the conclusions of Theorem 9.1 are semicontinuous on $A(h)$ (cf. [Arbarello and Cornalba 1981]), it is enough to show the existence of a single curve C for which the hypotheses of that theorem are satisfied. To check this, let h' denote the h-vector of a curve Γ linked by two surfaces of degrees s and $t = t_C$ to $C \in A(h)$. Note that h' may not be of decreasing type, but in any case $s_\Gamma \leq e_\Gamma + 3 < s$ by Lemma 6.5. By Theorem 7.21 a general curve Γ in $A(h')$ is reduced, its irreducible components are ACM, with biliaison type prescribed by λ_Γ; and, since $s > e_\Gamma + 3$, there exist smooth surfaces of any degree $\geq s-1$ containing Γ.

Now let h_2 be the h-vector of a curve C_2 linked to Γ by the complete intersection of two smooth surfaces of degree $s-1$ and s respectively. The flag Hilbert schemes parametrizing pairs (Γ, Y), where $\Gamma \in A(h')$ and Y is a complete intersection of type $(s-1, s)$, is irreducible [Martin-Deschamps and Perrin 1990, VII §3]. Thus a general Γ in $A(h')$ can be linked to a general $C_2 \in A(h_2)$. By Lemma 6.5 h_2 is of decreasing type, hence we may assume C_2 is smooth, and lies on smooth surfaces of degree $s-1$ and s. Since we are working over the complex numbers, we can use the Noether–Lefschetz type theorem of [Lopez 1991, II 3.1]. We apply this theorem to C_2 with $d = s$, $e = 1$, and T a smooth surface of degree $s-1$ through C_2 to conclude that, if X is a *very general* surface of degree s containing C_2, then $\mathrm{Pic}(X)$ is freely generated by the classes of a plane section H and of the irreducible components of Γ (here "very general" means, as usual, outside a countable union of proper subvarieties).

Now on X we can take for C a general curve in the linear system

$$|C_2 + (t-s+1)H| = |tH - \Gamma|.$$

The hypotheses of Theorem 9.1 are then satisfied for the smooth surface X and the curve C.

One can simplify the argument using a more recent result [Brevik and Nollet 2008, Theorem 1.1], which allows one to work directly with Γ rather than C_2. $\square$

Acknowledgements

The authors thank Pietro Pirola for many valuable discussions and suggestions, and Cecilia Rizzi for several conversations. We are grateful to Tony Geramita who pointed out that our results in Section 7 are related to those in [Davis et al. 1984] and [Davis 1985]. In particular, our Proposition 7.18 is a special case of [Davis 1985, Corollary 4.2].

References

[Arbarello and Cornalba 1981] E. Arbarello and M. Cornalba, "On a conjecture of Petri", *Comment. Math. Helv.* **56**:1 (1981), 1–38. MR 82k:14029

[Arbarello et al. 1985] E. Arbarello, M. Cornalba, P. A. Griffiths, and J. Harris, *Geometry of algebraic curves*, Grundlehren der Math. Wiss. **267**, Springer, New York, 1985. MR 86h:14019 Zbl 0559.14017

[Ballico 1997] E. Ballico, "On the gonality of curves in $\mathbf{P}^n$", *Comment. Math. Univ. Carolin.* **38**:1 (1997), 177–186. MR 98d:14041 Zbl 0888.14010

[Basili 1996] B. Basili, "Indice de Clifford des intersections complètes de l'espace", *Bull. Soc. Math. France* **124**:1 (1996), 61–95. MR 97f:14048 Zbl 0871.14027

[Bogomolov 1978] F. A. Bogomolov, "Holomorphic tensors and vector bundles on projective manifolds", *Izv. Akad. Nauk SSSR Ser. Mat.* **42**:6 (1978), 1227–1287. In Russian; translated in *Math. USSR, Izv.* **13** (1979), 499–555. MR 80j:14014 Zbl 0439.14002

[Brevik and Nollet 2008] J. Brevik and S. Nollet, "Extended Noether–Lefschetz loci", preprint, 2008. arXiv 0806.1243

[Ciliberto 1984] C. Ciliberto, "Alcune applicazioni di un classico procedimento di Castelnuovo", pp. 17–43 in *Seminari di geometria* (Bologna, 1982/1983), Univ. Stud. Bologna, 1984.

[Ciliberto and Lazarsfeld 1984] C. Ciliberto and R. Lazarsfeld, "On the uniqueness of certain linear series on some classes of curves", pp. 198–213 in *Complete intersections* (Acireale, 1983), edited by S. Greco and R. Strano, Lecture Notes in Math. **1092**, Springer, Berlin, 1984. MR 86b:14015 Zbl 0548.14016

[Coppens and Martens 1991] M. Coppens and G. Martens, "Secant spaces and Clifford's theorem", *Compositio Math.* **78**:2 (1991), 193–212. MR 92e:14021 Zbl 0741.14035

[Davis 1985] E. D. Davis, "Complete intersections of codimension 2 in $\mathbb{P}^r$: the Bezout–Jacobi–Segre theorem revisited", *Rend. Sem. Mat. Univ. Politec. Torino* **43**:2 (1985), 333–353. MR 88i:16045 Zbl 0675.14024

[Davis et al. 1984] E. D. Davis, A. V. Geramita, and P. Maroscia, "Perfect homogeneous ideals: Dubreil's theorems revisited", *Bull. Sci. Math.* (2) **108**:2 (1984), 143–185. MR 86m:13024 ZBL 0559.14034

[Ellia and Franco 2001] P. Ellia and D. Franco, "Gonality, Clifford index and multisecants", *J. Pure Appl. Algebra* **158**:1 (2001), 25–39. MR 2002k:14050 Zbl 1012.14011

[Ellingsrud 1975] G. Ellingsrud, "Sur le schéma de Hilbert des variétés de codimension 2 dans $\mathbf{P}^e$ à cône de Cohen–Macaulay", *Ann. Sci. École Norm. Sup.* (4) **8**:4 (1975), 423–431. MR 52 #13831 Zbl 0325.14002

[Farkas 2001] G. Farkas, "Brill–Noether loci and the gonality stratification of $\mathcal{M}_g$", *J. Reine Angew. Math.* **539** (2001), 185–200. MR 2002m:14024 Zbl 0994.14022

[Geramita and Migliore 1994] A. V. Geramita and J. C. Migliore, "A generalized liaison addition", *J. Algebra* **163**:1 (1994), 139–164. MR 94m:14066 Zbl 0798.14026

[Green 1998] M. L. Green, "Generic initial ideals", pp. 119–186 in *Six lectures on commutative algebra* (Bellaterra, 1996), edited by J. Elias et al., Progr. Math. **166**, Birkhäuser, Basel, 1998. MR 99m:13040 Zbl 0933.13002

[Gruson and Peskine 1978] L. Gruson and C. Peskine, "Genre des courbes de l'espace projectif", pp. 31–59 in *Algebraic geometry* (Tromsø, 1977), edited by L. D. Olsen, Lecture Notes in Math. **687**, Springer, Berlin, 1978. MR 81e:14019 Zbl 0412.14011

[Hartshorne 1986] R. Hartshorne, "Generalized divisors on Gorenstein curves and a theorem of Noether", *J. Math. Kyoto Univ.* **26**:3 (1986), 375–386. MR 87k:14036 Zbl 0613.14008

[Hartshorne 1994] R. Hartshorne, "Generalized divisors on Gorenstein schemes", *K-Theory* **8**:3 (1994), 287–339. MR 95k:14008 Zbl 0826.14005

[Hartshorne 2002] R. Hartshorne, "Clifford index of ACM curves in $\mathbb{P}^3$", *Milan J. Math.* **70** (2002), 209–221. MR 2003h:14054 Zbl 1051.14039

[Huybrechts and Lehn 1997] D. Huybrechts and M. Lehn, *The geometry of moduli spaces of sheaves*, Aspects of Mathematics **E31**, Vieweg, Braunschweig, 1997. MR 98g:14012 Zbl 0872.14002

[Lazarsfeld 1986] R. Lazarsfeld, "Brill–Noether–Petri without degenerations", *J. Differential Geom.* **23**:3 (1986), 299–307. MR 88b:14019 Zbl 0608.14026

[Lazarsfeld 1997] R. Lazarsfeld, "Lectures on linear series", pp. 161–219 in *Complex algebraic geometry* (Park City, UT, 1993), edited by J. Kollár, IAS/Park City Math. Ser. **3**, Amer. Math. Soc., Providence, RI, 1997. MR 98h:14008

[Le Barz 1987] P. Le Barz, "Quelques calculs dans les variétés d'alignements", *Adv. in Math.* **64**:2 (1987), 87–117. MR 88e:14065 Zbl 0621.14031

[Lopez 1991] A. F. Lopez, *Noether–Lefschetz theory and the Picard group of projective surfaces*, Mem. Amer. Math. Soc. **438**, American Mathematical Society, Providence, 1991. MR 91f:14030 Zbl 0736.14012

[Lopez and Pirola 1995] A. F. Lopez and G. P. Pirola, "On the curves through a general point of a smooth surface in $\mathbf{P}^3$", *Math. Z.* **219**:1 (1995), 93–106. MR 96h:14024 Zbl 0842.14008

[Maggioni and Ragusa 1988] R. Maggioni and A. Ragusa, "The Hilbert function of generic plane sections of curves of $\mathbf{P}^3$", *Invent. Math.* **91**:2 (1988), 253–258. MR 89g:14027 Zbl 0616.14026

[Martens 1996] G. Martens, "The gonality of curves on a Hirzebruch surface", *Arch. Math.* (*Basel*) **67**:4 (1996), 349–352. MR 97k:14036 Zbl 0872.14022

[Martin-Deschamps and Perrin 1990] M. Martin-Deschamps and D. Perrin, *Sur la classification des courbes gauches*, Astérisque **184-185**, 1990. MR 91h:14039 Zbl 0717.14017

[Migliore 1998] J. C. Migliore, *Introduction to liaison theory and deficiency modules*, Progress in Mathematics **165**, Birkhäuser, Boston, MA, 1998. MR 2000g:14058 Zbl 0921.14033

[Nollet 1998] S. Nollet, "Bounds on multisecant lines", *Collect. Math.* **49**:2-3 (1998), 447–463. MR 2000c:14072 Zbl 1069.49030

[Peskine and Szpiro 1974] C. Peskine and L. Szpiro, "Liaison des variétés algébriques, I", *Invent. Math.* **26** (1974), 271–302. MR 51 #526 Zbl 0298.14022

[Strano 2004] R. Strano, "Biliaison classes of curves in $\mathbf{P}^3$", *Proc. Amer. Math. Soc.* **132**:3 (2004), 649–658. MR 2005a:14065 Zbl 1053.14036

Received June 11, 2010.

ROBIN HARTSHORNE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720-3840
UNITED STATES
robin@math.berkeley.edu

ENRICO SCHLESINGER
DIPARTIMENTO DI MATEMATICA "F. BRIOSCHI"
POLITECNICO DI MILANO
PIAZZA LEONARDO DA VINCI, 32
I-20122 MILANO
ITALY
enrico.schlesinger@polimi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

UNIVERSAL INEQUALITIES FOR THE EIGENVALUES OF THE BIHARMONIC OPERATOR ON SUBMANIFOLDS

SAÏD ILIAS AND OLA MAKHOUL

We establish universal inequalities for the eigenvalues of the clamped plate problem on compact submanifolds of Euclidean space, of spheres and of real, complex and quaternionic projective spaces. We prove similar results for the biharmonic operator on domains of Riemannian manifolds that admit spherical eigenmaps (this includes compact homogeneous Riemannian spaces) and finally on domains of hyperbolic space.

1. Introduction

Let (M, g) be a Riemannian manifold of dimension n and let Δ be the Laplacian operator on M.

We will be concerned with the following eigenvalue problem for the Dirichlet biharmonic operator, called the clamped plate problem:

$$\begin{cases} \Delta^2 u = \lambda u & \text{in } \Omega, \\ u = \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-1}$$

where Ω is a bounded domain in M, Δ^2 is the biharmonic operator in M and ν is the outward unit normal. It is well known that the eigenvalues of this problem form a countable family $0 < \lambda_1 \leq \lambda_2 \leq \cdots \to +\infty$.

For the case when $M = \mathbb{R}^n$, Payne, Pólya and Weinberger [1956] established the following inequality, for each $k \geq 1$:

$$\lambda_{k+1} - \lambda_k \leq \frac{8(n+2)}{n^2 k} \sum_{i=1}^{k} \lambda_i .$$

This work was partially supported by the Agence Nationale de la Recherche through the FOG project (ANR-07-BLAN-0251-01).
MSC2000: 35P15, 58A10, 58C40, 58J50.
Keywords: eigenvalue, biharmonic operator, universal inequality, submanifold, eigenmap.

Implicit in [Payne et al. 1956], as noticed by Ashbaugh [1999], is the better inequality

$$\lambda_{k+1}-\lambda_k \le \frac{8(n+2)}{n^2k^2}\Bigl(\sum_{i=1}^{k}\lambda_i^{1/2}\Bigr)^2. \tag{1-2}$$

Later, Hile and Yeh [1984] extended ideas from earlier work on the Laplacian by Hile and Protter [1980] and proved the better bound

$$\frac{n^2k^{3/2}}{8(n+2)} \le \Bigl(\sum_{i=1}^{k}\frac{\lambda_i^{1/2}}{\lambda_{k+1}-\lambda_i}\Bigr)\Bigl(\sum_{i=1}^{k}\lambda_i\Bigr)^{1/2}.$$

Implicit in their work is the stronger inequality

$$\frac{n^2k^2}{8(n+2)} \le \Bigl(\sum_{i=1}^{k}\frac{\lambda_i^{1/2}}{\lambda_{k+1}-\lambda_i}\Bigr)\Bigl(\sum_{i=1}^{k}\lambda_i^{1/2}\Bigr),$$

which was proved independently by Hook [1990] and Chen and Qian [1990]; see also [Chen and Qian 1993a; 1993b; 1994].

Cheng and Yang [2006] obtained the bound

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i) \le \Bigl(\frac{8(n+2)}{n^2}\Bigr)^{1/2}\sum_{i=1}^{k}\bigl(\lambda_i(\lambda_{k+1}-\lambda_i)\bigr)^{1/2}. \tag{1-3}$$

Very recently, Cheng, Ichikawa and Mametsuka [2009b] obtained an inequality for eigenvalues of Laplacian with any order l on a bounded domain in $\mathbb{R}^n$. In particular, they showed that for $l=2$,

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \le \frac{8(n+2)}{n^2}\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\lambda_i. \tag{1-4}$$

For the case when $M=\mathbb{S}^n$, Wang and Xia [2007] showed that

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \le \frac{1}{n}\Bigl(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2\bigl(n^2+(2n+4)\lambda_i^{1/2}\bigr)\Bigr)^{1/2} \times\Bigl(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\ (n^2+4\lambda_i^{1/2})\Bigr)^{1/2}, \tag{1-5}$$

from which they deduced, using a variant of Chebyshev's inequality,

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \le \frac{1}{n^2}\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\bigl(2(n+2)\lambda_i^{1/2}+n^2\bigr)(4\lambda_i^{1/2}+n^2). \tag{1-6}$$

This last inequality was also obtained by a different method by Cheng, Ichikawa and Mametsuka [2009a].

On the other hand, Wang and Xia [2007] also considered the problem (1-1) on domains of an n-dimensional complete minimal submanifold M of $\mathbb{R}^m$ and proved

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \leq \Big(\frac{8(n+2)}{n^2}\Big)^{1/2}\Big(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2\lambda_i^{1/2}\Big)^{1/2}\Big(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\lambda_i^{1/2}\Big)^{1/2}, \tag{1-7}$$

from which they deduced the following generalization of inequality (1-4) to minimal Euclidean submanifolds:

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \leq \frac{8(n+2)}{n^2}\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\lambda_i. \tag{1-8}$$

Recently, Cheng, Ichikawa and Mametsuka [2010] extended this last inequality to any complete Riemannian submanifold M in $\mathbb{R}^m$ and showed

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \leq \frac{1}{n^2}\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\big(n^2\delta+2(n+2)\lambda_i^{1/2}\big)(n^2\delta+4\lambda_i^{1/2}), \tag{1-9}$$

with

$$\delta=\sup_\Omega |H|^2,$$

where H is the mean curvature of M.

The goal of Section 2 of this article is to study the relation between eigenvalues of the biharmonic operator and the local geometry of Euclidean submanifolds M of arbitrary codimension. The approach is based on an algebraic formula (see Theorem 2.3) we proved in [Ilias and Makhoul 2010]. This approach is useful for the unification and for the generalization of all the results in the literature. In fact, using this general algebraic inequality, we obtain (see Theorem 2.4) the inequality

$$\sum_{i=1}^{k} f(\lambda_i) \leq \frac{1}{n}\Big(\sum_{i=1}^{k} g(\lambda_i)\big(2(n+2)\lambda_i^{1/2}+n^2\delta\big)\Big)^{1/2} \times\Big(\sum_{i=1}^{k}\frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}(4\lambda_i^{1/2}+n^2\delta)\Big)^{1/2}, \tag{1-10}$$

where f and g are two functions satisfying some functional conditions (see Definition 2.1), $\delta=\sup_\Omega |H|^2$ and H is the mean curvature of M. The family of such pairs of functions is large. And particular choices for f and g lead to the known results. For instance, if we take $f(x)=g(x)=(\lambda_{k+1}-x)^2$, then (1-10) becomes

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2 \le \frac{1}{n}\Big(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)^2\big(2(n+2)\lambda_i^{1/2}+n^2\delta\big)\Big)^{1/2} \times\Big(\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)(4\lambda_i^{1/2}+n^2\delta)\Big)^{1/2}, \tag{1-11}$$

which gives easily (see Remark 2.2) inequality (1-9) of Cheng, Ichikawa and Mametsuka [2010].

In Section 3 we consider the case of manifolds admitting spherical eigenmaps and obtain similar results. As a consequence, we obtain universal inequalities for the clamped plate problem on domains of any compact homogeneous Riemannian manifold.

In Section 4, we show how one can easily obtain, from the algebraic techniques used in the previous sections, universal inequalities for eigenvalues of (1-1) on domains of hyperbolic space $\mathbb{H}^n$.

All our results hold if we add a potential to Δ^2 (that is, Δ^2+q where q is a smooth potential). For instance, in this case instead of inequality (1-10), we obtain

$$\sum_{i=1}^{k} f(\lambda_i) \le \frac{1}{n}\Big(\sum_{i=1}^{k} g(\lambda_i)\big(2(n+2)\bar\lambda_i^{1/2}+n^2\delta\big)\Big)^{1/2} \times\Big(\sum_{i=1}^{k}\frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}(4\bar\lambda_i^{1/2}+n^2\delta)\Big)^{1/2}, \tag{1-12}$$

where $\bar\lambda_i=\lambda_i-\inf_\Omega q$.

Finally, the case of the clamped problem with weight

$$\begin{cases}\Delta^2 u=\lambda\,\rho\,u & \text{in } \Omega,\\ u=\dfrac{\partial u}{\partial\nu}=0 & \text{on } \partial\Omega,\end{cases} \tag{1-13}$$

can be easily treated with minor changes.

2. Euclidean submanifolds

Before stating the main result of this section, we introduce a family of pairs of functions and a theorem obtained in [Ilias and Makhoul 2010], which will play an essential role in the proofs of all our results.

Definition 2.1. Let $\lambda\in\mathbb{R}$. A pair (f,g) of functions defined on $]-\infty,\lambda[$ belongs to $\Im_\lambda$ if f and g are positive and, for any distinct $x,\ y\in\,]-\infty,\lambda[$,

$$\Big(\frac{f(x)-f(y)}{x-y}\Big)^2+\Big(\frac{\big(f(x)\big)^2}{g(x)(\lambda-x)}+\frac{\big(f(y)\big)^2}{g(y)(\lambda-y)}\Big)\Big(\frac{g(x)-g(y)}{x-y}\Big)\le 0. \tag{2-1}$$

Remark 2.2. This definition of the family $\Im_\lambda$ differs slightly from that given in [Ilias and Makhoul 2010], but all the results there are still valid.

A direct consequence of our definition is that g must be nonincreasing.

If we multiply f and g of $\Im_\lambda$ by positive constants, the resulting functions are also in $\Im_\lambda$. In the case where f and g are differentiable, one can easily deduce from (2-1) the necessary condition

$$\big((\ln f(x))'\big)^2 \le \frac{-2}{\lambda - x}(\ln g(x))'.$$

This last condition helps us to find many pairs (f, g) satisfying the conditions of Definition 2.1, for example,

$$\{(1, (\lambda-x)^\alpha) \mid \alpha \ge 0\},$$
$$\{((\lambda-x), (\lambda-x)^\beta) \mid \beta \ge \tfrac{1}{2}\},$$
$$\{((\lambda-x)^\delta, (\lambda-x)^\delta) \mid 0 < \delta \le 2\}.$$

Let $\mathscr{H}$ be a complex Hilbert space with scalar product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\|$. For any two operators A and B, we denote by $[A, B]$ their commutator, defined by $[A, B] = AB - BA$.

Theorem 2.3. *Let $A : \mathfrak{D} \subset \mathscr{H} \to \mathscr{H}$ be a self-adjoint operator defined on a dense domain $\mathfrak{D}$, which is semibounded below and has a discrete spectrum*

$$\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots.$$

Let

$$\{T_p : \mathfrak{D} \to \mathscr{H}\}_{p=1}^n$$

be a collection of skew-symmetric operators and

$$\{B_p : T_p(\mathfrak{D}) \to \mathscr{H}\}_{p=1}^n$$

a collection of symmetric operators, leaving $\mathfrak{D}$ invariant. Denote by

$$\{u_i\}_{i=1}^\infty$$

a basis of orthonormal eigenvectors of A, u_i corresponding to λ_i. Let $k \ge 1$ and assume that $\lambda_{k+1} > \lambda_k$. Then, for any (f, g) in $\Im_{\lambda_{k+1}}$

$$\left(\sum_{i=1}^k \sum_{p=1}^n f(\lambda_i)\big\langle [T_p, B_p]u_i, u_i\big\rangle\right)^2 \le 4\left(\sum_{i=1}^k\sum_{p=1}^n g(\lambda_i)\big\langle [A, B_p]u_i, B_p u_i\big\rangle\right) \times \left(\sum_{i=1}^k\sum_{p=1}^n \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}\|T_p u_i\|^2\right). \tag{2-2}$$

Our first result is the following application of this inequality to the eigenvalues of the clamped plate problem (1-1) on a domain of a Euclidean submanifold:

Theorem 2.4. *Let $X : M \to \mathbb{R}^m$ be an isometric immersion of an n-dimensional Riemannian manifold M in $\mathbb{R}^m$. Let Ω be a bounded domain of M and consider the clamped plate problem (1-1) on Ω. Then for any $k \geq 1$ such that $\lambda_{k+1} > \lambda_k$ and for any (f, g) in $\Im_{\lambda_{k+1}}$, we have*

$$\sum_{i=1}^{k} f(\lambda_i) \leq \frac{2}{n}\Bigl(\sum_{i=1}^{k} g(\lambda_i)\bigl(2(n+2)\lambda_i^{1/2}+n^2\delta\bigr)\Bigr)^{1/2} \times \Bigl(\sum_{i=1}^{k} \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}\Bigl(\lambda_i^{1/2}+\frac{n^2}{4}\delta\Bigr)\Bigr)^{1/2}, \tag{2-3}$$

where $\delta = \sup_\Omega |H|^2$ and H be the mean curvature vector field of the immersion X (that is, which is given by $\frac{1}{n}$ trace h, where h is the second fundamental form of X).

Proof. We apply inequality (2-2) of Theorem 2.3 with $A = \Delta^2$, $B_p = X_p$ and $T_p = [\Delta, X_p]$, $p = 1, \dots, m$, where $X_1, \dots, X_m$ are the components of the immersion X. This gives

$$\Bigl(\sum_{i=1}^{k}\sum_{p=1}^{m} f(\lambda_i)\bigl\langle[[\Delta, X_p], X_p]u_i, u_i\bigr\rangle_{L^2}\Bigr)^2 \leq 4\Bigl(\sum_{i=1}^{k}\sum_{p=1}^{m} g(\lambda_i)\bigl\langle[\Delta^2, X_p]u_i, X_p u_i\bigr\rangle_{L^2}\Bigr) \times \Bigl(\sum_{i=1}^{k}\sum_{p=1}^{m} \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}\|[\Delta, X_p]u_i\|_{L^2}^2\Bigr), \tag{2-4}$$

where u_i are the L^2-normalized eigenfunctions. First we have, for $p = 1, \dots, m$,

$$[\Delta^2, X_p]u_i = \Delta^2 X_p u_i + 2\nabla\Delta X_p.\nabla u_i + 2\Delta(\nabla X_p\cdot\nabla u_i) + 2\Delta X_p \Delta u_i + 2\nabla X_p.\nabla\Delta u_i.$$

Thus

$$\begin{aligned}
\langle [\Delta^2, X_p]u_i, X_p u_i\rangle_{L^2}
&= \int_\Omega u_i^2 X_p \Delta^2 X_p + 2\int_\Omega X_p u_i \nabla\Delta X_p.\nabla u_i + 2\int_\Omega X_p u_i \Delta(\nabla X_p\cdot\nabla u_i) \\
&\qquad + 2\int_\Omega X_p u_i \Delta X_p \Delta u_i + 2\int_\Omega X_p u_i \nabla X_p.\nabla\Delta u_i \\
&= \int_\Omega \Delta X_p \Delta(X_p u_i^2) - 2\int_\Omega \operatorname{div}(X_p u_i \nabla u_i)\Delta X_p + 2\int_\Omega \Delta(X_p u_i)\nabla X_p\cdot\nabla u_i \\
&\qquad + 2\int_\Omega X_p \Delta X_p u_i \Delta u_i - 2\int_\Omega \operatorname{div}(X_p u_i \nabla X_p)\Delta u_i.
\end{aligned}$$

A straightforward calculation gives

$$\langle [\Delta^2, X_p]u_i, X_p u_i\rangle_{L^2} = 4\int_\Omega u_i \Delta X_p \nabla X_p \cdot \nabla u_i + \int_\Omega (\Delta X_p)^2 u_i^2 + 4\int_\Omega (\nabla X_p \cdot \nabla u_i)^2 - 2\int_\Omega |\nabla X_p|^2 u_i \Delta u_i. \tag{2-5}$$

Since X is an isometric immersion, we have

$$nH = (\Delta X_1, \dots, \Delta X_m) \tag{2-6}$$

and

$$\sum_{p=1}^m u_i \Delta X_p \nabla X_p \cdot \nabla u_i = 0, \quad \sum_{p=1}^m (\nabla X_p \cdot \nabla u_i)^2 = |\nabla u_i|^2. \tag{2-7}$$

Incorporating these identities in (2-5) and summing on p from 1 to m, we obtain

$$\begin{aligned}
\sum_{p=1}^m \langle [\Delta^2, X_p]u_i, X_p u_i\rangle_{L^2} &= 4\int_\Omega |\nabla u_i|^2 - 2n\int_\Omega u_i \Delta u_i + n^2 \int_\Omega |H|^2 u_i^2 \\
&= 2(n+2)\int_\Omega u_i(-\Delta u_i) + n^2\int_\Omega |H|^2 u_i^2 \\
&\le 2(n+2)\Big(\int_\Omega (-\Delta u_i)^2\Big)^{1/2}\Big(\int_\Omega u_i^2\Big)^{1/2} + n^2\int_\Omega |H|^2 u_i^2 \qquad (2\text{-}8) \\
&= 2(n+2)\lambda_i^{1/2} + n^2 \int_\Omega |H|^2 u_i^2 \\
&\le 2(n+2)\lambda_i^{1/2} + n^2\delta, \qquad (2\text{-}9)
\end{aligned}$$

where the Cauchy–Schwarz inequality gave (2-8) and where $\delta = \sup_\Omega |H|^2$.

On the other hand, we have

$$[\Delta, X_p]u_i = 2\nabla X_p \cdot \nabla u_i + u_i \Delta X_p.$$

Then

$$\begin{aligned}
\sum_{p=1}^m \|[\Delta, X_p]u_i\|_{L^2}^2 &= \sum_{p=1}^m \int_\Omega (2\nabla X_p \cdot \nabla u_i + u_i \Delta X_p)^2 \\
&= 4\sum_{p=1}^m \int_\Omega (\nabla X_p \cdot \nabla u_i)^2 + 4\sum_{p=1}^m \int_\Omega u_i \Delta X_p \nabla X_p \cdot \nabla u_i \\
&\quad + \sum_{p=1}^m \int_\Omega (\Delta X_p)^2 u_i^2.
\end{aligned}$$

Using the identities (2-6) and (2-7), we obtain

$$\begin{aligned}\sum_{p=1}^{m} \|[\Delta, X_p]u_i\|_{L^2}^2 &= 4\int_\Omega |\nabla u_i|^2 + n^2 \int_\Omega |H|^2 u_i^2 \\ &= 4\int_\Omega (-\Delta u_i)\cdot u_i + n^2 \int_\Omega |H|^2 u_i^2 \\ &\le 4\Big(\int_\Omega (-\Delta u_i)^2\Big)^{1/2}\Big(\int_\Omega u_i^2\Big)^{1/2} + n^2\delta \\ &= 4\lambda_i^{1/2} + n^2\delta.\end{aligned} \tag{2-10}$$

A direct calculation gives

$$\begin{aligned}\big\langle [[\Delta, X_p], X_p]u_i, u_i\big\rangle_{L^2} &= \int_\Omega \big(\Delta(X_p^2 u_i) - 2X_p\Delta(X_p u_i) + X_p^2 \Delta u_i\big)u_i \\ &= 2\int_\Omega |\nabla X_p|^2 u_i^2.\end{aligned}$$

Therefore

$$\sum_{p=1}^{m} \big\langle [[\Delta, X_p], X_p]u_i, u_i\big\rangle_{L^2} = 2\sum_{p=1}^{m}\int_\Omega |\nabla X_p|^2 u_i^2 = 2n. \tag{2-11}$$

To conclude, we simply use the estimates (2-9), (2-10) and (2-11) together with inequality (2-4). □

Remarks 2.5. • As indicated in the end of the introduction, Theorem 2.4 holds for a general operator $\Delta^2 + q$, where q is a smooth potential. Indeed, this is an immediate consequence of the fact that $[\Delta^2 + q, X_p] = [\Delta^2, X_p]$ and the entire proof of Theorem 2.4 works in this situation. The only modification is in the estimation of the term $\int_\Omega |\nabla u_i|^2$. In this case, letting $\bar\lambda_i = \lambda_i - \inf_\Omega q$, we have

$$\int_\Omega |\nabla u_i|^2 \le \Big(\int_\Omega (-\Delta u_i)^2\Big)^{1/2}\Big(\int_\Omega u_i^2\Big)^{1/2} = \Big(\lambda_i - \int_\Omega q u_i^2\Big)^{1/2} \le (\bar\lambda_i)^{1/2}.$$

Taking into account this modification in inequalities (2-8) and (2-10), we obtain inequality (1-12).

• If $f(x) = g(x) = (\lambda_{k+1} - x)^2$, then inequality (2-3) extends inequality (1-7) of Wang and Xia [2007] to any Riemannian submanifolds of $\mathbb{R}^m$. By using a Chebyshev inequality (for instance the one of [Cheng et al. 2009b, Lemma 1]), inequality (1-9) of Cheng, Ichikawa and Mametsuka [2010] can be easily deduced from inequality (2-3).

• If $f(x) = g(x)^2 = (\lambda_{k+1} - x)$, then inequality (2-3) generalizes inequality (1-3) of Cheng and Yang [2006] to the case of Euclidean submanifolds.

Using the standard embeddings of the rank one compact symmetric spaces in a Euclidean space (see for instance [El Soufi et al. 2009, Lemma 3.1] for the values of $|H|^2$ of these embeddings), we can extend easily the previous theorem to domains or submanifolds of these symmetric spaces and obtain:

Theorem 2.6. *Let $\overline{M}$ be the sphere $\mathbb{S}^m$, the real projective space $\mathbb{R}P^m$, the complex projective space $\mathbb{C}P^m$ or the quaternionic projective space $\mathbb{Q}P^m$ endowed with their respective metrics. Let (M, g) be a compact Riemannian manifold of dimension n and let $X : M \to \overline{M}$ be an isometric immersion of mean curvature H. Consider the clamped plate problem on a bounded domain Ω of M. For any $k \geq 1$ such that $\lambda_{k+1} > \lambda_k$ and for any $(f, g) \in \Im_{\lambda_{k+1}}$, we have*

$$\text{(2-12)}\quad \sum_{i=1}^{k} f(\lambda_i) \leq \frac{2}{n}\Big(\sum_{i=1}^{k} g(\lambda_i)\big(2(n+2)\lambda_i^{1/2}+n^2\delta'\big)\Big)^{1/2} \times \Big(\sum_{i=1}^{k} \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1}-\lambda_i)}\Big(\lambda_i^{1/2}+\frac{n^2}{4}\delta'\Big)\Big)^{1/2},$$

where

$$\delta' = \sup(|H|^2 + d(n)), \quad \text{where } d(n) = \begin{cases} 1 & \text{if } \overline{M} = \mathbb{S}^m, \\ 2(n{+}1)/n & \text{if } \overline{M} = \mathbb{R}P^m, \\ 2(n{+}2)/n & \text{if } \overline{M} = \mathbb{C}P^m, \\ 2(n{+}4)/n & \text{if } \overline{M} = \mathbb{Q}P^m. \end{cases}$$

Remarks 2.7. • As in [El Soufi et al. 2009, Remark 3.2], in some special geometrical situations, the constant $d(n)$ in the inequality of Theorem 2.6 can be replaced by a sharper one. For instance, when $\overline{M} = \mathbb{C}P^m$ and

– M is odd-dimensional, then $d(n)$ can be replaced by $d'(n) = (2/n)(n+2-1/n)$,

– $X(M)$ is totally real, then $d(n)$ can be replaced by $d'(n) = 2(n+1)/n$.

• When $f(x) = g(x) = (\lambda_{k+1} - x)^2$, and $\overline{M}$ is a sphere, (2-12) generalizes to submanifolds inequality (1-5) established by Wang and Xia for spherical domains.

• As for Theorem 2.4, the result of Theorem 2.6 holds for a more general operator $\Delta^2 + q$, with the same modification (that is, $\bar{\lambda}_i^{1/2}$ instead of $\lambda_i^{1/2}$).

3. Manifolds admitting spherical eigenmaps

In this section, as before, we let (M, g) be a Riemannian manifold and Ω be a bounded domain of M. A map $X : (M, g) \to \mathbb{S}^{m-1}$ is called an eigenmap if its components $X_1, X_2, \ldots, X_m$ are all eigenfunctions associated to the same eigenvalue λ of the Laplacian of (M, g). This is equivalent to say that the map X is a harmonic map from (M, g) into $\mathbb{S}^{m-1}$ with constant energy λ (that is, $\sum_{p=1}^{m} |\nabla X_p|^2 = \lambda$). The most important examples of such manifolds M are the compact homogeneous

Riemannian manifolds. In fact, they admit eigenmaps for all the positive eigenvalues of their Laplacian; see [Li 1980].

Theorem 3.1. *Let λ be an eigenvalue of the Laplacian of (M, g) and suppose that (M, g) admits an eigenmap X associated to this eigenvalue λ. Let Ω be a bounded domain of M and consider the clamped plate problem (1-1) on Ω. For any $k \geq 1$ such that $\lambda_{k+1} > \lambda_k$ and for any $(f, g) \in \Im_{\lambda_{k+1}}$, we have*

$$\sum_{i=1}^{k} f(\lambda_i) \leq \left(\sum_{i=1}^{k} g(\lambda_i)(\lambda + 6\lambda_i^{1/2})\right)^{1/2} \left(\sum_{i=1}^{k} \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1} - \lambda_i)}(\lambda + 4\lambda_i^{1/2})\right)^{1/2}. \tag{3-1}$$

Proof. As in the proof of Theorem 2.4, we apply Theorem 2.3 with $A = \Delta^2$, $B_p = X_p$ and $T_p = [\Delta, X_p]$, $p = 1, \ldots, m$, to obtain

$$\left(\sum_{i=1}^{k}\sum_{p=1}^{m} f(\lambda_i)\langle [[\Delta, X_p], X_p]u_i, u_i\rangle_{L^2}\right)^2 \leq 4\left(\sum_{i=1}^{k}\sum_{p=1}^{m} g(\lambda_i)\langle [\Delta^2, X_p]u_i, X_p u_i\rangle_{L^2}\right) \times \left(\sum_{i=1}^{k}\sum_{p=1}^{m} \frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1} - \lambda_i)} \|[\Delta, X_p]u_i\|_{L^2}^2\right), \tag{3-2}$$

where $\{u_i\}_{i=1}^{\infty}$ is a complete L^2-orthonormal basis of eigenfunctions of Δ^2 associated to $\{\lambda_i\}_{i=1}^{\infty}$. As in (2-11), and using the equality

$$\sum_{p=1}^{m} |\nabla X_p|^2 = \lambda,$$

we have

$$\sum_{p=1}^{m} \big\langle [[\Delta, X_p], X_p]u_i, u_i\big\rangle_{L^2} = 2\sum_{p=1}^{m} \int_{\Omega} |\nabla X_p|^2 u_i^2 = 2\lambda. \tag{3-3}$$

We further have

$$\begin{aligned}\sum_{p=1}^{m} \|[\Delta, X_p]u_i\|_{L^2}^2 &= \sum_{p=1}^{m} \int_{\Omega} \big([\Delta, X_p]u_i\big)^2 \\ &= 4\int_{\Omega} \sum_{p=1}^{m} (\nabla X_p \cdot \nabla u_i)^2 + \int_{\Omega} \sum_{p=1}^{m} (\Delta X_p)^2 u_i^2 + 4\int_{\Omega} \sum_{p=1}^{m} u_i \Delta X_p \nabla X_p \cdot \nabla u_i\end{aligned}$$

Applying Cauchy–Schwarz and the equalities

$$\sum_{p=1}^{m} X_p^2 = 1 \quad \text{and} \quad X_p = -\lambda X_p,$$

we then obtain

$$\begin{aligned}
&\sum_{p=1}^{m} \|[\Delta, X_p]u_i\|_{L^2}^2 \\
&\qquad \le 4\int_\Omega \sum_{p=1}^{m} |\nabla X_p|^2\, |\nabla u_i|^2 + \lambda^2 \int_\Omega \Bigl(\sum_{p=1}^{m} X_p^2\Bigr) u_i^2 - 2\lambda \int_\Omega u_i \nabla\Bigl(\sum_{p=1}^{m} X_p^2\Bigr)\cdot \nabla u_i \\
&\qquad = 4\lambda \int_\Omega (-\Delta u_i)u_i + \lambda^2 \le 4\lambda \Bigl(\int_\Omega (-\Delta u_i)^2\Bigr)^{1/2} \Bigl(\int_\Omega u_i^2\Bigr)^{1/2} + \lambda^2 \\
&\qquad = 4\lambda\lambda_i^{1/2} + \lambda^2.
\end{aligned}$$

Similarly, we infer from (2-5) that

$$\begin{aligned}
&\sum_{p=1}^{m} \langle [\Delta^2, X_p]u_i, X_p u_i \rangle_{L^2} \\
&\quad = \lambda^2 \int_\Omega u_i^2 - \lambda \int_\Omega \nabla\Bigl(\sum_{p=1}^{m} X_p^2\Bigr)\cdot \nabla u_i^2 + 4\sum_{p=1}^{m} \int_\Omega (\nabla X_p \cdot \nabla u_i)^2 + 2\lambda \int_\Omega (-\Delta u_i)u_i \\
&\quad \le \lambda^2 + 4\int_\Omega \sum_{p=1}^{m} |\nabla X_p|^2\, |\nabla u_i|^2 + 2\lambda \Bigl(\int_\Omega (-\Delta u)^2\Bigr)^{1/2} \Bigl(\int_\Omega u_i^2\Bigr)^{1/2} \\
&\quad \le \lambda^2 + 4\lambda\lambda_i^{1/2} + 2\lambda\lambda_i^{1/2} \\
&\quad = \lambda^2 + 6\lambda\lambda_i^{1/2}.
\end{aligned}$$

Incorporating these two bounds, together with (3-3), in inequality (3-2) gives the theorem. □

Corollary 3.2. *Let (M, g) be a compact homogeneous Riemannian manifold without boundary and let λ_1 be the first nonzero eigenvalue of its Laplacian. Then the inequality (3-1) of Theorem 3.1 holds with $\lambda = \lambda_1$.*

Remark 3.3. As before, one can get a similar result for the operator $\Delta^2 + q$.

4. Domains in hyperbolic space

We turn next to the case of a domain Ω of hyperbolic space. It is easy to establish a universal inequality for eigenvalues of the clamped plate problem (1-1) on Ω in the vein of the preceding ones. Unfortunately, until now we have not succeeded in obtaining a simple generalization for the case of domains of hyperbolic submanifolds. In what follows, we take the half-space model for $\mathbb{H}^n$, that is,

$$\mathbb{H}^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_n > 0\}$$

with the standard metric

$$ds^2 = \frac{dx_1^2 + dx_2^2 + \cdots + dx_n^2}{x_n^2}.$$

In terms of the coordinates $(x_i)_{i=1}^n$, the Laplacian of $\mathbb{H}^n$ is given by

$$\Delta = x_n^2 \sum_{j=1}^n \frac{\partial^2}{\partial x_j \partial x_j} + (2-n)x_n \frac{\partial}{\partial x_n}.$$

Theorem 4.1. *For any $k \geq 1$ such that $\lambda_{k+1} > \lambda_k$, the eigenvalues λ_i of the clamped problem (1-1) on the bounded domain Ω of $\mathbb{H}^n$ must satisfy for any $(f, g) \in \Im_{\lambda_{k+1}}$,*

$$\sum_{i=1}^k f(\lambda_i) \leq \left(\sum_{i=1}^k g(\lambda_i)(6\lambda_i^{1/2} - (n-1)^2) \right)^{1/2} \times \left(\sum_{i=1}^k \left(\frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1} - \lambda_i)} \right) (4\lambda_i^{1/2} - (n-1)^2) \right)^{1/2}. \tag{4-1}$$

Proof. Theorem 2.3 remains valid for $A = \Delta^2$, $B_p = F = \ln x_n$ and $T_p = [\Delta, F]$, for all $p = 1, \dots, n$. Thus, denoting by u_i the eigenfunction corresponding to λ_i, we have

$$\left(\sum_{i=1}^k f(\lambda_i) \big\langle [[\Delta, F], F]u_i, u_i \big\rangle_{L^2} \right)^2 \leq 4 \left(\sum_{i=1}^k g(\lambda_i) \langle [\Delta^2, F]u_i, Fu_i \rangle_{L^2} \right) \times \left(\sum_{i=1}^k \left(\frac{(f(\lambda_i))^2}{g(\lambda_i)(\lambda_{k+1} - \lambda_i)} \right) \|[\Delta, F]u_i\|_{L^2}^2 \right). \tag{4-2}$$

We start with the calculation of

$$\begin{aligned} \big\langle [[\Delta, F], F]u_i, u_i \big\rangle_{L^2} &= \int_\Omega \big([\Delta, F](Fu_i) - F[\Delta, F]u_i \big) u_i \\ &= \int_\Omega \big(\Delta(F^2 u_i) - 2F\Delta(Fu_i) + F^2 \Delta u_i \big) u_i. \end{aligned}$$

Note that

$$\Delta F = 1 - n \quad \text{and} \quad |\nabla F|^2 = 1. \tag{4-3}$$

Thus a direct calculation gives

$$\big\langle [[\Delta, F], F]u_i, u_i\big\rangle_{L^2} = 2\int_\Omega |\nabla F|^2 u_i^2 = 2. \tag{4-4}$$

On the other hand, using again the identities of (4-3), we obtain

$$\begin{aligned} \|[\Delta, F]u_i\|^2_{L^2} &= \int_\Omega (\Delta F u_i + 2\nabla F\cdot\nabla u_i)^2 \\ &= \int_\Omega (\Delta F)^2 u_i^2 + 4\int_\Omega (\nabla F\cdot\nabla u_i)^2 + 4\int_\Omega \Delta F u_i \nabla F\cdot\nabla u_i \\ &= (1-n)^2 + 4\int_\Omega (\nabla F\cdot\nabla u_i)^2 + 4(1-n)\int_\Omega u_i\nabla F\cdot\nabla u_i. \end{aligned} \tag{4-5}$$

But

$$\int_\Omega u_i\nabla F\cdot\nabla u_i = -\int_\Omega u_i\nabla F\cdot\nabla u_i - \int_\Omega u_i^2\Delta F,$$

hence

$$\int_\Omega u_i\nabla F\cdot\nabla u_i = \frac{n-1}{2}. \tag{4-6}$$

Then we infer from (4-3), (4-5) and (4-6) that

$$\begin{aligned} \|[\Delta, F]u_i\|^2_{L^2} &\le -(n-1)^2 + 4\int_\Omega |\nabla F|^2\,|\nabla u_i|^2 \\ &= -(n-1)^2 + 4\int_\Omega |\nabla u_i|^2 = -(n-1)^2 + 4\int_\Omega u_i(-\Delta u_i) \\ &\le -(n-1)^2 + 4\Big(\int_\Omega u_i^2\Big)^{1/2}\Big(\int_\Omega (-\Delta u_i)^2\Big)^{1/2} \\ &= 4\lambda_i^{1/2} - (n-1)^2. \end{aligned} \tag{4-7}$$

Now,

$$\begin{aligned} [\Delta^2, F]u_i &= \Delta^2(Fu_i) - F\Delta^2 u_i = \Delta(\Delta F u_i + 2\nabla F\cdot\nabla u_i + F\Delta u_i) - F\Delta^2 u_i \\ &= 2(1-n)\Delta u_i + 2\Delta(\nabla F\cdot\nabla u_i) + 2\nabla F\cdot\nabla\Delta u_i; \end{aligned} \tag{4-8}$$

thus

$$\begin{aligned} &\langle [\Delta^2, F]u_i, Fu_i\rangle_{L^2} \\ &\quad = 2(1-n)\int_\Omega Fu_i\Delta u_i + 2\int_\Omega Fu_i\Delta(\nabla F\cdot\nabla u_i) + 2\int_\Omega Fu_i\nabla F\cdot\nabla\Delta u_i \\ &\quad = 2(1-n)\int_\Omega Fu_i\Delta u_i + 2\int_\Omega \Delta(Fu_i)\nabla F\cdot\nabla u_i - 2\int_\Omega \operatorname{div}(Fu_i\nabla F)\Delta u_i \\ &\quad = 2\int_\Omega \Delta F u_i\nabla F\cdot\nabla u_i + 4\int_\Omega (\nabla F\cdot\nabla u_i)^2 - 2\int_\Omega |\nabla F|^2 u_i\Delta u_i. \end{aligned}$$

We infer from (4-3) and (4-6) that

$$
\begin{aligned}
\text{(4-9)}\quad \langle [\Delta^2, F]u_i, Fu_i\rangle_{L^2} &\le -(n-1)^2 + 4\int_\Omega |\nabla F|^2\,|\nabla u_i|^2 + 2\int_\Omega u_i(-\Delta u_i)\\
&= -(n-1)^2 + 6\int_\Omega u_i(-\Delta u_i)\\
&\le 6\Big(\int_\Omega u_i^2\Big)^{1/2}\Big(\int_\Omega (-\Delta u_i)^2\Big)^{1/2} - (n-1)^2\\
&= 6\lambda_i^{1/2} - (n-1)^2.
\end{aligned}
$$

Inequality (4-2) along with (4-4), (4-7) and (4-9) gives the theorem. □

Remarks 4.2. • It will be interesting to look for an extension of Theorem 4.1 to domains of hyperbolic submanifolds.

• Our method works for any bounded domain Ω of a Riemannian manifold admitting a function such that $|\nabla h|$ is constant and $|\Delta h| \le C$, where C is a constant.

• As before, we have the same statement as in Theorem 4.1 for the operator $\Delta^2 + q$; it suffices to replace $\lambda_i^{1/2}$ by $\bar{\lambda}_i^{1/2}$.

References

[Ashbaugh 1999] M. S. Ashbaugh, "Isoperimetric and universal inequalities for eigenvalues", pp. 95–139 in *Spectral theory and geometry* (Edinburgh, 1998), edited by B. Davies and Y. Safarov, London Math. Soc. Lecture Note Ser. **273**, Cambridge Univ. Press, Cambridge, 1999. MR 2001a:35131 Zbl 0937.35114

[Chen and Qian 1990] Z. C. Chen and C. L. Qian, "Estimates for discrete spectrum of Laplacian operator with any order", *J. China Univ. Sci. Tech.* **20**:3 (1990), 259–266. MR 92c:35087 Zbl 0748.35022

[Chen and Qian 1993a] Z. Chen and C. Qian, "Hile–Yeh estimates of eigenvalues of polyharmonic operators", *Acta Math. Sinica* (*N.S.*) **9**:4 (1993), 432–437. MR 96m:35239 Zbl 0812.35090

[Chen and Qian 1993b] Z. C. Chen and C. L. Qian, "On the difference of consecutive eigenvalues of uniformly elliptic operators of higher orders", *Chinese Ann. Math. Ser. B* **14**:4 (1993), 435–442. MR 95a:35105 Zbl 0782.35046

[Chen and Qian 1994] Z. C. Chen and C. L. Qian, "On the upper bound of eigenvalues for elliptic equations with higher orders", *J. Math. Anal. Appl.* **186**:3 (1994), 821–834. MR 95f:35181 Zbl 0814.35082

[Cheng and Yang 2006] Q.-M. Cheng and H. Yang, "Inequalities for eigenvalues of a clamped plate problem", *Trans. Amer. Math. Soc.* **358**:6 (2006), 2625–2635. MR 2006m:35263 Zbl 1096.35095

[Cheng et al. 2009a] Q.-M. Cheng, T. Ichikawa, and S. Mametsuka, "Estimates for eigenvalues of the poly-Laplacian with any order in a unit sphere", *Calc. Var. Partial Differential Equations* **36**:4 (2009), 507–523. MR 2010k:35332 Zbl 1180.35389

[Cheng et al. 2009b] Q.-M. Cheng, T. Ichikawa, and S. Mametsuka, "Inequalities for eigenvalues of Laplacian with any order", *Commun. Contemp. Math.* **11**:4 (2009), 639–655. MR 2010j:35113 Zbl 1181.35153

[Cheng et al. 2010] Q.-M. Cheng, T. Ichikawa, and S. Mametsuka, "Estimates for eigenvalues of a clamped plate problem on Riemannian manifolds", *J. Math. Soc. Japan* **62**:2 (2010), 673–686. MR 2011e:58039 Zbl 1191.35192

[El Soufi et al. 2009] A. El Soufi, E. M. Harrell, II, and S. Ilias, "Universal inequalities for the eigenvalues of Laplace and Schrödinger operators on submanifolds", *Trans. Amer. Math. Soc.* **361**:5 (2009), 2337–2350. MR 2010e:58032 Zbl 1162.58009

[Hile and Protter 1980] G. N. Hile and M. H. Protter, "Inequalities for eigenvalues of the Laplacian", *Indiana Univ. Math. J.* **29**:4 (1980), 523–538. MR 82c:35052 Zbl 0454.35064

[Hile and Yeh 1984] G. N. Hile and R. Z. Yeh, "Inequalities for eigenvalues of the biharmonic operator", *Pacific J. Math.* **112**:1 (1984), 115–133. MR 85k:35170 Zbl 0541.35059

[Hook 1990] S. M. Hook, "Domain-independent upper bounds for eigenvalues of elliptic operators", *Trans. Amer. Math. Soc.* **318**:2 (1990), 615–642. MR 90h:35075 Zbl 0727.35096

[Ilias and Makhoul 2010] S. Ilias and O. Makhoul, "Universal inequalities for the eigenvalues of a power of the Laplace operator", *Manuscripta Math.* **132**:1-2 (2010), 75–102. MR 2609289 Zbl 1189.35206

[Li 1980] P. Li, "Eigenvalue estimates on homogeneous manifolds", *Comment. Math. Helv.* **55**:3 (1980), 347–363. MR 81k:58067 Zbl 0451.53036

[Payne et al. 1956] L. E. Payne, G. Pólya, and H. F. Weinberger, "On the ratio of consecutive eigenvalues", *J. Math. and Phys.* **35** (1956), 289–298. MR 18,905c Zbl 0073.08203

[Wang and Xia 2007] Q. Wang and C. Xia, "Universal bounds for eigenvalues of the biharmonic operator on Riemannian manifolds", *J. Funct. Anal.* **245**:1 (2007), 334–352. MR 2008e:58033 Zbl 1113.58013

Received April 28, 2010.

SAÏD ILIAS
UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS
LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR CNRS 6083
PARC DE GRANDMONT
37200 TOURS
FRANCE

ilias@lmpt.univ-tours.fr

OLA MAKHOUL
UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS
LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR CNRS 6083
PARC DE GRANDMONT
37200 TOURS
FRANCE

ola.makhoul@lmpt.univ-tours.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

MULTIGRADED FUJITA APPROXIMATION

SHIN-YAO JOW

The original Fujita approximation theorem states that the volume of a big divisor D on a projective variety X can always be approximated arbitrarily closely by the self-intersection number of an ample divisor on a birational modification of X. One can also formulate it in terms of graded linear series as follows: Let $W_\bullet = \{W_k\}$ be the complete graded linear series associated to a big divisor D, where

$$W_k = H^0(X, \mathcal{O}_X(kD)).$$

For each fixed positive integer p, define $W_\bullet^{(p)}$ to be the graded linear subseries of $W_\bullet$ generated by W_p:

$$W_m^{(p)} = \begin{cases} 0 & \text{if } p \nmid m, \\ \operatorname{Image}(S^k W_p \to W_{kp}) & \text{if } m = kp. \end{cases}$$

Then the volume of $W_\bullet^{(p)}$ approaches the volume of $W_\bullet$ as $p \to \infty$. We will show that, under this formulation, the Fujita approximation theorem can be generalized to the case of multigraded linear series.

1. Introduction

Let X be an irreducible variety of dimension d over an algebraically closed field $\boldsymbol{K}$, and let D be a (Cartier) divisor on X. When X is projective, the following limit, which measures how fast the dimension of the section space $H^0(X, \mathcal{O}_X(mD))$ grows, is called the *volume* of D:

$$\operatorname{vol}(D) = \operatorname{vol}_X(D) = \lim_{m\to\infty} \frac{h^0(X, \mathcal{O}_X(mD))}{m^d/d!}.$$

One says that D is *big* if $\operatorname{vol}(D) > 0$. It turns out that the volume is an interesting numerical invariant of a big divisor [Lazarsfeld 2004a, Section 2.2.C], and it plays a key role in several recent works in birational geometry [Tsuji 2000; Boucksom et al. 2004; Hacon and McKernan 2006; Takayama 2006].

MSC2000: 14C20.
Keywords: Fujita approximation, multigraded linear series, Okounkov body.

When D is ample, one can show that $\mathrm{vol}(D) = D^d$, the self-intersection number of D. This is no longer true for a general big divisor D, since D^d may even be negative. However, Fujita [1994] showed that the volume of a big divisor can always be approximated arbitrarily closely by the self-intersection number of an ample divisor on a birational modification of X. This theorem, known as *Fujita approximation*, has several implications for the properties of volumes, and is also a crucial ingredient in [Boucksom et al. 2004] (see [Lazarsfeld 2004b, Section 11.4] for more details).

Lazarsfeld and Mustaţă [2009] (henceforth [LM]) recently obtained, among other things, a generalization of Fujita approximation to *graded linear series*. Recall that a graded linear series $W_\bullet = \{W_k\}$ on a (not necessarily projective) variety X associated to a divisor D consists of finite dimensional vector subspaces

$$W_k \subseteq H^0(X, \mathbb{O}_X(kD))$$

for each $k \geq 0$, with $W_0 = \boldsymbol{K}$, such that

$$W_k \cdot W_\ell \subseteq W_{k+\ell}$$

for all $k, \ell \geq 0$. Here the product on the left denotes the image of $W_k \otimes W_\ell$ under the multiplication map $H^0(X, \mathbb{O}_X(kD)) \otimes H^0(X, \mathbb{O}_X(\ell D)) \to H^0(X, \mathbb{O}_X((k+\ell)D))$. In order to state the Fujita approximation for $W_\bullet$, they defined, for each fixed positive integer p, a graded linear series $W_\bullet^{(p)}$ which is the subgraded linear series of $W_\bullet$ generated by W_p:

$$W_m^{(p)} = \begin{cases} 0 & \text{if } p \nmid m, \\ \mathrm{Im}(S^k W_p \to W_{kp}) & \text{if } m = kp. \end{cases}$$

Then under mild hypotheses, they showed that the volume of $W_\bullet^{(p)}$ approaches the volume of $W_\bullet$ as $p \to \infty$. See [LM, Theorem 3.5] for the precise statement, as well as [LM, Remark 3.4] for how this is equivalent to the original statement of Fujita when X is projective and $W_\bullet$ is the complete graded linear series associated to a big divisor D (that is, $W_k = H^0(X, \mathbb{O}_X(kD))$ for all $k \geq 0$).

The goal of this note is to generalize the Fujita approximation theorem to *multigraded linear series*. We will adopt the following notation from [LM, Section 4.3]: Let $D_1, \dots, D_r$ be divisors on X. For $\vec{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, write $\vec{m}D = \sum m_i D_i$, and put $|\vec{m}| = \sum |m_i|$.

Definition. A *multigraded linear series* $W_{\vec{\bullet}}$ on X associated to the D_i consists of finite-dimensional vector subspaces

$$W_{\vec{k}} \subseteq H^0(X, \mathbb{O}_X(\vec{k}D))$$

for each $\vec{k} \in \mathbb{N}^r$, with $W_{\vec{0}} = \boldsymbol{K}$, such that

$$W_{\vec{k}} \cdot W_{\vec{m}} \;\subseteq\; W_{\vec{k}+\vec{m}},$$

where the multiplication on the left denotes the image of $W_{\vec{k}} \otimes W_{\vec{m}}$ under the natural map

$$H^0(X, \mathcal{O}_X(\vec{k}D)) \otimes H^0(X, \mathcal{O}_X(\vec{m}D)) \to H^0(X, \mathcal{O}_X((\vec{k}+\vec{m})D)).$$

Given $\vec{a} \in \mathbb{N}^r$, denote by $W_{\vec{a},\bullet}$ the singly graded linear series associated to the divisor $\vec{a}D$ given by the subspaces $W_{k\vec{a}} \subseteq H^0(X, \mathcal{O}_X(k\vec{a}D))$. Then put

$$\mathrm{vol}_{W_{\vec{\bullet}}}(\vec{a}) \;=\; \mathrm{vol}(W_{\vec{a},\bullet})$$

(assuming that this quantity is finite). It will also be convenient for us to consider $W_{\vec{a},\bullet}$ when $\vec{a} \in \mathbb{Q}^r_{\geq 0}$, given by

$$W_{\vec{a},k} = \begin{cases} W_{k\vec{a}} & \text{if } k\vec{a} \in \mathbb{N}^r, \\ 0 & \text{otherwise.} \end{cases}$$

Our multigraded Fujita approximation, similar to the singly graded version, is going to state that (under suitable conditions) the volume of $W_{\vec{\bullet}}$ can be approximated by the volume of the following finitely generated submultigraded linear series of $W_{\vec{\bullet}}$:

Definition. Given a multigraded linear series $W_{\vec{\bullet}}$ and a positive integer p, define $W_{\vec{\bullet}}^{(p)}$ to be the submultigraded linear series of $W_{\vec{\bullet}}$ generated by all $W_{\vec{m}_i}$ with $|\vec{m}_i| = p$, or concretely,

$$W_{\vec{m}}^{(p)} = \begin{cases} 0 & \text{if } p \nmid |\vec{m}|, \\ \displaystyle\sum_{\substack{|\vec{m}_i|=p \\ \vec{m}_1+\cdots+\vec{m}_k=\vec{m}}} W_{\vec{m}_1} \cdots W_{\vec{m}_k} & \text{if } |\vec{m}| = kp. \end{cases}$$

We now state our multigraded Fujita approximation when $W_{\vec{\bullet}}$ is a complete multigraded linear series, since this is the case of most interest and allows for a more streamlined statement. The Remark on page 335 points out what assumptions on $W_{\vec{\bullet}}$ are actually needed in the proof.

Theorem. *Let X be an irreducible projective variety of dimension d, and let D_1, $D_2, \ldots, D_r$ be big divisors on X. Let $W_{\vec{\bullet}}$ be the complete multigraded linear series associated to the D_i, namely*

$$W_{\vec{m}} = H^0(X, \mathcal{O}_X(\vec{m}D))$$

for each $\vec{m} \in \mathbb{N}^r$. Then given any $\varepsilon > 0$, there exists an integer $p_0 = p_0(\varepsilon)$ having the property that if $p \geq p_0$, then

$$\left| 1 - \frac{\mathrm{vol}_{W_{\vec{\bullet}}^{(p)}}(\vec{a})}{\mathrm{vol}_{W_{\vec{\bullet}}}(\vec{a})} \right| < \varepsilon \tag{1}$$

for all $\vec{a} \in \mathbb{N}^r$.

2. Proof of the Theorem

The main tool in our proof is the theory of *Okounkov bodies* developed systematically in [Lazarsfeld and Mustaţă 2009]. Given a graded linear series $W_\bullet$ on a d-dimensional variety X, its Okounkov body $\Delta(W_\bullet)$ is a convex body in $\mathbb{R}^d$ that encodes many asymptotic invariants of $W_\bullet$, the most prominent one being the volume of $W_\bullet$, which is precisely $d!$ times the Euclidean volume of $\Delta(W_\bullet)$. The idea first appeared in Okounkov's papers [1996; 2003] in the case of complete linear series of ample line bundles on a projective variety. Later it was further developed and applied to much more general graded linear series by Lazarsfeld and Mustaţă [2009] and also independently by Kaveh and Khovanskii [2008; 2009].

Proof of the Theorem. Let $T = \{(a_1, \dots, a_r) \in \mathbb{R}^r_{\geq 0} \mid a_1 + \cdots + a_r = 1\}$, and let $T_{\mathbb{Q}}$ be the set of all points in T with rational coordinates. The fraction inside (1) is invariant under scaling of $\vec{a}$ due to homogeneity, hence it is enough to prove (1) for $\vec{a} \in T_{\mathbb{Q}}$.

Let $\Delta(W_{\vec{\bullet}}) \subseteq \mathbb{R}^d \times \mathbb{R}^r$ be the global Okounkov cone of $W_{\vec{\bullet}}$ as in [LM, Theorem 4.19], and let $\pi : \Delta(W_{\vec{\bullet}}) \to \mathbb{R}^r$ be the projection map. For each $\vec{a} \in T$, write $\Delta(W_{\vec{\bullet}})_{\vec{a}}$ for the fiber $\pi^{-1}(\vec{a})$. Define in a similar fashion the convex cone $\Delta(W_{\vec{\bullet}}^{(p)})$ and the convex bodies $\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}}$. By [LM, Theorem 4.19],

$$\Delta(W_{\vec{\bullet}})_{\vec{a}} = \Delta(W_{\vec{a},\bullet}) \quad \text{for all } \vec{a} \in T_{\mathbb{Q}}. \tag{2}$$

Although [LM, Theorem 4.19] requires $\vec{a}$ to be in the relative interior of T, here we know that (2) holds even for those $\vec{a}$ in the boundary of T because the big cone of X is open and $W_{\vec{\bullet}}$ was assumed to be the complete multigraded linear series. By the singly graded Fujita approximation, $\mathrm{vol}(W_{\vec{a},\bullet})$ can be approximated arbitrarily closely by $\mathrm{vol}(W_{\vec{a},\bullet}^{(p)})$ if p is sufficiently large. (Here by $W_{\vec{a},\bullet}^{(p)}$ we mean $W_{\vec{\bullet}}^{(p)}$ restricted to the $\vec{a}$ direction, which certainly contains $(W_{\vec{a},\bullet})^{(p)}$.) Hence given any finite subset $S \subset T_{\mathbb{Q}}$ and any $\varepsilon' > 0$, we have

$$\mathrm{vol}(\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}}) \geq \mathrm{vol}(\Delta(W_{\vec{\bullet}})_{\vec{a}}) - \varepsilon' \quad \text{for all } \vec{a} \in S$$

as soon as p is sufficiently large.

Because the function $\vec{a} \mapsto \mathrm{vol}(\Delta(W_{\vec{\bullet}})_{\vec{a}})$ is uniformly continuous on T, given any $\varepsilon' > 0$, we can partition T into a union of polytopes with disjoint interiors

$T = \bigcup T_i$, in such a way that the vertices of each T_i all have rational coordinates, and on each T_i we have a constant M_i such that

$$M_i \le \mathrm{vol}(\Delta(W_{\vec{\bullet}})_{\vec{a}}) \le M_i + \varepsilon' \quad \text{for all } \vec{a} \in T_i. \tag{3}$$

Let S be the set of vertices of all the T_i. Then as we saw in the end of the previous paragraph, as soon as p is sufficiently large we have

$$\mathrm{vol}(\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}}) \ge \mathrm{vol}(\Delta(W_{\vec{\bullet}})_{\vec{a}}) - \varepsilon' \quad \text{for all } \vec{a} \in S. \tag{4}$$

We claim that this implies

$$\mathrm{vol}(\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}}) \ge \mathrm{vol}(\Delta(W_{\vec{\bullet}})_{\vec{a}}) - 2\varepsilon' \quad \text{for all } \vec{a} \in T_{\mathbb{Q}}. \tag{5}$$

To show this, it suffices to verify it on each of the T_i. Let $\vec{v}_1, \dots, \vec{v}_k$ be the vertices of T_i. Then each $\vec{a} \in T_i$ can be written as a convex combination of the vertices: $\vec{a} = \sum t_j \vec{v}_j$ where each $t_j \ge 0$ and $\sum t_j = 1$. Since $\Delta(W_{\vec{\bullet}}^{(p)})$ is convex, we have

$$\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}} \supseteq \sum t_j \, \Delta(W_{\vec{\bullet}}^{(p)})_{\vec{v}_j},$$

where the sum on the right means the Minkowski sum. By (3) and (4), the volume of each $\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{v}_j}$ is at least $M_i - \varepsilon'$, hence by the Brunn–Minkowski inequality [Kaveh and Khovanskii 2008, Theorem 5.4], we have

$$\mathrm{vol}(\Delta(W_{\vec{\bullet}}^{(p)})_{\vec{a}}) \ge M_i - \varepsilon' \quad \text{for all } \vec{a} \in T_i \cap T_{\mathbb{Q}}.$$

This combined with (3) shows that (5) is true on $T_i \cap T_{\mathbb{Q}}$, hence it is true on $T_{\mathbb{Q}}$ since the T_i cover T.

Since (1) follows from (5) by choosing a suitable ε', the proof is complete. □

Remark. In the statement of the Theorem we assume that $W_{\vec{\bullet}}$ is the complete multigraded linear series associated to big divisors. But in fact since the main tool we used in the proof is the theory of Okounkov bodies established in [Lazarsfeld and Mustaţă 2009], in particular [LM, Theorem 4.19], the really indispensable assumptions on $W_{\vec{\bullet}}$ are the same as those in [LM] (which they called Conditions (A′) and (B′), or (C′)). The only place in the proof where we invoke that we are working with a complete multigraded linear series is the sentence right after (2), where we want to say that (2) holds not only in the relative interior of T but also in its boundary. Hence if $W_{\vec{\bullet}}$ is only assumed to satisfy Conditions (A′) and (B′), or (C′), then given any $\varepsilon > 0$ and any compact set C contained in $T \cap \mathrm{int}(\mathrm{supp}(W_{\vec{\bullet}}))$, there exists an integer $p_0 = p_0(C, \varepsilon)$ such that if $p \ge p_0$ then

$$\mathrm{vol}_{W_{\vec{\bullet}}^{(p)}}(\vec{a}) > \mathrm{vol}_{W_{\vec{\bullet}}}(\vec{a}) - \varepsilon$$

for all $\vec{a} \in C \cap T_{\mathbb{Q}}$.

Acknowledgments

The author would like to thank Robert Lazarsfeld for raising this question during an email correspondence.

References

[Boucksom et al. 2004] S. Boucksom, J.-P. Demailly, M. Paun, and T. Peternell, "The pseudo-effective cone of a compact Kähler manifold and varieties of negative Kodaira dimension", preprint, 2004. arXiv math/0405285

[Fujita 1994] T. Fujita, "Approximating Zariski decomposition of big line bundles", *Kodai Math. J.* **17**:1 (1994), 1–3. MR 95c:14053 Zbl 0814.14006

[Hacon and McKernan 2006] C. D. Hacon and J. McKernan, "Boundedness of pluricanonical maps of varieties of general type", *Invent. Math.* **166**:1 (2006), 1–25. MR 2007e:14022 Zbl 1121.14011

[Kaveh and Khovanskii 2008] K. Kaveh and A. Khovanskii, "Convex bodies and algebraic equations on affine varieties", preprint, 2008. arXiv 0804.4095

[Kaveh and Khovanskii 2009] K. Kaveh and A. Khovanskii, "Newton convex bodies, semigroups of integral points, graded algebras and intersection theory", preprint, 2009. arXiv 0904.3350

[Lazarsfeld 2004a] R. Lazarsfeld, *Positivity in algebraic geometry, I: Classical setting: line bundles and linear series*, Ergebnisse der Math. und ihrer Grenzgebiete (3) **48**, Springer, Berlin, 2004. MR 2005k:14001a Zbl 1093.14501

[Lazarsfeld 2004b] R. Lazarsfeld, *Positivity in algebraic geometry, II: Positivity for vector bundles, and multiplier ideals*, Ergebnisse der Math. und ihrer Grenzgebiete (3) **49**, Springer, Berlin, 2004. MR 2005k:14001b Zbl 1093.14500

[Lazarsfeld and Mustaţă 2009] R. Lazarsfeld and M. Mustaţă, "Convex bodies associated to linear series", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:5 (2009), 783–835. MR 2011e:14012 Zbl 1182.14004

[Okounkov 1996] A. Okounkov, "Brunn–Minkowski inequality for multiplicities", *Invent. Math.* **125**:3 (1996), 405–411. MR 99a:58074 Zbl 0893.52004

[Okounkov 2003] A. Okounkov, "Why would multiplicities be log-concave?", pp. 329–347 in *The orbit method in geometry and physics* (Marseille, 2000), edited by C. Duval et al., Progr. Math. **213**, Birkhäuser, Boston, 2003. MR 2004j:20022 Zbl 1063.22024

[Takayama 2006] S. Takayama, "Pluricanonical systems on algebraic varieties of general type", *Invent. Math.* **165**:3 (2006), 551–587. MR 2007m:14014 Zbl 1108.14031

[Tsuji 2000] H. Tsuji, "Effective birationality of pluricanonical systems", preprint, 2000. arXiv math/0006166

Received May 10, 2010.

SHIN-YAO JOW
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF PENNSYLVANIA
209 SOUTH 33RD STREET
PHILADELPHIA, PA 19104-6395
UNITED STATES

jows@math.upenn.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

SOME DIRICHLET PROBLEMS ARISING FROM CONFORMAL GEOMETRY

QI-RUI LI AND WEIMIN SHENG

We study the problem of finding complete conformal metrics determined by some symmetric function of the modified Schouten tensor on compact manifolds with boundary; which reduces to a Dirichlet problem. We prove the existence of the solution under some suitable conditions. In particular, we prove that every smooth compact n-dimensional manifold with boundary, with $n \geq 3$, admits a complete Riemannian metric g whose Ricci curvature Ric_g and scalar curvature R_g satisfy

$$\det(\mathrm{Ric}_g - R_g g) = \text{const}.$$

This result generalizes Aviles and McOwen's in the scalar curvature case.

1. Introduction

Let $(\overline{M}^n, g)$, for $n \geq 3$, be a compact, n-dimensional smooth Riemannian manifold with smooth boundary ∂M. Let $M = \overline{M} \setminus \partial M$ be the interior of $\overline{M}$, and denote the Ricci tensor and the scalar curvature by Ric and R (or Ric_g and R_g to emphasize the metric), respectively. In [2003], Gursky and Viaclovsky introduced the *modified Schouten tensor*

$$A_g^\tau := \frac{1}{n-2}\Big(\mathrm{Ric}_g - \frac{\tau}{2(n-1)} R_g g\Big),$$

where $\tau \in \mathbb{R}$. We are interested in deforming the metric in the conformal class $[g]$ of a fixed back ground metric g to certain *complete* metric $\bar{g}$ satisfying

$$\det(\bar{g}^{-1} A_{\bar{g}}^\tau) = \text{const in } M.$$

More generally, let Γ^+ be an open convex cone in $\mathbb{R}^n$ with vertex at the origin satisfying $\Gamma_n^+ \subset \Gamma^+ \subset \Gamma_1^+$, where

$$\Gamma_k^+ = \{\kappa = (\kappa_1, \ldots, \kappa_n) \in \mathbb{R}^n \mid \sigma_j(\kappa) > 0, 1 \leq j \leq k\},$$

The authors were supported by NSFC 10771189 and 10831008.
MSC2000: primary 53C21; secondary 53C23.
Keywords: modified Schouten tensor, Dirichlet problem, complete metric, prescribed curvature.

and

$$\sigma_k(\kappa) = \sum_{i_1<\cdots<i_k} \kappa_{i_1} \cdots \kappa_{i_k}.$$

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a smooth symmetric function that satisfies some structure conditions in Γ^+, to be listed later. We ask, Does there exist a complete metric $\bar{g}$ in the conformal class $[g]$ such that

$$F(\bar{g}^{-1} A^\tau_{\bar{g}}) = f(x) \quad \text{in } M \tag{1-1}$$

for some given smooth function $f \in C^\infty(\overline{M})$? In this paper, we give a partial answer in the case $\tau > n-1$. We remark that, if $F = \sigma_1$, then (1-1) becomes

$$\frac{(2-\tau)n-2}{2(n-1)(n-2)} R_{\bar{g}} = f(x).$$

In the case $\tau > n-1$ and $f(x)$ is positive, some results have appeared in [Aviles and McOwen 1988].

To find a complete conformal metric satisfying (1-1), we need to solve the Dirichlet problem for (1-1) with larger and larger boundary data. We first write this curvature equation as a partial differential equation. Recall the following formula for the transformation of A^τ under a conformal change of metric $\bar{g} = e^{2u} g$:

$$A^\tau_{\bar{g}} = \frac{\tau-1}{n-2}(\Delta u)g - \nabla^2 u + du \otimes du + \frac{\tau-2}{2}|\nabla u|^2 g + A^\tau_g. \tag{1-2}$$

From (1-2) we may write (1-1) as

$$F\Big(\frac{\tau-1}{n-2}(\Delta u)g - \nabla^2 u + du \otimes du + \frac{\tau-2}{2}|\nabla u|^2 g + A^\tau_g\Big) = f(x)e^{2u}.$$

In this paper, we study a more general equation. Let $h(x,z) : \overline{M}^n \times \mathbb{R}$ be some smooth positive function. Let's consider

$$F(\lambda(\Delta u)g - \nabla^2 u + a(x)du \otimes du + b(x)|\nabla u|^2 g + B) = h(x,u), \tag{1-3}$$

where $\lambda > 1$, B is a symmetric 2-tensor, and $a(x)$ and $b(x)$ are smooth functions on $\overline{M}$. Suppose F is homogeneous of degree one, $F = 0$ on $\partial\Gamma^+$, and F satisfies the following in Γ^+:

(C1) F is positive;

(C2) F is concave (that is, $\frac{\partial^2 F}{\partial\kappa_i \partial\kappa_j}$ is negative semidefinite);

(C3) F is monotone (that is, $\frac{\partial F}{\partial \kappa_i}$ is positive).

For convenience, we define

$$W[u] := \nabla^2_{\text{conf}} u + B,$$

and

$$\nabla^2_{\text{conf}} u = \lambda(\Delta u) g - \nabla^2 u + a\,du \otimes du + b|\nabla u|^2 g$$

in the sequel. We call u is *admissible* if $g^{-1}W[u] \in \Gamma^+$.

Theorem 1.1. *For $n \geq 3$, let $(\overline{M}^n, g)$ be a smooth, compact Riemannian manifold with boundary ∂M. If*

1. $B \in \Gamma^+$;
2. $h > 0$ *on* $\overline{M} \times \mathbb{R}$, $\partial_z h(x,z) > 0$ *on* $\overline{M} \times \mathbb{R}$, $\lim_{z\to+\infty} h(x,z) \to +\infty$ *and* $\lim_{z\to-\infty} h(x,z) \to 0$ *in* $M \times \mathbb{R}$; *and*
3. $a(x)$ *is positive on* $\overline{M}$ *and* $\lambda a(x) + b(x)$ *is nonnegative in* M,

then there exists a unique admissible function $u \in C^\infty(\overline{M})$ solving the Dirichlet problem

$$\begin{cases} F(W[u]) = h(x,u) & \text{in } M, \\ u = \varphi & \text{on } \partial M, \end{cases} \tag{1-4}$$

where φ is a smooth function defined on a neighborhood of ∂M.

We may apply Theorem 1.1 to the elementary symmetric functions and their quotients $(\sigma_k/\sigma_l)^{1/(k-l)}$ on Γ_k^+, with $0 \leq l < k \leq n$ and $\sigma_0 = 1$:

Corollary 1.2. *For $n \geq 3$, let $(\overline{M}^n, g)$ be a smooth, compact Riemannian manifold with boundary ∂M. Let $f \in C^\infty(\overline{M})$, let $f > 0$, and let S be a Riemannian metric on ∂M that is conformal to $g|_{\partial M}$. If $A_g^\tau \in \Gamma_k^+$ and $\tau > n-1$, then there exists a smooth metric $\hat{g} \in [g]$ on $\overline{M}$ satisfying*

$$\left(\frac{\sigma_k}{\sigma_l}\right)^{1/(k-l)} (A_{\hat{g}}^\tau) = f \quad \text{in } M \quad \text{and} \quad \hat{g}|_{\partial M} = S,$$

where $0 \leq l < k \leq n$.

Recently Gursky, Streets and Warren [2011] proved that any Riemannian manifold with boundary admits a negative Ricci curvature metric; see also Lohkamp [1994] and Guan [2008]. Once $\text{Ric}_g < 0$, we have $A_g^{2(n-1)} = \frac{1}{n-2}(\text{Ric}_g - R_g g) \in \Gamma_k^+$. Therefore:

Corollary 1.3. *For $n \geq 3$, every smooth compact n-dimensional manifold with boundary admits a Riemannian metric g with its Ricci tensor* Ric *and scalar curvature R satisfying*

$$\sigma_k(g^{-1}(\text{Ric} - Rg)) = \text{const} > 0,$$

where $1 \leq k \leq n$. In the case $k = n$, we have

$$\det(\text{Ric} - Rg) = \text{const} > 0.$$

By solving the infinite boundary data Dirichlet problem, we can produce complete metrics with constant σ_k-A_g^τ curvature, where $\tau > n-1$.

Theorem 1.4. *For $n \geq 3$, let $(\overline{M}^n, g)$ be a smooth, compact Riemannian manifold with boundary ∂M. Choose any smooth positive function $f \in C^\infty(\overline{M})$. If $B \in \Gamma^+$, $a(x)$ is positive on $\overline{M}$, and $\lambda a(x) + b(x)$ is nonnegative in M, then there exists an admissible solution $u \in C^\infty(M)$ to the equation*

$$\begin{cases} F(W[u]) = f(x)e^{2u} & \text{in } M, \\ u = +\infty & \text{on } \partial M. \end{cases} \tag{1-5}$$

Moreover, there exist some constants $C > 0$ and $0 < \gamma \leq 1$, depending on

$$n, \quad \lambda, \quad |f|_{C^2(\overline{M})}, \quad |a|_{L^\infty(\overline{M})}, \quad |b|_{L^\infty(\overline{M})}, \quad |B|_{g(\overline{M})}$$

and the geometry of $(\overline{M}, g)$, such that

$$-C - \gamma \log d(x) \leq u(x) \leq -\log d(x) + C \quad \text{near } \partial M,$$

where $d(x)$ denotes the distance to ∂M with respect to the metric g.

We can combine this with the result of [Gursky et al. 2011]:

Corollary 1.5. *For $n \geq 3$, every smooth compact n-dimensional manifold with boundary admits a complete metric g whose Ricci curvature satisfies*

$$\sigma_k(g^{-1}(\mathrm{Ric} - Rg)) = \mathrm{const} > 0,$$

where $1 \leq k \leq n$. In the case $k = n$, we have

$$\det(\mathrm{Ric} - Rg) = \mathrm{const} > 0.$$

When we consider the modified Schouten tensor with $\tau \leq 0$, it seems reasonable to consider the negative cone, by seeking a complete conformal metric $\bar{g}$ in the conformal class $[g]$, such that $\sigma_k(-\bar{g}A_{\bar{g}}^\tau) = \mathrm{const} > 0$. There are some interesting results, and we refer the reader to [Guan 2008] and [Gursky et al. 2011]. In the case $\tau = 1$, A_g^1 is just the classical Schouten tensor. In [2005], Schnürer fixes the metric at the boundary and realizes a prescribed value for the product of the eigenvalues of the Schouten tensor in the interior, provided there exists a subsolution. In [2007], Guan proved the existence of a conformal metric given its value on the boundary as a prescribed metric conformal to the (induced) background metric, with a prescribed curvature function of the Schouten tensor.

For compact manifolds without boundary, the problem of finding conformal metrics in Γ_k^+ of constant σ_k curvature (that is, of finding $g \in [g_0]$ such that $A_g^1 \in \Gamma_k^+$ and $\sigma_k(g^{-1}A_g^1) = \mathrm{const}$) — known as the higher order k-Yamabe problem for $k \geq 2$ — has attracted enormous interest since the work [Viaclovsky 2000] appeared. It can be viewed as a fully nonlinear version of the Yamabe problem, which was solved by Trudinger [1968], Aubin [1976] and Schoen [1984]. The solvability of the higher order k-Yamabe problem was shown for $k = 2$ in [Sheng et al. 2007] (see also [Chang et al. 2002; Ge and Wang 2006]), for $k = n/2$ in [Trudinger and Wang 2010], for $k > n/2$ in [Gursky and Viaclovsky 2007], and for locally conformally flat manifolds in [Guan and Wang 2003a; Li and Li 2003; Sheng et al. 2007]. For results concerning the modified Schouten tensor on closed manifolds, see [Gursky and Viaclovsky 2003; Li and Sheng 2005] for the case $\tau < 1$, and [Sheng and Zhang 2007] for the case $\tau \geq n-1$.

Our primary task is to solve the Dirichlet problem (1-4). The proof goes via the continuity method and a priori estimates. This paper is organized as follows. In Section 2, we show (1-3) is elliptic at any admissible solution. In Section 3, 4 and 5, we establish a priori estimates that are essential in proving the existence result. We then complete the proof of Theorem 1.1 in Section 6 and solve the infinite boundary data Dirichlet problem (1-5) in Section 7.

2. Ellipticity

In order to discuss the ellipticity properties of Equation (1-3), we define

$$\mathcal{A}[u] := F(g^{-1}W[u]) - h(x,u).$$

We then suppose that $u \in C^2(\overline{M})$ satisfies $\mathcal{A}[u] = 0$. Let $u_s = u + s\psi$, then the linearized operator of $\mathcal{A}$ is

$$\begin{aligned}\mathcal{L}\psi &:= \frac{d}{ds}\mathcal{A}[u_s]|_{s=0}\\ &= F(g^{-1}W[u])^{ij}(\lambda(\Delta\psi)g_{ij} - \psi_{ij} + 2au_i\psi_j + 2b\langle\nabla u, \nabla\psi\rangle g_{ij})\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - h_z(x,u)\psi.\end{aligned}$$

Defining

$$Q^{ij} = \lambda\sum_l (F^{ll})\delta^{ij} - F^{ij}, \tag{2-1}$$

we have

$$\mathcal{L}\psi = Q^{ij}\psi_{ij} + 2F^{ij}(au_i\psi_j + b\langle\nabla u, \nabla\psi\rangle g_{ij}) - h_z(x,u)\psi. \tag{2-2}$$

Proposition 2.1. *Equation (1-3) is elliptic at any admissible solution.*

Proof. Since F^{ij} is positive definite in Γ^+, we have

$$Q^{ij} \geq (\lambda - 1)\sum_l (F^{ll})\delta^{ij} > 0.$$

Therefore, (1-3) is elliptic by (2-2). □

If $\partial_z h(x,z)$ is positive on $\overline{M}\times\mathbb{R}$, then the coefficient of ψ in the zeroth-order term of (2-2) is strictly negative, and we have this:

Corollary 2.2. *If $\partial_z h(x,z)$ is positive on $\overline{M}\times\mathbb{R}$, then at any admissible solution of (1-3), the linearized operator $\mathscr{L}: C^{2,\alpha}(M)\to C^{\alpha}(M)$ is invertible.*

3. The global C^0 estimates

Proposition 3.1. *If $B\in\Gamma^+$ and $\lim_{z\to+\infty} h(x,z)\to+\infty$, $\lim_{z\to-\infty} h(x,z)\to 0$. Then there exists some positive constant C_0, depending only upon h, B and φ, such that for any $C^2(\overline{M})$ admissible solution u of (1-4), we have*

$$|u|_{C^0(\overline{M})}\le C_0.$$

Proof. Since $\overline{M}$ is compact, we may suppose $\tilde{x}$ is a minimum of the function u. If $\tilde{x}\in M$, we have

$$\begin{aligned} h(\tilde{x},u(\tilde{x})) &= F(\lambda(\Delta u)(\tilde{x})g-\nabla^2u(\tilde{x})+B(\tilde{x}))\\ &\ge \min_M F(B)>0. \end{aligned}$$

Using $\lim_{z\to-\infty} h(x,z)\to 0$, we get the lower bound of u. Otherwise $\tilde{x}\in\partial M$, we get $u\ge \min_{\partial M}\varphi$.

The upper bound of u follows by considering a maximum of the function u and using the fact that $\lim_{z\to+\infty} h(x,z)\to+\infty$. □

4. Gradient estimates

We first establish the interior gradient estimates.

Lemma 4.1. *Suppose $B\in\Gamma^+$ and $\lambda a(x)+b(x)$ is nonnegative in M. If $u\in C^3(B_r)$ is an admissible solution of (1-4) in a ball $B_r\subset M$, then there is a constant C depending only on $|a|_{C^1(M)}$, $|b|_{C^1(M)}$, $\max_{M\times[-C_0,C_0]}|h|_{C^1}$, $|g|_{C^2(M)}$, λ, $|B|_{C^1(M)}$ and $|u|_{C^0(B_r)}$, such that*

$$\sup_{B_{r/2}}|\nabla u|\le C.$$

Proof. Consider the auxiliary function

$$H(x)=\zeta(x)ve^{\phi(u)},$$

where $\zeta(x)\in C_0^\infty(B_r)$ is a cutoff function to be chosen later, $v=(1+\frac{1}{2}|\nabla u|_g^2)$, $\phi:\mathbb{R}\longrightarrow\mathbb{R}$ is a function of the form $\phi(s)=\alpha(\beta+s)^p$, and $|s|\le|u|_{C^0(B_r)}$. The constants α, β and p depend only on $|u|_{C^0(B_r)}$ and $|a|_{L^\infty}$, such that the function $\phi(s)$ satisfies $\phi'(s)>0$ and $\phi''(s)-\phi'^2(s)-|a|_{L^\infty}\phi'(s)\ge\varepsilon_1>0$ for some constant ε_1 depending on $|u|_{C^0(B_r)}$ and $|a|_{L^\infty}$. It is proved in [Gursky and Viaclovsky 2003] that such a function ϕ always exists in the case $|a|_{L^\infty} = 1$. With a slight modification, the proof still works for our case.

Suppose the maximum of H occurs at an interior point $\tilde{x} \in B_r$. Take a normal coordinate system $(x^1, \dots, x^n)$ at $\tilde{x}$ with respect to g such that $W[u]_{ij}(\tilde{x})$ is diagonal. Then at $\tilde{x}$ we have

$$0 = H_i = (v\zeta_i + \zeta u_{li}u_l + v\zeta\phi' u_i)e^{\phi(u)},$$

that is,

$$\zeta u_{li}u_l = -v(\zeta_i + \zeta\phi' u_i), \tag{4-1}$$

and

$$\begin{aligned} 0 \geq H_{ij} = \zeta(u_l u_{lij} + u_{li}u_{lj} + u_l(u_i u_{lj} + u_{li}u_j)\phi')e^{\phi(u)} \\ +v\zeta((\phi'^2 + \phi'')u_i u_j + \phi' u_{ij})e^{\phi(u)} \\ + u_l(u_{lj}\zeta_i + u_{li}\zeta_j)e^{\phi(u)} + v(\zeta_{ij} + \phi'(u_i\zeta_j + \zeta_i u_j))e^{\phi(u)}. \end{aligned} \tag{4-2}$$

Recall that $Q^{ij} = \lambda(\sum_l F^{ll})\delta^{ij} - F^{ij}$. Since F^{ij} is positive definite in Γ^+, one obtains $\lambda(\sum_l F^{ll})\delta^{ij} \geq Q^{ij} \geq \varepsilon_0(\sum_l F^{ll})\delta^{ij} > 0$, where $\varepsilon_0 = \lambda - 1$. Then (4-2) implies

$$\begin{aligned} 0 \geq \zeta Q^{ij}(u_l u_{lij} + u_{li}u_{lj} + 2u_i u_l u_{lj}\phi') \\ + v\zeta Q^{ij}((\phi'^2 + \phi'')u_i u_j + \phi' u_{ij}) \\ + 2u_l Q^{ij} u_{li}\zeta_j + vQ^{ij}(\zeta_{ij} + 2\phi' u_i\zeta_j). \end{aligned}$$

By the Ricci identity, we have $u_{lij} = u_{ijl} + R_{jlip}u_p$, where R_{ijlp} is the Riemannian curvature tensor of (M, g). Then

$$\begin{aligned} 0 \geq \zeta Q^{ij}\big(u_l u_{ijl} + R_{jlip}u_p u_l + 2u_l u_{li}u_j\phi' + v((\phi'^2 + \phi'')u_i u_j + \phi' u_{ij})\big) \\ + 2u_l Q^{ij} u_{li}\zeta_j + vQ^{ij}(\zeta_{ij} + 2\phi' u_i\zeta_j). \end{aligned} \tag{4-3}$$

Using $h(x, u) = F(W[u]) = F^{ij}W[u]_{ij}$ and $h_l + h_z u_l = F^{ij}W[u]_{ij;l}$, we obtain

$$Q^{ij}u_{ij} = -F^{ij}(au_i u_j + b|\nabla u|^2 g_{ij} + B_{ij}) + h(x, u), \tag{4-4}$$

and

$$\begin{aligned} u_l Q^{ij} u_{ijl} \\ = -F^{ij}(a_l u_l u_i u_j + 2au_i u_{jl}u_l + b_l u_l|\nabla u|^2 g_{ij} + 2bu_k u_{lk}u_l g_{ij} + u_l B_{ijl}) \\ + h_l u_l + h_z|\nabla u|^2. \end{aligned} \tag{4-5}$$

Plugging (4-4) and (4-5) into (4-3), we have

$$\begin{aligned}0\geq\ &-\zeta F^{ij}(a_l u_l u_i u_j+2au_iu_{jl}u_l+b_lu_l|\nabla u|^2g_{ij}+2bu_ku_{lk}u_lg_{ij}+u_lB_{ijl})\\&-\zeta v\phi' F^{ij}(au_iu_j+b|\nabla u|^2g_{ij}+B_{ij})\\&+\zeta Q^{ij}(R_{jlip}u_pu_l+2u_lu_{li}u_j\phi'+v(\phi'^2+\phi'')u_iu_j)\\&+\zeta(h_lu_l+h_z|\nabla u|^2+v\phi' h(x,u))\\&+2u_lQ^{ij}u_{li}\zeta_j+2v\phi' Q^{ij}u_i\zeta_j+vQ^{ij}\zeta_{ij}.\end{aligned}$$

Without loss of generality, we may assume $\frac{1}{2}|\nabla u|^2\leq v\leq|\nabla u|^2$, and using (4-1), we derive

$$\begin{aligned}0\geq\ &\zeta v\phi' F^{ij}(au_iu_j+b|\nabla u|^2g_{ij})+\zeta v(\phi''-\phi'^2)Q^{ij}u_iu_j\\&-\zeta F^{ij}(a_lu_lu_iu_j+b_lu_l|\nabla u|^2g_{ij}+u_lB_{ijl})\\&-\zeta v\phi' F^{ij}B_{ij}+\zeta Q^{ij}R_{jlip}u_pu_l\\&+\zeta(h_lu_l+h_z|\nabla u|^2+v\phi' h(x,u))\\&-2v\phi' Q^{ij}\zeta_iu_j+2v(aF^{ij}+b(\textstyle\sum F^{ll})\delta^{ij})\zeta_iu_j\\&+vQ^{ij}\zeta_{ij}-2(v/\zeta)Q^{ij}\zeta_i\zeta_j\\\geq\ &\zeta v(\phi''-\phi'^2-a\phi')Q^{ij}u_iu_j\\&+\zeta v\phi'(\lambda a(x)+b(x))(\textstyle\sum F^{ll})|\nabla u|^2-C\zeta(\sum F^{ll})(v^{3/2}+1)\\&-C\zeta(v+1)-C(\textstyle\sum F^{ll})(|\nabla\zeta|v^{3/2}+|\nabla^2\zeta|v+(|\nabla\zeta|^2/\zeta)v),\end{aligned}\tag{4-6}$$

in the second inequality, we have used the definition of Q^{ij} to get

$$a\zeta v\phi' F^{ij}u_iu_j=\lambda a\zeta\phi'(\textstyle\sum_l F^{ll})|\nabla u|^2-a\zeta v\phi' Q^{ij}u_iu_j.$$

Now we choose ζ to satisfy, as in [Guan and Wang 2003b],

$$0\leq\zeta\leq 1,\quad |\nabla\zeta|\leq b_0\zeta^{1/2},\quad |\nabla^2\zeta|\leq b_0$$

for some constant $b_0>0$ and

$$\zeta(x)=1\ \text{ in } B_{r/2}\quad\text{and}\quad \zeta(x)=0\ \text{ outside } B_r.$$

By virtue of (4-6), we then have

$$0\geq(\textstyle\sum_l F^{ll})(\varepsilon_0\varepsilon_1\zeta v^2-C\zeta v^{3/2}-C\zeta)-C\zeta(v+1)-C(\sum_l F^{ll})(\zeta^{1/2}v^{3/2}+v).$$

Multiplying by ζ on both sides and using that $0\leq\zeta\leq 1$, we have

$$0\geq(\textstyle\sum_l F^{ll})(\varepsilon_0\varepsilon_1\zeta^2v^2-C\zeta^{3/2}v^{3/2}-C\zeta v-C)-C(\zeta v+1).\tag{4-7}$$

Note that Euler formula and concavity of F imply

$$(\sum_l F^{ll})(\kappa) = F(\kappa) + \sum_i F^{ii}(\kappa)(1-\kappa_i) \geq F(e) > 0 \quad \text{in } \Gamma^+,$$

where $e = (1, \ldots, 1)$. From (4-7), if $\varepsilon_0\varepsilon_1\zeta^2 v^2 - C\zeta^{3/2}v^{3/2} - C\zeta v - C \leq 0$, we have $(\zeta v)(\tilde{x}) \leq C$. Otherwise, we have

$$0 \geq F(e)(\varepsilon_0\varepsilon_1\zeta^2 v^2 - C\zeta^{3/2}v^{3/2} - C\zeta v - C) - C(\zeta v + 1).$$

We then obtain $(\zeta v)(\tilde{x}) \leq C$. Hence $H \leq C$ in B_r; therefore $\sup_{B_{r/2}} |\nabla u| \leq C$. $\square$

We now derive a priori bounds for the boundary gradient of solutions to (1-4) with smooth Dirichlet data φ. Without loss of generality, we may assume that $\varphi \in C^\infty(\overline{M})$ in the sequel. The method is to construct barrier functions near ∂M using the boundary distance function. Let $d(x) = \mathrm{dist}_g(x, \partial M)$ for $x \in M$, and set

$$M_\delta = \{x \in M \mid d(x) < \delta\} \quad \text{for } \delta > 0.$$

Since ∂M is smooth and $|\nabla d| = 1$ on ∂M, we choose $\delta > 0$ sufficiently small so that d is smooth and $\frac{1}{2} \leq |\nabla d| \leq 2$ in M_δ.

Consider the locally defined auxiliary function

$$w^- := \varphi + \theta \log \frac{\delta^2}{d+\delta^2},$$

where θ is some small positive constant. We may directly check that

$$\begin{cases} w^-|_{\partial M} = \varphi, \\ \varphi + \theta \log(\delta/2) \leq w^-|_{\{d(x)=\delta\}} \leq \varphi + \theta \log \delta. \end{cases} \tag{4-8}$$

Since

$$\nabla w^- = \nabla\varphi - \frac{\theta}{d+\delta^2}\nabla d,$$
$$\nabla^2 w^- = \nabla^2\varphi - \frac{\theta}{d+\delta^2}\nabla^2 d + \frac{\theta}{(d+\delta^2)^2}\nabla d \otimes \nabla d,$$

we obtain

$$\begin{aligned} W[w^-]_{ij} &= \frac{(\lambda + b\theta)\theta}{(d+\delta^2)^2}|\nabla d|^2 g_{ij} + \frac{a\theta^2}{(d+\delta^2)^2} d_i d_j - \frac{\theta}{(d+\delta^2)^2} d_i d_j \\ &\qquad - \frac{\theta}{d+\delta^2}(\lambda \Delta d g_{ij} - d_{ij} + a(\varphi_j d_i + \varphi_i d_j) + 2b\,\langle \nabla\varphi, \nabla d\rangle\, g_{ij}) \\ &\qquad\qquad + \lambda\Delta\varphi g_{ij} - \varphi_{ij} + a\varphi_i\varphi_j + b|\nabla\varphi|^2 g_{ij} + B_{ij} \\ &\geq \frac{(\varepsilon_0 - (|a|_{L^\infty(\overline{M})} + |b|_{L^\infty(\overline{M})})\theta)\theta}{(d+\delta^2)^2}|\nabla d|^2 g_{ij} - \frac{\theta}{d+\delta^2} C' g_{ij} - C'' g_{ij}, \end{aligned}$$

where C' and C'' are some sufficiently large constants, depending only on $|\varphi|_{C^2(\bar{M})}$, λ, $|a|_{L^\infty(\overline{M})}$, $|b|_{L^\infty(\bar{M})}$, $|B|_{g(\bar{M})}$ and the geometric quantities of $(\overline{M}, g)$, independent of δ. Choosing

$$\theta \le \frac{\varepsilon_0}{2(|a|_{L^\infty(\overline{M})}+|b|_{L^\infty(\overline{M})})} \quad \text{and} \quad \delta \le \min\left\{1, \frac{\varepsilon_0}{16C'}, \frac{\varepsilon_0\theta}{64C''}\right\},$$

by virtue of $|\nabla d| > 1/2$ in M_δ, we derive

$$\begin{aligned} W[w^-]_{ij} &\ge \frac{\varepsilon_0\theta}{8(d+\delta^2)\delta} g_{ij} - \frac{\theta}{d+\delta^2} C' g_{ij} - C'' g_{ij} \\ &= \frac{\theta}{d+\delta^2}\Big(\frac{\varepsilon_0}{16\delta} - C'\Big) g_{ij} - C'' g_{ij} + \frac{\theta\varepsilon_0}{16\delta(d+\delta^2)} g_{ij} \\ &\ge \frac{\theta\varepsilon_0}{32\delta} g_{ij} - C'' g_{ij} \\ &= \frac{\theta\varepsilon_0}{64\delta} g_{ij} + \Big(\frac{\theta\varepsilon_0}{64\delta} - C''\Big) g_{ij} \ge \frac{\theta\varepsilon_0}{64\delta} g_{ij}, \end{aligned} \tag{4-9}$$

in the first inequality we have used the fact $d+\delta^2 \le 2\delta$, while in the second, we have used that $d+\delta^2 \le 2$.

To estimate the boundary gradient, we need the following maximum principle. We first give a standard definition.

Definition 4.2. We say a subsolution w of (1-3) is *admissible* and

$$F(W[w]) \ge h(x, w) \quad \text{in } M.$$

Changing the direction of the inequality, one gets the definition of the supsolution of (1-3).

Lemma 4.3. *Suppose that w_1 and w_2 are smooth sub- and supersolutions* (*respectively*) *of (1-3) with $w_1|_{\partial M} < w_2|_{\partial M}$. If $\partial_z h(x, z)$ is positive in $M\times\mathbb{R}$, then $w_1 \le w_2$ on $\overline{M}$.*

Proof. We argue by contradiction. Set $\tilde{w} = w_2 - w_1$. Suppose $\tilde{w}(\tilde{x}) = \min_{\overline{M}} \tilde{w} < 0$ for some $\tilde{x} \in \overline{M}$; then $\tilde{x}$ must be an interior point. At this point,

$$\nabla w_2(\tilde{x}) = \nabla w_1(\tilde{x}) \quad \text{and} \quad \nabla^2 w_2(\tilde{x}) \ge \nabla^2 w_1(\tilde{x}).$$

Consequently

$$\begin{aligned} F(W[w_2])(\tilde{x}) &= Q^{ij}\nabla^2_{ij} w_2(\tilde{x}) + F^{ij}(a\nabla_i w_2 \nabla_j w_2 + b|\nabla w_2|^2 g_{ij} + B_{ij})(\tilde{x}) \\ &\ge Q^{ij}\nabla^2_{ij} w_1(\tilde{x}) + F^{ij}(a\nabla_i w_1 \nabla_j w_1 + b|\nabla w_1|^2 g_{ij} + B_{ij})(\tilde{x}) \\ &= F(W[w_1])(\tilde{x}). \end{aligned}$$

We therefore have

$$h(\tilde{x}, w_2(\tilde{x})) \ge F(W[w_2])(\tilde{x}) \ge F(W[w_1])(\tilde{x}) \ge h(\tilde{x}, w_1(\tilde{x})),$$

which contradicts that $w_1(\tilde{x}) > w_2(\tilde{x})$ and $\partial_z h(x, z)$ is positive in $M\times\mathbb{R}$. $\square$

Let x_0 be an arbitrary point on ∂M. We pick local coordinates in M_δ so that ∂M is the plane $x_n = 0$, and let $\{e_\gamma, e_n\}_{\gamma=1}^{n-1}$ be the corresponding coordinate vector fields, where $e_n(x_0)$ denotes the interior normal vector and $e_\gamma(x_0)$ the tangential direction.

Lemma 4.4. *Let u be a $C^2(\overline{M})$ admissible solution of (1-4). If $B \in \Gamma^+$ and $\partial_z h(x, z)$ is positive in $M \times \mathbb{R}$, then there exists a constant C depending on*

$$C_0, \quad \lambda, \quad |\varphi|_{C^2(\overline{M})}, \quad |a|_{L^\infty(\overline{M})}, \quad |b|_{L^\infty(\overline{M})}, \quad |B|_{g(\overline{M})}$$

and the geometric quantities of $(\overline{M}, g)$, such that

$$\partial_n u|_{\partial M} > -C.$$

Proof. Recalling (4-8) and (4-9), we have

$$w^-|_{\partial M} = \varphi \quad \text{and} \quad F(W[w^-]) = F^{ij} W[w^-]_{ij} \geq \frac{\varepsilon_0 \theta}{64\delta} F(e) \quad \text{on } M_\delta.$$

We choose δ smaller, so that

$$F(W[w^-]) \geq \max_{\overline{M} \times [\min_{\overline{M}} \varphi, \max_{\overline{M}} \varphi]} h(x, z) \geq h(x, w^-) \quad \text{on } M_\delta.$$

Since $|u|_{C^0}(\overline{M}) < C_0$, we can regard w^- as a local subsolution of (1-3) on $\overline{M}_\delta = \{x \mid d(x) \leq \delta\}$. Applying Lemma 4.3 to $\overline{M}_\delta$, we have

$$\frac{u(x) - u(x_0)}{d(x, x_0)} \geq \frac{w^-(x) - w^-(x_0)}{d(x, x_0)} \quad \text{for any } x_0 \in \partial M.$$

That is, $\partial_n u|_{\partial M} \geq \partial_n w^-|_{\partial M}$, and our lemma follows. □

We next prove that the $\partial_n u$ have an upper bound; the boundary gradient estimates follow.

Lemma 4.5. *Let u be a $C^2(\overline{M})$ admissible solution of (1-4). If $B \in \Gamma^+$ and $\partial_z h(x, z)$ is positive in $M \times \mathbb{R}$, then we have*

$$\partial_n u(x_0) < C \quad \textit{for any point } x_0 \in \partial M,$$

where C is a positive constant depending on C_0, λ, $|\varphi|_{C^2(\overline{M})}$, $|a|_{L^\infty(\overline{M})}$, $|b|_{L^\infty(\overline{M})}$, $|B|_{g(\bar{M})}$ and the geometric quantities of $(\overline{M}, g)$.

Proof. Since u is admissible and $\Gamma^+ \subset \Gamma_1^+$, we have

$$c_1 \Delta u + c_2 |\nabla u|^2 + \operatorname{tr} B \geq (n\lambda - 1)\Delta u + (a + nb)|\nabla u|^2 + \operatorname{tr} B > 0,$$

where $c_1 = n\lambda - 1$ and $c_2 = |a|_{L^\infty} + n|b|_{L^\infty}$. Therefore the proof reduces to constructing a local supbarrier function of the equation

$$c_1 \Delta v + c_2 |\nabla v|^2 + \operatorname{tr} B = 0.$$

Let's consider $w^+ = \varphi + \theta \log((d + \delta^2/\delta^2))$ in M_δ; then

$$w_i^+ = \theta \frac{d_i}{d+\delta^2} + \varphi_i,$$
$$w_{ij}^+ = -\theta \frac{d_i d_j}{(d+\delta^2)^2} + \theta \frac{d_{ij}}{d+\delta^2} + \varphi_{ij}.$$

We therefore have

$$\begin{aligned} c_1 \Delta w^+ &+ c_2 |\nabla w^+|^2 + \operatorname{tr} B \\ &= -\theta(c_1 - c_2\theta) \frac{|\nabla d|^2}{(d+\delta^2)^2} + (c_1 \Delta d + 2c_2 \langle \nabla d, \nabla \varphi \rangle) \frac{\theta}{d+\delta^2} \\ &\qquad + c_1(\Delta \varphi) + c_2 |\nabla \varphi|^2 + \operatorname{tr} B. \end{aligned}$$

Now we choose $\theta < c_1/(2c_2)$. Then using $|\nabla d|^2 > \frac{1}{2}$ in M_δ, we derive

$$\begin{aligned} c_1 \Delta w^+ + c_2 |\nabla w^+|^2 + \operatorname{tr} B &\le -\frac{c_1 \theta}{4(d+\delta^2)^2} + C' \frac{\theta}{d+\delta^2} + C'' \\ &\le \Big(-\frac{c_1}{4\delta(1+\delta)} + C' \Big) \frac{\theta}{d+\delta^2} + C'' \quad \text{in } M_\delta, \end{aligned}$$

where C' and C'' are two positive constants depending on

$$|\varphi|_{C^2(\overline{M})}, \quad \lambda, \quad |a|_{L^\infty(\overline{M})}, \quad |b|_{L^\infty(\overline{M})}, \quad |B|_{g(\overline{M})}$$

and the geometric quantities of $(\overline{M}, g)$, independent of δ. Next we choose

$$\delta < \min\Big\{1, \frac{c_1}{8(C'+1)}, \frac{\theta}{2C''}\Big\};$$

then $c_1 \Delta w^+ + c_2 |\nabla w^+|^2 + \operatorname{tr} B < 0$ in M_δ.

Note that

$$\begin{cases} w^+|_{\partial M} = \varphi, \\ w^+|_{\{x \in M | d(x) = \delta\}} \ge \varphi + \theta \log(1/\delta). \end{cases}$$

Without loss of generality, we can assume δ is small; then $|u|_{C^0}(\overline{M}) < C_0$ and the maximum principle imply $u \le w^+$ in $\overline{M}_\delta$. Consequently, for any $x_0 \in \partial M$,

$$\frac{u(x) - u(x_0)}{d(x, x_0)} \le \frac{w^+(x) - w^+(x_0)}{d(x, x_0)}.$$

That is, $\partial_n u|_{\partial M} \le \partial_n w^+|_{\partial M}$, and our lemma follows. □

Combining Lemma 4.1, Lemma 4.4 and Lemma 4.5, we obtain this:

Proposition 4.6. *Suppose $B \in \Gamma^+$, $\lambda a(x) + b(x)$ is nonnegative in M and $\partial_z h(x, z)$ is positive in $M \times \mathbb{R}$. Then for any $C^3(\overline{M})$ admissible solution u of (1-4), there is* *a constant C_1 depending only on*

$$C_0,\quad \lambda,\quad |\varphi|_{C^2(\overline{M})},\quad |a|_{C^1(\overline{M})},\quad |b|_{C^1(\overline{M})},\quad \max_{M\times[-C_0,C_0]}|h|_{C^1},\quad |B|_{C^1(\overline{M})}$$

and the geometric quantities of $(\overline{M}, g)$, such that $|\nabla u| \le C_1$ on $\overline{M}$.

5. Estimates for the second derivative

As in Section 4, we begin by establishing the interior estimates.

Lemma 5.1. *Let $B \in \Gamma^+$ and $a(x)$ be positive on $\overline{M}$. Let $u \in C^4(B_r)$ be an admissible solution of (1-4) in a ball $B_r \subset M$; there is a constant C depending only on*

$$|a|_{C^2(M)},\quad |b|_{C^2(M)},\quad \max_{M\times[-C_0,C_0]}|h|_{C^2},\quad |g|_{C^2(M)},\quad |B|_{C^2(M)},\quad \lambda,\quad |u|_{C^1(B_r)}$$

such that $\sup_{B_{r/2}}|\nabla^2 u| \le C$.

Proof. Since $\Gamma^+ \subset \Gamma_1^+$, we obtain

$$0 < \operatorname{tr} W[u] = (n\lambda - 1)(\Delta u) + (a(x) + nb(x))|\nabla u|^2 + \operatorname{tr} B.$$

Consequently $\Delta u \ge -C$. For obtaining the upper bound of Δu, we consider the auxiliary function

$$G(x) = \zeta(x)(\Delta u + \Lambda a(x)|\nabla u|^2)$$

for some large constant $\Lambda > 1$, depending only on $|a|_{L^\infty}$, $|b|_{L^\infty}$ and λ, to be chosen later; here $\zeta(x) \in C_0^\infty(B_r)$ is a cutoff function as in Lemma 4.1.

Suppose G achieves a maximum at an interior point $\tilde{x} \in M$. We take a normal coordinate system $(x^1, \dots, x^n)$ with respect to g such that $W[u]_{ij}(\tilde{x})$ is diagonal. Without loss of generality, we may assume $G(\tilde{x}) \ge 1$ and $\tilde{x} \in B_r$. Then, at $\tilde{x}$, we have

$$0 = G_i = (\Delta u + \Lambda a|\nabla u|^2)\zeta_i + \zeta(u_{lli} + \Lambda a_i|\nabla u|^2 + 2\Lambda a u_l u_{li}),$$

that is,

$$\zeta u_{lli} = -\Lambda a_i \zeta|\nabla u|^2 - 2\Lambda a\zeta u_l u_{li} - (\Delta u + \Lambda a|\nabla u|^2)\zeta_i, \tag{5-1}$$

and

$$\begin{aligned} 0 \ge G_{ij} = \zeta(u_{llij} + \Lambda a_{ij}|\nabla u|^2 + 2\Lambda u_l(a_i u_{lj} + a_j u_{li}) + 2\Lambda a(u_{li}u_{lj} + u_l u_{lij})) \\ + (u_{lli} + \Lambda a_i|\nabla u|^2 + 2\Lambda a u_l u_{li})\zeta_j \\ + (u_{llj} + \Lambda a_j|\nabla u|^2 + 2\Lambda a u_l u_{lj})\zeta_i + (\Delta u + \Lambda a|\nabla u|^2)\zeta_{ij}. \end{aligned} \tag{5-2}$$

Recall that $Q^{ij} = \lambda(\sum_l F^{ll})\delta^{ij} - F^{ij}$. Since F^{ij} is positive definite in Γ^+, one obtains $\lambda(\sum_l F^{ll})\delta^{ij} \ge Q^{ij} \ge \varepsilon_0(\sum_l F^{ll})\delta^{ij} > 0$, where $\varepsilon_0 = \lambda - 1$. Notice that the Ricci identity gives $u_{lij} = u_{ijl} + O(|\nabla u|)$ and $u_{llij} = u_{ijll} + O(|\nabla^2 u| + |\nabla u|)$. Then (5-2) implies

$$
\begin{aligned}
0 &\ge Q^{ij} G_{ij} \\
&= \zeta Q^{ij}\big(u_{llij} + \Lambda a_{ij}|\nabla u|^2 + 4\Lambda u_l a_i u_{lj} + 2\Lambda a(u_{li}u_{lj} + u_l u_{lij})\big) \\
&\quad + 2Q^{ij}(u_{lli} + \Lambda a_i|\nabla u|^2 + 2\Lambda a u_l u_{li})\zeta_j + (\Delta u + \Lambda a|\nabla u|^2)Q^{ij}\zeta_{ij} \\
&\ge \zeta Q^{ij}(u_{ijll} + 2\Lambda a(u_{li}u_{lj} + u_l u_{ijl})) + 2Q^{ij}u_{lli}\zeta_j \\
&\qquad\qquad - C\Lambda\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1).
\end{aligned}
\tag{5-3}
$$

Using $h_{ll} + 2h_{lz}u_l + h_z u_{ll} = F^{ij}W[u]_{ij;ll} + F^{ij,rs}W[u]_{ij;l}W[u]_{rs;l}$ and the concavity of F, we obtain

$$
\begin{aligned}
Q^{ij}u_{ijll} \ge{}& -2aF^{ij}(u_{il}u_{jl} + u_i u_{jll}) - 2b\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u|^2 + u_{kll}u_k) \\
&- C\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1) + h_{ll} + 2h_{lz}u_l + h_z u_{ll}.
\end{aligned}
\tag{5-4}
$$

On the other hand, (4-5) implies

$$
2\Lambda a u_l Q^{ij}u_{ijl} \ge -C\Lambda\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1) + 2\Lambda a h_l u_l + 2\Lambda a h_z|\nabla u|^2.
\tag{5-5}
$$

Plugging (5-4) and (5-5) into (5-3), and employing (5-1) we have

$$
\begin{aligned}
0 &\ge 2\Lambda a\zeta Q^{ij}u_{li}u_{lj} - 2a\zeta F^{ij}(u_{il}u_{jl} + u_i u_{jll}) + 2Q^{ij}u_{lli}\zeta_j \\
&\qquad - 2b\zeta\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u|^2 + u_{kll}u_k) \\
&\qquad\qquad - C\Lambda\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1) - C\Lambda(|\nabla^2 u| + 1) \\
&\ge 2\zeta(\Lambda a\lambda - b)\big(\textstyle\sum_l F^{ll}\big)|\nabla^2 u|^2 - 2a\zeta(\Lambda + 1)F^{ij}u_{il}u_{jl} \\
&\qquad\qquad - C\Lambda\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1) - C\Lambda(|\nabla^2 u| + 1) \\
&\ge 2\zeta(\varepsilon_0 a\Lambda - a - b)\big(\textstyle\sum_l F^{ll}\big)|\nabla^2 u|^2 \\
&\qquad\qquad - C\Lambda\big(\textstyle\sum_l F^{ll}\big)(|\nabla^2 u| + 1) - C\Lambda(|\nabla^2 u| + 1).
\end{aligned}
$$

Since a is positive on $\overline{M}$, we assume $a(x) \ge \varepsilon_2 > 0$. We now choose $\Lambda > \max\{1, 2(|a|_{L^\infty} + |b|_{L^\infty})/(\varepsilon_0\varepsilon_2)\}$, and multiply ζ on both sides to produce

$$
0 \ge \Lambda\big(\textstyle\sum_l F^{ll}\big)(\varepsilon_0\varepsilon_2\zeta^2|\nabla^2 u|^2 - C\zeta|\nabla^2 u| - C) - C\Lambda(\zeta|\nabla^2 u| + 1).
\tag{5-6}
$$

It follows that $(\zeta|\nabla^2 u|)(\tilde{x}) \le C$. Therefore $\sup_{B_{r/2}} \Delta u \le C$.

If $\Gamma^+ \subset \Gamma_2^+$, then $\sup_{B_{r/2}} \Delta u \le C$ implies that $\sup_{B_{r/2}}|\nabla^2 u| \le C$. To get the Hessian bounds of u in general, we simply consider the maximum of

$$
\zeta(x) \max_{\xi \in (T_x M \cap \mathbb{S}^n)} (\nabla_\xi \nabla_\xi u + \Lambda a(x)(\nabla_\xi u)^2).
$$

The calculation is similar. □

We next derive a priori bounds for second derivatives of solutions to (1-4). The method we use is similar to that of [Guan 2007; Guan 2008; Gursky et al. 2011]. The notation below is the same as in Section 4.

We use a barrier function

$$v(x) = p(qd^2 - d) \quad \text{in } M_\delta,$$

where p and q are positive constants. Let's define a linear operator

$$\mathcal{P}(\psi) = Q^{ij}\psi_{ij} + 2F^{ij}(a(x)u_i\psi_j + b(x)\langle \nabla u, \nabla\psi\rangle g_{ij}). \tag{5-7}$$

Then

$$\mathcal{P}d = Q^{ij}d_{ij} + 2F^{ij}(au_i d_j + b\langle \nabla u, \nabla d\rangle g_{ij}),$$

and consequently

$$|\mathcal{P}d| \le C_\# \textstyle\sum_l F^{ll} \quad \text{in } M_\delta,$$

where $C_\#$ depends on λ, $|u|_{C^1(\overline{M})}$, $|a|_{L^\infty(\bar{M})}$, $|b|_{L^\infty(\overline{M})}$ and the geometric quantities of $(\overline{M}, g)$. On the other hand, we have in M_δ

$$\begin{aligned} \mathcal{P}d^2 &= 2Q^{ij}(d_i d_j) + 2d\mathcal{P}d \\ &\ge 2\varepsilon_0(\textstyle\sum_l F^{ll})|\nabla d|^2 - 2dC_\# \textstyle\sum_l F^{ll} \\ &\ge (\varepsilon_0 - 2C_\#\delta)\textstyle\sum_l F^{ll}, \end{aligned}$$

where $\varepsilon_0 = \lambda - 1$ as before. After we choose

$$q > 2(1 + C_\#)/\varepsilon_0 \quad \text{and} \quad \delta < \min\{\varepsilon_0/(4C_\#), 1/(2q)\},$$

the function v satisfies

$$\mathcal{P}v \ge p\{q(\varepsilon_0 - 2C_\#\delta) - C_\#\}\textstyle\sum_l F^{ll} \ge p\textstyle\sum_l F^{ll}, \tag{5-8}$$

and

$$v \le -\tfrac{1}{2}pd \quad \text{in } M_\delta. \tag{5-9}$$

Let x_0 be an arbitrary point on ∂M. Let $r(x) = \operatorname{dist}_g(x, x_0)$ to denote the distance from x to x_0 with respect to the background metric. Let $\Omega_\delta(x_0) = B_\delta(x_0) \cap M_\delta$, where $B_\delta(x_0) = \{x \in \overline{M} \mid r(x) < \delta\}$. Since δ is small, we assume r^2 is smooth in $\Omega_\delta(x_0)$. A similar calculation implies

$$\tfrac{1}{2}\varepsilon_0\textstyle\sum_l F^{ll} \le \mathcal{P}r^2 \le (2\lambda + \tfrac{1}{2}\varepsilon_0)\textstyle\sum_l F^{ll} \quad \text{in } \Omega_\delta(x_0). \tag{5-10}$$

Now we pick a local coordinates in M_δ so that ∂M is the plane $x_n = 0$, and we let $\{e_\gamma, e_n\}_{\gamma=1}^{n-1}$ be the corresponding coordinate vector fields, where $e_n(x_0)$ denotes the interior normal vector and $e_\gamma(x_0)$ the tangential direction. Fix some γ and consider the locally defined function $\phi = e_\gamma(u - \varphi)$, where u is a $C^3(\overline{M})$ admissible solution of (1-4). In order to derive the boundary estimates for second derivatives, we need the following lemma.

Lemma 5.2. *In the notation above, there exists a constant C, depending only on C_0, C_1, $|a|_{C^1(\bar{M})}$, $|b|_{C^1(\overline{M})}$, $|h|_{C^1(\overline{M}\times[-C_0,C_0])}$ and $|\varphi|_{C^3(M_\delta)}$, such that*

$$|\mathcal{P}\phi| \le C(1+\textstyle\sum_l F^{ll}).$$

Proof. Differentiating Equation (1-3) with respect to e_γ yields

$$Q^{ij}u_{ij\gamma}+2F^{ij}(au_{i\gamma}u_j+bu_lu_{l\gamma}g_{ij}) \\ = -F^{ij}(a_\gamma u_iu_j+b_\gamma|\nabla u|^2g_{ij}+B_{ij\gamma})+h_zu_\gamma+h_\gamma.$$

Exchanging derivatives implies

$$u_{ij\gamma}=u_{\gamma ij}+(Rm*\nabla u)_{ij\gamma}.$$

Combining these calculations yields

$$\begin{aligned}\mathcal{P}\phi &= Q^{ij}u_{\gamma ij}+2F^{ij}(au_iu_{\gamma j}+bu_ku_{\gamma k}g_{ij}) \\ &\qquad\qquad - Q^{ij}\varphi_{\gamma ij}-2F^{ij}(au_i\varphi_{\gamma j}+bu_k\varphi_{\gamma k}g_{ij}) \\ &= -F^{ij}(a_\gamma u_iu_j+b_\gamma|\nabla u|^2g_{ij}+B_{ij\gamma})+h_zu_\gamma+h_\gamma \\ &\qquad - Q^{ij}\varphi_{\gamma ij}-2F^{ij}(au_i\varphi_{\gamma j}+bu_k\varphi_{\gamma k}g_{ij})-Q^{ij}(Rm*\nabla u)_{ij\gamma}\end{aligned}$$

Therefore

$$|\mathcal{P}\phi|\le C(\textstyle\sum_l F^{ll})+C. \qquad \square$$

We are now ready to prove the boundary estimates for second derivatives.

Lemma 5.3. *Let $u\in C^3(\overline{M})$ be an admissible solution of (1-4). Then*

$$|\nabla^2u|\le C \quad \text{on } \partial M,$$

where the constant $C>0$ depends on

$$C_0,\quad C_1,\quad |a|_{C^1(\overline{M})},\quad |b|_{C^1(\overline{M})},\quad |h|_{C^1(\overline{M}\times[-C_0,C_0])},\quad |\varphi|_{C^3(M_\delta)},\quad |B|_{C^1(\overline{M})}$$

and the geometric quantities of $(\overline{M},g)$.

Proof. We require separate proofs for the different types $\nabla_\gamma\nabla_\eta u$, $\nabla_\gamma\nabla_n u$ and $\nabla_n\nabla_n u$ of boundary second derivatives.

Let x_0 be an arbitrary point on ∂M. Using that $u-\varphi=0$ on ∂M, we obtain

$$\nabla_\gamma\nabla_\eta(u-\varphi)(x_0)=-\nabla_n(u-\varphi)\Pi(e_\gamma,e_\eta)(x_0),$$

where $1\le\gamma,\eta\le n-1$ and Π denotes the second fundamental form of ∂M. We therefore have the estimates for the pure tangential second order derivatives.

Combining (5-8), (5-10) and Lemma 5.2, we have for any positive constant μ

$$\mathcal{P}(\phi - v + \mu r^2) \le (C - p + \mu(2\lambda + \tfrac{1}{2}\varepsilon_0)) \textstyle\sum_l F^{ll} + C.$$

Picking μ large enough and $p > \mu^2$, we get

$$\mathcal{P}(\phi - v + \mu r^2) \le -\tfrac{1}{2} p F(e) + C < 0.$$

Thus by the maximum principle, we conclude that the minimum of $\phi - v + \mu r^2$ occurs on the boundary of $\Omega_\delta(x_0)$. It remains to check these boundary values. There are two components of $\partial\Omega_\delta(x_0)$ to check. Firstly, since $\phi \equiv 0$ and $v \equiv 0$ on $\partial\Omega_\delta(x_0) \cap \partial M$, we get $\phi - v + \mu r^2 \ge 0$ on $\partial\Omega_\delta(x_0) \cap \partial M$ and $(\phi - v + \mu r^2)(x_0) = 0$. Since μ is large, (5-9) implies $\phi - v + \mu r^2 > \phi + (p/2)d + \mu r^2 > 0$ on $\partial\Omega_\delta(x_0) \setminus \partial M$. It follows that the normal derivative of $\phi - v + \mu r^2$ is nonnegative, and therefore we conclude

$$\begin{aligned}\nabla_n \nabla_\gamma u(x_0) &> \nabla_n(\nabla_\gamma \varphi + v - \mu r^2)(x_0)\\ &= \nabla_n \nabla_\gamma \varphi(x_0) - p > -C.\end{aligned}$$

However, using Lemma 5.2 again, it is clear that the same argument applies to $-\phi$, and one deduces the mixed second derivative estimates

$$|\nabla_n \nabla_\gamma u| < C.$$

Once we bound $\nabla_\gamma \nabla_\eta u$ and $\nabla_\gamma \nabla_n u$, to estimate the double normal second derivative $\nabla_n \nabla_n u$ we only need to bound Δu. Note that $W[u]_{ij} \in \Gamma_1^+$, that is,

$$(n\lambda - 1)(\Delta u) + (a(x) + nb(x))|\nabla u|^2 + \operatorname{tr} B > 0.$$

Consequently Δu is bounded from below and we have to establish an upper bound

$$u_{nn} \le C \quad \text{on } \partial M.$$

Without loss of generality, one can assume $u_{nn} \ge 0$ on ∂M (otherwise we are done). Orthogonally decompose the matrix W at $x_0 \in \partial M$ in terms of e_γ and e_n. Using the known bounds, we find

$$\begin{aligned}W[u]_{ij}(x_0) &= (\lambda \Delta u g_{ij} - u_{ij} + a u_i u_j + b|\nabla u|^2 g_{ij} + B_{ij})(x_0)\\ &\ge \begin{pmatrix} \lambda u_{nn} I_{n-1} & 0 \\ 0 & (\lambda - 1) u_{nn} \end{pmatrix}(x_0) - C\delta_{ij}\\ &\ge (\varepsilon_0 u_{nn}(x_0) - C)\delta_{ij},\end{aligned}$$

where C depends on $|u|_{C^1(\bar{M})}$, $|a|_{C^0(\bar{M})}$, $|b|_{C^0(\bar{M})}$, $|B|_{C^0(\bar{M})}$, $|\nabla_\gamma \nabla_\eta u|$ and $|\nabla_\gamma \nabla_n u|$. It is clear that

$$\begin{aligned} C &> \max_{M\times[-|u|_{C^0(\bar{M})},|u|_{C^0(\bar{M})}]} |h| \\ &\geq F^{ij}(x_0)W[u]_{ij}(x_0) \\ &\geq (\varepsilon_0 u_{nn}(x_0) - C)\textstyle\sum_l F^{ll}(x_0) \\ &\geq (\varepsilon_0 u_{nn}(x_0) - C)F(e). \end{aligned}$$

Thus we obtain the upper bound as desired. □

Combining Lemma 5.1 and Lemma 5.3, we have the global estimates for the second derivative.

Proposition 5.4. *Suppose $B \in \Gamma^+$ and $a(x)$ is positive on $\overline{M}$. Then for any $C^4(\overline{M})$ admissible solution u of (1-4), there is a constant C_2 depending only on C_0, C_1, λ, $|a|_{C^2(\bar{M})}$, $|b|_{C^2(\overline{M})}$, $|h|_{C^2(M\times[-C_0,C_0])}$, $|\varphi|_{C^3(\overline{M})}$, $|B|_{C^2(\overline{M})}$ and the geometric quantities of $(\overline{M}, g)$ such that*

$$|\nabla^2 u| \leq C_2 \quad \text{on } \overline{M}.$$

6. Proof of Theorem 1.1

The proof of Theorem 1.1 is standard. We only sketch it here. For $t \in [0, 1]$, we consider the equations

$$(\star_t) \qquad \begin{cases} F(\nabla^2_{\text{conf}} u + B^t) = h^t, \\ \qquad\qquad u|_{\partial M} = \varphi^t, \end{cases}$$

where

$$B^t = tB + \frac{1-t}{F(e)} g, \quad h^t = (1-t)e^{2u} + th(x, u), \quad \varphi^t = t\varphi.$$

For $t = 0$, the admissible solution is $u \equiv 0$ on $\overline{M}$; for $t = 1$, it is our desired Equation (1-4). It is direct to check that

- $B^t \in \Gamma^+$.
- $h^t > 0$ on $\overline{M} \times \mathbb{R}$, $\partial_z h^t(x, z) > 0$ on $\overline{M} \times \mathbb{R}$, $\lim_{z\to+\infty} h^t(x, z) \to +\infty$ and $\lim_{z\to-\infty} h^t(x, z) \to 0$ in $M \times \mathbb{R}$.
- There exists a uniform constant $C > 0$, independent of $t \in [0, 1]$, such that $|B^t|_{C^2(\bar{M})} < C$, $|h^t|_{C^2(\overline{M}\times[-C,C])} < C$ and $|\varphi^t|_{C^3(\overline{M})} < C$.

Applying our a priori estimates Proposition 3.1, 4.6 and 5.4 to $(\star_t)$ and noting that F is concave, we obtain, by Evans–Krylov estimates,

$$|u_t|_{C^{2,\alpha}(\overline{M})} \leq C \quad \text{for all } t \in [0, 1].$$

Combining this with Corollary 2.2, we see by standard degree theory that ($\star_t$) is solvable for $t = 1$. Uniqueness follows by Lemma 4.3.

7. Proof of Theorem 1.4

To solve the Dirichlet problem for large boundary conditions, we need to control the behavior of the solution near the boundary. We can do this by constructing barrier functions for some suitable equation.

Recall that F is concave, then

$$F(\kappa) \le \omega \sum \kappa_i \quad \text{in } \Gamma^+$$

for some uniform constant $\omega > 0$. For any $C^2(\overline{M})$ admissible function u satisfying

$$F(W[u]) = f(x)e^{2u} \quad \text{in } M,$$

u is a subsolution of the equation

$$b_1 \Delta u + b_2 |\nabla u|^2 + b_3 = e^{2u}, \tag{7-1}$$

where

$$b_1 = \frac{\omega(n\lambda - 1)}{\min_{\overline{M}} f}, \quad b_2 = \frac{\omega(|a|_{L^\infty} + n|b|_{L^\infty})}{\min_{\overline{M}} f} \quad b_3 = \frac{\omega|\operatorname{tr} B|_{L^\infty}}{\min_{\overline{M}} f}.$$

Before constructing a local supsolution of (7-1), we give some notation. Take a point $y_0 \in M_{\delta/4}$ near the boundary ∂M. Suppose $x_0 \in \partial M$ is the point that satisfies $d(y_0) = \operatorname{dist}_g(x_0, y_0)$. Consider a geodesic running from x_0, passing through y_0, and going out a small distance to a point z_0 with $\operatorname{dist}_g(z_0, x_0) = \eta$. We use $r(x)$ to denote the distance from z_0 to x with respect to the background metric g. We assume that δ and η are small enough that $r^2(x) = (\operatorname{dist}_g(x, z_0))^2$ is smooth in the ball $B_\eta(z_0)$. We may choose normal coordinates $\{e_k\}$. Then we have

$$\Delta r^2(z_0) = 2n.$$

We now assume

$$1 \le \Delta r^2 \le 3n \quad \text{in } B_\eta(z_0).$$

Consider the following auxiliary function defined in $B_\eta(z_0)$:

$$\overline{w}(x) = -\log(\eta^2 - r^2) + \theta \log \frac{\eta^2 - r^2 + \epsilon}{\epsilon} + \log 2 + \tfrac{1}{2}\log(nb_1 + b_2) + \log \eta,$$

where θ and ϵ are constants to be chosen later. It is easy to check that

$$\overline{w}_i = \frac{2rr_i}{\eta^2 - r^2} - \theta \frac{2rr_i}{\eta^2 - r^2 + \epsilon},$$

and

$$\overline{w}_{ij} = \frac{\nabla_{ij}^2 r^2}{\eta^2 - r^2} + \frac{4r^2 r_i r_j}{(\eta^2 - r^2)^2} - \theta \frac{\nabla_{ij}^2 r^2}{\eta^2 - r^2 + \epsilon} - \theta \frac{4r^2 r_i r_j}{(\eta^2 - r^2 + \epsilon)^2}.$$

Consequently, using $|\nabla r| = 1$ and $1 \le \Delta r^2 \le 3n$ in $B_\eta(z_0)$, we derive

$$\begin{aligned}
b_1 \Delta \overline{w} + b_2 |\nabla \overline{w}|^2 + b_3 \\
&= b_1 \frac{\Delta r^2}{\eta^2 - r^2} + \frac{4(b_1 + b_2) r^2}{(\eta^2 - r^2)^2} - \frac{b_1 \theta \Delta r^2}{\eta^2 - r^2 + \epsilon} - \frac{4(b_1 - b_2 \theta)\theta r^2}{(\eta^2 - r^2 + \epsilon)^2} \\
&\qquad\qquad - \frac{8 b_2 \theta r^2}{(\eta^2 - r^2)(\eta^2 - r^2 + \epsilon)} + b_3 \\
&\le \frac{3n b_1 \eta^2 + (3b_1 + 4b_2) r^2}{(\eta^2 - r^2)^2} - \frac{b_1 \theta}{\eta^2 - r^2 + \epsilon} - \frac{4(b_1 - b_2 \theta)\theta r^2}{(\eta^2 - r^2 + \epsilon)^2} + b_3.
\end{aligned}$$

Now choosing $\theta < b_1/(2b_2)$, $\eta < \sqrt{b_1 \theta/(2b_3)}$, $\epsilon < \eta^2$, and using $r \le \eta$, one obtains

$$b_1 \Delta \overline{w} + b_2 |\nabla \overline{w}|^2 + b_3 \le \frac{4(n b_1 + b_2)\eta^2}{(\eta^2 - r^2)^2} \le e^{2\overline{w}}.$$

Since $\overline{w}|_{\partial B_\eta(z_0)} = +\infty$, maximum principle implies

$$u \le \overline{w} \quad \text{in } B_\eta(z_0);$$

hence

$$u(y_0) \le -\log d(y_0) + \theta \log \frac{2\eta d(y_0) + \epsilon}{\epsilon} + \log 2 + \tfrac{1}{2} \log(n b_1 + b_2). \tag{7-2}$$

Now we complete the proof as follows.

Proof of Theorem 2. We use the notation of Section 4. The argument here is similar to that in [Guan 2008]. Let's consider the locally defined auxiliary functions

$$v_m^\gamma := \gamma \log \frac{m\delta^2}{md + \delta^2} \quad \text{in } M_\delta,$$

where γ is some small positive constant to be chosen later and $m = 1, 2, 3, \ldots$. It is direct to check that

$$\begin{gathered}
v_m^\gamma|_{\partial M} = \gamma \log m, \\
\gamma \log \tfrac{1}{2}\delta \le v_m^\gamma|_{\{d(x) = \delta\}} \le \gamma \log \delta.
\end{gathered} \tag{7-3}$$

By a direct computation, we obtain

$$\begin{aligned} W[v_m^\gamma]_{ij} &= \frac{(\lambda+b\gamma)\gamma m^2}{(md+\delta^2)^2}|\nabla d|^2 g_{ij} + \frac{a\gamma^2 m^2}{(md+\delta^2)^2} d_i d_j - \frac{\gamma m^2}{(md+\delta^2)^2} d_i d_j \\ &\qquad\qquad - \frac{\gamma m}{md+\delta^2}(\lambda \Delta d g_{ij} - d_{ij}) + B_{ij} \\ &\ge \frac{(\varepsilon_0 - (|a|_{L^\infty(\overline{M})} + |b|_{L^\infty(\overline{M})})\gamma)\gamma m^2}{(md+\delta^2)^2}|\nabla d|^2 g_{ij} \\ &\qquad\qquad - \frac{\gamma m}{md+\delta^2} C' g_{ij} - C'' g_{ij}, \end{aligned}$$

where C' and C'' are some large constants depending only on λ, $|B|_{g(\overline{M})}$ and the geometric quantities of $(\overline{M}, g)$, independent of δ. Choosing

$$\gamma \le \frac{\varepsilon_0}{2(|a|_{L^\infty(\overline{M})} + |b|_{L^\infty(\overline{M})})} \quad \text{and} \quad \delta \le \min\left\{1, \frac{\varepsilon_0}{16C'}, \frac{\varepsilon_0 \gamma}{64C''}\right\},$$

and observing that $|\nabla d| > 1/2$ in M_δ, we derive

$$\begin{aligned} W[v_m^\gamma]_{ij} &\ge \left(\frac{\varepsilon_0 m}{4(md+\delta^2)} - C'\right)\frac{\gamma m}{md+\delta^2} g_{ij} - C'' g_{ij} \\ &\ge \frac{\varepsilon_0 \gamma m^2}{8(md+\delta^2)^2} g_{ij} - C'' g_{ij} \\ &\ge \frac{\varepsilon_0 \gamma m^2}{16(md+\delta^2)^2} g_{ij}. \end{aligned}$$

Consequently, if $\gamma \le \min\{1, \frac{1}{2}\varepsilon_0/(|a|_{L^\infty(\overline{M})} + |b|_{L^\infty(\overline{M})})\}$ and δ is small enough, then

$$\begin{aligned} F(W[v_m^\gamma]) &\ge \frac{\varepsilon_0 \gamma m^2}{16(md+\delta^2)^2} F(e) \\ &= \frac{\varepsilon_0 \gamma F(e)}{16\delta^4} \exp(2v_m^\gamma/\gamma) \\ &\ge f(x) e^{2v_m^\gamma} \end{aligned} \tag{7-4}$$

in M_δ. For any integer $m \ge 1$, let $u_m \in C^\infty(\overline{M})$ be the admissible solution of the Dirichlet problem

$$\begin{cases} F(W[u]) = f(x)e^{2u} & \text{in } M, \\ u = \gamma \log m & \text{on } \partial M, \end{cases}$$

where γ is the constant has been fixed. Then (7-3), (7-4) and Lemma 4.3 imply

$$u_m \ge v_m^\gamma = \gamma \log \frac{m\delta^2}{md+\delta^2}. \tag{7-5}$$

Recalling (7-2), we obtain for any $m \geq 1$

$$u_m \leq -\log d + C. \tag{7-6}$$

Since $u_m \leq u_{m+1}$ for $m \geq 1$, and the u_m have the boundary control (7-5) and (7-6), the limit

$$u(x) := \lim_{m\to\infty} u_m(x)$$

exists for all $x \in M$ and satisfies

$$-C - \gamma \log d \leq u(x) \leq -\log d + C$$

near ∂M.

For any compact subset $K \subset M$, by the boundary control above and the a priori estimates of Proposition 3.1, Lemma 4.1 and Lemma 5.1, we obtain

$$|u_m|_{C^{2,\alpha}(K)} \leq C,$$

where $0 < \alpha < 1$, $C = C(K)$ is independent of m. Thus u is a solution of (1-5). □

References

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

[Aviles and McOwen 1988] P. Aviles and R. C. McOwen, "Complete conformal metrics with negative scalar curvature in compact Riemannian manifolds", *Duke Math. J.* **56**:2 (1988), 395–398. MR 89b:58224 Zbl 0645.53023

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

[Ge and Wang 2006] Y. Ge and G. Wang, "On a fully nonlinear Yamabe problem", *Ann. Sci. École Norm. Sup.* (4) **39**:4 (2006), 569–598. MR 2007k:53040 Zbl 1121.53027

[Guan 2007] B. Guan, "Conformal metrics with prescribed curvature functions on manifolds with boundary", *Amer. J. Math.* **129**:4 (2007), 915–942. MR 2008e:53053 Zbl 1130.53026

[Guan 2008] B. Guan, "Complete conformal metrics of negative Ricci curvature on compact manifolds with boundary", *Int. Math. Res. Not.* **2008** (2008), Art. ID rnn 105. MR 2009m:53091 Zbl 1160.53018

[Guan and Wang 2003a] P. Guan and G. Wang, "A fully nonlinear conformal flow on locally conformally flat manifolds", *J. Reine Angew. Math.* **557** (2003), 219–238. MR 2004e:53101 Zbl 1033.53058

[Guan and Wang 2003b] P. Guan and G. Wang, "Local estimates for a class of fully nonlinear equations arising from conformal geometry", *Int. Math. Res. Not.* **2003**:26 (2003), 1413–1432. MR 2003m:53055 Zbl 1042.53021

[Gursky and Viaclovsky 2003] M. J. Gursky and J. A. Viaclovsky, "Fully nonlinear equations on Riemannian manifolds with negative curvature", *Indiana Univ. Math. J.* **52**:2 (2003), 399–419. MR 2004a:53039 Zbl 1036.53025

[Gursky and Viaclovsky 2007] M. J. Gursky and J. A. Viaclovsky, "Prescribing symmetric functions of the eigenvalues of the Ricci tensor", *Ann. of Math.* (2) **166**:2 (2007), 475–531. MR 2008k:53068 Zbl 1142.53027

[Gursky et al. 2011] M. Gursky, J. Streets, and M. Warren, "Existence of complete conformal metrics of negative Ricci curvature on manifolds with boundary.", *Calc. Var. Partial Differ. Equ.* **41**:1-2 (2011), 21–43. Zbl 05884575

[Li and Li 2003] A. Li and Y. Li, "On some conformally invariant fully nonlinear equations", *Comm. Pure Appl. Math.* **56**:10 (2003), 1416–1464. MR 2004e:35072 Zbl 1155.35353

[Li and Sheng 2005] J. Li and W. Sheng, "Deforming metrics with negative curvature by a fully nonlinear flow", *Calc. Var. Partial Differential Equations* **23**:1 (2005), 33–50. MR 2005m:53121 Zbl 1069.53054

[Lohkamp 1994] J. Lohkamp, "Negative bending of open manifolds", *J. Differential Geom.* **40**:3 (1994), 461–474. MR 96k:53061 Zbl 0840.53023

[Schnürer 2005] O. C. Schnürer, "Schouten tensor equations in conformal geometry with prescribed boundary metric", *Electron. J. Differential Equations* **2005** (2005), No. 81. MR 2006h:53027 Zbl 1074.53032

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

[Sheng and Zhang 2007] W. Sheng and Y. Zhang, "A class of fully nonlinear equations arising from conformal geometry", *Math. Z.* **255**:1 (2007), 17–34. MR 2007k:53042 Zbl 1133.53036

[Sheng et al. 2007] W.-M. Sheng, N. S. Trudinger, and X.-J. Wang, "The Yamabe problem for higher order curvatures", *J. Differential Geom.* **77**:3 (2007), 515–553. MR 2008i:53048 Zbl 1133.53035

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

[Trudinger and Wang 2010] N. S. Trudinger and X.-J. Wang, "The intermediate case of the Yamabe problem for higher order curvatures", *Int. Math. Res. Not.* **2010**:13 (2010), 2437–2458. MR 2011g:53074 Zbl 1194.53033

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

Received May 14, 2010. Revised February 23, 2011.

QI-RUI LI
DEPARTMENT OF MATHEMATICS
ZHEJIANG UNIVERSITY
HANGZHOU 310027
CHINA
qirui-li@hotmail.com

WEIMIN SHENG
DEPARTMENT OF MATHEMATICS
ZHEJIANG UNIVERSITY
HANGZHOU 310027
CHINA
weimins@zju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

POLYCYCLIC QUASICONFORMAL MAPPING CLASS SUBGROUPS

KATSUHIKO MATSUZAKI

For a subgroup of the quasiconformal mapping class group of a Riemann surface in general, we give an algebraic condition which guarantees its discreteness in the compact-open topology. Then we apply this result to its action on the Teichmüller space.

1. Introduction

We consider a Riemann surface R in general, not necessarily topologically finite, and a subgroup G consisting of quasiconformal mapping classes of R. Such a group usually appears as acting on the infinite dimensional Teichmüller space of R and in particular discreteness of its orbit is often discussed. In this case, the discreteness of G is understood through the action on the Teichmüller space. In this paper however, we first start from a more basic viewpoint on G as surface homeomorphisms and then look into its action on the Teichmüller space.

Throughout this introduction, we assume that a Riemann surface R has no ideal boundary at infinity ∂R for the sake of simplicity. The quasiconformal mapping class group $\mathrm{MCG}(R)$ of R is the group of all quasiconformal automorphisms g of R modulo homotopy equivalence. We introduce a topology for this group induced by the compact-open topology of homeomorphisms of R. Then a subgroup G of $\mathrm{MCG}(R)$ is defined to be *discrete* if it is discrete in this topology. Our main theorem refers to a certain algebraic condition under which G is always discrete. Here we say that a group G is *polycyclic* if G is solvable and if every subgroup of G is finitely generated.

Theorem 2.4. *If a subgroup G of* $\mathrm{MCG}(R)$ *is polycyclic, then G is discrete.*

This result is sharp in a sense that there is a counterexample for either a finitely generated solvable group or an infinitely generated abelian group.

MSC2000: primary 30F60; secondary 37F30.
Keywords: quasiconformal mapping class group, Teichmüller space, discrete, polycyclic.

In the first of the application of this theorem, we deal with stationary mapping class subgroups and consider their action on Teichmüller spaces. The quasiconformal mapping class group $\mathrm{MCG}(R)$ acts on the Teichmüller space $T(R)$ of a Riemann surface R biholomorphically and isometrically. A subgroup $G \subset \mathrm{MCG}(R)$ is called *stationary* if there exists a compact subsurface V of R such that every representative g of every mapping class $[g] \in G$ satisfies $g(V) \cap V \neq \varnothing$.

A basic nature of stationary subgroups in connection with their discreteness in the compact-open topology and discontinuity of the action on the Teichmüller space is that, if $G \subset \mathrm{MCG}(R)$ is stationary and discrete, then G acts discontinuously on $T(R)$. Then we have the following consequence from the main theorem. Recall that we assume $\partial R = \varnothing$ until the end of this section.

Corollary 4.2. *If a polycyclic subgroup G of* $\mathrm{MCG}(R)$ *is stationary, then G acts discontinuously on $T(R)$.*

We expect that this result should be valid for every finitely generated stationary subgroup $G \subset \mathrm{MCG}(R)$.

In the second application of Theorem 2.4, we deal with asymptotically conformal mapping class subgroups. We say that a quasiconformal homeomorphism of a Riemann surface R is asymptotically conformal if its complex dilatation vanishes at infinity of R. We say that a subgroup $G \subset \mathrm{MCG}(R)$ is *asymptotically conformal* if there exists some $p \in T(R)$ such that every element of G can be realized as an asymptotically conformal automorphism of the Riemann surface R_p corresponding to p. We denote by $\mathrm{MCG}_p(R)$ the subgroup of $\mathrm{MCG}(R)$ having this property for $p \in T(R)$.

Theorem 5.1. *If an asymptotically conformal subgroup G of* $\mathrm{MCG}_p(R)$ *for $p \in T(R)$ is polycyclic, then the orbit $G(p)$ is a discrete set in $T(R)$.*

One may ask a question about how the algebraic assumption on G can be relaxed for this statement.

2. Discreteness of mapping class subgroups

We always assume that a Riemann surface R is hyperbolic, that is, R is represented by a Fuchsian group F acting on the unit disk $\mathbb{D}$ and is endowed with the hyperbolic metric. The quasiconformal mapping class group $\mathrm{MCG}(R)$ for R is the group of all homotopy classes $[g]$ of quasiconformal automorphisms g of R. Here the homotopy is considered to be relative to *the ideal boundary at infinity* ∂R of R, where $\partial R = (\partial\mathbb{D} - \Lambda(F))/F$ for the limit set $\Lambda(F)$ of F. This means that, when $\partial R \neq \varnothing$, two quasiconformal automorphisms g_0 and g_1 are regarded as homotopic if there is a homotopy $\Phi : R \times [0, 1] \to R$ between $g_0 = \Phi(\,\cdot\,, 0)$ and $g_1 = \Phi(\,\cdot\,, 1)$ such that its extension to each $x \in \partial R$ is constant over $[0, 1]$.

The compact-open topology on the space of all homeomorphic automorphisms of R induces a topology on $\mathrm{MCG}(R)$. More precisely, we say that a sequence of mapping classes $[g_n] \in \mathrm{MCG}(R)$ converges to a mapping class $[g] \in \mathrm{MCG}(R)$ in the compact-open topology if we can choose representatives $g_n \in [g_n]$ and $g \in [g]$ satisfying that g_n converges to g locally uniformly on R. When R has the ideal boundary at infinity ∂R, we further require that the extensions $\bar{g}_n$ of the quasiconformal automorphisms g_n to ∂R converge to the extension $\bar{g}$ of g in such a way that $\bar{g}_n$ is identical with $\bar{g}$ on a compact subset $W_n \subset \partial R$, where $\{W_n\}_{n=1}^{\infty}$ is some *compact exhaustion* of ∂R, that is, an increasing sequence of compact subsets of ∂R satisfying that the closure of the union of all W_n is ∂R. We call this topology on $\mathrm{MCG}(R)$ the *compact-open topology relative to the boundary*. If $[g_n]$ converges to $[g]$ in the compact-open topology relative to the boundary, then there are quasisymmetric automorphisms $\tilde{g}_n$ and $\tilde{g}$ of the unit circle $\partial\mathbb{D}$ corresponding to $[g_n]$ and $[g]$ respectively such that the sequence $\tilde{g}_n$ converges uniformly to $\tilde{g}$.

Definition. We say that a subgroup G of $\mathrm{MCG}(R)$ is *discrete* if it is a discrete set in $\mathrm{MCG}(R)$ with respect to the compact-open topology relative to the boundary. The discreteness is equivalent to the condition that, if a sequence of mapping classes $\{[g_n]\}_{n=1}^{\infty} \subset \mathrm{MCG}(R)$ converges to [id], then $[g_n] = [\mathrm{id}]$ for all sufficiently large n.

Concerning the discreteness of the full mapping class group $\mathrm{MCG}(R)$, we have a simple characterization.

Proposition 2.1. *The quasiconformal mapping class group* $\mathrm{MCG}(R)$ *is discrete if and only if R is analytically finite*, *that is*, *R is a compact Riemann surface from which at most finitely many points are removed.*

Proof. Assume that R is analytically finite. In this case, there are a finite number of simple closed geodesics $\{c_i\}_{i=1}^{k}$ such that, if $[g] \in \mathrm{MCG}(R)$ satisfies that $g(c_i)$ is freely homotopic to c_i for every i, then $[g] = [\mathrm{id}]$. If a sequence of mapping classes $\{[g_n]\}_{n=1}^{\infty}$ converges to [id], then $g_n(c_i)$ is freely homotopic to c_i for every i and for all sufficiently large n. This implies that $\mathrm{MCG}(R)$ is discrete.

Conversely, assume that R is not analytically finite. If R is topologically finite, that is, the fundamental group $\pi_1(R)$ of R is finitely generated, then R should have the ideal boundary at infinity and clearly $\mathrm{MCG}(R)$ is not discrete in this case. If R is not topologically finite, then there is an infinite sequence of simple closed geodesics $\{c_n\}_{n=1}^{\infty}$ diverging to the infinity of R, in other words, escaping from any compact subset of R. Let $[\tau_n]$ be the mapping class caused by the Dehn twist along c_n. Then $[\tau_n] \neq [\mathrm{id}]$ and $\{[\tau_n]\}_{n=1}^{\infty}$ converges to [id]. This implies that $\mathrm{MCG}(R)$ is not discrete. □

We will consider the discreteness of countable subgroups of $\mathrm{MCG}(R)$. Note that $\mathrm{MCG}(R)$ is uncountable in many cases when R is analytically infinite [Matsuzaki

2005]. An uncountable subgroup G of $\mathrm{MCG}(R)$ is not discrete, as the following proposition asserts.

Proposition 2.2. *Assume that R has no ideal boundary at infinity ∂R. If a subgroup $G \subset \mathrm{MCG}(R)$ is uncountable, then G is not discrete.*

Proof. Let $\{c_i\}_{i=1}^{\infty}$ be the family of (free homotopy classes of) all simple closed geodesics on R. We first consider the images of c_1 under G. Since G is uncountable whereas $\{c_i\}$ is countable, there are uncountably many elements of G that map c_1 to simple closed curves freely homotopic to each other. Then, by composing the inverse of one of these elements, we have uncountably many elements of G that keep c_1 in its free homotopy class. Next we consider the images of c_2 under this uncountable subset of G and obtain uncountably many elements of G that keep c_1 and c_2 in their free homotopy classes. By continuing this process and then by taking the diagonal, we can choose a sequence $\{[g_n]\}_{n=1}^{\infty}$ of elements in G such that $g_n(c_i)$ is freely homotopic to c_i for all $i = 1, 2, \ldots, n$ and for each n. This implies that $\{[g_n]\}$ converges to [id]. □

In this section, we investigate an algebraic condition on a countable subgroup G of $\mathrm{MCG}(R)$ under which G is always discrete. Our fundamental result is the following. The proof will be given in the next section.

Theorem 2.3. *If $G \subset \mathrm{MCG}(R)$ is a finitely generated abelian group, then G is discrete.*

Note that both assumptions that G is finitely generated and that G is abelian are necessary for the above theorem as examples below show. However, we cannot have the converse statement to the theorem. In fact, for any countable group G, there exists a discrete subgroup of $\mathrm{MCG}(R)$ for some Riemann surface R that is isomorphic to G. Indeed, we can construct R so that its conformal automorphism group, which is always discrete unless $\pi_1(R)$ is abelian, contains such a subgroup.

Examples. (1) First we give an indiscrete $G \subset \mathrm{MCG}(R)$ that is abelian but not finitely generated. Let R be a Riemann surface with an infinite family of mutually disjoint simple closed geodesics $\{c_n\}_{n=1}^{\infty}$ and G a subgroup of $\mathrm{MCG}(R)$ generated by all the mapping classes $[\tau_n]$ caused by the Dehn twist along c_n for each integer $n \geq 1$. Since $[\tau_n]$ converges to [id], G is not discrete though G is abelian.

(2) Next we give an indiscrete $G \subset \mathrm{MCG}(R)$ that is finitely generated but not abelian. Assume that there are a simple closed geodesic c_0 on R and a mapping class $[g] \in \mathrm{MCG}(R)$ such that the images $\{g^n(c_0)\}_{n\in\mathbb{Z}}$ of c_0 under the iteration of a representative $g \in [g]$ are mutually disjoint. Define c_n to be the simple closed geodesic freely homotopic to $g^n(c_0)$ and $[\tau_n]$ to be the mapping classes caused by the Dehn twist along c_n. Let G be a subgroup of $\mathrm{MCG}(R)$ generated by two elements $[g]$ and $[\tau_0]$. Since $[g]^n[\tau_0] = [\tau_n][g]^n$ for every integer $n \in \mathbb{Z}$, we see that

G contains the subgroup G' generated by all such $[\tau_n]$. Hence G is not discrete as Example (1) shows.

In the second example above, the group G is solvable since the commutator subgroup $[G, G]$ is contained in the abelian subgroup G'. Although G itself is finitely generated, G' is not, so G is not discrete. Hence we consider the following stronger condition than solvability which requires all its subgroups to be finitely generated.

Definition. We say that a group G is *polycyclic* if G is solvable and if every subgroup of G is finitely generated.

See [Wolf 1968] for other equivalent conditions for G to be polycyclic. This name comes from the fact that G is polycyclic if and only if G has a finite normal chain of subgroups $G = G_0 \triangleright G_1 \triangleright \cdots \triangleright G_m = \{1\}$ such that each quotient group G_{i-1}/G_i $(i = 1, \ldots, m)$ is cyclic. We can say that G is polycyclic when G is obtained in finitely many simple steps from finitely generated abelian groups.

Theorem 2.4. *If $G \subset \mathrm{MCG}(R)$ is a polycyclic group, then G is discrete.*

This extension of Theorem 2.3 is obtained by an inductive argument which is easily seen from the following assertion.

Lemma 2.5. *Assume that every subgroup of $G \subset \mathrm{MCG}(R)$ is finitely generated. If G is not discrete, then neither is the commutator subgroup $[G, G]$.*

Proof. Since G is not discrete, there is a sequence $\{[g_n]\}_{n=1}^{\infty}$ in G that converges to $[\mathrm{id}]$ as $n \to \infty$. Then we see that for every $n_0 \geq 1$, there exist $m, n \geq n_0$ such that $[g_m]$ and $[g_n]$ do not commute. Indeed, if not, there is n_0 such that $[g_m]$ and $[g_n]$ commute for any $m, n \geq n_0$. Then a subgroup G' of G generated by $\{[g_n]\}_{n \geq n_0}$ is abelian and G' is not discrete. By assumption, G' is finitely generated. However, this contradicts Theorem 2.3.

Fix some $n_0 \geq 1$. We choose $m_1, n_1 \geq n_0$ such that $[h_1] := [[g_{m_1}], [g_{n_1}]]$ is not the identity $[\mathrm{id}]$. Then we choose $m_2, n_2 \geq \max\{m_1, n_1\}$ such that $[h_2] := [[g_{m_2}], [g_{n_2}]]$ is not the identity. Inductively, for each $i \geq 1$, we choose $m_i, n_i \geq \max\{m_{i-1}, n_{i-1}\}$ such that $[h_i] := [[g_{m_i}], [g_{n_i}]]$ is not the identity. Then every $[h_i]$ belongs to the commutator subgroup $[G, G]$ of G and $[h_i]$ converges to $[\mathrm{id}]$ as $i \to \infty$. This implies that $[G, G]$ is not discrete. $\square$

3. Restraint of mapping class groups

In this section, we will prove Theorem 2.3. The proof uses a certain property of mapping class groups, not necessarily satisfied for abstract groups in general. We first explain this situation by the following example.

Example. Let $\mathfrak{S}_\infty$ be the infinite symmetric group acting on a countable set $X = \{1, 2, \ldots\}$ as permutations. We consider an element $g = (1)(23)(456)\cdots$ of $\mathfrak{S}_\infty$ which gives a cyclic permutation on mutually disjoint subsets of n points in X where n runs over all positive integers. Then we see that $g^{n!}$ converges to id in the compact-open topology with respect to the discrete topology on X. In particular, the cyclic subgroup $\langle g \rangle$ is not discrete.

Let $X = \{c_i\}_{i=1}^\infty$ be the family of (free homotopy classes of) all simple closed geodesics on a Riemann surface R. The quasiconformal mapping class group $\mathrm{MCG}(R)$ acts faithfully on the countable set X by the correspondence of the free homotopy class $g(c)$ to $[g] \cdot c$ for any $[g] \in \mathrm{MCG}(R)$ and for any $c \in X$. In this way, we can represent $\mathrm{MCG}(R)$ as a subgroup of $\mathfrak{S}_\infty$. As the above example shows, an arbitrary subgroup of $\mathfrak{S}_\infty$ cannot have the required property which we want to prove in Theorem 2.3. The nature in which $\mathrm{MCG}(R) \subset \mathfrak{S}_\infty$ originates from R gives a certain restriction on the action of $\mathrm{MCG}(R)$ and we must use this constraint in order to prove our theorem. The following lemma can be regarded as one of such properties of $\mathrm{MCG}(R)$.

Lemma 3.1. *For every element $[g] \in \mathrm{MCG}(R)$ of infinite order, there exists either a compact subsurface V in R or a compact subset V' in an arbitrarily given compact exhaustion of the ideal boundary at infinity ∂R such that either the restriction $g^n|_V$ is homotopic to $\mathrm{id}|_V$ on R or the extension $\bar{g}^n$ is the identity on V' for no positive integer $n \in \mathbb{N}$.*

Proof. Suppose to the contrary that there is no such compact subsurface V in R nor compact subset V' in the compact exhaustion of ∂R. Then, for any compact subsurface $V_1 \subset R$, there is $n_1 \in \mathbb{N}$ such that $g^{n_1}|_{V_1}$ is homotopic to $\mathrm{id}|_{V_1}$ on R. Also, for any compact subset V_1' in the compact exhaustion of ∂R, there is $n_1' \in \mathbb{N}$ such that $\bar{g}^{n_1'}$ is the identity on V'. Set $h = g^{n_1 n_1'}$. Since h is not homotopic to the identity on R relative to ∂R, there is either some compact subsurface $V_2 \subset R$ including V_1 such that $h|_{V_2}$ is not homotopic to $\mathrm{id}|_{V_2}$ on R or some compact subset V_2' in the compact exhaustion of ∂R including V_1' such that $\bar{h}$ is not the identity on V_2'. We assume that the first case occurs. The argument for the second case is similar.

For that compact subsurface V_2, there is $n_2 \in \mathbb{N}$ such that $g^{n_2}|_{V_2}$ is homotopic to $\mathrm{id}|_{V_2}$ on R. We may assume that n_2 is a proper multiple of $n_1 n_1'$, that is, $n_2 = k n_1 n_1'$ for some integer $k > 1$. Then $h|_{V_1} \sim \mathrm{id}|_{V_1}$, $h|_{V_2} \not\sim \mathrm{id}|_{V_2}$ and $h^k|_{V_2} \sim \mathrm{id}|_{V_2}$, where $\sim$ means that they are homotopic to each other on R. However, this is impossible, as we see in the following. Represent the Riemann surface R by a Fuchsian group F acting on the unit disk $\mathbb{D}$ and take a subgroup F_1 of F corresponding to the subsurface V_1. Choose a quasisymmetric automorphism $\tilde{h}$ of $\partial\mathbb{D}$ corresponding to h so that $\tilde{h}$ is the identity on the limit set $\Lambda(F_1) \subset \partial\mathbb{D}$ of F_1. Also, take a subgroup F_2 of F corresponding to the subsurface V_2 which contains F_1.

Then the quasisymmetric automorphism $\tilde{h}$ is not the identity on the limit set $\Lambda(F_2)$ containing $\Lambda(F_1)$. This implies that there is a point $x \in \Lambda(F_2) - \Lambda(F_1)$ that is moved by $\tilde{h}$. Since the movement of x is towards one direction in some interval contained in $\partial\mathbb{D} - \Lambda(F_1)$, it cannot return to the original place under the iteration of $\tilde{h}$. Thus $\tilde{h}^k(x) \neq x$, which violates the condition that $h^k|_{V_2} \sim \mathrm{id}|_{V_2}$. □

Although the following fact is not special for mapping class groups, the property of discreteness is shared with a subgroup of finite index as in usual arguments. We also use this fact in the proof of Theorem 2.3.

Proposition 3.2. *Let G' be a subgroup of $G \subset \mathrm{MCG}(R)$ of finite index. If G' is discrete, then so is G.*

Proof. If G is not discrete, there is a sequence of distinct elements $[g_n]$ of G that converges to [id]. Since the index of G' in G is finite, we may assume that the $[g_n]$ are all in the same coset, say, $G'[h]$ for some $[h] \in G$. Then $[g'_n] = [g_n] \cdot [h]^{-1}$ belong to G' and converge to $[h]^{-1}$. This contradicts the assumption that G' is discrete. □

Now we are ready to prove our fundamental result.

Proof of Theorem 2.3. By Proposition 3.2, we may assume that G is isomorphic to a free abelian group $\mathbb{Z}^m$ of rank $m \geq 1$. We will prove the statement of the theorem by induction with respect to m. First, we show that the statement is valid when $m = 1$. Assume that $G \cong \mathbb{Z}$ is not discrete, that is, there is a sequence of elements in G converging to [id]. When R has the ideal boundary at infinity ∂R, some compact exhaustion of ∂R is associated to this converging sequence. For a generator $[g] \in \mathrm{MCG}(R)$ of G, Lemma 3.1 gives either a compact subsurface V of R or a compact subset V' in the exhaustion of ∂R as in its statement. However, since G is not discrete, there is some $n \in \mathbb{N}$ such that $g^n|_V$ is homotopic to $\mathrm{id}|_V$ on R and the extension $\bar{g}^n$ of g^n to ∂R is the identity on V'. This contradicts the choice of V and V'.

We assume that the statement is true for any subgroup of $\mathrm{MCG}(R)$ isomorphic to $\mathbb{Z}^j$ for every integer j with $1 \leq j \leq m-1$. Let G be a subgroup of $\mathrm{MCG}(R)$ isomorphic to $\mathbb{Z}^m$; we prove that G is discrete. Suppose to the contrary that G is not discrete. Then we have a sequence $[g_n] \in G$ converging to [id] as well as a compact exhaustion of ∂R associated with this sequence. We will choose a subsequence of $[g_n]$ so that any m elements in the subsequence generates a subgroup isomorphic to $\mathbb{Z}^m$. To this end, first observe that all the elements $[g_n]$ in the convergent sequence cannot be contained in a finite union of subgroups of G that are isomorphic to $\mathbb{Z}^j$ with $1 \leq j \leq m-1$, by the induction assumption. Then choose a subsequence $[g_{n(i)}]$ in the following way. The first $m-1$ entries $[g_{n(1)}], \ldots, [g_{n(m-1)}]$ are chosen so that they are linearly independent over $\mathbb{Z}$. Suppose that we have already chosen l entries $G_l = \{[g_{n(1)}], \ldots, [g_{n(l)}]\}$ for $l \geq m-1$.

Then take the $(l+1)$-st entry $[g_{n(l+1)}]$ so that any $m-1$ elements of G_l together with $[g_{n(l+1)}]$ are linearly independent over $\mathbb{Z}$, in other words, $[g_{n(l+1)}]$ belongs to no maximal proper subgroup ($\cong \mathbb{Z}^{m-1}$) of G containing $m-1$ elements of G_l. The reason why we can choose such $[g_{n(l+1)}]$ is that, if not, all $[g_n]$ must be contained in the union of the finite number of subgroups of G determined by any $m-1$ elements of G_l. By this construction, it is clear that any m elements in the subsequence $[g_{n(i)}]$ generate a subgroup isomorphic to $\mathbb{Z}^m$.

Fix an arbitrary nontrivial element $[g_0] \in G$. By Lemma 3.1, we take either a compact subsurface V of R such that $g_0^n|_V \not\sim \mathrm{id}|_V$ or a compact subset V' in the exhaustion of ∂R such that $\bar{g}_0^n|_{V'} \neq \mathrm{id}|_{V'}$ for all $n \in \mathbb{N}$. We only consider the first case. The second case is similar. Since we are assuming that $[g_{n(i)}]$ converges to [id], there is some i_0 such that $g_{n(i)}|_V \sim \mathrm{id}|_V$ for every $i \geq i_0$. Take m arbitrary elements $[g_{n(i)}]$ with $i \geq i_0$ and rename them as $[g_i]$ $(i = 1, \ldots, m)$. Since they generate a subgroup of G isomorphic to $\mathbb{Z}^m$, a linear combination of $[g_i]$ $(i = 1, \ldots, m)$ over $\mathbb{Z}$ yields some multiple of any element of G. This implies that $[g_0]^n$ for some $n \in \mathbb{N}$ is represented by $[g_1]^{k_1} \cdots [g_m]^{k_m}$ for some $k_i \in \mathbb{Z}$. However, this forces $g_0^n|_V \sim \mathrm{id}|_V$, which contradicts the choice of V. □

4. Discontinuity of the action on the Teichmüller space

We apply our theorem to the action of mapping class subgroups on Teichmüller spaces. For a Riemann surface R, the Teichmüller space $T(R)$ is defined to be the set of all equivalence classes $[f]$ of quasiconformal homeomorphisms f of R. Here we say that two quasiconformal homeomorphisms f_1 and f_2 of R are equivalent if there exists a conformal homeomorphism $h : f_1(R) \to f_2(R)$ such that $f_2^{-1} \circ h \circ f_1$ is homotopic to the identity on R, where the homotopy is considered to be relative to the ideal boundary at infinity ∂R. The Teichmüller distance between two points $[f_1]$ and $[f_2]$ in $T(R)$ is defined by $d_T([f_1], [f_2]) = (1/2) \log K(f)$, where f is an extremal quasiconformal homeomorphism in the sense that its maximal dilatation $K(f)$ is minimal in the homotopy class of $f_2 \circ f_1^{-1}$. Then d_T is a complete distance on $T(R)$. The Teichmüller space $T(R)$ can be embedded in the complex Banach space of all bounded holomorphic quadratic differentials on R', where R' is the complex conjugate of R. In this way, $T(R)$ is endowed with a complex structure. Consult [Lehto 1987; Nag 1988; Gardiner and Lakic 2000] for the theory of Teichmüller spaces.

Each element $[g] \in \mathrm{MCG}(R)$ acts on $T(R)$ from the left as $[g] \cdot [f] = [f \circ g^{-1}]$ for $[f] \in T(R)$. It is evident from the definition that $\mathrm{MCG}(R)$ acts on $T(R)$ isometrically with respect to the Teichmüller distance. It also acts biholomorphically on $T(R)$. Except for few cases where the dimension of $T(R)$ is lower, the action of $\mathrm{MCG}(R)$ on $T(R)$ is faithful. Then $\mathrm{MCG}(R)$ can be represented in the group of all isometric biholomorphic automorphisms of $T(R)$.

We say that a subgroup $G \subset \mathrm{MCG}(R)$ acts at $p = [f] \in T(R)$ *discontinuously* if there exists a neighborhood U of p such that the number of the elements $[g] \in G$ satisfying $[g](U) \cap U \neq \varnothing$ is finite. We denote the orbit of p under G by $G(p)$ and the stabilizer subgroup of G at p by $\mathrm{Stab}_G(p)$. Then G acts discontinuously at p if and only if $G(p)$ is a discrete set and $\mathrm{Stab}_G(p)$ is a finite group. If G acts discontinuously at every point p in $T(R)$, then we say that G acts discontinuously on $T(R)$. When R is analytically finite, $\mathrm{MCG}(R)$ itself acts discontinuously on $T(R)$. However, for a Riemann surface in general, this is not always true. See [Fujikawa 2004] regarding the discontinuity of the action of mapping class groups on Teichmüller spaces.

We consider mapping class subgroups by imposing a stationary property on them in the following sense.

Definition. We call a subgroup G of $\mathrm{MCG}(R)$ *stationary* if there exists a compact subsurface V of R such that every representative g of every mapping class $[g] \in G$ satisfies $g(V) \cap V \neq \varnothing$.

The stationary property puts a certain normalization on a family of quasiconformal automorphisms of R. Under this condition, the discreteness of G in the compact-open topology affects the behavior of its orbit on the Teichmüller space.

Lemma 4.1. *Let G be a stationary subgroup of* $\mathrm{MCG}(R)$ *for a Riemann surface R with $\partial R = \varnothing$. If G is discrete then the orbit $G(p)$ for any $p \in T(R)$ diverges to the infinity of $T(R)$, and in particular, G acts discontinuously on $T(R)$.*

Proof. Compactness of a family of normalized quasiconformal homeomorphisms with uniformly bounded dilatations yields that if there is a sequence $[g_n]$ in a stationary subgroup G of $\mathrm{MCG}(R)$ such that $[g_n](p)$ is bounded in $T(R)$, then a subsequence of some representatives $g_n \in [g_n]$ converges to some quasiconformal automorphism of R locally uniformly. However, if G is discrete in the compact-open topology, then there is no such sequence. This implies that $[g_n](p)$ is bounded in $T(R)$ for no sequence $[g_n] \in G$, that is, the orbit $G(p)$ diverges to the infinity of $T(R)$. □

Combining Theorem 2.4 and Lemma 4.1 immediately yields the following.

Corollary 4.2. *Let G be a stationary subgroup of* $\mathrm{MCG}(R)$ *for a Riemann surface R with $\partial R = \varnothing$. If G is polycyclic, then G acts discontinuously on $T(R)$.*

We expect that this corollary is valid for every finitely generated stationary subgroup G of $\mathrm{MCG}(R)$.

Conjecture. *If a finitely generated subgroup $G \subset$* $\mathrm{MCG}(R)$ *is stationary, then G is discrete.*

If R is analytically finite, then $\mathrm{MCG}(R)$ is finitely generated and stationary. In this case, $\mathrm{MCG}(R)$ is discrete and acts on $T(R)$ discontinuously. The above conjecture can be regarded as a generalization of this property for mapping class groups of analytically finite Riemann surfaces.

There is an example of an infinitely generated (countable) stationary subgroup G such that G does not act discontinuously on $T(R)$. This is obtained similarly to Example (1) in Section 2 but we must further assume that the lengths of the simple closed geodesics c_n in the example tend to zero as $n \to \infty$.

Remark. If we assume a bounded geometry condition on the hyperbolic metric on R, then we do not have to impose any algebraic condition on a stationary subgroup G for the discontinuity of its action on $T(R)$. This result was proved in [Fujikawa 2004; Fujikawa et al. 2004]. See also these papers for the definition of the bounded geometry condition, to which we add $\partial R = \varnothing$.

5. Discreteness of the orbit on a fiber over the asymptotic Teichmüller space

In this section, we impose a certain analytic condition on a subgroup of the quasiconformal mapping class group and show the discreteness of its orbit in the Teichmüller space. Our condition also generalizes certain properties of the mapping class group of an analytically finite Riemann surface.

A quasiconformal homeomorphism f of a Riemann surface R is called *asymptotically conformal* if, for every $\varepsilon > 0$, there exists a compact subsurface V of R such that the maximal dilatation of f restricted to $R - V$ is less than $1 + \varepsilon$. The asymptotic Teichmüller space $AT(R)$ of R is defined by replacing the words "conformal automorphisms" with "asymptotically conformal automorphisms" in the definition of the Teichmüller space $T(R)$. Since a conformal automorphism is asymptotically conformal, there is a projection $\alpha : T(R) \to AT(R)$. We denote the fiber of α containing $p \in T(R)$ by T_p, that is, $T_p = \alpha^{-1}(\alpha(p))$. Consult [Earle et al. 2000; 2002; 2004; Gardiner and Lakic 2000] for the theory of asymptotic Teichmüller spaces.

The quasiconformal mapping class group $\mathrm{MCG}(R)$ acts on $T(R)$ preserving the fiber structure of α. Hence it acts on $AT(R)$. We define $\mathrm{MCG}_p(R)$ to be the subgroup of $\mathrm{MCG}(R)$ consisting of all elements keeping the fiber T_p invariant. Every element of $\mathrm{MCG}_p(R)$ can be realized as an asymptotically conformal automorphism of the Riemann surface R_p corresponding to p. We say that a subgroup G of $\mathrm{MCG}(R)$ is asymptotically conformal if G is a subgroup of $\mathrm{MCG}_p(R)$ for some $p \in T(R)$. When R is analytically finite, $AT(R)$ consists of a single point and $\mathrm{MCG}_p(R)$ coincides with the full $\mathrm{MCG}(R)$ for every $p \in T(R)$.

We will show the following theorem concerning the discreteness of the orbit of an asymptotically conformal subgroup.

Theorem 5.1. *For a Riemann surface R with $\partial R = \varnothing$, if an asymptotically conformal subgroup G of $\mathrm{MCG}_p(R)$ is polycyclic, then the orbit $G(p)$ is a discrete set in $T(R)$.*

We first prove this theorem in the case that G is a finitely generated abelian group. Before the proof, we give the definition of an escaping sequence of mapping classes. A sequence $\{[g_n]\}_{n=1}^{\infty}$ of mapping classes in $\mathrm{MCG}(R)$ is *stationary* if there exists a compact subsurface V of R such that every representative g_n of each mapping class $[g_n]$ satisfies $g_n(V) \cap V \neq \varnothing$. If a subgroup G of $\mathrm{MCG}(R)$ is stationary in the previous sense, then every sequence in G is stationary in this sense. On the contrary, a sequence $\{[g_n]\}_{n=1}^{\infty}$ is called *escaping* if, for every compact subsurface V of R, there exists some representative g_n of each mapping class $[g_n]$ such that $\{g_n(V)\}$ diverges to the infinity of R (that is, escapes from every compact subset of R) as $n \to \infty$. Remark that a sequence $\{[g_n]\} \subset \mathrm{MCG}(R)$ can be neither stationary nor escaping, but we can always choose a subsequence either stationary or escaping.

The following lemma is crucial for considering an escaping sequence in an asymptotically conformal mapping class group. The proof has been given in [Matsuzaki 2007; 2010, Theorem 5.6].

Lemma 5.2. *Assume that the fundamental group $\pi_1(R)$ of R is noncyclic. Let G be an abelian subgroup of $\mathrm{MCG}_p(R)$ having an escaping sequence $[g_n]$ such that $[g_n](p) \to p$ as $n \to \infty$. Then $[g](p) = p$ for every $[g] \in G$.*

Then the following inductive step gives the full statement of Theorem 5.1 as we have done in Section 2.

Lemma 5.3. *Assume that $\partial R = \varnothing$ and every subgroup of $G \subset \mathrm{MCG}_p(R)$ is finitely generated. If the orbit $G(p)$ is not a discrete set, then neither is the orbit $G_1(p)$ of the commutator subgroup $G_1 = [G, G]$.*

Proof of Theorem 5.1. Let G be a finitely generated abelian subgroup of $\mathrm{MCG}_p(R)$. If G is stationary, then Corollary 4.2 gives that G acts discontinuously on $T(R)$, and in particular, the orbit $G(p)$ is a discrete set in $T(R)$. This is also true for a stationary sequence in G. If G contains an escaping sequence $\{[g_n]\}$ such that $[g_n](p) \to p$ as $n \to \infty$, then Lemma 5.2 implies that $G(p) = \{p\}$ is a discrete set. Hence, if G is a finitely generated abelian subgroup, then the statement of the theorem is valid. For the general case that G is polycyclic, we apply Lemma 5.3 to obtain the statement. □

Proof of Lemma 5.3. If $G(p)$ is not a discrete set, then we find a sequence $\{[g_n]\}_{n=1}^{\infty} \subset G$ such that $[g_n](p) \neq p$ converges to p as $n \to \infty$. Then we can apply the same arguments as in the proof of Lemma 2.5. Namely, for every $n_0 \geq 1$, there exist $m, n \geq n_0$ such that $[g_m]$ and $[g_n]$ do not commute. Indeed, if not, there is n_0 such that $[g_m]$ and $[g_n]$ commute for any $m, n \geq n_0$. Then the finitely generated subgroup G' of G generated by $\{[g_n]\}_{n \geq n_0}$ is abelian and $G'(p)$ is not a discrete set. However, this contradicts Theorem 5.1 in the finitely generated abelian case. Note that this case has been proved without Lemma 5.3.

Fix some $n_0 \geq 1$. Choose $m_1, n_1 \geq n_0$ such that $[h_1] := \big[[g_{m_1}], [g_{n_1}]\big] \neq [\mathrm{id}]$. Then choose $m_2, n_2 \geq \max\{m_1, n_1\}$ such that $[h_2] := \big[[g_{m_2}], [g_{n_2}]\big] \neq [\mathrm{id}]$. Using induction, for each $i \geq 1$, choose $m_i, n_i \geq \max\{m_{i-1}, n_{i-1}\}$ such that $[h_i] := \big[[g_{m_i}], [g_{n_i}]\big] \neq [\mathrm{id}]$. Then every $[h_i]$ belongs to the commutator subgroup $[G, G]$ of G. Note that all $[h_i]$ are not necessarily distinct. We see that $[h_i](p) \to p$ as $i \to \infty$. Indeed,

$$d([h_i](p), p) \leq 2d([g_{m_i}](p), p) + 2d([g_{n_i}](p), p) \to 0$$

as $i \to \infty$. If $[h_i](p) \neq p$ for infinitely many i, then we are done by passing to a subsequence. Hence we have only to consider the case that all but finitely many $[h_i] \neq [\mathrm{id}]$ belong to the stabilizer subgroup $H = \mathrm{Stab}_G(p)$ of G for p, and in particular the case that H is not trivial.

We may assume that p is the base point of the Teichmüller space $T(R)$. Then there is a conformal automorphism group of R identified with H. Let $\mathrm{Fix}(H)$ be the fixed point locus of H in $T(R)$, which can be identified with the Teichmüller space $T(R/H)$ of the orbifold R/H. If $[g_n](p)$ does not lie in $\mathrm{Fix}(H)$, then there is some $[e_n] \in H$ such that $[e_n][g_n](p) \neq [g_n](p)$. Set $[h_n] = [e_n]^{-1}[g_n]^{-1}[e_n][g_n]$ for such n, which belongs to $[G, G]$ and satisfies $[h_n](p) \neq p$. If there are infinitely many such n, we have $[h_n](p) \to p$, which is the desired consequence. Hence we have only to consider the case that $[g_n](p)$ lies in $\mathrm{Fix}(H)$ for all but finitely many n.

The condition $[g_n](p) \in \mathrm{Fix}(H)$ is equivalent to $[g_n]^{-1}[e][g_n] \in H$ for every $[e] \in H$. This is satisfied if and only if the mapping class $[g_n] \in \mathrm{MCG}(R)$ descends to a mapping class $[\hat{g}_n]$ of R/H. Consider the subgroup of the mapping class group $\mathrm{MCG}(R/H)$ generated by all $\{[\hat{g}_n]\}_{n=1}^{\infty}$. Here $[\hat{g}_n]$ belongs to $\mathrm{MCG}_p(R/H)$ for $p \in T(R/H) = \mathrm{Fix}(H)$. In the case where H is a finite group, this is easily seen. In the case where H is an infinite group, the present situation is possible only when $[g_n]$ belongs to H. Indeed, this follows from the fact that $T_p \cap \mathrm{Fix}(H) = \{p\}$ for the infinite group H [Matsuzaki 2010, Theorem 4.2]. However, since we are dealing with the elements $[g_n] \in G$ satisfying $[g_n](p) \neq p$, this is not the case. Hence, by the same reason as before, we can choose a sequence $\{[h_i]\}$ in $[G, G]$ such that $[h_i](p) \to p$ as $i \to \infty$ and in addition that none of $[h_i]$ belongs to $H = \mathrm{Stab}_G(p)$. This implies $[h_i](p) \neq p$ converges to p as $i \to \infty$, which completes the proof. □

In the remark of the previous section, we mentioned that when R satisfies the bounded geometry condition, we do not have to impose any algebraic condition on G. In particular, G is not necessarily finitely generated. The corresponding statement for the discreteness of the orbit of an asymptotically conformal mapping class subgroup will be the following.

Proposition 5.4. *Assume that a Riemann surface R satisfies the bounded geometry condition. If a subgroup G of $\mathrm{MCG}_p(R)$ is solvable, then the orbit $G(p)$ is a discrete set in $T(R)$.*

However, if $G \subset \mathrm{MCG}_p(R)$ is an infinitely generated (countable) group, for instance, then the orbit is not necessarily a discrete set. Our question asks for some algebraic conditions upon G that guarantee this discreteness.

References

[Earle et al. 2000] C. J. Earle, F. P. Gardiner, and N. Lakic, "Asymptotic Teichmüller space, I: The complex structure", pp. 17–38 in *In the tradition of Ahlfors and Bers* (Stony Brook, NY, 1998), edited by I. Kra and B. Maskit, Contemp. Math. **256**, Amer. Math. Soc., Providence, RI, 2000. MR 2001m:32029 Zbl 0973.30032

[Earle et al. 2002] C. J. Earle, V. Markovic, and D. Saric, "Barycentric extension and the Bers embedding for asymptotic Teichmüller space", pp. 87–105 in *Complex manifolds and hyperbolic geometry* (Guanajuato, 2001), edited by C. J. Earle et al., Contemp. Math. **311**, Amer. Math. Soc., Providence, RI, 2002. MR 2003i:30072 Zbl 1020.30047

[Earle et al. 2004] C. J. Earle, F. P. Gardiner, and N. Lakic, "Asymptotic Teichmüller space, II: The metric structure", pp. 187–219 in *In the tradition of Ahlfors and Bers, III* (Storrs, CT, 2001), edited by W. Abikoff and A. Haas, Contemp. Math. **355**, Amer. Math. Soc., Providence, RI, 2004. MR 2006g:30078 Zbl 1112.30034

[Fujikawa 2004] E. Fujikawa, "Modular groups acting on infinite dimensional Teichmüller spaces", pp. 239–253 in *In the tradition of Ahlfors and Bers, III* (Storrs, CT, 2001), edited by W. Abikoff and A. Haas, Contemp. Math. **355**, Amer. Math. Soc., Providence, RI, 2004. MR 2006c:30052 Zbl 1072.30032

[Fujikawa et al. 2004] E. Fujikawa, H. Shiga, and M. Taniguchi, "On the action of the mapping class group for Riemann surfaces of infinite type", *J. Math. Soc. Japan* **56**:4 (2004), 1069–1086. MR 2005e:30075 Zbl 1064.30047

[Gardiner and Lakic 2000] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory*, Math. Surveys and Monogr. **76**, Amer. Math. Soc., Providence, RI, 2000. MR 2001d:32016 Zbl 0949.30002

[Lehto 1987] O. Lehto, *Univalent functions and Teichmüller spaces*, Graduate Texts in Math. **109**, Springer, New York, 1987. MR 88f:30073 Zbl 0606.30001

[Matsuzaki 2005] K. Matsuzaki, "A countable Teichmüller modular group", *Trans. Amer. Math. Soc.* **357**:8 (2005), 3119–3131. MR 2006f:30052 Zbl 1136.30011

[Matsuzaki 2007] K. Matsuzaki, "A classification of the modular transformations of infinite dimensional Teichmüller spaces", pp. 167–177 in *In the tradition of Ahlfors–Bers, IV* (Ann Arbor, MI, 2005), edited by D. Canary et al., Contemp. Math. **432**, Amer. Math. Soc., Providence, RI, 2007. MR 2009b:30094 Zbl 1170.30020

[Matsuzaki 2010] K. Matsuzaki, "Properties of asymptotically elliptic modular transformations of Teichmüller spaces", pp. 73–84 in *Infinite dimensional Teichmüller space and moduli space*, edited by E. Fujikawa, RIMS Kôkyûroku Bessatsu **B17**, Res. Inst. Math. Sci., Kyoto, 2010.

[Nag 1988] S. Nag, *The complex analytic theory of Teichmüller spaces*, Wiley-Interscience, New York, 1988. MR 89f:32040 Zbl 0667.30040

[Wolf 1968] J. A. Wolf, "Growth of finitely generated solvable groups and curvature of Riemanniann manifolds", *J. Differential Geometry* **2** (1968), 421–446. MR 40 #1939 Zbl 0207.51803

Received May 30, 2010.

KATSUHIKO MATSUZAKI
DEPARTMENT OF MATHEMATICS, SCHOOL OF EDUCATION
WASEDA UNIVERSITY
1-6-1 NISHI-WASEDA, SHINJUKU
TOKYO 169-8050
JAPAN
matsuzak@waseda.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

ON ZERO-DIVISOR GRAPHS OF BOOLEAN RINGS

ALI MOHAMMADIAN

The zero-divisor graph of a ring R is the graph whose vertices consist of the nonzero zero-divisors of R in which two distinct vertices a and b are adjacent if and only if either $ab = 0$ or $ba = 0$. In this paper, we investigate some properties of zero-divisor graphs of Boolean rings. Among other results, we prove that for any two rings R and S with $\Gamma(R) \simeq \Gamma(S)$, if R is Boolean and $|R| > 4$, then $R \simeq S$.

1. Introduction

Throughout the paper, R denotes a ring, not necessarily with identity, and $\mathcal{D}(R)$ denotes the set of all zero-divisors of R. If X is either an element or a subset of R, then the *left annihilator* of X is $\mathrm{Ann}_\ell(X) = \{a \in R \mid aX = 0\}$ and the right annihilator of X, denoted by $\mathrm{Ann}_r(X)$, is similarly defined. For any subset Y of R, we let $Y^* = Y \setminus \{0\}$. The *zero-divisor graph* of R, denoted by $\Gamma(R)$, is a graph with the vertex set $\mathcal{D}(R)^*$ such that two vertices x and y are joined by an undirected edge if and only if $x \neq y$ and either $xy = 0$ or $yx = 0$. Notice that a ring R is a domain if and only if $\Gamma(R)$ is the null graph. For a commutative ring R with identity, the definition of a zero-divisor graph of R that was first introduced in [Beck 1988] coincides with the above definition of $\Gamma(R)$. The zero-divisor graph concept for noncommutative rings was first defined in [Redmond 2002]. The zero-divisor graphs offer a graphical representation of rings so that we may discover some new algebraic properties of rings that are hidden from the viewpoint of classical ring theorists. For an instance, using the notion of a zero-divisor graph, it has been proven in [Redmond 2004] that for any finite ring R, $\sum_{x\in R} |\mathrm{Ann}_\ell(x) \setminus \mathrm{Ann}_r(x)|$ is even. A simple proof of this result is given in [Akbari and Mohammadian 2007].

Let us recall some definitions regarding graph theory and ring theory. For a vertex v of a graph G, $\mathcal{N}(v)$ denotes the set of all vertices of G adjacent to v, and the *degree* of v is defined by $|\mathcal{N}(v)|$. A graph G is called a *star* if G contains at least two vertices and there exists a vertex that is joined to all other vertices

This research was in part supported by a grant from IPM.
MSC2000: 05C25, 06E20, 16P10.
Keywords: Boolean ring, reduced ring, zero-divisor graph.

and G has no other edges. A *path* $\mathcal{P}$ in a graph G is a sequence of distinct vertices $v_1, v_2, \ldots, v_{k+1}$ in which every two consecutive vertices are adjacent. The number k is called the *length* of $\mathcal{P}$. For two vertices u and v in a graph G, the *distance* between u and v, denoted by $d(u, v)$, is the length of the shortest path between u and v, if such a path exists; otherwise, we define $d(u, v) = \infty$. The *diameter* of a graph G is defined by $\operatorname{diam} G = \sup\{d(u, v) \mid u \text{ and } v \text{ are distinct vertices of } G\}$. In [Redmond 2002] it was shown that for any ring R, $\operatorname{diam} \Gamma(R) \leqslant 3$. Furthermore, two graphs G_1 and G_2 are said to be *isomorphic* if there is a bijective map φ between the vertex set of G_1 and the vertex set of G_2 such that the adjacency relation is preserved. Finally, we recall that a ring is called *reduced* if it has no nonzero nilpotent elements. A ring whose elements are all idempotent is called *Boolean*. We denote by $\mathbb{Z}_n$ the ring of integers modulo n and by $\mathbb{F}_q$ the field with q elements.

In this article we study the zero-divisor graphs of Boolean rings. We show that for any reduced ring R that is not a domain, $\Gamma(R)$ is isomorphic to the zero-divisor graph of a nonreduced ring, provided that $\Gamma(R)$ is a star. As a consequence, we prove that Boolean rings with more than four elements are determined by their zero-divisor graphs.

2. The results

In [Akbari and Mohammadian 2006, Theorem 17], it is proven that for any finite ring R that is not a field, if $\Gamma(R)$ is isomorphic to the zero-divisor graph of a reduced ring S, then $R \simeq S$, unless $S \simeq \mathbb{Z}_2 \times \mathbb{F}_q$, where either $q = 2$ or $(q+1)/2$ is a prime power. Since for any finite field F, $\Gamma(\mathbb{Z}_2 \times F)$ is a star, the following theorem presents an analogue of this result for the general case.

Remark 1. Let $\{\mathcal{A}_i\}_{i \in I}$ and $\{\mathcal{B}_j\}_{j \in J}$ be two families of commutative domains with identity, where $|I| \geqslant 2$. In [Anderson et al. 2003, Theorem 2.1], it is shown that $\Gamma\big(\prod_{i \in I} \mathcal{A}_i\big) \simeq \Gamma\big(\prod_{j \in J} \mathcal{B}_j\big)$ if and only if there is a bijective map $\pi : I \to J$ such that $|\mathcal{A}_i| = |\mathcal{B}_{\pi(i)}|$ for all $i \in I$. Hence there are many examples of nonisomorphic pairs of infinite reduced commutative rings whose zero-divisor graphs are isomorphic.

Theorem 2. *Let S be a reduced ring such that S is not a domain and $\Gamma(S)$ is not a star. If R is a ring such that $\Gamma(R) \simeq \Gamma(S)$, then R is also a reduced ring.*

Proof. We recall a well-known fact about reduced rings: for all elements x and y of a reduced ring T, $xy = 0$ if and only if $yx = 0$. For this, note that if $xy = 0$ for some elements $x, y \in T$, then $(yx)^2 = 0$ and since T is reduced, we find that $yx = 0$. This fact implies that if two vertices u and v of $\Gamma(S)$ are adjacent, then $uv = vu = 0$. We use this property frequently in what follows. We also state two properties of $\Gamma(S)$:

(i) For every two adjacent vertices u and v of $\Gamma(S)$ with at least one common neighbor, $u+v$ is a vertex of $\Gamma(S)$ and $\mathcal{N}(u+v)=\mathcal{N}(u)\cap\mathcal{N}(v)$. For this, note that if $x\in\mathcal{N}(u)\cap\mathcal{N}(v)$, then $xu=xv=0$, and hence $x(u+v)=0$. Also, $u+v\neq 0$ since $uv=0$ and S is reduced. Therefore, $x\in\mathcal{N}(u+v)$. Conversely, if $x\in\mathcal{N}(u+v)$, then $(xu)u=x(u+v)u=0$ and thus $u(xu)=0$. Therefore $(xu)^2=0$ and so $xu=0$. This means that $x\in\mathcal{N}(u)$ and with a similar argument, we find that $x\in\mathcal{N}(v)$, as required.

(ii) For every three mutually adjacent vertices u, v and w of $\Gamma(S)$, we have $\mathcal{N}(u)\nsubseteq\mathcal{N}(v)\cup\mathcal{N}(w)$. Indeed, it easily seen that $v+w\in\mathcal{N}(u)\setminus(\mathcal{N}(v)\cup\mathcal{N}(w))$.

Suppose that R is a ring with $\Gamma(R)\simeq\Gamma(S)$. So properties (i) and (ii) also hold for $\Gamma(R)$. To the contrary, assume that $a^2=0$ for some element $a\in R^*$.

Since S is reduced, [Akbari and Mohammadian 2006, Corollary 4] yields that $\Gamma(R)$ has at least two vertices. Note that a is not adjacent to all other vertices of $\Gamma(R)$. To prove this, suppose otherwise. Since $\Gamma(R)$ is not a star, there exist two adjacent vertices $x, y\in\mathcal{N}(a)$. So $\mathcal{N}(x)\subseteq\mathcal{N}(a)\cup\mathcal{N}(y)$, which contradicts (ii). Moreover, we have $|\mathcal{N}(a)|\geqslant 2$. For this, suppose otherwise. Since $\Gamma(R)$ is a connected graph [Redmond 2002] with at least two vertices, we may assume that $\mathcal{N}(a)=\{b\}$ for some vertex b of $\Gamma(R)$. From $a+b\in\mathrm{Ann}_\ell(a)\cup\mathrm{Ann}_r(a)$, we conclude that $a+b=0$. Hence $b=-a$ and therefore $\Gamma(R)$ is a star on two vertices, a contradiction.

We claim that either $Ra=\{0,a\}$ or $aR=\{0,a\}$. Suppose that there exist two elements $b\in Ra\setminus\{0,a\}$ and $c\in aR\setminus\{0,a\}$. If $b\neq c$, then a, b and c are three mutually adjacent vertices and $\mathcal{N}(a)\subseteq\mathcal{N}(b)\cup\mathcal{N}(c)$, which contradicts (ii). Hence $b=c$. For some vertex $d\in\mathcal{N}(a)\setminus\{b\}$, the vertices a, b and d are mutually adjacent and $\mathcal{N}(a)\subseteq\mathcal{N}(b)\cup\mathcal{N}(d)$, which again contradicts (ii). Since $Ra\neq\{0\}$ and $aR\neq\{0\}$, the claim is proved.

We assume that $Ra=aR=\{0,a\}$. For any two vertices $x, y\notin\mathcal{N}(a)$, we have $xa=ya=a$. Thus $(xy)a=a$ and so $xy\neq 0$. This means that every edge of $\Gamma(R)$ has at least one endpoint in $\mathcal{N}(a)$. Working towards a contradiction, assume that no two vertices in $\mathcal{N}(a)$ are adjacent. This means that $\Gamma(R)$, and so $\Gamma(S)$ is a bipartite graph, and using [Akbari et al. 2003, Theorem 2.4], $\Gamma(S)$ and thus $\Gamma(R)$ is a complete bipartite graph. Let $r\notin\mathcal{N}(a)$ and $s\in\mathcal{N}(a)\cap\mathcal{N}(r)$. Since $\Gamma(R)$ is a complete bipartite graph and $a+s\in\mathcal{N}(a)$, r is adjacent to $a+s$. Therefore $a=r(a+s)r=0$, a contradiction. Hence there are two adjacent vertices $b, c\in\mathcal{N}(a)$. We now consider the two following cases.

Case I. Suppose that a together with one of the elements b, c are contained in one of the one-sided annihilators of the third element. Without loss of generality, assume that $\{a,c\}\in\mathrm{Ann}_\ell(b)$. By (i), there exists a vertex d in $\Gamma(R)$ such that $d\notin\{a,b\}$ and $\mathcal{N}(d)=\mathcal{N}(a)\cap\mathcal{N}(b)$. If $b\neq a+c$, then $a+c\in\mathcal{N}(a)\cap\mathcal{N}(b)$, and hence $a = d(a+c)d = 0$, a contradiction. Thus $b = a+c$, and it follows from $ab = 0$ that $ac = 0$. Moreover, if $ca \neq 0$, then $a = ca = cb - c^2 = -c^2$, which contradicts $d \in \mathcal{N}(c) \setminus \mathcal{N}(a)$. Therefore $ca = 0$ and so $c^2 = cb - ca = 0$. Since $b = a + c$, we find that the product of any two elements of $\{a, b, c\}$ is zero.

Suppose towards a contradiction that there is a vertex $r \in (\mathcal{N}(b) \cap \mathcal{N}(c)) \setminus \{a\}$. We have $rar = r(b-c)r = 0$ and by $Ra = aR = \{0, a\}$, we deduce that $r \in \mathcal{N}(a)$. By (i), there exists a vertex s in $\Gamma(R)$ such that $\mathcal{N}(s) = \mathcal{N}(a) \cap \mathcal{N}(r)$. This implies that $sas = s(b-c)s = 0$, since $\{b, c\} \subseteq \mathcal{N}(a) \cap \mathcal{N}(r)$. On the other hand, $s \notin \mathcal{N}(a)$ and $Ra = aR = \{0, a\}$ yields that $sas = a$, a contradiction. This establishes that $\mathcal{N}(b) \cap \mathcal{N}(c) = \{a\}$.

For convenience and without loss of generality, assume that $cd = 0$. From $\{b, c\} \subseteq \operatorname{Ann}_r(c)$, $d \notin \mathcal{N}(a) \cup \mathcal{N}(b)$ and $\mathcal{N}(b) \cap \mathcal{N}(c) = \{a\}$, we have $Rc = \{0, c\}$. Therefore $[R : \operatorname{Ann}_\ell(a) \cap \operatorname{Ann}_\ell(c)] \leqslant [R : \operatorname{Ann}_\ell(a)][R : \operatorname{Ann}_\ell(c)] = |Ra||Rc| = 4$. Since $\mathcal{N}(b) \cap \mathcal{N}(c) = \{a\}$ and the product of any two elements of $\{a, b, c\}$ is zero, we find that $\operatorname{Ann}_\ell(a) \cap \operatorname{Ann}_\ell(c) = \{0, a, b, c\}$. This yields that $|R| \leqslant 16$. Using (i), let e be a vertex of $\Gamma(R)$ in which $\mathcal{N}(e) = \mathcal{N}(a) \cap \mathcal{N}(c)$. It is not hard to see that

$$R = \{0, a, b, c\} \cup \bigl(d + \{0, a, b, c\}\bigr) \cup \bigl(e + \{0, a, b, c\}\bigr) \cup \bigl(d + e + \{0, a, b, c\}\bigr).$$

Therefore $\operatorname{Ann}_\ell(a) = \operatorname{Ann}_r(a) = \{0, a, b, c\} \cup (d + e + \{0, a, b, c\})$. Because $e \notin \mathcal{N}(a) \cup \mathcal{N}(c)$, $Ra = \{0, a\}$ and $Rc = \{0, c\}$, we conclude that $ea = a$ and $ec = c$. Therefore $eb = b$ and by $b \in \mathcal{N}(a) \cap \mathcal{N}(c)$, we obtain that $be = 0$. Furthermore, $e \notin \mathcal{N}(a) \cup \mathcal{N}(c)$ and $\mathcal{N}(b) \cap \mathcal{N}(c) = \{a\}$ yield that $Rb = \{0, b\}$. It follows from $d \notin \mathcal{N}(b)$ that $d + e \in \operatorname{Ann}_\ell(b)$, and so $\mathcal{N}(a) \subseteq \mathcal{N}(b) \cup \mathcal{N}(c)$, which contradicts (ii).

Case II. When Case I does not occur, by replacing b with c if necessary, we may assume that $ab = bc = ca = 0$ and none of ba, cb and ac is zero. We have $\{a, b\} \in \operatorname{Ann}_\ell(cb)$, and so, applying the argument in the first paragraph of Case I for cb and b instead of b and c, respectively, we obtain in particular that $ba = 0$, which is a contradiction.

Next, with no loss of generality, assume that $aR = \{0, a\}$ and there exists an element $g \in Ra \setminus \{0, a\}$. Since $aR = \{0, a\}$, $-a = a$ and so $-g = g$. Also, from $g \in Ra$ and $aR = \{0, a\}$, we easily obtain that $ag = ga = 0$. By (i), there exists a vertex h in $\Gamma(R)$ such that $\mathcal{N}(h) = \mathcal{N}(a) \cap \mathcal{N}(g)$. We claim that $\operatorname{Ann}_r(a) \subseteq \operatorname{Ann}_r(h) \cup \{0, a, g, a+g\}$. Suppose $x \in \operatorname{Ann}_r(a) \setminus (\operatorname{Ann}_r(h) \cup \{a, g\})$. Since $g \in Ra$ and $\mathcal{N}(h) = \mathcal{N}(a) \cap \mathcal{N}(g)$, we conclude that $x \in \operatorname{Ann}_\ell(h)$. Moreover, $h \notin \mathcal{N}(a)$ and $aR = \{0, a\}$, so $ah = a$. We have $a + g \in \mathcal{N}(a) \cap \mathcal{N}(g)$ and $(a+g)h = a + g$, and hence $h(a+g) = 0$. These equalities yield that $h(a+g+x) = hx \neq 0$ and $(a+g+x)h = a+g \neq 0$. On the other hand, $a(a+g+x) = 0$, so it follows from $g \in Ra$ and $\mathcal{N}(h) = \mathcal{N}(a) \cap \mathcal{N}(g)$ that $a + g + x = 0$. Therefore, $x = a + g$, and the claim is proved. Since $\operatorname{Ann}_r(a)$, $\operatorname{Ann}_r(h)$ and $\{0, a, g, a+g\}$

are three additive subgroups of R in which $\mathrm{Ann}_r(a) \subseteq \mathrm{Ann}_r(h) \cup \{0, a, g, a+g\}$ and $a \in \mathrm{Ann}_r(a) \setminus \mathrm{Ann}_r(h)$, we deduce that $\mathrm{Ann}_r(a) = \{0, a, g, a+g\}$. Applying (ii), there exists a vertex $y \in \mathcal{N}(a) \setminus (\mathcal{N}(g) \cup \mathcal{N}(a+g))$. We have $ya = 0$ and $ay \neq 0$. By $[R : \mathrm{Ann}_r(a)] = |aR|$, we conclude that $R = \mathrm{Ann}_r(a) \cup y + \mathrm{Ann}_r(a)$. It follows from $\mathrm{Ann}_r(a) = \{0, a, g, a+g\}$ that $Ra = \{0\}$, a contradiction. Now the proof is complete. □

Example 3. The condition on $\Gamma(S)$ in Theorem 2 is necessary. For examples involving infinite rings, let $\mathcal{S}$ be an arbitrary infinite domain, $\mathcal{R}$ be the polynomial ring in the set of variables $\{\mathrm{x}\} \cup \{\mathrm{x}_\alpha \mid \alpha \in \mathcal{S}\}$ with coefficients in $\mathbb{Z}_2$, and $\mathcal{I}$ be the ideal of $\mathcal{R}$ generated by $\{\mathrm{x}^2\} \cup \{\mathrm{x}\mathrm{x}_\alpha - \mathrm{x} \mid \alpha \in \mathcal{S}\} \cup \{\mathrm{x}_\alpha \mathrm{x} - \mathrm{x} \mid \alpha \in \mathcal{S}\}$. It is easy to verify that $\Gamma(\mathcal{R}/\mathcal{I})$ is a star on $|\mathcal{S}|$ vertices and $\mathrm{x} + \mathcal{I}$ is that vertex which is adjacent to all other vertices of the graph. Therefore $\Gamma(\mathcal{R}/\mathcal{I}) \simeq \Gamma(\mathbb{Z}_2 \times \mathcal{S})$, while $\mathcal{R}/\mathcal{I}$ is not reduced.

Remark 4. It is easy to establish that every reduced ring whose zero-divisor graph is a star is isomorphic to the direct product of $\mathbb{Z}_2$ and a domain. For this, let R be a reduced ring with $\Gamma(R)$ a star and let e be that vertex which is adjacent to all other vertices of $\Gamma(R)$. Obviously, e is idempotent, and using the fact that all idempotent elements of a reduced ring are central, we may write $R \simeq eR \times (1-e)R$. Since $\Gamma(R)$ is a star, we clearly conclude that $eR = \{0, e\}$ and $(1-e)R$ is a domain, as required. From this, Theorem 2, Example 3, and [Akbari and Mohammadian 2006, Theorem 17], we imply that for every reduced ring R that is not a domain, $\Gamma(R)$ is isomorphic to the zero-divisor graph of a nonreduced ring if and only if $\Gamma(R)$ is either an infinite star or a star with q vertices, where either $q = 2$ or both q and $(q+1)/2$ are prime powers.

In [LaGrange 2007, Theorem 4.1], it is shown that if R and S are two commutative rings with identity such that S is a Boolean ring with more than four elements and $\Gamma(R) \simeq \Gamma(S)$, then $R \simeq S$. In what follows, we generalize this result to every arbitrary ring R. We need the following easy lemmas.

Lemma 5. *Let R be a ring such that all elements in $\mathcal{D}(R)$ are idempotent. Then R is either a domain or a Boolean ring.*

Proof. Suppose that R is not a domain. By the hypotheses, R is reduced. Using the fact that all idempotent elements of a reduced ring are central, $\mathcal{D}(R)$ is contained in the center of R. Therefore, for every two elements $a \in R$ and $z \in \mathcal{D}(R)^*$, we have $az \in \mathcal{D}(R)$. Hence $(az)^2 = az$, and so $(a^2 - a)z = 0$. The latter equality shows that $a^2 - a \in \mathcal{D}(R)$ and also $\mathrm{Ann}_\ell(a^2 - a) = \mathcal{D}(R)$. Thus $a^2 - a = (a^2 - a)^2 = 0$ for each element $a \in R$, as desired. □

Lemma 6. *Let R be a Boolean ring with $|R| > 4$. Then $\Gamma(R)$ contains no vertex adjacent to all other vertices of the graph.*

Proof by contradiction. Suppose that a vertex r is adjacent to all other vertices of $\Gamma(R)$. Let $z \in \mathcal{D}(R) \setminus \{0, r\}$. We have $r(r+z) = r \neq 0$ and so $r + z$ is a nonzero-divisor idempotent of R. Thus $1 = r + z$ is the identity of R and so $R = \{0, 1, r, 1-r\}$, which contradicts $|R| > 4$. $\square$

Theorem 7. *Let S be a Boolean ring with $|S| > 4$. Suppose that R is a ring and $\varphi : \Gamma(R) \to \Gamma(S)$ is a graph isomorphism. Then φ is extendable to a ring isomorphism from R to S. In particular, $R \simeq S$.*

Proof. Recall that the characteristic of every Boolean ring is 2. We first state the following properties of $\Gamma(S)$.

(i) For every two vertices u and v of $\Gamma(S)$, if $\mathcal{N}(u) = \mathcal{N}(v)$, then $u = v$. For this, note that if $u \neq uv$, then $u + uv \in \mathcal{N}(v) \setminus \mathcal{N}(u)$, which is impossible. So we conclude that $u = uv$, and similarly $v = uv$, which yield that $u = v$, as desired.

(ii) For every two adjacent vertices u and v of $\Gamma(S)$, using (i) together with an easy argument, we find that $u + v$ is the unique vertex of $\Gamma(S)$ such that $\mathcal{N}(u+v) = \mathcal{N}(u) \cap \mathcal{N}(v)$, if $\mathcal{N}(u) \cap \mathcal{N}(v) \neq \varnothing$; and otherwise, $1 = u + v$ is the identity of S, because in this case $u + v$ is a nonzero-divisor idempotent of S. Moreover, if S has identity, then $v = 1 + u$ is the unique neighbor of u in $\Gamma(S)$ such that $\mathcal{N}(u) \cap \mathcal{N}(v) = \varnothing$. For uniqueness, note that for any vertex $x \in \mathcal{N}(u)$, if $x \neq 1 + u$, then $1 + u + x \in \mathcal{N}(u) \cap \mathcal{N}(x)$.

(iii) For every two nonadjacent vertices u and v of $\Gamma(S)$, $\mathcal{N}(u) \cup \mathcal{N}(v) \subseteq \mathcal{N}(uv)$; and if $\mathcal{N}(u) \cup \mathcal{N}(v) \subseteq \mathcal{N}(w)$ for some vertex w of $\Gamma(S)$, then $\mathcal{N}(uv) \subseteq \mathcal{N}(w)$. For the second statement, let $x \in \mathrm{Ann}_\ell(uv)$. We have $vx \in \mathrm{Ann}_\ell(u)$. Since $\mathcal{N}(u) \subseteq \mathcal{N}(w)$, $w(vx) = 0$ and so $wx \in \mathrm{Ann}_\ell(v)$. It follows from $\mathcal{N}(v) \subseteq \mathcal{N}(w)$ that $w(wx) = 0$ and thus $x \in \mathrm{Ann}_\ell(w)$, as required.

Since $\Gamma(R) \simeq \Gamma(S)$, the above properties also hold for $\Gamma(R)$. Using Theorem 2 and Lemma 6, R is reduced. It is easily checked that $\mathcal{N}(z^2) = \mathcal{N}(z)$ for each vertex z of $\Gamma(R)$. By (i), we have $z^2 = z$ for every element $z \in \mathcal{D}(R)$. Applying Lemma 5, R is a Boolean ring. Define $\varphi(0) = 0$. By (ii), $\Gamma(R)$ (respectively, $\Gamma(S)$) contains two adjacent vertices with no common neighbors if and only if R (respectively, S) has identity. Since $\Gamma(R) \simeq \Gamma(S)$, either both R and S have identity or neither of them has identity. When the first case occurs, we define $\varphi(1) = 1$. Furthermore, the properties (ii) and (iii) imply that for every two vertices u and v of $\Gamma(S)$, the elements

- $1 + u$, if S has identity;
- $u + v$, if u and v are adjacent and $u + v \neq 1$; and
- uv, if u and v are not adjacent,

can be determined by $\Gamma(S)$. We claim that for every two distinct nonadjacent vertices u and v of $\Gamma(S)$, the element $u+v$ can also be determined by $\Gamma(S)$. First assume that $uv \notin \{u, v\}$. Using (iii), we obtain that the element uv is determined by $\Gamma(S)$. By (i) and (ii), $u+uv$ is the unique vertex of $\Gamma(S)$ such that $\mathcal{N}(u) = \mathcal{N}(uv) \cap \mathcal{N}(u+uv)$. This and a similar argument establish that the elements $u+uv$ and $v+uv$ are determined by $\Gamma(S)$. Since the vertices $u+uv$ and $v+uv$ are adjacent, we are done using (ii). Next, with no loss of generality, suppose that $uv = u$. In this case, the vertices u and $u+v$ are adjacent, and so applying (i) and (ii), we find that $u+v$ is the unique vertex of $\Gamma(S)$ such that $\mathcal{N}(v) = \mathcal{N}(u) \cap \mathcal{N}(u+v)$. This proves the claim. Now, by $\Gamma(R) \simeq \Gamma(S)$ and the above reasonings, it is not hard to verify that $\varphi(a+b) = \varphi(a) + \varphi(b)$ and $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$, as desired. □

As an interesting fact, it is well-known that every isomorphism between multiplicative semigroups of two Boolean rings is a ring isomorphism. Obviously, Theorem 7 generalizes this fact. The following theorem asserts that the zero-divisor graph of a Boolean ring R determines whether R has identity or not.

Theorem 8. *Let R be a Boolean ring and $|R| > 4$. Then* $\operatorname{diam}\Gamma(R) = 3$ *if R has identity, and otherwise* $\operatorname{diam}\Gamma(R) = 2$.

Proof. We know from [Redmond 2002] that for any ring T, $\operatorname{diam}\Gamma(T) \leqslant 3$. First suppose that R has identity. Since $|R| > 4$, we can take an element $e \notin \{0, 1\}$. We have $R = eR \oplus (1-e)R$, so either $|eR| > 2$ or $|(1-e)R| > 2$. With no loss of generality, let $f \in eR \setminus \{0, e\}$. Since e and $1+e+f$ are two nonadjacent vertices with no common neighbors and $\operatorname{diam}\Gamma(R) \leqslant 3$, the result follows.

Next suppose that R has no identity. Applying Lemma 6, we find $\operatorname{diam}\Gamma(R) \geqslant 2$. Now, let a and b be two nonadjacent vertices of $\Gamma(R)$. Since R has no identity, there exists an element c such that $(a+b+ab)c \neq c$. We have $c+ac+bc+abc \in \mathcal{N}(a) \cap \mathcal{N}(b)$, which clearly completes the proof. □

It is well-known that every finite Boolean ring has identity. We generalize this fact in the following theorem.

Theorem 9. *Let R be a Boolean ring such that $\Gamma(R)$ has a vertex of finite degree. Then R has identity.*

Proof. Recall that the adjoint multiplication $\circ$ of an arbitrary ring T is defined by $x \circ y = x + y + xy$ for any two elements $x, y \in T$. Suppose that a is a vertex of finite degree of $\Gamma(R)$ and $\mathcal{N}(a) = \{a_1, \ldots, a_n\}$ for some integer $n \geqslant 1$. Let $b = a_1 \circ \cdots \circ a_n$. Clearly, $ab = 0$ and $a_i b = a_i$ for all i. We show that $a+b$ is the identity of R. Indeed, it is enough to prove that $a+b$ is a nonzero-divisor. Toward a contradiction, assume that $(a+b)z = 0$ for some element $z \in R^*$. Multiplying this equality by a, we find that $az = 0$, and hence $z = a_j$ for some $j \in \{1, \ldots, n\}$.

Also, multiplying the equality $(a+b)z = 0$ by a_j yields that $a_j z = 0$, which is impossible. This completes the proof. $\square$

Remark 10. The converse of Theorem 9 is not true. Let $\mathcal{R}$ be the set consisting of the empty set together with all finite unions of all left-closed right-open intervals and all left-unbounded right-open intervals of real numbers. Clearly, $\mathcal{R}$ is a Boolean ring with identity with respect to symmetric difference as the addition operation and intersection as the multiplication operation, while obviously every vertex of $\Gamma(\mathcal{R})$ has infinite degree.

We conclude the paper with the following theorem on the polynomial rings over Boolean rings.

Theorem 11. *Let R and S be two Boolean rings such that $\Gamma(R[\mathrm{x}]) \simeq \Gamma(S[\mathrm{x}])$. Then $R \simeq S$.*

Proof. Let T be an arbitrary Boolean ring. $\Gamma(T[\mathrm{x}])$ is the null graph if and only if $T \simeq \mathbb{Z}_2$. Hence we may assume that $\mathcal{D}(R)^*$ and $\mathcal{D}(S)^*$ are both nonempty. Using Theorem 7, it suffices to establish that $\Gamma(R) \simeq \Gamma(S)$. Since finitely generated one-sided ideals of von Neumann regular rings, including Boolean rings, are principal [Lam 2001, (4.23)], for each finitely generated ideal I of T, there exists a unique element e such that $I = (e)$. For a polynomial $f(\mathrm{x}) = a_n\mathrm{x}^n + \cdots + a_0 \in T[\mathrm{x}]$, let $\widehat{f(\mathrm{x})}$ be the unique element of T such that $(a_0, \ldots, a_n) = (\widehat{f(\mathrm{x})})$. From [Armendariz 1974, Lemma 1], every reduced ring is Armendariz, and hence it is not hard to see that for any polynomial $f(\mathrm{x}) \in \mathcal{D}(T[\mathrm{x}])^*$, $\widehat{f(\mathrm{x})}$ is the unique element of T such that $\mathcal{N}(f(\mathrm{x})) = \mathcal{N}(\widehat{f(\mathrm{x})})$.

Now, assume that $\phi : \Gamma(R[\mathrm{x}]) \to \Gamma(S[\mathrm{x}])$ is a graph isomorphism. We define $\psi : \Gamma(R) \to \Gamma(S)$ by $\psi(a) = \widehat{\phi(a)}$ for all $a \in \mathcal{D}(R)^*$, and we claim that ψ is a graph isomorphism. If a and b are two adjacent vertices of $\Gamma(R)$, then $\phi(a) \in \mathcal{N}(\phi(b)) = \mathcal{N}(\psi(b))$. This yields that $\psi(b) \in \mathcal{N}(\phi(a)) = \mathcal{N}(\psi(a))$ and therefore $\psi(a)$ and $\psi(b)$ are adjacent in $\Gamma(S)$. The converse is clearly true, and so ψ preserves the adjacency relation. Moreover, if $\psi(a) = \psi(b)$ for two vertices a and b of $\Gamma(R)$, then $\mathcal{N}(\phi(a)) = \mathcal{N}(\phi(b))$ and thus $\mathcal{N}(a) = \mathcal{N}(b)$. In particular, $\mathcal{N}(a) \cap R = \mathcal{N}(b) \cap R$. Using the property (i) of the zero-divisor graphs of Boolean rings given in the proof of Theorem 7, we deduce that $a = b$. This concludes the injectivity of ψ. Finally, we prove that ψ is surjective. Suppose $s \in \mathcal{D}(S)^*$ and let

$$r = \widehat{\phi^{-1}(s)}.$$

Since $\mathcal{N}(\phi^{-1}(s)) = \mathcal{N}(r)$, we find that $\mathcal{N}(s) = \mathcal{N}(\phi(r)) = \mathcal{N}(\psi(r))$ and hence $s = \psi(r)$. This establishes the claim and completes the proof. $\square$

Remark 12. Let $n \geqslant 2$ and $\mathcal{R}$ and $\mathcal{S}$ be two rings which each of them is the direct product of n arbitrary finite fields. Using the result mentioned in Remark 1, it is easily checked that $\Gamma(\mathscr{R}[x]) \simeq \Gamma(\mathscr{S}[x])$. Therefore the conclusion of Theorem 11 is not true if one of R and S is not Boolean.

References

[Akbari and Mohammadian 2006] S. Akbari and A. Mohammadian, "Zero-divisor graphs of non-commutative rings", *J. Algebra* **296**:2 (2006), 462–479. MR 2007a:16046 Zbl 1113.16038

[Akbari and Mohammadian 2007] S. Akbari and A. Mohammadian, "On zero-divisor graphs of finite rings", *J. Algebra* **314**:1 (2007), 168–184. MR 2008f:16048 Zbl 1122.16015

[Akbari et al. 2003] S. Akbari, H. R. Maimani, and S. Yassemi, "When a zero-divisor graph is planar or a complete r-partite graph", *J. Algebra* **270**:1 (2003), 169–180. MR 2004h:13025 Zbl 1032.13014

[Anderson et al. 2003] D. F. Anderson, R. Levy, and J. Shapiro, "Zero-divisor graphs, von Neumann regular rings, and Boolean algebras", *J. Pure Appl. Algebra* **180**:3 (2003), 221–241. MR 2003m:13007 Zbl 1076.13001

[Armendariz 1974] E. P. Armendariz, "A note on extensions of Baer and P.P.-rings", *J. Austral. Math. Soc.* **18** (1974), 470–473. MR 51 #3224 Zbl 0292.16009

[Beck 1988] I. Beck, "Coloring of commutative rings", *J. Algebra* **116**:1 (1988), 208–226. MR 89i:13006 Zbl 0654.13001

[LaGrange 2007] J. D. LaGrange, "Complemented zero-divisor graphs and Boolean rings", *J. Algebra* **315**:2 (2007), 600–611. MR 2008h:13012 Zbl 1133.13005

[Lam 2001] T. Y. Lam, *A first course in noncommutative rings*, 2nd ed., Graduate Texts in Mathematics **131**, Springer, New York, 2001. MR 2002c:16001 Zbl 0980.16001

[Redmond 2002] S. P. Redmond, "The zero-divisor graph of a non-commutative ring", pp. 39–47 in *Commutative rings*, edited by A. Badawi, Nova Sci. Publ., Hauppauge, NY, 2002. MR 2037657 Zbl 1195.16038

[Redmond 2004] S. P. Redmond, "Structure in the zero-divisor graph of a noncommutative ring", *Houston J. Math.* **30**:2 (2004), 345–355. MR 2005d:16002 Zbl 1064.16033

Received June 5, 2010. Revised August 31, 2010.

ALI MOHAMMADIAN
SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM)
NIAVARAN SQUARE
P.O. BOX 19395-5746
TEHRAN
IRAN
ali_m@ipm.ir

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

RATIONAL CERTIFICATES OF POSITIVITY ON COMPACT SEMIALGEBRAIC SETS

VICTORIA POWERS

Let $\mathbb{R}[X]$ denote the real polynomial ring $\mathbb{R}[X_1, \dots, X_n]$ and write $\sum \mathbb{R}[X]^2$ for the set of sums of squares in $\mathbb{R}[X]$. Given $g_1, \dots, g_s \in \mathbb{R}[X]$ such that the semialgebraic set $K := \{x \in \mathbb{R}^n \mid g_i(x) \geq 0 \text{ for all } i\}$ is compact, Schmüdgen's theorem says that if $f \in \mathbb{R}[X]$ such that $f > 0$ on K, then f is in the preordering in $\mathbb{R}[X]$ generated by the g_i's, i.e., f can be written as a finite sum of elements $\sigma g_1^{e_1} \dots g_s^{e_s}$, where σ is a sum of squares in $\mathbb{R}[X]$ and each $e_i \in \{0, 1\}$. Putinar's theorem says that under a condition on the set of generators $\{g_1, \dots, g_s\}$ (which is a stronger condition than the compactness of K), any $f > 0$ on K can be written $f = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s$, where $\sigma_i \in \sum \mathbb{R}[X]^2$. Both of these theorems can be viewed as statements about the existence of certificates of positivity on compact semialgebraic sets. In this note we show that if the defining polynomials $g_1, \dots, g_s$ and polynomial f have coefficients in $\mathbb{Q}$, then in Schmüdgen's theorem we can find a representation in which the σ's are sums of squares of polynomials over $\mathbb{Q}$. We prove a similar result for Putinar's theorem assuming that the set of generators contains $N - \sum X_i^2$ for some $N \in \mathbb{N}$.

1. Introduction

We write $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{Q}$ for the set of natural, real, and rational numbers. Let $n \in \mathbb{N}$ be fixed and let $\mathbb{R}[X]$ denote the polynomial ring $\mathbb{R}[X_1, \dots, X_n]$. We denote by $\sum \mathbb{R}[X]^2$ the set of sums of squares in $\mathbb{R}[X]$.

For $S = \{g_1, \dots, g_s\} \subseteq \mathbb{R}[X]$, the *basic closed semialgebraic set* generated by S, denoted K_S, is

$$\{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \dots, g_s(x) \geq 0\}.$$

Associated to S are two algebraic objects: The *quadratic module generated by S*, denoted M_S, is the set of $f \in \mathbb{R}[X]$ which can be written

$$f = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s,$$

MSC2010: primary 11E25, 12D15, 13J30, 14P10; secondary 14Q20.
Keywords: rational sums of squares, certificates of positivity, Schmüdgen's theorem, Putinar's theorem.

where each σ_i lies in $\sum \mathbb{R}[X]^2$, and the *preordering generated by* S, denoted T_S, is the quadratic module generated by all products of elements in S. In other words, T_S is the set of $f \in \mathbb{R}[X]$ which can be written as a finite sum of elements

$$\sigma g_1^{e_1} \dots g_s^{e_s}, \quad \text{for } \sigma \in \mathbb{R}[X] \text{ and each } e_i \in \{0, 1\}.$$

A polynomial $f \in \sum \mathbb{R}[X]^2$ is obviously globally nonnegative in $\mathbb{R}^n$ and writing f explicitly as a sum of squares gives a "certificate of positivity" for the fact that f takes only nonnegative values in $\mathbb{R}^n$. (Note: To avoid having to write "nonnegativity or positivity" we use the term "positivity" to mean either.) More generally, for a basic closed semialgebraic set K_S, if $f \in T_S$ or $f \in M_S$, then f is nonnegative on K_S and an explicit representation of f in M_S or T_S gives a certificate of positivity for f on K_S.

Schmüdgen [1991] showed that if the semialgebraic set K_S is compact, then any $f \in \mathbb{R}[X]$ which is strictly positive on K_S is in the preordering T_S. A preordering or quadratic module is *archimedean* if it contains $N - \sum X_i^2$ for some $N \in \mathbb{N}$. We note that if M_S is archimedean, then it follows immediately that K_S is compact, however the converse is not true in general. Putinar [1993] showed that if M_S is archimedean then any $f \in \mathbb{R}[X]$ which is strictly positive on K_S is in M_S. In other words, these results say that under the given conditions a certificate of positivity for f on K_S exists.

Recently, techniques from semidefinite programming combined with Schmüdgen's and Putinar's theorems have been used to give numerical algorithms for applications such as optimization of polynomials on semialgebraic sets. However since these algorithms are numerical they might not produce exact certificates of positivity. With this in mind, Sturmfels asked whether any $f \in \mathbb{Q}[X]$ which is a sum of squares in $\mathbb{R}[X]$ is a sum of squares in $\mathbb{Q}[X]$. Hillar [2009] showed that the answer is yes in the case where f is known to be a sum of squares over a totally real field K. The general question remains unsolved.

It is natural to ask a similar question for Schmüdgen's and Putinar's theorems: If the polynomials defining the semialgebraic set and the positive polynomial f have rational coefficients, is there a certificate of positivity for f in which the sums of squares have rational coefficients? In this note, we show that in the case of Schmüdgen's theorem the answer is yes. This follows from an algebraic proof of the theorem, originally due to T. Wörmann [1998]. In the case of Putinar's theorem, we show that the answer is also yes as long as the generating set contains $N - \sum X_i^2$ for some $N \in \mathbb{N}$. This follows easily from an algorithmic proof of the theorem, due to Schweighofer [2005]. For Lasserre's method [2001] for optimization of polynomials on compact semialgebraic sets, it is usual in concrete cases to add a polynomial of the type $N - \sum X_i^2$ to the generators in order to insure that Putinar's theorem holds. Thus our assumption in this case is reasonable.

2. Rational certificates of for Schmüdgen's theorem

Fix $S = \{g_1, \ldots, g_s\} \subseteq \mathbb{R}[X]$ and define K_S and T_S as above.

Theorem 1 (Schmüdgen). *Suppose that K_S is compact. If $f \in \mathbb{R}[X]$ and $f > 0$ on K_S, then $f \in T_S$.*

In this section we show that if f and the generating polynomials $g_1, \ldots, g_s$ are in $\mathbb{Q}[X]$, then f has a representation in T_S in which all sums of squares σ_ϵ are in $\sum \mathbb{Q}[X]^2$. This follows from T. Wörmann's algebraic proof of the theorem using the classical Abstract Positivstellensatz, and a generalization of Wörmann's crucial lemma due to M. Schweighofer.

The abstract Positivstellensatz. We will need a version of the abstract Positivstellensatz, a result traditionally attributed to Kadison and Dubois, but now thought to have been proved earlier by Krivine or Stone. For details on its history, see [Prestel and Delzell 2001, Section 5.6]. The setting is preordered commutative rings, and we state the version we need as Theorem 2 below.

Let A be a commutative ring with $\mathbb{Q} \subseteq A$. A subset $T \subseteq A$ is a *preordering* if $T + T \subseteq T$, $T \cdot T \subseteq T$, and $-1 \notin T$. For $S = \{a_1, \ldots, a_k\} \subseteq A$, we define the *preordering generated by S*, T_S, exactly as for $A = \mathbb{R}[X]$.

An *ordering* in A is a preordering P such that $P \cup -P = A$ and $P \cap -P$ is a prime ideal. Any $a \in A$ has a unique sign in $\{-1, 0, 1\}$ with respect to a fixed ordering P and we use the notation $a \geq_P 0$ if $a \in P$, $a >_P 0$ if $a \in P \setminus (P \cap -P)$, etc.

Fix a preordered ring (A, T) and denote by Sper A the real spectrum of (A, T), i.e., the set of orderings of A which contain T. Then define

$$H(A) = \{a \in A \mid \text{ there exists } n \in \mathbb{N} \text{ with } n \pm a \geq_P 0 \text{ for all } P \in \operatorname{Sper} A\},$$

the *ring of geometrically bounded elements in* (A, T), and

$$H'(A) = \{a \in A \mid \text{ there exists } n \in \mathbb{N} \text{ with } n \pm a \in T\},$$

the *ring of arithmetically bounded elements in* (A, T). Clearly, $H'(A) \subseteq H(A)$. The preordering T is *archimedean* if $H'(A) = A$.

Theorem 2 [Schweighofer 2002, Theorem 1]. *Given the preordered ring (A, T) as above and suppose $A = H'(A)$. For any $a \in A$, if $a >_P 0$ for all $P \in$ Sper A, then $a \in T$.*

Consider the case where $A = \mathbb{R}[X]$ and $T = T_S$ for $S = \{g_1, \ldots, g_s\} \subseteq \mathbb{R}[X]$. Let $K = K_S$, then K embeds densely in Sper A and hence $H(A) = \{f \in \mathbb{R}[X] \mid f$ is bounded on $S\}$. If S is compact, this implies $H(A) = A$ and Schmüdgen's theorem follows from the following result:

Lemma 3 [Berr and Wörmann 2001, Lemma 1]. *With A, T, and S as above, if $H(A) = A$ then $H'(A) = A$.*

Our result follows from a generalization of this lemma:

Theorem 4 [Schweighofer 2002, Theorem 4.13]. *Let F be a subfield of $\mathbb{R}$ and (A, T) a preordered F-algebra such that $F \subseteq H'(A)$ and A has finite transcendence degree over F. Then*

$$A = H(A) \implies A = H'(A).$$

We can now prove the existence of rational certificates of positivity in Schmüdgen's theorem. The argument is exactly that of the proof of the general theorem above.

Theorem 5. *Given $S = \{g_1, \dots, g_s\} \subseteq \mathbb{Q}[X]$ and suppose $K_S \subseteq \mathbb{R}^n$ is compact. Then for any and $f \in \mathbb{Q}[X]$ such that $f > 0$ on K_S, there is a representation of f in the preordering T_S,*

$$f = \sum_{e \in \{0,1\}^s} \sigma_e g_1^{e_1} \dots g_s^{e_s},$$

with all $\sigma_e \in \sum \mathbb{Q}[X]^2$.

Proof. Let T be the preordering in $\mathbb{Q}[X]$ generated by S. Since K_S is compact, every element of $\mathbb{Q}[X]$ is bounded on K_S. Then K_S dense in Sper A implies that $H(\mathbb{Q}[X]) = \mathbb{Q}[X]$, hence by Theorem 4 we have $\mathbb{Q}[X] = H'(A)$. Note that the condition $F \subseteq H'(A)$ holds in this case since $\mathbb{Q}^+ = \sum \mathbb{Q}^2$. The result follows from Theorem 2. □

3. Rational certificates for Putinar's theorem

Given $S = \{g_1, \dots, g_s\}$, recall that the quadratic module generated by S, M_S, is the set of elements in the preordering K_S with a "linear" representation, i.e.,

$$M_S = \left\{\sigma_0 + \sigma_1 g_1 + \dots + \sigma_s g_s \mid \sigma_i \in \sum \mathbb{R}[X]^2\right\}.$$

In order to guarantee representations of positive polynomials in the quadratic module, we need a condition stronger than compactness of K_S, namely, we need M_S to be archimedean.

The quadratic module M_S is archimedean if all elements of $\mathbb{R}[X]$ are bounded by a positive integer with respect to M_S, i.e., if for every $f \in \mathbb{R}[X]$ there is some $N \in \mathbb{N}$ such that $N - f \in M_S$. It is not too hard to show that M_S is archimedean if there is some $N \in \mathbb{N}$ such that $N - \sum X_i^2 \in M_S$. Clearly, if M_S is archimedean, then K_S is compact; the polynomial $N - \sum X_i^2$ can be thought of as a "certificate of compactness". However, the converse is not true; see [Prestel and Delzell 2001, Example 6.3.1]. The key to the algebraic proof of Schmüdgen's theorem from the previous section is showing that in the case of the preordering generated by a finite set of elements from $\mathbb{R}[X]$, the compactness of the semialgebraic set implies that the corresponding preordering is archimedean.

Putinar [1993] showed that if the quadratic module M_S is archimedean, we can replace the preordering T_S by the quadratic module M_S.

Theorem 6 (Putinar). *Suppose that the quadratic module M_S is archimedean. Then for every $f \in \mathbb{R}[X]$ with $f > 0$ on K_S, $f \in M_S$.*

Lasserre's method [2001] for minimizing a polynomial on a compact semialgebraic set involves defining a sequence of semidefinite programs corresponding to representations of bounded degree in M_S whose solutions converge to the minimum. In this context, if M_S is archimedean then Putinar's theorem implies the convergence of the semidefinite programs. In practice, it is not clear how to decide if M_S is archimedean for a given set of generators S, however in concrete cases a polynomial $N - \sum X_i^2$ can be added to the generators if an appropriate N is known or can be computed.

Using an algorithmic proof of Putinar's theorem due to M. Schweighofer [2005] we can show that rational certificates exist for the theorem as long as we have a polynomial $N - \sum X_i^2$ as one of our generators:

Theorem 7. *Suppose $S = \{g_1, \dots, g_s\} \subseteq \mathbb{Q}[X]$ and $N - \sum X_i^2 \in M_S$ for some $N \in \mathbb{N}$. Then given any $f \in \mathbb{Q}[X]$ such that $f > 0$ on K_S, there exist $\sigma_0 \dots \sigma_s, \sigma \in \sum \mathbb{Q}[X]^2$ so that*

$$f = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s + \sigma\big(N - \textstyle\sum X_i^2\big).$$

Proof. The idea of Schweighofer's proof is to reduce to Pólya's theorem. We follow the proof, making sure that each step preserves rationality.

Let

$$\Delta = \big\{ y \in [0, \infty)^{2n} \mid y_1 + \cdots + y_{2n} = 2n(N + \tfrac{1}{4}) \big\} \subseteq \mathbb{R}^{2n}$$

and let C be the compact subset of $\mathbb{R}^n$ defined by $C = l(\Delta)$, where $l : \mathbb{R}^{2n} \to \mathbb{R}^n$ is defined by

$$y \mapsto \left(\frac{y_1 - y_{n+1}}{2}, \dots, \frac{y_n - y_{2n}}{2} \right).$$

Scaling the g_i's by positive elements in $\mathbb{Q}$, we can assume that $g_i \le 1$ on C for all i. The key to the proof is the observation that *there exists* $\lambda \in \mathbb{R}^+$ *such that* $q := f - \lambda \sum (g_i - 1)^{2k} g_i > 0$ *on* C [Schweighofer 2005, Lemma 2.3]. Since we can always replace λ by a smaller value, we can assume $\lambda \in \mathbb{Q}$, whence $q \in \mathbb{Q}[X]$.

Let $d = \deg q$ and let Q_i be the homogeneous part of q of degree i, so $q = \sum_{i=1}^{d} Q_i$. Let $Y = (Y_1, \dots, Y_{2n})$ and define in $\mathbb{Q}[Y]$

$$F(Y_1, \dots, Y_{2n}) := \sum_{i=1}^{d} Q_i \left(\frac{Y_1 - Y_{n+1}}{2}, \dots, \frac{Y_n - Y_{2n}}{2} \right) \left(\frac{Y_1 + \cdots + Y_{2n}}{2n(N + \frac{1}{4})} \right)^{d-i}.$$

Then F is homogenous and $F > 0$ on $[0, \infty)^{2n} \setminus \{0\}$. By Pólya's theorem, there is some $k \in \mathbb{N}$ so that

$$G := \left(\frac{Y_1 + \cdots + Y_{2n}}{2n(N + \frac{1}{4})} \right)^k F$$

has nonnegative coefficients as a polynomial in $\mathbb{R}[Y]$. Furthermore, since $F \in \mathbb{Q}[Y_1, \dots, Y_{2n}]$, it is easy to see that $G \in \mathbb{Q}[Y]$.

Define $\phi : \mathbb{Q}[Y_1, \dots, Y_{2n}] \to \mathbb{Q}[X]$ by

$$\phi(Y_i) = N + \tfrac{1}{4} + X_i, \quad \phi(Y_{n+i}) = (N + \tfrac{1}{4}) - X_i \quad \text{for } i = 1, \dots, n$$

and note that $\phi(G) = q$ and

$$\begin{aligned} \phi(Y_i) &= (N + \tfrac{1}{4}) \pm X_i \\ &= \sum_{j \neq i} \big(X_j^2 + (X_i \pm \tfrac{1}{2})^2 \big) + \big(N - \sum X_j^2 \big) \in \sum \mathbb{Q}[X]^2 + \big(N - \sum X_j^2 \big). \end{aligned}$$

Thus $\phi(G) = q$ implies there is a representation of q of the required type and then, since $f = q + \lambda \sum (g_i - 1)^{2k} g_i$ with $\lambda \in \mathbb{Q}$, we are done. $\square$

Remark 8. In the preordering case (Schmüdgen's theorem), as noted above if the semialgebraic set K_S is compact, then it follows that the preordering T_S in $\mathbb{Q}[X]$ is archimedean. However it is more subtle in the quadratic module case since it is not always clear how to decide if M_S is archimedean for a given set of generators S. Thus an open question is the following: Suppose $S \subseteq \mathbb{Q}[X]$ is a finite set of polynomials and M_S is archimedean as a quadratic module in $\mathbb{R}[X]$. Is it true that M_S is archimedean as a quadratic module in $\mathbb{Q}[X]$? To put it more concretely, suppose $S = \{g_1, \dots, g_s\} \subseteq \mathbb{Q}[X]$ and we know that there is some $N \in \mathbb{N}$ such that

$$N - \sum X_i^2 = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s,$$

with $\sigma_i \in \sum \mathbb{R}[X]^2$. Does there exist a representation with $\sigma_i \in \sum \mathbb{Q}[X]^2$? Equivalently, does there exist $N \in \mathbb{N}$ such that for each $i = 1, \dots, n$ we can write

$$N \pm X_i = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s,$$

with $\sigma_i \in \sum \mathbb{Q}[X]^2$?

References

[Berr and Wörmann 2001] R. Berr and T. Wörmann, "Positive polynomials on compact sets", *Manuscripta Math.* **104**:2 (2001), 135–143. MR 2002a:14061 Zbl 0992.14021

[Hillar 2009] C. J. Hillar, "Sums of squares over totally real fields are rational sums of squares", *Proc. Amer. Math. Soc.* **137**:3 (2009), 921–930. MR 2009h:12009 Zbl 1163.12005

[Lasserre 2001] J. B. Lasserre, "Global optimization with polynomials and the problem of moments", *SIAM J. Optim.* **11**:3 (2001), 796–817. MR 2002b:90054 Zbl 1010.90061

[Prestel and Delzell 2001] A. Prestel and C. N. Delzell, *Positive polynomials: from Hilbert's 17th problem to real algebra*, Springer, Berlin, 2001. MR 2002k:13044 Zbl 0987.13016

[Putinar 1993] M. Putinar, "Positive polynomials on compact semi-algebraic sets", *Indiana Univ. Math. J.* **42**:3 (1993), 969–984. MR 95h:47014 Zbl 0796.12002

[Schmüdgen 1991] K. Schmüdgen, "The K-moment problem for compact semi-algebraic sets", *Math. Ann.* **289**:2 (1991), 203–206. MR 92b:44011

[Schweighofer 2002] M. Schweighofer, *Iterated rings of bounded elements and generalizations of Schmüdgen's theorem*, Ph.D. thesis, Universität Konstanz, 2002, available at http://tinyurl.com/SchweighoferPhD.

[Schweighofer 2005] M. Schweighofer, "Optimization of polynomials on compact semialgebraic sets", *SIAM J. Optim.* **15**:3 (2005), 805–825. MR 2006d:90136 Zbl 1114.90098

[Wörmann 1998] T. Wörmann, *Strikt positive polynome in der semialgebraischen geometrie*, Ph.D. thesis, Universität Dortmund, 1998, available at http://www.lulu.com/items/volume_63/2227000/2227267/2/print/diss.pdf.

Received January 2, 2011. Revised February 3, 2011.

VICTORIA POWERS
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY
ATLANTA, GA 30322
UNITED STATES

vicki@mathcs.emory.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

QUIVER GRASSMANNIANS, QUIVER VARIETIES AND THE PREPROJECTIVE ALGEBRA

ALISTAIR SAVAGE AND PETER TINGLEY

Quivers play an important role in the representation theory of algebras, with a key ingredient being the path algebra and the preprojective algebra. Quiver grassmannians are varieties of submodules of a fixed module of the path or preprojective algebra. In the current paper, we study these objects in detail. We show that the quiver grassmannians corresponding to submodules of certain injective modules are homeomorphic to the lagrangian quiver varieties of Nakajima which have been well studied in the context of geometric representation theory. We then refine this result by finding quiver grassmannians which are homeomorphic to the Demazure quiver varieties introduced by the first author, and others which are homeomorphic to the graded/cyclic quiver varieties defined by Nakajima. The Demazure quiver grassmannians allow us to describe injective objects in the category of locally nilpotent modules of the preprojective algebra. We conclude by relating our construction to a similar one of Lusztig using projectives in place of injectives. In an appendix added after the first version of the current paper was released, we show how subsequent results of Shipman imply that the above homeomorphisms are in fact isomorphisms of algebraic varieties.

Savage's research was supported by the Natural Sciences and Engineering Research Council of Canada. Tingley was supported by Australian Research Council grant DP0879951 and NSF grant DMS-0902649.

MSC2010: primary 16G20; secondary 17B10.

Keywords: quiver, preprojective algebra, quiver grassmannian, quiver variety, Kac–Moody algebra, Demazure module.

Introduction

Quivers play a fundamental role in the theory of associative algebras and their representations. Gabriel's theorem, which states a precise relationship between indecomposable representations of certain quivers and root systems of associated Lie algebras, indicated that the representation theory of quivers was also intimately connected to the representation theory of Kac–Moody algebras. This eventually lead to the Ringel–Hall construction of quantum groups and the quiver variety constructions of Lusztig and Nakajima.

Fix a quiver (directed graph) $Q = (Q_0, Q_1)$ with vertex set Q_0 and arrow set Q_1. The corresponding path algebra $\mathbb{C}Q$ is the algebra spanned by the set of directed paths, with multiplication given by concatenation. There is a natural grading $\mathbb{C}Q = \bigoplus_n (\mathbb{C}Q)_n$ of the path algebra by length of paths. Representations of a quiver are equivalent to representations (or modules) of its path algebra. Note that $(\mathbb{C}Q)_0$-modules are simply Q_0-graded vector spaces, and in particular all $\mathbb{C}Q$-modules are Q_0-graded. For a $\mathbb{C}Q$-module V and $u \in \mathbb{N}Q_0$, the associated *quiver grassmannian* is the variety $\mathrm{Gr}_Q(u, V)$ of all $\mathbb{C}Q$-submodules of V of graded dimension u. These natural objects (or closely related ones) can be found in several places in the literature. For instance, they appear in [Crawley-Boevey 1996; Schofield 1992] in the study of spaces of morphisms of $\mathbb{C}Q$-modules and in [Caldero and Chapoton 2006; Caldero and Keller 2006; Derksen et al. 2009] in connection with the theory of cluster algebras. Geometric properties have been studied in [Caldero and Reineke 2008; Szántó 2009; Wolf 2009] and representation theoretic properties in [Fedotov 2010; Geiss et al. 2006; Lusztig 1998; 2000; Nakajima 2003; Reineke 2008].

Let $\mathfrak{g}$ be the Kac–Moody algebra whose Dynkin diagram is the underlying graph of Q (the graph obtained by forgetting the orientation of all arrows) and let $\tilde{Q}$ be the *double quiver* obtained from Q by adding an oppositely oriented arrow $\bar{a}$ for every $a \in Q_1$. One is often interested in modules of the preprojective algebra $\mathcal{P} = \mathcal{P}(Q)$, which is a certain natural quotient of the path algebra $\mathbb{C}\tilde{Q}$ and inherits the grading. In particular, $\mathcal{P}$-modules are also $\mathbb{C}\tilde{Q}$-modules. To each vertex $i \in Q_0$, we have an associated one-dimensional simple $\mathcal{P}$-module s^i. For

$$w = \sum_i w_i i \in \mathbb{N}Q_0,$$

we let $s^w = \bigoplus_i (s^i)^{\oplus w_i}$ be the corresponding semisimple module. By Baer's Theorem, the category of $\mathcal{P}$-modules has enough injectives, so we can define q^w to be the injective hull of s^w. One of the main results of the current paper is that the quiver grassmannian $\mathrm{Gr}_{\tilde{Q}}(v, q^w)$ is homeomorphic to the lagrangian Nakajima quiver variety $\mathfrak{L}(v, w)$ used to give a geometric realization of irreducible highest weight representations of $\mathfrak{g}$; [Nakajima 1994; 1998]. In addition, for each σ in the Weyl group of $\mathfrak{g}$, there is a natural finite-dimensional submodule $q^{w,\sigma}$ of q^w such that the quiver grassmannian $\mathrm{Gr}_{\tilde{Q}}(v, q^{w,\sigma})$ is homeomorphic to the Demazure quiver variety $\mathfrak{L}_\sigma(v, w)$ defined in [Savage 2006d]. Since Nakajima's realization of highest weight representations and the first author's realization of Demazure modules depend only on the topological information of the spaces involved, such homeomorphisms allow one to replace quiver varieties by quiver grassmannians in the constructions. This change of setting affords some advantages. In particular, it avoids the description as a moduli space. One can view it as a uniform way of picking a representative from each orbit in the original moduli space descriptions.

Quiver grassmannians admit natural group actions. We describe these actions and show that certain special cases agree, under the homeomorphisms described above, with well-studied groups actions on Nakajima quiver varieties. In this way, we are able to give a quiver grassmannian realization of the cyclic/graded quiver varieties used by Nakajima [2004] to define t-analogs of q-characters of quantum affine algebras.

The injective modules q^w are locally nilpotent if and only if the quiver Q is of finite or affine type. However, it turns out that the submodules $q^{w,\sigma}$ are always nilpotent. The limit $\tilde{q}^w$ of these submodules is the injective hull of the semisimple module s^w in the category of locally nilpotent $\mathcal{P}$-modules, giving us a description of the indecomposable injectives in this category.

Lusztig has previously presented a canonical bijection between the points of the lagrangian Nakajima quiver variety and the points of a type of quiver grassmannian inside a projective (as opposed to injective) object. In finite type, the projective objects are also injective. It turns out that, on the level of geometric realizations of representations of finite type $\mathfrak{g}$, the two constructions are related by the Chevalley involution. Outside of finite type, there are some other subtle yet important differences between the two constructions. In particular, the description in terms of projective objects requires one to impose a nilpotency condition in the definitions. However, the description in terms of injectives given in the current paper requires no such condition and is in this way simpler. Furthermore, through the use of the distinguished modules $q^{w,\sigma}$ mentioned above, one can always consider quiver grassmannians of submodules of a fixed *finite-dimensional* module of the preprojective algebra. Thus, one can avoid working with infinite-dimensional objects.

Motivated by an earlier version of the current paper [Savage and Tingley 2009], I. Shipman [2010] has recently proven that the canonical bijection given by Lusztig and mentioned above is, in fact, an isomorphism of algebraic varieties. We have added an Appendix explaining how this result allows us to conclude that the maps between quiver grassmannians and lagrangian Nakajima quiver varieties described in the current paper are also isomorphisms of algebraic varieties.

Throughout this paper, we work over the field $\mathbb{C}$ of complex numbers. While many results hold in more generality, this assumption will streamline the exposition and several results we quote in the literature are stated over $\mathbb{C}$. We will always use the Zariski topology and do not assume that algebraic varieties are irreducible. We let $\mathbb{N} = \mathbb{Z}_{\geq 0}$ and denote the fundamental weights and simple roots of a Kac–Moody algebra by ω_i and α_i respectively.

This paper is organized as follows. In Section 1 we review some results on quivers, path algebras and preprojective algebras. In Section 2 we discuss various module categories of these objects and introduce our main object of study, the quiver grassmannian. We review the definition of the quiver varieties of Lusztig and Nakajima in Section 3 and realize these as quiver grassmannians in Section 4. In Section 5 we introduce a natural group action and show how it can be used to recover group actions typically constructed on quiver varieties. We also define graded/cyclic versions of quiver grassmannians. In Section 6 we use quiver grassmannians to give a geometric realization of integrable highest weight representations of a symmetric Kac–Moody algebra and discuss the compatibility of this construction with the natural nesting of quiver grassmannians. Finally, in Section 7 we discuss a precise relationship between our construction and a similar one due to Lusztig. The Appendix, added after the appearance of [Shipman 2010], provides a proof that the maps between quiver grassmannians and quiver varieties described in the current paper are isomorphisms of algebraic varieties.

1. Quivers, path algebras, and preprojective algebras

We briefly review the relevant definitions concerning quivers. We refer the reader to [Deng et al. 2008; Ringel 1998; Savage 2006a] for further details.

A *quiver* is a directed graph. That is, it is a quadruple $Q = (Q_0, Q_1, s, t)$ where Q_0 and Q_1 are sets and s and t are maps from Q_1 to Q_0. We call Q_0 and Q_1 the sets of *vertices* and *directed edges* (or *arrows*) respectively. For an arrow $a \in Q_1$, we call $s(a)$ the *source* of a and $t(a)$ the *target* of a. Usually we will write $Q = (Q_0, Q_1)$, leaving the maps s and t implied. The quiver Q is said to be *finite* if Q_0 and Q_1 are finite. A *loop* is an arrow a with $s(a) = t(a)$. In this paper, all quivers will be assumed to be finite and without loops. A quiver is said to be of *finite type* if the underlying graph of Q (i.e the graph obtained from Q by forgetting the orientation of the edges) is a Dynkin diagram of finite ADE type. Similarly, it is of *affine* (or *tame*) *type* if the underlying graph is a Dynkin diagram of affine type and of *indefinite* (or *wild*) *type* if the underlying graph is a Dynkin diagram of indefinite type.

A *path* in Q is a sequence $\beta = a_l a_{l-1} \cdots a_1$ of arrows such that $t(a_i) = s(a_{i+1})$ for $1 \leq i \leq l-1$. We call l the *length* of the path. We let $s(\beta) = s(a_1)$ and

$t(\beta) = t(a_l)$ denote the initial and final vertices of the path β. For each vertex $i \in I$, we have a trivial path e_i with $s(e_i) = t(e_i) = i$.

The *path algebra* $\mathbb{C}Q$ associated to a quiver Q is the $\mathbb{C}$-algebra whose underlying vector space has basis the set of paths in Q, and with the product of paths given by concatenation. More precisely, if $\beta = a_l \cdots a_1$ and $\beta' = b_m \cdots b_1$ are two paths in Q, then $\beta\beta' = a_l \cdots a_1 b_m \cdots b_1$ if $t(\beta') = s(\beta)$ and $\beta\beta' = 0$ otherwise. This multiplication is associative. There is a natural grading

$$\mathbb{C}Q = \bigoplus_{n \geq 0} (\mathbb{C}Q)_n$$

where $(\mathbb{C}Q)_n$ is the span of the paths of length n.

Given a quiver $Q = (Q_0, Q_1)$, we define the *double quiver* associated to Q to be the quiver $\tilde{Q} = (Q_0, \tilde{Q}_1)$ where

$$\tilde{Q}_1 = \bigcup_{a \in Q_1} \{a, \bar{a}\}, \quad \text{where} \quad s(\bar{a}) = t(a),\ t(\bar{a}) = s(a).$$

We then have a natural involution $\tilde{Q}_1 \to \tilde{Q}_1$ given by $a \mapsto \bar{a}$ (where $\bar{\bar{a}} = a$). The algebra

$$\mathcal{P} = \mathcal{P}(Q) = \mathbb{C}\tilde{Q} / \sum_{a \in Q_1} (a\bar{a} - \bar{a}a)$$

is called the *preprojective algebra* associated to Q. It inherits a grading

$$\mathcal{P} = \bigoplus_{n \geq 0} \mathcal{P}_n$$

from the grading on $\mathbb{C}Q$. Up to isomorphism, the preprojective algebra $\mathcal{P}(Q)$ depends only on the underlying graph of Q. See [Lusztig 1991, §12.15] for details.

2. Modules of the path algebra and quiver grassmannians

2A. *Module categories.* For an associative algebra A, let A-Mod denote the category of A-modules and A-mod the category of finite-dimensional A-modules. We will use the notation $V \in A$-Mod (resp. $V \in A$-mod) to indicate that V is an object in the category A-Mod (resp. A-mod). Note that $\mathcal{P}_0$-mod is equivalent to the category of finite-dimensional Q_0-graded vector spaces whose morphisms are linear maps preserving the grading, and we will often blur the distinction between these two categories. Up to isomorphism, the objects of $\mathcal{P}_0$-mod are classified by their graded dimension. We denote the graded dimension of a module V by $\dim_{Q_0} V = \sum_i (\dim V_i) i \in \mathbb{N}Q_0$ and let $\dim_{\mathbb{C}} V = \sum_{i \in Q_0} \dim V_i \in \mathbb{N}$. We will sometimes view the graded dimension $\dim_{Q_0} V$ of V as its isomorphism class.

For $V, W \in \mathcal{P}_0$-mod, we denote the set of $\mathcal{P}_0$-module morphisms from V to W by $\mathrm{Hom}_{\mathcal{P}_0}(V, W)$. Under the equivalence of categories above, $\mathrm{Hom}_{\mathcal{P}_0}(V, W)$ is identified with $\bigoplus_{i\in Q_0}\mathrm{Hom}_{\mathbb{C}}(V_i, W_i)$. We define $\mathrm{End}_{\mathcal{P}_0}V$ to be $\mathrm{Hom}_{\mathcal{P}_0}(V, V)$ and $\mathrm{GL}_V=\prod_{i\in Q_0}GL(V_i)$ to be group of invertible elements of $\mathrm{End}_{\mathcal{P}_0}V$. For $V\in\mathcal{P}_0$-mod, we will write $U\subseteq V$ to mean that U is a $\mathcal{P}_0$-submodule of V. This is the same as a Q_0-graded subspace. Note that any $\mathcal{P}$-module becomes a $\mathcal{P}_0$-module by restriction, and thus can be thought of as a Q_0-graded vector space.

Suppose $A=\bigoplus_{n\geq 0}A_n$ is a graded algebra and V is an A-module. Then V is *nilpotent* if there exists an $n\in\mathbb{N}$ such that $A_k\cdot V=0$ for all $k\geq n$. We say V is *locally nilpotent* if for all $v\in V$, there exists $n\in\mathbb{N}$ such that $A_k\cdot v=0$ for all $k\geq n$. We denote by A-lnMod the category of locally nilpotent A-modules. For $n\geq 0$, we define $A_{\geq n}=\bigoplus_{k\geq n}A_k$ and we let $A_+=A_{\geq 1}$.

Proposition 2.1. *For a quiver Q, the following are equivalent*:

(i) *$\mathcal{P}(Q)$ is finite-dimensional*,

(ii) *all finite-dimensional $\mathcal{P}(Q)$-modules are nilpotent*,

(iii) *all finite-dimensional $\mathcal{P}(Q)$-modules are locally nilpotent*, *and*

(iv) *Q is of finite type.*

Proof. The equivalence of (i) and (iv) is well-known; see [Reiten 1997], for example. That (ii) implies (iv) was proven in [Crawley-Boevey 2001] and the converse was proven by Lusztig [Lusztig 1991, Proposition 14.2]. Since a finite-dimensional module is nilpotent if and only if it is locally nilpotent, (ii) is equivalent to (iii). □

2B. *Simple objects.* For each $i\in Q_0$, let s^i be the simple $\mathbb{C}\tilde{Q}$-module given by $s^i_i=\mathbb{C}$ and $s^i_j=0$ for $i\neq j$. Then s^i is also naturally a $\mathcal{P}$-module which we also denote by s^i.

Lemma 2.2. *The set $\{s^i\}_{i\in Q_0}$ is a set of representatives of the isomorphism classes of simple objects of $\mathbb{C}\tilde{Q}$-lnMod and $\mathcal{P}$-lnMod. In particular*, *if Q is of finite type*, *then $\{s^i\}_{i\in Q_0}$ is a set of representatives of the isomorphism classes of simple objects of $\mathbb{C}\tilde{Q}$-mod and $\mathcal{P}$-mod.*

Proof. Any nonzero element of a simple locally nilpotent module M generates a finite-dimensional module which must be all of M. Therefore M is finite-dimensional and hence nilpotent. Then $(\mathbb{C}\tilde{Q})_+$ and $\mathcal{P}_+$ are two-sided ideals of $\mathbb{C}\tilde{Q}$ and $\mathcal{P}$ respectively that act nilpotently on any nilpotent module. Therefore, simple nilpotent $\mathbb{C}\tilde{Q}$-modules and $\mathcal{P}$-modules are the same as simple $\mathbb{C}\tilde{Q}/(\mathbb{C}\tilde{Q})_+$-modules and $\mathcal{P}/\mathcal{P}_+$-modules respectively. Since

$$\mathbb{C}\tilde{Q}/(\mathbb{C}\tilde{Q})_+\cong\mathcal{P}/\mathcal{P}_+\cong\bigoplus_{i\in I}\mathbb{C}e_i,$$

the first statement follows. The second statement then follows from Proposition 2.1.
□

Lemma 2.3. *Fix a quiver Q and let A be either $\mathbb{C}\tilde{Q}$ or $\mathcal{P}(Q)$. If $V \in A$-lnMod, then the socle of V is $\{v \in V \mid A_+ \cdot v = 0\}$.*

Proof. It is clear that $\{v \in V \mid A_+ \cdot v = 0\}$ is a sum of simple subrepresentations of V and is thus contained in the socle of V. Similarly, by Lemma 2.2, any simple subrepresentation of (V, x) is contained in $\{v \in V \mid A_+ \cdot v = 0\}$. □

2C. *Projective covers.* Recall that if A is an associative algebra and V is an A-module, then a *projective cover* of V is a pair (P, f) such that P is a projective A-module and $f : P \to V$ is a superfluous epimorphism of A-modules. This means that $f(P) = V$ and $f(P') \neq V$ for all proper submodules P' of P. We often omit the homomorphism f and simply call P a projective cover of V.

Definition 2.4. For $i \in Q_0$, let $p^i = \mathcal{P}e_i$.

Lemma 2.5. *Assume Q is a quiver of finite type. For $i \in Q_0$, $\{p^i\}_{i \in Q_0}$ is a set of representatives of the isomorphism classes of indecomposable projective $\mathcal{P}$-modules. Furthermore, p^i is a projective cover of s^i.*

Proof. This follows from [Auslander et al. 1995, Proposition 4.8]. □

Lemma 2.6. *Assume Q is a quiver of affine (tame) or indefinite (wild) type. Then there exist $i \in Q_0$ for which the simple module s^i does not have a projective cover.*

Proof. Since the module s^i is obviously cyclic, by [Anderson and Fuller 1992, Lemma 27.3] it has a projective cover if and only if $s^i \cong \mathcal{P}e/Ie$ for some idempotent $e \in \mathcal{P}$ and some left ideal I contained in the Jacobson radical of $\mathcal{P}$. Assume this is true for some idempotent e and ideal I. Then we must have $e = e_i$ and then I would have to contain $\mathcal{P}_+e_i$, the ideal consisting of all paths of length at least one starting at vertex i. We identify $\mathbb{Z}Q_0$ with the root lattice via $\sum v_j j \leftrightarrow \sum v_j \alpha_j$. Let β be a minimal positive imaginary root and let i be in the support of β (i.e., $\beta = \sum \beta_j \alpha_j$ with $\beta_i > 0$). By [Crawley-Boevey 2001, Theorem 1.2], there is a simple module T of $\mathcal{P}$ whose dimension vector is β and so, in particular, $\dim T_i \neq 0$. Since the simple module T cannot be killed by $\mathcal{P}_+e_i$ (since then T_i would be a proper submodule), $\mathcal{P}_+e_i$ is not contained in the Jacobson radical of $\mathcal{P}$. This contradicts the fact that I is contained in the Jacobson radical. □

2D. *Injective hulls.* Recall that if A is an associative algebra and V is an A-module, then an *injective hull* of V is an injective A-module E that is an essential extension of V (that is, V is a submodule of E and any nonzero submodule of E intersects V nontrivially). By Baer's Theorem [1940], the category $\mathcal{P}$-Mod has enough injectives. In particular, the simple modules s^i have injective hulls. Here we give an explicit description of these injective hulls in the finite type case, and study some of their properties in the more general case.

Definition 2.7. Assume Q is a quiver of finite type. For $i \in Q_0$, let

$$q^i = \mathrm{Hom}_{\mathbb{C}}(e_i\mathcal{P}, \mathbb{C})$$

be the dual space of the right $\mathcal{P}$-module $e_i\mathcal{P}$. Define a left $\mathcal{P}$-module structure on q^i by setting $a \cdot f(x) = f(xa)$, for $a \in \mathcal{P}$, $f \in q^i$, and $x \in e_i\mathcal{P}$.

Lemma 2.8. *If Q is a quiver of finite type, then $\{q^i\}_{i\in Q_0}$ is a set of representatives of the isomorphism classes of indecomposable injective $\mathcal{P}$-modules. Furthermore, q^i is an injective hull of s^i.*

Proof. If Q is of finite type, then $\mathcal{P}$ is finite-dimensional by Proposition 2.1. The result then follows from Lemma 2.5 and a well-known fact about modules over finite-dimensional algebras; see, for example, [Lam 1999, Corollary 3.66]. □

For $w = \sum_i w_i i \in \mathbb{N}Q_0$, define the semisimple $\mathcal{P}$-module

$$s^w = \bigoplus_{i\in Q_0} (s^i)^{\oplus w_i}.$$

Let q^i be the injective hull of s^i in the category $\mathcal{P}$-Mod (if Q is a quiver of finite type, this agrees with the notation of Definition 2.7). Then

$$q^w = \bigoplus_{i\in I} (q^i)^{\oplus w_i}$$

is the injective hull of s^w.

Lemma 2.9. *For $w \in \mathbb{N}Q_0$, any finite-dimensional submodule of q^w is nilpotent.*

Proof. Let V be a finite-dimensional submodule of q^w. Then we have the chain of submodules $V = \mathcal{P}_{\geq 0}V \supseteq \mathcal{P}_{\geq 1}V \supseteq \mathcal{P}_{\geq 2}V \supseteq \cdots$. Since q^w is an essential extension of s^w, we have $s^w \cap \mathcal{P}_{\geq n}V \neq 0$ for all $n \in \mathbb{N}$ such that $\mathcal{P}_{\geq n}V \neq 0$. Because $\mathcal{P}_1$ acts trivially on s^w, we have $\dim \mathcal{P}_{\geq n+1}V < \dim \mathcal{P}_{\geq n}V$ for all $n \in \mathbb{N}$ such that $\mathcal{P}_{\geq n}V \neq 0$. Thus $\mathcal{P}_{\geq n}V = 0$ for n large enough. □

Remark 2.10. It follows from Lemma 2.9 and Proposition 7.10 that if Q is a quiver of finite type, then p^w (and q^w) is nilpotent. However, in general the p^w are not nilpotent.

Proposition 2.11. *If Q is of affine (tame) type, then q^w is locally nilpotent for all $w \in \mathbb{N}Q_0$. If Q is connected and of indefinite (wild) type, then q^w is not locally nilpotent for any $w \in \mathbb{N}Q_0$, $w \neq 0$.*

The following proof was explained to us by W. Crawley-Boevey.

Proof. It suffices to consider the case where $w = i$ for some $i \in Q_0$. We identify $\mathbb{Z}Q_0$ with the root lattice via $\sum v_j j \leftrightarrow \sum v_j \alpha_j$. We first assume that Q is connected of wild type. Let β be a minimal positive imaginary root. Thus $(\beta, j) \leq 0$

for all $j \in Q_0$. Suppose the support of β is all of Q_0. Since Q is wild, β cannot be a radical vector (see [Kac 1990, Theorem 4.3]), so $(\beta, j) < 0$ for some $j \in Q_0$. If, on the other hand, the support of β is not all of Q_0, we take $j \in Q_0$ to be a vertex not in the support of β but connected to it by an arrow and we again have $(\beta, j) < 0$. By [Crawley-Boevey 2001, Theorem 1.2], there is a simple module T for the preprojective algebra of dimension β. By [Crawley-Boevey 2000, Lemma 1], $\mathrm{Ext}^1(T, s^j)$ is nonzero. Let V be a nontrivial extension of T by s^j. This module must embed in the injective hull q^j of s^j and thus q^j cannot be locally nilpotent. Thus the result holds whenever $(\beta, i) < 0$. For general i, choose a shortest path from i to some j with $(\beta, j) < 0$ and consider the corresponding nilpotent module U with head s^j and socle s^i. Then, as above, there is a nontrivial extension of T by U, which must embed into q^i. So q^i is not locally nilpotent.

Now assume that Q is of tame type. Since the preprojective algebra of a tame quiver is a finitely generated $\mathbb{C}$-algebra, noetherian, and a polynomial identity ring [Baer et al. 1987, Theorem 6.5] (see [Ringel 1998] for a proof that the preprojective algebra considered there is the same as the one considered here), any simple module is finite-dimensional; see [McConnell and Robson 2001, Theorem 13.10.3]. By [Jategaonkar 1976, Theorem 2], the injective hull of a simple $\mathcal{P}$-module is artinian. In particular, finitely generated submodules of injective hulls of simple modules are artinian and noetherian. Thus they are of finite length and hence finite-dimensional. Now, the dimension vectors of simple $\mathcal{P}$-modules are the coordinate vectors $i \in Q_0$ and the minimal imaginary root δ. Since $(\delta, i) = 0$ for all $i \in Q_0$, there are no nontrivial extensions between simples of dimension δ and the one-dimensional simples. Therefore, the composition factors of the finite-dimensional submodules of the injective hull q^i of s^i are all one-dimensional simple modules. Thus q^i is locally nilpotent. □

Remark 2.12. In types A and D, there exist simple and explicit descriptions of the representations q^i, $i \in Q_0$, in terms of classical combinatorial objects such as Young diagrams; see [Frenkel and Savage 2003; Savage 2006b; 2006c]. This allows one to give simple and explicit descriptions of the injective modules q^w for any $w \in \mathbb{N}Q_0$ when the underlying graph of the corresponding quiver is of type A or D.

2E. *Quiver grassmannians.*

Definition 2.13 (quiver grassmannian). For a $\mathbb{C}Q$-module V, let $\mathrm{Gr}_Q(V)$ be the variety of all $\mathbb{C}Q$-submodules of V. We have a natural decomposition

$$\mathrm{Gr}_Q(V) = \bigsqcup_{u \in \mathbb{N}Q_0} \mathrm{Gr}_Q(u, V), \quad \mathrm{Gr}_Q(u, V) = \{U \in \mathrm{Gr}_Q(V) \mid \dim U = u\}.$$

We call $\mathrm{Gr}_Q(u, V)$ a *quiver grassmannian*. Note that $\mathrm{Gr}_Q(u, V)$ is a closed subset of the usual grassmannian of dimension u subspaces of V and thus is a projective variety. If V is a $\mathcal{P}$-module, then $\mathcal{P}$-submodules of V are the same as $\mathbb{C}\tilde{Q}$-submodules of V. Hence one can think of $\mathrm{Gr}_{\tilde{Q}}(V)$ as the variety of all $\mathcal{P}$-submodules of V. Therefore, we will often write $\mathrm{Gr}_{\mathcal{P}}(V)$ and $\mathrm{Gr}_{\mathcal{P}}(u, V)$ for $\mathrm{Gr}_{\tilde{Q}}(V)$ and $\mathrm{Gr}_{\tilde{Q}}(u, V)$ when V is a $\mathcal{P}$-module.

Example 2.14 (grassmannians). If Q is the quiver with a single vertex and no arrows, then $\mathcal{P} = \mathbb{C}$ and $\mathcal{P}$-modules are simply vector spaces. Then $\mathrm{Gr}_{\mathcal{P}}(u, V) = \mathrm{Gr}(u, V)$ is the usual grassmannian of dimension u subspaces of V.

Example 2.15 (partial flag varieties). Let Q be the quiver with $Q_0 = \{1, 2, \ldots, n\}$ and $Q_1 = \{a_1, \ldots, a_{n-1}\}$, where $s(a_i) = i$, $t(a_i) = i+1$ for all $i = 1, \ldots, n-1$. Fix a positive integer d and set $V_i = \mathbb{C}^d$ for all $i = 1, \ldots, n$. For each $1 \le i \le n-1$, let a_i act by the identification $V_i \cong V_{i+1}$. Then for $u \in \mathbb{N}Q_0$ with $u_1 \le u_2 \le \cdots \le u_n \le d$, the quiver grassmannian $\mathrm{Gr}_{\mathcal{P}}(u, V)$ is isomorphic to the partial flag variety

$$\{0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq \mathbb{C}^d \mid \dim F_i = u_i\}.$$

Definition 2.16. For $V \in \mathcal{P}$-Mod, we define a natural action of $\mathrm{Aut}_{\mathcal{P}} V$ on $\mathrm{Gr}_{\mathcal{P}}(u, V)$ by

$$(g, U) \mapsto g(U), \quad g \in \mathrm{Aut}_{\mathcal{P}} V, \quad U \in \mathrm{Gr}_{\mathcal{P}}(u, V).$$

3. Quiver varieties

We briefly recall certain quiver varieties defined by Lusztig and Nakajima, referring the reader to [Lusztig 1991; Nakajima 1994; 1998] for further details, as well as the Demazure quiver varieties introduced in [Savage 2006d]. We fix a quiver $Q = (Q_0, Q_1)$ and let $\mathcal{P} = \mathcal{P}(Q)$ denote its preprojective algebra.

3A. ***Lusztig and Nakajima quiver varieties.*** For $V \in \mathcal{P}_0$-mod, define

$$\mathrm{Rep}_{\tilde{Q}} V = \bigoplus_{a \in \tilde{Q}_1} \mathrm{Hom}_{\mathbb{C}}(V_{s(a),t(a)}).$$

For a path $\beta = a_l \cdots a_1$ in Q and $x = (x_a)_{a \in \tilde{Q}_1} \in \mathrm{Rep}_{\tilde{Q}} V$, we define $x_\beta = x_{a_l} \cdots x_{a_1}$. For an element $\sum_j c_j \beta_j \in \mathbb{C}Q$, we define

$$x_{\sum_j c_j \beta_j} = \sum_j c_j x_{\beta_j}.$$

Thus each $x \in \mathrm{Rep}_{\tilde{Q}} V$ defines a representation $\mathbb{C}\tilde{Q} \to \mathrm{End}_{\mathbb{C}} V$ of graded dimension $\dim_{Q_0} V$ (i.e., whose induced representation of $(\mathbb{C}Q)_0$ is in the isomorphism class determined by $\dim_{Q_0} V$). Furthermore, each such representation comes from an element of $x \in \mathrm{Rep}_{\tilde{Q}} V$. These two statements are simply the equivalence of categories between the representations of the quiver and of the path algebra. We say that x is *nilpotent* if there exists $N > 0$ such that $x_\beta = 0$ for all paths β of length greater than N.

Definition 3.1 (Lusztig nilpotent variety). For $V \in \mathcal{P}_0$-mod, define $\Lambda(V) = \Lambda_Q(V)$ to be the set of all nilpotent $\mathcal{P}$-module structures on V compatible with its $\mathcal{P}_0$-module structure. More precisely,

$$\Lambda(V) = \left\{ x \in \operatorname{Rep}_{\tilde{Q}} V \;\middle|\; \sum_{\substack{a \in Q_1, \\ t(a)=i}} x_a x_{\bar{a}} - \sum_{\substack{a \in Q_1, \\ s(a)=i}} x_{\bar{a}} x_a = 0 \;\; \forall i \in Q_0,\; x \text{ nilpotent} \right\}.$$

We call $\Lambda(V)$ a *Lusztig nilpotent variety.*

As above, elements of $\Lambda(V)$ are in natural one-to-one correspondence with nilpotent representations $\mathcal{P} \to \operatorname{End}_{\mathbb{C}} V$ of graded dimension $\dim_{Q_0} V$.

For $V, W \in \mathcal{P}_0$-mod, let $\Lambda(V, W) = \Lambda(V) \times \operatorname{Hom}_{\mathcal{P}_0}(V, W)$. We say that $(x, t) \in \Lambda(V, W)$ is *stable* if there exists no nontrivial x-invariant $\mathcal{P}_0$-submodule of V contained in $\ker t$. This is equivalent to the condition that $\ker((x, t)|_{V_i}) = 0$ for all $i \in Q_0$ (see [Frenkel and Savage 2003, Lemma 3.4] — while the statement there is for type A, the proof carries over to the more general case). We denote the set of stable elements by $\Lambda(V, W)^{\text{st}}$. There is a natural action of GL_V on $\Lambda(V, W)$ and the restriction to $\Lambda(V, W)^{\text{st}}$ is free; see [Nakajima 1994; 1998]. We denote the GL_V-orbit through a point (x, t) by $[x, t]$.

Definition 3.2 (lagrangian Nakajima quiver variety). For $V, W \in \mathcal{P}_0$-mod, let $\mathfrak{L}(V, W) = \Lambda(V, W)^{\text{st}}/\mathrm{GL}_V$. We call $\mathfrak{L}(V, W)$ a *lagrangian Nakajima quiver variety*. Up to isomorphism, this variety depends only on $v = \dim_{Q_0} V$ and $w = \dim_{Q_0} W$ and so we will sometimes denote it by $\mathfrak{L}(v, w)$.

Remark 3.3. The quiver varieties defined above are lagrangian subvarieties of what are usually called the Nakajima quiver varieties [Nakajima 1994; 1998].

3B. *Group actions.* Let $G_{\mathcal{P}}$ be the group of algebra automorphisms of $\mathcal{P}$ that fix $\mathcal{P}_0$. The group GL_W acts naturally on $\operatorname{Hom}_{\mathcal{P}_0}(V, W)$. As above, we identify elements of $\Lambda(V)$ with nilpotent representations $\mathcal{P} \to \operatorname{End}_{\mathbb{C}} V$ of graded dimension $\dim_{Q_0} V$. Then

$$(h, (x, t)) \mapsto (h \star x, t), \quad h \star x = x \circ h^{-1}, \quad h \in G_{\mathcal{P}},$$

defines a $G_{\mathcal{P}}$-action on $\Lambda(V, W)$. The actions of GL_W and $G_{\mathcal{P}}$ commute and both commute with the GL_V-action. Since they also preserve the stability condition, they define a $\mathrm{GL}_W \times G_{\mathcal{P}}$-action on $\mathfrak{L}(v, w)$.

We can use this action to define $\mathrm{GL}_W \times \mathbb{C}^*$-actions on $\mathfrak{L}(v, w)$ as follows. Suppose a function $m : \tilde{Q}_1 \to \mathbb{Z}$ is given such that $m(a) = -m(\bar{a})$ for all $a \in \tilde{Q}_1$.

Then the map $a \mapsto z^{m(a)+1}a$, $z \in \mathbb{C}^*$, extends to an automorphism of $\mathcal{P}$ fixing $\mathcal{P}_0$. We denote this automorphism by $h_m(z)$. Thus h_m defines a group homomorphism $\mathbb{C}^* \to G_{\mathcal{P}}$. Then the homomorphism

$$\mathrm{GL}_W \times \mathbb{C}^* \to \mathrm{GL}_W \times G_{\mathcal{P}}, \quad (g, z) \mapsto (zg, h_m(z)) \tag{3-1}$$

defines a $\mathrm{GL}_W \times \mathbb{C}^*$-action on $\mathfrak{L}(v, w)$ which we denote by $\star_m$.

We give two important examples of this action [Nakajima 2001, §2.7; 2004]. First, for each pair $i, j \in Q_0$ connected by at least one edge, let b_{ij} denote the number of arrows in Q_1 joining i and j. We fix a numbering $a_1, \dots, a_{b_{ij}}$ of these arrows, which induces a numbering $\bar{a}_1, \dots, \bar{a}_{b_{ij}}$ of the corresponding arrows in $\bar{Q}_1$. Define $m_1 : H \to \mathbb{Z}$ by

$$m_1(a_p) = b_{ij} + 1 - 2p, \quad m_1(\bar{a}_p) = -b_{ij} - 1 + 2p.$$

For the second action, we define $m_2(a) = 0$ for all $a \in Q_1$.

3C. *Demazure quiver varieties.* Let $\mathfrak{g}$ be the Kac–Moody algebra corresponding to the underlying graph of Q (the one whose Dynkin diagram is this graph) and let $\mathcal{W}$ be its Weyl group. Recall that $\mathcal{W}$ acts naturally on the weight lattice of $\mathfrak{g}$. For $u \in \mathbb{Z}Q_0$, we define elements of the weight and root lattice by

$$\omega_u = \sum_{i \in Q_0} u_i \omega_i, \quad \alpha_u = \sum_{i \in Q_0} u_i \alpha_i.$$

Proposition/Definition 3.4 [Savage 2006d, Proposition 5.1]. The lagrangian Nakajima quiver variety $\mathfrak{L}(v, w)$ is a point if and only if $\omega_w - \alpha_v = \sigma(\omega_w)$ for some $\sigma \in \mathcal{W}$ (i.e., $\omega_w - \alpha_v$ is an extremal weight of the irreducible representation of highest weight ω_w, equivalently v is w-extremal in the sense of Definition 4.7). In this case, we let $(x^{w,\sigma}, t^{w,\sigma})$ be a representative (unique up to isomorphism) of the GL_V-orbit corresponding to this point. So $\mathfrak{L}(v, w) = \{[x^{w,\sigma}, t^{w,\sigma}]\}$ when $\omega_w - \alpha_v = \sigma(\omega_w)$.

Definition 3.5 (Demazure quiver variety). For $\sigma \in \mathcal{W}$ and $v, w \in \mathbb{N}Q_0$, let $\mathfrak{L}_\sigma(v, w)$ be the subvariety consisting of all $[x, t] \in \mathfrak{L}(v, w)$ such that (x, t) is isomorphic to a subrepresentation of $(x^{w,\sigma}, t^{w,\sigma})$. We call $\mathfrak{L}_\sigma(v, w)$ a *Demazure quiver variety*.

Remark 3.6. It follows from the uniqueness assertion in Proposition/Definition 3.4 that the $\mathrm{GL}_W \times G_{\mathcal{P}}$-action on $\mathfrak{L}(v, w)$ fixes $\mathfrak{L}_\sigma(v, w)$ for all $\sigma \in \mathcal{W}$. Thus we have an induced $\mathrm{GL}_W \times G_{\mathcal{P}}$-action on the Demazure quiver varieties.

4. Quiver varieties as quiver grassmannians

4A. *Lagrangian Nakajima quiver varieties as quiver grassmannians.* We will now show that certain quiver grassmannians are homeomorphic to the lagrangian Nakajima quiver varieties. We begin with a key technical proposition.

Proposition 4.1. *Suppose $A = \bigoplus_{n\geq 0} A_n$ is a graded algebra and V is a locally nilpotent A-module. Furthermore, suppose S is a semisimple locally nilpotent A-module with injective hull E.*

(i) *Let $\pi : E \to S$ be an A_0-linear retract for the canonical embedding $\iota : S \to E$ (that is, an A_0-linear map such that $\pi\iota = \mathrm{id}$) and let $\tau : V \to S$ be a homomorphism of A_0-modules. Then there exists a unique A-module homomorphism $\gamma : V \to E$ such that the following diagram commutes:*

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ V \arrow[ur, "\gamma"] \arrow[r, "\tau"'] & S \end{tikzcd}$$

Furthermore, the map γ is injective if and only if $\tau|_{\mathrm{socle}\, V}$ is injective.

(ii) *Suppose $\pi_1, \pi_2 : E \to S$ are A_0-linear retracts for the canonical embedding $\iota : S \to E$. Then there exists a unique $\gamma \in \mathrm{Aut}_A E$ such that $\pi_2 = \pi_1\gamma$. The map γ fixes S pointwise. Conversely, given an A_0-linear retract $\pi : E \to S$ and any $\gamma \in \mathrm{Aut}_A E$ fixing S pointwise, $\pi\gamma : E \to S$ is also a A_0-linear retract.*

Proof. Since V is locally nilpotent, we have a filtration

$$0 = V^{(0)} \subseteq V^{(1)} = \mathrm{socle}\, V \subseteq V^{(2)} \subseteq V^{(3)} \subseteq \cdots$$

of V where $V^{(n)} = \{m \in V \mid A_{\geq n} \cdot m = 0\}$. We prove by induction on n that there exists a unique homomorphism $\gamma_n : V^{(n)} \to E$ such that the diagram

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ V^{(n)} \arrow[ur, "\gamma_n"] \arrow[r, "\tau_n"'] & S \end{tikzcd} \tag{4-1}$$

commutes, where $\tau_n = \tau|_{V^{(n)}}$. Since $V^{(1)} = \mathrm{socle}\, V$ and $A_+ \cdot \mathrm{socle}\, V = 0$, we must have $\gamma_1(V^{(1)}) \subseteq S$ and so the unique choice for γ_1 is τ_1. Suppose the statement holds for $n = k$. Since E is injective, there exists an A-module homomorphism $\hat{\gamma}_{k+1}$ such that the following diagram commutes:

$$\begin{tikzcd} V^{(k+1)} \arrow[r, "\hat{\gamma}_{k+1}"] & E \\ V^{(k)} \arrow[u, hook] \arrow[ur, "\gamma_k"'] & \end{tikzcd}$$

Define γ_{k+1} by

$$\gamma_{k+1} = \hat{\gamma}_{k+1} - \pi \circ \hat{\gamma}_{k+1} + \tau.$$

It is then clear that the diagram (4-1) commutes (with $n = k+1$). Note also that $\gamma_{k+1}|_{V^{(k)}} = \gamma_k$. We claim that γ_{k+1} is a homomorphism of A-modules. Since it is an A_0-module homomorphism by definition, it suffices to show it commutes with the action of A_+.

For $r \in A_+$ and $m \in V^{(k+1)}$, we have $r \cdot m \in V^{(k)}$. Also, $A_+ \cdot S = 0$. Then

$$\begin{aligned} r \cdot \gamma_{k+1}(m) &= r \cdot (\hat{\gamma}_{k+1}(m) - \pi \circ \hat{\gamma}_{k+1}(m) + \tau(m)) \\ &= r \cdot \hat{\gamma}_{k+1}(m) = \hat{\gamma}_{k+1}(r \cdot m) = \gamma_k(r \cdot m) \\ &= \gamma_{k+1}(r \cdot m), \end{aligned}$$

as desired.

Now suppose that γ'_{k+1} is another $\mathcal{P}$-module homomorphism making (4-1) commute (with $n = k+1$). By the inductive hypothesis, we have $\gamma_{k+1}|_{V^{(k)}} = \gamma'_{k+1}|_{V^{(k)}}$. For all $r \in A_+$ and $m \in V^{(k+1)}$, we have

$$r \cdot \gamma_{k+1}(m) = \gamma_{k+1}(r \cdot m) = \gamma'_{k+1}(r \cdot m) = r \cdot \gamma'_{k+1}(m).$$

Thus $\gamma_{k+1}(m) - \gamma'_{k+1}(m)$ lies in S. Therefore

$$\begin{aligned} \gamma_{k+1}(m) - \gamma'_{k+1}(m) &= \pi(\gamma_{k+1}(m) - \gamma'_{k+1}(m)) \\ &= \pi(\gamma_{k+1}(m)) - \pi(\gamma'_{k+1}(m)) = \tau(m) - \tau(m) = 0. \end{aligned}$$

The induction is complete and we obtain the desired map γ by taking the limit.

Note that $\gamma|_{\mathrm{socle}\, V} = \tau|_{\mathrm{socle}\, V}$. Since a homomorphism of modules is injective if and only if its restriction to the socle is injective, it follows that γ is injective if and only if $\tau|_{\mathrm{socle}\, V}$ is injective.

We now prove (ii). By (i), there exists a unique A-module homomorphism $\gamma : E \to E$ such that $\pi_2 = \pi_1 \gamma$. Similarly, there exists a unique A-module automorphism $\tilde{\gamma} : E \to E$ such that $\pi_1 = \pi_2 \tilde{\gamma}$ and $\gamma\tilde{\gamma} = \tilde{\gamma}\gamma = \mathrm{id}$ by the uniqueness assertion in (i). Thus γ is an A-automorphism of E. The converse statement is trivial. $\square$

Remark 4.2. The retract $\pi : E \to S$ in Proposition 4.1 is equivalent to choosing an A_0-module decomposition $E = S \oplus T$. The second part of the proposition states that any two such decompositions are related by a unique A-module automorphism of E fixing S.

Definition 4.3. Let V be a $\mathcal{P}_0$-module of graded dimension v. Define $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ to be the variety of injective $\mathcal{P}_0$-module homomorphisms $\gamma : V \to q^w$ whose image is a $\mathcal{P}$-submodule of q^w.

Theorem 4.4. *Fix $v, w \in \mathbb{N}Q_0$. Then there is a bijective GL_V-equivariant algebraic map from $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ to $\Lambda(v, w)^{\mathrm{st}}$ and a bijective algebraic map from*

$\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ *to* $\mathfrak{L}(v, w)$. *In particular*, $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ *is homeomorphic to* $\Lambda(v, w)^{\mathrm{st}}$ *and* $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ *is homeomorphic to* $\mathfrak{L}(v, w)$.

Remark 4.5. Lusztig [1998; 2000] has described a canonical bijection between the lagrangian Nakajima quiver varieties and grassmannian type varieties inside the projective modules p^w (see Section 7). In several places in the literature, it was claimed that the varieties defined by Lusztig are isomorphic (as algebraic varieties) to the lagrangian Nakajima quiver varieties. However, the authors were not aware of a proof existing in the literature. Most references for this statement were to [Lusztig 1998; 2000], where the points of the two varieties are shown to be in canonical bijection (similar to the situation in the current paper). Lusztig informed the authors that he was not aware of a proof that the varieties are isomorphic. After the appearance of an earlier version of the current paper [Savage and Tingley 2009], Shipman [2010] proved that the varieties are indeed isomorphic. From now on, we will incorporate Shipman's work, as it allows us to strengthen several results; in particular (see Corollary A.6 in the Appendix) the map $\bar{\iota}$ in the proof below is an isomorphism of algebraic varieties.

Proof of Theorem 4.4. Fix $V \in \mathcal{P}_0$-mod of graded dimension v and a $\mathcal{P}_0$-module homomorphism $\pi : q^w \to s^w$ that is the identity on s^w. We identify s^w with the W appearing in the definition of the quiver varieties. A point $\gamma \in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ defines an embedding of V into q^w, hence a $\mathcal{P}$-module structure on V satisfying the stability condition and so a point of $\Lambda(v, w)^{\mathrm{st}}$. More precisely, $\gamma \in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ corresponds to the point $(\gamma^{-1} x^w \gamma, \pi\gamma) \in \Lambda(v, w)^{\mathrm{st}}$, where x^w is the element of $\mathrm{Rep}_{\bar{Q}}\, q^w$ corresponding to the $\mathcal{P}$-module q^w. Thus we have a map

$$\iota : \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w) \to \Lambda(V, W)^{\mathrm{st}},$$

which is clearly algebraic and GL_V-equivariant. By Proposition 4.1, ι is bijective. Passing to the quotient by GL_V we also obtain a bijective algebraic map $\bar{\iota}$ from $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ to $\mathfrak{L}(v, w)$.

Now, $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ and $\mathfrak{L}(v, w)$ are both projective. By, for example, [Hartshorne 1977, Theorem 4.9 and Exercise 4.4], the image of a projective variety under an algebraic map is always closed, so $\bar{\iota}$ takes closed subsets to closed subsets. Since $\bar{\iota}$ is a bijection, this implies that $\bar{\iota}^{-1}$ is continuous. Hence $\bar{\iota}$ is a homeomorphism. Since $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ and $\Lambda(v, w)^{\mathrm{st}}$ are principal G-bundles over $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ and $\mathfrak{L}(v, w)$, the map ι also induces a homeomorphism. $\square$

Remark 4.6.

(i) The role of the retract π in Proposition 4.1 is to ensure the uniqueness of γ.

(ii) When Q is of finite type, the injective module q^w is also projective (see Proposition 7.10) and thus Theorem 4.4 follows from [Lusztig 2000, §2.1].

(iii) The isomorphisms of Theorem 4.4 depend on the choice of the retract $\pi : q^w \to s^w$. By Proposition 4.1(ii), isomorphisms coming from different retracts are related by an automorphism of q^w fixing s^w.

(iv) In Lusztig's grassmannian type realization of the lagrangian Nakajima quiver varieties [Lusztig 1998; 2000], one must require that the submodules contain all paths of large enough length (this corresponds to the nilpotency condition in the definition of the quiver varieties). In the current approach using injective modules, no such condition is required due to Lemma 2.9.

4B. ***Demazure quiver grassmannians.*** As before, let $\mathfrak{g}$ be the Kac–Moody algebra corresponding to the underlying graph of Q and let $\mathcal{W}$ be its Weyl group with Bruhat order $\preceq$.

Definition 4.7. For each $w \in \mathbb{N}Q_0$, we define an action of $\mathcal{W}$ on $\mathbb{Z}Q_0$ as follows. For $v \in \mathbb{Z}Q_0$ and $\sigma \in \mathcal{W}$, define $\sigma \cdot_w v = u$ where u is the unique element of $\mathbb{Z}Q_0$ satisfying

$$\sigma(\omega_w - \alpha_v) = \omega_w - \alpha_u.$$

We say that $v \in \mathbb{N}Q_0$ is *w-extremal* if $v \in \mathcal{W} \cdot_w 0$.

Lemma 4.8. *If $v, w \in \mathbb{N}Q_0$ and $\omega_w - \alpha_v$ is a weight of the irreducible highest weight representation of $\mathfrak{g}$ of highest weight ω_w (i.e the corresponding weight space is nonzero), then $\sigma \cdot_w v \in \mathbb{N}Q_0$ for all $\sigma \in \mathcal{W}$. In particular $\mathcal{W} \cdot_w 0 \subseteq \mathbb{N}Q_0$.*

Proof. This follows easily from the fact that $\mathcal{W}$ acts on the weights of highest weight irreducible representations and the weight multiplicities are invariant under this action. □

Proposition 4.9. *For $v \in \mathbb{N}Q_0$, the following statements are equivalent*:

(i) *v is w-extremal*,

(ii) *$\mathfrak{L}(v, w)$ consists of a single point*,

(iii) *$\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ consists of a single point*, *and*

(iv) *there is a unique submodule of q^w of graded dimension v.*

Proof. The equivalence of (i) and (ii) is given in [Savage 2006d, Proposition 5.1]. The equivalence of (ii), (iii) and (iv) follows from Theorem 4.4. □

Definition 4.10 (Demazure quiver grassmannian). For $\sigma \in \mathcal{W}$, we let $q^{w,\sigma}$ denote the unique submodule of q^w of graded dimension $\sigma \cdot_w 0$. We call $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma})$ a *Demazure quiver grassmannian*.

Proposition 4.11. *If $\sigma_1, \sigma_2 \in \mathcal{W}$ with $\sigma_1 \preceq \sigma_2$, then q^{w,σ_2} has a unique submodule of graded dimension $\sigma_1 \cdot_w 0$ and this submodule is isomorphic to q^{w,σ_1}.*

Proof. Since $\sigma_1 \preceq \sigma_2$, we have $L_{\omega_w,\sigma_1} \subseteq L_{\omega_w,\sigma_2}$, where L_{ω_w,σ_i} is the Demazure module corresponding to L_{ω_w} (the irreducible integrable highest weight $\mathfrak{g}$-module with highest weight ω_w) and σ_i. It then follows from [Savage 2006d, Theorem 7.1] that q^{w,σ_1} is (isomorphic to) a submodule of q^{w,σ_2}. Since any submodule of q^{w,σ_2} is also a submodule of q^w, uniqueness follows directly from Proposition 4.9. □

Proposition 4.12. *Fix $\sigma \in \mathcal{W}$ and $v, w \in \mathbb{N}Q_0$. Then $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma})$ is isomorphic (as an algebraic variety) to the Demazure quiver variety $\mathfrak{L}_\sigma(v, w)$.*

Proof. This follows immediately from Definitions 3.5 and 4.10 and the description of the homeomorphism $\mathrm{Gr}_{\mathcal{P}}(v, q^w) \cong \mathfrak{L}(v, w)$ given in Theorem 4.4, which is actually an isomorphism of algebraic varieties by Corollary A.6. □

Remark 4.13. Note that if Q is a quiver of finite type and σ_0 is the longest element of $\mathcal{W}$, then $\mathfrak{L}_{\sigma_0}(v, w) = \mathfrak{L}(v, w)$ and $\mathrm{Gr}(v, q^{w,\sigma_0}) = \mathrm{Gr}(v, q^w)$ for all $v, w \in \mathbb{N}Q_0$.

The $(q^{w,\sigma})_{\sigma\in\mathcal{W}}$ form a directed system under the Bruhat order. Let $\tilde{q}^w$ be the direct limit of this system.

Lemma 4.14. *Any locally nilpotent submodule V of q^w is contained in $\tilde{q}^w$.*

Proof. First note that for $n \in \mathbb{N}$, the submodule $(q^w)^{(n)} = \{v \in q^w : \mathcal{P}_{\geq n} \cdot v = 0\}$ of q^w is finite-dimensional. This follows from the fact that q^i is a submodule of $\mathrm{Hom}_{\mathbb{C}}(e_i\mathcal{P}, \mathbb{C})$ (since this is an injective module containing s^i), which has this property, and $q^w = \bigoplus_{i\in I}(q^i)^{\oplus w_i}$.

Since V is locally nilpotent, we have a filtration

$$0 = V^{(0)} \subseteq V^{(1)} = \operatorname{socle} V \subseteq V^{(2)} \subseteq \cdots$$

where $V^{(n)} = \{v \in V : \mathcal{P}_{\geq n} \cdot v = 0\}$. Local nilpotency of V ensures that $\bigcup_n V^{(n)} = V$. It suffices to show that each $V^{(n)}$ is contained in $\tilde{q}^w$. Since $V^{(n)} \subseteq (q^w)^{(n)}$, it follows that $V^{(n)}$ is finite-dimensional. Choose a linear retract $\pi : q^w \to s^w$. By Theorem 4.4, V corresponds to a point of $\mathfrak{L}(v, w)$. Choose $\sigma \in \mathcal{W}$ sufficiently large so that the $(\omega_w - \alpha_v)$-weight space of the representation L_{ω_w} is contained in the Demazure module $L_{\omega_w,\sigma}$ (we can always do this since the weight space is finite-dimensional). Then by Proposition 4.12, we have that $V \subseteq q^{w,\sigma} \subseteq \tilde{q}^w$. □

Theorem 4.15. *We have that $\tilde{q}^w$ is the injective hull of s^w in the category $\mathcal{P}$-lnMod.*

Proof. Since each $q^{w,\sigma}$ is nilpotent, it follows that $\tilde{q}^w$ is locally nilpotent and thus belongs to the category $\mathcal{P}$-lnMod. Furthermore, it is clear that $\tilde{q}^w$ has socle s^w and that it is an essential extension of s^w. It remains to show that $\tilde{q}^w$ is an injective object of $\mathcal{P}$-lnMod. Suppose M and N are locally nilpotent $\mathcal{P}$-modules and we have a homomorphism $M \to \tilde{q}^w$ and an injection $M \hookrightarrow N$. Since q^w is injective in the category of $\mathscr{P}$-modules, there exists a homomorphism $h : N \to q^w$ such that the following diagram commutes:

$$\begin{tikzcd} N \arrow[drr, "h"] & & \\ M \arrow[u, hook] \arrow[r] & \tilde{q}^w \arrow[r, hook] & q^w \end{tikzcd}$$

Since N is locally nilpotent, $h(N)$ is a locally nilpotent submodule of q^w. Therefore the map h factors through $\tilde{q}^w$ by Lemma 4.14. □

Corollary 4.16. *We have that $\tilde{q}^w \cong q^w$ if and only if Q is of finite or affine* (*tame*) *type.*

Proof. This follows immediately from Theorem 4.15 and Proposition 2.11. □

We see from the above that $\{q^{w,\sigma}\}_{\sigma \in \mathscr{W}}$ is a "rigid" filtration of $\tilde{q}^w$ (rigid in the sense of the uniqueness of submodules of the given w-extremal graded dimensions). Proposition 4.12 can be seen as a representation theoretic interpretation of this filtration. It corresponds to the filtration by Demazure modules of the irreducible highest-weight representation of $\mathfrak{g}$ of highest weight ω_w. If the quiver Q is of finite type, the Weyl group $\mathscr{W}$, and hence this filtration, is finite. Otherwise they are infinite. In the infinite case, we have a filtration of the infinite-dimensional $\tilde{q}^w$ by finite-dimensional submodules $q^{w,\sigma}$, $\sigma \in \mathscr{W}$.

5. Group actions and graded quiver grassmannians

We now define a natural $\mathrm{GL}_W \times G_{\mathscr{P}}$-action on the quiver grassmannians and show that the maps of Theorem 4.4 are equivariant. We then define graded/cyclic quiver grassmannians and show they are isomorphic to the graded/cyclic quiver varieties of Nakajima [2001, §4.1; 2004, §4].

5A. $\mathbf{GL}_{w} \times G_{\mathscr{P}}$*-action and equivariance.* Let $\mathrm{GL}_w = \mathrm{GL}_{s^w}$ and recall that $G_{\mathscr{P}}$ is the group of algebra automorphisms of $\mathscr{P}$ that fix $\mathscr{P}_0$ pointwise. For a $\mathscr{P}$-module V and $h \in G_{\mathscr{P}}$, denote by ${}^hV$ the $\mathscr{P}$-module with action given by $(a, v) \mapsto h^{-1}(a) \cdot v$. Now, fix $(g, h) \in \mathrm{GL}_w \times G_{\mathscr{P}}$ and a $\mathscr{P}_0$-module retract $\pi : q^w \to s^w$. By Proposition 4.1, there exists a unique $\mathscr{P}$-module homomorphism $\gamma_{(g,h)} : {}^hq^w \to q^w$ such that the following diagram commutes:

$$\begin{tikzcd} {}^hq^w \arrow[r, "\gamma_{(g,h)}"] \arrow[d, "\pi"'] & q^w \arrow[d, "\pi"] \\ s^w \arrow[r, "g"] & s^w \end{tikzcd}$$

The uniqueness assertion of Proposition 4.1 ensures that $\gamma_{(g,h)}$ is bijective with inverse $\gamma_{(g^{-1},h^{-1})}$. Note that since the action of $\mathcal{P}_0$ on ${}^h q^w$ and q^w is the same, $\gamma_{(g,h)}$ can be considered as a $\mathcal{P}_0$-automorphism of q^w. This defines a group homomorphism $\mathrm{GL}_w \times G_{\mathcal{P}} \to \mathrm{GL}_{q^w}$, $(g,h) \mapsto \gamma_{(g,h)}$. In other words, it defines an action of $\mathrm{GL}_w \times G_{\mathcal{P}}$ on q^w by $\mathcal{P}_0$-module automorphisms. This in turn defines an action on $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ and $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ given by

$$\begin{aligned} (g,h) \star \gamma &= \gamma_{(g,h)}\gamma, & \gamma &\in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w) \\ (g,h) \star U &= \gamma_{(g,h)}(U), & U &\in \mathrm{Gr}_{\mathcal{P}}(v, q^w). \end{aligned}$$

Proposition 5.1. *The isomorphisms of Theorem 4.4 are* $\mathrm{GL}_w \times G_{\mathcal{P}}$*-equivariant.*

Proof. Let $(x,t) \mapsto \gamma(x,t)$ be the map $\Lambda(v,w)^{\mathrm{st}} \xrightarrow{\cong} \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ of Theorem 4.4. Fix $(x,t) \in \Lambda(v,w)^{\mathrm{st}}$. Recall that for $(g,h) \in \mathrm{GL}_w \times G_{\mathcal{P}}$, we have $(g,h) \star (x,t) = (h\star x, gt)$. Let V^x be the $\mathcal{P}$-module corresponding to x. Then ${}^h V^x$ is the $\mathcal{P}$-module corresponding to $h \star x$. We have the commutative diagram

$$\begin{tikzcd} & q^w \arrow[d, "\pi"] \\ V^x \arrow[ur, "\gamma(x,t)"] \arrow[r, "t"'] & s^w \end{tikzcd}$$

It follows that the diagram

$$\begin{tikzcd} & {}^h q^w \arrow[r, "\gamma_{(g,h)}"] \arrow[d, "\pi"] & q^w \arrow[d, "\pi"] \\ {}^h V^x \arrow[ur, "\gamma(x,t)"] \arrow[r, "t"'] & s^w \arrow[r, "g"'] & s^w \end{tikzcd}$$

commutes. By the uniqueness statement in Proposition 4.1, we have

$$\gamma((g,h) \star (x,t)) = \gamma(h \star x, gt) = \gamma_{(g,h)}\gamma(x,t) = (g,h) \star \gamma(x,t),$$

which proves that the map $\Lambda(v,w)^{\mathrm{st}} \cong \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ is equivariant. The remaining claim follows from the fact that the isomorphism $\mathfrak{L}(v,w) \cong \mathrm{Gr}_{\mathcal{P}}(v, q^w)$ is obtained from the map $\Lambda(v,w)^{\mathrm{st}} \xrightarrow{\cong} \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ by taking quotients by GL_V. □

5B. ***Graded/cyclic quiver grassmannians.*** Fix an abelian reductive subgroup A and a group homomorphism $\rho : A \to \mathrm{GL}_w \times G_{\mathcal{P}}$, defining an action of A on q^w by $\mathcal{P}_0$-module automorphisms. The weight space corresponding to $\lambda \in \mathrm{Hom}(A, \mathbb{C}^*)$ is

$$q^w(\lambda) \stackrel{\mathrm{def}}{=} \{v \in q^w \mid \rho(a)(v) = \lambda(a)v \ \ \forall a \in A\}. \tag{5-1}$$

We define

$$\begin{aligned}\mathrm{Gr}_{\mathcal{P}}(q^w)^A &= \{U \in \mathrm{Gr}_{\mathcal{P}}(q^w) \mid \ \rho(a) \star U = U \ \ \forall a \in A\},\\ \mathrm{Gr}_{\mathcal{P}}(u, q^w)^A &= \mathrm{Gr}_{\mathcal{P}}(q^w)^A \cap \mathrm{Gr}_{\mathcal{P}}(u, q^w).\end{aligned}$$

Then for all $U \in \mathrm{Gr}_{\mathcal{P}}(q^w)^A$, we have the map $\rho_U : A \to \mathrm{GL}_U$, $a \mapsto \rho(a)|_U$. In other words, ρ_U is a representation of A in the category of $\mathcal{P}_0$-modules. If ρ_1 and ρ_2 are two such representations, we write $\rho_1 \cong \rho_2$ when ρ_1 and ρ_2 are isomorphic. That is, $\rho_1 \cong \rho_2$ for $\rho_i : A \to \mathrm{GL}_{U_i}$, if there exists a $\mathcal{P}_0$-module isomorphism $\xi : U_1 \to U_2$ such that $\rho_2 = \xi\rho_1\xi^{-1}$, where $\xi\rho_U\xi^{-1}$ denotes the homomorphism $a \mapsto \xi\rho_U(a)\xi^{-1}$. Then, for $\rho_1 : A \to \mathrm{GL}_U$, U a $\mathcal{P}_0$-module, we define

$$\mathrm{Gr}_{\mathcal{P}}(\rho_1, q^w)^A = \{U' \in \mathrm{Gr}_{\mathcal{P}}(q^w)^A \mid \rho_{U'} \cong \rho_1\}.$$

Note that $\mathrm{Gr}_{\mathcal{P}}(\rho_1, q^w)^A$ depends only on the isomorphism class of ρ_1.

Recall the action of $\mathrm{GL}_w \times G_{\mathcal{P}}$ on $\Lambda(V, W)^{\mathrm{st}}$ and $\mathfrak{L}(v, w)$ described in Section 3B (where we now identify W with s^w, $w = \dim_{Q_0} W$). Define

$$\begin{aligned}\mathfrak{L}(w)^A &= \{[x, t] \in \mathfrak{L}(v, w) \mid \rho(a) \star [x, t] = [x, t] \ \forall\, a \in A\},\\ \mathfrak{L}(v, w)^A &= \mathfrak{L}(w)^A \cap \mathfrak{L}(v, w).\end{aligned}$$

Fix a point $[x, t] \in \mathfrak{L}(v, w)^A$. For every $a \in A$, there exists a unique $\rho_1(a) \in \mathrm{GL}_V$ such that

$$\rho(a) \star (x, t) = \rho_1^{-1}(a) \cdot (x, t), \tag{5-2}$$

and the map $\rho_1 : A \to \mathrm{GL}_V$ is a homomorphism. Let $\mathfrak{L}(\rho_1, w)^A \subseteq \mathfrak{L}(v, w)^A$ be the set of A-fixed points y such that (5-2) holds for some representative (x, t) of y.

Theorem 5.2. *Let V be a $\mathcal{P}_0$-module and $\rho_1 : A \to \mathrm{GL}_V$ a group homomorphism. Then $\mathrm{Gr}_{\mathcal{P}}(\rho_1, q^w)^A$ is isomorphic to $\mathfrak{L}(\rho_1, w)^A$ as an algebraic variety.*

Proof. Choose $[x, t] \in \mathfrak{L}(\rho_1, w)^A$. Let $U = \gamma(x, t)(V)$ be the corresponding point of $\mathrm{Gr}_{\mathcal{P}}(v, q^w)^A$. We want to show that $\rho_1 \cong \rho_U$. Let $(g, h) \in A$ and consider the commutative diagram

$$\begin{tikzcd}
 & {}^h q^w \arrow[rr, "\gamma_{(g,h)}"] \arrow[d, "\pi"] & & q^w \arrow[d, "\pi"'] & \\
{}^h V^x \arrow[ur, "\gamma(x,t)"] \arrow[r, "t"'] & s^w \arrow[rr, "g"'] & & s^w & V \arrow[l, "t"] \arrow[ul, "\gamma(x,t)"']
\end{tikzcd}\ .$$

Then $\rho_U(g, h) = \gamma_{(g,h)}|_U$. Note that $\gamma(x, t)$ is an isomorphism when its codomain is restricted to U and we denote by $\gamma(x, t)^{-1}$ the inverse of this restriction. We claim that $\rho_1 = \tilde{\rho} \overset{\text{def}}{=} \gamma(x,t)^{-1}\left(\gamma_{(g,h)}|_U\right)\gamma(x,t)$. It suffices to show that

$$(h \star x, gt) = (g,h) \star (x,t) = \tilde{\rho}^{-1} \cdot (x,t) = (\tilde{\rho}^{-1} x \tilde{\rho}, t\tilde{\rho}).$$

We have

$$\begin{aligned}
\tilde{\rho}^{-1}x &= \gamma(x,t)^{-1}(\gamma_{(g,h)}|_U)^{-1}\gamma(x,t)x \\
&= \gamma(x,t)^{-1}(\gamma_{(g,h)}|_U)^{-1}x\gamma(x,t) \\
&= \gamma(x,t)^{-1}(h \star x)(\gamma_{(g,z)}|_U)^{-1}\gamma(x,t) \\
&= (h \star x)\gamma(x,t)^{-1}(\gamma_{(g,z)}|_U)^{-1}\gamma(x,t) \\
&= (h \star x)\tilde{\rho}^{-1},
\end{aligned}$$

so $\tilde{\rho}^{-1}x\tilde{\rho} = h \star x$. Similarly, $t\tilde{\rho} = t\gamma(x,t)^{-1}\left(\gamma_{(g,h)}|_U\right)\gamma(x,t) = gt$ and we are done. □

We now restrict to a special case of this construction that has been studied by Nakajima. In particular, we define $\mathrm{GL}_w \times \mathbb{C}^*$-actions on the quiver grassmannians corresponding to the actions on quiver varieties described in Section 3B.

For any function $m : \tilde{Q}_1 \to \mathbb{Z}$ such that $m(a) = -m(\bar{a})$ for all $a \in \tilde{Q}_1$, the group homomorphism (3-1) defines a $\mathrm{GL}_w \times \mathbb{C}^*$-action on q^w, $\widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w)$ and $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ which we again denote by $\star_m$. If A is any abelian reductive subgroup of $\mathrm{GL}_w \times \mathbb{C}^*$, we can consider the weight decompositions as above. For the remainder of this section, we fix $m = m_2$ (see Section 3B). That is, $m(a) = 0$ for all $a \in Q_1$. We also write $\star$ for $\star_m$. Recall the definition (5-1) of $q^w(\lambda)$. For $x \in \mathcal{P}_n$, $v \in q^w(\lambda)$ and $(g,z) \in A$, we have

$$\rho(g,z)(x \cdot v) = \gamma_{(zg, h_m(z))}(x \cdot v) = z^{-n} x \cdot \gamma_{(zg, h_m(z))}(v) = z^{-n}\lambda(g,z)v.$$

Thus $\mathcal{P}_n : q^w(\lambda) \to q^w(l^{-n}\lambda)$, where we write $l^{-n}\lambda$ for the element $L(-n) \otimes \lambda$ of $\mathrm{Hom}(A, \mathbb{C}^*)$ and $L(-n) = \mathbb{C}$ with $\mathbb{C}^*$-module structure given by $z \cdot v = z^{-n}v$.

Now let (g,z) be a semisimple element of A and define

$$\begin{aligned}
\mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)} &= \{U \in \mathrm{Gr}_{\mathcal{P}}(q^w) \mid (g,z) \star U = U\}, \\
\mathrm{Gr}_{\mathcal{P}}(u, q^w)^{(g,z)} &= \mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)} \cap \mathrm{Gr}_{\mathcal{P}}(u, q^w).
\end{aligned}$$

The module q^w has an eigenspace decomposition with respect to the action of (g,z) given by

$$q^w = \bigoplus_{a \in \mathbb{C}^*} q^w(a), \quad q^w(a) = \{v \in q^w \mid (g,z) \star v = av\}.$$

Then $\mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)}$ consists of those $U \in \mathrm{Gr}_{\mathcal{P}}(q^w)$ that are direct sums of subspaces of the weight spaces $q^w(a)$, $a \in \mathbb{C}^*$. Thus, each $U \in \mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)}$ inherits a weight space decomposition, or $\mathbb{C}^*$-grading,

$$U = \bigoplus_{a\in\mathbb{C}^*} U(a), \qquad U(a) = \{v \in U \mid (g,z)\star v = av\}.$$

As above we see that $\mathcal{P}_n : q^w(a) \to q^w(az^{-n})$ and $\mathcal{P}_n : U(a) \to U(az^{-n})$. We also regard s^w as an A-module via the composition

$$A \hookrightarrow \mathrm{GL}_w \times \mathbb{C}^* \xrightarrow{\text{projection}} \mathrm{GL}_w = \mathrm{GL}_{s^w}.$$

Thus s^w also inherits a $\mathbb{C}^*$-grading as above. For a $Q_0 \times \mathbb{C}^*$-graded vector space $V = \bigoplus_{i\in Q_0,\, a\in\mathbb{C}^*} V_{i,a}$, define the graded dimension (or character)

$$\operatorname{char} V = \sum_{\substack{i\in Q_0,\\ a\in\mathbb{C}*}} (\dim V_{i,a}) X_{i,a} \in \mathbb{N}[X_{i,a}]_{i\in Q_0,\, a\in\mathbb{C}^*}.$$

Recall that a $\mathcal{P}_0$-module is equivalent to an Q_0-graded vector space. Thus q^w, s^w, and elements of $\mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)}$ have natural $Q_0 \times \mathbb{C}^*$-gradings and we can consider their graded dimensions.

Definition 5.3 (graded/cyclic quiver grassmannian). For a graded dimension $\mathbf{d} \in \mathbb{N}[X_{i,a}]_{i\in Q_0,\, a\in\mathbb{C}^*}$, define

$$\mathrm{Gr}_{\mathcal{P}}(\mathbf{d}, q^w)^{(g,z)} = \{U \in \mathrm{Gr}_{\mathcal{P}}(q^w)^{(g,z)} \mid \operatorname{char} U = \mathbf{d}\}.$$

We call $\mathrm{Gr}_{\mathcal{P}}(\mathbf{d}, q^w)^{(g,z)}$ a *cyclic quiver grassmannian* if z is a root of unity, and a *graded quiver grassmannian* otherwise.

Theorem 5.4. *Let V be a $Q_0 \times \mathbb{C}^*$-graded vector space. For a semisimple element $(g,z) \in \mathrm{GL}_w \times \mathbb{C}^*$, the graded/cyclic quiver grassmannian $\mathrm{Gr}_{\mathcal{P}}(\operatorname{char} V, q^w)^{(g,z)}$ is isomorphic to the lagrangian graded/cylic quiver variety $\mathfrak{L}^\bullet(V, s^w)$ defined in* [Nakajima 2004, §4], *where s^w is considered as a $Q_0 \times \mathbb{C}^*$-graded vector space as above.*

Proof. This follows immediately from Proposition 5.1 since $\mathfrak{L}^\bullet(V, W)$ is simply the set of points of $\mathfrak{L}(V, W)$ fixed by a semisimple element (g,z) of $\mathrm{GL}_w \times \mathbb{C}^*$. □

Remark 5.5. Nakajima [2004] assumes the quiver Q is of ADE type. However, the definitions in §4 of that article extend naturally to the more general case.

6. Geometric construction of representations of Kac–Moody algebras and compatibility with nested quiver grassmannians

Since certain quiver grassmannians are isomorphic to lagrangian Nakajima quiver varieties, one can translate Nakajima's geometric construction of representations of Kac–Moody algebras into the quiver grassmannian setting. Having done this, one sees that the quiver grassmannian construction is compatible with a natural nesting of these varieties — a property which seems to have no analog in the setting of quiver varieties. One benefit of this nesting compatibility is that it allows one to always work with quiver grassmannians in *finite-dimensional* modules, even though the injective objects q^w themselves may be infinite-dimensional (outside of finite type).

For the remainder of this section, we fix a Kac–Moody algebra $\mathfrak{g}$ with symmetric Cartan matrix and let $\mathcal{W}$ be its Weyl group. Let $Q = (Q_0, Q_1)$ be a quiver whose underlying graph is the Dynkin graph of $\mathfrak{g}$ and let $\mathcal{P} = \mathcal{P}(Q)$ denote the corresponding path algebra. We also fix a $\mathcal{P}_0$-module retract $\pi : q^w \to s^w$, allowing us to identify $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ with $\mathfrak{L}(v, w)$ as in Theorem 4.4.

6A. ***Constructible functions.*** Recall that for a topological space X, a *constructible set* is a subset of X that is obtained from open sets by a finite number of the usual set theoretic operations (complement, union and intersection). A *constructible function* on X is a function that is a finite linear combination of characteristic functions of constructible sets. For a complex variety X, let $M(X)$ denote the $\mathbb{C}$-vector space of constructible functions on X with values in $\mathbb{C}$. We define $M(\varnothing) = 0$. For a continuous map $p : X \to X'$, define

$$p^* : M(X') \to M(X), \quad (p^* f')(x) = f'(p(x)), \quad f' \in M(X')$$

and

$$p_! : M(X) \to M(X'), \quad (p_! f)(x) = \sum_{a \in \mathbb{Q}} a\chi(p^{-1}(x) \cap f^{-1}(a)), \quad f \in M(X),$$

where χ denotes the Euler characteristic of cohomology with compact support.

Lemma 6.1. *Suppose X is a constructible subset of a topological space Y and let $\iota : X \hookrightarrow Y$ be the inclusion map. Then*

(i) $\iota^*(f) = f|_X$ *for* $f \in M(Y)$, *and*

(ii) *for* $f \in M(X)$, $\iota_!(f)$ *is the extension of* f *by zero. That is,*

$$\iota_!(f)(x) = \begin{cases} f(x) & \text{if } x \in X, \\ 0 & \text{if } x \in Y \setminus X. \end{cases}$$

The proof is straightforward and will be omitted.

6B. ***Raising and lowering operators.*** Let V be a $\mathcal{P}$-module. For $u, u' \in \mathbb{N}Q_0$ with $u \le u'$ (i.e., $u = \sum u_i i$ and $u' = \sum u'_i i$ where $u_i \le u'_i$ for all $i \in Q_0$), define

$$\mathrm{Gr}_{\mathcal{P}}(u, u', V) = \{(U, U') \in \mathrm{Gr}_{\mathcal{P}}(u, V) \times \mathrm{Gr}_{\mathcal{P}}(u', V) \mid U \subseteq U'\}, \tag{6-1}$$

and let

$$\mathrm{Gr}_{\mathcal{P}}(u, V) \xleftarrow{\pi_1} \mathrm{Gr}_{\mathcal{P}}(u, u', V) \xrightarrow{\pi_2} \mathrm{Gr}_{\mathcal{P}}(u', V)$$

be the natural projections given by $\pi_1(U, U') = U$ and $\pi_2(U, U') = U'$. For each $i \in I$, define the operators

$$\begin{aligned} \hat{E}_i &: \quad M(\mathrm{Gr}_{\mathcal{P}}(u+i, V)) \to M(\mathrm{Gr}_{\mathcal{P}}(u, V)), \quad \hat{E}_i f = (\pi_1)_!(\pi_2^* f), \\ \hat{F}_i &: \quad M(\mathrm{Gr}_{\mathcal{P}}(u, V)) \to M(\mathrm{Gr}_{\mathcal{P}}(u+i, V)), \quad \hat{F}_i f = (\pi_2)_!(\pi_1^* f), \end{aligned} \tag{6-2}$$

where the maps π_1 and π_2 are as in (6-1) with $u' = u + i$.

6C. *Compatibility with nested quiver grassmannians.* Suppose $V_1 \subseteq V_2$ are $\mathcal{P}$-modules. Then we have the commutative diagram

$$\begin{tikzcd} \mathrm{Gr}_{\mathcal{P}}(u, V_1) \arrow[d, hook, "\iota_u"] & \mathrm{Gr}_{\mathcal{P}}(u, u', V_1) \arrow[l, "\pi_1^1"'] \arrow[r, "\pi_2^1"] \arrow[d, hook, "\iota_{u,u'}"'] & \mathrm{Gr}_{\mathcal{P}}(u', V_1) \arrow[d, hook, "\iota_{u'}"'] \\ \mathrm{Gr}_{\mathcal{P}}(u, V_2) & \mathrm{Gr}_{\mathcal{P}}(u, u', V_2) \arrow[l, "\pi_1^2"'] \arrow[r, "\pi_2^2"] & \mathrm{Gr}_{\mathcal{P}}(u', V_2) \end{tikzcd}$$

where ι_u, $\iota_{u'}$ and $\iota_{u,u'}$ denote the canonical inclusions. Denote by $\hat{E}_i^j$ and $\hat{F}_i^j$, $j = 1, 2$, the operators defined in (6-2) for $V = V_j$.

Proposition 6.2. *We have*

(i) $\hat{E}_i^1 = \iota_u^* \circ \hat{E}_i^2 \circ (\iota_{u+i})_!$, *and*

(ii) $\hat{F}_i^1 = \iota_{u+i}^* \circ \hat{F}_i^2 \circ (\iota_u)_!$.

Proof. Let $u' = u + i$. By linearity, it suffices to prove the first statement for functions of the form 1_X where X is a constructible subset of $\mathrm{Gr}_{\mathcal{P}}(u', V_1)$. Then $(\iota_{u'})_! 1_X = 1_X$, where on the right-hand side, X is viewed as a subset of $\mathrm{Gr}_{\mathcal{P}}(u', V_2)$. We have

$$(\pi_2^2)^* \circ (\iota_{u'})_! 1_X = (\pi_2^2)^* 1_X = 1_{(\pi_2^2)^{-1}(X)}$$

and

$$(\iota_{u,u'})_! (\pi_2^1)^* 1_X = (\iota_{u,u'})_! 1_{(\pi_2^1)^{-1}(X)} = 1_{(\pi_2^1)^{-1}(X)}.$$

Since $X \subseteq \mathrm{Gr}_{\mathcal{P}}(u', V_1)$, we have $(\pi_2^2)^{-1}(X) = (\pi_2^1)^{-1}(X)$ and thus

$$(\pi_2^2)^* \circ (\iota_{u'})_! 1_X = (\iota_{u,u'})_! \circ (\pi_2^1)^* 1_X.$$

Therefore

$$\begin{aligned}
\iota_u^* \circ \hat{E}_i^2 \circ (\iota_{u'})_! 1_X &= \iota_u^* \circ (\pi_1^2)_! \circ (\pi_2^2)^* \circ (\iota_{u'})_! 1_X \\
&= \iota_u^* \circ (\pi_1^2)_! \circ (\iota_{u,u'})_! \circ (\pi_2^1)^* 1_X \\
&= \iota_u^* \circ (\pi_1^2 \circ \iota_{u,u'})_! \circ (\pi_2^1)^* 1_X \\
&= \iota_u^* \circ (\iota_u \circ \pi_1^1)_! \circ (\pi_2^1)^* 1_X \\
&= \iota_u^* \circ (\iota_u)_! \circ (\pi_1^1)_! \circ (\pi_2^1)^* 1_X \\
&= (\pi_1^1)_! \circ (\pi_2^1)^* 1_X \\
&= \hat{E}_i^1 1_X,
\end{aligned}$$

where the sixth equality holds since $\iota_u^* \circ (\iota_u)_!$ is the identity on $M(\mathrm{Gr}_{\mathcal{P}}(u, V_1))$.

We now prove the second statement. Again, it suffices to prove it for functions of the form 1_X where X is a constructible subset of $\mathrm{Gr}_{\mathcal{P}}(u, V_1)$. Now, for $U \in \mathrm{Gr}_{\mathcal{P}}(u', V_1)$, we have

$$\begin{aligned}
\iota_{u'}^* \circ \hat{F}_i^2 \circ (\iota_u)_! 1_X(U) &= \iota_{u'}^* \circ (\pi_2^2)_! \circ (\pi_1^2)^* \circ (\iota_u)_! 1_X(U) \\
&= \iota_{u'}^* \circ (\pi_2^2)_! \circ (\pi_1^2)^* 1_X(U) \\
&= \iota_{u'}^* \circ (\pi_2^2)_! \circ 1_{(\pi_1^2)^{-1}(X)}(U) \\
&= \chi\left((\pi_2^2)^{-1}(U) \cap (\pi_1^2)^{-1}(X)\right) \\
&= \chi\left((\pi_2^1)^{-1}(U) \cap (\pi_1^1)^{-1}(X)\right) \\
&= (\pi_2^1)_! 1_{(\pi_1^1)^{-1}(X)}(U) \\
&= (\pi_2^1)_! \circ (\pi_1^1)^* 1_X(U) \\
&= \hat{F}_i^1 1_X(U),
\end{aligned}$$

where the fifth equality holds since $U \in \mathrm{Gr}_{\mathcal{P}}(u', V_1)$. □

It follows from Proposition 4.12 that the Demazure quiver grassmannians stabilize in the following sense.

Corollary 6.3. *For $u, w \in \mathbb{N}Q_0$, there exists $\sigma \in \mathcal{W}$, such that $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma'})$ is isomorphic to $\mathfrak{L}(v, w)$ for all $\sigma' \succeq \sigma$.*

Proof. It follows from [Savage 2006d, Proposition 6.1] that there exists a $\sigma \in \mathcal{W}$ such that $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma}) \cong \mathfrak{L}_\sigma(v, w) = \mathfrak{L}(v, w)$. It follows from the same proposition that for $\sigma' \succeq \sigma$, we have $\mathfrak{L}_{\sigma'}(v, w) = \mathfrak{L}(v, w)$. The result then follows from Proposition 4.12. □

Corollary 6.4. *For $v, w \in \mathbb{N}Q_0$, let $\sigma^{v,w} \in \mathcal{W}$ be minimal among the $\sigma \in \mathcal{W}$ such that $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma})$ is isomorphic to $\mathfrak{L}(v, w)$. Then $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma}) \cong \mathrm{Gr}_{\mathcal{P}}(v, q^w)$ for all $\sigma \succeq \sigma^{v,w}$. In particular, every submodule of the injective module q^w of graded dimension v is a submodule of $q^{w,\sigma}$ for $\sigma \succeq \sigma^{v,w}$.*

Remark 6.5. In the case when $\mathfrak{g}$ is of finite type, we can take $\sigma = \sigma_0$, where σ_0 is the longest element of the Weyl group. Then $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ is isomorphic to $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma_0})$ for all $v \in \mathbb{N}Q_0$.

Lemma 6.6. *Suppose $w, v, v' \in \mathbb{N}Q_0$ with $v \leq v'$ and $\sigma \in \mathcal{W}$. Then the diagram*

$$\begin{tikzcd}
\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma}) \arrow[d, hook] & \mathrm{Gr}_{\mathcal{P}}(v, v', q^{w,\sigma}) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] \arrow[d, hook] & \mathrm{Gr}_{\mathcal{P}}(v', q^{w,\sigma}) \arrow[d, hook] \\
\mathrm{Gr}_{\mathcal{P}}(v, q^{w}) & \mathrm{Gr}_{\mathcal{P}}(v, v', q^{w}) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] & \mathrm{Gr}_{\mathcal{P}}(v', q^{w})
\end{tikzcd}$$

commutes, where the vertical arrows are the natural inclusions. If $\sigma \succeq \sigma^{v,w}, \sigma^{v',w}$, then the vertical arrow are isomorphisms.

Proof. This follows immediately from Corollary 6.4. □

6D. *Quiver grassmannian realization of representations.* For each $i \in I$, define

$$H_i : M(\mathrm{Gr}_{\mathcal{P}}(v, q^w)) \to M(\mathrm{Gr}_{\mathcal{P}}(v, q^w)), \quad H_i f = (w - Cv)_i f, \tag{6-3}$$

where C is the Cartan matrix of $\mathfrak{g}$. Also, in the special case when $V = q^w$ for some w, we denote the operators $\hat{E}_i$ and $\hat{F}_i$ by E_i and F_i respectively.

Proposition 6.7. *The operators E_i, F_i, H_i define an action of $\mathfrak{g}$ on*

$$\bigoplus_u M(\mathrm{Gr}_{\mathcal{P}}(u, q^w)).$$

Proof. Throughout this proof, for varieties X and Y, the notation $X \cong Y$ means that X and Y are homeomorphic. In [Nakajima 1994, §10], Nakajima defines the variety

$$\mathfrak{F}(v, w; i) \stackrel{\text{def}}{=} \tilde{\mathfrak{F}}(v, w; i)/\mathrm{GL}_V,$$

where

$$\tilde{\mathfrak{F}}(v, w; i) = \{(x, t, Z) \mid (x, t) \in \Lambda(V, W)^{\mathrm{st}},\ Z \subseteq V,\ x(Z) \subseteq Z,\ \dim Z = v - i\}.$$

Using the homeomorphism of Theorem 4.4, we have

$$\tilde{\mathfrak{F}}(v, w; i) \cong \{(\gamma, Z) \mid \gamma \in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w),\ Z \subseteq V,\ \dim Z = v - i,\ \mathcal{P} \cdot \gamma(Z)) \subseteq \gamma(Z)\}.$$

The map from the set

$$\big\{(\gamma, Z) \mid \gamma \in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w),\ Z \subseteq V,\ \dim Z = v - i,\ \mathcal{P} \cdot \gamma(Z) \subseteq \gamma(Z)\big\}$$

into $\mathrm{Gr}_{\mathcal{P}}(v - i, v, q^w)$ given by

$$(\gamma, Z) \mapsto (\gamma(Z), \gamma(V))$$

is a principal GL_V-bundle and thus

$$\begin{aligned}
&\mathfrak{F}(v, w; i)\\
&\quad = \tilde{\mathfrak{F}}(v, w; i)/\mathrm{GL}_V\\
&\quad \cong \left\{(\gamma, Z) \mid \gamma \in \widehat{\mathrm{Gr}}_{\mathcal{P}}(v, q^w),\ Z \subseteq V,\ \dim Z = v - i,\ \mathcal{P} \cdot \gamma(Z) \subseteq \gamma(Z)\right\}/\mathrm{GL}_V\\
&\quad = \mathrm{Gr}_{\mathcal{P}}(u - i, u, q^w).
\end{aligned}$$

Therefore, the following diagram commutes:

$$\begin{tikzcd}
\mathrm{Gr}_{\mathcal{P}}(v-i, q^w) \arrow[d, "\cong"] & \mathrm{Gr}_{\mathcal{P}}(v-i, v, q^w) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] \arrow[d, "\cong"] & \mathrm{Gr}_{\mathcal{P}}(v, q^w) \arrow[d, "\cong"] \\
\mathfrak{L}(v-i, w) & \mathfrak{F}(v, w; i) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] & \mathfrak{L}(v, w)
\end{tikzcd} \tag{6-4}$$

where the maps π_1 and π_2 appearing on the bottom row are described in §10 of [Nakajima 1994]. The result then follows immediately from Proposition 10.12 of the same reference. □

Let $U(\mathfrak{g})^-$ be the lower half of the enveloping algebra of $\mathfrak{g}$. Let α be the constant function on $\mathrm{Gr}_{\mathcal{P}}(0, q^w)$ with value 1 and let

$$L_w \stackrel{\text{def}}{=} U(\mathfrak{g})^- \cdot \alpha \subseteq \bigoplus_v M(\mathrm{Gr}_{\mathcal{P}}(v, q^w)), \tag{6-5}$$

$$L_w(v) \stackrel{\text{def}}{=} M(\mathrm{Gr}_{\mathcal{P}}(v, q^w)) \cap L_w. \tag{6-6}$$

Theorem 6.8. *The operators E_i, F_i, H_i preserve L_w and L_w is isomorphic to the irreducible highest-weight integrable representation of $\mathfrak{g}$ with highest weight ω_w. The summand $L_w(v)$ in the decomposition $L_w = \bigoplus_v L_w(v)$ is a weight space with weight $\omega_w - \alpha_v$.*

Proof. In light of the commutative diagram (6-4), the result follows immediately from [Nakajima 1994, Theorem 10.14]. □

Remark 6.9. It follows from Proposition 6.2 and Lemma 6.6 that we can always work with $\mathrm{Gr}_{\mathcal{P}}(v, q^{w,\sigma})$ for large enough σ. Therefore, we can avoid quiver grassmannians in infinite-dimensional injectives if desired.

From the realization of irreducible highest-weight representations given in Theorem 6.8, we obtain some natural automorphisms of these representations. Recall from Definition 2.16 the natural action of $\mathrm{Aut}_{\mathcal{P}}\, q^w$ on $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ for any v given by $(g, V) \mapsto g(V)$. This induces an action on $\bigoplus_v M(\mathrm{Gr}_{\mathcal{P}}(v, q^w))$ given by

$$(g, f) \mapsto f \circ g^{-1}, \quad f \in \bigoplus_v M(\mathrm{Gr}_{\mathcal{P}}(v, q^w)), \quad g \in \mathrm{Aut}_{\mathcal{P}}\, q_w.$$

This action clearly commutes with the operators E_i and F_i and thus induces an action on L_w. Such actions do not seem to be clear in the original quiver variety picture. Similar actions were considered in [Lusztig 2000, §1.22] in the case when Q is of finite type.

7. Relation to Lusztig's grassmannian realization

Lusztig [1998; 2000] gave a grassmannian type realization of the lagrangian Nakajima quiver varieties inside the projective modules p^w. In the case when Q is a quiver of finite type, the injective hulls of the simple objects are also projective covers (of different simple objects). Thus, Lusztig's and our construction are closely related. In this section, we extend Lusztig's construction to give a realization of the Demazure quiver varieties. We then give a precise relationship between his construction and ours in the finite type case. We will see that the natural identification of the two constructions corresponds to the Chevalley involution on the level of representations of the Lie algebra $\mathfrak{g}$ associated to our quiver.

7A. *Lusztig's construction and Demazure quiver varieties.*

Definition 7.1. For $V \in \mathcal{P}$-Mod, define

$$\tilde{\mathrm{Gr}}_{\mathcal{P}}(V) = \{U \in \mathrm{Gr}_{\mathcal{P}}(V) \mid \mathcal{P}_n \cdot V \subseteq U \text{ for some } n \in \mathbb{N}\}.$$

In other words, $\tilde{\mathrm{Gr}}_{\mathcal{P}}(V)$ consists of all $\mathcal{P}$-submodules of V such that the quotient V/U is nilpotent. For $u \in \mathbb{N}Q_0$, we define

$$\tilde{\mathrm{Gr}}_{\mathcal{P}}(u, V) = \{U \in \tilde{\mathrm{Gr}}_{\mathcal{P}}(V) \mid \dim_{Q_0}(V/U) = u\}.$$

Proposition 7.2. *Fix $v, w \in \mathbb{N}Q_0$. Then $\mathfrak{L}(v, w)$ is isomorphic to $\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$ as an algebraic variety.*

Proof. This is proven in Corollary 3.2 of [Shipman 2010]. Note that, in that article, a different stability condition is used in the definition of $\mathfrak{L}(v, w)$. However, it is well-known that the different stability conditions give rise to isomorphic varieties. We refer the reader to [Nakajima 1996] for a discussion of various stability conditions. □

Proposition 7.3. *For $v \in \mathbb{N}Q_0$, the following statements are equivalent*:

(i) *v is w-extremal.*

(ii) *$\mathfrak{L}(v, w)$ consists of a single point.*

(iii) *$\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$ consists of a single point.*

(iv) *There is a unique $\mathcal{P}$-submodule V of p^w of codimension v such that p^w/V is nilpotent.*

Proof. The equivalence of (i) and (ii) is given in [Savage 2006d, Proposition 5.1]. The equivalence of (ii) and (iii) follows from Proposition 7.2. Finally, the equivalence of (iii) and (iv) follows directly from Definition 7.1 □

Definition 7.4. For $\sigma \in \mathcal{W}$, we let $p^{w,\sigma}$ denote the unique submodule of p^w of graded codimension $\sigma \cdot_w 0$ and define

$$\tilde{\mathrm{Gr}}_{Q,\sigma}(v, p^w) = \{V \in \tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w) \mid p^{w,\sigma} \subseteq V\}.$$

Proposition 7.5. *Fix $\sigma \in \mathcal{W}$ and $v, w \in \mathbb{N}Q_0$. Then $\tilde{\mathrm{Gr}}_{Q,\sigma}(v, p^w)$ is isomorphic to the Demazure quiver variety $\mathfrak{L}_\sigma(v, w)$.*

Proof. This follows directly from Definitions 3.5 and 7.4 and Proposition 7.2. □

7B. ***Relation between the projective and injective constructions.*** We now suppose Q is of finite type and let $\mathfrak{g}$ be the Kac–Moody algebra whose Dynkin diagram is the underlying graph of Q. Let σ_0 be the longest element of the Weyl group of $\mathfrak{g}$. There is a unique Dynkin diagram automorphism θ such that $-w_0(\alpha_i) = \alpha_{\theta(i)}$. Extend θ to an automorphism of the root lattice $\bigoplus_{i \in Q_0} \mathbb{Z}\alpha_i$ by linearly extending the map $\alpha_i \mapsto \alpha_{\theta(i)}$. We also have an involution of $\mathbb{N}Q_0$ given by $w \mapsto \theta(w)$ where $\theta(w)_i = w_{\theta(i)}$.

Definition 7.6 (Chevalley involution). The *Chevalley involution* ζ of $\mathfrak{g}$ is given by

$$\zeta(E_i) = F_i, \quad \zeta(F_i) = E_i, \quad \zeta(H_i) = -H_i.$$

For any representation V of $\mathfrak{g}$, let ${}^\zeta V$ be the representation with the same underlying vector space as V, but with the action of $\mathfrak{g}$ twisted by ζ. More precisely, the $\mathfrak{g}$-action on ${}^\zeta V$ is given by $(a, v) \mapsto \zeta(a) \cdot v$.

For a dominant weight λ of $\mathfrak{g}$, let L_λ denote the corresponding irreducible highest-weight representation and let v_λ be a highest weight vector. Recall that an isomorphism of irreducible representations is uniquely determined by the image of v_λ. The following lemma is well known.

Lemma 7.7. *The lowest weight of L_λ is $\sigma_0(\lambda) = -\theta(\lambda)$. If $v_{-\theta(\lambda)}$ denotes a lowest weight vector, then the map $v_\lambda \mapsto v_{-\theta(\lambda)}$ induces an isomorphism ${}^\zeta L_\lambda \cong L_{\theta(\lambda)}$.*

Lemma 7.8. *We have* $\dim_{Q_0} p^w = \dim_{Q_0} q^w = \sigma_0 \cdot_w 0$.

Proof. Since the lowest weight of the representation $L(w)$ is $\sigma_0(w)$, the result follows immediately from Theorem 4.4 and Proposition 7.2. □

Lemma 7.9. *For $w \in \mathbb{N}Q_0$, we have $\sigma_0 \cdot_w 0 = \sigma_0 \cdot_{\theta(w)} 0$. Furthermore, $\theta(\sigma_0 \cdot_w 0) = \sigma_0 \cdot_w 0$.*

Proof. Let $v = \sigma_0 \cdot_w 0$. Then $\alpha_v = \omega_w - \sigma_0(\omega_w) = \omega_w + \theta(\omega_w)$ and the results follow easily from the fact that $\theta^2 = \mathrm{Id}$. □

Proposition 7.10. *If Q is a quiver of finite type and $w \in \mathbb{N}Q_0$, then $p^w \cong q^{\theta(w)}$.*

Proof. Since $p^w = \bigoplus_{i \in Q_0} (p^i)^{\oplus w_i}$ and $q^w = \bigoplus_{i \in Q_0} (q^i)^{\oplus w_i}$, it suffices to prove the result for w equal to i for arbitrary $i \in Q_0$.

Let $v = \sigma_0 \cdot_w 0 = \dim_{Q_0} p^i$. In the geometric realization of crystals via quiver varieties [Saito 2002], the point $\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w) \cong \mathfrak{L}(v, w)$ corresponds to the lowest weight element of the crystal B_{ω_i}. The lowest weight of the representation L_{ω_i} is $\sigma_0(\omega_i) = -\omega_{\theta(i)}$. Therefore, it follows from the geometric description of the crystals that $\dim_{Q_0} \operatorname{socle} p^i = \theta(i)$. By Lemmas 7.8 and 7.9, we have

$$\dim_{Q_0} p^i = \sigma_0 \cdot_w 0 = \sigma_0 \cdot_{\theta(w)} 0 = \dim_{Q_0} q^{\theta(i)}.$$

Thus, by Proposition 4.9, we have $p^i \cong q^{\theta(i)}$. □

Corollary 7.11. *Suppose Q is a quiver of finite type, $w \in \mathbb{N}Q_0$, and $\sigma \in \mathcal{W}$. Then $q^{w,\sigma} \cong p^{\theta(w),\sigma\sigma_0}$.*

Proof. Let $\tau = \sigma\sigma_0$ (and so $\sigma = \tau\sigma_0$). In light of Propositions 4.9, 7.3 and 7.10 and Definitions 4.10 and 7.4, it suffices to prove that the codimension of $q^{w,\sigma}$ in q^w is $\tau \cdot_{\theta(w)} 0$.

Let $y = \tau \cdot_{\theta(w)} 0$, so that $\tau(\theta(w)) = \theta(w) - \alpha_y$, that is,

$$\alpha_y = \theta(w) - \tau(\theta(w)).$$

Next, let

$$v = \dim_{Q_0} q^w = \sigma_0 \cdot_w 0 \quad \text{and} \quad u = \dim_{Q_0} q^{w,\sigma} = \sigma \cdot_w 0,$$

which implies $\sigma_0(w) = w - \alpha_v$ and $\sigma(w) = w - \alpha_u$. Then

$$\sum_{i \in Q_0} (v_i - u_i)\alpha_i = -\sigma_0(w) + \sigma(w) = \theta(w) + \tau\sigma_0(w) = \theta(w) - \tau(\theta(w)),$$

and so $y = v - u$ as desired. □

Proposition 7.12. *If Q is a quiver of finite type, then*

$$\mathrm{Gr}_{\mathcal{P}}(u, q^w) \cong \tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0 \cdot_w 0) - u, p^{\theta(w)}).$$

Proof. Let (x, V) be the quiver representation corresponding to the $\mathcal{P}$-module q^w and let $v = \dim_{Q_0} V = \sigma_0 \cdot_w 0$. By Proposition 7.10, (x, V) also corresponds to the $\mathcal{P}$-module $p^{\theta(w)}$. By Remark 2.10, $\mathcal{P}_n \cdot p^w = 0$ for sufficiently large n. Therefore

$$\begin{aligned} \mathrm{Gr}_{\mathcal{P}}(u, q^w) &= \{U \subseteq V \mid x(U) \subseteq U,\ \dim U = u\} \\ &= \{U \subseteq V \mid x(U) \subseteq U,\ \dim_{Q_0} V/U = v - u\} \\ &\cong \tilde{\mathrm{Gr}}_{\mathcal{P}}(v - u, p^{\theta(w)}). \end{aligned}$$

□

By Proposition 7.12, we have

(7-1) $$\mathfrak{L}(u,w) \xleftarrow[\cong]{\phi_w(u)} \mathrm{Gr}_{\mathcal{P}}(u,q^w) \cong \tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0 \cdot_w 0)-u,\, p^{\theta(w)}) \xrightarrow[\cong]{\psi_{\theta(w)}((\sigma_0\cdot_w 0)-u)} \mathfrak{L}((\sigma_0 \cdot_w 0)-u,\theta(w)),$$

where $\phi_w(u)$ is the isomorphism of Theorem 4.4 (see Corollary A.6), and $\psi_{\theta(w)}(u)$ is the isomorphism of Proposition 7.2. Define

$$\phi_w = (\phi_w(u))_u : \mathrm{Gr}_{\mathcal{P}}(q^w) \to \bigsqcup_u \mathfrak{L}(u,w),$$
$$\psi_w = (\psi_w(u))_u : \tilde{\mathrm{Gr}}_{\mathcal{P}}(p^w) \to \bigsqcup_u \mathfrak{L}(u,w).$$

Theorem 7.13. *The isomorphism $\psi_{\theta(w)} \circ \phi_w^{-1}$ induces the involution ζ. More precisely, we have $a \circ (\psi_{\theta(w)} \circ \phi_w^{-1})^* = (\psi_{\theta(w)} \circ \phi_w^{-1})^* \circ \zeta(a)$, $a \in \mathfrak{g}$, as operators on L_w, where $(\psi_{\theta(w)} \circ \phi_w^{-1})^*$ denotes the pullback of functions along $\psi_{\theta(w)} \circ \phi_w^{-1}$.*

Proof. For $u, u' \in \mathbb{N}Q_0$, define

$$\tilde{\mathrm{Gr}}_{\mathcal{P}}(u,u',p^{\theta(w)}) = \{(U,U') \in \tilde{\mathrm{Gr}}_{\mathcal{P}}(u,p^{\theta(w)}) \times \tilde{\mathrm{Gr}}_{\mathcal{P}}(u',p^{\theta(w)}) \mid U' \subseteq U\}.$$

The map $\psi_{\theta(w)}$ induces a isomorphism

$$\tilde{\mathrm{Gr}}_{\mathcal{P}}(u,u',p^{\theta(w)}) \xrightarrow{\cong} \mathfrak{F}(u,\theta(w);u-u')$$

for all $u, u' \in \mathbb{N}Q_0$ and we will also denote this collection of isomorphisms by $\psi_{\theta(w)}$. Then we have the commutative diagram

$$\begin{tikzcd}
\mathfrak{L}(u-i,w) & \mathfrak{F}(u,w;i) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] & \mathfrak{L}(u,w) \\
\mathrm{Gr}_{\mathcal{P}}(u-i,q^w) \arrow[u, "\phi_w", "\cong"'] \arrow[d, "\cong"] & \mathrm{Gr}_{\mathcal{P}}(u-i,u,q^w) \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] \arrow[u, "\phi_w", "\cong"'] \arrow[d, "\cong"] & \mathrm{Gr}_{\mathcal{P}}(u,q^w) \arrow[u, "\phi_w", "\cong"'] \arrow[d, "\cong"] \\
\tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0\cdot_w 0)-(u-i),p^{\theta(w)}) \arrow[d, "\psi_{\theta(w)}"', "\cong"] & \Xi \arrow[l, "\pi_2"'] \arrow[r, "\pi_1"] \arrow[d, "\psi_{\theta(w)}"', "\cong"] & \tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0\cdot_w 0)-u,p^{\theta(w)}) \arrow[d, "\psi_{\theta(w)}"', "\cong"] \\
\mathfrak{L}((\sigma_0\cdot_w 0)-(u-i),\theta(w)) & \mathfrak{F}((\sigma_0\cdot_w 0)-u,\theta(w);i) \arrow[l, "\pi_2"'] \arrow[r, "\pi_1"] & \mathfrak{L}((\sigma_0\cdot_w 0)-u,\theta(w))
\end{tikzcd}$$

where $\Xi = \tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0 \cdot_w 0)-u, (\sigma_0 \cdot_w 0)-(u-i), p^{\theta(w)})$. It follows that, for f in $\bigoplus_u M(\mathfrak{L}(u,w))$, we have

$$E_i \circ (\psi_{\theta(w)} \circ \phi_w^{-1})^*(f) = (\psi_{\theta(w)} \circ \phi_w^{-1})^* \circ F_i(f),$$
$$F_i \circ (\psi_{\theta(w)} \circ \phi_w^{-1})^*(f) = (\psi_{\theta(w)} \circ \phi_w^{-1})^* \circ E_i(f).$$

Furthermore, $(\psi_{\theta(w)} \circ \phi_w^{-1})^*$ maps the constant function on $\mathfrak{L}(0, w)$ with value one to the constant function on $\mathfrak{L}(\sigma_0 \cdot_w 0, \theta(w))$ with value one. The result follows. $\square$

Remark 7.14. Note that the middle isomorphism in (7-1) depends on our identification of q^w and $p^{\theta(w)}$. The isomorphism $\phi_w(u)$ also depends on our fixed retract $\pi : q^w \to s^w$. By Proposition 4.1, all such choices are related by the natural action of $\mathrm{Aut}_{\mathcal{P}}\, q^w$; see Definition 2.16. A similar group action appears in the identification of $\tilde{\mathrm{Gr}}_{\mathcal{P}}((\sigma_0 \cdot_w 0) - u,\, p^{\theta(w)})$ with $\mathfrak{L}((\sigma_0 \cdot_w 0) - u, \theta(w))$; see [Lusztig 2000]. Via the isomorphisms $\phi_w(u)$, the group $\mathrm{Aut}_{\mathcal{P}}\, q^w$ acts on the space of constructible functions on $\bigsqcup_v \mathfrak{L}(v, w)$ and L_w is a subspace of the space of invariant functions. The pullback $(\psi_{\theta(w)} \circ \phi_w^{-1})^*$ acting on the space of invariant functions is independent of the choice of π and the chosen identification of q^w with $p^{\theta(w)}$.

Appendix: Isomorphisms of varieties

After an earlier version of the current paper was released [Savage and Tingley 2009], Shipman proved [2010] that the grassmannian type varieties $\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$ defined by Lusztig are indeed isomorphic as algebraic varieties to the lagrangian Nakajima quiver varieties $\mathfrak{L}(v, w)$. A simple "duality" map gives an isomorphism of varieties between the quiver grassmannian $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ and $\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$. The purpose of this appendix is to describe this map precisely, and from there to conclude that the map from $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ to $\mathfrak{L}(v, w)$ constructed in Theorem 4.4 is in fact an isomorphism of algebraic varieties. An alternative approach (not pursued here) would be an injective version of the argument of [Shipman 2010] that would directly show that $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ is isomorphic to $\mathfrak{L}(v, w)$.

Let $i \in Q_0$ and fix a nondegenerate bilinear pairing

$$\langle \cdot\, , \cdot \rangle_{s^i} : s^i \times s^i \to \mathbb{C},$$

and a retract $\pi : q^i \to s^i$ of $\mathcal{P}_0$-modules. For a path $\beta = a_1 \cdots a_n$ in the double quiver $\tilde{Q}$, let

$$\beta^\vee = \bar{a}_n \cdots \bar{a}_1 \tag{A-1}$$

be the reverse path. Extending by linearity, this defines an algebra anti-involution of $\mathbb{C}\tilde{Q}$ that induces an algebra anti-involution of $\mathcal{P}$. Then define a bilinear pairing

$$\langle \cdot\, , \cdot \rangle : \tilde{q}^i \times p^i \to \mathbb{C}, \quad \langle v, \beta e_i \rangle = \langle \pi(\beta^\vee v), e_i \rangle_{s^i}. \tag{A-2}$$

For $n \geq 0$, let

$$p_n^i = \mathcal{P}_{\geq n} e_i \subseteq p^i,$$
$$q_n^i = \{v \in q^i \mid \mathcal{P}_n \cdot v = 0\} = \{v \in \tilde{q}^i \mid \mathcal{P}_n \cdot v = 0\},$$

where the last equality holds since $\tilde{q}^i$ contains all nilpotent elements of q^i by Lemma 4.14. Note that each q^i_n is finite-dimensional. We have the obvious inclusions

$$q^i_0 \subseteq q^i_1 \subseteq q^i_2 \subseteq \cdots,$$

and it follows from Lemma 4.14 and Theorem 4.15 that $\tilde{q}^i = \bigcup_{n=0}^{\infty} q^i_n$. It is clear from the definitions that

$$\langle q^i_n, p^i_{n+1} \rangle = 0, \quad \text{for all } n \geq 0.$$

Thus we have the induced bilinear pairing on $q^i_n \times (p^i / p^i_{n+1})$.

Lemma A.1. *The pairing*

$$\langle\,\cdot\,,\,\cdot\,\rangle : q^i_n \times (p^i / p^i_{n+1}) \to \mathbb{C}$$

is nondegenerate.

Proof. Since q^i_n is nilpotent of degree n and has socle s^i, for all nonzero $v \in q^i_n$, there exists $\beta \in \mathcal{P}_{\leq n}$ such that $0 \neq \beta \cdot v \in s^i$. Then $\langle v, \beta^\vee e_i \rangle \neq 0$. Thus, it suffices to show that $\dim(p^i / p^i_{n+1}) \leq \dim q^i_n$. Now, $(p^i / p^i_{n+1})^*$ is naturally a right $\mathcal{P}$-module. Via the anti-involution (A-1), this becomes a nilpotent left $\mathcal{P}$-module with socle s^i. Therefore, by Proposition 4.1, $(p^i / p^i_{n+1})^*$ injects into $\tilde{q}^i$. It is clear that the image of this injection is contained in q^i_n and thus the result follows since q^i_n is finite-dimensional. □

We then have the following corollary, whose proof is immediate.

Corollary A.2. *The pairing (A-2) is nondegenerate. Furthermore,*

$$\tilde{q}^i \cong \{ f \in \operatorname{Hom}_{\mathbb{C}}(p^i, \mathbb{C}) \mid f|_{p^i_n} = 0 \text{ for } n \gg 0 \}$$

as $\mathcal{P}$-modules, where the $\mathcal{P}$-module structure on the right-hand side is given by

$$(\beta \cdot f')(v) = f'(\beta^\vee \cdot v),$$

for $\beta \in \mathcal{P}$, $v \in p^i$, and $f' \in \{ f \in \operatorname{Hom}_{\mathbb{C}}(p^i, \mathbb{C}) \mid f|_{p^i_n} = 0 \text{ for } n \gg 0 \}$.

Remark A.3. One should compare this result to Definition 2.7 and Lemma 2.8 in finite type.

Recall that, for $w = \sum_i w_i i \in \mathbb{N}Q_0$, we have

$$s^w = \bigoplus_i (s^i)^{\oplus w_i}, \quad p^w = \bigoplus_i (p^i)^{\oplus w_i}, \quad \tilde{q}^w = \bigoplus_i (\tilde{q}^i)^{\oplus w_i}.$$

By declaring distinct summands to be orthogonal, we have a nondegenerate bilinear pairing

$$\langle\,\cdot\,,\,\cdot\,\rangle : \tilde{q}^w \times p^w \to \mathbb{C}. \tag{A-3}$$

For a subspace U of $\tilde{q}^w$, define the subspace

$$U^\perp = \{v \in p^w \mid \langle v', v\rangle = 0 \text{ for all } v' \in U\}$$

of p^w. Similarly, for a subspace U of p^w, define the subspace $U^\perp$ of $\tilde{q}^w$.

Proposition A.4. *For $U \in \mathrm{Gr}_{\mathcal{P}}(v, \tilde{q}^w)$, we have $U^\perp \in \tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$, and the map*

$$\mathrm{Gr}_{\mathcal{P}}(v, \tilde{q}^w) \to \tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w), \quad U \mapsto U^\perp,$$

is an isomorphism of algebraic varieties.

Proof. It follows from the definition of the pairing (A-3) that U is a submodule of $\tilde{q}^w$ if and only if $U^\perp$ is a submodule of p^w. Also, note that $U \subseteq \tilde{q}^w$ is finite-dimensional if and only if $U \subseteq q^w_n$ for some n. Therefore, it follows from Lemma A.1 that the maps $U \mapsto U^\perp$ (in either direction) are mutually inverse bijections between $\mathrm{Gr}_{\mathcal{P}}(v, \tilde{q}^w)$ and $\tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w)$. Since these maps are clearly algebraic, the result follows. □

Theorem A.5. *The quiver grassmannian $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ is isomorphic to the lagrangian Nakajima quiver variety $\mathfrak{L}(v, w)$ as an algebraic variety.*

Proof. This follows from the isomorphisms of algebraic varieties

$$\mathrm{Gr}_{\mathcal{P}}(v, q^w) = \mathrm{Gr}_{\mathcal{P}}(v, \tilde{q}^w) \cong \tilde{\mathrm{Gr}}_{\mathcal{P}}(v, p^w) \cong \mathfrak{L}(v, w).$$

Recall that all finite-dimensional submodules of q^w are submodules of $\tilde{q}^w$. This gives the first equality. The first isomorphism is Proposition A.4 and the second is Proposition 7.2. □

Corollary A.6. *The map $\bar{\iota} : \mathrm{Gr}_{\mathcal{P}}(v, q^w) \to \mathfrak{L}(v, w)$ of Theorem 4.4 is an isomorphism of algebraic varieties.*

Proof. By Theorem A.5, we know that $\mathrm{Gr}_{\mathcal{P}}(v, q^w)$ and $\mathfrak{L}(v, w)$ are isomorphic as algebraic varieties. Since $\bar{\iota}$ is a bijective algebraic map by Theorem 4.4, the result follows by [Kaliman 2005, Lemma 1] (while the result there is stated for irreducible varieties, the proof applies to reducible ones — the only difference is that the normalization is now a disjoint union of components). □

Acknowledgements

The authors would like to thank B. Leclerc who, after hearing some of the preliminary results of the current paper, suggested extending these results to graded/cyclic versions. They are also grateful to W. Crawley-Boevey for many helpful discussions and for suggesting the proof of Proposition 2.11. Furthermore, they would like to thank P. Etingof, A. Hubery, H. Nakajima, M. Roth, O. Schiffmann, and I. Shipman for useful conversations and S.-J. Kang, Y.-T. Oh, and the Korean

Mathematical Society for the invitation to participate in the 2008 Global KMS International Conference in Jeju, Korea, where the ideas in this paper were originally developed.

References

[Anderson and Fuller 1992] F. W. Anderson and K. R. Fuller, *Rings and categories of modules*, 2nd ed., Graduate Texts in Mathematics **13**, Springer, New York, 1992. MR 94i:16001 Zbl 0765.16001

[Auslander et al. 1995] M. Auslander, I. Reiten, and S. O. Smalø, *Representation theory of Artin algebras*, Cambridge Studies in Advanced Mathematics **36**, Cambridge University Press, Cambridge, 1995. MR 98e:16011 Zbl 0834.16001

[Baer 1940] R. Baer, "Abelian groups that are direct summands of every containing abelian group", *Bull. Amer. Math. Soc.* **46** (1940), 800–806. MR 2,126i Zbl 0024.14902

[Baer et al. 1987] D. Baer, W. Geigle, and H. Lenzing, "The preprojective algebra of a tame hereditary Artin algebra", *Comm. Algebra* **15**:1-2 (1987), 425–457. MR 88i:16036 Zbl 0612.16015

[Caldero and Chapoton 2006] P. Caldero and F. Chapoton, "Cluster algebras as Hall algebras of quiver representations", *Comment. Math. Helv.* **81**:3 (2006), 595–616. MR 2008b:16015 ZBL 1119.16013

[Caldero and Keller 2006] P. Caldero and B. Keller, "From triangulated categories to cluster algebras. II", *Ann. Sci. École Norm. Sup.* (4) **39**:6 (2006), 983–1009. MR 2008m:16031 Zbl 1115.18301

[Caldero and Reineke 2008] P. Caldero and M. Reineke, "On the quiver Grassmannian in the acyclic case", *J. Pure Appl. Algebra* **212**:11 (2008), 2369–2380. MR 2009f:14102 Zbl 1153.14032

[Crawley-Boevey 1996] W. Crawley-Boevey, "On homomorphisms from a fixed representation to a general representation of a quiver", *Trans. Amer. Math. Soc.* **348**:5 (1996), 1909–1919. MR 96i:16022 Zbl 0855.16014

[Crawley-Boevey 2000] W. Crawley-Boevey, "On the exceptional fibres of Kleinian singularities", *Amer. J. Math.* **122**:5 (2000), 1027–1037. MR 2001f:14009 Zbl 1001.14001

[Crawley-Boevey 2001] W. Crawley-Boevey, "Geometry of the moment map for representations of quivers", *Compositio Math.* **126**:3 (2001), 257–293. MR 2002g:16021 Zbl 1037.16007

[Deng et al. 2008] B. Deng, J. Du, B. Parshall, and J. Wang, *Finite dimensional algebras and quantum groups*, Mathematical Surveys and Monographs **150**, American Mathematical Society, Providence, RI, 2008. MR 2457938 Zbl 1154.17003

[Derksen et al. 2009] H. Derksen, J. Weyman, and A. Zelevinsky, "Quivers with potentials and their representations, II: applications to cluster algebras", preprint, 2009. arXiv 0904:0676

[Fedotov 2010] S. Fedotov, "Framed moduli and Grassmannians of submodules", preprint, 2010. arXiv 1010.4761

[Frenkel and Savage 2003] I. B. Frenkel and A. Savage, "Bases of representations of type A affine Lie algebras via quiver varieties and statistical mechanics", *Int. Math. Res. Not.* **2003**:28 (2003), 1521–1547. MR 2004j:17010 Zbl 1023.17010

[Geiss et al. 2006] C. Geiss, B. Leclerc, and J. Schröer, "Verma modules and preprojective algebras", *Nagoya Math. J.* **182** (2006), 241–258. MR 2007m:17037 Zbl 1137.17021

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Jategaonkar 1976] A. V. Jategaonkar, "Certain injectives are Artinian", pp. 128–139 in *Noncommutative ring theory* (Kent, OH, 1975), Lecture Notes in Math. **545**, Springer, Berlin, 1976. MR 55 #5688 Zbl 0343.16021

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

[Kaliman 2005] S. Kaliman, "On a theorem of Ax", *Proc. Amer. Math. Soc.* **133**:4 (2005), 975–977. MR 2006a:14104 Zbl 1060.14002

[Lam 1999] T. Y. Lam, *Lectures on modules and rings*, Graduate Texts in Mathematics **189**, Springer, New York, 1999. MR 99i:16001 Zbl 0911.16001

[Lusztig 1991] G. Lusztig, "Quivers, perverse sheaves, and quantized enveloping algebras", *J. Amer. Math. Soc.* **4**:2 (1991), 365–421. MR 91m:17018 Zbl 0738.17011

[Lusztig 1998] G. Lusztig, "On quiver varieties", *Adv. Math.* **136**:1 (1998), 141–182. MR 2003f:16016 Zbl 0915.17008

[Lusztig 2000] G. Lusztig, "Remarks on quiver varieties", *Duke Math. J.* **105**:2 (2000), 239–265. MR 2002b:20072 Zbl 1017.20040

[McConnell and Robson 2001] J. C. McConnell and J. C. Robson, *Noncommutative Noetherian rings*, rev. ed., Graduate Studies in Mathematics **30**, American Mathematical Society, Providence, RI, 2001. MR 2001i:16039 Zbl 0980.16019

[Nakajima 1994] H. Nakajima, "Instantons on ALE spaces, quiver varieties, and Kac–Moody algebras", *Duke Math. J.* **76**:2 (1994), 365–416. MR 95i:53051 Zbl 0826.17026

[Nakajima 1996] H. Nakajima, "Varieties associated with quivers", pp. 139–157 in *Representation theory of algebras and related topics* (Mexico City, 1994), edited by R. Bautista et al., CMS Conf. Proc. **19**, Amer. Math. Soc., Providence, RI, 1996. MR 97m:16022 Zbl 0870.16008

[Nakajima 1998] H. Nakajima, "Quiver varieties and Kac–Moody algebras", *Duke Math. J.* **91**:3 (1998), 515–560. MR 99b:17033 Zbl 0970.17017

[Nakajima 2001] H. Nakajima, "Quiver varieties and finite-dimensional representations of quantum affine algebras", *J. Amer. Math. Soc.* **14**:1 (2001), 145–238. MR 2002i:17023 Zbl 0981.17016

[Nakajima 2003] H. Nakajima, "Reflection functors for quiver varieties and Weyl group actions", *Math. Ann.* **327**:4 (2003), 671–721. MR 2004k:16036 Zbl 1060.16017

[Nakajima 2004] H. Nakajima, "Quiver varieties and t-analogs of q-characters of quantum affine algebras", *Ann. of Math.* (2) **160**:3 (2004), 1057–1097. MR 2006k:17029 Zbl 1140.17015

[Reineke 2008] M. Reineke, "Framed quiver moduli, cohomology, and quantum groups", *J. Algebra* **320**:1 (2008), 94–115. MR 2009d:16021 Zbl 1153.14033

[Reiten 1997] I. Reiten, "Dynkin diagrams and the representation theory of algebras", *Notices Amer. Math. Soc.* **44**:5 (1997), 546–556. MR 98i:16014 Zbl 0940.16009

[Ringel 1998] C. M. Ringel, "The preprojective algebra of a quiver", pp. 467–480 in *Algebras and modules* (Geiranger, 1996), vol. 2, edited by I. Reiten et al., CMS Conf. Proc. **24**, Amer. Math. Soc., Providence, RI, 1998. MR 99i:16031 Zbl 0928.16012

[Saito 2002] Y. Saito, "Crystal bases and quiver varieties", *Math. Ann.* **324**:4 (2002), 675–688. MR 2004a:17023 Zbl 1041.17018

[Savage 2006a] A. Savage, "Finite-dimensional algebras and quivers", pp. 313–320 in *Encyclopedia of Mathematical Physics*, vol. 2, edited by J.-P. Françoise et al., Elsevier, Oxford, 2006.

[Savage 2006b] A. Savage, "Geometric and combinatorial realizations of crystal graphs", *Algebr. Represent. Theory* **9**:2 (2006), 161–199. MR 2238365 Zbl 1138.17006

[Savage 2006c] A. Savage, "A geometric realization of spin representations and Young diagrams from quiver varieties", *J. Algebra* **297**:1 (2006), 186–207. MR 2206854 Zbl 1141.17004

[Savage 2006d] A. Savage, "Quiver varieties and Demazure modules", *Math. Ann.* **335**:1 (2006), 31–46. MR 2008j:16052 Zbl 1101.17010

[Savage and Tingley 2009] A. Savage and P. Tingley, "Quiver grassmannians, quiver varieties and the preprojective algebra", preprint, 2009. arXiv 0909.3746v1

[Schofield 1992] A. Schofield, "General representations of quivers", *Proc. London Math. Soc.* (3) **65**:1 (1992), 46–64. MR 93d:16014 Zbl 0795.16008

[Shipman 2010] I. Shipman, "On representation schemes and grassmanians of finite dimensional algebras and a construction of Lusztig", *Math. Res. Lett.* **17**:5 (2010), 969–976.

[Szántó 2009] C. Szántó, "On the cardinalities of Kronecker quiver grassmannians", preprint, 2009. arXiv 0903.1928

[Wolf 2009] S. Wolf, "The Hall algebra and the composition monoid", preprint, 2009. arXiv0907.1106

Received June 23, 2010. Revised January 21, 2011.

ALISTAIR SAVAGE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
585 KING EDWARD AVENUE
OTTAWA, ON K1N 6N5
CANADA

alistair.savage@uottawa.ca
http://www.mathstat.uottawa.ca/~asavag2

PETER TINGLEY
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 MASSACHUSETTS AVENUE
CAMBRIDGE, MA 02139-4307
UNITED STATES

ptingley@math.mit.edu
http://www-math.mit.edu/~ptingley/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

NONAUTONOMOUS SECOND ORDER HAMILTONIAN SYSTEMS

MARTIN SCHECHTER

We study the existence of periodic solutions for a second order nonautonomous dynamical system. We make no assumptions on the gradient other than continuity. This allows both sublinear and superlinear problems. We also study the existence of nonconstant solutions.

1. Introduction

We consider the following problem. One wishes to solve

$$-\ddot{x}(t) = \nabla_x V(t, x(t)), \tag{1-1}$$

where

$$x(t) = (x_1(t), \dots, x_n(t)) \tag{1-2}$$

is a map from $I = [0, T]$ to $\mathbb{R}^n$ such that each component $x_j(t)$ is a periodic function in H^1 with period T, and the function $V(t, x) = V(t, x_1, \dots, x_n)$ is continuous from $\mathbb{R}^{n+1}$ to $\mathbb{R}$ with

$$\nabla_x V(t, x) = (\partial V/\partial x_1, \dots, \partial V/\partial x_n) \in C(\mathbb{R}^{n+1}, \mathbb{R}^n). \tag{1-3}$$

For each $x \in \mathbb{R}^n$, the function $V(t, x)$ is periodic in t with period T.

We shall study this problem under several sets of assumptions. First, we make no assumption on $\nabla_x V(t, x)$ other than (1-3). This allows both sublinear and superlinear problems.

Theorem 1.1. *Assume*:

(1) *The function V satisfies*

$$0 \le \int_0^T V(t, x)\, dt \to \infty \quad \text{as } |x| \to \infty,\ x \in \mathbb{R}^n.$$

MSC2000: 35J20, 35J25, 47J30, 49J40, 58E05.

Keywords: Hamiltonian, second order, nonautonomous, critical point, linking, dynamical system, periodic solution.

(2) *There are positive constants* α, m *such that*

$$\int_0^T V(t,x)\,dt \le \alpha, \quad |x| \le m,\ x \in \mathbb{R}^n.$$

Then the system

$$-\ddot{x}(t) = \beta \nabla_x V(t, x(t)) \tag{1-4}$$

has a solution for almost all values of β *satisfying* $\beta \le 6m^2/\alpha T$. *If, in addition, there are a constant* $\gamma > 0$ *and a function* $W(t) \in L^1(I)$ *such that*

$$V(t,x) \ge \gamma |x|^2 - W(t),$$

then the system (1-4) *has a nonconstant solution for almost all* β *satisfying*

$$\frac{2\pi^2}{\gamma T^2} \le \beta \le \frac{6m^2}{\alpha T}.$$

Corollary 1.2. *Assume*:

(1) *The function* V *satisfies*

$$0 \le \int_0^T V(t,x)\,dt \to \infty \quad \text{as } |x| \to \infty,\ x \in \mathbb{R}^n.$$

(2) *There are positive constants* α, m *such that*

$$V(t,x) \le \alpha, \quad |x| \le m,\ x \in \mathbb{R}^n.$$

Then the system (1-4) *has a solution for almost all values of* β *satisfying* $0 \le \beta \le 6m^2/\alpha T^2$.

Theorem 1.3. *Assume*:

(1) *The function* V *satisfies*

$$0 \le \int_0^T V(t,x)\,dt \to \infty \quad \text{as } |x| \to \infty,\ x \in \mathbb{R}^n.$$

(2) *There is a constant* $q > 2$ *such that*

$$V(t,x) \le C(|x|^q + 1), \quad t \in I,\ x \in \mathbb{R}^n.$$

(3) *there are constants* $m > 0$, $\alpha > 0$ *such that*

$$V(t,x) \le \alpha |x|^2, \quad |x| \le m,\ t \in I,\ x \in \mathbb{R}^n.$$

Then the system (1-4) *has a solution for almost all* β *satisfying* $0 \le \beta \le 2\pi^2/\alpha T^2$.

Theorem 1.4. *Assume*:

(1) *The function* V *satisfies*

$$0 \le \int_0^T V(t,x)\,dt \to \infty \quad \text{as } |x| \to \infty,\ x \in \mathbb{R}^n.$$

(2) *There are a constant* $\alpha > 0$ *and a function* $W(t) \in L^1(I)$ *such that*

$$V(t,x) \le \alpha |x|^2 + W(t), \quad t \in I,\ x \in \mathbb{R}^n.$$

Then the system (1-4) *has a solution for almost all* $0 \le \beta \le 2\pi^2/\alpha T^2$. *If we assume*

$$B := \int_I W(t)\,dt < 0,$$

then (1-4) *has a nonconstant solution for almost all such* β.

Theorem 1.5. *The conclusions of Theorem 1.4 are valid if we replace condition* (2) *with*:

(2′) *There is a constant* $\alpha > 0$ *such that*

$$\sup_{|x|<m} \int_0^T V(t,x)\,dt \le \alpha m^2 + B \quad \text{for every } m > 0,$$

and require $0 \le \beta \le 6/\alpha T$.

The advantage of these theorems is that we obtain solutions under very weak hypotheses. In fact, we make no assumption on $\nabla_x V(t,x)$ other than (1-3). The disadvantage is that we do not obtain a solution for any particular value of β. If we wish to prove existence for every such β, we will have to make assumptions concerning $\nabla_x V(t,x)$ as well. We now present additional hypotheses which guarantee existence of solutions for all values of β in the given intervals. We do this for Theorems 1.1 and 1.3. The hypotheses are:

(1) $0 \le V(t,x)/|x|^2 \to \infty$ as $|x| \to \infty$.

(2) There are a constant C and a function $W(t) \in L^1(I)$ such that

$$H(t,\theta x) \le C(H(t,x) + W(t)), \quad 0 \le \theta \le 1,\ t \in I,\ x \in \mathbb{R}^n,$$

where

$$H(t,x) := \nabla_x V(t,x) \cdot x - 2V(t,x).$$

Theorem 1.6. *Assume*:

(1) $0 \le V(t,x)/|x|^2 \to \infty$ as $|x| \to \infty$.

(2) *There are positive constants* α, m *such that*

$$\int_0^T V(t,x)\,dt \le \alpha, \quad |x| \le m,\ x \in \mathbb{R}^n.$$

(3) *There are a constant C and a function $W(t) \in L^1(I)$ such that*

$$H(t, \theta x) \leq C(H(t, x) + W(t)), \quad 0 \leq \theta \leq 1, \ t \in I, \ x \in \mathbb{R}^n.$$

Then the system (1-4) has a solution for all values of β satisfying $0 < \beta < 6m^2/\alpha T$.

Theorem 1.7. *Assume*:

(1) $0 \leq V(t, x)/|x|^2 \to \infty$ as $|x| \to \infty$.

(2) *There is a constant $q > 2$ such that*

$$V(t, x) \leq C(|x|^q + 1), \quad t \in I, \ x \in \mathbb{R}^n.$$

(3) *There are constants $m > 0$, $\alpha > 0$ such that*

$$V(t, x) \leq \alpha |x|^2, \quad |x| \leq m, \ t \in I, \ x \in \mathbb{R}^n.$$

(4) *There are a constant C and a function $W(t) \in L^1(I)$ such that*

$$H(t, \theta x) \leq C(H(t, x) + W(t)), \quad 0 \leq \theta \leq 1, \ t \in I, \ x \in \mathbb{R}^n.$$

Then the system (1-4) has a solution for all β satisfying $0 < \beta < 2\pi^2/\alpha T^2$.

The periodic nonautonomous problem

$$\ddot{x}(t) = \nabla_x V(t, x(t)) \tag{1-5}$$

has an extensive history in the case of singular systems (see, for example, [Ambrosetti and Coti Zelati 1993]). The first to consider it for potentials satisfying (1-3) were Berger and the author [1977]. We proved the existence of solutions to (1-4) under the condition that

$$V(t, x) \to \infty \quad \text{as } |x| \to \infty$$

uniformly for a.e. $t \in I$. Subsequently, Willem [1981], Mawhin [1987], Mawhin and Willem [1989], Tang [1995; 1998], Tang and Wu [1999; 2001; 2002] and others (see the references therein) proved existence under various conditions.

The periodic problem (1-1) was studied by Mawhin and Willem [1986; 1989], Long [1995], Tang and Wu [2003] and others. Tang and Wu [2003] proved existence of solutions of problem (1-1) under the following hypotheses:

(I) $V(t, x) \to \infty$ as $|x| \to \infty$ uniformly for a.e. $t \in I$.

(II) There exist $a \in C(\mathbb{R}^+, \mathbb{R}^+)$, $b \in L^1(0, T, \mathbb{R}^+)$ such that

$$|V(t, x)| + |\nabla V(t, x)| \leq a(|x|)b(t) \quad \text{for all } x \in \mathbb{R}^n \text{ and a.e. } t \in [0, T].$$

and the superquadraticity condition:

(III) There exist $0 < \mu < 2$, $M > 0$ such that

$$V(t, x) > 0, \ H_\mu := \nabla V(t, x) \cdot x - \mu V(t, x) \leq 0 \quad \text{for all } |x| \geq M \text{ and a.e. } t \in [0, T].$$

Rabinowitz [1980] proved existence under stronger hypotheses. In particular, in place of (I), he assumed:

(I′) There exist constants $a_1, a_2 > 0$, $\mu_0 > 1$ such that

$$V(t, x) \geq a_1|x|^{\mu_0} + a_2 \quad \text{for all } x \in \mathbb{R}^n \text{ and a.e. } t \in [0, T]$$

In place of (III), he assumed:

(III′) There exist $0 < \mu < 2$, $M > 0$ such that

$$0 < \nabla V(t, x) \cdot x \leq \mu V(t, x) \quad \text{for all } |x| \geq M \text{ and a.e. } t \in [0, T].$$

Mawhin and Willem [1986] proved existence for the case of convex potentials, while Long [1995] studied the problem for even potentials. They assumed that $V(t, x)$ is subquadratic in the sense that

$$\begin{aligned}&\text{there exist } a_3 < (2\pi/T)^2 \text{ and } a_4 \text{ such that}\\ &|V(t, x)| \leq a_3|x|^2 + a_4 \text{ for all } x \in \mathbb{R}^n \text{ and a.e. } t \in [0, T].\end{aligned}$$

Mawhin and Willem [1989] also studied the problem for a bounded nonlinearity. Tang and Wu [2003] also proved existence of solutions if one replaces (I) with

$$\int_0^T V(t, x)\, dt \to \infty \quad \text{as } |x| \to \infty$$

and $V(t, x)$ is γ-subadditive with $\gamma > 0$ for a.e. $t \in [0, T]$. All of these authors studied only the existence of solutions.

All of the results mentioned above concerned the existence of solutions, which might be constants. Little was done concerning *nonconstant* solutions of problem (1-1). For the homogeneous case, Ben-Naoum, Troestler and Willem [Ben-Naoum et al. 1994] proved the existence of a nonconstant solution. For the case $T = 2\pi$, Theorem 1.7, with substantially stronger hypotheses, was proved by Nirenberg; see [Ekeland and Ghoussoub 2002]. Among other things, they assumed

$$V(t, x) \leq \frac{3}{2\pi^2}, \quad |x| \leq 1,\ t \in \mathbb{R},\ x \in \mathbb{R}^n,$$

and the superquadraticity condition

$$V(t, x) > 0,\ H_\mu(t, x) \leq 0, \quad |x| \geq C,\ t \in \mathbb{R},\ x \in \mathbb{R}^n,$$

for some $\mu > 2$, which implies our hypotheses, and

$$V(t, x) \geq C|x|^\mu - C', \quad x \in \mathbb{R}^n, C > 0,$$

among other things. These results were generalized in [Schechter 2006a; 2006b]. Further results, involving some of the hypotheses used in these last two papers, were obtained in [Wang et al. 2009].

We shall prove Theorems 1.1–1.5 in Section 5, and Theorems 1.6 and 1.7 in Section 7. We use linking and sandwich methods of critical point theory and then apply the monotonicity trick introduced by Struwe [1988; 1996] for minimization problems. (This trick was also used by others to solve Landesman–Lazer type problems, for bifurcation problems, for Hamiltonian systems and Schrödinger equations.)

Jeanjean [1999] shows that for a specific class of functionals having a mountain-pass (MP) geometry, almost every functional in this class has a bounded Palais–Smale sequence at the (MP) level. This theorem is used to obtain, for a given functional, a special Palais–Smale sequence possessing extra properties that help to ensure its convergence. Subsequently, these abstract results are applied to prove the existence of a positive solution for a problem of the form (P) $-\Delta u + Ku = f(x,u)$, $u \in H^1(R^N)$, $K > 0$. He assumed that the functional associated to (P) has an (MP) geometry. His results cover the case where the nonlinearity f satisfies (i) $f(x,s)s^{-1} \to a \in (0,\infty]$ as $s \to +\infty$ and (ii) $f(x,s)s^{-1}$ is nondecreasing as a function of $s \geq 0$, a.e. $x \in R^N$.

Here, we obtain a bounded Palais–Smale sequences for functionals that need not have (MP) geometry. We then apply the theory to situations in which the (MP) geometry is not present. In particular, we apply it to situations where there is linking without the (MP) geometry. We also apply it to situations in which there are sandwich pairs which do not link.

The theory of sandwich pairs began in [Silva 1991; Schechter 1992; 1993] and was developed in subsequent publications such as [Schechter 2008; 2009].

2. Flows

Let E be a Banach space, and let Σ be the set of all continuous maps $\sigma = \sigma(t)$ from $E \times [0,1]$ to E such that

(1) $\sigma(0)$ is the identity map,

(2) for each $t \in [0,1]$, $\sigma(t)$ is a homeomorphism of E onto E,

(3) $\sigma'(t)$ is piecewise continuous on [0,1] and satisfies

$$\|\sigma'(t)u\| \leq \text{constant}, \quad u \in E. \tag{2-1}$$

The mappings in Σ are called *flows*.

Remark 2.1. If $\sigma_1,\ \sigma_2$ are in Σ, define $\sigma_3 = \sigma_1 \circ \sigma_2$ by

$$\sigma_3(s) = \begin{cases} \sigma_1(2s) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ \sigma_2(2s-1)\sigma_1(1) & \text{if } \frac{1}{2} < s \leq 1. \end{cases}$$

Then $\sigma_1 \circ \sigma_2 \in \Sigma$.

3. Sandwich systems

Let E be a Banach space. Define a nonempty collection $\mathcal{K}$ of nonempty subsets $K \subset E$ to be a *sandwich system* if $\mathcal{K}$ has the following property:

$$\sigma(1)K \in \mathcal{K}, \quad \sigma \in \Sigma, \ K \in \mathcal{K}.$$

Theorem 3.1. *Let $\mathcal{K}$ be a sandwich system, and let $G(u)$ be a C^1 functional on E. Define*

$$a := \inf_{K \in \mathcal{K}} \sup_K G, \tag{3-1}$$

and assume that a is finite. Assume, in addition, that there is a constant C_0 such that for each $\delta > 0$ there is a $K \in \mathcal{K}$ satisfying

$$\sup_K G \le a + \delta, \tag{3-2}$$

such that the inequality

$$G(u) \ge a - \delta, \quad u \in K, \tag{3-3}$$

implies $\|u\| \le C_0$. Then there is a bounded sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to a, \quad \|G'(u_k)\| \to 0. \tag{3-4}$$

Theorem 3.2. *Let $\mathcal{K}$ be a sandwich system, and let $G(u)$ be a C^1 functional on E. Assume that there are subsets A, B of E such that*

$$a_0 := \sup_A G < \infty, \quad b_0 := \inf_B G > -\infty, \tag{3-5}$$

$A \in \mathcal{K}$ and

$$B \cap K \neq \varnothing, \quad K \in \mathcal{K}. \tag{3-6}$$

Assume, in addition, that there is a constant C_0 such that for each $\delta > 0$ there is a $K \in \mathcal{K}$ satisfying (3-2) such that the inequality (3-3) implies $\|u\| \le C_0$. Then the value a given by (3-1) satisfies $b_0 \le a \le a_0$ and there is a bounded sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to a, \quad \|G'(u_k)\| \to 0. \tag{3-7}$$

Definition 3.3. We shall say that sets A, B in E form a *sandwich pair* if A is a member of a sandwich system $\mathcal{K}$ and B satisfies (3-6).

Theorem 3.4. *Let N be a finite dimensional subspace of a Banach space E, and let p be any point of N. Let F be a continuous map of E onto N such that $F = I$ on N. Then $A = N$ and $B = F^{-1}(p)$ form a sandwich pair.*

Corollary 3.5. *Let N be a closed subspace of a Hilbert space E and let $M = N^{\perp}$. Assume that at least one of the subspaces M, N is finite dimensional. Then M, N form a sandwich pair.*

Corollary 3.6. *Let N be a finite dimensional subspace of a Hilbert space E with complement $M' = M \oplus \{v_0\}$, where v_0 is an element in E having unit norm, and let δ be any positive number. Let $\varphi(t) \in C^1(\mathbb{R})$ be such that*

$$0 \le \varphi(t) \le 1,\ \varphi(0) = 1 \quad \text{and} \quad \varphi(t) = 0,\ |t| \ge 1.$$

Let

$$F(v + w + sv_0) = v + \big(s + \delta - \delta\varphi(\|w\|^2/\delta^2)\big)v_0, \quad v \in N,\ w \in M,\ s \in \mathbb{R}. \tag{3-8}$$

Then $A = N' = N \oplus \{v_0\}$ and $B = F^{-1}(\delta v_0)$ form a sandwich pair.

Proof. One checks that the mapping F given by (3-8) satisfies the hypotheses of Theorem 3.4 for N'. □

4. The parameter problem

Let E be a reflexive Banach space with norm $\|\cdot\|$, and let A, B be two closed subsets of E. Suppose that $G \in \mathscr{C}^1(E, \mathbb{R})$ is of the form $G(u) := I(u) - J(u)$, $u \in E$, where I, $J \in \mathscr{C}^1(E, \mathbb{R})$ map bounded sets to bounded sets. Define

$$G_\lambda(u) = \lambda I(u) - J(u), \quad \lambda \in \Lambda,$$

where Λ is an open interval contained in $(0, +\infty)$. Assume one of the following alternatives holds.

(H_1) $I(u) \ge 0$ for all $u \in E$ and $I(u) + |J(u)| \to \infty$ as $\|u\| \to \infty$.

(H_2) $I(u) \le 0$ for all $u \in E$ and $|I(u)| + |J(u)| \to \infty$ as $\|u\| \to \infty$.

Furthermore, we suppose that $\mathscr{K}$ is a sandwich system satisfying

(H_3) $a(\lambda) := \inf_{K \in \mathscr{K}} \sup_K G_\lambda$ is finite for each $\lambda \in \Lambda$.

Theorem 4.1. *Assume that (H_1) (or (H_2)) and (H_3) hold.*

(1) *For almost all $\lambda \in \Lambda$ there exists a constant $k_0(\lambda) := k_0$ (depending only on λ) such that for each $\delta > 0$ there exists a $K \in \mathscr{K}$ such that*

$$\sup_K G_\lambda \le a(\lambda) + \delta,$$

$$\|u\| \le k_0 \quad \text{whenever} \quad u \in K \quad \text{and} \quad G_\lambda(u) \ge a(\lambda) - \delta. \tag{4-1}$$

(2) *For almost all $\lambda \in \Lambda$ there exists a bounded sequence $u_k(\lambda) \in E$ such that*

$$\|G'_\lambda(u_k)\| \to 0, \quad G_\lambda(u_k) \to a(\lambda) := \inf_{K \in \mathscr{K}} \sup_K G_\lambda \quad \text{as } k \to \infty.$$

Corollary 4.2. *The conclusions of Theorem 4.1 hold if we replace Hypothesis (H_3) with*:

(H_3') *There is a sandwich pair A, B such that for each $\lambda \in \Lambda$,*

$$a_0 := \sup_A G_\lambda < \infty, \quad b_0 := \inf_B G_\lambda > -\infty. \tag{4-2}$$

Corollary 4.3. *The conclusions of Theorem 4.1 hold if we replace Hypothesis (H_3) with*:

(H_3'') *There are sets A, B such that A links B and for each $\lambda \in \Lambda$,*

$$a_0 := \sup_A G_\lambda \le b_0 := \inf_B G_\lambda. \tag{4-3}$$

5. Proofs of the theorems

We now give the proof of Theorem 1.4.

Proof. Let X be the set of vector functions $x(t)$ described above. It is a Hilbert space with norm satisfying

$$\|x\|_X^2 = \sum_{j=1}^n \|x_j\|_{H^1}^2.$$

We also write

$$\|x\|^2 = \sum_{j=1}^n \|x_j\|^2,$$

where $\|\cdot\|$ is the $L^2(I)$ norm.

Let

$$N = \{x(t) \in X : x_j(t) \equiv \text{constant for } 1 \le j \le n\},$$

and set $M = N^\perp$. The dimension of N is n, and $X = M \oplus N$. See, for example, [Mawhin and Willem 1989, Proposition 1.3] for details on the following lemma.

Lemma 5.1. *If $x \in M$, then*

$$\|x\|_\infty^2 \le \frac{T}{12}\|\dot{x}\|^2 \quad \textit{and} \quad \|x\| \le \frac{T}{2\pi}\|\dot{x}\|.$$

Define

$$G(x) = \|\dot{x}\|^2 - 2\int_I V(t, x(t))\,dt, \quad x \in X. \tag{5-1}$$

For each $x \in X$ write $x = v + w$, where $v \in N$, $w \in M$. For convenience, we shall follow [Mawhin and Willem 1989] and use the equivalent norm for X:

$$\|x\|_X^2 = \|\dot{w}\|^2 + \|v\|^2.$$

Let

$$I(x) = \|\dot{x}\|^2, \quad J(x) = 2\int_I V(t, x(t))\,dt.$$

By Hypothesis (1),

$$J(v) \to \infty \quad \text{as } \|v\| \to \infty, \ v \in N.$$

Hence,

$$I(x) + |J(x)| \to \infty \quad \text{as } \|x\|_X \to \infty.$$

Let

(5-2) $$G_\lambda(x) = \lambda\|\dot{x}\|^2 - 2\int_I V(t, x(t))\,dt = \lambda I(x) - J(x), \quad x \in X.$$

Hypothesis (1) implies

(5-3) $$\sup_N G_\lambda(v) = -\inf_N J(v) < \infty.$$

If $x \in M$, we have by Hypothesis (2) and Lemma 5.1 that

(5-4) $$\begin{aligned} G_\lambda(x) &\geq \lambda\|\dot{x}\|^2 - 2\int \alpha|x(t)|^2\,dt - B \\ &\geq \Big(\frac{4\pi^2\lambda}{T^2} - 2\alpha\Big)\|x\|^2 - B \geq -B, \end{aligned}$$

provided

(5-5) $$\lambda \geq \alpha T^2/2\pi^2.$$

By Corollary 3.5, M and N form a sandwich pair. Then by Corollary 4.2, for almost every λ satisfying (5-5) there is a bounded sequence $\{x^{(k)}\} \subset X$ such that

(5-6) $$G_\lambda(x^{(k)}) = \lambda\|\dot{x}^{(k)}\|^2 - 2\int_I V(t, x^{(k)}(t))\,dt \to c \geq -B,$$

(5-7) $$(G'_\lambda(x^{(k)}), z)/2 = \lambda(\dot{x}^{(k)}, \dot{z}) - \int_I \nabla_x V(t, x^{(k)}) \cdot z(t)\,dt \to 0, \quad z \in X,$$

(5-8) $$(G'_\lambda(x^{(k)}), x^{(k)})/2 = \lambda\|\dot{x}^{(k)}\|^2 - \int_I \nabla_x V(t, x^{(k)}) \cdot x^{(k)}\,dt \to 0.$$

Since

$$\rho_k = \|x^{(k)}\|_X \leq C,$$

there is a renamed subsequence such that $x^{(k)}$ converges to a limit $x \in X$ weakly in X and uniformly on I. From (5-7) we see that

$$(G'_\lambda(x), z)/2 = \lambda(\dot{x}, \dot{z}) - \int_I \nabla_x V(t, x(t)) \cdot z(t)\,dt = 0, \quad z \in X,$$

from which we conclude easily that x is a solution of (1-4) with $\beta = 1/\lambda$, proving the first statement of the theorem. To prove the second, note that (5-4) implies

$$G_\lambda(x) \geq -B, \quad x \in M.$$

Consequently, if $B < 0$, we see that

$$b_0 = \inf_M G_\lambda(x) > 0.$$

Thus, the solution x satisfies $G_\lambda(x) \geq b_0 > 0$. If x were a constant, we would have $G_\lambda(x) = -J(x) \leq 0$, a contradiction. This gives the result. □

The proof of Theorem 1.5 is similar to that of Theorem 1.4 with the exception of the inequality (5-4) resulting from Hypothesis (2). In its place we reason as follows: If $x \in M$ and $\|\dot{x}\|^2 = 12m^2/T$, then $|x| \leq m$ by Lemma 5.1. Thus, we have by Hypothesis (2′),

$$\begin{aligned} G_\lambda(x) &\geq \lambda\|\dot{x}\|^2 - 2\alpha m^2 - B \\ &\geq (12\lambda - 2\alpha T)m^2/T - B \geq -B, \end{aligned}$$

provided $\lambda \geq \alpha T/6$. The remainder of the proof is essentially the same.

In proving Theorem 1.1, we follow the proof of Theorem 1.4. Hypothesis (1) implies

$$G_\lambda(v) \leq 0, \quad v \in N. \tag{5-9}$$

If $x \in M$ and

$$\|\dot{x}\|^2 = \rho^2 = \frac{12}{T}m^2,$$

then Lemma 5.1 implies that $\|x\|_\infty \leq m$, and we have by Hypothesis (2) that $\int_0^T V(t,x)\,dt \leq \alpha$. Hence,

$$\begin{aligned} G_\lambda(x) &\geq \lambda\|\dot{x}\|^2 - 2\int_0^T V(t,x)\,dt \\ &\geq \lambda\rho^2 - 2\alpha \geq 0, \end{aligned} \tag{5-10}$$

provided $\lambda \geq \alpha T/6m^2$.

If we take

$$A = M \cap B_\rho, \quad B = N,$$

then A links B by [Schechter 1999, Corollary 13.5]. Thus, we see that Hypothesis (H_3'') of Corollary 4.3 holds with G_λ replaced with $-G_\lambda$. By that corollary, there is a bounded sequence satisfying (5-6)–(5-8). The first result now follows as before. To prove the second, let

$$y(t) = v + sw_0,$$

where $v \in N$, $s \geq 0$, and

$$w_0 = (\sin(2\pi t/T), 0, \dots, 0).$$

Then $w_0 \in M$, and

$$\|w_0\|^2 = T/2, \quad \|\dot{w}_0\|^2 = 2\pi^2/T.$$

Note that

$$\|y\|^2 = \|v\|^2 + s^2T/2 = T|v|^2 + Ts^2/2.$$

Consequently,

$$\begin{aligned} G_\lambda(y) = \lambda s^2\|\dot{w}_0\|^2 - 2\int_I V(t, y(t))\,dt &\leq 2\lambda\pi^2 s^2/T - 2\gamma\int_I |y(t)|^2\,dt + B \\ &\leq 2\lambda\pi^2 s^2/T - 2\gamma(\|v\|^2 + Ts^2/2) + B \\ &\leq (2\lambda\pi^2 - \gamma T^2)s^2/T - 2T\gamma|v|^2 + B \to -\infty \text{ as } s^2 + |v|^2 \to \infty. \end{aligned}$$

Take

$$\begin{aligned} &A = \{v \in N : |v\| \leq R\} \cup \{sw_0 + v : v \in N, s \geq 0, \|sw_0 + v\| = R\}, \\ &B = \partial B_\rho \cap M,\ 0 < \rho < R, \end{aligned}$$

where

$$B_\sigma = \{x \in X : \|x\|_X < \sigma\}.$$

By [Schechter 1999, Example 3, page 38], A links B. Moreover, if R is sufficiently large,

$$\sup_A G_\lambda \leq 0 \leq \inf_B G_\lambda. \tag{5-11}$$

Hence, we may apply [Schechter 1999, Corollary 2.8.2] and Corollary 4.3 to conclude that there is a sequence $\{x^{(k)}\} \subset X$ such that

$$G_\lambda(x^{(k)}) = \lambda\|\dot{x}^{(k)}\|^2 - 2\int_I V(t, x^{(k)}(t))\,dt \to c \geq 0, \tag{5-12}$$

$$(G'_\lambda(x^{(k)}), z)/2 = \lambda(\dot{x}^{(k)}, \dot{z}) - \int_I \nabla_x V(t, x^{(k)}) \cdot z(t)\,dt \to 0, \quad z \in X, \tag{5-13}$$

$$(G'_\lambda(x^{(k)}), x^{(k)})/2 = \lambda\|\dot{x}^{(k)}\|^2 - \int_I \nabla_x V(t, x^{(k)}) \cdot x^{(k)}\,dt \to 0. \tag{5-14}$$

Since

$$\rho_k = \|x^{(k)}\|_X \leq C,$$

there is a renamed subsequence such that $x^{(k)}$ converges to a limit $x \in X$ weakly in X and uniformly on I. From (5-13) we see that

$$(G'_\lambda(x), z)/2 = \lambda(\dot{x}, \dot{z}) - \int_I \nabla_x V(t, x(t)) \cdot z(t)\,dt = 0, \quad z \in X,$$

from which we conclude easily that x is a solution of (1-1). By (5-12) we see that

$$G_\lambda(x) \geq c \geq 0,$$

showing that $x(t)$ is not a constant. For if $c > 0$ and $x \in N$, then

$$G_\lambda(x) = -2 \int_I V(t, x(t))\, dt \leq 0.$$

If $c = 0$, we know that $d(x^{(k)}, B) \to 0$ by [Schechter 1999, Theorem 2.1.1]. Hence, there is a sequence $\{y^{(k)}\} \subset B$ such that $x^{(k)} - y^{(k)} \to 0$ in X. If $v \in N$, then

$$(x, v) = (x - x^{(k)}, v) + (x^{(k)} - y^{(k)}, v) \to 0,$$

since $y^{(k)} \in M$. Thus $x \in M$. This completes the proof.

To prove Theorem 1.3, note that Hypothesis (1) implies

$$G_\lambda(v) \leq 0, \quad v \in N. \tag{5-15}$$

If $x \in M$, we have by Hypothesis (2)

$$\begin{aligned}
G_\lambda(x) &\geq \lambda \|\dot{x}\|^2 - 2 \int_{|x|<m} \alpha |x(t)|^2\, dt - C \int_{|x|>m} (|x|^q + 1)\, dt \\
&\geq \lambda \|\dot{x}\|^2 - 2\alpha \|x\|^2 - C(1 + m^{2-q} + m^{-q}) \int_{|x|>m} |x|^q\, dt \\
&\geq \|\dot{x}\|^2 \big(\lambda - (2\alpha T^2/4\pi^2)\big) - C' \int_{|x|>m} |x|^q\, dt \\
&\geq \big(\lambda - (\alpha T^2/2\pi^2)\big) \|x\|_X^2 - C'' \int_I \|x\|_X^q\, dt \\
&\geq \big(\lambda - (\alpha T^2/2\pi^2)\big) \|x\|_X^2 - C''' \|x\|_X^q = \big(\lambda - (\alpha T^2/2\pi^2) - C''' \|x\|_X^{q-2}\big) \|x\|_X^2.
\end{aligned}$$

Hence,

$$G_\lambda(x) \geq \varepsilon \|x\|_X^2, \quad \|x\|_X \leq \rho,\ x \in M \tag{5-16}$$

for $\rho > 0$ sufficiently small, where $\varepsilon < \lambda - (\alpha T^2/2\pi^2)$ is positive. If we take

$$A = M \cap B_\rho, \quad B = N,$$

then A links B by [Schechter 1999, Corollary 13.5]. Thus, Hypothesis (H_3'') of Corollary 4.3 holds with G_λ replaced with $-G_\lambda$. By that corollary, there is a bounded sequence satisfying (5-6)–(5-8). The result now follows as before.

6. Finding the sequences

Proof of Theorem 3.1. Let $M = C_0 + 1$. Then

$$\|\sigma(1)v\| \le M$$

whenever $\sigma \in \Sigma$ satisfies $\|\sigma'(t)\| \le 1$ and $v \in E$ satisfies $\|v\| \le C_0$. If the theorem were false, then there would be a $\delta > 0$ such that

$$\|G'(u)\| \ge 3\delta \tag{6-1}$$

when

$$u \in \{u \in E : \|u\| \le M+1, |G(u)-a| \le 3\delta\}. \tag{6-2}$$

Take $\delta < 1/3$. Since $G \in C^1(E, \mathbb{R})$, for each $\theta < 1$ there is a locally Lipschitz continuous mapping $Y(u)$ of $\hat{E} = \{u \in E : G'(u) \neq 0\}$ into E such that

$$\|Y(u)\| \le 1, \quad \theta\|G'(u)\| \le (G'(u), Y(u)), \quad u \in \hat{E} \tag{6-3}$$

(see, for example, [Schechter 1999]). Take $\theta > 2/3$. Let

$$\begin{aligned} Q_0 &= \{u \in E : \|u\| \le M+1, |G(u)-a| \le 2\delta\}, \\ Q_1 &= \{u \in E : \|u\| \le M, |G(u)-a| \le \delta\}, \\ Q_2 &= E \setminus Q_0, \\ \eta(u) &= d(u, Q_2)/\big(d(u, Q_1) + d(u, Q_2)\big). \end{aligned}$$

It is easily checked that $\eta(u)$ is locally Lipschitz continuous on E and satisfies

$$\begin{cases} \eta(u) = 1 & \text{if } u \in Q_1, \\ \eta(u) = 0 & \text{if } u \in \bar{Q}_2, \\ \eta(u) \in (0, 1) & \text{otherwise.} \end{cases} \tag{6-4}$$

Let

$$W(u) = -\eta(u)Y(u).$$

Then

$$\|W(u)\| \le 1, \quad u \in E.$$

By [Schechter 2009, Theorem 4.5], for each $v \in E$ there is a unique solution $\sigma(t)v$ of the system

$$\sigma'(t) = W(\sigma(t)), \ t \in \mathbb{R}^+, \quad \sigma(0) = v. \tag{6-5}$$

We have

$$\begin{aligned} dG(\sigma(t)v)/dt &= -\eta(\sigma(t)v)(G'(\sigma(t)v), Y(\sigma(t)v)) \\ &\le -\theta\eta(\sigma)\|G'(\sigma)\| \le -3\theta\delta\eta(\sigma). \end{aligned} \tag{6-6}$$

Let $K \in \mathcal{K}$ satisfy the hypotheses of the theorem. Let v be any element of $K \cap Q_1$. Then $\|v\| \le C_0$. If there is a $t_1 \le 1$ such that $\sigma(t_1)v \notin Q_1$, then

$$G(\sigma(1)v) < a - \delta, \tag{6-7}$$

since $\|\sigma(1)v\| \le M$,

$$G(\sigma(1)v) \le G(\sigma(t_1)v)$$

and the right hand side cannot be greater than $a + \delta$ by (6-6). On the other hand, if $\sigma(t)v \in Q_1$ for all $t \in [0, 1]$, then we have by (6-6)

$$G(\sigma(1)v) \le a + \delta - 3\delta\theta < a - \delta.$$

If $v \in K \backslash Q_1$, then we must have

$$G(\sigma(1)v) \le G(v) < a - \delta,$$

since $G(v) \ge a - \delta$ would put v into Q_1. Hence

$$G(\sigma(1)v) < a - \delta, \quad v \in K. \tag{6-8}$$

By hypothesis, $\widetilde{K} = \sigma(1)K \in \mathcal{K}$. This means that

$$G(w) < a - \delta, \quad w \in \widetilde{K}. \tag{6-9}$$

But this contradicts the definition (3-1) of a. Hence (6-1) cannot hold for u satisfying (6-2). This proves the theorem. □

Proof of Theorem 3.2. Since $A \in \mathcal{K}$, clearly $a \le a_0$. Moreover, for any $K \in \mathcal{K}$, we have

$$b_0 = \inf_B G_\lambda \le \inf_{B\cap K} G_\lambda \le \sup_{B\cap K} G_\lambda \le \sup_K G_\lambda.$$

Hence, $b_0 \le a$. Apply Theorem 3.1. □

Proof of Theorem 3.4. Define

$$\mathcal{K} = \{\sigma(1)A : \sigma \in \Sigma\}.$$

Then $\mathcal{K}$ is a sandwich system. To see this, let $K = \widetilde{\sigma}(1)A$ be a set in $\mathcal{K}$. If $\sigma \in \Sigma$, then $\sigma \circ \widetilde{\sigma}$ is also in Σ. Thus, $\mathcal{K}$ is a sandwich system. Let $B = F^{-1}(p)$. If we can show that B satisfies (3-6), then the result will follow from Theorem 3.2. Now (3-6) is equivalent to

$$F^{-1}(p) \cap \sigma(1)N \ne \varnothing, \quad \sigma \in \Sigma.$$

Let $\Omega_R(p)$ be a ball in N with radius R and center p, and let $\sigma(t)$ be any flow in Σ. Since

$$\sigma(t)u - u = \int_0^t \sigma'(\tau)u \, d\tau, \tag{6-10}$$

we have

$$\|\sigma(t)u - \sigma(s)u\| \leq C|t-s|.$$

If $u \in A_R = \partial\Omega_R(p)$, and $v \in B$, we have

$$h(s) := d(\sigma(s)u, B) \leq \|\sigma(s)u - v\| \leq \|\sigma(t)u - v\| + C|t-s|.$$

This implies

$$h(s) \leq h(t) + C|t-s|. \tag{6-11}$$

Moreover, by [Schechter 2009, Lemmas 4.3 and 4.8], $h(s)$ satisfies

$$h(s) \geq m(R) \to \infty \quad \text{as } R \to \infty, \quad 0 \leq s \leq 1,\ u \in \partial\Omega_R(p).$$

Thus,

$$\|\sigma(s)u - F^{-1}(p)\| \geq h(s) \geq m(R) \to \infty, \quad u \in A_R.$$

Consequently,

$$F^{-1}(p) \cap \sigma(1)A_R = \varnothing, \quad \sigma \in \Sigma, \tag{6-12}$$

for R sufficiently large. Now A_R links B; see, for example, [Schechter 1999]. For $\Gamma \in \Phi$, define

$$\Gamma_1(s) = \begin{cases} \sigma(2s) & \text{if } 0 \leq s \leq \frac{1}{2}, \\ \sigma(1)\Gamma(2s-1) & \text{if } \frac{1}{2} < s \leq 1. \end{cases}$$

Clearly, $\Gamma_1 \in \Phi$. Consequently, there is a $t_0 \in [0, 1]$ such that

$$\Gamma_1(t_0)A_R \cap B \neq \varnothing.$$

If $t_0 \leq \frac{1}{2}$, then

$$\sigma(2t_0)A_R \cap B \neq \varnothing,$$

contradicting (6-12). If $t_0 > \frac{1}{2}$, then

$$\sigma(1)\Gamma(2t_0 - 1)A_R \cap B \neq \varnothing.$$

Take $\Gamma(s)u = (1-s)u$. Then $\Gamma \in \Phi$ and $\Gamma(2t_0 - 1)A_R \subset N$. Hence,

$$\sigma(1)N \cap B \neq \varnothing.$$

Thus (3-6) holds, and the theorem is proved. $\square$

7. The monotonicity trick

Proof of Theorem 4.1. We prove conclusion (1) assuming the first of the alternative hypotheses, (H_1).

By (H_1), the map $\lambda \mapsto a(\lambda)$ is nondecreasing. Hence, $a'(\lambda) := da(\lambda)/d\lambda$ exists for almost every $\lambda \in \Lambda$. From this point on, we consider those λ where $a'(\lambda)$ exists. For fixed $\lambda \in \Lambda$, let $\lambda_n \in (\lambda, 2\lambda) \cap \Lambda$, $\lambda_n \to \lambda$ as $n \to \infty$. Then there exists $\bar{n}(\lambda)$ such that

$$a'(\lambda) - 1 \le \frac{a(\lambda_n) - a(\lambda)}{\lambda_n - \lambda} \le a'(\lambda) + 1 \quad \text{for } n \ge \bar{n}(\lambda). \tag{7-1}$$

Next, there exist $K_n \in \mathcal{K}_Q$, $k_0 := k_0(\lambda) > 0$ such that

$$\|u\| \le k_0 \quad \text{whenever} \;\; G_\lambda(u) \ge a(\lambda) - (\lambda_n - \lambda). \tag{7-2}$$

In fact, by the definition of $a(\lambda_n)$, there exists K_n such that

$$\sup_{K_n} G_\lambda(u) \le \sup_{K_n} G_{\lambda_n}(u) \le a(\lambda_n) + (\lambda_n - \lambda). \tag{7-3}$$

If $G_\lambda(u) \ge a(\lambda) - (\lambda_n - \lambda)$ for some $u \in K_n$, then, by (7-1) and (7-3), we have that

$$\begin{aligned} I(u) &= \frac{G_{\lambda_n}(u) - G_\lambda(u)}{\lambda_n - \lambda} \\ &\le \frac{a(\lambda_n) + (\lambda_n - \lambda) - a(\lambda) + (\lambda_n - \lambda)}{\lambda_n - \lambda} \\ &\le a'(\lambda) + 3, \end{aligned} \tag{7-4}$$

and it follows that

$$\begin{aligned} J(u) &= \lambda_n I(u) - G_{\lambda_n}(u) \\ &\le \lambda_n (a'(\lambda) + 3) - G_\lambda(u) \\ &\le \lambda_n (a'(\lambda) + 3) - a(\lambda) + (\lambda_n - \lambda) \\ &\le 2\lambda (a'(\lambda) + 3) - a(\lambda) + \lambda. \end{aligned} \tag{7-5}$$

On the other hand, by (H_1), (7-1), and (7-3),

$$\begin{aligned} J(u) &= \lambda_n I(u) - G_{\lambda_n}(u) \\ &\ge -G_{\lambda_n}(u) \\ &\ge -(a(\lambda_n) + (\lambda_n - \lambda)) \\ &\ge -(a(\lambda) + (\lambda_n - \lambda)(a'(\lambda) + 2)) \\ &\ge -a(\lambda) - \lambda |a'(\lambda) + 2|. \end{aligned} \tag{7-6}$$

Combining (7-4)–(7-7) and (H_1), we see that there exists $k_0(\lambda) := k_0$ (depending only on λ) such that (7-2) holds.

By the choice of K_n and (7-1), we see that

$$\begin{aligned} G_\lambda(u) &\le G_{\lambda_n}(u) \le \sup_{K_n} G_{\lambda_n}(u) \\ &\le a(\lambda_n) + (\lambda_n - \lambda) \\ &\le (a'(\lambda)+1)(\lambda_n - \lambda) + a(\lambda) + (\lambda_n - \lambda) \\ &\le a(\lambda) + (a'(\lambda)+2)(\lambda_n - \lambda) \end{aligned}$$

for all $u \in K_n$. Take n sufficiently large to ensure that $|a'(\lambda)+2|(\lambda_n - \lambda) < \delta$. This proves conclusion (1). Conclusion (2) now follows from Theorem 3.1. The proof under Hypothesis (H_2) is similar, and is omitted. □

In proving Corollary 4.3, we shall make use of the following results of linking. Let E be a Banach space. The set Φ of mappings $\Gamma(t) \in C(E \times [0, 1], E)$ is to have following properties:

(a) For each $t \in [0, 1)$, $\Gamma(t)$ is a homeomorphism of E onto itself and $\Gamma(t)^{-1}$ is continuous on $E \times [0, 1)$.

(b) $\Gamma(0) = I$.

(c) For each $\Gamma(t) \in \Phi$ there is a $u_0 \in E$ such that $\Gamma(1)u = u_0$ for all $u \in E$ and $\Gamma(t)u \to u_0$ as $t \to 1$ uniformly on bounded subsets of E.

(d) For each $t_0 \in [0, 1)$ and each bounded set $A \subset E$ we have

$$\sup_{\substack{0 \le t \le t_0 \\ u \in A}} \{\|\Gamma(t)u\| + \|\Gamma^{-1}(t)u\|\} < \infty.$$

A subset A of E links a subset B of E if $A \cap B = \varnothing$ and, for each $\Gamma(t) \in \Phi$, there is a $t \in (0, 1]$ such that $\Gamma(t)A \cap B \neq \varnothing$.

Theorem [Schechter 1999, Theorem 2.1.1]. *Let G be a C^1-functional on E, and let A, B be subsets of E such that A links B and*

$$a_0 := \sup_A G \le b_0 := \inf_B G.$$

Assume that

$$a := \inf_{\Gamma \in \Phi} \sup_{\substack{0 \le s \le 1 \\ u \in A}} G(\Gamma(s)u)$$

is finite. Then there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to a, \quad G'(u_k) \to 0.$$

If $a = b_0$, then we can also require that

$$d(u_k, B) \to 0.$$

Proof of Corollary 4.3. Let

$$\mathcal{K} = \{\Gamma(s)A : \Gamma \in \Phi,\ s \in I\}.$$

Then $\mathcal{K}$ is a sandwich system. In fact, if $\sigma \in \Sigma$ and $\Gamma \in \Phi$, define

$$\Gamma_1(s) = \begin{cases} \sigma(2s) & \text{if } 0 \le s \le \frac{1}{2}, \\ \sigma(1)\Gamma(2s-1) & \text{if } \frac{1}{2} < s \le 1. \end{cases}$$

Then $\Gamma_1 \in \Phi$. Thus,

$$\sigma(1)K \in \mathcal{K}, \quad \sigma \in \Sigma,\ K \in \mathcal{K}.$$

Since A links B, we have for each $\Gamma(t) \in \Phi$, there is a $t \in (0, 1]$ such that $\Gamma(t)A \cap B \ne \varnothing$. Consequently,

$$B \cap K \ne \varnothing, \quad K \in \mathcal{K}. \tag{7-7}$$

Thus, A, B form a sandwich pair. Let

$$a(\lambda) := \inf_{\Gamma \in \Phi} \sup_{\substack{0 \le s \le 1 \\ u \in A}} G_\lambda(\Gamma(s)u).$$

Then $a(\lambda) := \inf_{K \in \mathcal{K}} \sup_K G_\lambda$ is finite for any $\lambda \in \Lambda$. This shows that Hypothesis (H_3'') implies Hypothesis (H_3). We can now apply Theorem 4.1. □

Proof of Theorem 1.6. Take $\lambda = 1/\beta$. Let $\lambda_0 = \alpha T/6m^2$, and let $\nu < \infty$. By Theorem 1.1, for a.e. $\lambda \in (\lambda_0, \nu)$, there exists u_λ such that $G_\lambda'(u_\lambda) = 0$, $G_\lambda(u_\lambda) = a(\lambda) \ge a(\lambda_0)$. Let λ satisfy $\lambda_0 < \lambda < \nu$. Choose $\lambda_n \to \lambda$, $\lambda_n > \lambda$. Then there exists x_n such that

$$G_{\lambda_n}'(x_n) = 0, \quad G_{\lambda_n}(x_n) = a(\lambda_n) \ge a(\lambda_0).$$

Therefore,

$$\int_\Omega \frac{2V(t, x_n)}{\|x_n\|_X^2}\, dt \le C.$$

Now we prove that $\{x_n\}$ is bounded. If $\|x_n\|_X \to \infty$, let $w_n = x_n/\|x_n\|_X$. Then there is a renamed subsequence such that $w_n \to w$ weakly in X, strongly in $L^\infty(\Omega)$ and a.e. in Ω.

Let Ω_0 be the set where $w \ne 0$. Then $|x_n(t)| \to \infty$ for $t \in \Omega_0$. If Ω_0 had positive measure, then we would have

$$C \ge \int_\Omega \frac{2V(t, x_n)}{\|x_n\|_X^2}\, dt = \int_\Omega \frac{2V(t, x_n)}{x_n^2} |w_n|^2\, dt \ge \int_{w \ne 0} \frac{2V(t, x_n)}{x_n^2} |w_n|^2\, dt \to \infty,$$

showing that $w = 0$ a.e. in Ω. Hence, $w_n \to 0$. Since

$$\|\dot{w}_n\|^2 + \|w_n\|^2 = 1,$$

we have $\|\dot{w}_n\| \to 1$. Define $\theta_n \in [0, 1]$ by

$$G_{\lambda_n}(\theta_n x_n) = \max_{\theta \in [0,1]} G_{\lambda_n}(\theta x_n).$$

For any $c > 0$ and $\overline{w}_n = c w_n$, we have

$$\int_\Omega V(t, \overline{w}_n)\, dt \to 0$$

(see, for example, [Schechter 2008, page 64]). Thus,

$$G_{\lambda_n}(\theta_n x_n) \geq G_{\lambda_n}(c w_n) = c^2 \lambda_n \|\dot{w}_n\|^2 - 2 \int_\Omega V(t, \overline{w}_n)\, dt \to \lambda c^2, \quad n \to \infty.$$

Hence, $G_{\lambda_n}(\theta_n x_n) \geq \lambda c^2/2$ for n sufficiently large. That is, $\lim_{n\to\infty} G_{\lambda_n}(\theta_n x_n) = \infty$. If there is a renamed subsequence such that $\theta_n = 1$, then

$$G_{\lambda_n}(x_n) \to \infty. \tag{7-8}$$

If $0 \leq \theta_n < 1$ for all n, then we have $(G'_{\lambda_n}(\theta_n x_n), x_n) \leq 0$. Therefore,

$$\begin{aligned}\int_\Omega H(t, \theta_n x_n)\, dt &= \int_\Omega \big(\nabla_x V(t, \theta_n x_n)\theta_n x_n - 2V(t, \theta_n x_n)\big)\, dt \\ &= G_{\lambda_n}(\theta_n x_n) - (G'_{\lambda_n}(\theta_n x_n), \theta_n x_n) \\ &\geq G_{\lambda_n}(\theta_n x_n) \to \infty.\end{aligned}$$

By hypothesis,

$$G_{\lambda_n}(x_n) = \int_\Omega H(t, x_n)\, dx \geq \int_\Omega H(t, \theta_n x_n)\, dt / C - \int_\Omega W(t)\, dt \to \infty.$$

Thus, (7-8) holds in any case. But

$$G_{\lambda_n}(x_n) = a(\lambda_n) \leq a(\nu) < \infty,$$

Thus, $\|x_n\|_X \leq C$. It now follows that for a renamed subsequence,

$$G'_\lambda(x_n) \to 0, \quad G_\lambda(x_n) \to a(\lambda) \geq a(\lambda_0).$$

Applying [Schechter 1999, Theorem 3.4.1, page 64] gives the desired solution. □

Proof of Theorem 1.7. This time we take $\lambda_0 = \alpha T^2/2\pi^2$, apply Theorem 1.3 and follow the proof of Theorem 1.6. □

References

[Ambrosetti and Coti Zelati 1993] A. Ambrosetti and V. Coti Zelati, *Periodic solutions of singular Lagrangian systems*, Progress in Nonlinear Differentials Equations and their Appl. **10**, Birkhäuser, Boston, 1993. MR 95b:58054 Zbl 0785.34032

[Ben-Naoum et al. 1994] A. K. Ben-Naoum, C. Troestler, and M. Willem, "Existence and multiplicity results for homogeneous second order differential equations", *J. Differential Equations* **112**:1 (1994), 239–249. MR 95g:58036 Zbl 0808.58013

[Berger and Schechter 1977] M. S. Berger and M. Schechter, "On the solvability of semilinear gradient operator equations", *Advances in Math.* **25**:2 (1977), 97–132. MR 58 #17997 Zbl 0354.47025

[Ekeland and Ghoussoub 2002] I. Ekeland and N. Ghoussoub, "Selected new aspects of the calculus of variations in the large", *Bull. Amer. Math. Soc.* (*N.S.*) **39**:2 (2002), 207–265. MR 2003b:35048 Zbl 1064.35054

[Jeanjean 1999] L. Jeanjean, "On the existence of bounded Palais–Smale sequences and application to a Landesman–Lazer-type problem set on $\mathbf{R}^N$", *Proc. Roy. Soc. Edinburgh Sect. A* **129**:4 (1999), 787–809. MR 2001c:35034 Zbl 0935.35044

[Long 1995] Y. M. Long, "Nonlinear oscillations for classical Hamiltonian systems with bi-even subquadratic potentials", *Nonlinear Anal.* **24** (1995), 1665–1671. MR 96h:34079 Zbl 0824.34042

[Mawhin 1987] J. Mawhin, "Semicoercive monotone variational problems", *Acad. Roy. Belg. Bull. Cl. Sci.* (5) **73**:3-4 (1987), 118–130. MR 89e:49005 Zbl 0647.49007

[Mawhin and Willem 1986] J. Mawhin and M. Willem, "Critical points of convex perturbations of some indefinite quadratic forms and semilinear boundary value problems at resonance", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **3**:6 (1986), 431–453. MR 88a:35023 Zbl 0678.35091

[Mawhin and Willem 1989] J. Mawhin and M. Willem, *Critical point theory and Hamiltonian systems*, Applied Math. Sciences **74**, Springer, New York, 1989. MR 90e:58016 Zbl 0676.58017

[Rabinowitz 1980] P. H. Rabinowitz, "On subharmonic solutions of Hamiltonian systems", *Comm. Pure Appl. Math.* **33**:5 (1980), 609–633. MR 81k:34032 Zbl 0425.34024

[Schechter 1992] M. Schechter, "A generalization of the saddle point method with applications", *Ann. Polon. Math.* **57**:3 (1992), 269–281. MR 94c:58028 Zbl 0780.35001

[Schechter 1993] M. Schechter, "New saddle point theorems", pp. 213–219 in *Generalized functions and their applications* (Varanasi, 1991), edited by R. S. Pathak, Plenum, New York, 1993. MR 94i:58034 Zbl 0846.46027

[Schechter 1999] M. Schechter, *Linking methods in critical point theory*, Birkhäuser, Boston, 1999. MR 2001f:58032 Zbl 0915.35001

[Schechter 2006a] M. Schechter, "Periodic nonautonomous second-order dynamical systems", *J. Differential Equations* **223**:2 (2006), 290–302. MR 2007d:34081 Zbl 1099.34042

[Schechter 2006b] M. Schechter, "Periodic solutions of second-order nonautonomous dynamical systems", *Bound. Value Probl.* **2006** (2006), Art. ID 25104. MR 2007g:34085 Zbl 1140.34366

[Schechter 2008] M. Schechter, "Sandwich pairs in critical point theory", *Trans. Amer. Math. Soc.* **360**:6 (2008), 2811–2823. MR 2008k:35090 Zbl 1143.35061

[Schechter 2009] M. Schechter, *Minimax systems and critical point theory*, Birkhäuser, Boston, 2009. MR 2010e:58009 Zbl 1186.35043

[Silva 1991] E. A. d. B. e. Silva, "Linking theorems and applications to semilinear elliptic problems at resonance", *Nonlinear Anal.* **16**:5 (1991), 455–477. MR 92d:35108 Zbl 0731.35042

[Struwe 1988] M. Struwe, "The existence of surfaces of constant mean curvature with free boundaries", *Acta Math.* **160**:1-2 (1988), 19–64. MR 89a:53012 Zbl 0646.53005

[Struwe 1996] M. Struwe, *Variational methods, Applications to nonlinear partial differential equations and Hamiltonian systems*, 2nd ed., Ergebnisse der Math. und ihrer Grenzgebiete (3) **34**, Springer, Berlin, 1996. MR 98f:49002 Zbl 0864.49001

[Tang 1995] C. Tang, "Periodic solutions of non-autonomous second order systems with γ-quasisubadditive potential", *J. Math. Anal. Appl.* **189**:3 (1995), 671–675. MR 96a:34090 Zbl 0824.34043

[Tang 1998] C.-L. Tang, "Periodic solutions for nonautonomous second order systems with sublinear nonlinearity", *Proc. Amer. Math. Soc.* **126** (1998), 3263–3270. MR 99a:34122 Zbl 0902.34036

[Tang and Wu 2001] C.-L. Tang and X.-P. Wu, "Periodic solutions for second order systems with not uniformly coercive potential", *J. Math. Anal. Appl.* **259**:2 (2001), 386–397. MR 2002e:34070 Zbl 0999.34039

[Tang and Wu 2002] C.-L. Tang and X.-P. Wu, "Periodic solutions for a class of nonautonomous subquadratic second order Hamiltonian systems", *J. Math. Anal. Appl.* **275**:2 (2002), 870–882. MR 2003j:37092 Zbl 1043.34045

[Tang and Wu 2003] C.-L. Tang and X.-P. Wu, "Notes on periodic solutions of subquadratic second order systems", *J. Math. Anal. Appl.* **285**:1 (2003), 8–16. MR 2004g:34074 Zbl 1054.34075

[Wang et al. 2009] Z. Wang, J. Zhang, and Z. Zhang, "Periodic solutions of second order nonautonomous Hamiltonian systems with local superquadratic potential", *Nonlinear Anal.* **70**:10 (2009), 3672–3681. MR 2010b:37168 Zbl 1179.34037

[Willem 1981] M. Willem, "Oscillations forcées systèmes hamiltoniens", Public. Sémin. Analyse Non Linéarie, University of Besançon, 1981.

[Wu and Tang 1999] X.-P. Wu and C.-L. Tang, "Periodic solutions of a class of nonautonomous second order systems", *J. Math. Anal. Appl.* **236** (1999), 227–235. MR 2000g:34069 Zbl 0971.34027

Received May 4, 2010.

MARTIN SCHECHTER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE CA 92697-3875
UNITED STATES

mschecht@math.uci.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

GENERIC FUNDAMENTAL POLYGONS FOR FUCHSIAN GROUPS

AKIRA USHIJIMA

A Dirichlet fundamental polygon for a Fuchsian group is said to be generic if its combinatorial shape is stable under any small permutation of the center of the polygon. Almost all points in the hyperbolic plane are known to be centers of generic fundamental polygons. We prove that the same property holds for points in the boundary of the hyperbolic plane.

1. Introduction

For a given topological space with a group action, a fundamental region is a subset consisting of representatives of the orbits of a given point by the action. In general, it is chosen to be connected. Such regions are used for the study of groups and their actions on spaces; they give tessellation of the spaces, which imply presentations of the groups.

When a metric is given to the space and the group is discrete, the *Dirichlet domain* (also known as the *Voronoi cell*) is an example of a fundamental region; for a point p free under the group action, the Dirichlet domain for p is the set of all points closer to p than any other point in the orbit of p. For discrete groups acting on the hyperbolic plane, such domains are also called Dirichlet fundamental polygons with center p. We simply call them fundamental polygons in what follows.

One interesting question about fundamental polygons is how many different combinatorial shapes of such polygons are obtained from a given hyperbolic surface. This problem was considered for closed surfaces of genus two by Fricke and Klein [1897], and, independently, by Jørgensen and Näätänen [1982]. They showed that there were exactly eight types of "generic" fundamental polygons. Though the precise definition will be given in Section 3, a generic fundamental polygon has a property of stability of its combinatorial shape under any small perturbation of its center. Generic fundamental polygons are therefore in a sense far from the so-called canonical polygons of Fricke [Fricke and Klein 1897; Keen

This work was supported by a Grant-in-Aid for Young Scientists (B) 21740047, Japan.
MSC2000: primary 20H10; secondary 57M60.
Keywords: Fuchsian group, fundamental polygon.

1966]. For closed surfaces of genus two, each generic fundamental polygon has 18 edges, while each canonical polygon has 8 edges.

Besides genus two, there are known facts about numbers of combinatorial shapes of admissible generic fundamental polygons for closed surfaces. The complete list of generic fundamental polygons of genus three was obtained in [Nakamura 2004]. Each such polygon has 30 edges. The formula to calculate possible numbers of combinatorial shapes of generic fundamental polygons for closed surfaces of any genus was obtained in [Bacher and Vdovina 2002]. Counting number of possible types is related to the study of extremal discs in a surface. For further results on this subject, see [Girondo and Nakamura 2007; Vdovina 2008].

Once we have known the number of combinatorial shapes of admissible generic fundamental polygons for a surface, it is also interesting to think about how these fundamental polygons are related to each other. Such a question was proposed in [Näätänen and Penner 1991] as follows: what kind of decomposition is given on a surface by a relation that two points on the surface are equivalent if they are centers of the fundamental polygon with the same combinatorial shape.

A local figure of such a decomposition of a closed surface of genus two was given in [Näätänen 1985]. In this figure, the set of points corresponding to non-generic fundamental polygons seems to have measure zero. Beardon proved [1983, Theorem 9.4.5] that this is true for any Fuchsian group; for any given such group, almost all points in the hyperbolic plane are centers of generic fundamental polygons. A corresponding result for three-dimensional hyperbolic geometry, that is, for Kleinian groups, was proposed in [Jørgensen and Marden 1988]. However, the proof of Lemma 3.1 in that article, which plays an important role in the proof of the main result, is incomplete.

As a first try to give a complete proof of Jørgensen and Marden's result, we applied their strategy to the case for Fuchsian groups in [Díaz and Ushijima 2009]. We obtained an alternative proof of the result of Beardon there.

The idea of fundamental polygons can be generalized to the case where the center lies on the boundary of the hyperbolic plane. The main purpose of this paper is to show, again following the strategy of Jørgensen and Marden, that Beardon's result holds even when the centers lie in the boundary of the hyperbolic plane.

2. Preliminaries

Let $\mathbb{H}^2 := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ be the *upper half-plane model* of the two-dimensional hyperbolic space. It is given as a subset of the complex plane $\mathbb{C}$, but it is also regarded as contained in the *Riemann sphere* $\widehat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$, where ∞ denotes the *point at infinity*. The boundary $\partial\mathbb{H}^2$ of $\mathbb{H}^2$ is considered in $\widehat{\mathbb{C}}$ so that it consists of the real axis $\mathbb{R}$ plus ∞. We set $\overline{\mathbb{H}^2} := \mathbb{H}^2 \cup \partial\mathbb{H}^2$.

A *circle in* $\widehat{\mathbb{C}}$ means either a Euclidean circle in $\mathbb{C}$ or a Euclidean line in $\mathbb{C}$, as a circle through ∞. *Hyperbolic lines* in $\mathbb{H}^2$ (resp. in $\overline{\mathbb{H}^2}$) are obtained as the intersection of $\mathbb{H}^2$ (resp. $\overline{\mathbb{H}^2}$) and circles in $\widehat{\mathbb{C}}$ which are perpendicular to $\mathbb{R}$. For two distinct points z and w in $\overline{\mathbb{H}^2}$, we denote by $[z, w]$ the hyperbolic line segment with endpoints z and w in $\overline{\mathbb{H}^2}$.

The orientation-preserving isometry group of $\mathbb{H}^2$ is known to be isomorphic to the following projective special linear group:

$$\mathrm{PSL}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \, a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\} \Big/ \{\pm I\},$$

where I denotes the identity matrix. The action of an element T in $\mathrm{PSL}_2(\mathbb{R})$ on $\widehat{\mathbb{C}}$ is a Möbius transformation

$$T(z) := \frac{az+b}{cz+d}, \quad z \in \widehat{\mathbb{C}}.$$

The restriction of this action on $\mathbb{H}^2$ is orientation-preserving and isometric with respect to the hyperbolic metric. We denote the set of the fixed points of T in $\overline{\mathbb{H}^2}$ by $\mathrm{Fix}(T)$. Any nontrivial element of $\mathrm{PSL}_2(\mathbb{R})$ is classified into three types according to the number of elements in $\mathrm{Fix}(T)$; a nontrivial element T in $\mathrm{PSL}_2(\mathbb{R})$ is said to be *elliptic* if $\mathrm{Fix}(T)$ coincides with $\mathrm{Fix}(T) \cap \mathbb{H}^2$ that is a one point set, *parabolic* if $\mathrm{Fix}(T)$ coincides with $\mathrm{Fix}(T) \cap \partial\mathbb{H}^2$ that is a one point set, and *hyperbolic* if $\mathrm{Fix}(T)$ coincides with $\mathrm{Fix}(T) \cap \partial\mathbb{H}^2$ that consists of two points. For hyperbolic T, the *axis* $\mathrm{Ax}(T)$ is defined to be the hyperbolic line whose endpoints are the fixed points of T.

For an element T in $\mathrm{PSL}_2(\mathbb{R})$ and a point z in $\mathbb{H}^2 - \mathrm{Fix}(T)$, let

$$\mathrm{B}(z; T) := \left\{ w \in \mathbb{H}^2 \mid d(w, z) = d(w, T(z)) \right\}$$

be the set of points in $\mathbb{H}^2$ that are equidistant from z and $T(z)$ with respect to the hyperbolic distance $d(\cdot, \cdot)$. It is a hyperbolic line, the perpendicular bisector of $[z, T(z)]$. We remark that our definition of $\mathrm{B}(z; T)$, as in [Díaz and Ushijima 2009], differs from the one given in [Jørgensen and Marden 1988]; there $\mathrm{B}(z; T)$ was defined as the perpendicular bisector of $[z, T^{-1}(z)]$.

The definition of $\mathrm{B}(z; T)$ generalizes to the case that the point z lies in $\partial\mathbb{H}^2$. For a point p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$, a hyperbolic line $\mathrm{B}(p; T)$ is defined to be the limit of $\mathrm{B}(z; T)$ as z converges to p. In particular, if p is taken to be ∞, then $\mathrm{B}(\infty; T)$ is the *isometric semicircle* for T^{-1}.

We denote by $\overline{\mathrm{B}(z; T)}$ the closure of $\mathrm{B}(z; T)$ in $\overline{\mathbb{H}^2}$. For further properties and a proof of the following proposition [Jørgensen and Marden 1988, Section 2].

Proposition 1. *Let T be a nontrivial element in* $\mathrm{PSL}_2(\mathbb{R})$.

(1) *If T is elliptic, then* $\mathrm{B}(z;T)$ *contains the fixed point of T for any point z in* $\overline{\mathbb{H}^2}-\mathrm{Fix}(T)$.

(2) *If T is parabolic, then* $\overline{\mathrm{B}(z;T)}$ *contains the fixed point of T for any point z in* $\overline{\mathbb{H}^2}-\mathrm{Fix}(T)$. *If the point z approaches the fixed point ζ of T conically, then* $\overline{\mathrm{B}(z;T)}$ *converges to ζ.*

(3) *If T is hyperbolic, then* $\overline{\mathrm{B}(z;T)}$ *does not contain any fixed point of T for any point z in* $\overline{\mathbb{H}^2}-\mathrm{Fix}(T)$. *Furthermore it intersects perpendicularly with* $\mathrm{Ax}(T)$. *If the point z approaches a fixed point ζ of T, then* $\overline{\mathrm{B}(z;T)}$ *converges to ζ.*

Fuchsian groups are discrete subgroups of the orientation-preserving isometry group of $\mathbb{H}^2$. We regard them as subgroups of $\mathrm{PSL}_2(\mathbb{R})$ in what follows. For a given Fuchsian group Γ and a point w in $\overline{\mathbb{H}^2}$, we define a subset $\mathcal{P}_0(w)$ in $\mathbb{H}^2$ as follows:

$$\mathcal{P}_0(w) := \left\{ z \in \mathbb{H}^2 \;\middle|\; d(z,w) \le d(z,T(w)) \text{ for all } T \in \Gamma \right\}$$

when w is in $\mathbb{H}^2$, and

$$\mathcal{P}_0(w) := \left\{ z \in \mathbb{H}^2 \;\middle|\; \begin{array}{l} \text{for any } T \in \Gamma - \{I\}, \text{ the point } z \text{ lies in the closure of} \\ \text{the component of } \mathbb{H}^2 - \mathrm{B}(w;T) \text{ that is adjacent to } w \end{array} \right\}$$

when w is in $\partial\mathbb{H}^2$. The subset $\mathcal{P}_0(w)$ is a fundamental polygon for Γ when w is taken from $\mathbb{H}^2 - \bigcup_{T\in\Gamma}\mathrm{Fix}(T)$. Then $\mathcal{P}_0(w)$ is called a (*Dirichlet*) *fundamental polygon* for Γ. For a point w in $\partial\mathbb{H}^2$, on the other hand, the subset $\mathcal{P}_0(w)$ is not always a fundamental polygon. Let $\Omega(\Gamma)$ be the *ordinary set* for Γ in $\widehat{\mathbb{C}}$. It is shown in [Beardon 1983, Theorem 9.5.2] that $\mathcal{P}_0(w)$ is a fundamental polygon when w is taken from $\Omega(\Gamma)$. When such w is taken to be ∞, the fundamental polygon $\mathcal{P}_0(w)$ is known as the *Ford fundamental region*. Set $\mathcal{P}(w) := \overline{\mathcal{P}_0(w)} \cap \Omega(\Gamma)$, where $\overline{\mathcal{P}_0(w)}$ means the closure of $\mathcal{P}(w)$ in $\overline{\mathbb{H}^2}$. The point w is called the *center* of $\mathcal{P}_0(w)$, or of $\mathcal{P}(w)$.

3. Generic fundamental polygons and the main result

Since the polygon $\mathcal{P}(w)$ is defined to be the intersection of $\overline{\mathcal{P}_0(w)}$ with $\Omega(\Gamma)$, the vertices of $\mathcal{P}_0(w)$, or the endpoints in $\Omega(\Gamma)$ of edges of $\mathcal{P}_0(w)$, are vertices of $\mathcal{P}(w)$. Fixed points of elliptic elements of order two on edges of $\mathcal{P}_0(w)$ are also called vertices of $\mathcal{P}(w)$. Vertices in $\mathbb{H}^2$ are called *inner vertices*, and those in $\Omega(\Gamma)$ are called *boundary vertices*.

A *cusp* of $\mathcal{P}(w)$ is a parabolic fixed point lying in $\overline{\mathcal{P}(w)}$. Cusps are not boundary vertices, since any parabolic fixed point belongs to the limit set, which is the complement of $\Omega(\Gamma)$ in $\widehat{\mathbb{C}}$.

The edges of $\mathcal{P}(w)$ are either those of $\mathcal{P}_0(w)$ or closed segments in $\mathcal{P}(w)\cap\partial\mathbb{H}^2$. An edge of $\mathcal{P}_0(w)$ as a hyperbolic polygon is decomposed into two edges of $\mathcal{P}(w)$

if it has a vertex corresponding to the fixed point of an elliptic element of order two. The edges of $\mathcal{P}(w)$ which come from those of $\mathcal{P}_0(w)$ are also called *inner edges*. We denote by $\ell(e)$ the hyperbolic line containing an inner edge e of $\mathcal{P}_0(w)$.

An inner vertex v of $\mathcal{P}(w)$ said to have a *vertex cycle of length k* if there is a sequence $T_1 = I$, T_2, $T_3, \ldots, T_k$, $T_{k+1} = T_1 = I$ of elements in Γ such that the sequence $T_1(\mathcal{P}(w)) = \mathcal{P}(w)$, $T_2(\mathcal{P}(w)), \ldots, T_k(\mathcal{P}(w))$ of polygons is a cyclic arrangement around v in the Γ-orbit of $\mathcal{P}(w)$. In other words, the length of a vertex cycle is the number of disjoint vertices of $\mathcal{P}_0(w)$ that are equivalent to v under Γ if v is not fixed by elliptic elements in Γ. The sequence $T_1, T_2, \ldots, T_k$ is called the *vertex cycle* of v.

Definition. For a Fuchsian group Γ, the fundamental polygon $\mathcal{P}(w)$ centered at w in $\mathbb{H}^2 \cup \Omega(\Gamma)$ is said to be *generic* if it satisfies the following conditions:

(1) For an inner vertex, if the length of its vertex cycle is greater than three, then the vertex is the fixed point of an elliptic element in Γ.

(2) For an inner edge e, if $\overline{\ell(e)}$ contains the fixed point of an elliptic or a parabolic element in Γ, the element of Γ defining $\ell(e)$ as the bisector is an elliptic or a parabolic element fixing the point in question. Similarly, if $\ell(e)$ intersects perpendicularly with the axis of a hyperbolic element in Γ, the element of Γ defining $\ell(e)$ as the bisector is a hyperbolic element fixing the axis in question.

(3) Every boundary vertex is an endpoint of exactly one inner edge.

(4) If two inner edges share an endpoint on $\partial\mathbb{H}^2$, then the endpoint is a cusp that is the fixed point of a parabolic element gluing these inner edges.

Analogous notions have been studied before. Our definition of generic fundamental polygons is the two-dimensional counterpart of the definition of generic fundamental polyhedra in [Jørgensen and Marden 1988]. Our Conditions (1), (3), and (4) correspond to those defining Dirichlet polygons in [Beardon 1983, Theorem 9.4.5].

The conditions for generic fundamental polygons have geometric interpretations. Condition (1) and (2) together imply that, if a vertex v is fixed by an elliptic element in Γ, its vertex cycle coincides with the cyclic elliptic subgroup with fixed point v. Another interpretation is that any cone singularity of the surface $\mathbb{H}^2/\Gamma$ is cut by the image of exactly one inner edge. Similarly, Conditions (2) and (3) mean that any border (open end) of $\mathbb{H}^2/\Gamma$ is also cut by the image of exactly one inner edge, and Condition (4) means that any cusp of $\mathbb{H}^2/\Gamma$ is also cut by the image of exactly one inner edge.

Theorem. *For a Fuchsian group Γ, there is a subset $\mathcal{N}_\Gamma$ in $\partial\mathbb{H}^2$ of measure zero such that, for any point w in $(\Omega(\Gamma) \cap \partial\mathbb{H}^2) - \mathcal{N}_\Gamma$, the fundamental polygon $\mathcal{P}(w)$ is generic.*

Proof. To obtain the subset $\mathcal{N}_\Gamma$, we define three families of subsets in $\partial\mathbb{H}^2$. For T_1, T_2, T_3 in $\mathrm{PSL}_2(\mathbb{R})$, define a subset $\mathcal{V}_{T_1,T_2,T_3}$ in $\partial\mathbb{H}^2$ as

$$\mathcal{V}_{T_1,T_2,T_3} := \bigl\{p \in \partial\mathbb{H}^2 \bigm| \overline{\mathrm{B}(p;\, T_1)} \cap \overline{\mathrm{B}(p;\, T_2)} \cap \overline{\mathrm{B}(p;\, T_3)} \neq \varnothing\bigr\}.$$

For T_1, T_2 in $\mathrm{PSL}_2(\mathbb{R})$, define $\mathcal{T}_{T_1,T_2}$ as

$$\mathcal{T}_{T_1,T_2} := \bigl\{p \in \partial\mathbb{H}^2 \bigm| \overline{\mathrm{B}(p;\, T_1)} \cap \overline{\mathrm{B}(p;\, T_2)} \cap \partial\mathbb{H}^2 \neq \varnothing\bigr\}.$$

For T_1, T_2 in $\mathrm{PSL}_2(\mathbb{R})$, define $\mathcal{F}_{T_1,T_2}$ as

$$\mathcal{F}_{T_1,T_2} := \bigl\{p \in \partial\mathbb{H}^2 \bigm| \mathrm{Fix}(T_1) \cap \overline{\mathrm{B}(p;\, T_2)} \neq \varnothing\bigr\}$$

when T_1 is elliptic or parabolic, and

$$\mathcal{F}_{T_1,T_2} := \bigl\{p \in \partial\mathbb{H}^2 \bigm| \mathrm{B}(p;\, T_2) \text{ intersects perpendicularly with the axis of } T_1\bigr\}$$

when T_1 is hyperbolic.

Using these subsets, we define

$$\mathcal{N}_\Gamma := \Bigl(\bigcup_{\{T_1,T_2,T_3\}} \mathcal{V}_{T_1,T_2,T_3}\Bigr) \cup \Bigl(\bigcup_{\{T_4,T_5\}} \mathcal{T}_{T_4,T_5}\Bigr) \cup \Bigl(\bigcup_{(T_6,T_7)} \mathcal{F}_{T_6,T_7}\Bigr),$$

where $\{T_1, T_2, T_3\}$ runs over all triples in Γ that are mutually distinct, nontrivial and neither elliptic with a common fixed point nor parabolic with a common fixed point, $\{T_4, T_5\}$ runs over all pairs in Γ that are distinct, nontrivial and not parabolic with a common fixed point, and (T_6, T_7) runs over all ordered pairs of nontrivial elements in Γ such that $\mathrm{Fix}(T_7)$ and $\mathrm{Fix}(T_6)$ are different.

Each of the indexes in the definition of $\mathcal{N}_\Gamma$ runs over countably many triples and pairs, because Γ is a discrete group (see [Beardon 1983, Exercise 2.3.3], for example,). To see that $\mathcal{N}_\Gamma$ has measure zero, it is thus enough to see that the subsets $\mathcal{V}_{T_1,T_2,T_3}$, $\mathcal{T}_{T_4,T_5}$ and $\mathcal{F}_{T_6,T_7}$ have measure zero. These are shown in Propositions 5, 7 and 9, respectively.

We next use case analysis to show that the fundamental polygon $\mathcal{P}(w)$ is generic for any w in $(\Omega(\Gamma) \cap \partial\mathbb{H}^2) - \mathcal{N}_\Gamma$.

For Condition (1), take a vertex v of $\mathcal{P}_0(w)$. Let S_v be the vertex cycle of v, and k_v the length of S_v. Suppose $k_v > 3$. Then there are at least three nontrivial and distinct elements, say T_a, T_b and T_c, in S_v. Though w is not in $\bigcup_{\{T_1,T_2,T_3\}} \mathcal{V}_{T_1,T_2,T_3}$, the bisectors $\mathrm{B}(w;\, T_a)$, $\mathrm{B}(w;\, T_b)$ and $\mathrm{B}(w;\, T_c)$ contain v. This means that the three elements T_a, T_b and T_c are either elliptic with a common fixed point or parabolic with a common fixed point. If they are parabolic, then $\overline{\mathrm{B}(w;\, T_a)}$, $\overline{\mathrm{B}(w;\, T_b)}$ and $\overline{\mathrm{B}(w;\, T_c)}$ contain a common fixed point in $\partial\mathbb{H}^2$. This contradicts the assumption that v is in $\mathbb{H}^2$. If they are elliptic with a common fixed point, S_v is in a cyclic subgroup of Γ by the discreteness of Γ. Then the cyclic subgroup is generated by an elliptic element, which fixes v.

For Condition (2), take an inner edge e of $\mathcal{P}(w)$. Let T_e be an element in Γ such that $\ell(e)$ coincides with $\mathrm{B}(w; T_e)$. Suppose that an endpoint of $\ell(e)$ coincided with the fixed point of a parabolic element, say T. This means that $\mathrm{Fix}(T) \cap \overline{\mathrm{B}(w; T_e)}$ is nonempty. Since w is not in $\bigcup_{(T_6,T_7)} \mathcal{F}_{T_6,T_7}$, the set $\mathrm{Fix}(T_e)$ coincides with $\mathrm{Fix}(T)$. The same argument is applied to the cases that T_e is either elliptic or hyperbolic.

For Conditions (3), and (4), take a vertex (in the ordinary sense) v^* of $\overline{\mathcal{P}(w)}$ that is in $\partial\mathbb{H}^2$. Suppose that v^* is the endpoint of two inner edges, say a and b, of $\mathcal{P}_0(w)$. Let T_a and T_b are elements in Γ such that $\ell(a)$ and $\ell(b)$ coincide with $\mathrm{B}(w; T_a)$ and $\mathrm{B}(w; T_b)$ respectively. Since w is not in $\bigcup_{\{T_4,T_5\}} \mathcal{T}_{T_4,T_5}$, the elements T_a and T_b are parabolic with a common fixed point. Then v^* is their common fixed point so that it is a cusp of $\mathcal{P}(w)$ given as the endpoint of a and b that are glued together by $T_a = {T_b}^{-1}$.

Furthermore, the argument above implies that a vertex v^* in $\Omega(\Gamma)$ is an endpoint of exactly one inner edge, for any parabolic fixed point is not in $\Omega(\Gamma)$. Such a vertex is a boundary vertex by definition.

We have thus shown that $\mathcal{P}(w)$ is generic for any w in $(\Omega(\Gamma) \cap \partial\mathbb{H}^2) - \mathcal{N}_\Gamma$. $\square$

This theorem is a generalization of [Beardon 1983, Theorem 9.4.5] and [Díaz and Ushijima 2009, Corollary 3.11]. The algebraic equations defining $\mathcal{V}_{T_1,T_2,T_3}$, $\mathcal{T}_{T_4,T_5}$ and $\mathcal{F}_{T_6,T_7}$ can be regarded as defined on $\mathbb{C}$, so the set $\mathcal{N}_\Gamma$ can be extended to one in $\overline{\mathbb{H}^2}$. Both $\mathcal{N}_\Gamma$ and its extension have measure zero, so their complements are dense in $\partial\mathbb{H}^2$ and $\overline{\mathbb{H}^2}$. The fundamental polygon $\mathcal{P}(w)$ is thus generic for almost every point w in $\mathbb{H}^2$ and $\Omega(\Gamma) \cap \partial\mathbb{H}^2$.

4. Propositions

To prove the propositions used in the proof of the theorem, we use the *projective disc model* D^2 of two-dimensional hyperbolic space. As a set it is the open unit disc centered at the origin in two-dimensional real projective space $\mathbb{RP}^2$. We choose the isometry from $\mathbb{H}^2$ to D^2 so that $0, 1, \infty \in \partial\mathbb{H}^2$ are mapped, respectively, to $(0, -1)$, $(1, 0)$, $(0, 1)$ in $\mathbb{RP}^2$.

Given an element $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\mathrm{PSL}_2(\mathbb{R})$ and a point z in $\overline{\mathbb{H}^2} - \mathrm{Fix}(T)$, let $\mathrm{C}(z; T)$ be the pole in $\mathbb{RP}^2$ of the projective line containing the image of $\mathrm{B}(z; T)$ in D^2. For a point $p \in \mathbb{R} \subset \partial\mathbb{H}^2$, a formula in [Jørgensen and Marden 1988, §2.9] tells us the coordinate of $\mathrm{C}(p; T)$:

$$(*)\quad \mathrm{C}(p; T) = \frac{1}{(ap+b)^2 + (cp+d)^2 - p^2 - 1} \times \bigl(2((ap+b)(cp+d) - p),\ (ap+b)^2 - (cp+d)^2 - p^2 + 1\bigr).$$

Using this formula, we have the following proposition, which is a key in proving others.

Proposition 2. *For distinct and nontrivial elements* T_1, T_2 *and* T_3 *in* $\mathrm{PSL}_2(\mathbb{R})$, *the following two conditions are equivalent*:

(1) *The three points* $\mathrm{C}(p;\,T_1)$, $\mathrm{C}(p;\,T_2)$ *and* $\mathrm{C}(p;\,T_3)$ *are collinear in* $\mathbb{RP}^2$ *for any point* p *in* $\partial\mathbb{H}^2 - \bigcup_{i=1}^{3} \mathrm{Fix}(T_i)$.

(2) *The elements* T_1, T_2 *and* T_3 *either*

  (a) *have the same fixed point set*, *or*

  (b) *up to conjugation by an element of* $\mathrm{PSL}_2(\mathbb{R})$, *satisfy*

$$T_1(z) = az, \quad T_2(z) = bz + 1 - b, \quad T_3(z) = \frac{az}{(a-b)z + b},$$

  for some a, b *in* $\mathbb{R} - \{0, 1\}$.

Proof. This result is closely related to Theorem 4.3 in [Díaz and Ushijima 2009], which says that Condition (2) of the proposition holds if and only if, for any $z \in \mathbb{H}^2$, the points z, $T_1(z)$, $T_2(z)$ and $T_3(z)$ are cocyclic in $\widehat{\mathbb{C}}$. We will see in Proposition 3 that this condition is equivalent to $\mathrm{C}(z;\,T_1)$, $\mathrm{C}(z;\,T_2)$ and $\mathrm{C}(z;\,T_3)$ being collinear for any $z \in \mathbb{H}^2 - \bigcup_{i=1}^{3} \mathrm{Fix}(T_i)$. This shows the implication (2) $\Rightarrow$ (1).

Moreover the collinearity of $\mathrm{C}(z;\,T_1)$, $\mathrm{C}(z;\,T_2)$ and $\mathrm{C}(z;\,T_3)$ is an algebraic condition (see proof of the theorem just cited). If T_1, T_2 and T_3 are not one of the triples listed in Condition (2), the algebraic equation is proper, that is, the solution set is nowhere dense in $\mathbb{C}$. This, however, does not guarantee that $\mathbb{R}$ is not contained in the solution set. Thus the implication (1) $\Rightarrow$ (2) is not proved yet.

Let T_1, T_2, T_3 be distinct nontrivial elements of $\mathrm{PSL}_2(\mathbb{R})$ satisfying (1). Then the determinant $\Delta(p)$ of the 2×2 matrix with columns $\mathrm{C}(p;\,T_1) - \mathrm{C}(p;\,T_2)$ and $\mathrm{C}(p;\,T_1) - \mathrm{C}(p;\,T_3)$ vanishes wherever it is defined — that is, for any $p \in \mathbb{R}$ such that the points $\mathrm{C}(p;\,T_1)$, $\mathrm{C}(p;\,T_2)$, $\mathrm{C}(p;\,T_3)$ are not on the line at infinity of $\mathbb{RP}^2$. (Here of course the $\mathrm{C}(p;\,T_i)$ are given by the formula (*).)

Now, as p runs over $\partial\mathbb{H}$, each $\mathrm{B}(p;\,T_i)$ describes a projective line; we can assume without loss of generality that none of these three lines is the line at infinity. (If it is, we conjugate T_1, T_2, T_3 by an element of $\mathrm{PSL}_2(\mathbb{R})$, which is allowed since desired conclusion, Condition (2), is insensitive to conjugation.) Thus the determinant $\Delta(p)$ is defined — and, by assumption, vanishes — for all but finitely many values of p. Hence, in any expression $N(p)/D(p)$ of $\Delta(p)$ as a rational function of p, the numerator $N(p)$ is the zero polynomial. We will show, by analyzing the possible cases, that this implies the desired conclusion.

Case 1. One of T_i (say T_1) is hyperbolic and another (say T_2) is not elliptic.

We first consider the case that the set $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2)$ is not empty. By conjugation, we can assume that T_1 fixes 0 and ∞, and T_2 fixes ∞. The matrix presentations in $\mathrm{SL}_2(\mathbb{R})$ of these elements are

$$T_1 = \begin{pmatrix} a_1 & 0 \\ 0 & 1/a_1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} a_2 & b_2 \\ 0 & 1/a_2 \end{pmatrix}, \quad T_3 = \begin{pmatrix} a_3 & b_3 \\ c_3 & d_3 \end{pmatrix}.$$

The constant term of N is then $4a_2b_2b_3(a_2b_2d_3 - b_3)$. Since a_2 is not zero, we have either $b_2 = 0$, $b_3 = 0$ or $a_2b_2d_3 = b_3$.

When $b_2 = 0$, the coefficient of p^2 is $-4b_3d_3({a_1}^2 - {a_2}^2)({a_2}^2 - 1)$. We have $T_1 = T_2$ in $\mathrm{PSL}_2(\mathbb{R})$ when ${a_1}^2 = {a_2}^2$, we have T_1 is trivial when ${a_2}^2 = 1$, and the case that $b_3 = 0$ will be discussed later. So d_3 is to be 0, which together with $\det T_3 = 1$ implies $-b_3c_3 = 1$. The coefficient of p^3 is then $8({a_1}^2 - {a_2}^2)({a_2}^2 - 1)$. All of the possible cases when it is zero have already been discussed.

When $b_3 = 0$, we have $d_3 = 1/a_3$, for $\det T_3 = 1$. The coefficient of p^3 is then $8a_2b_2c_3({a_2}^2{a_3}^2 - {a_1}^2)/a_3$. It is a straightforward calculation that there will be no new possible cases when either b_2 or c_3 is zero. We thus assume $a_3 = a_1/a_2$ without loss of generality. The polynomial N is then expressed as

$$N(p) = \frac{4}{a_2}\big(({a_1}^2 - {a_2}^2)({a_2}^2 - 1) + a_1{a_2}^2b_2c_3\big)(b_2 - a_1c_3z^2)p^2.$$

We thus have $c_3 = ({a_1}^2 - {a_2}^2)(1 - {a_2}^2)/(a_1{a_2}^2b_2)$. The point $a_2b_2/(1 - {a_2}^2)$ is fixed by T_2. After normalizing this fixed point to be 1, three elements T_1, T_2 and T_3 coincide with the ones in (2b).

When $a_2b_2d_3 = b_3$, we have $c_3 = (a_3d_3 - 1)/(a_2b_2d_3)$, for $\det T_3 = 1$. The coefficient of p is then $8{a_2}^2{b_2}^2({a_2}^2{d_3}^2 - 1)$, which implies $T_2 = T_3$, contrary to assumption.

We next consider the case that $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2)$ is the empty set; here T_1 can be assumed to fix 0 and ∞, and T_2 to fix 1. The matrix presentations in $\mathrm{SL}_2(\mathbb{R})$ of these elements are

$$T_1 = \begin{pmatrix} a_1 & 0 \\ 0 & \dfrac{1}{a_1} \end{pmatrix}, \quad T_2 = \begin{pmatrix} a_2 & \dfrac{1}{a_2 - c_2} - a_2 \\ c_2 & \dfrac{1}{a_2 - c_2} - c_2 \end{pmatrix}, \quad T_3 = \begin{pmatrix} a_3 & \dfrac{a_3d_3 - 1}{c_3} \\ c_3 & d_3 \end{pmatrix}.$$

The coefficient of p^6 is then $-4{a_1}^2c_2{c_3}^3(a_2 - c_2)^2(a_2c_3 - a_3c_2)$. When it is zero, we have $a_2c_3 - a_3c_2 = 0$, otherwise we are in the case already considered. This implies $T_2 = T_3$.

Case 2. Two of T_i (say T_1 and T_2) are parabolic.

We first consider the case that $\mathrm{Fix}(T_1)$ coincides with $\mathrm{Fix}(T_2)$; the common fixed point of T_1 and T_2 is assumed to be ∞. The matrix presentations in $\mathrm{SL}_2(\mathbb{R})$ of these elements are

$$T_1 = \begin{pmatrix} 1 & b_1 \\ 0 & 1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & b_2 \\ 0 & 1 \end{pmatrix}, \quad T_3 = \begin{pmatrix} a_3 & b_3 \\ c_3 & d_3 \end{pmatrix},$$

where $a_3d_3 - b_3c_3 = 1$. It is then a straightforward calculation that there will be no new possible cases.

We next consider the case that the set $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2)$ is empty; T_1 is assumed to fix ∞ and T_2 is assumed to fix 0. The matrix presentations in $\mathrm{SL}_2(\mathbb{R})$ of these elements are

$$T_1 = \begin{pmatrix} 1 & b_1 \\ 0 & 1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & 0 \\ c_2 & 1 \end{pmatrix}, \quad T_3 = \begin{pmatrix} a_3 & b_3 \\ \dfrac{a_3d_3-1}{b_3} & d_3 \end{pmatrix}.$$

The constant term is then $8{b_3}^3(b_1d_3 - b_3)$, which implies $T_1 = T_3$.

Case 3. Two of the T_i are elliptic.

We assume T_1 fixes $\sqrt{-1}$ and T_2 fixes $h\sqrt{-1}$, for some $h > 0$. The matrix presentations in $\mathrm{SL}_2(\mathbb{R})$ of these elements are

$$T_1 = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} \cos\theta_2 & h\sin\theta_2 \\ -\frac{1}{h}\sin\theta_2 & \cos\theta_2 \end{pmatrix}, \quad T_3 = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $\theta_1, \theta_2 \in \mathbb{R}$ and $ad - bc = 1$. Recall $\sin\theta_i \neq 0$ for $i = 1, 2$ since both T_1 and T_2 are not trivial.

Let $p_{2\,1} := h(-1+\cos\theta_2)/\sin\theta_2$ and $p_{2\,2} := h(1+\cos\theta_2)/\sin\theta_2$. These points satisfy $\sqrt{-1} \in \mathrm{B}(p_{2\,i}; T_2)$ for $i = 1, 2$. Since we are assuming (1), we also have $\sqrt{-1} \in \mathrm{B}(p_{2\,i}; T_3)$ for $i = 1, 2$, which means $\mathrm{C}(p_{2\,i}; T_3) \notin \mathbb{R}^2 \subset \mathbb{RP}^2$.

Let D_i be the denominator of $\mathrm{C}(p_{2\,i}; T_3)$ for $i = 1, 2$. They then satisfy

$$D_1 - D_2 = \frac{-4h}{\sin^2\theta_2} d_1, \qquad D_1 + D_2 = \frac{4h\cos\theta_2}{\sin^2\theta_2} d_1 + 2d_2,$$

where $d_1 := (ab + cd)\sin\theta_2 + h(a^2 + c^2 - 1)\cos\theta_2$ and $d_2 := h^2(a^2 + c^2 - 1) + (b^2 + d^2 - 1)$. Since $\mathrm{C}(p_{2\,i}; T_3) \notin \mathbb{R}^2$, we have $D_1 = D_2 = 0$. This implies, by the equations above, $d_1 = d_2 = 0$.

Similarly, there are points $p_{1\,i}$ in $\partial\mathbb{H}^2$ satisfying $h\sqrt{-1} \in \mathrm{B}(p_{1\,i}; T_1)$ for $i = 1, 2$. Let

$$H := \begin{pmatrix} \sqrt{h} & 0 \\ 0 & 1/\sqrt{h} \end{pmatrix}.$$

We then have $\mathrm{Fix}(H^{-1}T_1H) = \{\sqrt{-1}/h\}$ and $\mathrm{Fix}(H^{-1}T_2H) = \{\sqrt{-1}\}$. Applying similar calculations as before we have $d_3 = d_4 = 0$, where

$$d_3 := (ab + cdh)\sin\theta_1 + (a^2 + c^2h^2 - 1)\cos\theta_1,$$
$$d_4 := (a^2 + b^2 - 1) + (c^2 + d^2 - 1)h^2.$$

When $c = 0$, we have $a = 1/d$ and $b = h(d^2 - 1)\cos\theta_2/(d\sin\theta_2)$. Substitute them for both d_2 and d_4 and we have $d_4 - d_2 = (d^4 - 1)(h^2 - 1)\sin^2\theta_2$. The condition $d_4 - d_2 = 0$ implies that T_3 is trivial.

When $c \neq 0$, we have $b = (ad - 1)/c$. Substitute it for both d_2 and d_4 and we have $d_2 - d_4 = c^2(a^2 - d^2)(h^2 - 1)$. When $h^2 = 1$ or $a = d$, a straightforward calculation shows that T_1, T_2 and T_3 have a common fixed point. When $a = -d$, we have the following expression of d_1:

$$d_1 = -a(a^2 + c^2 + 1)\sin\theta_2 + ch(a^2 + c^2 - 1)\cos\theta_2.$$

Suppose $a^2 + c^2 - 1 \neq 0$, otherwise we have no new triple. Then h is expressed as

$$h = \frac{a(a^2 + c^2 + 1)\sin\theta_2}{c(a^2 + c^2 - 1)\cos\theta_2}.$$

Substitute them for d_2 and we have the expression of $\sin^2\theta_2$ as

$$\sin^2\theta_2 = \frac{(a^2 + c^2 - 1)\big((a^2 + c^2 - 1)(a^2 - 1) + 4a^2\big)}{4c^2 - (a^2 + c^2 + 1)}.$$

Similarly, substitute them for d_3 and we have the expression of $\sin^2\theta_1$ as

$$\sin^2\theta_1 = \frac{(2ac)^2}{(a^2 + c^2 + 1) - 4c^2}.$$

These expressions imply, after a few more calculations, that there will be no new triple in this case. This concludes the proof of Proposition 2. $\square$

Proposition 3. *For $i = 1, 2$, let B_i be the perpendicular bisector of the hyperbolic line segment $[z_0, z_i]$ with endpoints z_0 and z_i in $\mathbb{H}^2$, and C_i the pole in $\mathbb{RP}^2$ of the projective line containing the image of B_i. In each of the following triples of statement, the three conditions are equivalent*:

(1-1) *The points z_0, z_1 and z_2 are on a hyperbolic circle with center $w \in \mathbb{H}^2$.*

(1-2) *The projective lines B_1 and B_2 intersect at $w \in \mathbb{H}^2$.*

(1-3) *The projective line in $\mathbb{RP}^2$ through C_1 and C_2 does not intersect $\overline{D^2}$, and its pole corresponds to $w \in \mathbb{H}^2$.*

(2-1) *The points z_0, z_1 and z_2 are on a horocycle with center $w \in \partial\mathbb{H}^2$.*

(2-2) *The projective lines $\overline{\mathrm{B}_1}$ and $\overline{\mathrm{B}_2}$ intersect at $w \in \partial\mathbb{H}^2$.*

(2-3) *The projective line in $\mathbb{RP}^2$ through C_1 and C_2 touches ∂D^2, and the tangential point corresponds to $w \in \partial\mathbb{H}^2$.*

(3-1) *The points* z_0, z_1 *and* z_2 *are on an equidistant point set whose axis is* ℓ.

(3-2) *The projective lines* $\overline{\mathrm{B}_1}$ *and* $\overline{\mathrm{B}_2}$ *intersect perpendicularly with* ℓ.

(3-3) *The projective line in* $\mathbb{RP}^2$ *through* C_1 *and* C_2 *intersects with* D^2, *and it contains the image of* ℓ.

Proof. For $j = 1, 2, 3$, the equivalence between (j-2) and (j-3) comes from the duality for the pole and the projective line in $\mathbb{RP}^2$.

Suppose that (1-1) holds. Since any point on B_1 is equidistant from z_0 and z_1, the center w lies on B_1, and so does on B_2. Thus (1-3) holds, and the converse is also true by the same argument. The equivalence between (2-1) and (2-2) comes by a continuity argument from the equivalence above. The equivalence between (3-1) and (3-2) comes from Proposition 1 (3). □

Lemma 4. *Let* T_1, T_2 *and* T_3 *be mutually distinct and nontrivial elements in a Fuchsian group. Suppose that three points* $\mathrm{C}(p;\,T_1)$, $\mathrm{C}(p;\,T_2)$ *and* $\mathrm{C}(p;\,T_3)$ *are collinear in* $\mathbb{R}P^2$ *for any* p *in* $\partial\mathbb{H}^2 - \bigcup_{i=1}^3 \mathrm{Fix}(T_i)$, *and that there is a point* p_0 *in* $\partial\mathbb{H}^2 - \bigcup_{i=1}^3 \mathrm{Fix}(T_i)$ *such that the set*

$$\mathrm{B}(p_0;\,T_1) \cap \mathrm{B}(p_0;\,T_2) \cap \mathrm{B}(p_0;\,T_3) \quad (\textit{resp.}\ \overline{\mathrm{B}(p_0;\,T_1)} \cap \overline{\mathrm{B}(p_0;\,T_2)} \cap \overline{\mathrm{B}(p_0;\,T_3)})$$

is not empty. Then T_1, T_2 *and* T_3 *belong to a cyclic elliptic* (*resp. parabolic*) *subgroup.*

Proof. Suppose that $\mathrm{C}(p;\,T_1)$, $\mathrm{C}(p;\,T_2)$ and $\mathrm{C}(p;\,T_3)$ are collinear in $\mathbb{R}P^2$ for any p in $\partial\mathbb{H}^2 - \bigcup_{i=1}^3 \mathrm{Fix}(T_i)$. Since Proposition 2(2b) does not occur from elements in a Fuchsian group, the elements T_1, T_2 and T_3 have the same fixed point set. Furthermore, they belong to a cyclic subgroup generated by, say T, since they come from a Fuchsian group.

We first consider the case that T is hyperbolic; in particular, the axis of T is assumed to be contained in the imaginary axis of $\mathbb{C}$. The circle in $\widehat{\mathbb{C}}$ through p, $T_1(p)$, $T_2(p)$ and $T_3(p)$ is a Euclidean line through the origin for any p in $\partial\mathbb{H}^2 - \{0, \infty\}$. Then the hyperbolic lines $\mathrm{B}(p;\,T_i)$ are ultraparallel by Proposition 3 so that there is no such point p_0 in question.

We next consider the case that T is parabolic; in particular, T is assumed to fix ∞. For each $i = 1, 2, 3$, the hyperbolic line $\mathrm{B}(p;\,T_i)$ is contained in a vertical Euclidean line for any p in $\partial\mathbb{H}^2 - \{\infty\}$. The set

$$\overline{\mathrm{B}(p;\,T_1)} \cap \overline{\mathrm{B}(p;\,T_2)} \cap \overline{\mathrm{B}(p;\,T_3)}$$

then coincides with $\{\infty\}$.

Finally, if T is elliptic, then the four points are on a hyperbolic circle centered at the fixed point of T for any p in $\partial\mathbb{H}^2$. □

Proposition 5. *Let T_1, T_2 and T_3 be elements in a Fuchsian group. Suppose that they are mutually distinct, nontrivial and neither elliptic with a common fixed point nor parabolic with a common fixed point. Then $\mathcal{V}_{T_1,T_2,T_3}$ has measure zero.*

Proof. We first consider the case that T_1, T_2 and T_3 belong to a cyclic subgroup of a given Fuchsian group Γ. By the assumption, the subgroup is generated by a hyperbolic element T. By Proposition 1, the set $\mathcal{V}_{T_1,T_2,T_3}$ coincides with the fixed point set $\mathrm{Fix}(T)$ of T, which consists of two points. So $\mathcal{V}_{T_1,T_2,T_3}$ has measure zero.

We next consider the case that T_1, T_2 and T_3 do not belong to any cyclic subgroup of Γ. Define

$$\mathcal{V}'_{T_1,T_2,T_3} := \left\{ p \in \partial\mathbb{H}^2 \;\middle|\; \mathrm{C}(p;\, T_1), \mathrm{C}(p;\, T_2) \text{ and } \mathrm{C}(p;\, T_3) \text{ are collinear in } \mathbb{R}P^2 \right\}.$$

As is mentioned in the proof of Proposition 2, the set $\mathcal{V}'_{T_1,T_2,T_3}$ is the solution set of a real algebraic equation with variable p. By the assumption together with Proposition 2, the equation is proper. So $\mathcal{V}'_{T_1,T_2,T_3}$ has measure zero in $\partial\mathbb{H}^2$. By Proposition 3 the set $\mathcal{V}_{T_1,T_2,T_3}$ is contained in $\mathcal{V}'_{T_1,T_2,T_3}$; when the projective line through $\mathrm{C}(p;\, T_1)$, $\mathrm{C}(p;\, T_2)$ and $\mathrm{C}(p;\, T_3)$ intersects with D^2, the point p is not contained in $\mathcal{V}_{T_1,T_2,T_3}$. The set $\mathcal{V}_{T_1,T_2,T_3}$ then has measure zero in $\partial\mathbb{H}^2$ as well. □

Lemma 6. *Let T_1 and T_2 be distinct and nontrivial elements in* $\mathrm{PSL}_2(\mathbb{R})$. *Suppose that they are not parabolic with a common fixed point. Then there is a point p_0 in $\partial\mathbb{H}^2 - (\mathrm{Fix}(T_1) \cup \mathrm{Fix}(T_2))$ such that $\overline{\mathrm{B}(p_0;\, T_1)}$ and $\overline{\mathrm{B}(p_0;\, T_2)}$ do not share their endpoints.*

Proof. We first consider the case that both T_1 and T_2 are elliptic. Take the hyperbolic line ℓ through their fixed points (choose any hyperbolic line which contains the fixed point if their fixed points are identical). Then there is a point p_0 in $\partial\mathbb{H}^2$ such that ℓ coincides with $\mathrm{B}(p_0;\, T_1)$. On the other hand, $\mathrm{B}(p_0;\, T_2)$ always contains the fixed point of T_2 by Proposition 1. So $\mathrm{B}(p_0;\, T_2)$ intersects with $\mathrm{B}(p_0;\, T_1) = \ell$ at the fixed point of T_2. If ℓ happen to coincide with $\mathrm{B}(p_0;\, T_2)$, then take another point p_0 satisfying the same condition, or switch T_1 and T_2 and do the same argument. Then you can find a point in question on $\partial\mathbb{H}^2$ since T_1 and T_2 are different.

We next consider the case that T_1 is elliptic and that T_2 is either parabolic or hyperbolic. For any point in $\mathbb{H}^2 - \mathrm{Fix}(T_2)$, there is p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T_2)$ such that $\mathrm{B}(p;\, T_2)$ contains the point. So there is a point p_0 in $\partial\mathbb{H}^2$ such that $\mathrm{B}(p_0;\, T_2)$ contains the fixed point of T_1, which is on $\mathrm{B}(p_0;\, T_1)$ since T_1 is elliptic.

We then consider the case that both T_1 and T_2 are not elliptic. Suppose first that their fixed point sets are different. Let p_1 be a fixed point of T_1 which is not fixed by T_2, for example. Since T_1 and T_2 are not elliptic, $\mathrm{B}(p;\, T_1)$ converges to p_1 when p approaches to p_1. A point p close enough to p_1 is mapped to another point $T_2(p)$ by T_2, which is not close to p_1. Then the endpoints of $\overline{\mathrm{B}(p;\, T_2)}$ are not close to p_1 as well, which means that $\overline{\mathrm{B}(p;T_1)}$ and $\overline{\mathrm{B}(p;T_2)}$ do not share their endpoints.

If T_1 and T_2 are not elliptic and their fixed point sets coincide with each other, then they are hyperbolic by the assumption. Then $\overline{\mathrm{B}(p;T_1)}$ and $\overline{\mathrm{B}(p;T_2)}$ are ultraparallel for any p in $\partial\mathbb{H}^2 - (\mathrm{Fix}(T_1) \cup \mathrm{Fix}(T_2))$ by Proposition 3. □

Proposition 7. *Let T_1 and T_2 be distinct and nontrivial elements in* $\mathrm{PSL}_2(\mathbb{R})$*. Suppose that they are not parabolic with a common fixed point. Then $\mathcal{T}_{T_1,T_2}$ has measure zero.*

Proof. The set $\mathcal{T}_{T_1,T_2}$ is defined as the solution set of an algebraic equation. Actually it is given by calculating the Euclidean distance between the origin in D^2 and the projective line through $\mathrm{C}(p;T_1)$ and $\mathrm{C}(p;T_2)$. See [Jørgensen and Marden 1988, Lemma 3.3]. Lemma 6 means that the equation is proper. So $\mathcal{T}_{T_1,T_2}$ has measure zero. □

Lemma 8. *Let T be a nontrivial element in* $\mathrm{PSL}_2(\mathbb{R})$*. If there is a point w in $\overline{\mathbb{H}^2}$ such that $\overline{\mathrm{B}(p;T)}$ contains w for any p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$, then T is either elliptic or parabolic and w is its fixed point. Similarly, for a hyperbolic line ℓ, if $\mathrm{B}(p;T)$ intersects perpendicularly with ℓ for any p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$, then T is hyperbolic and ℓ is its axis.*

Proof. Suppose that T fixes w. Then the statement holds only if T is elliptic or parabolic by Proposition 1.

Suppose that T does not fix w. If T is elliptic, there is a point p_0 in $\partial\mathbb{H}^2$ such that $\mathrm{B}(p_0;T)$ does not contain w, for the fixed point of T is the unique point that lies in $\mathrm{B}(p;T)$ for any p in $\partial\mathbb{H}^2$. If T is parabolic, then $\overline{\mathrm{B}(p;T)}$ does not contain w for any point p close enough to the fixed point of T, for $\overline{\mathrm{B}(p;T)}$ converges to the fixed point. The same argument holds when T is hyperbolic.

For a hyperbolic line ℓ, consider its image in D^2. Let C_ℓ be the pole of the image of ℓ. Then we interpret the assumption as the projective line L_p containing the image of $\mathrm{B}(p;T)$ contains C_ℓ for any p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$. If T is elliptic or parabolic, then L_p contains the image of its fixed point for any p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$. If T is hyperbolic, then L_p contains the pole of the image of its axis for any p in $\partial\mathbb{H}^2 - \mathrm{Fix}(T)$. This implies that, if the assumption holds, then T is hyperbolic with axis ℓ. □

Proposition 9. *Let T_1 and T_2 be distinct and nontrivial elements in* $\mathrm{PSL}_2(\mathbb{R})$*. Suppose that* $\mathrm{Fix}(T_1)$ *and* $\mathrm{Fix}(T_2)$ *are different. Then $\mathcal{F}_{T_1,T_2}$ has measure zero.*

Proof. Let z_1 be a point in $\mathbb{RP}^2$ which corresponds to the fixed point of T_1 if it is elliptic or parabolic, or to the pole of the axis of T_1 if it is hyperbolic. We denote by ℓ_2 the projective line in $\mathbb{RP}^2$ containing the image of $\mathrm{B}(p;T_2)$. Then $\mathcal{F}_{T_1,T_2}$ is interpreted as

$$\mathcal{F}_{T_1,T_2} = \{p \in \partial\mathbb{H}^2 \mid \ell_2 \text{ contains } z_1\}.$$

Using this interpretation, the set $\mathcal{F}_{T_1,T_2}$ is defined as the solution set of an algebraic equation. To see it, use the Minkowski space model. The equation is given by the Minkowski inner product of p_1 and the normal vector of ℓ_2. Lemma 8 means that this algebraic equation is proper. So the set $\mathcal{F}_{T_1,T_2}$ has measure zero. □

Acknowledgement

The author thanks Professor Raquel Díaz and Professor Makoto Sakuma for fruitful discussion on this research. The author is grateful to the hospitality of Professor Caroline Series and the University of Warwick, and Professor Albert Marden and the University of Minnesota. Thank you also to the referee for his/her valuable suggestions.

References

[Bacher and Vdovina 2002] R. Bacher and A. Vdovina, "Counting 1-vertex triangulations of oriented surfaces", *Discrete Math.* **246**:1-3 (2002), 13–27. MR 2003d:05106 Zbl 0994.05058

[Beardon 1983] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics **91**, Springer, New York, 1983. MR 85d:22026 Zbl 0528.30001

[Díaz and Ushijima 2009] R. Díaz and A. Ushijima, "On the properness of some algebraic equations appearing in Fuchsian groups", *Topology Proc. Auburn Univ.* **33** (2009), 81–106. MR 2010i:30056 Zbl 1179.30039

[Fricke and Klein 1897] R. Fricke and F. Klein, *Vorlesungen über die Theorie der automorphen Functionen. Erster Band; Die gruppentheoretischen Grundlagen*, B. G. Teubner, Leipzig, 1897.

[Girondo and Nakamura 2007] E. Girondo and G. Nakamura, "Compact non-orientable hyperbolic surfaces with an extremal metric disc", *Conform. Geom. Dyn.* **11** (2007), 29–43. MR 2007k:30087 Zbl 1185.30047

[Jørgensen and Marden 1988] T. Jørgensen and A. Marden, "Generic fundamental polyhedra for Kleinian groups", pp. 69–85 in *Holomorphic functions and moduli, II* (Berkeley, 1986), edited by D. Drasin, Math. Sci. Res. Inst. Publ. **11**, Springer, New York, 1988. MR 89h:30061 Zbl 0673.30034

[Jørgensen and Näätänen 1982] T. Jørgensen and M. Näätänen, "Surfaces of genus 2: generic fundamental polygons", *Quart. J. Math. Oxford Ser.* (2) **33**:132 (1982), 451–461. MR 84c:51029

[Keen 1966] L. Keen, "Canonical polygons for finitely generated Fuchsian groups", *Acta Math.* **115** (1966), 1–16. MR 32 #1349 Zbl 0144.34101

[Näätänen 1985] M. Näätänen, "On the stability of identification patterns for Dirichlet regions", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **10** (1985), 411–417. MR 86k:30054

[Näätänen and Penner 1991] M. Näätänen and R. C. Penner, "The convex hull construction for compact surfaces and the Dirichlet polygon", *Bull. London Math. Soc.* **23**:6 (1991), 568–574. MR 93c:30056

[Nakamura 2004] G. Nakamura, "Generic fundamental polygons for surfaces of genus three", *Kodai Math. J.* **27**:1 (2004), 88–104. MR 2005a:30069 Zbl 1059.30032

[Vdovina 2008] A. Vdovina, "On the number of optimal surfaces", pp. 557–567 in *The Zieschang Gedenkschrift*, edited by M. Boileau et al., Geom. Topol. Monogr. **14**, Geom. Topol., Coventry, 2008. MR 2010a:53055 Zbl 1138.53037

Received May 6, 2010. Revised February 16, 2011.

AKIRA USHIJIMA
FACULTY OF MATHEMATICS AND PHYSICS
INSTITUTE OF SCIENCE AND ENGINEERING
KANAZAWA UNIVERSITY
KANAZAWA 920-1192
JAPAN
ushijima@kenroku.kanazawa-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

STABILITY OF THE KÄHLER–RICCI FLOW IN THE SPACE OF KÄHLER METRICS

KAI ZHENG

We prove that on a Fano manifold M admitting a Kähler–Ricci soliton (ω, X), if the initial Kähler metric ω_{φ_0} is close to ω in a certain weak sense, then the weak Kähler–Ricci flow exists globally and converges in the sense of Cheeger and Gromov. In particular, φ_0 is not assumed to be K_X-invariant. The methods used are based on the metric geometry of the space of the Kähler metrics and are potentially applicable to other stability problems of geometric flows near the corresponding critical metrics.

1. Introduction

The Ricci flow was first introduced by Hamilton [1982] and now plays an important role in understanding the geometric and topological structure of the manifolds it lives on. We call the Ricci flow a *Kähler–Ricci flow* if the underlying manifold is a Kähler manifold. The *normalized Kähler–Ricci flow* is given by

$$\text{(1-1)}\qquad \begin{cases} \dfrac{\partial}{\partial t}\omega = -\operatorname{Ric} + \lambda\omega, \\ \omega(0) = \omega_{\varphi_0}, \end{cases}$$

where $\omega(0)$ stays in the canonical class $2\pi C_1(M)$ and λ is the sign of the first Chern class. Cao [1985] first showed that the Kähler–Ricci flow (1-1) has long-time existence. He also proved that the Kähler–Ricci flow converges to a Kähler–Einstein metric when the first Chern class is negative or zero. Now we restrict ourselves to the case where the first Chern class is positive. Since the Kähler–Ricci flow preserves the Kähler class, we can rewrite the Kähler–Ricci flow in terms of the Ricci potential:

$$\text{(1-2)}\qquad \begin{cases} \dfrac{\partial\varphi}{\partial t} = \log\dfrac{\omega_\varphi^n}{\omega^n} + \varphi - h_\omega + a(t), \\ \varphi(0) = \varphi_0, \end{cases}$$

where a is a function of t and the Ricci potential h_ω of the reference metric ω is

MSC2000: primary 32Q20, 53C25; secondary 53C55, 58E11.
Keywords: Kähler–Ricci flow, space of Kähler metrics, stability.

defined by

$$\sqrt{-1}\,\partial\bar{\partial}h_\omega = \mathrm{Ric}(\omega) - \omega \quad \text{and} \quad \int_M e^{h_\omega}\omega^n = \mathrm{Vol}(M). \tag{1-3}$$

The convergence of the Kähler–Ricci flow has been studied by many authors. Chen and Tian [2002; 2006] proved it for Kähler–Einstein manifolds under the assumption of positivity of the Ricci curvature along the flow. Perelman (unpublished; see detailed proof in [Sesum and Tian 2008]) obtained an estimate of the Kähler–Ricci flow and proved that it converges to a Kähler–Einstein metric in the sense of Cheeger–Gromov, when one exists for any initial Kähler metric. Tian and Zhu [2007] extended this to the case of a Kähler–Ricci soliton for a K_X-invariant initial metric; a *Kähler–Ricci soliton* is a Kähler metric such that there is a holomorphic vector field X satisfying

$$L_X\omega = \mathrm{Ric} - \omega. \tag{1-4}$$

Since the right side of (1-4) is real-valued, we obtain $L_{\mathrm{Im}\,X}\omega = 0$ and $\mathrm{Im}\,X$, the imaginary part of X, generates a one-parameter isometry group K_X. Without assuming that M admits a Kähler–Einstein metric or a Kähler–Ricci soliton, the analytic and geometric conditions of the convergence of the Kähler–Ricci flow are studied in [Phong and Sturm 2006; Phong et al. 2009, 2008; 2011, Tosatti 2010; Székelyhidi 2010; Munteanu 2009; Pali 2009; Chen and Wang 2010; Rubinstein 2009.

In order to study the asymptotic behavior of the Kähler–Ricci flow, we consider the flow's stability problem. That is, on a Kähler manifold M admitting a Kähler–Ricci soliton (ω, X), for what kind of neighborhood of ω does the Kähler–Ricci flow with initial datum in that neighborhood converge (in some sense — maybe exponentially) to the Kähler–Ricci soliton?

This stability problem has been investigated by many people; for references, see [Chen and Li 2009]. That work and [Tian and Zhu 2008] consider perturbing both the initial metric and the complex structure near a Kähler–Einstein metric.

In this paper, we focus on perturbing the initial metric near the Kähler–Ricci soliton without changing the complex structure. The main results of this paper are as follows, where $\mathcal{N}(\epsilon_0; B, p)$ is a small neighborhood of the zero function, to be specified in Section 6.

First we will give a direct proof, based on the geometry of the space of Kähler metrics, of long-time existence and convergence in the Cheeger–Gromov sense, within the frame of Donaldson's program [2004]. The result is this:

Theorem 1.1. *If a Kähler manifold admits a Kähler–Ricci soliton* (ω, X), *there exists a positive constant* ϵ_0 *such that, if the initial potential* φ_0 *stays in* $\mathcal{N}(\epsilon_0; B, p)$, *the weak Kähler–Ricci flow exists globally and converges in the Cheeger–Gromov* *sense. If, moreover,* φ_0 *is* K_X*-invariant, the weak modified Kähler–Ricci flow converges exponentially to a unique Kähler–Ricci soliton nearby.*

When the Futaki invariant vanishes, it is obvious that the holomorphic vector fields X is zero and the Kähler–Ricci soliton is a Kähler–Einstein metric.

Theorem 1.2. *On a Kähler–Einstein manifold, there exists a positive constant* ϵ_0 *such that, if the initial potential* φ_0 *stays in* $\mathcal{N}(\epsilon_0; B, p)$, *the weak Kähler–Ricci flow exists globally and converges exponentially to a unique Kähler–Einstein metric nearby.*

Simon [1983] studied the asymptotic behavior of the gradient flow of the variation problem via the Łojasiewicz–Simon inequality, which compares the distance to the critical set with the norm of the gradient of the functional in the L^2 space under the condition that the functional should be analytic. The underlying idea is to reduce the infinite-dimensional problem to a finite-dimensional problem. Perelman [2002] introduced a new functional, called the μ functional, and pointed out that the Ricci flow is the gradient flow of the μ functional up to a diffeomorphism.

We will not apply the Łojasiewicz–Simon inequality to the μ functional directly. Instead, we provide a new approach to the study of the asymptotic behavior of the flow which is merely a pseudogradient flow of some functional, since in the Kähler setting, geometry gives us more information. To be precise, the critical set in the space of Kähler metrics is a finite-dimensional Riemannian symmetric space, which we will explain later.

Since the Kähler–Ricci flow is the pseudogradient flow of the K-energy, in order to make the mechanism of our proof more clear, we first prove Theorem 1.2 under the assumption that the $C^{2,\alpha}$ norm of φ_0 is small. Then we generalize our approach to the case of a Kähler–Ricci soliton (Theorem 1.1).

A sketch of the proofs goes as follows. We first prove that the Kähler–Ricci flow (1-2), after pullback by the corresponding holomorphic transformations, will always stay in a small neighborhood of the background Kähler–Einstein metric. When M has no nontrivial holomorphic vector field, it is not necessary to find the transformations; Section 3 gives a proof of this. However, in general, when M admits nontrivial holomorphic vector fields, we need a new method, developed in Section 4A, to pick up the appropriate transformations following the trace of the Kähler–Ricci flow in the space of normalized Kähler potential $\mathcal{H}_0$; see (2-2). It has been shown by Mabuchi [1987], Donaldson [1999] and Semmes [1992] independently that $\mathcal{H}_0$ is an infinite-dimensional symmetry space of negative curvature. Later, Chen [2000b] proved $\mathcal{H}_0$ is also a metric space. Since the space $\mathcal{E}_0$ of potentials of Kähler–Einstein metrics is a totally geodesic submanifold in $\mathcal{H}_0$, the projection ρ minimizing the distance function from the Kähler–Ricci flow to $\mathcal{E}_0$ is uniquely determined. The Bando–Mabuchi uniqueness theorem [1987] on the Kähler–Einstein metric implies that ω_ρ is different from the reference Kähler–Einstein metric by a holomorphic transformation. The projection Kähler–Einstein metric is exactly the new reference metric we've acquired.

Another way to derive a holomorphic transformation (Section 7) of $\varphi \in \mathscr{H}_0$ is to minimize in $\mathscr{E}_0$ the $I - J$ functional, introduced in [Bando and Mabuchi 1987] to prove the uniqueness of the Kähler–Einstein metric. However, this method cannot be applied in our case directly, since in general the hessian of the $I - J$ functional is not strictly positive; that is, the minimizer is not unique. Nevertheless, as we observed when the $C^{2,\alpha}$ norm of φ is small, the hessian of the $I - J$ functional is indeed strictly positive. Therefore, the holomorphic transformation is uniquely determined.

In Section 5, we prove stability of the Kähler–Ricci flow near a Kähler–Ricci soliton (ω, X), similarly to the case of Kähler–Einstein metric. We use (ω, X) as the background metric. We first prove that the Kähler–Ricci flow (1-2) with small $C^{2,\alpha}$ initial Kähler potential will always stay in a small neighborhood of ω in the Cheeger–Gromov sense. The key idea is to use Perelman's μ functional [2002] instead of the K-energy, since the hessian of the μ functional is nonnegative at a Kähler–Ricci soliton within the canonical class [Tian and Zhu 2008]. Furthermore, we reparametrize the Kähler–Ricci flow (1-1) by the automorphisms $\varsigma(t)$ generated by the real part $\operatorname{Re} X$ of X such that

$$
\begin{cases}
\dfrac{\partial}{\partial t}\omega_\phi = -\operatorname{Ric}(\omega_\phi) + \omega_\phi + L_{\operatorname{Re} X}\omega_\phi, \\
\omega_{\phi(0)} = \omega_{\varphi_0}.
\end{cases}
\tag{1-5}
$$

It is obvious that the Kähler–Ricci soliton is the stationary solution of the modified Kähler–Ricci flow (1-5). Since the Kähler–Ricci soliton (ω, X) is K_X-invariant and the Kähler–Ricci flow is also invariant under the holomorphic diffeomorphism, we assume without loss of generality that the initial datum is K_X-invariant. Then the exponential convergence of the modified Kähler–Ricci flow follows from [Phong et al. 2011].

Finally, in Section 6, at a fixed time, we show that the $C^{2,\alpha}$ norm of the potential is small when the initial value is small under certain weak conditions. The main idea is to use the estimate introduced in [Chen et al. 2008].

As a corollary of Theorem 1.1, we deduce that the limit metric of the Kähler–Ricci flow is unique. Let $\{\varphi(t_i)\}$ be a sequence of solutions of the Kähler–Ricci flow converging to a Kähler–Einstein metric or Kähler–Ricci soliton g_∞, if one exists; then there exists some $\varphi \in \{\varphi(t_i)\}$ satisfying the stability condition of Theorem 1.1. According to that condition, the Kähler–Ricci flow with initial value φ converges exponentially to a Kähler–Einstein metric g^1_∞ or Kähler–Ricci soliton, respectively. Further, since we assume that $\{\varphi(t_i)\} \to g_\infty$, we must have $g^1_\infty = g_\infty$.

We stress that the approach used to prove Theorem 1.1 is also applicable to the case of the general pseudogradient flow: neither the condition that the flow is the gradient flow of some functional, nor Perelman's deep estimate, nor a prior long-time existence of the flow is required. It is possible that our method can be used to solve other, similar problems of geometric flow, such as the stability of the pseudo-Calabi flow near a constant scalar curvature Kähler metric in [Chen and Zheng 2010] and of the Calabi flow near a extremal metric in [Huang and Zheng 2010].

The paper is organized as follows: in Section 2 we review known results on the space of Kähler metrics and the well-posedness of the pseudo-Calabi flow obtained in [Chen and Zheng 2010] — see (2-10). In Sections 3 and 4 we prove Theorem 1.2 under the assumption that the $C^{2,\alpha}$ norm of the initial Kähler potential is small. Then we prove Theorem 1.1 under the same assumption in Section 5. Finally, in Section 6 we explain how to weaken the initial condition to the one stated in Theorem 1.2 and Theorem 1.1. In Section 7, we explain another method to choose the holomorphic transformation.

2. Notation and basic results

Let M be a compact Kähler manifold of complex dimension n with positive first Chern class $C_1(M)$ and let ω be a Kähler form representing the canonical class $2\pi C_1(M)$. In a local holomorphic coordinate $z_1, z_2, \dots, z_n$, the form ω is expressed by

$$\omega = \sqrt{-1}\sum_{i=1}^{n} g_{i\bar{j}}\, dz^i \wedge dz^{\bar{j}}.$$

The corresponding Riemannian metric is given by

$$g = \sum_{i=1}^{n} g_{i\bar{j}}\, dz^i \otimes dz^{\bar{j}}.$$

For a Kähler metric ω, the volume form is

$$dV = \omega^n = (\sqrt{-1})^n \det(g_{i\bar{j}})\, dz^1 \wedge dz^{\bar{1}} \wedge \cdots \wedge dz^n \wedge dz^{\bar{n}}.$$

The Ricci form

$$\mathrm{Ric} = \sqrt{-1}\sum_{i=1}^{n} R_{i\bar{j}}\, dz^i \wedge dz^{\bar{j}} = -\sqrt{-1}\,\partial\bar{\partial}\log\det\omega^n$$

is a closed real $(1,1)$-form and belongs to $2\pi C_1(M)$. Accordingly, the scalar curvature satisfies

$$S\omega^n = n\,\mathrm{Ric}\wedge\omega^{n-1}.$$

A direct calculation gives the average scalar curvature:

$$\underline{S} = \frac{1}{V}\int_M S\, dV = \frac{n}{V}\int_M \mathrm{Ric}\wedge\omega^{n-1} = n.$$

Let $\mathcal{K}$ be the set of Kähler forms on M representing $2\pi C_1(M)$ and let $\mathcal{E}$ be the set of Kähler–Einstein metrics in $\mathcal{K}$. According to the $\partial\bar{\partial}$ lemma, for any Kähler metric ω' in $\mathcal{K}$ there exists a smooth real-valued function φ such that $\omega' = \omega + \sqrt{-1}\,\partial\bar{\partial}\varphi$. Then the space of Kähler potentials of $\mathcal{K}$ is given by

$$\mathcal{H} = \bigl\{\varphi \in C^\infty(M, \mathbb{R}) \mid \omega + \sqrt{-1}\,\partial\bar{\partial}\varphi \in \mathcal{K}\bigr\}.$$

Apparently, we have an isomorphism

$$T\mathcal{H} \cong \mathcal{H} \times C^\infty(M, \mathbb{R}).$$

Mabuchi [1987], Donaldson [1999] and Semmes [1992] independently defined a Riemannian metric on $\mathcal{H}$ by

$$\int_M f_1 f_2 \omega_\varphi^n,$$

for any $f_1, f_2 \in T_\varphi\mathcal{H}$. For any path $\varphi(t)$ $(0 \le t \le 1)$ in $\mathcal{H}$, the length is given by

$$L(\varphi(t)) = \int_0^1 \sqrt{\int_M \varphi'(t)^2 \omega_{\varphi(t)}^n}\, dt, \tag{2-1}$$

and the geodesic equation is

$$\varphi''(t) - \tfrac{1}{2}|\nabla_t \varphi'(t)|_{\varphi(t)}^2 = 0,$$

where $'$ denotes differentiation in t and ∇_t denotes the covariant derivative for the metric $g_{\varphi(t)}$. The geodesic equation enables us to define the connection on the tangent bundle. For any tangent vector field $\psi(t)$ along the path $\varphi(t)$, the covariant derivative along $\varphi(t)$ is defined by

$$D_t\psi = \frac{\partial\psi}{\partial t} - \tfrac{1}{2}(\nabla_t\psi, \nabla_t\varphi')_{g_\varphi}.$$

Then the connection at φ is given by

$$G(X|Y)(\psi_1, \psi_2) = -\tfrac{1}{2}(\nabla\psi_1, \nabla\psi_2)_{g_\varphi},$$

for any ψ_1 and ψ_2 in $T_\varphi\mathcal{H}$. $G(X|Y)$ is torsion-free and metric-compatible.

Theorem 2.1 [Mabuchi 1987; Donaldson 1999; Semmes 1992]. *The Riemannian manifold $\mathcal{H}$ is an infinite-dimensional symmetric space; it admits a Levi-Civita connection whose curvature is covariant constant. At a point $\varphi \in \mathcal{H}$ the curvature is given by*

$$R_\varphi(\delta_1\varphi, \delta_2\varphi)\delta_3\varphi = -\tfrac{1}{4}\bigl\{\{\delta_1\varphi, \delta_2\varphi\}_\varphi, \delta_3\varphi\bigr\}_\varphi,$$

where $\{\,,\,\}_\varphi$ is the Poisson bracket on $C^\infty(M)$ of the symplectic form ω_φ.

Theorem 2.2 [Chen 2000b]. *$\mathcal{H}$ is a metric space, and is convex by $C^{1,1}$ geodesics.*

Calabi and Chen [2002] proved $\mathcal{H}$ is negatively curved in the sense of Alexandroff. We denote the space of normalized Kähler potentials by

$$\mathcal{H}_0 = \{\varphi \in C^\infty(M, R) \mid \omega + \sqrt{-1}\,\partial\bar{\partial}\varphi > 0 \text{ and } I(\varphi) = 0\}, \tag{2-2}$$

where

$$I(\varphi) = \frac{1}{V}\sum_{p=0}^{n} \frac{1}{(p+1)!\,(n-p)!}\int_M \varphi\omega^{n-p} \wedge (\partial\bar{\partial}\varphi)^p.$$

In fact, $\mathcal{H}$ can be naturally split as

$$\mathcal{H} = \mathcal{H}_0 \times \mathbb{R}.$$

This leads to a decomposition of the tangent space:

$$T_\varphi = \left\{ f \;\middle|\; \int_M f\omega_\varphi^n = 0 \right\} \oplus \mathbb{R}.$$

On a Kähler–Einstein manifold (M, ω), choose ω be the reference metric. It is clear from the definition (1-3) that $h_\omega = 0$. Substituting this into the potential equation (1-2) of the Kähler–Ricci flow, we obtain

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = \log \dfrac{\omega_\varphi^n}{\omega^n} + \varphi + a(t), \\ \varphi(0) = \varphi_0. \end{cases} \tag{2-3}$$

If we choose the normalization constant in (2-3) appropriately, namely,

$$a(t) = -\frac{1}{V}\int_M \left(\log \frac{\omega_\varphi^n}{\omega^n} + \varphi\right)\omega_\varphi^n, \tag{2-4}$$

we see that

$$\partial_t I(\varphi) = \frac{1}{V}\int_M \partial_t\varphi\omega_\varphi^n = 0. \tag{2-5}$$

We first assume that $\varphi_0 \in \mathcal{H}_0$ satisfies $I(\varphi_0) = 0$; the general case will be treated in Section 6. Then (2-5) implies $I(\varphi) = 0$, which ensures that the solution φ of (2-3) always stays in $\mathcal{H}_0$.

For any $\varphi \in \mathcal{H}$, Mabuchi [1986] defined the K-energy of (M, ω) as

$$\nu(\omega, \omega_\varphi) = -\frac{1}{V}\int_0^1 \int_M \dot{\varphi}(\tau)(S_{\varphi(\tau)} - \underline{S})\omega_{\varphi(\tau)}^n\, d\tau, \tag{2-6}$$

where $\varphi(\tau)$ is an arbitrary piecewise smooth path from 0 to φ. An explicit expression of the K-energy is formulated in [Chen 2000a; Tian 2000] as

$$\begin{aligned}(2\text{-}7)\quad \nu_\omega(\varphi) = \frac{1}{V}\int_M \log\frac{\omega_\varphi^n}{\omega^n}\omega_\varphi^n + \frac{\underline{S}n!}{V} I(\varphi)\\ - \frac{1}{V}\sum_{i=0}^{n-1}\frac{n!}{(i+1)!(n-i-1)!}\int_M \varphi\,\mathrm{Ric}\wedge\omega^{n-1-i}\wedge(\partial\bar\partial\varphi)^i.\end{aligned}$$

In later sections we will simply write $\nu(\varphi)$ instead of $\nu_\omega(\varphi)$.

Theorem 2.3 [Mabuchi 1987]. *If ω is a critical point of $\nu(\varphi)$, the second variation of the K-energy satisfies*

$$\frac{d^2}{dt^2}\nu(\theta_t)|_{t=0}\geq 0$$

for every smooth path $\{\theta_t \mid -\epsilon\leq t\leq\epsilon\}$ in $\mathcal{H}$ such $\theta_0=\omega$.

Let $\mathrm{Aut}(M)$ be the group of holomorphic automorphisms of M and $\mathrm{Aut}_0(M)$ its identity component.

Theorem 2.4 [Bando and Mabuchi 1987; Bando 1987]. *Assume $\mathcal{E}\neq\varnothing$.*

(i) *The K-energy is bounded from below on $\mathcal{H}$ and takes its absolute minimum exactly on $\mathcal{E}$.*

(ii) *$\mathcal{E}$ consists a single $\mathrm{Aut}_0(M)$-orbit.*

Indeed, the normalization constant $a(t)$ can be estimated by the K-energy.

Lemma 2.5. *Let φ be the solution of (2-3). The relation between $a(t)$ and the K-energy $\nu(\varphi)$ is given by*

$$(2\text{-}8)\qquad a(t)+\nu(\varphi)=a(0)+\nu(\varphi_0).$$

Proof. We calculate the evolution of $a(t)$ along the Kähler–Ricci flow directly:

$$V\frac{d}{dt}a(t) = -\int_M(\triangle_\varphi+1)\dot\varphi\omega_\varphi^n - \int_M\Big(\log\frac{\omega_\varphi^n}{\omega^n}+\varphi\Big)\triangle_\varphi\dot\varphi\omega_\varphi^n.$$

By Stokes' theorem and (2-5) the first term vanishes identically. Integration by parts and the use of (2-3) gives for the second term

$$\int_M(S_\varphi-n)\dot\varphi\omega_\varphi^n.$$

Since (2-6) implies

$$(2\text{-}9)\qquad \frac{d}{dt}\nu(t) = -\frac{1}{V}\int_M(S_\varphi-n)\dot\varphi\omega_\varphi^n,$$

we obtain $\frac{d}{dt}a(t)=-\frac{d}{dt}\nu(t)$. We conclude by integrating both sides with respect to t. □

Since the K-energy is decreasing along the Kähler–Ricci flow, we immediately conclude the following according to Theorem 2.4.

Corollary 2.6. *On a Kähler–Einstein manifold, $a(t)$ is uniformly bounded along the Kähler–Ricci flow.*

Set

$$X = C^0([0,T), C^{2+\alpha}(M,g)) \cap C^1([0,T), C^{\alpha}(M,g)).$$

The following theorems asserting short-time existence, regularity and continuous dependence on initial data for the Kähler–Ricci flow were proved by Chen and the author, who defined a new second-order Monge–Ampère flow, called the pseudo-Calabi flow and coinciding with the Kähler–Ricci flow when the initial datum is restricted in the canonical Kähler class:

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = -f(\varphi), \\ \triangle_{\varphi} f(\varphi) = S(\varphi) - \underline{S}. \end{cases} \tag{2-10}$$

Theorem 2.7 [Chen and Zheng 2010]. *Let $\varphi_0 \in C^{2,\alpha}(M,g)$ be such that*

$$\lambda\omega \le \omega_{\varphi_0} \le \Lambda\omega,$$

for two positive constants λ and Λ. Then the pseudo-Calabi flow has a unique solution $\varphi(x,t) \in X$, where T is the maximal existence time.

Theorem 2.8 [Chen and Zheng 2010]. *The solution of the pseudo-Calabi flow $\varphi \in X$ is smooth for any $t > 0$. More precisely, if $|\varphi(t)|_{C^{2,\alpha}} \le A$ for any $0 \le t \le T$, there exists a constant C (depending on A, g, t_0 and k) such that $|\varphi(t)|_{C^{k,\alpha}} \le C$ for any $T - t_0 \le t \le t_0 < T$.*

Theorem 2.9 [Chen and Zheng 2010]. *If ϕ is the solution of the pseudo-Calabi flow for an initial datum ϕ_0 on $[0,T]$, there is a neighborhood U of ϕ_0 such that the pseudo-Calabi flow has a solution $\varphi(t)$ on $[0,T]$ for any $\varphi_0 \in U$ and the mapping $\varphi_0 \mapsto \varphi(t)$ is C^k for $k = 0, 1, 2, \dots$.*

As a corollary of the continuous dependence on initial data we have:

Theorem 2.10 [Chen and Zheng 2010]. *Suppose M admits a constant scalar curvature Kähler metric ω. Let $\varphi_0 \in C^{2,\alpha}(M,g)$ be such that $\lambda\omega \le \omega_{\varphi_0} \le \Lambda\omega$ for positive constants λ and Λ. Then for any $T > 0$ there exits a positive constant $\epsilon_0(T)$ such that, if $|\varphi_0|_{C^{2,\alpha}(M,g)} \le \epsilon_0(T)$, the pseudo-Calabi flow has a unique solution on $[0,T]$ and*

$$|\dot\varphi|_{C^{\alpha}(M,g)} + |\varphi|_{C^{2,\alpha}(M,g)} \le C\epsilon_0(T) \quad \textit{for all } t \in [0,T],$$

where C depends on M, g and T. As T goes to infinity, $\epsilon_0(T)$ goes to zero.

3. No nontrivial holomorphic vector fields

Let $\eta(M)$ be the set of all holomorphic vector fields on M. We start with the case $\eta(M)=\varnothing$. We shall prove the following proposition in this section.

Proposition 3.1 [Tian and Zhu 2008; Zhu 2009]. *Assume M admits a Kähler–Einstein metric ω and has no holomorphic vector fields. There exits a small positive constant ϵ_0 such that, if the initial datum satisfies*

$$|\varphi_0|_{C^{2,\alpha}(M)} \le \epsilon_0,$$

then the Kähler–Ricci flow g_φ converges smoothly to g.

Proof. We at first show that under the assumption of the proposition, the solution of (2-3) always stays in some small ϵ_1-neighborhood of the zero function.

Lemma 3.2. *For any $\epsilon_1>0$, there exits a small positive constant ϵ_0 such that, if $|\varphi_0|_{C^{2,\alpha}(M)} \le \epsilon_0$, then $|\varphi(t)|_{2,\alpha} \le \epsilon_1$ for all $t\in[0,+\infty)$.*

Proof. Suppose that the conclusion fails; then there exists a sequence of initial data φ_s^0 such that

$$|\varphi_s^0|_{C^{2,\alpha}} \le \frac{1}{s}.$$

By virtue of Theorem 2.10, we get a sequence of solutions $\varphi_s(t)$ satisfying the flow equations (2-3) with $\varphi_s(0)=\varphi_s^0$. Let T_s be the first time such that

$$|\varphi_s(T_s)|_{C^{2,\alpha}} = \epsilon_1 \qquad \text{and} \qquad |\varphi_s(t)|_{C^{2,\alpha}} < \epsilon_1 \ \text{ for } 0\le t<T_s. \tag{3-1}$$

According to Theorem 2.10 again, we have $T_s \ge T_1 > 0$. Moreover, we apply Theorem 2.8 to (2-3) on $[T_s-2a, T_s]$ for fixed a such that $0<a<T_s/2-T_1/4$, then we obtain a uniform higher-order bound for the sequence of solutions:

$$|\varphi_s|_{C^{k,\alpha}(M)} \le C(k,\epsilon_1,a) \quad \text{ on } [T_s-a, T_s], \quad \text{for all } k\ge 0.$$

Consequently, there is a subsequence of $\phi_s=\varphi_s(T_s)$ converges smoothly to ϕ_∞ satisfying

$$|\phi_\infty|_{C^{2,\alpha}} = \epsilon_1. \tag{3-2}$$

It is obvious that g_{ϕ_∞} is still a Kähler metric. Since the K-energy is not only well defined for φ_s^0 by (2-7) but also decreasing along the Kähler–Ricci flow, Theorem 2.4 implies that

$$0 \le \nu_\omega(\phi_s) \le \nu_\omega(\varphi_s(0)) \le \frac{C}{s}.$$

By passing the limit we obtain

$$\lim_{s\to\infty} \nu_\omega(\varphi_s) = \nu_\omega(\varphi_\infty) = 0.$$

Using Theorem 2.4, we obtain that g_{ϕ_∞} is a Kähler–Einstein metric. From the same theorem we deduce that ϕ_∞ must be a constant. Furthermore the normalization condition $I(\phi_\infty) = 0$ leads to $\phi_\infty = 0$, which contradicts to (3-2). The lemma follows. $\square$

According to Theorem 2.8 and Lemma 3.2, we have $|\varphi(t)|_{C^k} \le C_k$ for any $k \ge 3$ away from $t = 0$. It follows that for any sequence t_i there is a subsequence such that $\phi(t_i)$ converges smoothly to a limit function φ_∞. Moreover, since the K-energy has a lower bound and it decays along the flow, ω_{φ_∞} must be a Kähler–Einstein metric. This, together with Theorem 2.4 and the normalization condition, implies that $\varphi_\infty = 0$. Because the t_i can be chosen arbitrarily, we conclude that the Kähler–Ricci flow converges smoothly to the original Kähler–Einstein metric. $\square$

4. *M* admits nontrivial holomorphic vector fields

4A. ***Choice and estimate of holomorphic transformations.*** When M admits holomorphic vector fields, we need to find an appropriate holomorphic transformation. Let $\mathscr{E}_0 \subset \mathscr{H}_0$ be the space of Kähler potentials of Kähler–Einstein metrics.

Let $\sigma_t^* \omega$ be any curve with $\sigma_0 = \mathrm{id}$ in $\mathscr{E}_0$. The tangent vector at ω is

$$\frac{d}{dt}\sigma_t^*\Big|_{t=0}\omega = L_X\omega.$$

Here $X = (\sigma_t)_*^{-1}\,\partial_t\sigma_t|_{t=0}$ is the real part of some holomorphic vector field. Since $C_1(M) > 0$ implies that M is simply connected by [Kobayashi 1961], we obtain $L_X\omega = \sqrt{-1}\,\partial\bar\partial\theta_X$ for some function θ_X. Hence, the finite-dimensionalness of the space of holomorphic vector fields implies that of $\mathscr{E}_0$. Moreover, according to [Mabuchi 1987], $\mathscr{E}_0$ is also a totally geodesic submanifold of $\mathscr{H}_0$. Then the point $\rho \in \mathscr{E}_0$ realizes the shortest distance between φ, and $\mathscr{E}_0$ is uniquely determined. In fact, according to Theorem 2.4, we obtain a holomorphic diffeomorphism $\sigma \in \mathrm{Aut}_0(M)$ such that $\sigma^*\omega = \omega + \sqrt{-1}\,\partial\bar\partial\rho$. The K-energy is invariant under holomorphic transformations:

Lemma 4.1 [Mabuchi 1986]. $\nu(\omega, \omega_{(\sigma^{-1})^*(\varphi-\rho)}) = \nu(\omega, \omega_\varphi) = \nu(\omega_\rho, \omega_\varphi)$.

Proof. Since ω and ω_ρ are both Kähler–Einstein metrics, Lemma (5.4.1) and Theorem (5.3) of [Mabuchi 1986] yield, respectively, the equalities

$$\begin{aligned}\nu(\omega, \omega_{(\sigma^{-1})^*(\varphi-\rho)}) &= \nu(\sigma^*\omega, \omega_\varphi) = \nu(\omega_\rho, \omega_\varphi)\\ &= \nu(\omega_\rho, \omega) + \nu(\omega, \omega_\varphi) = \nu(\omega, \omega_\varphi).\end{aligned}$$ $\square$

We next state two lemmas from [Chen and Zheng 2010] regarding the metric
rephrased for economy
geometry of the space of constant scalar curvature Kähler metrics. They show that when metrics stay close to ω, their projection metrics are uniformly bounded.

Lemma 4.2 [Chen and Zheng 2010]. *There exists a positive constant ϵ such that $|\rho|_{C^{3,\alpha}} \le C_2\epsilon$ for any ρ satisfying $d(0,\rho)\le\epsilon$.*

Proof. Since $\mathscr{E}_0$ is a finite-dimensional Riemannian symmetric space, a small ϵ neighborhood near $\rho=0$ in this submanifold can be pulled back by the exponential map $\exp_0$ to the tangent space $T_0(\mathscr{E}_0)$ at 0. Set $\psi=\exp_0^{-1}(\rho)$. Then the length from ψ to 0 is ϵ. The norm induced by the distance on $T_0(\mathscr{E}_0)$ is equivalent to the $C^{2,\alpha}$ norm, since all norms on a finite-dimensional vector space are equivalent. Thus $|\exp_0^{-1}(\rho)|_{C^{2,\alpha}}$ is bounded by $C_1\epsilon$. Since the exponential map is a diffeomorphism in the ϵ neighborhood near $\rho=0$, we obtain $|\rho|_{C^{2,\alpha}}\le C_2\epsilon$ for some constant C_2. The lemma follows by an appropriate choice of ϵ. □

We can improve this conclusion for C^k for fixed $k\ge 0$, not only for $C^{3,\alpha}$ norm.

Lemma 4.3 [Chen and Zheng 2010]. *There exists a positive constant ϵ_1 such that $|\varphi|_{C^{2,\alpha}}\le\epsilon_1$ implies*

$$|\rho|_{C^{3,\alpha}}\le C_4 \quad \textit{and} \quad |\sigma|_h\le C_5,$$

where h is the left invariant metric in $\mathrm{Aut}(M)$.

Proof. Choose a path $\gamma_t=t\varphi-I(t\varphi)\in\mathscr{H}_0$ for $0\le t\le 1$. Denote $d(0,\varphi)$ the distance between 0 and φ. Using (2-1), we compute

$$\begin{aligned} d(0,\varphi)\le L(\gamma_t) &= \int_0^1\left(\int_M\left(\frac{\partial\gamma_t}{\partial t}\right)^2\omega_{\gamma_t}^n\right)^{1/2}dt\\ &= \int_0^1\left(\int_M(\varphi-\partial_t I(t\varphi))^2\omega_{\gamma_t}^n\right)^{1/2}dt\le C_3\epsilon_1\end{aligned}$$

for $|\varphi|_{C^{2,\alpha}}\le\epsilon_1$. Moreover, the choice of the ρ implies

$$d(0,\rho)\le d(0,\varphi)+d(\varphi,\rho)\le 2d(0,\varphi)\le C_3\epsilon_1,$$

by the triangle inequality. From Lemma 4.2, it follows that $|\rho|_{C^{3,\alpha}}\le C_4=C_2C_3\epsilon_1$. Using [Chen and Tian 2006, Lemma 4.6], we derive $|\sigma|_h\le C_5$ and the lemma follows. □

Remark. Alternatively the holomorphic transformation can be derived by minimizing the $I-J$ functional, as in [Bando and Mabuchi 1987], which will be further discussed in Section 7. Those authors use this minimizer to prove the uniqueness of the Kähler–Einstein metric when the first Chern class is positive. The minimizer of the $I-J$ functional is not unique in general, since the second variation of this functional is not strictly positive. However, when the potential is small enough, the minimizer is unique. We mention also that Corollary 7.2 provides an estimate similar to Lemma 4.3.

4B. ***Long time existence and Cheeger–Gromov convergence.*** Set

$$\mathcal{S}(\epsilon_1, C(k,\epsilon_1)) = \left\{\varphi \;\middle|\; \varphi|_{C^{2,\alpha}} \le \epsilon_1;\, |\varphi|_{C^{k,\alpha}(M)} \le C(k,\epsilon_1)\right\}.$$

It is obvious that $0 \in \mathcal{S}$. We will show that when the initial potential is small, the solution of (2-3) always stays in $\mathcal{S}$ after pulling back by a sequence of holomorphic transformations.

Lemma 4.4. *For any $\epsilon > 0$, there is a small positive constant o depends on ϵ and $\mathcal{S}$ such that, for any $\varphi \in \mathcal{S}$, if $\nu_\omega(\varphi) \le o$, then $|(\sigma^{-1})^*(\varphi - \rho)|_{C^{2,\alpha}} < \epsilon$.*

Proof. If the conclusion fails, we take a positive constant ϵ and a sequence of $\varphi_s \in \mathcal{S}$ satisfying

$$\nu_\omega(\varphi_s) \le \frac{1}{s}$$

and such that

$$|(\sigma_s^{-1})^*(\varphi_s - \rho_s)|_{C^{2,\alpha}} \ge \epsilon. \tag{4-1}$$

Since $\varphi_s \in \mathcal{S}(\epsilon_1, C(k,\epsilon_1))$, we obtain a subsequence φ_{s_j} of φ_s converging smoothly to φ_∞. Let $\hat{\varphi}_s = (\sigma_s^{-1})^*(\varphi_s - \rho_s)$. Lemma 4.3 gives

$$|\rho_s|_{C^{3,\alpha}} \le C_4 \quad \text{and} \quad |\sigma_s|_h \le C_5,$$

which implies that there are, by the Arzelà–Ascoli theorem and the Bolzano–Weierstrass theorem respectively, subsequences of ρ_{s_j} and σ_{s_j} for which (using the same notation)

$$\rho_{s_j} \to \rho_\infty \text{ in } C^{3,\beta} \quad \text{for any } \beta < \alpha$$

and

$$\sigma_{s_j} \to \sigma_\infty \text{ in the left invariant metric.}$$

Combining with Lemma 4.1, which implies that

$$\nu_\omega(\varphi_\infty) = \nu_\omega(\hat{\varphi}_\infty) = 0,$$

we derive that $\hat{\varphi}_{s_j}$ converges to $\hat{\varphi}_\infty = (\sigma_\infty^{-1})^*(\varphi_\infty - \rho_\infty) \in \mathcal{E}_0$ in $C^{3,\beta}$ and $\sigma_\infty^*\omega = \omega + \partial\bar{\partial}\rho_\infty$. Moreover, according to Theorem 2.4, we have $\hat{\varphi}_\infty, \varphi_\infty \in \mathcal{E}_0$.

We claim that

$$d(\varphi_\infty, \rho_\infty) = 0.$$

Otherwise, for some sufficient large N, when $s_j > N$,

$$d(\varphi_{s_j}, \rho_{s_j}) = d(\varphi_{s_j}, \mathcal{E}_0)$$

has a strictly positive lower bound. Since the distance function is at least C^1 (see [Chen 2000b]), we have $d(\varphi_\infty, \mathcal{E}_0) > 0$, contradicting $\varphi_\infty \in \mathcal{E}_0$. Consequently, the claim holds. It follows that $\hat{\varphi}_\infty = 0$, in contradiction with the lower bound $|\hat{\varphi}_\infty|_{C^{2,\alpha}} \ge \epsilon$ of (4-1). ◻

Proposition 4.5. *Assume M admits a Kähler–Einstein metric ω and has nontrivial holomorphic vector fields. There is a small positive constant ϵ_0 such that, if $|\varphi_0|_{C^{2,\alpha}(M)} \leq \epsilon_0$, there is a unique solution $\varphi(t)$ and a corresponding holomorphic transformation $\varrho(t)$ such that the normalization potential of $\varrho(t)^*\omega(t)$ always stays in $\mathcal{S}$. Moreover, any sequence t_j has is a subsequence (still denoted by t_j) such that $\varrho(t_j)^*\omega(t_j)$ converges smoothly to a Kähler–Einstein metric ω_∞.*

Proof. We prove this proposition by contradiction. Let ϵ_1 be as in Lemma 4.3. Using Theorem 2.10, we assume there is a maximal time T such that

$$|\varphi|_{C^{2,\alpha}} < \epsilon_1 \text{ on } [0, T) \quad \text{and} \quad |\varphi(T)|_{C^{2,\alpha}} = \epsilon_1.$$

According to Theorem 2.8 we obtain $|\varphi(T)|_{C^{k,\alpha}} \leq C(k, \epsilon_1, t_0, g)$ on $[T - t_0, T]$ for a fixed $T/2 \leq t_0 \leq T$. Let the constant $C(k, \epsilon_1)$ be $C(k, \epsilon_1, t_0, g)$. So we get

$$\varphi(T) \in \mathcal{S}(\epsilon_1, C(k, \epsilon_1)).$$

There are two situations. If $\varphi(T)$ is a Kähler–Einstein metric, the flow will stop here and our theorem is proved. Otherwise, we will extend the flow as follows.

We first choose ϵ_0 small enough to guarantee

$$\nu_\omega(\varphi_0) \leq o\Big(\frac{\epsilon_1}{2}, \mathcal{S}(\epsilon_1, C(k, \epsilon_1))\Big),$$

where the constant $o(\epsilon_1/2, \mathcal{S}(\epsilon_1, C(k, \epsilon_1)))$ is determined in Lemma 4.4. Let the holomorphic transformation σ be the projection of $\varphi(T)$ in $\mathcal{E}_0$ with

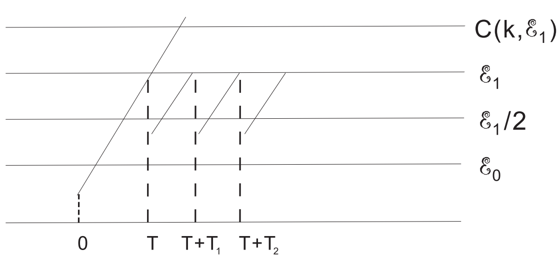
$$\sigma^*\omega = \omega + \sqrt{-1}\,\partial\bar{\partial}\rho.$$

We set φ_1^0 be the Kähler potential of the metric pulled back by σ, that is,

$$(\sigma^{-1})^*\omega_{\varphi(T)} = \omega + \sqrt{-1}\,\partial\bar{\partial}[(\sigma^{-1})^*(\varphi(T) - \rho)] = \omega + \sqrt{-1}\,\partial\bar{\partial}\varphi_1^0.$$

Since the K-energy decreases along the Kähler–Ricci flow, Lemma 4.1 yields

$$\nu_\omega(\varphi_1^0) \leq o\Big(\frac{\epsilon_1}{2}, \mathcal{S}(\epsilon_1, C(k, \epsilon_1))\Big). \tag{4-2}$$



Idea of the proof of Proposition 4.5: solving the equation after pulling back.

Letting $\psi = (\sigma^{-1})^*(\varphi(T) - \rho)$, Lemma 4.4 implies that

$$|\psi|_{C^{2,\alpha}(g)} = \left|(\sigma^{-1})^*(\varphi(T)-\rho)\right|_{C^{2,\alpha}(g)} < \frac{\epsilon_1}{2}. \tag{4-3}$$

We next show that the Kähler–Ricci flow is invariant under the transformation. Let $\varphi_1 = (\sigma^{-1})^*(\varphi(t) - \rho)$. We compute

$$\begin{aligned}
\frac{\partial}{\partial t}\varphi_1 &= (\sigma^{-1})^*\left[\log\frac{\omega_\varphi^n}{\omega^n} + \varphi - \frac{1}{V}\int_M\Big(\log\frac{\omega_\varphi^n}{\omega^n} + \varphi\Big)\omega_\varphi^n\right]\\
&= (\sigma^{-1})^*\left[\log\frac{\omega_\varphi^n}{\omega_\rho^n} + \varphi - \rho - \frac{1}{V}\int_M\Big(\log\frac{\omega_\varphi^n}{\omega_\rho^n} + \varphi - \rho\Big)\omega_\varphi^n\right]\\
&= \left[\log\frac{\omega_{\varphi_1}^n}{\omega^n} + \varphi_1 - \frac{1}{V}\int_M\Big(\log\frac{\omega_{\varphi_1}^n}{\omega^n} + \varphi_1\Big)\omega_{\varphi_1}^n\right].
\end{aligned}$$

The second equality follows form the fact that ω_ρ is a Kähler–Einstein metric. We conclude that φ_1 is the solution of an equation of the form

$$\begin{cases} \dfrac{\partial}{\partial t}\varphi_1 &= \log\dfrac{\omega_{\varphi_1}^n}{\omega^n} + \varphi_1 + a(t),\\ \varphi_1(0) &= \varphi_1^0 = (\sigma^{-1})^*(\varphi(T)-\rho), \end{cases} \tag{4-4}$$

where (4-3) and (4-2) hold. Again, Theorem 2.10 implies (4-4) has a solution on $[0, T_1]$ with $T_1 \geq T$ such that

$$|\varphi_1(T_1)|_{C^{2,\alpha}} = \epsilon_1.$$

According to Theorem 2.8, we also obtain

$$|\varphi(T_1)|_{C^{k,\alpha}} \leq C(k, \epsilon_1, t_0, g) \quad \text{ on } [T_1 - t_0, T_1].$$

So we still have $\varphi(T_1) \in \mathcal{S}(\epsilon_1, C(k, \epsilon_1))$. Moreover, if we let

$$\varphi(t) = \sigma^*\varphi_1(t - T) + \rho \quad \text{ on } [T, T+T_1),$$

the new $\varphi(t)$ is the solution of (2-3) on $[0, T+T_1]$.

We repeat the same steps inductively for

$$\varphi_{s-1}(T_{s-1}) \in \mathcal{S}(\epsilon_1, C(k, \epsilon_1)),$$

with $T_{s-1} \geq T$ obtained in Theorem 2.10, until φ_s becomes a Kähler–Einstein at time T_s, with $T_s < \infty$. If this does not happen, the Kähler–Ricci flow has long-time existence and the solution $\varphi(t)$ for all $t \geq 0$ is given by

$$\omega_{\varphi(t)} = \prod_{i=0}^{s-1}\sigma_i^*\omega_{\varphi_s(t)} \quad \text{ on } \left[\sum_{i=0}^{s-1}T_i, \sum_{i=0}^{s}T_i\right).$$

Finally, we prove the convergence of the Kähler–Ricci flow. For any sequence $\{\varphi_{t_j}\}$, there is s such that $\sum_{i=0}^{s-1} T_i \le t_j \le \sum_{i=0}^{s} T_i$. Let

$$\varrho_j = \left(\prod_{i=0}^{s-1} \sigma_i\right)^{-1}.$$

We have

$$|\varrho_j^* \omega_{\varphi_{t_j}} - \omega|_{C^\alpha} \le \epsilon_1 \quad \text{and} \quad |\varrho_j^* \omega_{\varphi_{t_j}} - \omega|_{C^k} \le C(k, \epsilon_1).$$

Therefore all metrics are equivalent and their derivatives are bounded. We set

$$\omega_{\psi_{t_j}} = \varrho_j^* \omega_{\varphi_{t_j}}.$$

It follows that there is a subsequence of $\omega_{\psi_{t_j}}$ that converges to a limit metric ω_∞ (which depends on the choice of the subsequence). Since the K-energy is bounded below, we have $\lim_{s\to\infty} \nu(\omega, \omega_{\psi_{t_j}}) = 0$. It follows from Theorem 2.4 that g_∞ is a Kähler–Einstein metric. The proposition is proved. □

Let $t_s = \sum_{i=0}^{s} T_i$. Following the argument in [Chen and Tian 2006], we can first connect each pair of points φ_{t_s} and $\varphi_{t_{s+1}}$ by a geodesic in the space of Kähler–Einstein metrics, so

$$\varrho(t) = \varrho(s) \exp((t-s) X_s) \quad \text{for all } t \in [s, s+1],$$

with X_s uniformly bounded by Lemma 4.3. We then smooth the corner at each point t_s by replacing the broken line by a smooth curve in a small neighborhood of t_s without changing the value and the t derivative at the endpoints. Hence we have extended the holomorphic transformation to all t, while ensuring Lipschitz continuity in t.

Let $\omega_{\psi(t)} = \varrho(t)^* \omega_{\varphi(t)}$. We have already seen that the Kähler–Ricci flow converges to a Kähler–Einstein metric in Cheeger–Gromov sense; i.e., for any sequence $g(t_i)$, there is a subsequence $g(t_{i_j})$ and a holomorphic transformation $\varrho(t_{i_j})$ such that $\varrho(t_{i_j})^* g(t_{i_j})$ converges smoothly to a Kähler–Einstein metric g_∞. So we have

$$\lim_{t\to\infty} \mathrm{Ric}(g_{\psi_t}) - \omega_{\psi_t} = 0,$$

which leads to the convergence of the eigenvalue. To obtain the compactness of the sequence of holomorphic transformations $\varrho(t)$ and the exponential convergence of the Kähler–Ricci flow, we use an auxiliary result:

Theorem 4.6 [Phong et al. 2009, Theorem 2 and Remark (7)]. *If the Kähler–Ricci flow converges to a Kähler–Einstein metric in Cheeger–Gromov sense. Then the Kähler–Ricci flow must converge exponentially to a unique Kähler–Einstein metric nearby.*

5. Kähler–Ricci solitons

In this section we generalize our argument to the Kähler–Ricci solitons. According to [Fujiki 1978], the identity part of holomorphic transformation group $\mathrm{Aut}_0(M)$ is meromorphically isomorphic to a linear algebraic group $L(M)$ and the quotient $\mathrm{Aut}_0(M)/L(M)$ is a complex torus. Futaki and Mabuchi [1995] used the Chevalley decomposition to $L(M)$ to obtain a semidirect decomposition

$$\mathrm{Aut}_0(M) = \mathrm{Aut}_r(M) \ltimes R_u.$$

Here $\mathrm{Aut}_r(M)$ is the reductive algebra group, which is the complexification of a maximal compact subgroup K, and R_u is the unipotent radical of $\mathrm{Aut}_0(M)$. Let η_r be the Lie algebra of $\mathrm{Aut}_r(M)$. Recall that a Kähler metric ω is called a *Kähler–Ricci soliton* if there is a holomorphic vector field X such that

$$L_X\omega = \mathrm{Ric} - \omega. \tag{5-1}$$

Tian and Zhu [2000] proved the uniqueness of Kähler–Ricci solitons for a fixed X in the Lie algebra of $\mathrm{Aut}_0(M)$.

Theorem 5.1 [Tian and Zhu 2000]. *If (ω, X) and (ω', X) are Kähler–Ricci solitons, there are holomorphic transformations $\sigma \in \mathrm{Aut}_0(M)$ and $\tau \in \mathrm{Aut}_r(M)$ such that $\sigma^*\omega = \tau^*\sigma^*\omega'$ and $\sigma^*X \in \eta_r$.*

Theorem 5.2 [Tian and Zhu 2002]. *If (ω, X) and (ω', X') are two Kähler–Ricci solitons, then there is a holomorphic transformation group $\sigma \in \mathrm{Aut}_0(M)$ such that $\omega = \sigma^*\omega'$ and $X = \sigma_*^{-1}X'$.*

Since $L_{\mathrm{Im}\,X}\omega = 0$, $\mathrm{Im}\,X$ generates a one-parameter isometric group K_X. We further choose K such that $K_X \subseteq K$. According to Proposition 2.1 of [Tian and Zhu 2002], X lies in the center of η_r.

Now we fix a holomorphic vector field X. By the Hodge theory there is a real value function θ_X such that $L_X\omega = \sqrt{-1}\,\partial\bar\partial\theta_X$ with $\int_M e^{\theta_X}\omega^n = V$. Then the potential equation of the Kähler–Ricci flow (1-2) is

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = \log\dfrac{\omega_\varphi^n}{\omega^n} + \varphi - \theta_X + a(t), \\ \varphi(0) = \varphi_0. \end{cases} \tag{5-2}$$

We choose

$$a(t) = -\frac{1}{V}\int_M \Big(\log\frac{\omega_\varphi^n}{\omega^n} + \varphi - \theta_X\Big)\omega_\varphi^n;$$

moreover $I(\varphi_0) = 0$, so the Kähler–Ricci flow stays in $\mathcal{H}_0$.

Recall the W-functional of [Perelman 2002], defined by

$$\mathcal{W}(g, f, \tau) = (4\pi\tau)^{-n/2} \int_M \big[\tau(|\nabla f|^2 + S) + f - n\big] e^{-f}\, dV,$$

and invariant under diffeomorphisms σ and scaling C:

$$\mathcal{W}(C\sigma^* g, \sigma^* f, C\tau) = \mathcal{W}(g, f, \tau). \tag{5-3}$$

Recall also Perelman's μ functional, defined by

$$\mu(g, \tau) = \inf\big\{\mathcal{W}(g, f, \tau) \,\big|\, (4\pi\tau)^{-n/2} \textstyle\int_M e^{-f}\, dV = 1\big\},$$

and also invariant under diffeomorphism. Its minimum is achieved by some smooth function f satisfying $\tau[(2\triangle f - |\nabla f|^2) + S] + f - n = \mu(g, \tau)$. The first variation of $\mu(g, \tau)$ at $g'_{ij} = v_{ij}$ for fixed τ is

$$\mu'(v_{ij}, \tau) = (4\pi\tau)^{-n/2} \int_M \Big\{-\tau\Big(v_{ij}, \mathrm{Ric} + D^2 f - \frac{1}{2\tau} g\Big)\Big\} e^{-f}\, dV_g.$$

So the (shrinking) Kähler–Ricci soliton is the critical point of $\mu(g, \tau = \frac{1}{2})$. The gradient flow of the μ functional equals to (1-1) with $\lambda = 1$ up to a diffeomorphism generated by ∇f. So the μ functional is nondecreasing along the Ricci flow. The second variation of this functional near a Kähler–Ricci soliton in the canonical class has been calculated:

Theorem 5.3 [Tian and Zhu 2008, Proposition 2.1]. *We have*

$$\frac{\partial^2}{dt^2} \mu\big(\omega + \sqrt{-1}\, \partial\bar{\partial}\varphi\big)|_{t=0} \le 0, \tag{5-4}$$

and equality holds if and only if $\dot{\varphi}(0)$ is the real part of the holomorphic potential of some holomorphic vector field.

So the only directions in which the Kähler–Ricci soliton ω in (5-4) vanishes are those tangent to the orbit of ω under the action of $\mathrm{Aut}_0(M)$. We thus obtain the following local property of the μ functional:

Lemma 5.4. *A Kähler–Ricci soliton is a local maximum of $\mu(g)$ in $\mathcal{H}_0$.*

Proof. Near a Kähler–Ricci soliton g, the tangent space $T_\omega(\mathcal{H}_0)$ splits as $\eta(M) \oplus N$, where N is the orthonormal part. Due to Theorem 5.3, $\mu(g') < \mu(g)$ along any direction in N. Moreover, since $\sigma^* g$ is still a Kähler–Ricci soliton for any $\sigma \in \mathrm{Aut}_0(m)$, we have $\mu(g') \equiv \mu(g)$ along any direction in $\eta(M)$. □

As a result we deduce that a Kähler metric that achieves the maximum value of the $\mu(g)$ functional near a Kähler–Ricci soliton must be a Kähler–Ricci soliton.

Let $\mathcal{E}_0 \subset \mathcal{H}_0$ be the space of potentials of Kähler–Einstein solitons with respect to the holomorphic vector field X. Due to Theorem 5.1, $\mathcal{E}_0$ is a single orbit under the action of $\mathrm{Aut}_r(M)$.

Lemma 5.5. *$\mathscr{E}_0$ is a finite-dimensional totally geodesic submanifold of $\mathscr{H}_0$.*

Proof. Analogously to the case of the extremal metric in [Calabi 1985], Lemma A.2 and Theorem A of [Tian and Zhu 2000] imply that the identity component of the holomorphic isometric group of the Kähler–Ricci soliton (ω, X) is a maximal compact subgroup of $\mathrm{Aut}_r(M)$ containing K_X. So $(\mathrm{Aut}_r(M), K)$ is a Riemannian symmetric pair and $\mathscr{E}_0$ is $\mathrm{Aut}_r(M)$-equivariantly diffeomorphic to the Riemannian symmetric space $\mathrm{Aut}_r(M)/K$. Then for any $\omega \in \mathscr{E}_0$, each geodesic starting at ω in $\mathscr{E}_0$ can be written in the form

$$\gamma(t) = \exp(t \operatorname{Re} Y)^* \omega,$$

for some nonzero Y whose imagine part is a Killing vector field. Then Theorem 3.5 and Remark 3.3 in [Mabuchi 1987] show that $\gamma(t)$ is also a geodesic in $\mathscr{H}_0$. □

Now choose $\omega_\rho = \omega + \partial\bar{\partial}\rho$ such that ρ realizes the shortest distance between ψ and $\mathscr{E}_0$. Clearly, ρ is uniquely determined. In fact, due to Theorem 5.1 we obtain a holomorphic diffeomorphism $\sigma \in \mathrm{Aut}_r(M)$ such that

$$\sigma^*\omega = \omega_\rho = \omega + \sqrt{-1}\,\partial\bar{\partial}\rho,$$

with $\rho \in \mathscr{E}_0$. By an argument analogous to the one in Proposition 4.5, but using the μ functional instead of the K-energy, we obtain:

Lemma 5.6. *For any $\epsilon > 0$, There is a small positive constant o depends on ϵ and $\mathscr{S}$ such that for any $\varphi \in \mathscr{S}$, if $\mu(\omega_\varphi) \geq \mu(\omega) - o$, then $|(\sigma^{-1})^*(\varphi - \rho)|_{C^{2,\alpha}} < \epsilon$.*

Proof. If the conclusion fails, take a positive ϵ and a sequence of $\varphi_s \in \mathscr{S}$ satisfying

$$\mu(\omega_{\varphi_s}) \geq \mu(\omega) - \frac{1}{s} \quad \text{and} \quad \left|(\sigma_s^{-1})^*(\varphi_s - \rho_s)\right|_{C^{2,\alpha}} \geq \epsilon.$$

Since $\varphi_s \in \mathscr{S}$, we obtain a subsequence φ_{s_j} of φ_s converging smoothly to φ_∞. Lemma 4.3 gives

$$|\rho_s|_{C^{3,\alpha}} \leq C_4 \quad \text{and} \quad |\sigma_s|_h \leq C_5,$$

which implies that $(\sigma_{s_j}^{-1})^*(\varphi_{s_j} - \rho_{s_j})$ converges in $C^{3,\beta}$ towards

$$\hat{\varphi}_\infty = (\sigma_\infty^{-1})^*(\varphi_\infty - \rho_\infty) \in \mathscr{E}_0, \quad \text{with } \sigma_\infty^*\omega = \omega + \partial\bar{\partial}\rho_\infty.$$

Then (5-3) implies that $\mu(\omega_{\varphi_\infty}) = \mu(\omega_{\hat{\varphi}_\infty}) = \mu(\omega)$. The rest of this proof is the same as for Lemma 4.4. □

Proposition 5.7. *Assume M admits a Kähler–Ricci soliton (ω, X). There exits a small constant ϵ_0 such that, if $|\varphi_0|_{C^{2,\alpha}(M)} \leq \epsilon_0$, there is a unique solution $\varphi(t)$ and a corresponding holomorphic transformation $\varrho(t) \in \mathrm{Aut}_r(M)$ such that the normalization potential of $\varrho(t)^*\omega_\varphi(t)$ always stays in $\mathscr{S}$. Moreover, for any sequence t_i, there is a subsequence t_{i_j} such that $\varrho(t_{i_j})^* g_{\varphi(t_{i_j})}$ converges smoothly to g_∞.*

Proof. The proof, by contradiction, is similar to that of Proposition 4.5. Let ϵ_1 be as in Lemma 4.3. Applying Theorem 2.10 to the potential equation (5-2), we assume there is a maximal time T such that

$$|\varphi|_{C^{2,\alpha}} < \epsilon_1 \text{ on } [0, T) \quad \text{and} \quad |\varphi(T)|_{C^{2,\alpha}} = \epsilon_1.$$

From Theorem 2.8 we obtain $|\varphi(T)|_{C^{k,\alpha}} \leq C(k, \epsilon_1, t_0, g)$ on $[T - t_0, T]$ for a fixed $T/2 \leq t_0 \leq T$. Let the constant $C(k, \epsilon_1)$ be $C(k, \epsilon_1, t_0, g)$. So we get

$$\varphi(T) \in \mathcal{S}(\epsilon_1, C(k, \epsilon_1)).$$

There are two situations. If $\varphi(T)$ is a Kähler–Ricci soliton, the flow will stop here and our theorem is proved. Otherwise, we will extend the flow as in the proof in Proposition 4.5.

We first choose ϵ_0 small enough to guarantee that

$$\mu(\omega_{\varphi_0}) \geq \mu(\omega) - o\Big(\frac{\epsilon_1}{2}, \mathcal{S}(\epsilon_1, C(k, \epsilon_1))\Big),$$

where the constant $o(\epsilon_1/2, \mathcal{S}(\epsilon_1, C(k, \epsilon_1)))$ is determined in Lemma 4.4. Let the holomorphic transformation σ be the projection of $\varphi(T)$ in $\mathscr{E}_0$ with

$$\sigma^*\omega = \omega + \sqrt{-1}\,\partial\bar{\partial}\rho.$$

Let φ_1^0 be the Kähler potential of the metric pulled back by σ, that is,

$$(\sigma^{-1})^*\omega_{\varphi(T)} = \omega + \sqrt{-1}\,\partial\bar{\partial}\big[(\sigma^{-1})^*(\varphi(T) - \rho)\big] = \omega + \sqrt{-1}\,\partial\bar{\partial}\varphi_1^0.$$

Since the μ functional is nondecreasing along the Kähler–Ricci flow, we obtain

$$\mu(\omega_{\varphi_1^0}) \geq \mu(\omega) - o\Big(\frac{\epsilon_1}{2}, \mathcal{S}(\epsilon_1, C(k, \epsilon_1))\Big). \tag{5-5}$$

Lemma 4.4 implies that

$$\big|(\sigma^{-1})^*(\varphi(T) - \rho)\big|_{C^{2,\alpha}(g)} < \frac{\epsilon_1}{2}. \tag{5-6}$$

Set $\varphi_1(t) = (\sigma^{-1})^*(\varphi(t) - \rho)$. Combining (5-2) and (5-1), we obtain that φ_1 is the solution of an equation of the form

$$\begin{cases} \dfrac{\partial}{\partial t}\varphi_1 = \log \dfrac{\omega_{\varphi_1}^n}{\omega^n} + \varphi_1 - \theta_X + a(t), \\ \varphi_1(0) = \varphi_1^0 = (\sigma^{-1})^*(\varphi(T) - \rho), \end{cases} \tag{5-7}$$

where (5-6) and (5-5) hold.

Again, Theorem 2.10 implies that (5-7) has a solution on $[0, T_1]$, with $T_1 \geq T$, such that

$$|\varphi_1(T_1)|_{C^{2,\alpha}} = \epsilon_1.$$

From Theorem 2.8, we also obtain $|\varphi(T_1)|_{C^{k,\alpha}} \le C(k,\epsilon_1,t_0,g)$ on $[T_1-t_0, T_1]$. So we still have

$$\varphi(T_1) \in \mathcal{S}(\epsilon_1, C(k,\epsilon_1)).$$

If we set $\varphi(t) = \sigma^*\varphi_1(t-T)+\rho$ on $[T, T+T_1)$, the new $\varphi(t)$ is the solution of (5-2) on $[0, T+T_1]$.

We repeat the same steps inductively for

$$\varphi_{s-1}(T_{s-1}) \in \mathcal{S}(\epsilon_1, C(k,\epsilon_1)),$$

with $T_{s-1} \ge T$ obtained in Theorem 2.10. We thus obtain a sequence of holomorphic transformations σ_i and the solution $\varphi(t)$ for all $t \ge 0$ given by

$$\omega_{\varphi(t)} = \prod_{i=0}^{s-1} \sigma_i^* \omega_{\varphi_s(t)} \quad \text{on } \Bigl[\sum_{i=0}^{s-1} T_i, \sum_{i=0}^{s} T_i\Bigr).$$

Set $\varrho_j = \Bigl(\prod_{i=0}^{s-1}\sigma_i\Bigr)^{-1}$. We have

$$|\varrho_j^*\omega_{\varphi_{t_j}} - \omega|_{C^\alpha} \le \epsilon_1 \quad \text{and} \quad |\varrho_j^*\omega_{\varphi_{t_j}} - \omega|_{C^k} \le C(k,\epsilon_1).$$

It follows that there is a subsequence of $\varrho_j^*\omega_{\varphi_{t_j}}$ converging to a limit metric ω_∞. According to Lemma 5.4, the μ functional is bounded above and ω_∞ is a Kähler–Ricci soliton. □

Assume ς is generated by $\operatorname{Re} X$:

$$\operatorname{Re} X = (\varsigma^{-1})_* \frac{\partial}{\partial t}\varsigma.$$

Let ϱ and ϕ satisfy $\varsigma^*\omega = \omega_\varrho$ and $\phi = \varsigma^*\varphi + \varrho$. We obtain the modified Kähler–Ricci flow of the form

$$\begin{cases} \dfrac{\partial}{\partial t}\omega_\phi = -\operatorname{Ric}(\omega_\phi) + \omega_\phi + L_{\operatorname{Re} X}\omega_\phi, \\ \omega_{\varphi(0)} = \omega_{\varphi_0}. \end{cases}$$

We apply [Phong et al. 2011, Theorem 1] to obtain:

Theorem 5.8. *If the Kähler–Ricci flow converges to a Kähler–Ricci soliton in the Cheeger–Gromov sense and the initial Kähler potential is K_X-invariant, then the modified Kähler–Ricci flow converges exponentially to a unique Kähler–Ricci soliton nearby.*

Zhu [2009] also discussed the stability of Kähler–Ricci flow near a Kähler–Ricci soliton by using Perelman's estimate (unpublished) and Chen and Tian's energy method [2002; 2006].

6. Weak flow

In this section we weaken the initial condition. Let $a(t)=0$ in (2-3); the potential equation then reads

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = \log \dfrac{\omega_\varphi^n}{\omega^n} + \varphi, \\ \varphi(0) = \varphi_0. \end{cases} \tag{6-1}$$

We defined φ_0 is the limit of $\varphi_s \in \mathrm{PSH}(M, \omega) \cap L^\infty(M)$ in L^∞ norm. Meanwhile, $\omega_{\varphi_0} \geq 0$ in the sense of currents. Let the weak solution be a limit of a sequence of approximation solutions by

$$\varphi(t) = \lim_{s\to 0} \varphi(s, t).$$

The Kähler–Ricci flow with weak initial data was studied in [Chen and Ding 2007; Chen and Tian 2008; 2008]. We also have:

Theorem 6.1 [Song and Tian 2009, Proposition 3.2]. *If φ_0 is defined above with $|\varphi_0|_{L^\infty} \leq A$ and $|\omega^n_{\varphi_0}/\omega^n|_{L^p(M,\omega)} \leq B$ for $p > 1$, there is a unique smooth solution $g_\varphi(t)$ of (1-1) for $t > 0$ such that*

$$\lim_{t\to 0^+} \varphi(t) = \varphi_0.$$

The estimate in Song and Tian's proof is that

$$|\varphi(t)|_{C^k} \leq C(t, T, k, A, B) \quad \text{on } (0, T]. \tag{6-2}$$

For fixed B and p, introduce the space

$$\mathcal{N}(\epsilon_0; B, p) = \left\{ \varphi \;\middle|\; \varphi|_{L^\infty} \leq \epsilon_0, \left|\frac{\omega_\varphi^n}{\omega^n}\right|_{L^p(M,\omega)} \leq B \text{ for some } p > 1 \right\},$$

Here B and p should be chosen such that $\mathcal{N}(\epsilon_0; B, p)$ is not the empty set. Clearly, if $|\varphi_0|_{C^{1,1}} \leq \epsilon_0$, then $\varphi_0 \in \mathcal{N}\big(\epsilon_0, 1 + (2^n - 1)\epsilon_0, \infty\big)$. Actually, we have:

Lemma 6.2. *Fix $t_0 \in (0, T]$. For any $\epsilon_1 > 0$ there is a small ϵ_0 such that for any $\varphi_0 \in \mathcal{N}(\epsilon_0; B, p)$ we have $|\varphi(t_0)|_{C^{2,\alpha}} \leq \epsilon_1$.*

Proof. If the conclusion fails, choose a sequence of φ_s such that

$$|\varphi_s|_{L^\infty} \leq \frac{1}{s} \quad \text{and} \quad \left|\frac{\omega^n_{\varphi_s}}{\omega^n}\right|_{L^p(M,\omega)} \leq B.$$

For each corresponding solution $\varphi_s(t)$ constructed by Theorem 6.1, we have

$$|\varphi_s(t_0)|_{C^{2,\alpha}} > \epsilon_1. \tag{6-3}$$

Setting $g_{a\varphi i\bar{j}} = \int_0^t (g_{i\bar{j}} + a\varphi_{i\bar{j}})\,da > 0$, we rewrite (6-1) as follows

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = \triangle_{g_{a\varphi}}\varphi + \varphi, \\ \varphi_s(0) = \varphi_s. \end{cases}$$

From the maximum principle we obtain

$$\sup_M |\varphi_s(t_0)| \le e^{t_0} \sup_M |\varphi_s|. \tag{6-4}$$

By (6-2), we can pass a subsequence of $\varphi_{s_i}(t_0)$ such that $\lim_{i\to\infty} \varphi_{s_i}(t_0) = \varphi_\infty(t_0)$ in C^k for $k \ge 0$. Let $s = s_i$ in (6-4). Then the limit approaches $\sup_M |\varphi_\infty(t_0)| \le 0$, which contradicts (6-3). □

Now we have a $C^{2,\alpha}$ small initial datum $\varphi(t_0)$; we normalize it to be $\varphi_0 - I(\varphi_0)$ which is also $C^{2,\alpha}$ small. Then we can solve Equation (2-3) with this initial datum. Combining Propositions 3.1 and 4.5, Theorem 4.6, and Lemma 6.2, we obtain Theorem 1.2. Analogously, we apply Proposition 5.7, Theorem 5.8 and Lemma 6.2 to obtain Theorem 1.1.

7. Another choice of holomorphic transformations

In this section, we follow the arguments in [Bando and Mabuchi 1987; Chen and Tian 2002] to find a good holomorphic transformation. The I and J functionals are defined as

$$I(\omega,\omega_\varphi) = \frac{1}{V}\int_M \varphi(\omega^n - \omega_\varphi^n),$$

$$J(\omega,\omega_\varphi) = \frac{1}{V}\sum_{i=0}^{n-1}\int_M \frac{i+1}{n+1}\sqrt{-1}\,\partial\varphi\wedge\bar\partial\varphi\wedge\omega^i\wedge\omega_\varphi^{n-1-i}.$$

From [Aubin 1998] we know that I and J are both semipositive functionals and satisfy

$$0 \le I(\omega,\omega_\varphi) \le (n+1)(I(\omega,\omega_\varphi) - J(\omega,\omega_\varphi)) \le nI(\omega,\omega_\varphi) \quad \text{for } \varphi\in\mathcal{H}. \tag{7-1}$$

Fix $\varphi\in\mathcal{H}_0$. Consider the functional

$$\Psi(\sigma) = (I-J)(\omega_\varphi, \sigma^*\omega) = (I-J)(\omega_\varphi,\omega_\rho),$$

which is defined for any σ in the reductive subgroup $\mathrm{Aut}(M)$ with $\sigma^*\omega = \omega + \partial\bar\partial\rho$. Since ω_ρ is a Kähler–Einstein metric, it satisfies

$$\log\frac{\omega_\rho^n}{\omega^n} + \rho = 0 \quad \text{and} \quad I(\rho) = 0. \tag{7-2}$$

If ω_ρ is the minimal point of Ψ, for any $u \in \Lambda_1(\omega_\rho)$, we have

$$\int_M (\rho-\varphi)u\omega_\rho^n = 0. \tag{7-3}$$

It is known that $\eta(M) \cong \Lambda_1(\omega)$ for any Kähler–Einstein metric ω [Matsushima 1957]. To prove that the minimizer of Ψ is always attained, it suffices to prove:

Proposition 7.1 [Bando and Mabuchi 1987]. *For all*

$$\rho \in \{\rho \mid \sigma^*\omega = \omega_\rho, \sigma \in \mathrm{Aut}_r(M), \Psi(\sigma) \le r\},$$

we have

$$|\varphi-\rho|_{C^{2,\alpha}(g_\varphi)} \le C(|\varphi|_{C^{4,\alpha}}).$$

Proof. Clearly,

$$-\triangle_\varphi(\rho-\varphi) < n \quad \text{and} \quad -\triangle_\rho(\rho-\varphi) > -n.$$

A lower bound for the Green function is given by

$$G_\varphi \ge -\gamma \frac{D_\varphi^2}{\mathrm{Vol}_\varphi} \doteq -A_\varphi, \tag{7-4}$$

since the volume is constant in a fixed Kähler class and the diameter of g_φ is bounded by $C\,\mathrm{diam}(g)$ when $|\varphi|_{C^2} \le C$. Using Green's formula and (7-4), we obtain

$$\begin{aligned} &\sup_M(\rho-\varphi) \\ &\quad= \frac{1}{V}\int_M(\rho-\varphi)\omega_\varphi^n - \frac{1}{V}\int_M \triangle_\varphi(\rho-\varphi)(y)(G_\varphi(x,y)+A_\varphi)\omega_\varphi^n(y) \\ &\quad\le \frac{1}{V}\int_M(\rho-\varphi)\omega_\varphi^n + nA_\varphi. \end{aligned} \tag{7-5}$$

Similarly, we deduce that

$$\begin{aligned} &\inf_M(\rho-\varphi) \\ &\quad= \frac{1}{V}\int_M(\rho-\varphi)\omega_\rho^n - \frac{1}{V}\int_M \triangle_\rho(\rho-\varphi)(y)(G_\rho(x,y)+A_\rho)\omega_\rho^n(y) \\ &\quad\ge \frac{1}{V}\int_M(\rho-\varphi)\omega_\rho^n - nA_\rho. \end{aligned} \tag{7-6}$$

Because $\mathrm{Ric}(\rho)=\omega_\rho$, we have $\mathrm{diam}(g_\rho) \le \sqrt{2n-1}\pi$ Myers' theorem. Combining (7-5) and (7-6) we get

$$\mathrm{Osc}_M(\rho-\varphi) \ge \frac{1}{V}\int_M(\rho-\varphi)(\omega_\varphi^n-\omega_\rho^n) + C(|\varphi|_{C^2}). \tag{7-7}$$

From (7-1) we obtain

$$\frac{1}{V}\int_M (\rho-\varphi)(\omega_\varphi^n-\omega_\rho^n) = I(\omega_\varphi,\omega_\rho) \le (n+1)(I-J)(\omega_\varphi,\omega_\rho) \le (n+1)r.$$

Since ω_ρ is a Kähler–Einstein metric, we have

$$\big(\omega_\varphi+\sqrt{-1}\,\partial\bar\partial(\rho-\varphi)\big)^n = e^{-(\rho-\varphi)+h_\varphi}\omega_\varphi^n, \tag{7-8}$$

with

$$\sqrt{-1}\,\partial\bar\partial h_\varphi = \mathrm{Ric}(\omega_\varphi)-\omega_\varphi \quad \text{and} \quad \int_M e^{h_\varphi}\omega_\varphi^n = \mathrm{Vol}(M).$$

By using the second-order estimate in [Yau 1978], we get

$$\begin{aligned} n+\triangle_\varphi(\rho-\varphi) &\le e^{C\,\mathrm{Osc}_M(\rho-\varphi)}C\big(\sup_M(\inf_{i\neq k}|R_{\varphi i\bar i k\bar k}|),\, \inf_M S_\varphi,\, \sup_M h_\varphi\big)\\ &\le e^{C\,\mathrm{Osc}_M(\rho-\varphi)}C(|\varphi|_{C^4}). \end{aligned}$$

Then the Krylov estimate shows that $\rho-\varphi$ has $C^{2,\alpha}$ bound. □

Thus we also obtain a uniform bound for the gauge ρ. Our previous discussion implies:

Corollary 7.2. *If $|\varphi|_{C^{4,\alpha}}$ is bounded and ρ is the minimizer of Ψ, then $|\varphi-\rho|_{C^{2,\alpha}}$ and $|\rho|_{C^{2,\alpha}}$ are both bounded.*

This implies that g_ρ is equivalent to g.

We now turn to the uniqueness of the critical points of the functional Ψ when φ is small. The second variation of Ψ at ρ is given by

$$D^2\Psi_\rho(u,v) = \frac{1}{V}\int_M \Big(1+\tfrac{1}{2}\triangle_\rho\rho\Big)uv\omega_\rho^n. \tag{7-9}$$

Lemma 7.3. *For all $|\varphi|_{C^{2,\alpha}} \le \epsilon_1$ and $u \in \Lambda_1(\omega_\rho)$, the bilinear form $D^2\Psi_\rho(u,u)$ is positive definite. Hence ρ is unique.*

Proof. Note that (7-8) can be rewritten as

$$(\omega_\rho+\sqrt{-1}\,\partial\bar\partial(\varphi-\rho))^n = e^{-(\varphi-\rho)-h_\varphi}\omega_\rho^n. \tag{7-10}$$

By definition, h_φ is given by

$$h_\varphi = -\log\frac{\omega_\varphi^n}{\omega^n} - \varphi - \log\bigg(\frac{1}{V}\int_M e^{-\varphi}\bigg)\omega^n.$$

We conclude that

$$|h_\varphi|_{C^{2,\alpha}(g_\rho)} \le C\epsilon_1 \le \delta, \tag{7-11}$$

by the assumption on φ. Let

$$C^{2,\alpha}_{\perp}(M)=\left\{\varphi\in C^{2,\alpha}(M)\;\middle|\;\int_M \varphi u\omega_\rho^n \;\text{ for all } u\in\Lambda_1(\omega_\rho)\right\}.$$

Define the operator of (7-10) by

$$\Phi(a,b)=\log\frac{(\omega_\rho+\sqrt{-1}\,\partial\bar\partial a)^n}{\omega_\rho^n}+a+b,\quad C^{2,\alpha}_{\perp}(M)\times C^{\alpha}(M)\to C^{\alpha}(M).$$

It is clear from (7-10) that $\Phi(\varphi-\rho,h_\varphi)=0$. The linearized operator of (7-10) at $(a,b)=(0,0)$ is given by

$$\delta_a\Phi(v)=\triangle_\rho v+v.$$

We infer that $\delta_a\Phi$ is invertible from $C^{2,\alpha}_{\perp}(M)$ to $C^{\alpha}_{\perp}(M)$. The implicit function theorem implies that there is a small δ neighborhood of 0 in $C^{\alpha}(M)$ such that when $|h_\varphi|_{C^{2,\alpha}(g_\rho)}\le\delta$, we have from (7-3) that

$$|\varphi-\rho|_{C^{2,\alpha}(g_\rho)}\le C\delta.$$

Then we use Corollary 7.2 and (7-11) to obtain

$$|\rho|_{C^{2,\alpha}}\le|\varphi-\rho|_{C^{2,\alpha}}+|\varphi|_{C^{2,\alpha}}\le C\epsilon_1.$$

Hence $|\rho|_{C^{2,\alpha}}$ is small if we choose appropriate ϵ_1 and the bilinear form $D^2\Psi_\rho(u,u)$ is positive definite. □

Acknowledgements

The author is grateful to Professor Xiuxiong Chen, who introduced him to this problem and to Kähler geometry. He is also grateful to Professor Weiyue Ding for his constant encouragement and support. He also thanks Professor Xiaohua Zhu for his interest in this problem and many helpful discussions.

References

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 99i:58001 Zbl 0896.53003

[Bando 1987] S. Bando, "The K-energy map, almost Einstein Kähler metrics and an inequality of the Miyaoka–Yau type", *Tohoku Math. J.* (2) **39**:2 (1987), 231–235. MR 88k:53092

[Bando and Mabuchi 1987] S. Bando and T. Mabuchi, "Uniqueness of Einstein Kähler metrics modulo connected group actions", pp. 11–40 in *Algebraic geometry* (Sendai, 1985), edited by T. Oda, Adv. Stud. Pure Math. **10**, North-Holland, Amsterdam, 1987. MR 89c:53029 Zbl 0641.53065

[Calabi 1985] E. Calabi, "Extremal Kähler metrics, II", pp. 95–114 in *Differential geometry and complex analysis*, edited by I. Chavel and H. E. Rauch, Springer, Berlin, 1985. MR 86h:53067 Zbl 0574.58006

[Calabi and Chen 2002] E. Calabi and X. X. Chen, "The space of Kähler metrics, II", *J. Differential Geom.* **61**:2 (2002), 173–193. MR 2004i:32039 Zbl 1067.58010

[Cao 1985] H. D. Cao, "Deformation of Kähler metrics to Kähler–Einstein metrics on compact Kähler manifolds", *Invent. Math.* **81**:2 (1985), 359–372. MR 87d:58051 Zbl 0574.53042

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

[Chen 2000b] X. Chen, "The space of Kähler metrics", *J. Differential Geom.* **56**:2 (2000), 189–234. MR 2003b:32031 Zbl 1041.58003

[Chen and Ding 2007] X. Chen and W. Ding, "Ricci flow on surfaces with degenerate initial metrics", *J. Partial Differential Equations* **20**:3 (2007), 193–202. MR 2008g:53081 Zbl 1174.53025

[Chen and Li 2009] X. Chen and H. Li, "Stability of Kähler–Ricci Flow", *J. Geom. Anal.* **20**:2 (2009), 306–334. Zbl 1185.53074

[Chen and Tian 2002] X. X. Chen and G. Tian, "Ricci flow on Kähler–Einstein surfaces", *Invent. Math.* **147**:3 (2002), 487–544. MR 2003c:53095 Zbl 1047.53043

[Chen and Tian 2006] X. X. Chen and G. Tian, "Ricci flow on Kähler–Einstein manifolds", *Duke Math. J.* **131**:1 (2006), 17–73. MR 2007c:53083 Zbl 1097.53045

[Chen and Tian 2008] X. X. Chen and G. Tian, "Geometry of Kähler metrics and foliations by holomorphic discs", *Publ. Math. Inst. Hautes Études Sci.* **107** (2008), 1–107. MR 2009g:32048 Zbl 1182.32009

[Chen and Wang 2010] X. Chen and B. Wang, "Remarks on Kähler Ricci flow", *J. Geom. Anal.* **20**:2 (2010), 335–353. MR 2011b:32036 Zbl 1185.53075

[Chen and Zheng 2010] X. Chen and K. Zheng, "The pseudo-Calabi flow", preprint, 2010. to appear in *J. Reine Angew. Math.* arXiv 1004.2663

[Chen et al. 2008] X. X. Chen, G. Tian, and Z. Zhang, "On the weak Kähler-Ricci flow", preprint, 2008. arXiv 0802.0809

[Donaldson 1999] S. K. Donaldson, "Symmetric spaces, Kähler geometry and Hamiltonian dynamics", pp. 13–33 in *Northern California Symplectic Geometry Seminar*, edited by Y. Eliashberg et al., Amer. Math. Soc. Transl. Ser. 2 **196**, Amer. Math. Soc., Providence, RI, 1999. MR 2002b:58008 Zbl 0972.53025

[Donaldson 2004] S. K. Donaldson, "Conjectures in Kähler geometry", pp. 71–78 in *Strings and geometry*, edited by M. Douglas et al., Clay Math. Proc. **3**, Amer. Math. Soc., Providence, RI, 2004. MR 2005i:32024 Zbl 1161.32010

[Fujiki 1978] A. Fujiki, "On automorphism groups of compact Kähler manifolds", *Invent. Math.* **44**:3 (1978), 225–258. MR 58 #1285 Zbl 0367.32004

[Futaki and Mabuchi 1995] A. Futaki and T. Mabuchi, "Bilinear forms and extremal Kähler vector fields associated with Kähler classes", *Math. Ann.* **301**:2 (1995), 199–210. MR 95m:32039 Zbl 0831.53042

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR 84a:53050 Zbl 0504.53034

[Huang and Zheng 2010] H. Huang and K. Zheng, "Stability of Calabi flow near extremal metric", preprint, 2010. to appear in *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* arXiv 1007.4571

[Kobayashi 1961] S. Kobayashi, "On compact Kähler manifolds with positive definite Ricci tensor", *Ann. of Math.* (2) **74** (1961), 570–574. MR 24 #A2922 Zbl 0107.16002

[Mabuchi 1986] T. Mabuchi, "K-energy maps integrating Futaki invariants", *Tohoku Math. J.* (2) **38**:4 (1986), 575–593. MR 88b:53060 Zbl 0619.53040

[Mabuchi 1987] T. Mabuchi, "Some symplectic geometry on compact Kähler manifolds. I", *Osaka J. Math.* **24**:2 (1987), 227–252. MR 88m:53126 Zbl 0645.53038

[Matsushima 1957] Y. Matsushima, "Sur la structure du groupe d'homéomorphismes analytiques d'une certaine variété kählérienne", *Nagoya Math. J.* **11** (1957), 145–150. MR 20 #995 ZBL 0091.34803

[Munteanu 2009] O. Munteanu, "A sharp estimate for the bottom of the spectrum of the Laplacian on Kähler manifolds", *J. Differential Geom.* **83**:1 (2009), 163–187. MR 2010j:58071 Zbl 1183.53068

[Pali 2009] N. Pali, "On the solutions of the Aubin equation and the K-energy of Einstein–Fano manifolds", *Complex Var. Elliptic Equ.* **54**:11 (2009), 1019–1054. MR 2011c:32039 Zbl 1181.53038

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv 0211159v1

[Phong and Sturm 2006] D. H. Phong and J. Sturm, "On stability and the convergence of the Kähler–Ricci flow", *J. Differential Geom.* **72**:1 (2006), 149–168. MR 2007a:53128 Zbl 1125.53048

[Phong et al. 2008] D. H. Phong, J. Song, J. Sturm, and B. Weinkove, "The Kähler–Ricci flow with positive bisectional curvature", *Invent. Math.* **173**:3 (2008), 651–665. MR 2009f:53103 ZBL 1145.53050

[Phong et al. 2009] D. H. Phong, J. Song, J. Sturm, and B. Weinkove, "The Kähler–Ricci flow and the $\overline{\partial}$ operator on vector fields", *J. Differential Geom.* **81**:3 (2009), 631–647. MR 2010b:32035 Zbl 1162.32014

[Phong et al. 2011] D. H. Phong, J. Song, J. Sturm, and B. Weinkove, "On the convergence of the modified Kähler–Ricci flow and solitons", *Comment. Math. Helv.* **86**:1 (2011), 91–112.

[Rubinstein 2009] Y. A. Rubinstein, "On the construction of Nadel multiplier ideal sheaves and the limiting behavior of the Ricci flow", *Trans. Amer. Math. Soc.* **361**:11 (2009), 5839–5850. MR 2010i:32021 Zbl 1185.53077

[Semmes 1992] S. Semmes, "Complex Monge–Ampère and symplectic manifolds", *Amer. J. Math.* **114**:3 (1992), 495–550. MR 94h:32022 Zbl 0790.32017

[Sesum and Tian 2008] N. Sesum and G. Tian, "Bounding scalar curvature and diameter along the Kähler Ricci flow (after Perelman)", *J. Inst. Math. Jussieu* **7**:3 (2008), 575–587. MR 2009c:53092

[Simon 1983] L. Simon, "Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems", *Ann. of Math.* (2) **118**:3 (1983), 525–571. MR 85b:58121

[Song and Tian 2009] J. Song and G. Tian, "The Kähler–Ricci flow through singularities", preprint, 2009. arXiv 0909.4898

[Székelyhidi 2010] G. Székelyhidi, "The Kähler-Ricci flow and K-polystability", *Amer. J. Math.* **132**:4 (2010), 1077–1090. MR 2663648 Zbl 1206.53075

[Tian 2000] G. Tian, *Canonical metrics in Kähler geometry*, Birkhäuser, Basel, 2000. Notes taken by Meike Akveld. MR 2001j:32024 Zbl 0978.53002

[Tian and Zhu 2000] G. Tian and X. Zhu, "Uniqueness of Kähler–Ricci solitons", *Acta Math.* **184**:2 (2000), 271–305. MR 2001h:32040 Zbl 1036.53052

[Tian and Zhu 2002] G. Tian and X. Zhu, "A new holomorphic invariant and uniqueness of Kähler–Ricci solitons", *Comment. Math. Helv.* **77**:2 (2002), 297–325. MR 2003i:32042 Zbl 1036.53053

[Tian and Zhu 2007] G. Tian and X. Zhu, "Convergence of Kähler–Ricci flow", *J. Amer. Math. Soc.* **20**:3 (2007), 675–699. MR 2007k:53107 Zbl 1185.53078

[Tian and Zhu 2008] G. Tian and X. Zhu, “Perelman’s W-functional and stability of Kähler–Ricci flow”, preprint, 2008. arXiv 0801.3504

[Tosatti 2010] V. Tosatti, “Kähler–Ricci flow on stable Fano manifolds”, *J. Reine Angew. Math.* **640** (2010), 67–84. MR 2011e:32030 Zbl 1189.53069

[Yau 1978] S. T. Yau, “On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation. I”, *Comm. Pure Appl. Math.* **31**:3 (1978), 339–411. MR 81d:53045 Zbl 0369.53059

[Zhu 2009] X. Zhu, “Stability on Kähler–Ricci flow, I”, preprint, 2009. arXiv 0908.1488

Received May 23, 2010. Revised November 18, 2010.

KAI ZHENG
ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
BEIJING, 100190
CHINA

kaizheng@amss.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 2, 2011

THE SECOND VARIATION OF THE RICCI EXPANDER ENTROPY

MENG ZHU

The critical points of the $\mathcal{W}_+$ functional introduced by M. Feldman, T. Ilmanen and L. Ni are the expanding Ricci solitons, which are special solutions of the Ricci flow. On compact manifolds, expanding solitons coincide with Einstein metrics. In this paper, we compute the first and second variations of the entropy functional of the $\mathcal{W}_+$ functional, and briefly discuss the linear stability of compact hyperbolic space forms.

1. Introduction

Perelman [2002] introduced two important functionals, denoted by $\mathcal{F}$ and $\mathcal{W}$. The corresponding entropy functionals λ and ν are monotone along the Ricci flow $\partial g_{ij}/\partial t = -2R_{ij}$ and are constant precisely on steady and shrinking solitons. H.-D. Cao, R. Hamilton and T. Ilmanen [Cao et al. 2004] presented the second variations of both entropy functionals and studied the linear stabilities of certain closed Einstein manifolds of nonnegative scalar curvature.

To find the corresponding variational structure for the expanding case, M. Feldman, T. Ilmanen and L. Ni [Feldman et al. 2005] introduced the functional $\mathcal{W}_+$. Let (M^n, g) be a compact Riemannian manifold, f a smooth function on M, and $\sigma > 0$. Define

$$\mathcal{W}_+(g, f, \sigma) = (4\pi\sigma)^{-n/2} \int_M e^{-f} \big(\sigma(|\nabla f|^2 + R) - f + n\big)\, dV,$$

$$\mu_+(g, \sigma) = \inf\left\{ W_+(g, f, \sigma) \;\middle|\; f \in C^\infty(M) \text{ with } (4\pi\sigma)^{-n/2} \int_M e^{-f}\, dV = 1 \right\},$$

$$\nu_+(g) = \sup_{\sigma > 0} \mu_+(g, \sigma).$$

Then ν_+ is nondecreasing along the Ricci flow and constant precisely on expanding solitons.

Research is partially supported by NSF grant DMS-0354621.

MSC2000: 53C21, 53C25, 53C44, 58J60.

Keywords: entropy functional, ν_+ functional, $\mathcal{W}_+$ functional, linear stability, linear variation, negative Einstein manifold, second variation.

In this note, analogous to [Cao et al. 2004], we present the first and second variations of the entropy ν_+. By computing the first variation of ν_+, one can see that the critical points are expanding solitons, which are actually negative Einstein manifolds (see [Cao and Zhu 2006], for example). Our main result is this:

Theorem 1.1. *Let (M^n, g) be a compact negative Einstein manifold. Let h be a symmetric 2-tensor. Consider the variation of metric $g(s) = g + sh$. Then the second variation of ν_+ is*

$$\frac{\mathrm{d}^2\nu_+(g(s))}{\mathrm{d}s^2}\Big|_{s=0} = \frac{\sigma}{\operatorname{Vol} g}\int_M \langle N_+h, h\rangle,$$

where

$$N_+h := \tfrac{1}{2}\Delta h + \operatorname{div}^*\operatorname{div} h + \tfrac{1}{2}\nabla^2 v_h + \operatorname{Rm}(h,\cdot\,) + \frac{g}{2n\sigma\operatorname{Vol} g}\int_M \operatorname{tr} h;$$

here tr *is the trace with respect to g and v_h is the unique solution of*

$$\Delta v_h - \frac{v_h}{2\sigma} = \operatorname{div}(\operatorname{div} h), \qquad \int_M v_h = 0.$$

In this case, we may still define the concept of linear stability. We say that an expanding soliton is *linearly stable* if $N_+ \le 0$; otherwise it is *linearly unstable*. Similar to [Cao et al. 2004], the N_+ operator is nonpositive definite if and only if the maximal eigenvalue of the Lichnerowicz Laplacian acting on the space of transverse traceless 2-tensors has a certain upper bound. Using the results in [Delay 2002; 2008] or [Lee 2006], one can then see that compact hyperbolic spaces are linearly stable. But unlike the positive Einstein case, it seems hard to find other examples of negative Einstein manifolds which are either linear stable or linear unstable.

2. The first variation of the expander entropy

Recall that in [Perelman 2002], the $\mathcal{F}$ functional is defined by

$$\mathcal{F}(f, g) = \int_M (|\nabla f|^2 + R)e^{-f}\,dV,$$

and its entropy $\lambda(g)$ is

$$\lambda(g) = \inf\left\{\mathcal{F}(f, g)\;\middle|\; f \in C^\infty(M) \text{ with } \int_M e^{-f} = 1\right\},$$

where R is the scalar curvature. By [Feldman et al. 2005, Theorem 1.7], we know that $\mu_+(g, \sigma)$ is attained by some function f. Moreover, if $\lambda(g) < 0$, then $\nu_+(g)$ can be attained by some positive number σ.

Lemma 2.1. *If $\nu_+(g)$ is realized by some f and σ, it is necessary that the pair (f,σ) solves the equations*

$$\sigma(-2\Delta f+|\nabla f|^2-R)+f-n+\nu_+=0 \tag{1}$$

and

$$(4\pi\sigma)^{-n/2}\int_M fe^{-f}\,dV=\frac{n}{2}-\nu_+. \tag{2}$$

Proof. For fixed $\sigma>0$, suppose that $\mu_+(g,\sigma)$ is attained by some function f. Using the Lagrange multiplier method, consider the following functional

$$\begin{aligned}&L(g,f,\sigma,\lambda)\\&=(4\pi\sigma)^{-n/2}\int_M e^{-f}\big(\sigma(|\nabla f|^2+R)-f+n\big)\,dV+\lambda\Big((4\pi\sigma)^{-n/2}\int_M e^{-f}\,dV-1\Big).\end{aligned}$$

Denote by δf the variation of f. Then the variation of L is

$$\begin{aligned}0&=\delta L\\&=(4\pi\sigma)^{-n/2}\int_M e^{-f}(-\delta f)\big(\sigma(|\nabla f|^2+R)-f+n\big)\,dV\\&\qquad+(4\pi\sigma)^{-n/2}\int_M e^{-f}\big(2\sigma\nabla f\nabla(\delta f)-\delta f\big)\,dV-(4\pi\sigma)^{-n/2}\int_M\lambda(\delta f)e^{-f}\,dV\\&=(4\pi\sigma)^{-n/2}\int_M e^{-f}(\delta f)\big(\sigma(-2\Delta f+|\nabla f|^2-R)\big)\,dV\\&\qquad\qquad+(4\pi\sigma)^{-n/2}\int_M e^{-f}(\delta f)(f-n-1-\lambda)\,dV\end{aligned}$$

Therefore,

$$\sigma(-2\Delta f+|\nabla f|^2-R)+f-n-1-\lambda=0.$$

Integrating both sides with respect to the measure $(4\pi\sigma)^{-n/2}e^{-f}\,dV$, we get

$$-\lambda-1=(4\pi\sigma)^{-n/2}\int_M e^{-f}\big(\sigma(|\nabla f|^2+R)-f+n\big)\,dV=\mu_+(g,\sigma).$$

When σ and f realize $\nu_+(g)$, this is just Equation (1).

Now we consider the variations $\delta\sigma$ and δf of both σ and f. We have

$$\begin{aligned}0=&(4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\delta\sigma-\delta f\Big)\big(\sigma(|\nabla f|^2+R)-f+n\big)\,dV\\&+(4\pi\sigma)^{-n/2}\int_M e^{-f}\big(\delta\sigma(|\nabla f|^2+R)+2\sigma\nabla f\nabla(\delta f)-\delta f\big)\,dV\end{aligned} \tag{3}$$

and

$$(4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\delta\sigma-\delta f\Big)\,dV=0. \tag{4}$$

Using (1) and (4), we can write (3) as

$$
\begin{aligned}
0 &= (4\pi\sigma)^{-n/2}\int_M e^{-f}\big(\delta\sigma(|\nabla f|^2+R)-\delta f\big)\,dV\\
&= (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\frac{1}{\sigma}\delta\sigma(\nu_+ + f - n)+\frac{n}{2\sigma}\delta\sigma\Big)\,dV\\
&= (\delta\sigma)\frac{1}{\sigma}(4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\nu_+ + f-\frac{n}{2}\Big)\,dV,
\end{aligned}
$$

which gives (2). □

Before computing the variations of the ν_+ functional, let's recall some variation formulas for curvatures. By direct computation, we have:

Lemma 2.2. *Suppose that h is a symmetric 2-tensor and $g(s)=g+sh$ is a variation of g. Then*

$$
\frac{\partial R}{\partial s}\Big|_{s=0} = -h_{kl}R_{kl}+\nabla_p\nabla_k h_{pk}-\Delta\,\mathrm{tr}\,h \tag{5}
$$

and

$$
\begin{aligned}
\frac{\partial^2 R}{\partial s^2}\Big|_{s=0} &= 2h_{kp}h_{pl}R_{kl}-2h_{kl}\frac{\partial R_{kl}}{\partial s}\Big|_{s=0}+g^{kl}\frac{\partial^2 R_{kl}}{\partial s^2}\Big|_{s=0}\\
&= 2h_{kp}h_{pl}R_{kl}-h_{kl}(2\nabla_p\nabla_k h_{pl}-\Delta h_{kl}-\nabla_k\nabla_l\,\mathrm{tr}\,h)\\
&\quad -\nabla_p\big(h_{pq}(2\nabla_k h_{kq}-\nabla_q\,\mathrm{tr}\,h)\big)+\nabla_k(h_{pq}\nabla_k h_{pq})\\
&\quad +\tfrac{1}{2}\nabla_p\,\mathrm{tr}\,h(2\nabla_k h_{kp}-\nabla_p\,\mathrm{tr}\,h)+\tfrac{1}{2}(\nabla_k h_{pq}\nabla_k h_{pq}-2\nabla_p h_{kq}\nabla_q h_{kp}),
\end{aligned} \tag{6}
$$

where ∇ is the Levi-Civita connection of g and $\mathrm{tr}\,h$ *is the trace of h taken with respect to g.*

Now we are ready to compute the first variation of $\nu_+(g)$.

Proposition 2.3. *Let (M^n, g) be a compact Riemannian manifold with $\lambda(g)<0$. Let h be any symmetric covariant 2-tensor on M, and consider the variation*

$$
g(s)=g+sh.
$$

Then the first variation of $\nu_+(g(s))$ is

$$
\frac{d\nu_+(g(s))}{ds}\Big|_{s=0} = (4\pi\sigma)^{-n/2}\int_M \sigma e^{-f}\Big(-R_{ij}-\nabla_i\nabla_j f-\frac{1}{2\sigma}g_{ij}\Big)h_{ij}\,dV,
$$

where the smooth function f and $\sigma>0$ realize $\nu_+(g)$.

Proof. By taking derivatives directly, we have

$$\frac{\partial \nu_+}{\partial s} = (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)\big(\sigma(|\nabla f|^2 + R)\big)\,dV$$
$$+ (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)(-f+n)\,dV$$
$$+ (4\pi\sigma)^{-n/2}\int_M e^{-f}\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R)\,dV$$
$$- (4\pi\sigma)^{-n/2}\int_M e^{-f}(\sigma g^{ip}g^{jq}h_{pq}\nabla_i f\nabla_j f)\,dV$$
$$+ (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\sigma\Big(2g^{ij}\nabla_i f\nabla_j\frac{\partial f}{\partial s} + \frac{\partial R}{\partial s}\Big) - \frac{\partial f}{\partial s}\Big)\,dV. \tag{7}$$

Since $(4\pi\sigma)^{-n/2}\int_M e^{-f}\,dV = 1$, we have

$$(4\pi\sigma)^{-n/2}\int_M \Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)e^{-f}\,dV = 0. \tag{8}$$

Substituting (1), (2) and (8) in (7), we obtain

$$\frac{\partial\nu_+(s)}{\partial s}\Big|_{s=0}$$
$$= (4\pi\sigma)^{-n/2}\int_M \big(2\sigma(|\nabla f|^2 - \Delta f) + \nu_+(0)\big)\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)e^{-f}\,dV$$
$$+ (4\pi\sigma)^{-n/2}\int_M \Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) - \frac{\partial f}{\partial s} - \sigma h_{ij}\nabla_i f\nabla_j f\Big)e^{-f}\,dV$$
$$+ (4\pi\sigma)^{-n/2}\int_M \sigma\Big(2\frac{\partial f}{\partial s}(|\nabla f|^2 - \Delta f) + \nabla_i\nabla_j h_{ij} - \Delta\operatorname{tr}h - h_{ij}R_{ij}\Big)e^{-f}\,dV$$
$$= (4\pi\sigma)^{-n/2}\int_M \Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) - \frac{\partial f}{\partial s} - \sigma(h_{ij}\nabla_i\nabla_j f + h_{ij}R_{ij})\Big)e^{-f}\,dV$$
$$= (4\pi\sigma)^{-n/2}\int_M \Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) + \frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}\Big)e^{-f}\,dV$$
$$- (4\pi\sigma)^{-n/2}\int_M \sigma h_{ij}\Big(R_{ij} + \nabla_i\nabla_j f + \frac{1}{2\sigma}g_{ij}\Big)e^{-f}\,dV$$
$$= (4\pi\sigma)^{-n/2}\int_M \frac{1}{\sigma}\frac{\partial\sigma}{\partial s}\Big(f(0) - \frac{n}{2} + \nu_+(0) - 2\sigma(|\nabla f|^2 - \Delta f)\Big)e^{-f}\,dV$$
$$-(4\pi\sigma)^{-n/2}\int_M \sigma h_{ij}\Big(R_{ij} + \nabla_i\nabla_j f + \frac{1}{2\sigma}g_{ij}\Big)e^{-f}\,dV$$
$$= -(4\pi\sigma)^{-n/2}\int_M \sigma h_{ij}\Big(R_{ij} + \nabla_i\nabla_j f + \frac{1}{2\sigma}g_{ij}\Big)e^{-f}\,dV.$$

Hence, the first variation of ν_+ is

$$\frac{d\nu_+(g(s))}{ds}\Big|_{s=0} = (4\pi\sigma)^{-n/2}\int_M \sigma e^{-f}\Big(-R_{ij} - \nabla_i\nabla_j f - \frac{1}{2\sigma}g_{ij}\Big)h_{ij}\,dV. \qquad \square$$

From the proposition, we can see that a critical point of $\nu_+(g)$ satisfies

$$\mathrm{Rc} + \nabla^2 f + \frac{1}{2\sigma}g = 0,$$

which means that (M, g) is a gradient expanding soliton.

3. The second variation

Now we compute the second variation of ν_+. Since any compact expanding soliton is Einstein (see [Cao and Zhu 2006], for example), f is a constant. After adding a constant to f we may assume that $f = n/2$.

In the following, as in [Cao et al. 2004], we set $\mathrm{Rm}(h, h) = R_{ijkl}h_{ik}h_{jl}$, $\operatorname{div}\omega = \nabla_i\omega_i$, $(\operatorname{div} h)_i = \nabla_j h_{ji}$, and $(\operatorname{div}^*\omega)_{ij} = -(\nabla_i\omega_j + \nabla_j\omega_i) = -\frac{1}{2}L_{\omega^\#}g_{ij}$, where h is a symmetric 2-tensor, ω is a 1-tensor, $\omega^\#$ is the dual vector field of ω, and $L_{\omega^\#}$ is the Lie derivative.

Proof of Theorem 1.1. Let (M, g) be a compact negative Einstein manifold with $f = n/2$ and $R_{ij} = -1/(2\sigma)g_{ij}$. For any symmetric 2-tensor h, consider the variation $g(s) = g + sh$. By Proposition 2.3, we know that $(d\nu_+/ds)|_{s=0} = 0$.

From (1) and (2), we get

$$\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}(0) - 2\sigma\Delta\frac{\partial f}{\partial s}(0) - \sigma\frac{\partial R}{\partial s}(0) + \frac{\partial f}{\partial s}(0) = 0, \tag{9}$$

and

$$(4\pi\sigma)^{-n/2}\int_M e^{-n/2}\bigg(\frac{n}{2}\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}(0) - \frac{\partial f}{\partial s}(0) + \frac{1}{2}\operatorname{tr} h\Big) + \frac{\partial f}{\partial s}(0)\bigg)\,dV = 0.$$

It follows by (8) that

$$(4\pi\sigma)^{-n/2}\int_M \frac{\partial f}{\partial s}(0)e^{-n/2}\,dV = 0 \tag{10}$$

and

$$\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}(0) = \frac{1}{\operatorname{Vol} g}\int_M \frac{1}{2}\operatorname{tr} h\,dV, \tag{11}$$

where $(4\pi\sigma)^{-n/2}e^{-n/2} = \dfrac{1}{\operatorname{Vol} g}$. Thus

$$
\begin{aligned}
\frac{\mathrm{d}\nu_+}{\mathrm{d}s} &= (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)\big(\sigma(|\nabla f|^2 + R) - f + n\big)\,dV\\
&\qquad + (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) - \frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad + (4\pi\sigma)^{-n/2}\int_M \sigma e^{-f}\Big(-g^{ip}g^{jq}h_{pq}\nabla_i f\nabla_j f + 2g^{ij}\nabla_i f\nabla_j\frac{\partial f}{\partial s} + \frac{\partial R}{\partial s}\Big)\,dV\\
&= (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}g^{ij}h_{ij}\Big)\big(2\sigma(|\nabla f|^2 - \Delta f) + \nu_+\big)\,dV\\
&\qquad + (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) - \frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad + (4\pi\sigma)^{-n/2}\int_M \sigma e^{-f}\Big(-g^{ip}g^{jq}h_{pq}\nabla_i f\nabla_j f + 2g^{ij}\nabla_i f\nabla_j\frac{\partial f}{\partial s} + \frac{\partial R}{\partial s}\Big)\,dV\\
&= (4\pi\sigma)^{-n/2}\int_M \sigma e^{-f}g^{ij}h_{ij}(|\nabla f|^2 - \Delta f)\,dV\\
&\qquad + (4\pi\sigma)^{-n/2}\int_M e^{-f}\Big(\sigma\Big(-g^{ip}g^{jq}h_{pq}\nabla_i f\nabla_j f + \frac{\partial R}{\partial s}\Big) - \frac{1}{2}g^{ij}h_{ij}\Big)\,dV,
\end{aligned}
$$

where we note that

$$
\int_M 2\sigma e^{-f}g^{ij}\nabla_i f\nabla_j\frac{\partial f}{\partial s}\,dV = \int_M 2\sigma e^{-f}\frac{\partial f}{\partial s}(|\nabla f|^2 - \Delta f)\,dV
$$

and

$$
\begin{aligned}
&\int_M e^{-f}\Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) - \frac{\partial f}{\partial s}\Big)\,dV\\
&\quad = \int_M e^{-f}\Big(\frac{\partial\sigma}{\partial s}(|\nabla f|^2 + R) + \frac{n}{2\sigma}\frac{\partial\sigma}{\partial s} - \frac{1}{2}g^{ij}h_{ij}\Big)\,dV\\
&\quad = \int_M e^{-f}\Big(\frac{1}{\sigma}\frac{\partial\sigma}{\partial s}\Big(\sigma(|\nabla f|^2 + R) + \frac{n}{2}\Big) - \frac{1}{2}g^{ij}h_{ij}\Big)\,dV\\
&\quad = \int_M e^{-f}\frac{1}{\sigma}\frac{\partial\sigma}{\partial s}\Big(\sigma(2|\nabla f|^2 - 2\Delta f) + f - \frac{n}{2} + \nu_+\Big)\,dV - \int_M e^{-f}\cdot\tfrac{1}{2}g^{ij}h_{ij}\,dV\\
&\quad = -\int_M e^{-f}\cdot\tfrac{1}{2}g^{ij}h_{ij}\,dV.
\end{aligned}
$$

Since $f(0) = \dfrac{n}{2}$, we have

$$\frac{d^2\nu_+}{ds^2}\Big|_{s=0} = -\frac{1}{\operatorname{Vol} g}\int_M \sigma \operatorname{tr} h \Delta \frac{\partial f}{\partial s} dV + \frac{1}{\operatorname{Vol} g}\int_M \Big(-\frac{n}{2\sigma}\frac{\partial \sigma}{\partial s} - \frac{\partial f}{\partial s} + \frac{1}{2}\operatorname{tr} h\Big)\Big(\sigma \frac{\partial R}{\partial s} - \frac{1}{2}\operatorname{tr} h\Big) dV + \frac{1}{\operatorname{Vol} g}\int_M \Big(\frac{\partial \sigma}{\partial s}\frac{\partial R}{\partial s} + \sigma \frac{\partial^2 R}{\partial s^2} + \frac{1}{2}|h_{ij}|^2\Big) dV. \tag{12}$$

In the following, all quantities are evaluated at $s = 0$. First, we have

$$\begin{aligned} &\frac{1}{\operatorname{Vol} g}\int_M \sigma \frac{\partial^2 R}{\partial s^2} dV \\ &= \frac{\sigma}{\operatorname{Vol} g}\int_M \Big(-\frac{1}{\sigma}|h_{ij}|^2 - h_{kl}(2\nabla_p\nabla_k h_{pl} - \Delta h_{kl} - \nabla_k\nabla_l \operatorname{tr} h) \\ &\quad - \nabla_p\big(h_{pq}(2\nabla_k h_{kq} - \nabla_q \operatorname{tr} h)\big) + \nabla_k(h_{pq}\nabla_k h_{pq}) \\ &\quad + \tfrac{1}{2}\nabla_p \operatorname{tr} h(2\nabla_k h_{kp} - \nabla_p \operatorname{tr} h) + \tfrac{1}{2}(\nabla_k h_{pq}\nabla_k h_{pq} - 2\nabla_p h_{kq}\nabla_q h_{kp})\Big) dV \\ &= \frac{\sigma}{\operatorname{Vol} g}\int_M \Big(-\frac{1}{\sigma}|h_{ij}|^2 - h_{kl}\nabla_p\nabla_k h_{pl} - \frac{1}{2}|\nabla h|^2 - \frac{1}{2}|\nabla \operatorname{tr} h|^2\Big) dV \\ &= \frac{\sigma}{\operatorname{Vol} g}\int_M \Big(-\frac{1}{\sigma}|h_{ij}|^2 - \frac{1}{2}|\nabla h|^2 - \frac{1}{2}|\nabla \operatorname{tr} h|^2\Big) dV \\ &\qquad - \frac{\sigma}{\operatorname{Vol} g}\int_M h_{kl}(\nabla_k\nabla_p h_{pl} + R_{kq}h_{ql} + R_{pkql}h_{pq}) dV \\ &= -\frac{1}{\operatorname{Vol} g}\int_M \frac{1}{2}|h_{ij}|^2 dV \\ &\qquad + \frac{\sigma}{\operatorname{Vol} g}\int_M \big(|\operatorname{div} h|^2 + \operatorname{Rm}(h,h) - \tfrac{1}{2}|\nabla h|^2 - \tfrac{1}{2}|\nabla \operatorname{tr} h|^2\big) dV. \end{aligned} \tag{13}$$

Moreover,

$$\begin{aligned} \frac{1}{\operatorname{Vol} g}\int_M \frac{\partial \sigma}{\partial s}\frac{\partial R}{\partial s} dV &= \frac{\sigma}{n}\frac{1}{\operatorname{Vol} g}\int_M \operatorname{tr} h\, dV \frac{1}{\operatorname{Vol} g}\int_M \frac{\partial R}{\partial s} dV \\ &= \frac{1}{2n}\Big(\frac{1}{\operatorname{Vol} g}\int_M \operatorname{tr} h\, dV\Big)^2. \end{aligned} \tag{14}$$

Let v_h be the solution to the equation

$$\Delta v_h - \frac{v_h}{2\sigma} = \operatorname{div}\operatorname{div} h = \nabla_p\nabla_q h_{pq}, \qquad \int_M v_h = 0.$$

Then

$$
\begin{aligned}
&\frac{1}{\operatorname{Vol} g}\int_M\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}-\frac{\partial f}{\partial s}+\frac{1}{2}\operatorname{tr}h\Big)\sigma\frac{\partial R}{\partial s}dV\\
&\quad=\frac{\sigma}{\operatorname{Vol} g}\int_M\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}-\frac{\partial f}{\partial s}+\frac{1}{2}\operatorname{tr}h\Big)\Big(\Delta v_h-\frac{v_h}{2\sigma}+\frac{1}{2\sigma}\operatorname{tr}h-\Delta\operatorname{tr}h\Big)\,dV\\
&\quad=-\Big(\frac{1}{\operatorname{Vol} g}\int_M\frac{1}{2}\operatorname{tr}h\,dV\Big)^2+\frac{\sigma}{\operatorname{Vol} g}\int_M v_h\Big(-\Delta\frac{\partial f}{\partial s}+\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad\qquad+\frac{\sigma}{\operatorname{Vol} g}\int_M\operatorname{tr}h\Big(\Delta\frac{\partial f}{\partial s}-\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad\qquad+\frac{\sigma}{\operatorname{Vol} g}\int_M\frac{1}{2}\operatorname{tr}h\Big(\Delta v_h-\frac{v_h}{2\sigma}+\frac{1}{2\sigma}\operatorname{tr}h-\Delta\operatorname{tr}h\Big)\,dV,
\end{aligned}
$$

where we have used (11) to derive the first term in the last equality. Meanwhile,

$$
\begin{aligned}
&-\frac{1}{\operatorname{Vol} g}\int_M\frac{1}{2}\operatorname{tr}h\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}-\frac{\partial f}{\partial s}+\frac{1}{2}\operatorname{tr}h\Big)\\
&\qquad\qquad=-\frac{1}{\operatorname{Vol} g}\int_M\frac{1}{2}\operatorname{tr}h\Big(-2\sigma\,\Delta\frac{\partial f}{\partial s}-\sigma\frac{\partial R}{\partial s}+\frac{1}{2}\operatorname{tr}h\Big).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\frac{1}{\operatorname{Vol} g}\int_M\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}-\frac{\partial f}{\partial s}+\frac{1}{2}\operatorname{tr}h\Big)\Big(\sigma\frac{\partial R}{\partial s}-\frac{1}{2}\operatorname{tr}h\Big)\,dV\\
&\quad=\frac{1}{\operatorname{Vol} g}\int_M\sigma\operatorname{tr}h\Delta\frac{\partial f}{\partial s}dV-\frac{1}{\operatorname{Vol} g}\int_M\frac{1}{4}(\operatorname{tr}h)^2dV\\
&\qquad-\Big(\frac{1}{\operatorname{Vol} g}\int_M\frac{1}{2}\operatorname{tr}h\,dV\Big)^2+\frac{\sigma}{\operatorname{Vol} g}\int_M v_h\Big(-\Delta\frac{\partial f}{\partial s}+\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad+\frac{\sigma}{\operatorname{Vol} g}\int_M\operatorname{tr}h\Big(\Delta\frac{\partial f}{\partial s}-\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV\\
&\qquad+\frac{\sigma}{\operatorname{Vol} g}\int_M\operatorname{tr}h\Big(\Delta v_h-\frac{v_h}{2\sigma}+\frac{1}{2\sigma}\operatorname{tr}h-\Delta\operatorname{tr}h\Big)\,dV.
\end{aligned}
$$

Now since

$$
\begin{aligned}
\frac{\sigma}{\operatorname{Vol} g}\int_M v_h\Big(-\Delta\frac{\partial f}{\partial s}+\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV&=\frac{\sigma}{\operatorname{Vol} g}\int_M v_h\Big(-\frac{n}{4\sigma^2}\frac{\partial\sigma}{\partial s}+\frac{1}{2}\frac{\partial R}{\partial s}\Big)\,dV\\
&=\frac{\sigma}{\operatorname{Vol} g}\int_M\frac{1}{2}v_h\Big(\Delta v_h-\frac{v_h}{2\sigma}+\frac{1}{2\sigma}\operatorname{tr}h-\Delta\operatorname{tr}h\Big)\,dV\\
&=\frac{\sigma}{\operatorname{Vol} g}\int_M-\frac{1}{2}|\nabla v_h|^2-\frac{v_h^2}{4\sigma}+\frac{v_h}{4\sigma}\operatorname{tr}h-\frac{1}{2}v_h\Delta\operatorname{tr}h\,dV
\end{aligned}
$$

and

$$\frac{\sigma}{\operatorname{Vol} g}\int_M \operatorname{tr} h\Big(\Delta\frac{\partial f}{\partial s}-\frac{1}{2\sigma}\frac{\partial f}{\partial s}\Big)\,dV$$
$$=\frac{\sigma}{\operatorname{Vol} g}\int_M \operatorname{tr} h\Big(\frac{n}{4\sigma^2}\frac{\partial\sigma}{\partial s}-\frac{1}{2}\frac{\partial R}{\partial s}\Big)\,dV$$
$$=\Big(\frac{1}{\operatorname{Vol} g}\int_M \frac{1}{2}\operatorname{tr} h\;dV\Big)^2-\frac{\sigma}{\operatorname{Vol} g}\int_M \frac{1}{2}\operatorname{tr} h\Big(\Delta v_h-\frac{v_h}{2\sigma}+\frac{1}{2\sigma}\operatorname{tr} h-\Delta\operatorname{tr} h\Big)\,dV,$$

we have

$$\frac{1}{\operatorname{Vol} g}\int_M\Big(-\frac{n}{2\sigma}\frac{\partial\sigma}{\partial s}-\frac{\partial f}{\partial s}+\frac{1}{2}\operatorname{tr} h\Big)\Big(\sigma\frac{\partial R}{\partial s}-\frac{1}{2}\operatorname{tr} h\Big)\,dV \tag{15}$$
$$=\frac{1}{\operatorname{Vol} g}\int_M \sigma\operatorname{tr} h\Delta\frac{\partial f}{\partial s}dV+\frac{\sigma}{\operatorname{Vol} g}\int_M\Big(-\frac{1}{2}|\nabla v_h|^2-\frac{v_h^2}{4\sigma}+\frac{1}{2}|\nabla\operatorname{tr} h|^2\Big)\,dV.$$

Substituting (13), (14) and (15) in (12), we get

$$\frac{d^2\nu_+}{ds^2}\Big|_{s=0}=\frac{\sigma}{\operatorname{Vol} g}\left(\int_M\Big(|\operatorname{div} h|^2+\operatorname{Rm}(h,h)-\frac{1}{2}|\nabla h|^2-\frac{1}{2}|\nabla v_h|^2-\frac{v_h^2}{4\sigma}\Big)\,dV\right)+\frac{1}{2n}\left(\frac{1}{\operatorname{Vol} g}\int_M\operatorname{tr} h\,dV\right)^2$$
$$=\frac{\sigma}{\operatorname{Vol} g}\int_M\langle N_+h,h\rangle. \qquad\square$$

As a simple application, we discuss briefly the linear stability of negative Einstein manifolds. In analogy with [Cao et al. 2004], we say that a negative Einstein manifold is linearly stable if $N_+\le 0$, otherwise it is linearly unstable. As in that paper, decompose the space of symmetric 2-tensors as

$$\ker\operatorname{div}\oplus\operatorname{im}\operatorname{div}^*,$$

and further decompose $\ker\operatorname{div}$ as

$$(\ker\operatorname{div})_0\oplus\mathbb{R}g,$$

where $(\ker\operatorname{div})_0$ is the space of divergence free 2-tensors h with $\int_M\operatorname{tr} h=0$. It is easy to see that N_+ vanishes on $\operatorname{im}\operatorname{div}^*$, and on $(\ker\operatorname{div})_0$

$$N_+=\frac{1}{2}\Big(\Delta_L-\frac{1}{\sigma}\Big),$$

where $\Delta_L=\Delta+2\operatorname{Rm}(\cdot\,,\cdot)-2\operatorname{Rc}$ is the Lichnerowicz Laplacian on symmetric 2-tensors.

Moreover, we may write $(\ker \operatorname{div})_0$ as

$$(\ker \operatorname{div})_0 = S_0 \oplus S_1,$$

where S_0 is the subspace of trace free 2-tensors and

$$S_1 = \left\{ h \in (\ker \operatorname{div})_0 \,\middle|\, h_{ij} = \left(-\frac{1}{2\sigma}u + \Delta u\right) g_{ij} - \nabla_i \nabla_j u,\ u \in C^\infty(M) \text{ and } \int_M u = 0 \right\};$$

see [Buzzanca 1984], for example.

Define

$$Tu := \left(-\frac{1}{2\sigma}u + \Delta u\right) g_{ij} - \nabla_i \nabla_j u.$$

Since $\Delta_L(Tu) = T(\Delta u)$ for all smooth functions u and $\ker T = \{0\}$, we can see that the Lichnerowicz Laplacian and the Laplacian on function space have the same eigenvalues. Thus N_+ is always negative on S_1. Therefore, to study the linear stability of negative Einstein manifolds, it remains to look at the behavior of Δ_L acting on S_0 which is the space of transverse traceless 2-tensors.

Example. Suppose that M is an n dimensional compact real hyperbolic space with $n \geq 3$. By [Delay 2002] or [Lee 2006], the biggest eigenvalue of Δ_L on trace free symmetric 2-tensors on real hyperbolic space is $-\frac{1}{4}(n-1)(n-9)$. Since on M we have $\mathrm{Rc} = -(n-1)g$, we obtain

$$\frac{1}{\sigma} = 2(n-1).$$

Thus the biggest eigenvalue of N_+ on S_0 is not greater than $-\frac{1}{8}(n-1)^2$. This implies that M is linearly stable for $n \geq 3$.

Remarks. (1) When $n = 3$, D. Knopf and A. Young [2009] proved that closed 3-folds with constant negative curvature are geometrically stable under certain normalized Ricci flow. R. Ye [1993] had obtained a more powerful stability result earlier.

(2) For $n = 2$, R. Hamilton [1988] proved that when the average scalar curvature is negative, the solution of the normalized Ricci flow with any initial metric converges to a metric with constant negative curvature. In particular, they are linearly stable. On the other hand, in [Delay 2008] we see that the biggest eigenvalue of the Lichnerowicz Laplacian on trace free symmetric 2-tensors is 2. Thus N_+ is nonpositive definite on $(\ker \operatorname{div})_0$, which also implies the linear stability.

(3) For the noncompact case, V. Suneeta [2009] proved certain geometric stability of $\mathbb{H}^n$ using different methods.

Acknowledgement

The author thanks his advisor, Professor Huai-Dong Cao, for encouragement and suggestions.

References

[Buzzanca 1984] C. Buzzanca, "The Lichnerowicz Laplacian on tensors", *Boll. Un. Mat. Ital. B* (6) **3**:2 (1984), 531–541. MR 86c:58132

[Cao and Zhu 2006] H.-D. Cao and X.-P. Zhu, "A complete proof of the Poincaré and geometrization conjectures: application of the Hamilton–Perelman theory of the Ricci flow", *Asian J. Math.* **10**:2 (2006), 165–492. MR 2008d:53090 Zbl 1200.53057

[Cao et al. 2004] H.-D. Cao, R. Hamilton, and T. Ilmanen, "Gaussian densities and stability for some Ricci solitons", preprint, 2004. arXiv math/0404165

[Delay 2002] E. Delay, "Essential spectrum of the Lichnerowicz Laplacian on two tensors on asymptotically hyperbolic manifolds", *J. Geom. Phys.* **43**:1 (2002), 33–44. MR 2003i:58055 Zbl 1146.58301

[Delay 2008] E. Delay, "Spectrum of the Lichnerowicz Laplacian on asymptotically hyperbolic surfaces", *Math. Phys. Anal. Geom.* **11**:3-4 (2008), 365–379. MR 2010e:58031 Zbl 1178.35276

[Feldman et al. 2005] M. Feldman, T. Ilmanen, and L. Ni, "Entropy and reduced distance for Ricci expanders", *J. Geom. Anal.* **15**:1 (2005), 49–62. MR 2006b:53091 Zbl 1071.53040

[Hamilton 1988] R. S. Hamilton, "The Ricci flow on surfaces", pp. 237–262 in *Mathematics and general relativity* (Santa Cruz, CA, 1986), edited by J. A. Isenberg, Contemp. Math. **71**, Amer. Math. Soc., Providence, RI, 1988. MR 89i:53029 Zbl 0663.53031

[Knopf and Young 2009] D. Knopf and A. Young, "Asymptotic stability of the cross curvature flow at a hyperbolic metric", *Proc. Amer. Math. Soc.* **137**:2 (2009), 699–709. MR 2009j:53085 Zbl 1168.53034

[Lee 2006] J. M. Lee, *Fredholm operators and Einstein metrics on conformally compact manifolds*, Mem. Amer. Math. Soc. **183**:864, Amer. Math. Soc., Providence, RI, 2006. MR 2007m:53047 Zbl 1112.53002

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv math.DG/0211159

[Suneeta 2009] V. Suneeta, "Investigating the off-shell stability of anti-de Sitter space in string theory", *Classical Quantum Gravity* **26**:3 (2009), 035023. MR 2010c:83119 Zbl 1159.83023

[Ye 1993] R. Ye, "Ricci flow, Einstein metrics and space forms", *Trans. Amer. Math. Soc.* **338**:2 (1993), 871–896. MR 93j:58029 Zbl 0804.53054

Received January 30, 2010.

MENG ZHU
DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PA 18015
UNITED STATES

mez206@lehigh.edu

CONTENTS

Volume 251, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a 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 251 No. 2 June 2011
