Pacific Journal of Mathematics

Volume 272 No. 2 December 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

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

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

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.257

MARGINALLY TRAPPED SUBMANIFOLDS IN SPACE FORMS WITH ARBITRARY SIGNATURE

HENRI ANCIAUX

Dedicated to the memory of Franki Dillen (1963–2013).

We give explicit representation formulas for marginally trapped submanifolds of codimension two in pseudo-Riemannian spaces with arbitrary signature and constant sectional curvature.

Introduction

Let $(\mathcal{N}, g)$ be a pseudo-Riemannian manifold and $\mathcal{S}$ a submanifold of $(\mathcal{N}, g)$ with nondegenerate induced metric. We shall say that $\mathcal{S}$ is *marginally trapped* if its mean curvature vector is null, that is, $g(\vec{H}, \vec{H})$ vanishes. When $(\mathcal{N}, g)$ is a Lorentzian four-manifold and $\mathcal{S}$ is spacelike, the marginally trapped condition has an interpretation in terms of general relativity: it describes the horizon of a black hole [Penrose 1965; Chruściel et al. 2010]. The equation $g(\vec{H}, \vec{H}) = 0$ is nevertheless interesting in whole generality from the geometric viewpoint, being actually the simplest curvature equation which is purely pseudo-Riemannian: in the Riemannian case this equation implies minimality.

In [Anciaux and Godoy 2012], marginally trapped submanifolds with codimension two have been locally characterized in several simple Lorentzian spaces: the Minkowski space $\mathbb{R}^{n+2}$, the Lorentzian space forms $d\mathbb{S}^{n+2}$ and $\mathrm{Ad}\,\mathbb{S}^{n+2}$, and the Lorentzian products $\mathbb{S}^{n+1} \times \mathbb{R}$ and $\mathbb{H}^{n+1} \times \mathbb{R}$. Little has been done about marginally trapped surfaces in the case of a manifold with a non Lorentzian metric. In [Chen 2009], flat marginally trapped surfaces of $\mathbb{R}^4$ endowed with the neutral metric $dx_1^2 + dx_2^2 - dx_3^2 - dx_4^2$ have been studied, while Lagrangian marginally trapped surfaces of complex space forms of complex dimension two were characterized in [Chen and Dillen 2007]. Recently marginally trapped surfaces of certain spaces of oriented geodesics have been investigated [Georgiou and Guilfoyle 2014].

The purpose of the present paper is to extend the results of [Anciaux and Godoy 2012] to the case of codimension-two submanifolds in constant curvature spaces

Supported by CNPq (PQ 306154/2011-0) and Fapesp (2011/21362-2).
MSC2010: 53A10, 53C42.
Keywords: marginally trapped submanifolds, pseudo-Riemannian manifolds, mean curvature vector.

with arbitrary signature, that is, (i) in the pseudo-Euclidean space $\mathbb{R}^{n+2}_{p+1}$ equipped with the inner product of signature $(p+1, q+1)$, and (ii) in the space form $\mathbb{S}^{n+2}_{p+1}$ of signature $(p+1, q+1)$ and sectional curvature 1 (see next section for more precise definition and notation). As in [Anciaux and Godoy 2012], we rely on the use of the contact structure of the set of null geodesics of the ambient space. The congruence of null lines which are normal to a submanifold of codimension two is a Legendrian submanifold with respect to this contact structure. Conversely, given a null line congruence $\mathcal{L}$ which is Legendrian, there exists an infinite-dimensional family of submanifolds, parametrized by the set of real maps $\tau \in C^2(\mathcal{L})$, such that the congruence is normal to them. In order to obtain our characterization results, we prove that, given a Legendrian, null line congruence $\mathcal{L}$, the submanifold parametrized by τ is marginally trapped if and only if the real map $\tau \in C^2(\mathcal{L})$ is a root of certain polynomial map with coefficients valued in $C^2(\mathcal{L})$.

The paper is organized as follows: Section 1 introduces some notation and gives the precise statements of the results; Section 2 gives a characterization of those submanifolds whose second fundamental tensor is null (Theorem 1), while Section 3 provides a local representation formula which is similar to that of [Anciaux and Godoy 2012] (Theorem 2). In Section 4, an alternative, more global representation formula is given, under certain maximal rank assumption (Theorems 3 and 4). Finally Section 5 attempts to shed light on the ideas in this paper by providing an interpretation of the general construction in terms of contact geometry and explains also the relation between Theorems 2 and 3 in the Lorentzian case.

1. Statement of results

We fix throughout three integers p, q and n such that $p + q = n \geq 1$. We shall denote by $\mathbb{R}^{n+2}_{p+1}$ the $(n+2)$-dimensional real vector space equipped with the inner product of signature $(p+1, q+1)$ given by

$$\langle \cdot , \cdot \rangle = \sum_{i=1}^{p+1} dx_i^2 - \sum_{i=p+2}^{n+2} dx_i^2 .$$

A nonvanishing vector v of $\mathbb{R}^{n+2}_{p+1}$ is said to be *null* if $\langle v, v\rangle = 0$. We furthermore introduce the hyperquadric

$$\mathbb{S}^{n+2}_{p+1} := \{x \in \mathbb{R}^{n+3}_{p+2} \, \langle x, x\rangle = 1\}.$$

The induced metric of $\mathbb{S}^{n+2}_{p+1}$, still denoted by $\langle \cdot , \cdot \rangle$, has signature $(p+1, q+1)$ and constant sectional curvature 1. Conversely it is well known (see [Kriele 1999]) that a simply connected $(n+2)$-dimensional manifold endowed with a pseudo-Riemannian metric with signature $(p+1, q+1)$ and constant sectional curvature is, up to isometries and scaling, $\mathbb{R}^{n+2}_{p+1}$ or $\mathbb{S}^{n+2}_{p+1}$. We shall call these spaces *pseudo-Riemannian space forms.*

We shall be concerned with submanifolds Σ of $\mathbb{R}^{n+2}_{p+1}$ and $\mathbb{S}^{n+2}_{p+1}$ with nondegenerate induced metric g and whose normal bundle $N\Sigma$ (i) is two-dimensional (so that Σ has dimension n), and (ii) has indefinite (Lorentzian) metric (so that the induced metric on Σ has signature (p, q)). We recall that the *second fundamental form* h of Σ is the symmetric tensor $h : T\Sigma \times T\Sigma \to N\Sigma$ defined by $h(X, Y) := (D_X Y)^\perp$, where $(\cdot)^\perp$ denotes the projection onto the normal space $N\Sigma$ and D is the Levi-Civita connection of ambient space. If ν is a normal vector field along Σ, we define the *shape operator* of Σ with respect to ν to be the endomorphism of $T\Sigma$ defined by $A_\nu X = -(D_X \nu)^\top$, where $(\cdot)^\top$ denotes the projection onto $T\Sigma$. The relation $\langle h(X, Y), \nu\rangle = \langle A_\nu X, Y\rangle$ shows that the second fundamental form and the shape operator carry the same information.

The *mean curvature vector* $\vec{H}$ of the immersion is the trace of h with respect to the induced metric of Σ divided by n. Our first result is the characterization of n-dimensional submanifolds of space forms with *null second fundamental form*, that is, such that $h(X, Y)$ is null for all $X, Y \in T\Sigma$:

Theorem 1. *Let ν be a constant, null vector of $\mathbb{R}^{n+2}_{p+1}$ and Σ an n-dimensional submanifold with nondegenerate induced metric which is contained in the hyperplane $\nu^\perp$. Then Σ has null second fundamental form and is therefore marginally trapped. Moreover both the tangent and the normal bundles of Σ are flat.*

Analogously let ν be a constant, null vector of $\mathbb{R}^{n+3}_{p+2}$ and Σ an n-dimensional submanifold of $\mathbb{S}^{n+2}_{p+1}$ with nondegenerate induced metric which is contained in the hypersurface $\nu^\perp \cap \mathbb{S}^{n+2}_{p+1}$. Then Σ has null second fundamental form and is therefore marginally trapped. Moreover Σ has constant scalar curvature and flat normal bundle.

Conversely any submanifold of $\mathbb{R}^{n+2}_{p+1}$ or $\mathbb{S}^{n+2}_{p+1}$ with null second fundamental form is locally described in this way.

Quite surprisingly, the method introduced in [Anciaux and Godoy 2012] in the Lorentzian case can be used here, in the case of marginally trapped submanifolds whose second fundamental form is not null, providing local parametrizations:

Theorem 2. *Let σ be an immersion of class C^4 of an n-dimensional manifold $\mathcal{M}$ into $\mathbb{R}^{n+1}_{p+1}$ (respectively, $\mathbb{S}^{n+1}_{p+1}$) whose induced metric is nondegenerate and has signature (p, q). Denote by ν the Gauss map of σ, which is therefore $\mathbb{S}^n_p$-valued (respectively, $\mathbb{S}^{n+1}_{p+1}$-valued), by $A = -d\nu$ the corresponding shape operator, and by τ_i the roots of the polynomial of degree $n-1$*

$$P(\tau) := \operatorname{tr}(\mathrm{Id} - \tau A)^{-1}.$$

Then the immersions $\varphi_i : \mathcal{M} \to \mathbb{R}^{n+2}_{p+1} = \mathbb{R}^{n+1}_{p+1} \times \mathbb{R}$ (respectively, $\mathbb{S}^{n+2}_{p+1} \subset \mathbb{R}^{n+2}_{p+2} \times \mathbb{R}$)

defined by

$$\varphi_i = (\sigma + \tau_i \nu, \tau_i)$$

are marginally trapped.

Conversely any n-dimensional marginally trapped submanifold of $\mathbb{R}^{n+2}_{p+1}$ (respectively, $\mathbb{S}^{n+1}_{p+1}$) whose second fundamental form is not null is locally congruent to the image of such an immersion.

Remark 1. If the shape operator A of σ is diagonalizable (which is not always the case since the induced metric on σ is not definite) the polynomial P takes the form

$$P(\tau) := \sum_{i=1}^{p} m_i \prod_{j\neq i}^{p} (\kappa_j^{-1} - \tau),$$

where $\kappa_1, \dots, \kappa_p,\ p \geq 2$ are the p distinct, nonvanishing principal curvatures of σ with multiplicity m_i.

In order to state the next theorem, we introduce some more notation: writing $x = (x', x'') \in \mathbb{R}^{n+2} = \mathbb{R}^{p+1} \times \mathbb{R}^{q+1}$, where $x' \in \mathbb{R}^{p+1}$ and $x'' \in \mathbb{R}^{q+1}$, we introduce the conjugation map $\overline{(x', x'')} := (x', -x'')$, as well as the $n \times n$ diagonal matrix $\overline{\mathrm{Id}}_n$ whose (p, q)-block decomposition is $\overline{\mathrm{Id}}_n = \begin{pmatrix} \mathrm{Id}_p & 0 \\ 0 & -\mathrm{Id}_q \end{pmatrix}$.

Since the normal spaces $N\Sigma$ are assumed to be two-dimensional and Lorentzian, the marginally trapped assumption $\langle \vec{H}, \vec{H} \rangle = 0$ is equivalent to the fact that $\vec{H}$ is contained in one of the two null lines of $N\Sigma$. We shall call *mean Gauss map*, and denote by $\nu = (\nu', \nu'')$, the null vector which is collinear to $\vec{H}$ and normalized by the condition $\nu \in \mathbb{S}^p \times \mathbb{S}^q \subset \mathbb{R}^{p+1} \times \mathbb{R}^{q+1}$. The next two theorems give a global description of those marginally trapped submanifolds whose mean Gauss map has maximal rank. We observe that this is a generic property and that it is a stronger assumption than requiring the mean curvature vector $\vec{H}$ to have itself maximal rank.

Theorem 3. *Let Ω be an open subset of the universal covering of $\mathbb{S}^p \times \mathbb{S}^q$ and $\sigma \in C^4(\Omega)$. Denote by τ_i the roots of the polynomial of degree $n - 1$*

$$P(\tau) = \mathrm{tr}\big((\tau\, \mathrm{Id}_n + \sigma\, \overline{\mathrm{Id}}_n + 2\, \mathrm{Hess}(\sigma))^{-1}\big).$$

Then the immersions

$$\varphi_i : \Omega \to \mathbb{R}^{n+2}_{p+1}, \qquad \nu \mapsto \tau_i \nu + \sigma \bar{\nu} + 2\nabla\sigma,$$

are marginally trapped.

Conversely any connected, marginally trapped n-dimensional submanifold of $\mathbb{R}^{n+2}_{p+1}$ whose mean Gauss map ν has maximal rank is the image of such an immersion.

When $n = 2$, the condition of maximal rank on ν is equivalent to the fact that the second fundamental form is not null. Hence Theorems 1 and 3 provide a complete characterization of marginally trapped surfaces of $\mathbb{R}^4$ with arbitrary signature. Since the Minkowski case has already been discussed in [Anciaux and Godoy 2012], we detail the case $(p, q) = (1, 1)$, that is, of a Lorentzian surface in $\mathbb{R}^4_2$. Observe first that $\mathbb{R}^4_2$ is endowed with

(i) a natural pseudo-Kähler structure, with complex structure $J(x_1, x_2, x_3, x_4) := (-x_2, x_1, -x_4, x_3)$ and symplectic form $\omega = \langle J\,\cdot, \cdot\rangle = dx_1 \wedge dx_2 - dx_3 \wedge dx_4$; this corresponds to the identification of $\mathbb{R}^4_2$ with $\mathbb{C}^2$ through the formula $(z_1, z_2) = (x_1 + ix_2, x_3 + ix_4)$;

(ii) a natural para-Kähler structure, with paracomplex[1] structure

$$K(x_1, x_2, x_3, x_4) := (x_3, x_4, x_1, x_2)$$

and symplectic form $\omega' = \langle K\,\cdot, \cdot\rangle = dx_1 \wedge dx_3 + dx_2 \wedge dx_4$; this corresponds to the identification of $\mathbb{R}^4_2$ with $\mathbb{D}^2$, where $\mathbb{D} = \{a + eb \mid (a, b) \in \mathbb{R}^2\}$ is the ring of paracomplex numbers, through the formula $(w_1, w_2) = (x_1 + ex_3, x_2 + ex_4)$;

Corollary 1. *Let Ω be an open subset of $\mathbb{R}^2$ endowed with the Lorentzian metric $du^2 - dv^2$ and $\sigma \in C^4(\Omega)$. Denote by subscript u or v the partial derivative with respect to the corresponding variable. Then the immersion*

$$\varphi : \Omega \to \mathbb{R}^4_2 \simeq \mathbb{C}^2, \quad (u, v) \mapsto \big((\sigma - \sigma_{uu} + \sigma_{vv} + 2i\sigma_u)e^{iu}, (-\sigma - \sigma_{uu} + \sigma_{vv} - 2i\sigma_v)e^{iv}\big),$$

is weakly conformal and its null points are characterized by $\sigma + \sigma_{uu} + \sigma_{vv} = \pm 2\sigma_{uv}$. Moreover, away from its null points, φ is marginally trapped.

Conversely any connected, marginally trapped surface of $\mathbb{R}^4_2$ whose second fundamental form is not null is the image of such an immersion.

In [Chen 2009] and [Chen and Dillen 2007], marginally trapped surfaces of $\mathbb{R}^4_2$ which are in addition, respectively, flat and Lagrangian with respect to ω have been characterized. These additional conditions may be readily seen in terms of the formula given above:

Corollary 2. *The marginally trapped immersion φ of Corollary 1 is in addition*

(i) *flat if and only if $(\partial_{uu} - \partial_{vv})\big((\sigma + \sigma_{uu} + \sigma_{vv})^2 - 4\sigma_{uv}^2\big) = 0$;*

(ii) *Lagrangian with respect to the symplectic form ω if and only if*

$$\sigma_u + \sigma_v + \sigma_{vvv} - \sigma_{uuv} - \sigma_{uvv} + \sigma_{uuu} = 0.$$

Moreover there is no marginally trapped surface which is in addition Lagrangian with respect to the symplectic form ω'.

In the next theorem we give a characterization of marginally trapped submanifolds whose mean Gauss map has maximal rank in $\mathbb{S}^{n+2}_{p+1}$.

[1] We refer the reader to [Alekseevsky et al. 2009] or [Cruceanu et al. 1996] for material about paracomplex geometry, also called split geometry.

Theorem 4. *Let $\sigma : \mathcal{M} \to \mathbb{S}^{p+1} \times \mathbb{S}^q$ be an immersed, oriented hypersurface of class C^4 whose induced metric has signature (p, q). Denote by ν its Gauss map (hence a $\mathbb{S}^{n+2}_{p+1}$-valued map) and by $A = -d\nu$ the corresponding shape operator. Denote by τ_i the roots of the polynomial of degree $n-1$*

$$P(\tau) = \operatorname{tr}(\tau \operatorname{Id} - A)^{-1}.$$

Then the immersions $\varphi_i : \mathcal{M} \to \mathbb{S}^{n+2}_{p+1}$ defined by $\varphi_i := \nu + \tau_i \sigma$ are marginally trapped.

Conversely any connected, marginally trapped n-dimensional submanifold of $\mathbb{S}^{n+2}_{p+1}$ whose mean Gauss map has maximal rank is the image of such an immersion.

Like in the flat case $\mathbb{R}^4_2$, a marginally trapped surface of $\mathbb{S}^4_2$ has either null second fundamental form, or a mean Gauss map with maximal rank. Therefore Theorems 1 and 4 provide a complete characterization in this case. It enjoys, moreover, a more explicit description:

Corollary 3. *Let σ be an immersion of class C^4 of a surface $\mathcal{M}$ into $\mathbb{S}^2 \times \mathbb{S}^1$ with Lorentzian induced metric. Denote by ν the Gauss map of σ (hence a $\mathbb{S}^4_2$-valued map) and by H the (scalar) mean curvature of σ with respect to ν. Then the immersion $\varphi : \mathcal{M} \to \mathbb{S}^4_2$ defined by*

$$\varphi = \nu + H\sigma$$

is marginally trapped.

Conversely any connected marginally trapped surface of $\mathbb{S}^4_2$ whose second fundamental form is not null is the image of such an immersion.

2. Submanifolds with null second fundamental form: proof of Theorem 1

Let Σ be an n-dimensional submanifold of $\mathbb{R}^{n+2}_{p+1}$ such that the induced metric on the normal bundle $N\Sigma$ is Lorentzian. Since the intersection of the light cone of $\mathbb{R}^{n+2}_{p+1}$ with $N\Sigma$ is made of two null lines, there exists a null normal frame, that is, a pair of normal, null vector fields along Σ such that $\langle \nu, \nu \rangle = \langle \xi, \xi \rangle = 0$ and $\langle \nu, \xi \rangle = 2$. So, given a normal vector N, we have

$$N = \tfrac{1}{2}(\langle N, \xi \rangle \nu + \langle N, \nu \rangle \xi).$$

Lemma 1. *The second fundamental form h is collinear to ν (so in particular it is null) if and only if the mean curvature vector $\vec{H}$ is collinear to ν and ν has rank at most* 1.

Proof. We denote by $(e_1, \dots, e_n)$ a local, orthonormal, tangent frame along Σ and we set

$$h^1_{ij} := \langle h(e_i, e_j), \nu \rangle = -\langle d\nu(e_i), e_j \rangle.$$

Then we have, taking into account that $\langle d\nu, \nu\rangle = 0$,

$$d\nu(e_i) = -\sum_{i=1}^{n} h^1_{ij} e_j + \tfrac{1}{2}\langle d\nu(e_i), \xi\rangle \nu.$$

Assume first that h is collinear to ν. Then clearly its trace $n\vec{H}$ is collinear to ν as well. Moreover all the coefficients h^1_{ij} vanish, so by the equation above, for all i, $1 \le i \le n$, the vector $d\nu(e_i)$ is collinear to ν, and hence $d\nu$ has rank at most 1.

Conversely, if $d\nu$ has rank 1, then $d\nu(e_i)$ and $d\nu(e_j)$ are proportional for any pair (i, j), $i \neq j$. Taking into account the symmetry of the tensor h^1_{ij}, an elementary calculation implies that there exist $n+1$ real constants $c, \lambda_1, \lambda_2, \dots, \lambda_n$ such that $h^1_{ij} = c\lambda_i\lambda_j$. If in addition Σ is marginally trapped, that is, $\langle n\vec{H}, \nu\rangle = \operatorname{tr}\,[h^1_{ij}]_{1\le i,j\le n} = c\sum_{i=1}^{n}\lambda_i^2 = 0$, then either $c = 0$ or $(\lambda_1, \dots, \lambda_n) = (0, \dots, 0)$ and in both cases the whole tensor h^1_{ij} vanishes, that is, h is collinear to ν. □

We come back to the proof of Theorem 1, observing that under the assumption of the lemma above, $d\nu$ is collinear to ν. This implies the existence of a map $\lambda \in C^1(\Sigma)$ such that $\nu = e^{\lambda}\nu_0$, where ν_0 is a constant, null vector of $\mathbb{R}^{n+2}_{p+1}$ or $\mathbb{R}^{n+3}_{p+2}$. We conclude that $\Sigma \subset \nu_0^{\perp}$.

We now write the Gauss and the Ricci equations in the flat case:

$$\langle R(X,Y)Z, W\rangle + \langle h(X,Z), h(Y,W)\rangle - \langle h(X,W), h(Y,Z)\rangle = 0,$$
$$\langle R^{\perp}(X,Y)\nu, \xi\rangle - \langle [A_\nu, A_\xi]X, Y\rangle = 0,$$

If h is collinear to ν, both terms $\langle h(X,Z), h(Y,W)\rangle$ and $\langle h(X,W), h(Y,Z)\rangle$ vanish, hence the curvature of the tangent bundle vanishes. Moreover, if h is collinear to ν, then A_ν vanishes as well and the normal bundle is flat.

In the case of $\mathbb{S}^{n+2}_{p+1}$, the Gauss and the Ricci equations become

$$\begin{aligned}\langle R(X,Y)Z, W\rangle + \langle h(X,Z), h(Y,W)\rangle - \langle h(X,W), h(Y,Z)\rangle \\ = \langle X,Z\rangle\langle Y,W\rangle - \langle X,W\rangle\langle Y,Z\rangle,\end{aligned}$$
$$\langle R^{\perp}(X,Y)\nu, \xi\rangle - \langle [A_\nu, A_\xi]X, Y\rangle = 0.$$

Again, if h is collinear to ν, the terms $\langle h(X,Z), h(Y,W)\rangle$ and $\langle h(X,W), h(Y,Z)\rangle$ vanish. It follows that the scalar curvature of the induced metric is constant and equal to 1. Analogously, the fact that h is collinear to ν implies the vanishing of A_ν and therefore the flatness of the normal bundle.

3. Parametrizations by hypersurfaces: proof of Theorem 2

3.1. ***The flat case.*** Let $\varphi = (\psi, \tau)$ be an immersion of an n-dimensional manifold $\mathcal{M}$ into $\mathbb{R}^{n+2}_{p+1}$ whose induced metric $\tilde{g} := \varphi^*\langle\cdot,\cdot\rangle$ has signature (p, q). In particular the induced metric on the normal space of $\bar{\varphi}$ is Lorentzian. Let $\tilde{\nu}$ be one of the two normalized, null normal fields along φ. Since the discussion is local, there is no loss of generality in assuming that, modulo congruence, its last component ν_{n+3} does not vanish, so that we may normalize $\tilde{\nu} = (\nu, 1)$.

Lemma 2. *We set $\sigma := \psi - \tau\nu$. Then the map $(\sigma, \nu) : \mathcal{M} \to \mathbb{R}^{n+1}_{p+1} \times \mathbb{S}^n_p$ is an immersion.*

Proof. Suppose (φ, ν) is not an immersion, so that there exists a nonvanishing vector $v \in T\mathcal{M}$ such that $(d\varphi(v), d\nu(v)) = (0, 0)$. Since $d\psi = d\sigma + \tau d\nu + d\tau\nu$, it follows that

$$d\varphi(v) = (d\psi(v), d\tau(v)) = (d\tau(v)\nu, d\tau(v)) = d\tau(v)\tilde{\nu},$$

which is a normal to φ. However the immersion φ is pseudo-Riemannian and therefore a vector cannot be tangent and normal at the same time, so we get the required contradiction. □

Lemma 3. $$\langle d\sigma, \nu\rangle = 0.$$

Proof. Using again that $d\psi = d\sigma + \tau d\nu + d\tau\nu$ and observing that $\langle \nu, d\nu\rangle = 0$, we have

$$0 = \langle d\varphi, \tilde{\nu}\rangle = \langle (d\psi, d\tau), (\nu, 1)\rangle = \langle d\psi, \nu\rangle - d\tau = \langle d\sigma, \nu\rangle. \qquad \square$$

Lemma 4. *Given $\epsilon > 0$, there exists $t_0 \in (-\epsilon, \epsilon)$ such that $\sigma + t_0\nu$ is an immersion, and ν is its Gauss map.*

Proof. This follows from the fact that the set $\{t \in \mathbb{R} \mid \sigma + t\nu \text{ is not an immersion}\}$ contains at most n elements. To see this, observe that given a pair of distinct real numbers (t, t'), we have

$$\operatorname{Ker}(d\sigma + t d\nu) \cap \operatorname{Ker}(d\sigma + t' d\nu) = \{0\}$$

(otherwise we would have a contradiction with the fact that (σ, ν) is an immersion). Hence there cannot be more than n distinct values t such that $\operatorname{Ker}(d\sigma + t d\nu) \neq \{0\}$. The fact that ν is the Gauss map of $\sigma + t_0\nu$ comes from Lemma 3:

$$\langle d(\sigma + t_0\nu), \nu\rangle = \langle d\sigma, \nu\rangle + t_0\langle d\nu, \nu\rangle = 0. \qquad \square$$

Lemma 4 shows that there is no loss of generality in assuming that σ is an immersion: if it is not the case, we may translate the immersion φ along the vertical direction, setting $\varphi_{t_0} := \varphi - (0, t_0)$. Of course φ is marginally trapped if and only if φ_{t_0} is so, and moreover the vector field $\tilde{\nu}$ is still normal to φ_{t_0}. Finally observe that the map $\sigma_{t_0} : \mathcal{M} \to \mathbb{R}^{n+1}_{p+1}$ associated to φ_{t_0} is

$$\sigma_{t_0} = \psi - (\tau - t_0)\nu = \psi - \tau\nu + t_0\nu = \sigma + t_0\nu,$$

hence an immersion.

We now describe the first fundamental form of φ and its second fundamental form with respect to $\tilde{\nu}$, both in terms of the geometry of the immersion σ:

Lemma 5. *Denote by $g := \sigma^*\langle \cdot , \cdot \rangle$ the metric induced on $\mathcal{M}$ by σ and A the shape operator associated to ν.*

Then the metric $\tilde{g} := \varphi^*\langle \cdot , \cdot \rangle$ induced on $\mathcal{M}$ by φ is given by the formula

$$\tilde{g} = g(\cdot , \cdot) - 2\tau g(A \cdot , \cdot) + \tau^2 g(A \cdot , A \cdot).$$

In particular, the nondegeneracy assumption on $\tilde{g}$ implies that τ^{-1} is not equal to any principal curvature of φ. Moreover the second fundamental form of φ with respect to $\tilde{\nu}$ is given by

$$\tilde{h}_{\tilde{\nu}} := \langle \tilde{h}(\cdot , \cdot), \tilde{\nu} \rangle = g(\cdot , A \cdot) - \tau g(A \cdot , A \cdot).$$

Proof. Since $\langle d\sigma, \nu\rangle = \langle d\nu, \nu\rangle = 0$, given $v_1, v_2 \in T\mathcal{M}$, we have

$$\begin{aligned}
\tilde{g}(v_1, v_2) &= \langle d\varphi(v_1), d\varphi(v_2)\rangle \\
&= \langle d\sigma(v_1), d\sigma(v_2)\rangle + \tau\langle d\sigma(v_1), d\nu(v_2)\rangle + \tau\langle d\nu(v_1), d\sigma(v_2)\rangle \\
&\qquad + \tau^2\langle d\nu(v_1), d\nu(v_2)\rangle + d\tau(v_1)d\tau(v_2)\langle \nu, \nu\rangle - d\tau(v_1)d\tau(v_2) \\
&= g(v_1, v_2) - \tau(g(v_1, Av_2) + g(Av_1, v_2)) + \tau^2 g(Av_1, Av_2) \\
&= g(v_1, v_2) - 2\tau g(Av_1, v_2) + \tau^2 g(Av_1, Av_2).
\end{aligned}$$

We calculate the second fundamental form of φ with respect to $\tilde{\nu} = (\nu, 1)$:

$$\begin{aligned}
\tilde{h}_{\tilde{\nu}} &= -\langle d\varphi, d\tilde{\nu}\rangle \\
&= -\langle d\sigma + \tau d\nu + d\tau\nu, d\nu\rangle = -\langle d\sigma, d\nu\rangle - \tau\langle d\nu, d\nu\rangle \\
&= g(\cdot , A \cdot) - \tau g(A \cdot , A \cdot).
\end{aligned}$$ □

The proof of Theorem 2 follows easily: denoting by $\tilde{A}_{\tilde{\nu}}$ the shape operator of φ with respect to $\tilde{\nu}$, we have from Lemma 5

$$g(\tilde{A}_{\tilde{\nu}}(\mathrm{Id} - \tau A)\cdot , (\mathrm{Id} - \tau A)\cdot) = g(\cdot , (\mathrm{Id} - \tau A)\cdot).$$

It follows that

$$\tilde{A}_{\tilde{\nu}} := (\mathrm{Id} - \tau A)^{-1}$$

and that $\vec{H}$ is collinear to $\tilde{\nu}$ if and only if $\tilde{A}_{\tilde{\nu}}$ is trace-free, that is, τ is the root of the polynomial $P(\tau) = \mathrm{tr}(\mathrm{Id} - \tau A)^{-1}$.

Remark 2. If φ is minimal, $\tau = 0$ is a root of $P(\tau)$. The corresponding immersion $\varphi = (\sigma, 0)$ is not only marginally trapped but minimal.

3.2. *The $\mathbb{S}^{n+2}_{p+1}$ case.* Let $\varphi = (\psi, \tau) : \mathcal{M} \to \mathbb{S}^{n+2}_{p+1}$ an immersion such that the induced metric $\tilde{g} := \varphi^*\langle\cdot, \cdot\rangle$ has signature (p, q). Let $\tilde{\nu}$ be one of the two normalized, null normal fields along φ. Since the discussion is local, there is no loss of generality to assume that, modulo congruence, its last component ν_{n+3} does not vanish, so that we may normalize $\tilde{\nu} = (\nu, 1)$.

We define the *null projection* of φ to be $\sigma := \psi - \tau\nu$. The fact that $(\nu, 1) \in T_\varphi \mathbb{S}^{n+2}_{p+1}$, that is, $0 = \langle(\psi, \tau), (\nu, 1)\rangle = \langle\psi, \nu\rangle - \tau$, implies that $\langle\psi, \nu\rangle = \tau$. Hence

$$\begin{aligned}\langle\sigma, \sigma\rangle &= \langle\psi, \psi\rangle - 2\tau\langle\psi, \nu\rangle + \tau^2\langle\nu, \nu\rangle \\ &= \langle\psi, \psi\rangle - \tau^2 \\ &= \langle\varphi, \varphi\rangle \\ &= 1,\end{aligned}$$

which shows that σ is $\mathbb{S}^{n+1}_p$-valued. The proofs of the next two lemmas are omitted, since they are similar to the flat case:

Lemma 6. *The map $(\sigma, \nu) : \mathcal{M} \to \mathbb{S}^{n+1}_{p+1} \times \mathbb{S}^{n+1}_{p+1}$ is an immersion.*

Lemma 7.
$$\langle\sigma, \nu\rangle = 0 \quad \text{and} \quad \langle d\sigma, \nu\rangle = 0.$$

Unlike the flat case, there is no vertical translation in $\mathbb{S}^{n+2}_{p+1}$. We may however, up to an arbitrarily small perturbation, assume that σ is an immersion.

Lemma 8. *Given $\epsilon > 0$, there exists $\alpha \in (-\epsilon, \epsilon)$ and a hyperbolic rotation R^α of angle α such that the null projection σ^α of $\varphi^\alpha := R^\alpha\varphi$ is an immersion.*

Proof. Set

$$R^\alpha = \begin{pmatrix} \cosh\alpha & & \sinh\alpha \\ & \mathrm{Id} & \\ \sinh\alpha & & \cosh\alpha \end{pmatrix} \in \mathrm{SO}(p+2, q+1)$$

and $\varphi^\alpha := R^\alpha\varphi$, $\tilde{\nu}^\alpha := R^\alpha\tilde{\nu}$. Observe that $\tilde{\nu}^\alpha := (\nu^\alpha, \nu^\alpha_{n+3})$ is no longer normalized a priori, since its last component ν^α_{n+3} is equal to $\cosh(\alpha) + \sinh(\alpha)\nu_1$, where ν_1 is the first component of the vector $\tilde{\nu}$.

Nevertheless the null geodesic passing through the point φ^α and directed by the vector $\tilde{\nu}^\alpha$ crosses the slice $d\mathbb{S}^{n+2}_{p+2} \cap \{x_{n+3} = 0\}$ at the point

$$(\sigma^\alpha, 0) := \left(\psi^\alpha - \frac{\tau^\alpha}{\nu^\alpha_{n+3}}\nu^\alpha, 0\right).$$

Clearly σ^α is an immersion if and only if

$$R^{-\alpha}\sigma^\alpha = \psi - \tau^\alpha\nu/\nu^\alpha_{n+3} = \varphi + (\tau - \tau^\alpha/\nu^\alpha_{n+3})\nu$$

is so. Observe that

$$\tau - \frac{\tau^\alpha}{\nu^\alpha_{n+3}} = \tau - \frac{\cosh(\alpha)\tau + \sinh(\alpha)\psi_1}{\cosh(\alpha) + \sinh(\alpha)\nu_1}$$
$$= \tanh(\alpha)(\psi_1 - \tau\nu_1) + o(\alpha)$$
$$= \sigma_1\alpha + o(\alpha).$$

Now assume that $R^{-\alpha}\sigma^\alpha$ is not an immersion for $\alpha \in (-\epsilon, \epsilon)$. Hence there exists a one-parameter family of unit tangent vectors v^α such that

$$0 = d(R^{-\alpha}\sigma^\alpha)(v^\alpha) = d\sigma(v^\alpha) + (d\sigma_1(v^\alpha)\nu + \sigma_1 d\nu(v^\alpha))\alpha + o(\alpha)$$

for all $\alpha \in (-\epsilon, \epsilon)$. Thus

$$\begin{cases} d\sigma(v^\alpha) = 0, \\ d\sigma_1(v^\alpha)\nu + \sigma_1 d\nu(v^\alpha) = \sigma_1 d\nu(v^\alpha) = 0. \end{cases}$$

By Lemma 6, $d\nu(v^\alpha)$ and $d\sigma(v^\alpha)$ cannot vanish simultaneously, therefore σ_1 vanishes. Repeating the argument with suitable rotations yields that all the other coordinates of σ vanish, a contradiction since $\sigma \in \mathbb{S}^{n+1}_{p+1}$. □

By the previous lemma we may assume that σ is an immersion. The remainder of the proof follows the lines of that of the flat case; in particular, Lemma 5 still holds.

4. Parametrization by the mean Gauss map

4.1. ***The flat case: proof of Theorem 3.*** In this section Σ denotes an n-dimensional submanifold of $\mathbb{R}^{n+2}_{p+1}$ whose induced metric has signature (p, q) and such that the normalized vector $\nu \in \mathbb{S}^p \times \mathbb{S}^q \subset \mathbb{R}^{n+2}$ has rank n. We may therefore parametrize Σ locally by ν, that is, by a map $\varphi : \Omega \to \mathbb{R}^{n+2}_{p+1}$, where Ω is an open subset of the universal covering of $\mathbb{S}^p \times \mathbb{S}^q$. We set $\sigma(\nu) := \frac{1}{2}\langle\varphi(\nu), \nu\rangle$ and $\tau(\nu) := \frac{1}{2}\langle\varphi(\nu), \bar{\nu}\rangle$.[2]

Lemma 9. *We have*

$$\varphi = \tau\nu + \sigma\bar{\nu} + 2\nabla\sigma,$$

where ∇ is the gradient with respect to the induced metric on $\mathbb{S}^p \times \mathbb{S}^q$ (that is, $\nabla\sigma = (\nabla'\sigma, -\nabla''\sigma)$, where ∇' and ∇'' are respectively the gradients on $\mathbb{S}^p$ and $\mathbb{S}^q$).

Proof. Since ν and $\bar{\nu}$ are null and $\langle\nu, \bar{\nu}\rangle = 2$, we clearly have $\varphi = \tau\nu + \sigma\bar{\nu} + V$, where $V \in T_\nu(\mathbb{S}^p \times \mathbb{S}^q) = T_{\bar{\nu}}(\mathbb{S}^p \times \mathbb{S}^q) = T_{\nu'}\mathbb{S}^p \times T_{\nu''}\mathbb{S}^q$. In order to determine V, we use the assumption $\langle d\varphi, \nu\rangle = 0$. Taking into account that

$$d\varphi = d\tau\,\nu + \tau\,d\nu + d\sigma\,\bar{\nu} + \sigma\,d\bar{\nu} + dV$$

[2]The pair (σ, τ) may be regarded as a generalization of the *support function* of a hypersurface.

and that $\langle \nu, \nu\rangle$, $\langle d\nu, \nu\rangle$ and $\langle d\bar{\nu}, \nu\rangle$ vanish, we have

$$\langle d\varphi, \nu\rangle = d\sigma\,\langle\bar{\nu}, \nu\rangle + \langle dV, \nu\rangle = 2d\sigma + \langle dV, \nu\rangle.$$

On the other hand, from $0 = d(\langle V, \nu\rangle) = \langle dV, \nu\rangle + \langle V, d\nu\rangle$, we conclude, observing that $d\nu = \mathrm{Id}$,

$$\langle V, W\rangle = \langle V, d\nu(W)\rangle = -\langle dV(W), \nu\rangle = 2d\sigma(W) \quad \text{for all } W \in T_\nu(\mathbb{S}^p \times \mathbb{S}^q),$$

which, by the very definition of the gradient, proves that $V = 2\nabla\sigma$. □

We now complete the proof of Theorem 3: define the endomorphism A_ν on $T_\nu(\mathbb{S}^p \times \mathbb{S}^q)$ by

$$\langle A_\nu\,\cdot\,, \cdot\,\rangle = h_\nu.$$

Hence, using that $d\nu$ is the identity map of $T_\nu(\mathbb{S}^p \times \mathbb{S}^q)$, we have

$$\langle d\varphi \circ A_\nu\,\cdot\,, d\varphi\,\cdot\,\rangle = -\langle d\nu\,\cdot\,, d\varphi\,\cdot\,\rangle = -\langle \Pi\,\cdot\,, d\varphi\,\cdot\,\rangle,$$

where Π is the restriction to $T_\nu(\mathbb{S}^p \times \mathbb{S}^q)$ of the normal projection $\mathbb{R}^{n+2}_{p+1} \to T_{\varphi(\nu)}\mathcal{S}$. It follows that

$$d\varphi \circ A_\nu = -\Pi,$$

and therefore

$$A_\nu^{-1} = -\Pi^{-1} \circ d\varphi$$

(the maximal rank assumption on ν implies that Π is one-to-one). In order to calculate the trace of A_ν we introduce an orthonormal basis $(e_1, \dots, e_n)$ of $T_\nu(\mathbb{S}^p \times \mathbb{S}^q)$, such that $\langle e_i, e_i\rangle = 1$ if $1 \le i \le p$ and $\langle e_i, e_i\rangle = -1$ if $p+1 \le i \le n$. We define the coefficients a_{ij} by

$$d\varphi(e_i) = \sum_{j=1}^{n} a_{ij}\,\Pi e_j.$$

Clearly

$$A_\nu^{-1} = [a_{ij}]_{1 \le i, j \le n}.$$

To determine the coefficients a_{ij} explicitly, we calculate

$$d\varphi = d\tau\;\nu + \tau\;d\nu + d\sigma\;\bar{\nu} + \sigma\;d\bar{\nu} + 2d\nabla\sigma. \tag{1}$$

Then we introduce a null, normal vector field ξ along $\mathcal{S}$ such that (ν, ξ) is a null frame of $N\Sigma = T\Sigma^\perp$, which is in addition normalized, that is, $\langle \nu, \xi\rangle = 2$. Then the projection of a vector V of $\mathbb{R}^{n+2}_{p+1}$ onto $N\Sigma$ is given by the formula

$$\tfrac{1}{2}(\langle V, \xi\rangle\nu + \langle V, \nu\rangle\xi).$$

It follows that

$$\Pi V = V - \tfrac{1}{2}(\langle V, \xi\rangle\nu + \langle V, \nu\rangle\xi). \tag{2}$$

For $1 \leq i \leq p$, using (1) and observing that $d\nu(e_i) = d\bar{\nu}(e_i) = e_i$, we have

$$d\varphi(e_i) = (\tau + \sigma)e_i + d\tau(e_i)\nu + d\sigma(e_i)\bar{\nu} + 2d(\nabla\sigma)(e_i)$$

Using (2) and the fact that $\langle d(\nabla\sigma)(e_i), e_j\rangle = \mathrm{Hess}(\sigma)(e_i, e_j)$, we conclude that, for $1 \leq i \leq p$,

$$a_{ij} = \delta_{ij}(\tau + \sigma) + 2\,\mathrm{Hess}(\sigma)(e_i, e_j).$$

Analogously we get, if $p + 1 \leq i \leq n$,

$$a_{ij} = \delta_{ij}(\tau - \sigma) + 2\,\mathrm{Hess}(\sigma)(e_i, e_j).$$

The conclusion of the proof of Theorem 3 follows easily.

4.2. *The case $(p, q) = (1, 1)$: proof of Corollaries 1 and 2.* We use the natural identification $\mathbb{R}^4_2 \simeq \mathbb{C}^2$ and denote by (u, v) the natural coordinates on $\mathbb{S}^1 \times \mathbb{S}^1$, so that $\nu := (e^{iu}, e^{iv})$. In particular, the metric on $\mathbb{S}^1 \times \mathbb{S}^1$ is $du^2 - dv^2$. Hence

$$A_\nu = -\begin{pmatrix} \tau + \sigma + 2\sigma_{uu} & 2\sigma_{uv} \\ -2\sigma_{uv} & \tau - \sigma - 2\sigma_{vv} \end{pmatrix}^{-1},$$

whose trace is $2/\det A_\nu(\tau + \sigma_{uu} - \sigma_{vv})$. Hence φ is marginally trapped if and only if $\tau = \sigma_{vv} - \sigma_{uu}$.

We now study the induced metric $\varphi^*\langle\cdot,\cdot\rangle$. Since

$$\varphi = \tau(e^{iu}, e^{iv}) + \sigma(e^{iu}, -e^{iv}) + 2(i\sigma_u e^{iu}, -i\sigma_v e^{iv}),$$

we have

$$\begin{aligned}
\varphi_u &= \big(((\tau - \sigma)_u + i(2\sigma_{uu} + \tau + \sigma))e^{iu},\ ((\tau - \sigma)_u - 2i\sigma_{uv})e^{iv}\big),\\
\varphi_v &= \big(((\tau + \sigma)_v + 2i\sigma_{uv})e^{iu},\ ((\tau + \sigma)_v + i(-2\sigma_{vv} + \tau - \sigma))e^{iv}\big).
\end{aligned}$$

By a straightforward calculation the coefficients of the first fundamental form $\varphi^*\langle\cdot,\cdot\rangle$ are

$$\begin{aligned}
E &:= (2\sigma_{uu} + \tau + \sigma)^2 - 4\sigma_{uv}^2,\\
F &:= 4\sigma_{uv}(\sigma_{uu} - \sigma_{vv} + 2\tau),\\
G &:= -(2\sigma_{vv} - \tau + \sigma)^2 + 4\sigma_{uv}^2.
\end{aligned}$$

The marginally trapped assumption $\tau = \sigma_{vv} - \sigma_{uu}$ implies

$$E = -G = (2\sigma_{uu} + \tau + \sigma)^2 - 4\sigma_{uv}^2 = (\sigma + \sigma_{uu} + \sigma_{vv})^2 - 4\sigma_{uv}^2$$

and the vanishing of F, so that φ is weakly conformal (and conformal whenever E does not vanish).

It is well known that the induced metric of a surface with isothermic coordinates is flat if and only if its conformal factor is harmonic. Here we are dealing with the

Lorentzian metric $du^2 - dv^2$, whose Laplacian operator is $\partial_{uu} - \partial_{vv}$. Hence the induced metric is flat if and only if

$$(\partial_{uu} - \partial_{vv})E = (\partial_{uu} - \partial_{vv})\big((\sigma + \sigma_{uu} + \sigma_{vv})^2 - 4\sigma_{uv}^2\big).$$

Marginally trapped Lagrangian surfaces. We recall that $J(z_1, z_2) = (iz_1, iz_2)$, so

$$J\varphi_u = \big(-(2\sigma_{uu} + \tau + \sigma + i(\tau - \sigma)_u)e^{iu}, (2\sigma_{uv} + i(\tau - \sigma)_u)e^{iv}\big).$$

Hence, using the usual formula $\omega = \langle J\,\cdot, \cdot\rangle$,

$$\begin{aligned}
\omega(\varphi_u, \varphi_v) &= \langle J\varphi_u, \varphi_v\rangle \\
&= -(2\sigma_{uu} + \tau + \sigma)(\tau + \sigma)_v + 2\sigma_{uv}(\tau - f)_u - 2\sigma_{uv}(\tau + \sigma)_v \\
&\qquad - (\tau - \sigma)_u(-2\sigma_{vv} + \tau - \sigma) \\
&= -(\tau + \sigma)_v(\sigma + \sigma_{uu} + \sigma_{vv} + 2\sigma_{uv}) + (\tau - \sigma)_u(\sigma + \sigma_{uu} + \sigma_{vv} + 2\sigma_{uv}) \\
&= (\sigma + \sigma_{uu} + \sigma_{vv} + 2\sigma_{uv})(-\sigma_v - \sigma_u - \sigma_{vvv} + \sigma_{uuv} + \sigma_{uvv} - \sigma_{uuu}).
\end{aligned}$$

The first factor does not vanish except at degenerate points, so φ is Lagrangian with respect to ω if and only if $\sigma_v + \sigma_u + \sigma_{vvv} - \sigma_{uuv} - \sigma_{uvv} + \sigma_{uuu} = 0$.

Recalling that the paracomplex structure is given by

$$K(x_1, x_2, x_3, x_4) := (x_3, x_4, x_1, x_2),$$

we have

$$K\varphi_u = \big(((\tau - \sigma)_u - 2i\sigma_{uv})e^{iv}, ((\tau - \sigma)_u + i(2\sigma_{uu} + \tau + \sigma))e^{iu}\big),$$

and so

$$\begin{aligned}
\omega'(\varphi_u, \varphi_v) &= \langle K\varphi_u, \varphi_v\rangle \\
&= \cos(u - v)\big(-4\sigma_{uv}^2 - (\tau - \sigma - 2\sigma_{vv})(2\sigma_{uu} + \tau + \sigma)\big) \\
&= \cos(u - v)\big((-4\sigma_{uv}^2 + (\sigma + \sigma_{uu} + \sigma_{vv})\big)^2 \\
&= \cos(u - v)E.
\end{aligned}$$

Hence φ is Lagrangian with respect to ω' if and only if the induced metric is totally null, which is incompatible with the marginally trapped assumption.

4.3. ***The case of $\mathbb{S}^{n+2}_{p+1}$: proof of Theorem 4.*** Let $\varphi : \mathcal{M} \to \mathbb{S}^{n+2}_{p+1}$ be an immersed submanifold of codimension two of $\mathbb{S}^{n+2}_{p+1}$. Let σ be a normal, null vector field along φ which is normalized in such way that $\sigma \in \mathbb{S}^{p+1} \times \mathbb{S}^q$. We moreover assume that σ has maximal rank, that is, $\sigma : \mathcal{M} \to \mathbb{S}^{p+1} \times \mathbb{S}^q$ is an immersed hypersurface.

Lemma 10 [Godoy and Salvai 2013]. *There exists a unique pair (ν, τ), where $\nu : \mathcal{M} \to \mathbb{S}^{n+2}_{p+1}$ and $\tau \in C^2(\mathcal{M})$ are such that*

$$\varphi = \nu + \tau\sigma.$$

Moreover the map $\nu : \mathcal{M} \to \mathbb{S}^{n+2}_{p+1}$ *is the Gauss map of* σ; *that is*, $\nu \in T_\sigma(\mathbb{S}^{p+1} \times \mathbb{S}^q)$ *and* $\langle d\sigma, \nu\rangle = 0$.

Proof. For an arbitrary $\tau \in C^2(\mathcal{M})$, we have $\nu := \varphi - \tau\sigma \in \mathbb{S}^{n+2}_{p+1}$. Hence we shall determine $\tau \in C^2(\mathcal{M})$ by the condition $\nu \in T_\nu(\mathbb{S}^{p+1} \times \mathbb{S}^q)$. Recalling the decomposition $\mathbb{R}^{n+3} = \mathbb{R}^{p+2} \times \mathbb{R}^{q+1}$, and writing $\nu = (\nu', \nu'')$, $\sigma = (\sigma', \sigma'')$ accordingly, this condition amounts to $\langle \nu', \sigma'\rangle_{p+2} = 0$ and $\langle \nu'', \sigma''\rangle_{q+1} = 0$, where $\langle \cdot\,, \cdot\rangle_{p+2}$ and $\langle \cdot\,, \cdot\rangle_{q+1}$ denote the Euclidean inner products of $\mathbb{R}^{p+2}$ and $\mathbb{R}^{q+1}$, respectively. These two equations yield $\tau = \langle \nu', \sigma'\rangle_{p+1}$ and $\tau = \langle \nu'', \sigma''\rangle_{q+1}$, which are actually two equivalent requirements since $\langle \varphi, \sigma\rangle = \langle \varphi', \sigma'\rangle_{p+2} - \langle \varphi'', \sigma''\rangle_{q+1}$ vanishes. Therefore τ is uniquely determined by the condition $\nu \in T_\sigma(\mathbb{S}^{p+1} \times \mathbb{S}^q) = T_{\sigma'}\mathbb{S}^{p+1} \times \mathbb{T}_{\sigma''}\mathbb{S}^q$.

It remains to check that ν is the Gauss map of σ. For this purpose we differentiate $\varphi = \nu + \tau\sigma$ and remember that σ is normal to φ, so that

$$0 = \langle d\varphi, \sigma\rangle = \langle d\nu, \sigma\rangle + d\tau\langle \sigma, \sigma\rangle + \tau\langle d\sigma, \sigma\rangle = \langle d\nu, \sigma\rangle.$$

Hence $\langle d\nu, \sigma\rangle$ vanishes. Since $0 = d(\langle \nu, \sigma\rangle) = \langle d\nu, \sigma\rangle + \langle \nu, d\sigma\rangle$, we deduce that $\langle d\sigma, \nu\rangle$ vanishes as well. □

Observe that the lemma above implies furthermore that the induced metric $g := \sigma^*\langle \cdot\,, \cdot\rangle$ is nondegenerate, since $\sigma(\mathcal{M})$ is a hypersurface and admits a unit normal vector field.

Lemma 11. *Denote by* $g := \sigma^*\langle \cdot\,, \cdot\rangle$ *the metric induced on* $\mathcal{M}$ *by* σ, *and by* A *the shape operator associated to* ν, *so* $A(v) := -d\nu(v)$ *for all* $v \in T\mathcal{M}$. *Then the metric* $\tilde{g} := \varphi^*\langle \cdot\,, \cdot\rangle$ *induced on* $\mathcal{M}$ *by* φ *is given by the formula*

$$\tilde{g} = \tau^2 g(\cdot\,, \cdot) - 2\tau g(A\,\cdot\,, \cdot) + g(A\,\cdot\,, A\,\cdot).$$

In particular, the nondegeneracy assumption on $\tilde{g}$ *implies that* τ *is not equal to any principal curvature of* ν. *Moreover the second fundamental form of* φ *with respect to* σ *is given by*

$$h_\sigma := \langle h(\cdot\,, \cdot), \sigma\rangle = g(A\,\cdot\,, \cdot) - \tau g(\cdot\,, \cdot).$$

Proof. Taking into account that $d\varphi = d\nu + d\tau\,\sigma + \tau d\sigma$, we have

$$\begin{aligned}\tilde{g} = \langle d\varphi, d\varphi\rangle &= \langle d\nu, d\nu\rangle + 2\tau\langle d\nu, d\sigma\rangle + \tau^2\langle d\sigma, d\sigma\rangle \\ &= g(A\,\cdot\,, A\,\cdot) - 2\tau g(A\,\cdot\,, \cdot) + \tau^2 g(\cdot\,, \cdot)\end{aligned}$$

and

$$h_\sigma = -\langle d\varphi, d\sigma\rangle = -\langle d\nu, d\sigma\rangle - \tau\langle d\sigma, d\sigma\rangle = g(A\,\cdot\,, \cdot) - \tau g(\cdot\,, \cdot). \qquad \square$$

The proof of Theorem 4 is now straightforward: if φ is marginally trapped, we may assume without loss of generality that its mean curvature vector $\vec{H}$ is collinear to σ. By the maximal rank assumption on σ we may use Lemmas 10 and 11.

Denote by $\tilde{A}_\sigma$ the shape operator of φ with respect to σ. Then, from Lemma 11 above, we have

$$g(\tilde{A}_\sigma(\tau\,\mathrm{Id}-A)\,\cdot\,,(\tau\,\mathrm{Id}-A)\,\cdot\,)=g(\,\cdot\,,(\tau\,\mathrm{Id}-A)\,\cdot\,).$$

It follows that $\tilde{A}_\sigma := (\tau\,\mathrm{Id}-A)^{-1}$ and that $\vec{H}$ is collinear to ν if and only if A_ν is trace-free, that is, τ is the root of the polynomial $P(\tau)=\mathrm{tr}(\tau\,\mathrm{Id}-A)^{-1}$.

4.4. ***The case $(p,q)=(1,1)$: proof of Corollary 3.*** It is straightforward that if M is a 2×2 matrix, then $\mathrm{tr}\,M^{-1}=(\det M)^{-1}\,\mathrm{tr}\,M$. Hence $\mathrm{tr}(\tau\,\mathrm{Id}-A)^{-1}$ vanishes if and only if $\mathrm{tr}(\tau\,\mathrm{Id}-A)=2\tau-\mathrm{tr}\,A$ does. Hence φ is marginally trapped if and only if $\tau=\mathrm{tr}\,A/2:=H$, the (scalar) mean curvature of the immersion σ. This proves Corollary 3.

5. Further remarks

5.1. ***Interpretation of the result in terms of contact geometry.*** The constructions in the previous sections come from the natural contact structure enjoyed by the spaces of null geodesics of the ambient spaces and from the fact that the set of null geodesics which are normal to a submanifold of codimension two is Legendrian with respect to this contact structure.

The proof of Theorem 2 is based on the following fact: Let $\mathcal{U}$ be the dense, open subset of null geodesics of $\mathbb{R}^{n+2}_{p+1}$ that cross the horizontal hyperplane $\{x_{n+2}=0\}$ (in the Minkowski case $(p,q)=(n,0)$, all null geodesics cross the horizontal hyperplane). Then the correspondence $\{(\sigma,0)+t(\nu,1)\mid t\in\mathbb{R}\}\mapsto(\sigma,\nu)$ defines a contactomorphism between $\mathcal{U}$ and the unit tangent bundle $T^1\mathbb{R}^{n+1}_p$. The canonical contact structure α of the unit tangent of a pseudo-Riemannian manifold $(\mathcal{M},g)$ is given by the expression $\alpha=g(d\sigma,\nu)$, where ν is a unit vector tangent to $\mathcal{M}$ at the point σ. Hence, given an immersion $x\mapsto(\sigma(x),\nu(x))$ of an n-dimensional manifold such that $x\mapsto\sigma(x)$ is an immersion as well (a generic assumption), the Legendre condition $g(d\sigma_x,\nu(x))=0$ simply means that ν is the Gauss map of σ or, equivalently, ν is a unit vector field normal to the immersed hypersurface σ.

The interpretation of the proof of Theorem 3 in terms of contact geometry is as follows: The space of null geodesics of $\mathbb{R}^{n+2}_{p+1}$ may be identified with space of one-jets on $\mathbb{S}^p\times\mathbb{S}^q$, that is, the space $T(\mathbb{S}^p\times\mathbb{S}^q)\times\mathbb{R}$ such that to the triple $(\nu,V,z)\in T(\mathbb{S}^p\times\mathbb{S}^q)\times\mathbb{R}$, we associate the null line $\{V+z\bar{\nu}+t\nu\mid t\in\mathbb{R}\}\subset\mathbb{R}^{n+2}_{p+1}$. The natural contact structure on the space of one-jets $T\mathcal{M}\times\mathbb{R}$, where $(\mathcal{M},g)$ is a pseudo-Riemannian manifold, is given by $\alpha:=\psi-dz$, where ψ is the *Liouville* form[3] or *tautological* form on $T\mathcal{M}$. Moreover a generic Legendrian immersion

[3] To be more precise, the Liouville form is canonically defined on the cotangent bundle $T^*\mathcal{M}$ of a differentiable manifold $\mathcal{M}$. If $\mathcal{M}$ is moreover equipped with a pseudo-Riemannian metric (as it is the case of $\mathbb{S}^p\times\mathbb{S}^q$), we may identify $T^*\mathcal{M}$ and $T\mathcal{M}$ and therefore speak of a Liouville form on $T\mathcal{M}$.

in $T\mathcal{M}\times\mathbb{R}$ is locally a section and takes the form $\nu\mapsto(\nu,\nabla\sigma(\nu),\sigma(\nu))$, where $\sigma\in C^2(\mathcal{M})$ and ∇ denotes the gradient of the metric g. It follows, in the case $\mathcal{M}=\mathbb{S}^p\times\mathbb{S}^q$, that a generic Legendrian congruence of null lines of $\mathbb{R}^{n+2}_{p+1}$ takes the form

$$\nu\mapsto\{\nabla\sigma(\nu)+\sigma(\nu)\bar{\nu}+t\nu\mid t\in\mathbb{R}\},$$

where $\sigma\in C^2(\mathbb{S}^p\times\mathbb{S}^q)$. The choice of real function $\tau\in C^2(\mathbb{S}^p\times\mathbb{S}^q)$ determines an n-dimensional submanifold parametrized by $\nu\mapsto\nabla\sigma(\nu)+\sigma(\nu)\bar{\nu}+\tau(\nu)\nu$, one of whose null normals is ν. These observations inspired the proof of Theorem 3.

Finally the proof of Theorem 4 comes from the fact, proved in [Godoy and Salvai 2013], that the space of null geodesics of $\mathbb{S}^{n+2}_{p+1}$ can be identified with $T^1(\mathbb{S}^p\times\mathbb{S}^q)$, the unit tangent bundle of $\mathbb{S}^p\times\mathbb{S}^q$, as follows: to the pair $(\nu,\psi)\in T^1(\mathbb{S}^p\times\mathbb{S}^q)$, we associate the null line $\{\psi+t\nu\mid t\in\mathbb{R}\}\subset\mathbb{S}^{n+2}_{p+1}$.

5.2. *Relation between Theorems 2 and 3 in the case $(p,q)=(n,0)$*. In the Lorentzian case $(p,q)=(n,0)$, it is easy to relate the formulas of Theorems 2 and 3. To avoid confusion, all mathematical quantities from Theorem 2 will be written with subscript 2, and those from Theorem 3 with subscript 3. We start by writing $\nu_3=(\nu_2,1)\in\mathbb{S}^n\times\mathbb{S}^0\simeq\mathbb{S}^n\times\{1,-1\}$, so that $\bar{\nu}_3=(\nu_2,-1)$. Hence the main formula of Theorem 3 becomes

$$\varphi=\big((\tau_3+\sigma_3)\nu_2+2\nabla\sigma_3,\,\tau_3-\sigma_3\big),$$

where $\sigma_3\in C^4(\mathbb{S}^n\times\mathbb{S}^0)\simeq C^4(\mathbb{S}^n)$ and τ_3 depends on the second derivatives of σ_3. Introducing $\sigma_2:=2\sigma_3\,\nu_2+2\nabla\sigma_3$ and $\tau_2:=\tau_3-\sigma_3$, we obtain

$$\varphi=(\sigma_2+\tau_2\nu_2,\,\tau_2),$$

which is exactly the main formula of Theorem 2. Observe that $\langle d\sigma_2,\nu_2\rangle$ vanishes, that is, ν_2 is normal to the immersion σ_2, which is therefore parametrized by its Gauss map. Moreover $\langle\sigma_2,\nu_2\rangle=2\sigma_3$, that is, $2\sigma_3$ is the support function of the immersion σ_2.

References

[Alekseevsky et al. 2009] D. V. Alekseevsky, K. Medori, and A. Tomassini, "Однородные пара-кэлеровы многообразия Эйнштейна", *Uspekhi Mat. Nauk* **64**:1(385) (2009), 3–50. Translated as "Homogeneous para-Kähler Einstein manifolds" in *Russian Math. Surveys* **64**:1 (2009), 1–43. MR 2010k:53068 Zbl 1179.53050

[Anciaux and Godoy 2012] H. Anciaux and Y. Godoy, "Marginally trapped submanifolds in Lorentzian space forms and in the Lorentzian product of a space form by the real line", preprint, 2012. arXiv 1209.5118

[Chen 2009] B.-Y. Chen, "Classification of marginally trapped surfaces of constant curvature in Lorentzian complex plane", *Hokkaido Math. J.* **38**:2 (2009), 361–408. MR 2010f:53092 Zbl 1187.53056

[Chen and Dillen 2007] B.-Y. Chen and F. Dillen, "Classification of marginally trapped Lagrangian surfaces in Lorentzian complex space forms", *J. Math. Phys.* **48**:1 (2007), Art. ID #013509. MR 2008d:53097 Zbl 1121.53047

[Chruściel et al. 2010] P. T. Chruściel, G. J. Galloway, and D. Pollack, "Mathematical general relativity: a sampler", *Bull. Amer. Math. Soc.* (*N.S.*) **47**:4 (2010), 567–638. MR 2011j:53152 Zbl 1205.83002

[Cruceanu et al. 1996] V. Cruceanu, P. Fortuny, and P. M. Gadea, "A survey on paracomplex geometry", *Rocky Mountain J. Math.* **26**:1 (1996), 83–115. MR 97c:53112 Zbl 0856.53049

[Georgiou and Guilfoyle 2014] N. Georgiou and B. Guilfoyle, "Marginally trapped surfaces in spaces of oriented geodesics", *J. Geom. Phys.* **82** (2014), 1–12. MR 3206636 Zbl 06300570 arXiv 1305.6600

[Godoy and Salvai 2013] Y. Godoy and M. Salvai, "The canonical contact structure on the space of oriented null geodesics of pseudospheres and products", *Adv. Geom.* **13**:4 (2013), 713–722. MR 3181543 Zbl 06229240

[Kriele 1999] M. Kriele, *Spacetime: foundations of general relativity and differential geometry*, Lecture Notes in Physics **59**, Springer, Berlin, 1999. MR 2001g:53126 Zbl 0933.83001

[Penrose 1965] R. Penrose, "Gravitational collapse and space-time singularities", *Phys. Rev. Lett.* **14** (1965), 57–59. MR 30 #2897 Zbl 0125.21206

Received September 23, 2013. Revised February 20, 2014.

HENRI ANCIAUX
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO
RUA DO MATÃO, 1010
CIDADE UNIVERSITÁRIA
05508-090 SÃO PAULO, SP
BRAZIL

henri@ime.usp.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.275

ONE LINE COMPLEX KLEINIAN GROUPS

WALDEMAR BARRERA, ANGEL CANO AND JUAN PABLO NAVARRETE

We give an algebraic description of those subgroups of PGL(3, $\mathbb{C}$) acting on $\mathbb{P}^2_{\mathbb{C}}$ with Kulkarni limit set equal to one complex projective line. Conversely, we prove that the Kulkarni limit set of a group $G \le$ PGL(3, $\mathbb{C}$) acting properly and discontinuously on the complement of one line in $\mathbb{P}^2_{\mathbb{C}}$ is equal to one or two lines.

1. Introduction

The Kleinian groups are discrete subgroups of PSL(2, $\mathbb{C}$) acting on $\mathbb{S}^2 \cong \mathbb{P}^1_{\mathbb{C}}$ in such way that its limit set is not all of $\mathbb{S}^2$. They are classified in elementary and nonelementary groups. The elementary groups are those Kleinian groups whose limit set is equal to zero, one or two points, and they are classified (see [Maskit 1988]). The nonelementary groups are those Kleinian groups whose limit set contains more than two points and in this case its limit set is a perfect set.

Our interest relies on the study of *complex Kleinian groups*. These are discrete subgroups of PGL(3, $\mathbb{C}$) acting properly and discontinuously on some open subset of $\mathbb{P}^2_{\mathbb{C}}$. In this setting, there is no standard definition of limit set, however, in [Barrera Vargas et al. 2011] it is proved that under some mild hypothesis on the dynamics of the group, Kulkarni's definition of limit set is an appropriate definition (see Definition 2.1).

In [Cano and Seade 2014] it is proved that the Kulkarni limit set of a complex Kleinian group contains a complex projective line whenever the group is infinite. Moreover, in [Barrera Vargas et al. 2011] it is proved that under some mild hypothesis on the group, the Kulkarni limit set is a union of complex projective lines. The definition of elementary group in this case is that the Kulkarni limit set consists of a finite union of complex projective subspaces (see [Cano et al. 2013]).

An interesting problem consists of classifying all elementary complex Kleinian groups, and one natural step consists of classifying those discrete subgroups of PGL(3, $\mathbb{C}$) whose Kulkarni limit set consists of one complex projective line. In

Research supported by CONACYT Project Number 176680 and SEP grant P/PIFI-2011-31MSU0098J-15.
MSC2010: 37F30, 22E40.
Keywords: complex Kleinian groups, Kulkarni limit set, complex projective plane.

this paper we prove that the complex Kleinian group $G \leq \mathrm{PGL}(3, \mathbb{C})$ is virtually nilpotent whenever its Kulkarni limit set is equal to one complex projective line. In fact we prove the following:

Theorem 1.1. *If G is a subgroup of* $\mathrm{PGL}(3, \mathbb{C})$ *such that its Kulkarni limit set $\Lambda(G)$ consists of precisely one complex projective line ℓ, then*:

(i) *If G contains a loxoparabolic element then G is a finite cyclic extension of order* $1, 2, 3, 4$ *or* 6 *of $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}_{n_0}$, where $n_0 \in \mathbb{N}$ is arbitrary. The $\mathbb{Z} \oplus \mathbb{Z}$ is a group of rank two generated by a loxoparabolic element and another element which can be loxoparabolic or parabolic. Also, the $\mathbb{Z}_{n_0}$ is a group of complex reflections.*

(ii) *If G does not contain any loxoparabolic elements and the group G does not contain any element which acts as a parabolic element on the complex line $\Lambda(G) = \ell$, then G is a group of isometries of $\mathbb{C}^2$ and it contains a free abelian normal subgroup of finite index and of rank less than or equal to four.*

(iii) *If G does not contain any loxoparabolic elements but it does contain an element which acts as a parabolic element on the complex line $\Lambda(G) = \ell$, then G does not contain any irrational ellipto-parabolic elements and it is a finite extension of a unipotent subgroup* (*this subgroup consists of unipotent parabolic maps*)*. Hence it is a finite extension of a group of the form $\mathbb{Z}$, $\mathbb{Z}^2$, $\mathbb{Z}^3$, $\mathbb{Z}^4$, Δ_k or Γ_k, where*

$$\Delta_k = \langle A, B, C, D \,:\, C, D \text{ are central and } [A, B] = C^k \rangle, \quad k \in \mathbb{N},$$

and

$$\Gamma_k = \langle A, B, C \,:\, C \text{ is central and } [A, B] = C^k \rangle, \quad k \in \mathbb{N}.$$

The outline of the proof of Theorem 1.1 is as follows: Since the group acts properly and discontinuously on the complement of one complex projective line in $\mathbb{P}^2_{\mathbb{C}}$, the dynamics of each element in the group are restricted in some way; see Remark 2.4. In fact, the elements in the group are elliptic, parabolic or loxoparabolic according to the classification given in [Navarrete 2008].

If the group contains a loxoparabolic element, then there are restrictions on the group G, as shown in Lemma 3.1. The proof of Theorem 1.1(i) follows from the fact that there exists an invariant complex projective line where the action of the group is properly discontinuous except in one point. Hence the group acts as a Euclidean group on this line.

If the group does not contain any loxoparabolic elements, then we consider the following two cases:

If G acts on the limit set $\ell \cong \hat{\mathbb{C}} \cong \mathbb{S}^2$ without parabolic elements then G can be considered as a group of Euclidean isometries of $\mathbb{R}^4$.

If some element in G acts on the limit set ℓ as parabolic element then the group can be identified with a group of triangular matrices (see Proposition 5.2). The existence of irrational ellipto-parabolic elements in the group is ruled out by Propositions 5.6 and 5.7. Finally, there exists a unipotent subgroup of finite index (see Proposition 5.8.)

We remark that not every finite extension of those nilpotent groups given in Theorem 1.1(ii) and (iii) can occur as a group with Kulkarni limit set equal to one line. Which of them can occur is a more delicate question. However, Theorem 1.1 gives a qualitative description according to the dynamics of the kind of elements contained in the group.

We are not restricting here to the case where the quotient space $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ by the group G is compact. The case where the action of $G \leq \mathrm{PGL}(3, \mathbb{C})$ on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ is free, properly discontinuous and the quotient is compact is handled in [Fillmore and Scheuneman 1973; Scheuneman 1974; Suwa 1975].

Finally, if $G \leq \mathrm{PGL}(3, \mathbb{C})$ satisfies $\Lambda(G)$ is equal to one line, then G acts properly and discontinuously on the complement of one line in $\mathbb{P}^2_{\mathbb{C}}$, so G can be considered as a discrete subgroup of $\mathrm{Aff}(\mathbb{C}^2)$ acting properly and discontinuously on $\mathbb{C}^2$. The converse statement is not true as we show in the following:

Theorem 1.2. *Let $G \leq \mathrm{PGL}(3, \mathbb{C})$ be an infinite group which acts properly and discontinuously on the complement of the line $\ell \subset \mathbb{P}^2_{\mathbb{C}}$.*

(i) *If G contains a loxoparabolic element then,*
 - *the Kulkarni limit set $\Lambda(G)$ is equal to the union of ℓ and another line whenever G contains a cyclic subgroup of finite index generated by a loxoparabolic element, or*
 - *the Kulkarni limit set $\Lambda(G)$ is equal to ℓ whenever G contains a finite-index free abelian subgroup of rank two generated by a loxoparabolic element and another element, which can be either loxoparabolic or parabolic.*

(ii) *If G does not contain any loxoparabolic elements then $\Lambda(G) = \ell$.*

2. Preliminaries

2.1. ***Projective geometry.*** Recall that the complex projective plane $\mathbb{P}^2_{\mathbb{C}}$ is defined as

$$\mathbb{P}^2_{\mathbb{C}} := (\mathbb{C}^3 \setminus \{\mathbf{0}\})/\mathbb{C}^*,$$

where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ acts on $\mathbb{C}^3 \setminus \{\mathbf{0}\}$ by the usual scalar multiplication. This is a compact connected complex 2-dimensional manifold. Let $[\,\cdot\,] : \mathbb{C}^3 \setminus \{\mathbf{0}\} \to \mathbb{P}^2_{\mathbb{C}}$ be the quotient map. If $\beta = \{e_1, e_2, e_3\}$ is the standard basis of $\mathbb{C}^3$, we write $[e_j] = e_j$ and if $\mathbf{z} = (z_1, z_2, z_3) \in \mathbb{C}^3 \setminus \{\mathbf{0}\}$ then we write $[\mathbf{z}] = [z_1 : z_2 : z_3]$. Also, $\ell \subset \mathbb{P}^2_{\mathbb{C}}$ is said to be a complex line if $[\ell]^{-1} \cup \{\mathbf{0}\}$ is a complex linear subspace of dimension two.

Given distinct points $[\mathbf{z}], [\mathbf{w}] \in \mathbb{P}^2_{\mathbb{C}}$, there is a unique complex projective line passing through $[\mathbf{z}]$ and $[\mathbf{w}]$. Such a complex projective line is called a *line*, for short, and it is denoted by $\overleftrightarrow{[\mathbf{z}], [\mathbf{w}]}$. Consider the action of $\mathbb{C}^*$ on $\mathrm{GL}(3, \mathbb{C})$ given by the usual scalar multiplication. Then

$$\mathrm{PGL}(3, \mathbb{C}) = \mathrm{GL}(3, \mathbb{C})/\mathbb{C}^*$$

is a Lie group. The elements of this Lie group are called projective transformations. Let $[[\,\cdot\,]] : \mathrm{GL}(3, \mathbb{C}) \to \mathrm{PGL}(3, \mathbb{C})$ be the quotient map, $g \in \mathrm{PGL}(3, \mathbb{C})$ and $\mathbf{g} \in \mathrm{GL}(3, \mathbb{C})$. We say that $\mathbf{g}$ is a lift of g if $[[\mathbf{g}]] = g$. One can show that $\mathrm{PGL}(3, \mathbb{C})$ is a Lie group that acts transitively, effectively and by biholomorphisms on $\mathbb{P}^2_{\mathbb{C}}$ by $[[\mathbf{g}]]([\mathbf{w}]) = [\mathbf{g}(\mathbf{w})]$, where $\mathbf{w} \in \mathbb{C}^3 \setminus \{\mathbf{0}\}$ and $\mathbf{g} \in \mathrm{GL}(3, \mathbb{C})$.

The Fubini–Study metric on $\mathbb{P}^2_{\mathbb{C}}$ is a useful tool in the computation of the Kulkarni limit set of cyclic subgroups of $\mathrm{PGL}(3, \mathbb{C})$ acting on $\mathbb{P}^2_{\mathbb{C}}$ (see [Navarrete 2008]). The Fubini–Study distance $d([\mathbf{z}], [\mathbf{w}])$ between $[\mathbf{z}], [\mathbf{w}] \in \mathbb{P}^2_{\mathbb{C}}$ satisfies the equation

$$\cos^2\bigl(d([\mathbf{z}], [\mathbf{w}])\bigr) = \frac{|z_1\bar{w}_1 + z_2\bar{w}_2 + z_3\bar{w}_3|^2}{(|z_1|^2 + |z_2|^2 + |z_3|^2)(|w_1|^2 + |w_2|^2 + |w_3|^2)}. \tag{1}$$

We denote by $M_{3\times 3}(\mathbb{C})$ the space of all 3×3 matrices with entries in $\mathbb{C}$ equipped with the standard topology. The quotient space

$$(M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\})/\mathbb{C}^*$$

is called the space of *pseudo-projective maps of* $\mathbb{P}^2_{\mathbb{C}}$ and it is naturally identified with the projective space $\mathbb{P}^8_{\mathbb{C}}$. Since $\mathrm{GL}(3, \mathbb{C})$ is an open, dense, $\mathbb{C}^*$-invariant set of $M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\}$, we obtain that the space of pseudo-projective maps of $\mathbb{P}^2_{\mathbb{C}}$ is a compactification of $\mathrm{PGL}(3, \mathbb{C})$. As in the case of projective maps, if $\mathbf{s}$ is an element in $M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\}$, then $[\mathbf{s}]$ denotes the equivalence class of the matrix $\mathbf{s}$ in the space of pseudo-projective maps of $\mathbb{P}^2_{\mathbb{C}}$. Also, we say that $\mathbf{s} \in M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\}$ is a lift of the pseudo-projective map S whenever $[\mathbf{s}] = S$.

Let S be an element in $(M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\})/\mathbb{C}^*$ and $\mathbf{s}$ a lift to $M_{3\times 3}(\mathbb{C}) \setminus \{\mathbf{0}\}$ of S. The matrix $\mathbf{s}$ induces a nonzero linear transformation $s : \mathbb{C}^3 \to \mathbb{C}^3$, which is not necessarily invertible. Let $\mathrm{Ker}\, s \subsetneq \mathbb{C}^3$ be its kernel and let $\mathrm{Ker}\, S$ denote its projectivization to $\mathbb{P}^2_{\mathbb{C}}$, taking into account that $\mathrm{Ker}\, S := \varnothing$ whenever $\mathrm{Ker}\, s = \{(0, 0, 0)\}$. We refer to [Cano and Seade 2010] for more details about this subject.

2.2. ***Complex Kleinian groups.*** We recall that a point $z \in \hat{\mathbb{C}} \cong \mathbb{S}^2 \cong \mathbb{P}^1_{\mathbb{C}}$ is a *limit point* of the discrete subgroup G of $\mathrm{PSL}(2, \mathbb{C})$ if it is a cluster point of some orbit $\{gx : g \in G\}$, where $x \in \hat{\mathbb{C}}$. The set $\Lambda(G) = \{z \in \hat{\mathbb{C}} : z \text{ is a limit point of } G\}$ is called the *limit set* of G (see [Marden 2007]).

It is a known fact that the action of G on the complement of the limit set $\hat{\mathbb{C}} \setminus \Lambda(G)$ is properly discontinuous. However, when working on higher dimensions, this is no longer valid. See Example 2.3 below.

Kulkarni considers a very general setting of discrete group actions on a topological space X, and the Kulkarni limit set provides a canonical choice of a closed G-invariant set in X such that the G-action on its complement is properly discontinuous.

Definition 2.1. Let $G \subset \mathrm{PGL}(n+1, \mathbb{C})$ be a subgroup. The set $L_0(G)$ is defined as the closure of the points in $\mathbb{P}^n_{\mathbb{C}}$ with infinite isotropy group. In other words, $L_0(G) = \overline{\{x \in \mathbb{P}^n_{\mathbb{C}} : \mathrm{Stab}_G(x) \text{ is an infinite group}\}}$ (see [Kulkarni 1978]). The set $L_1(G)$ is the closure of the set of cluster points of the G-orbit of z, where z runs over $\mathbb{P}^n_{\mathbb{C}} \setminus L_0(G)$. Recall that q is a cluster point of the family of sets $\{g(K) : g \in G\}$, where $K \subset \mathbb{P}^n_{\mathbb{C}}$ is a nonempty set, if there is a sequence $(k_m)_{m\in\mathbb{N}} \subset K$ and a sequence of distinct elements $(g_m)_{m\in\mathbb{N}} \subset G$ such that $g_m(k_m) \xrightarrow[m\to\infty]{} q$. The set $L_2(G)$ is defined as the closure of the union of cluster points of $\{g(K) : g \in G\}$, where K runs over all the compact sets in $\mathbb{P}^n_{\mathbb{C}} \setminus (L_0(G) \cup L_1(G))$. The *Kulkarni limit set* for G is defined as the G-invariant closed set

$$\Lambda(G) = L_0(G) \cup L_1(G) \cup L_2(G).$$

The *discontinuity region in the sense of Kulkarni* of G is defined as

$$\Omega(G) = \mathbb{P}^n_{\mathbb{C}} \setminus \Lambda(G).$$

If $\Omega(G) \neq \varnothing$ then we say that G is a *complex Kleinian group*.

In the case of a cyclic group $\langle g\rangle$, we write $L_0(g)$, $L_1(g)$, etc. instead of $L_0(\langle g\rangle)$, $L_1(\langle g\rangle)$, etc.

The following lemma is a useful tool for the computation of Kulkarni limit sets, and we include it here for reader's convenience. See [Navarrete 2008] for a proof.

Lemma 2.2. *Let G be a subgroup of* $\mathrm{PGL}(3, \mathbb{C})$. *If C is a closed set such that for every compact set $K \subset \mathbb{P}^2_{\mathbb{C}} \setminus C$, the cluster points of the family of compact sets $\{g(K)\}_{g\in G}$ are contained in $L_0(G) \cup L_1(G)$, then $L_2(G) \subset C$.*

Example 2.3. If $g \in \mathrm{PGL}(3, \mathbb{C})$ is induced by the matrix

$$\mathbf{g} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \lambda \in \mathbb{C}^*,\ |\lambda| < 1,$$

then:

(i) $L_0(g) = \{e_1, e_2\}$.

(ii) $g^{-n}(\cdot) \xrightarrow[n\to\infty]{} e_2$ uniformly on compact subsets of $\mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$.

(iii) $g^n(\cdot) \xrightarrow[n\to\infty]{} e_1$ uniformly on compact subsets of $\mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$.

(iv) $L_1(g) = \{e_1, e_2\}$.

(v) $L_2(g) = \overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3}$.

(vi) The action of the cyclic group generated by g on $\mathbb{P}^2_{\mathbb{C}} \setminus \{e_1, e_2\}$ is not properly discontinuous.

(vii) The cyclic group generated by g acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$.

Proof.

(i) The proof follows straightforwardly from the fact that $\{(z_1, 0, 0) : z_1 \in \mathbb{C}\}$ and $\{(0, z_2, 0) : z_2 \in \mathbb{C}\}$ are the only possible eigenspaces for each matrix of the form

$$\mathbf{g}^n = \begin{pmatrix} 1 & 0 & n \\ 0 & \lambda^n & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad n \in \mathbb{Z} \setminus \{0\}.$$

(ii) If K is a compact subset of $\mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$, then every point in K can be written as $[\mathbf{z}] = [z_1 : z_2 : z_3]$, where $|z_1|^2 + |z_2|^2 + |z_3|^2 = 1$, $|z_2| \geq \epsilon$ and $|z_3| \geq \epsilon$ for some fixed $\epsilon > 0$. If $d(g^{-n}[\mathbf{z}], e_2)$ denotes the Fubini–Study distance between $g^{-n}[\mathbf{z}]$ and e_2, then

$$\begin{aligned} \cos^2\big(d(g^{-n}([\mathbf{z}]), e_2)\big) &= \frac{|\lambda^{-n} z_2|^2}{|z_1 - n z_3|^2 + |\lambda^{-n} z_2|^2 + |z_3|^2} \\ &\geq \frac{|\lambda^{-n} z_2|^2}{(n\epsilon + 1)^2 + |\lambda^{-n} z_2|^2 + 1} \\ &\geq \frac{|\lambda|^{-2n} \epsilon^2}{(n\epsilon + 1)^2 + |\lambda|^{-2n} \epsilon^2 + 1}, \end{aligned}$$

and the last expression goes to 1 as $n \to \infty$. Therefore, $g^{-n}(\cdot) \xrightarrow[n\to\infty]{} e_2$ uniformly on compact subsets of $\mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$.

(iii) The proof is analogous to (ii)

(iv) Since g acts on the invariant line $\overleftrightarrow{e_1, e_2}$ as a loxodromic element of $\mathrm{PGL}(2, \mathbb{C})$ with fixed points e_1 and e_2, the orbits of the points in $\overleftrightarrow{e_1, e_2} \setminus \{e_1, e_2\}$ accumulate at e_1 and e_2.

Also, g acts on the invariant line $\overleftrightarrow{e_1, e_3}$ as a parabolic element of $\mathrm{PGL}(2, \mathbb{C})$ with fixed point at e_1, so the orbits of the points in $\overleftrightarrow{e_1, e_3} \setminus \{e_1\}$ accumulate at e_1.

Finally, by (ii) and (iii), the set of cluster points of the orbits of points in $\mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$ is equal to $\{e_1, e_2\}$.

(v) By (ii) and (iii), for every compact set $K \subset \mathbb{P}^2_{\mathbb{C}} \setminus (\overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3})$, the cluster points of the family of compact sets $\{g^n(K)\}_{n\in\mathbb{Z}}$ is equal to $\{e_1, e_2\} \subset L_0(g) \cup L_1(g)$. It follows, by Lemma 2.2, that $L_2(g) \subset \overleftrightarrow{e_1, e_2} \cup \overleftrightarrow{e_1, e_3}$.

Conversely, for every $z \in \mathbb{C}^*$, the compact set

$$K_z = \left\{ \left[z + \lambda^n - \frac{1}{n} : 1 : -\frac{z}{n} \right] \,\middle|\, n \in \mathbb{N} \right\} \cup \{[z : 1 : 0]\}$$

is contained in $\mathbb{P}^2_{\mathbb{C}} \setminus \{e_1, e_2\} = \mathbb{P}^2_{\mathbb{C}} \setminus (L_0(g) \cup L_1(G))$. Since

$$\begin{aligned} g^n\left(\left[z + \lambda^n - \frac{1}{n} : 1 : -\frac{z}{n} \right] \right) &= \left[\lambda^n - \frac{1}{n} : \lambda^n : -\frac{z}{n} \right] \\ &= [n\lambda^n - 1 : n\lambda^n : -z] \xrightarrow[n \to \infty]{} [1 : 0 : z], \end{aligned}$$

it follows that $\overleftrightarrow{e_1, e_3} \subset L_2(g)$.

Similarly, for every $z \in \mathbb{C}^*$, consider the compact set

$$\left\{ [n : nz : -\lambda^{-n}] \,\middle|\, n \in \mathbb{N} \right\} \cup \{e_3\} \subset \mathbb{P}^2_{\mathbb{C}} \setminus (L_0(g) \cup L_1(G)).$$

Since $g^{-n}([n : nz : -\lambda^{-n}]) = [\lambda^n + 1 : z : -1/n] \xrightarrow[n \to \infty]{} [1 : z : 0]$, it follows that $\overleftrightarrow{e_1, e_2} \subset L_2(g)$.

(vi) If K_z is as in the proof of (v), then $g^n(K_z)$ intersects any compact neighborhood of $[1 : 0 : z]$ for infinitely many values of $n \in \mathbb{Z}$.

(vii) The set $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$ is naturally identified with $\mathbb{C}^2$ by the map $[z_1 : z_2 : 1] \mapsto (z_1, z_2)$, and the action is now $g(z_1, z_2) = (z_1 + 1, \lambda z_2)$.

If for some fixed $R > 0$, $(z_1, z_2) \in \mathbb{C}^2$ satisfies

$$\|(z_1, z_2)\|_1 = |z_1| + |z_2| \leq R, \tag{2}$$

and

$$\|g^n(z_1, z_2)\|_1 = |z_1 + n| + |\lambda^n z_2| \leq R, \tag{3}$$

then $|n| \leq |z_1 + n| + |z_1| \leq 2R$. Hence (2) and (3) are satisfied for finitely many values of $n \in \mathbb{Z}$. Therefore the cyclic group $\langle g \rangle$ acts properly discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$. □

By conformal properties, we have that the Kulkarni limit set of a discrete subgroup of $\mathrm{PGL}(2, \mathbb{C})$ acting on $\hat{\mathbb{C}}$ agrees with its classical limit set. In fact, $L_0 = L_1 = L_2 = \Lambda$ in that case. However, when working in higher dimensional projective geometry, the sets L_0, L_1 and L_2 can be quite different amongst themselves. Moreover, the set $\Omega(G)$ is not always the maximal open subset where the action is properly discontinuous, as illustrated in Example 2.3(vii). Nevertheless, when G acts on $\mathbb{P}^2_{\mathbb{C}}$ without fixed points nor invariant lines, it is possible to show that $\Omega(G)$ is the maximal open set where the action is properly discontinuous (see [Barrera Vargas et al. 2011]).

2.3. ***Classification of automorphisms of*** $\mathbb{P}^2_{\mathbb{C}}$**.** The nontrivial elements of $\mathrm{PGL}(3, \mathbb{C})$ can be classified as elliptic, parabolic or loxodromic (see [Navarrete 2008]).

The *elliptic* elements in $\mathrm{PGL}(3,\mathbb{C})$ are those elements g that have a lift to $\mathrm{GL}(3,\mathbb{C})$ whose Jordan canonical form is

$$\begin{pmatrix} e^{2\pi i\theta_1} & 0 & 0 \\ 0 & e^{2\pi i\theta_2} & 0 \\ 0 & 0 & e^{2\pi i\theta_3} \end{pmatrix}, \qquad \text{where } \theta_1,\theta_2,\theta_3 \in \mathbb{R}.$$

The limit set $\Lambda(g)$ for an elliptic element g is $\varnothing$ or all of $\mathbb{P}^2_{\mathbb{C}}$ according to whether the order of g is finite or infinite. Those subgroups of $\mathrm{PGL}(3,\mathbb{C})$ containing an elliptic element of infinite order cannot be discrete.

The *parabolic* elements in $\mathrm{PGL}(3,\mathbb{C})$ are those elements g such that the limit set $\Lambda(g)$ is equal to a single complex line. If g is parabolic then it has a lift to $\mathrm{GL}(3,\mathbb{C})$ whose Jordan canonical form is one of the following matrices:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} e^{2\pi i\theta} & 1 & 0 \\ 0 & e^{2\pi i\theta} & 0 \\ 0 & 0 & e^{-4\pi i\theta} \end{pmatrix}, \qquad \theta \in \mathbb{R}\setminus\mathbb{Z}.$$

In the first case, $\Lambda(g)$ is the complex line consisting of all the fixed points of g. In the second case, $\Lambda(g)$ is the unique g-invariant complex line. In the last case, $\Lambda(g)$ is the complex line determined by the two fixed points of g.

There are four kinds of *loxodromic* elements in $\mathrm{PGL}(3,\mathbb{C})$:

- The *complex homotheties* are those elements $g \in \mathrm{PGL}(3,\mathbb{C})$ that have a lift to $\mathrm{GL}(3,\mathbb{C})$ whose Jordan canonical form is

  $$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^{-2} \end{pmatrix}, \qquad \lambda \in \mathbb{C}, \quad |\lambda| \neq 1,$$

  and its limit set $\Lambda(g)$ is the set of fixed points of g, consisting of one line ℓ and a point not lying in ℓ. Moreover, in this case, $L_0(g)\cup L_1(g) = \Lambda(g)$ is not contained in one line.

- The *screws* are those elements $g \in \mathrm{PGL}(3,\mathbb{C})$ that have a lift to $\mathrm{GL}(3,\mathbb{C})$ whose Jordan canonical form is

  $$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & (\lambda\mu)^{-1} \end{pmatrix}, \qquad \lambda,\mu \in \mathbb{C}, \quad \lambda \neq \mu, \quad |\lambda| = |\mu| \neq 1,$$

  and its limit set $\Lambda(g)$ consists of the line ℓ on which g acts as an elliptic transformation of $\mathrm{PSL}(2,\mathbb{C})$ and the fixed point of g not lying in ℓ. In this case, $L_0(g)\cup L_1(g) = \Lambda(g)$ is not contained in one line.

- The *loxoparabolic* elements $g \in \mathrm{PGL}(3, \mathbb{C})$ have a lift to $\mathrm{GL}(3, \mathbb{C})$ whose Jordan canonical form is
$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \lambda \in \mathbb{C}, \quad |\lambda| \neq 1,$$
and the limit set $\Lambda(g)$ consists of two g-invariant complex lines. The element g acts on one of these complex lines as a parabolic element of $\mathrm{PSL}(2, \mathbb{C})$ and on the other as a loxodromic element of $\mathrm{PSL}(2, \mathbb{C})$. In this case $L_0(g) \cup L_1(g)$ is contained in one line.
- The *strongly loxodromic* elements $g \in \mathrm{PGL}(3, \mathbb{C})$ have a lift to $\mathrm{GL}(3, \mathbb{C})$ whose Jordan canonical form is
$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \qquad \lambda_1, \lambda_2, \lambda_3 \in \mathbb{C}, \quad |\lambda_1| < |\lambda_2| < |\lambda_3|.$$
This kind of transformation has three fixed points: one of them is attracting, another is repelling and the last one is a saddle. The limit set $\Lambda(g)$ is equal to the union of the complex line determined by the attracting and saddle points and the complex line determined by the saddle and repelling points. In this case $L_0(g) \cup L_1(g)$ consists of three points in general position, so $L_0(g) \cup L_1(g)$ is not contained in one line.

Remark 2.4. If $g \in \mathrm{PGL}(3, \mathbb{C})$ satisfies that $L_0(g) \cup L_1(g)$ is contained in one line then g is elliptic, parabolic or loxoparabolic.

2.4. ***Groups acting properly and discontinuously on*** $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$**.** If G is a subgroup of $\mathrm{PGL}(3, \mathbb{C})$ acting properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$, where $\ell \subset \mathbb{P}^2_{\mathbb{C}}$ is a line, then we can assume, from now on, that $\ell = \overleftrightarrow{e_1, e_2}$. So every element $g \in G$ can be induced by a matrix of the form

$$\begin{pmatrix} a & b & v \\ c & d & w \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

When convenient, we shall write $a(g), b(g), c(g), \dots$ instead of $a, b, c, \dots$.

We can regard $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ as $\mathbb{C}^2$, and (4) means that g can be considered as the affine automorphism
$$\mathbf{z} \mapsto A\mathbf{z} + \mathbf{v},$$
where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v \\ w \end{pmatrix}. \tag{5}$$

The projection onto the linear part of the affine map above is denoted by

$$\phi : G \to \mathrm{PGL}(2, \mathbb{C}), \quad \phi(g) = A, \tag{6}$$

and it is a group homomorphism.

On the other hand, the map

$$\psi : G \to \mathrm{PGL}(2, \mathbb{C}) \quad \psi(g) = \begin{pmatrix} a & v \\ 0 & 1 \end{pmatrix} \tag{7}$$

is not necessarily a group homomorphism. However, it will be useful in Section 3.

Given that G acts properly discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$, then for every $g \in G$, the cyclic group $\langle g \rangle$ acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$. So one has $L_0(g) \cup L_1(g) \subset \ell$. By Remark 2.4, G contains only elliptic, parabolic or loxoparabolic elements.

In Section 3, we assume that G contains a loxoparabolic element and we prove Theorem 1.1(i) together with some other results that will be useful for the proof of Theorem 1.2 in Section 4.

When G does not contain any loxoparabolic elements, the group $\phi(G)$ contains only elliptic or parabolic elements. In the first part of Section 5, we consider the case when G acts on ℓ without parabolic elements. In other words, $\phi(G)$ does not contain any parabolic element, and we prove Theorem 1.1(ii). Finally, in the last part of the same section, we consider the case when $\phi(G)$ contains a parabolic element, and we finish the proof of Theorem 1.1.

3. *G* contains a loxoparabolic element

Lemma 3.1. *Let ℓ be a line in $\mathbb{P}^2_{\mathbb{C}}$. If G is a discrete subgroup of* $\mathrm{PGL}(3, \mathbb{C})$ *acting properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ and G contains a loxoparabolic element, then there exists a conjugate of G such that every element in this conjugate group has a representative in* $\mathrm{GL}(3, \mathbb{C})$ *of the form*

$$\begin{pmatrix} a & 0 & v \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{8}$$

where a is a root of unity of order 1, 2, 3, 4 *or* 6. *Moreover, this conjugate group acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$.*

Proof. Every element $g \in G$ has a representative matrix of the form (4) and we can assume that the matrix that induces a loxoparabolic element h_0 in G has the form

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad 0 < |\lambda| < 1.$$

The commutator $g_n = [h_0^n, g]$ and its inverse are induced by matrices of determinant one and their traces are equal to

$$\tau_n = \frac{-3ad + bc(1 + \lambda^{-n} + \lambda^n)}{(bc - ad)}.$$

If $bc \neq 0$ then, by Theorem 6.3 in [Navarrete 2008], $g_n \in G$ is a strongly loxodromic element for all sufficiently large n. Hence, for every $g \in G$, $b(g) = 0$ or $c(g) = 0$. It is not hard to check that $b(g) = 0$ for every $g \in G$ or $c(g) = 0$ for every $g \in G$. Therefore we can assume, conjugating if necessary, that $c(g) = 0$ for every $g \in G$.

If $g \in G$ satisfies that $a(g) = 1$ then, by Lemma 3.3 in [Fillmore and Scheuneman 1973], g commutes with h_0. It follows that for every h in the normal subgroup $H = \{g \in G : a(g) = 1\}$, one has $b(h) = w(h) = 0$.

Let g be an arbitrary element in G. Then

$$gh_0g^{-1} = \begin{pmatrix} 1 & \frac{b(-1+\lambda)}{d\lambda} & a - \frac{d+bw(-1+\lambda)}{d} \\ 0 & d & w - w\lambda \\ 0 & 0 & 1 \end{pmatrix} \in H.$$

It follows that $b(-1+\lambda)/(d\lambda) = 0 = w - w\lambda$. Hence $b = 0 = w$.

The line $\overleftrightarrow{e_1, e_3}$ is G-invariant because $b(g) = c(g) = w(g) = 0$. Moreover, G acts on it as a classic elementary group with limit point e_1. In fact, the action of G on this line is the action on $\mathbb{P}^1_{\mathbb{C}}$ of the group

$$\psi(G) = \left\{ \begin{pmatrix} a(g) & v(g) \\ 0 & 1 \end{pmatrix} : g \in G \right\},$$

where ψ is defined as in (7). It follows, by well-known facts on Euclidean groups (see [Maskit 1988]), that $a(g)$ is a root of unity of order 1, 2, 3, 4 or 6. □

Lemma 3.2. *If $G \leq \mathrm{PGL}(3, \mathbb{C})$ acts properly and discontinuously on the complement of the line $\ell \subset \mathbb{P}^2_{\mathbb{C}}$ and G contains a loxoparabolic element then G contains a normal abelian subgroup H isomorphic to $\mathbb{Z} \oplus \mathbb{Z}_{n_0}$ or to $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}_{n_0}$ for some $n_0 \in \mathbb{N}$. Moreover, the index of H in G is equal to* 1, 2, 3, 4 *or* 6.

Proof. We can assume that every element of G is induced by a matrix of the form (8). In this case, the map $\psi : G \to \mathrm{PGL}(2, \mathbb{C})$, defined as in (7), is a homomorphism and its image is a Euclidean group of $\mathrm{PGL}(2,\mathbb{C})$. The kernel of this homomorphism consists of all those transformations in G induced by a matrix of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

but these transformations are necessarily elliptic or the identity. Hence $\operatorname{Ker}\psi$ is a finite group (because it is discrete and every element has finite order). Moreover, it is a cyclic group of some order n_0. Let us denote by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & d_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

the generator of $\operatorname{Ker}\psi$.

Let H be the normal abelian subgroup of G consisting of those elements $g \in G$ induced by a matrix of the form

$$\begin{pmatrix} 1 & 0 & v \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9}$$

We notice that $H/\operatorname{Ker}\psi$ is a free abelian subgroup of the Euclidean group $G/\operatorname{Ker}\psi$, consisting of all parabolic elements. Furthermore, $H/\operatorname{Ker}\psi$ has rank equal to one or two.

Since $H/\operatorname{Ker}\psi$ has index equal to 1, 2, 3, 4 or 6 in $G/\operatorname{Ker}\psi$, it follows that H has index equal to 1, 2, 3, 4 or 6 in G. □

Lemma 3.3. *Let $G \leq \mathrm{PGL}(3,\mathbb{C})$ be a group that acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$. If G contains a loxoparabolic element and the abelian normal subgroup H has rank equal to one (where H is as in Lemma 3.2), then $L_0(G) = L_1(G)$ consists of two points in ℓ and $L_2(G)$ is equal to the union of ℓ and one other line. In particular, the Kulkarni limit set of G is equal to the union of two lines.*

Proof. Let ψ be as in (7) and H be defined as the subgroup of G induced by matrices of the form (9). There are two possible cases:

- If $H/\operatorname{Ker}\psi = G/\operatorname{Ker}\psi$, then we can assume that G is generated by the two elements induced by the matrices

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & d_0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

  where $0 < |\lambda| < 1$ and d_0 is an n_0-th root of unity.

- If $H/\operatorname{Ker}\psi \subsetneq G/\operatorname{Ker}\psi$, then we can assume that G is generated by three elements induced by the matrices

$$\begin{pmatrix}1&0&1\\0&\lambda&0\\0&0&1\end{pmatrix},\quad \begin{pmatrix}1&0&0\\0&d_0&0\\0&0&1\end{pmatrix},\quad \begin{pmatrix}-1&0&0\\0&d_0^m&0\\0&0&1\end{pmatrix},$$

where $0<|\lambda|<1$, d_0 is an n_0-th root of unity and m is an integer. (If the parabolic subgroup of a Euclidean group E has rank one and it is not equal E, then E is the dihedral infinite group $\{z\mapsto \pm z+n,\ n\in\mathbb{Z}\}$.)

In any case, the subgroup N of G generated by

$$\begin{pmatrix}1&0&1\\0&\lambda&0\\0&0&1\end{pmatrix},\quad 0<|\lambda|<1,$$

is normal and it has finite index in G. Moreover, $L_0(N)=L_0(G)=\{e_1,e_2\}$ and $L_1(G)=\bigcup_{i=1}^m g_i(L_1(N))$, where $\{g_1N,\dots,g_mN\}$ are all the cosets of N in G, but $L_1(N)=\{e_1,e_2\}$ and each g_i, $i=1,\dots,m$ fixes e_1 and e_2. Hence, $L_1(G)=\{e_1,e_2\}$.

Since $L_0(G)=L_0(N)=\{e_1,e_2\}=L_1(N)=L_1(G)$, it follows that

$$L_2(G)=\bigcup_{i=1}^m g_i(L_2(N))=\bigcup_{i=1}^m g_i(\overleftrightarrow{e_1,e_2}\cup\overleftrightarrow{e_1,e_3}).$$

We conclude that $L_2(G)=\overleftrightarrow{e_1,e_2}\cup\overleftrightarrow{e_1,e_3}$ because $\overleftrightarrow{e_1,e_2}\cup\overleftrightarrow{e_1,e_3}$ is G-invariant. $\square$

The proof of Theorem 1.1(i) follows immediately from Lemmas 3.2 and 3.3. Now we prove the converse.

Lemma 3.4. *Let $G\le \mathrm{PGL}(3,\mathbb{C})$ be a group that acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}}\setminus\ell$. If G contains a loxoparabolic element and the abelian normal subgroup H has rank equal to two* (*where H is as in Lemma 3.2*), *then $L_0(G)\cup L_1(G)=\ell$ and $L_2(G)\subset\ell$. In particular*, *the Kulkarni limit set of G is equal to ℓ.*

Proof. We can assume that the two transformations induced by

$$\gamma_1=\begin{pmatrix}1&0&1\\0&\lambda&0\\0&0&1\end{pmatrix},\quad 0<|\lambda|<1,\qquad\text{and}\qquad \gamma_2=\begin{pmatrix}1&0&\nu\\0&\delta&0\\0&0&1\end{pmatrix},\quad \nu\in\mathbb{C}\setminus\mathbb{R},\ \delta\in\mathbb{C}^*,$$

generate a subgroup N of finite index in the abelian group H. Hence, $L_0(N)=L_0(H)=L_0(G)$ and $L_1(G)=\bigcup_{i=1}^m g_iL_1(N)$, where $\{g_1N,\dots,g_mN\}$ are all distinct left cosets of N in G. It follows that $L_0(G)\cup L_1(G)=\ell$ whenever $L_0(N)\cup L_1(N)=\ell$ because ℓ is G-invariant.

Now we consider all possible cases:

If $|\delta| = 1$, let us say $\delta = e^{2i\pi\theta}$, $\theta \in \mathbb{R}$, then $L_0(\gamma_2) = \ell$ whenever $\theta \in \mathbb{Q}$ or $L_1(\gamma_2) = \ell$ whenever $\theta \notin \mathbb{Q}$. In any case, $L_0(N) \cup L_1(N) = \ell$.

If $|\delta| \neq 1$ then there are two cases depending on whether $\log|\lambda|$ and $\log|\delta|$ are rationally independent or not. In the first case, the action of N on ℓ is not discrete. Hence $L_0(N) \cup L_1(N) = \ell$. In the second case, there exists an element $\gamma \in N$ such that $L_0(\gamma) \cup L_1(\gamma) = \ell$. Thus $L_0(N) \cup L_1(N) = \ell$.

Since G acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ and $L_0(G) \cup L_1(G) = \ell$, it follows (by Lemma 2.2) that $L_2(G) \subset \ell$. □

We notice that the proof of Theorem 1.2(i) follows from Lemmas 3.3 and 3.4.

Examples. In these examples, λ is a fixed complex number such that $0 < |\lambda| < 1$.

(i) The abelian group G_1, generated by the projective transformations

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

acts properly and discontinuously on $\mathbb{C}^2 = \mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$, and, by Lemma 3.4, $\Lambda(G_1) = \overleftrightarrow{e_1, e_2}$.

(ii) Let θ be a fixed real number. The abelian group G_2, generated by the projective transformations

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & i \\ 0 & e^{2\pi i\theta} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

acts properly and discontinuously on $\mathbb{C}^2 = \mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$ and, by Lemma 3.4, $\Lambda(G_2) = \overleftrightarrow{e_1, e_2}$.

(iii) Let θ be a fixed real number and $n_0 \in \mathbb{N}$ a fixed natural number. Denote by G_3 the abelian group G_3 generated by the projective transformations

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & i \\ 0 & e^{2\pi i\theta} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{2\pi i/n_0} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since G_2 is a finite-index subgroup of G_3, it follows that G_3 acts properly and discontinuously on $\mathbb{C}^2 = \mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$. By Lemma 3.4, $\Lambda(G_3) = \overleftrightarrow{e_1, e_2}$.

(iv) In an analogous way, the group generated by G_3 and the projective transformation

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

is a group whose Kulkarni limit set is equal to the line $\overleftrightarrow{e_1, e_2}$.

(v) If $\theta \in \mathbb{R}$ and $n_0 \in \mathbb{N}$ are fixed numbers, the reader can check that the group generated by the projective transformations

$$\begin{pmatrix} e^{\pi i/3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & e^{\pi i/3} \\ 0 & e^{2\pi i\theta} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{2\pi i/n_0} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is a group whose Kulkarni limit set is equal to the line $\overleftrightarrow{e_1, e_2}$.

4. Proof of Theorem 1.2

The main purpose of this section is to prove Theorem 1.2. First, we prove some technical results in order to achieve our goal. In what follows we assume that G does not contain any loxoparabolic elements.

Lemma 4.1. *Let G be a subgroup of* $\mathrm{PGL}(3, \mathbb{C})$ *acting properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$. If G does not contain any loxoparabolic elements and $L_0(G) \cup L_1(G) \subsetneq \ell$, then $L_2(G) = \ell$.*

Proof. Since $L_0(G) \cup L_1(G) \subsetneq \ell$, it follows that the group $\phi(G)$, where ϕ is defined as in (6), is a classical Kleinian group containing only parabolic or elliptic elements. Hence it acts as an elementary group on ℓ. It follows that this group has a fixed point in ℓ, so we can assume this fixed point is e_1, and it implies that every element in G can be represented by a matrix of the form

$$\begin{pmatrix} a & b & v \\ 0 & d & w \\ 0 & 0 & 1 \end{pmatrix}.$$

The kernel of the homomorphism $\phi : G \to \mathrm{PGL}(2, \mathbb{C})$ consists of all the transformations in G induced by matrices of the form

$$\begin{pmatrix} a & 0 & v \\ 0 & a & w \\ 0 & 0 & 1 \end{pmatrix}.$$

Since $L_0(G) \cup L_1(G) \subsetneq \ell$, every infinite order element in G has canonical form

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

It follows that every element in $\operatorname{Ker}\phi$ is elliptic, and so $\operatorname{Ker}\phi$ is finite. Moreover, it is cyclic because it is isomorphic to a finite subgroup of $\mathbb{S}^1$. This isomorphism is given by

$$\begin{pmatrix} a & 0 & v \\ 0 & a & w \\ 0 & 0 & 1 \end{pmatrix} \mapsto a.$$

Lemma 4.2. *Let G be a subgroup of* $\mathrm{PGL}(3,\mathbb{C})$ *acting properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$. If G does not contain any loxoparabolic elements and $L_0(G) \cup L_1(G) \subsetneq \ell$, then the matrix*

$$Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

induces the only pseudo-projective limit of the group G, also denoted by Q. In consequence, by Lemma 2.2, $L_2(G) \subset \ell$.

Proof. Let g_n be a sequence of distinct elements in G such that $g_n \to R$ when $n \to \infty$ (where R is a pseudo-projective transformation)

Case 1. The sequence contains a subsequence which consists only of parabolic elements.

In this case, we can assume that the elements are induced by matrices of the form

$$\begin{pmatrix} 1 & b_n & v_n \\ 0 & 1 & w_n \\ 0 & 0 & 1 \end{pmatrix},$$

but the set of all these transformations lying in G form a subgroup of G which is abelian because the commutator of two such elements is equal to the identity or to a parabolic element having a line of fixed points (and the group G does not contain this kind of parabolic elements). Moreover, the restriction of the homomorphism ϕ to this abelian subgroup is an isomorphism onto the "parabolic subgroup" of the Euclidean group $\{A(g) : g \in G\}$. Hence, this free abelian subgroup has rank at most two. If the group has rank one, then we can assume (conjugating by an upper triangular matrix) that this group is generated by an element induced by a matrix of the form

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

and it is not hard to check that the pseudo-projective limit R is induced by the matrix Q.

If the rank is equal to two, then we can assume (conjugating by an upper triangular matrix) that the parabolic group is generated by

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & \beta & \nu \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{pmatrix}, \quad \beta \in \mathbb{C} \setminus \mathbb{R}.$$

It is not hard to check, by means of an analysis of the general element $A^m B^n$ that any pseudo-projective limit must be induced by the matrix Q.

Case 2. The sequence (g_n) consists (except for finitely many values of n) of elliptic elements.

Since $G/\operatorname{Ker}\phi$ is a Euclidean group and $\operatorname{Ker}\phi$ is finite, there is an $n_0 \in \mathbb{N}$ such that for every elliptic element in G, $g^{n_0} = \mathrm{Id}$. Therefore, there exists a subsequence of (g_n), still denoted (g_n), and a diagonal matrix h such that $(h^{-1}g_n)$ is a sequence of distinct parabolic elements (because these transformations are represented by upper triangular matrices with ones in the diagonal entries). By Case 1 above, it follows that the pseudo-projective limit of $h^{-1}g_n$ is induced by the matrix Q. Therefore, the pseudo-projective limit of the sequence g_n is induced by the same matrix. □

Finally, if ℓ_0 is a line not intersecting $L_0(G) \cup L_1(G)$ and $g \in G$ is a parabolic element, then the family of compact sets $\{g^n(\ell_0)\}$ accumulates in ℓ. Therefore, $\ell \subset L_2(G)$. □

Proof of Theorem 1.2. The proof of (i) follows from Lemmas 3.3 and 3.4.

Now, for (ii), the group G contains a parabolic element of infinite order. Since G acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$, it follows that $L_0(G) \cup L_1(G) \subset \ell$.

Then, we consider two cases according to whether $L_0(G) \cup L_1(G) = \ell$ or $L_0(G) \cup L_1(G) \subsetneq \ell$.

If $L_0(G) \cup L_1(G) = \ell$, then, by Lemma 2.2, $L_2(G) \subset \ell$. Therefore, $\ell = \Lambda(G)$.

If $L_0(G) \cup L_1(G) \subsetneq \ell$, then by Lemma 4.1, $L_2(G) = \ell$. □

The next corollary follows from Lemma 4.2.

Corollary 4.3. *If $G \subset \mathrm{PGL}(3, \mathbb{C})$ acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$ and G does not contain any loxoparabolic elements, then one and only one of the following statements is verified*:

- *$L_0(G) \cup L_1(G) = \ell$ and $L_2(G) \subset \ell$, or*
- *$L_0(G) = L_1(G)$ is a point in ℓ and $L_2(G) = \ell$.*

Proposition 4.4. *If $G \subset \mathrm{PGL}(3, \mathbb{C})$ is a discrete subgroup such that $\Lambda(G)$ is equal to a line ℓ and G does not contain loxoparabolic elements, then for any finite extension G_1 of G (i.e., G is a finite-index normal subgroup of $G_1 \leq \mathrm{PGL}(3, \mathbb{C})$), $\Lambda(G_1) = \Lambda(G) = \ell$.*

Proof. (a) Since G is a finite-index normal subgroup of G_1, we have $L_0(G) = L_0(G_1)$ and $L_1(G_1) = g_1(L_1(G)) \cup \cdots \cup g_k(L_1(G))$, where $[g_1], \ldots, [g_k]$ are all the distinct cosets in G_1/G.

(b) If $L_0(G_1) = L_0(G)$ and $L_1(G_1) = L_1(G)$ then

$$L_2(G_1) = g_1(L_2(G)) \cup \cdots \cup g_k(L_2(G)),$$

where $[g_1], \ldots, [g_k]$ are all the distinct cosets in G_1/G.

(c) If $L_0(G) \cup L_1(G) = \ell$ then we consider two cases.

Case 1. The set $L_0(G)$ consists of a single point, called p. In this case, $L_0(G_1) \cup L_1(G_1)$ is a G_1-invariant set and it is a finite union of lines passing through p (the line ℓ is one of such lines). Let $g \in G_1$ be a parabolic element with p as its only fixed point. Then there exists $n_0 \in \mathbb{N}$ such that every line in $L_0(G_1) \cup L_1(G_1)$ is g^{n_0}-invariant. Since g^{n_0} is parabolic with a single fixed point, it has a single invariant line. Hence $L_0(G_1) \cup L_1(G_1) = \ell$.

Case 2. The set $L_0(G)$ contains more than one point. In this case, the set $L_0(G_1) = L_0(G)$ determines the line ℓ and is G_1-invariant. Then ℓ is G_1-invariant and

$$L_0(G_1) \cup L_1(G_1) = g_1(L_0(G) \cup L_1(G)) \cup \cdots \cup g_k(L_0(G) \cup L_1(G)) = \ell.$$

In any case, it is not hard to check that G_1 acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \ell$, and by Lemma 2.2 we obtain that $L_2(G_1) \subset \ell$.

(d) If $L_0(G) \cup L_1(G) \subsetneq \ell$, then by Corollary 4.3, $L_0(G) = L_1(G)$ is equal to one point. Thus we can assume that $L_2(G) = \ell$ and $L_0(G) = L_1(G) = \{e_1\}$. In this case, $L_0(G_1) = L_0(G) = \{e_1\}$ and it follows that every element in G_1 fixes e_1. Hence (by item (a)), $L_1(G_1) = L_1(G) = \{e_1\}$. It follows (by item (b)) that $L_2(G_1) = g_1(\ell) \cup \cdots \cup g_k(\ell)$, where $[g_1], \ldots, [g_k]$ are all distinct cosets in G_1/G. Therefore, $\Lambda(G_1) = \ell_1 \cup \cdots \cup \ell_k$, but there exists a parabolic element $g_0 \in G \subset G_1$ with Jordan canonical form equal to

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

Thus, every line ℓ_j, $j = 1, \ldots, k$, is invariant under some fixed power of g_0, but every power of g_0 has a single invariant line, and this line is equal to ℓ. We conclude that $\Lambda(G_1) = \ell$. □

5. Proof of Theorem 1.1

We recall that the proof of Theorem 1.1(i) follows from Lemmas 3.2 and 3.3.

Proof of Theorem 1.1(ii). Since G does not contain any element which acts as a parabolic element on ℓ, it follows that $\phi(G)\setminus\{\mathrm{Id}\}$ induces only elliptic elements, where ϕ is defined as in (6). We notice that the determinant of the matrix

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has modulus equal to one. Hence, every element in $\phi(G)$ can be written as a matrix of the form

$$\lambda\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \lambda\in\mathbb{S}^1 \quad \text{and} \quad ad-bc=1.$$

Since $\phi(G)$ induces a purely elliptic group acting on $\hat{\mathbb{C}}$, there is a global fixed point for the action of $\phi(G)$ on $\mathbb{H}^3_{\mathbb{R}}$. We can assume that this global fixed point is $(0, 0, 1)$ (in the upper half-space model). Then every element in $\phi(G)$ can be written as a matrix of the form

$$\lambda\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \qquad \lambda\in\mathbb{S}^1 \quad \text{and} \quad |a|^2+|b|^2=1.$$

Therefore, every element $g\in G$ can be written in the form

$$g=\begin{pmatrix} A & \mathbf{v} \\ 0 & 1 \end{pmatrix}, \qquad \text{where } A\in U(2) \quad \text{and} \quad \mathbf{v}, 0^{\mathrm{T}}\in\mathbb{C}^2.$$

Hence, G is a discrete subgroup of isometries of $\mathbb{R}^4$. By Theorem 5.4.5 in [Ratcliffe 1994], G contains a normal finite-index free abelian subgroup of rank less than or equal to four. □

Remark 5.1. If $\phi(G)\setminus\{\mathrm{Id}\}$ induces only elliptic elements, then G is conjugate in $\mathrm{PGL}(3,\mathbb{C})$ to some group such that every element in a finite-index subgroup is induced by an upper triangular matrix.

Examples. (i) Let G be the group generated by the two matrices

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & e^{2\pi i\theta_1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & i \\ 0 & e^{2\pi i\theta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where θ_1,θ_2 are fixed real numbers. It is not hard to check that G acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}}\setminus\overleftrightarrow{e_1, e_2}$. It follows from Theorem 1.2(ii) that $\Lambda(G)=\overleftrightarrow{e_1, e_2}$. In this case $\phi(G)$ does not contain any parabolic elements.

(ii) Let G_1 be the group generated by the matrices

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{pmatrix}.$$

It is not hard to check that G_1 acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$. It follows from Theorem 1.2(ii) that $\Lambda(G_1) = \overleftrightarrow{e_1, e_2}$. In this case $\phi(G_1)$ consists of the identity element.

(iii) If k is a fixed natural number, then the group G_2 generated by the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

acts properly and discontinuously on $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$ and the Kulkarni limit set $\Lambda(G_2)$ is equal to $\overleftrightarrow{e_1, e_2}$ by Theorem 1.2(ii). In this case, $\phi(G_2)$ is a cyclic group generated by a parabolic element. We notice that G_2 is a 2-step nilpotent group, so it is a uniform lattice of $H_3 \times \mathbb{R}$, where H_3 is the 3-dimensional real Heisenberg group (see [Dekimpe 1996, Corollary 6.2.5]). If Γ is a subgroup of G_2 such that $0 < \operatorname{rank} \Gamma < 4$ then $\Lambda(\Gamma) = \overleftrightarrow{e_1, e_2}$. However, the quotient of $\mathbb{P}^2_{\mathbb{C}} \setminus \overleftrightarrow{e_1, e_2}$ by Γ is not compact.

In what follows we develop some tools that will be useful in the proof of Theorem 1.1(iii).

Proposition 5.2. *If $G \subset \mathrm{PGL}(3, \mathbb{C})$ is a discrete subgroup such that $\Lambda(G)$ is equal to a line ℓ, G does not contain any loxoparabolic elements and G contains an element which acts as a parabolic element on ℓ, then G is conjugate in* $\mathrm{PGL}(3, \mathbb{C})$ *to some group such that every element in it is induced by a matrix of the form*

$$\begin{pmatrix} a & b & v \\ 0 & d & w \\ 0 & 0 & 1 \end{pmatrix}, \quad |a| = |d| = 1. \tag{10}$$

(If the group G does contain any loxoparabolic elements, then the statement is still valid with the exception that the eigenvalues are not unitary complex numbers. See Lemma 3.1)

Proof. Since the set $\phi(G)$ contains a parabolic element, we can assume (conjugating, if needed, by an element that preserves the line ℓ) that such parabolic element is induced by a matrix of the form

$$A_1 = \mu \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{for some } \mu \in \mathbb{C}.$$

If we assume there is a matrix

$$A_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{such that } c \neq 0,$$

which induces an element in $\phi(G)$, then $\mathrm{Tr}^2(A_1^n A_2 A_1^{-n} A_2^{-1}) = (2+(cn)^2)^2$. Thus, $A_1^n A_2 A_1^{-n} A_2^{-1}$ is a loxodromic element for n large enough, and it contradicts that $\phi(G)$ contains only elliptic or parabolic elements. $\square$

Lemma 5.3. *Let V be an abelian subgroup of $\mathbb{C}^2$ of rank $r = 1, 2, 3$ or 4 and u an m-th root of unity that generates a cyclic group of order m. Assume that $uv \in V$ whenever $v \in V$. Then $\varphi(m) \leq r$ (where φ denotes the Euler's totient function). Hence, $m \leq 12$.*

Proof. This proof is contained in the proof of Theorem 4.1 in [Fillmore and Scheuneman 1973], and we include it here for reader's convenience. Let $v_1, \ldots, v_r$ be a basis of V. Expressing uv_i in terms of this basis and taking a determinant, we obtain a polynomial of degree r with integer coefficients which is satisfied by u. Hence the field generated by u over the rationals is of degree at most r. This field is generated by a primitive m-th root of unity, so it has degree $\varphi(m)$, where φ is Euler's totient. Thus $\varphi(m) \leq r$. It follows that $m = 1, 2, 3, 4, 5, 6, 8, 10$ or 12. $\square$

The hypothesis in the lemma above can be slightly modified to obtain the following:

Lemma 5.4. *Let V be an abelian subgroup of $\mathbb{C}^2$ of rank $r = 1, 2$ or 3 and $u \in \mathbb{S}^1$. Assume $uv \in V$ whenever $v \in V$. Then u is a root of unity of order $1, 2, 3, 4$ or 6.*

The proof is almost the same as the one given for Lemma 5.3. One just uses the fact that $u \in \mathbb{S}^1$ is a root of unity whenever there is a monic polynomial of degree $r \leq 3$ with integer coefficients which is satisfied by u.

Remark 5.5. Lemmas 5.3 and 5.4 can be applied to abelian subgroups of

$$(\mathbb{C}, +) \cong (\{0\} \times \mathbb{C}, +) \cong (\mathbb{C} \times \{0\}, +).$$

If we assume that every element in G can be represented by an upper triangular matrix of the form (10) and G contains an irrational ellipto-parabolic element then we can assume that it is induced by one of the following two upper triangular matrices:

$$E_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & d_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, E_2 = \begin{pmatrix} a_0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad a_0 = d_0 = e^{2\pi i \theta},\ \theta \in \mathbb{R} \setminus \mathbb{Q}.$$

Proposition 5.6. *Let G be a discrete subgroup of $\mathrm{PGL}(3, \mathbb{C})$ such that every element can be induced by a matrix of the form (10) and $\Lambda(G) = \ell$. If G contains the irrational ellipto-parabolic element induced by E_1, then the normal subgroup N_1*

consisting of the elements of the form

$$\begin{pmatrix} 1 & b & v \\ 0 & d & w \\ 0 & 0 & 1 \end{pmatrix}$$

is a (finitely generated) finite-index abelian subgroup. In fact, $b = w = 0$ for every element in N_1.

Proof. Every element in the subgroup N_1 commutes with E_1 (by Lemma 3.3 in [Fillmore and Scheuneman 1973]), and hence it has the form

$$h = \begin{pmatrix} 1 & 0 & v(h) \\ 0 & d(h) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{11}$$

Therefore, N_1 is a finitely generated abelian subgroup. Moreover, the set

$$V = \{\nu \in \mathbb{C} \mid \exists h \in N_1 \text{ with } v(h) = \nu\}$$

is an abelian subgroup of $(\mathbb{C}, +)$.

Now, let g be any element in G. By considering the upper right entry of ghg^{-1}, it is possible to check that $a(g)\nu \in V$ whenever $\nu \in V$. It follows, by Lemma 5.4, that $a(g)$ is a root of unity of order $1, 2, 3, 4$ or 6. Therefore N_1 has finite index in G. Since every element in the discrete abelian group N_1 has the form (11), it is not hard to check that N_1 is finitely generated and its rank is less or equal to two. □

Proposition 5.7. *Let G be a discrete subgroup of* $\mathrm{PGL}(3, \mathbb{C})$ *such that every element can be induced by a matrix of the form (10) and $\Lambda(G) = \ell$. If G contains the irrational ellipto-parabolic element induced by E_2, then the normal subgroup N_2 consisting of those elements of the form*

$$\begin{pmatrix} a & b & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix}$$

is a (finitely generated) finite-index abelian subgroup. In fact, $b = v = 0$ for every element in N_2.

Proof. Let us denote by L the set consisting of those $(\beta, \nu) \in \mathbb{C}^2$ for which there exists an element in G of the form

$$\begin{pmatrix} 1 & \beta & \nu \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

It is not hard to check that L is a discrete subgroup of $(\mathbb{C}^2, +)$. If we assume that $b \neq 0$ for some

$$f = \begin{pmatrix} a & b & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix} \in N_2,$$

then $L \neq \mathbf{0}$ because the commutator of E_2 and f is equal to

$$[E_2, f] = \begin{pmatrix} 1 & (a_0 - 1)b & (a_0 - 1)v - (a_0 - 1)bw - a_0 b \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If $(\beta, \nu) \in L$ then $(a_0\beta, a_0\nu - a_0\beta) \in L$ (conjugate the element of the form (12) by E_2). In other words, the $\mathbb{C}$-linear map induced by

$$\begin{pmatrix} a_0 & 0 \\ -a_0 & a_0 \end{pmatrix}$$

preserves L. If we assume rank $L = 4$, then a_0 is a root of a polynomial with integer coefficients of the form $p(x) = (x^2 - 2x\cos(2\pi\theta) + 1)^2$ and it implies that a_0 is a root of unity, which is a contradiction. Therefore, rank $L \leq 3$. Hence, applying Lemma 5.4 to the abelian group L and to the unitary complex number a_0, we obtain that a_0 is a root of unity, a contradiction. It follows that $b = 0$ for all $f \in N_2$.

Analogously, it can be proved that $v = 0$ for all $f \in N_2$. Thus every $f \in N_2$ has the form

$$\begin{pmatrix} a & 0 & 0 \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix},$$

and it follows that N_2 is an abelian group of rank less than or equal to two.

Let W denote the subset of $\mathbb{C}$ consisting of those $\omega \in \mathbb{C}$ such that there exist $f \in N_2$ of the form

$$f = \begin{pmatrix} a & 0 & 0 \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{pmatrix}.$$

We notice that W is an abelian discrete subgroup of $\mathbb{C}$ and $W \neq 0$ because $E_2 \in N_2$.

If $g \in G$ is an arbitrary element, then it is not hard to check that $d(g)\omega \in W$ whenever $\omega \in W$. It follows, by Lemma 5.4, that $d(g)$ is a root of unity of order 1, 2, 3, 4 or 6. Therefore N_2 has finite index in G. □

Proposition 5.8. *Let G be a discrete subgroup of* $\mathrm{PGL}(3, \mathbb{C})$ *such that* $\Lambda(G) = \ell$ *and every element $g \in G$ is induced by a matrix of the form* (10). *Assume that for every $g \in G$, there exists $n \in \mathbb{N}$, depending on g, such that $a(g), d(g)$ are n-th roots* *of unity. If N denotes the subset of G consisting of those elements in G induced by matrices of the form*

$$\begin{pmatrix} 1 & b & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix},$$

then N is a finite-index torsion free normal subgroup of G (it is also finitely generated).

Proof. Let B be the abelian subgroup of $\mathbb{C}$ that consists of all those $\beta \in \mathbb{C}$ for which there exists an element in N of the form

$$\begin{pmatrix} 1 & \beta & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix}.$$

Analogously, let W be the abelian subgroup of $\mathbb{C}$ consisting of those $\omega \in \mathbb{C}$ such that exists an element in N of the form

$$\begin{pmatrix} 1 & b & v \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, let V be the subset of $\mathbb{C}$ (it is not necessarily a subgroup) consisting of those $\nu \in \mathbb{C}$ such that there exists an element in G of the form

$$\begin{pmatrix} 1 & b & \nu \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix}.$$

Since G acts properly and discontinuously on $\mathbb{C}^2$, so does N. Moreover, the nilpotent group N acts freely on $\mathbb{C}^2$, so N is generated by at most four elements (see [Cartan and Eilenberg 1999] and [Fillmore and Scheuneman 1973]). It follows that rank $B \leq 4$ and rank $W \leq 4$. Moreover, when V is an abelian group, rank $V \leq 4$.

We consider the following cases:

Case 1. $B \neq 0$ and $W \neq 0$.

In this case, by Lemma 5.3 applied to W and to the root of unity $d(g)$, there exists $n_1 \in \mathbb{N}$ (not depending on g) such that for every $g \in G$, $d(g)^{n_1} = 1$.

Similarly, by Lemma 5.3 applied to B and to the root of unity $a(g)/d(g)$, there exists $n_2 \in \mathbb{N}$ (not depending on g) such that for every $g \in G$, $(a(g)/d(g))^{n_2} = 1$.

Therefore, there exists $n_0 \in \mathbb{N}$ such that $g^{n_0} \in N$ for every $g \in G$.

Case 2. $B \neq 0$, $W = 0$ and $V \neq 0$.

In this case, by Lemma 5.3 applied to the abelian group V and to the root of unity $a(g)/d(g)$, there exists $n_1 \in \mathbb{N}$ such that $(a(g)/d(g))^{n_1} = 1$ for every $g \in G$.

Let us fix an element

$$g_0 = \begin{pmatrix} a_0 & b_0 & v_0 \\ 0 & d_0 & w_0 \\ 0 & 0 & 1 \end{pmatrix} \in G,$$

with $d_0 \neq 1$ (if such an element does not exist then the proof ends). If $g \in G$ is an arbitrary element, then the commutator $[g_0, g]$ has the form

$$\begin{pmatrix} 1 & \beta & \nu \\ 0 & 1 & (d_0-1)w-(d-1)w_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then $(d_0-1)w-(d-1)w_0 = 0$. Therefore, conjugating G by the projective transformation induced by

$$h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & d_0-1 & w_0 \\ 0 & 0 & 1 \end{pmatrix},$$

we can assume that $w = 0$ for all $g \in G$.

Now, if $\nu \in V$ and g is an arbitrary element in G as above, then $a\nu \in V$ (just conjugate

$$\begin{pmatrix} 1 & \beta & \nu \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

by g). It follows, by Lemma 5.3, that there exists $n_2 \in \mathbb{N}$ such that $a^{n_2} = 1$ for every $g \in G$. Therefore, there exists $n_0 \in \mathbb{N}$ such that $g^{n_0} \in N$ for every $g \in G$.

Case 3. $B \neq 0$, $W = 0$ and $V = 0$.

This case cannot happen, otherwise the Kulkarni limit set of N (and the Kulkarni limit set of G) would be equal to $\overleftrightarrow{e_1, e_3}$.

Case 4. $B = 0$, $W \neq 0$ and $V \neq 0$.

In this case, V is an abelian subgroup of $\mathbb{C}$. By Lemma 5.3, there exists $n_1 \in \mathbb{N}$ such that $d^{n_1} = 1$ for every $g \in G$. Let us fix an element

$$g_0 = \begin{pmatrix} a_0 & b_0 & v_0 \\ 0 & d_0 & w_0 \\ 0 & 0 & 1 \end{pmatrix} \in G,$$

with $a_0 \neq d_0$ (if such element does not exist then the proof ends).

If $g \in G$ is an arbitrary element, then the commutator $[g_0, g]$ has the form

$$\begin{pmatrix} 1 & \frac{b(a_0-d_0)-b_0(a-d)}{dd_0} & \nu \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{pmatrix}.$$

Then $[g_0, g] \in N$, so $b(a_0 - d_0) - b_0(a - d) = 0$.

Hence, we can assume, conjugating G by

$$h = \begin{pmatrix} a_0 - d_0 & b_0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

that $b = 0$ for all $g \in G$.

It is not hard to check that if $g \in G$, and $\nu \in V$, then $a\nu \in V$ (just conjugate by g). Lemma 5.3 implies that there exists $n_2 \in \mathbb{N}$ such that $a^{n_2} = 1$ for every $g \in G$.

Case 5. $B = 0$, $W \neq 0$ and $V = 0$.

In this case, W is an abelian subgroup of $\mathbb{C}$ and it is not hard to check that there exists $n_1 \in \mathbb{N}$ such that $d^{n_1} = 1$ for every $g \in G$. Now we consider the normal subgroup N_1 consisting of all those elements in G of the form

$$\begin{pmatrix} a & b & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to check by straightforward computation that the commutator of any two elements in G has the form

$$\begin{pmatrix} 1 & \beta & \nu \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since $B = V = 0$, it follows that N_1 is an abelian group.

We can assume, conjugating by a suitable matrix, that every element in N_1 has the form

$$\begin{pmatrix} a & 0 & 0 \\ 0 & 1 & w \\ 0 & 0 & 1 \end{pmatrix}.$$

It follows that N_1 is a (finitely generated) finite-index abelian subgroup of G.

Case 6. $B = 0$, $W = 0$ and $V \neq 0$.

In this case, $V \neq 0$ is an abelian subgroup of $\mathbb{C}$ and there is a fixed natural number $n_1 \in \mathbb{N}$ such that $a^{n_1} = 1$ for every $g \in G$.

Now, let us denote by N_2 the normal subgroup of G consisting of those elements of the form

$$\begin{pmatrix} 1 & b & v \\ 0 & d & w \\ 0 & 0 & 1 \end{pmatrix}.$$

Let us fix an element

$$h_0 = \begin{pmatrix} 1 & b_0 & v_0 \\ 0 & d_0 & w_0 \\ 0 & 0 & 1 \end{pmatrix} \in N_2, \quad d_0 \neq 1,$$

(if $d = 1$ for every element in N_2, then the proof ends). If

$$h = \begin{pmatrix} 1 & b & v \\ 0 & d & w \\ 0 & 0 & 1 \end{pmatrix} \in N_2,$$

then $[h, h_0] \in N$. Hence

$$(1-d)b_0 + (1-d_0)b = 0 = (1-d)w_0 + (1-d_0)w.$$

Thus, conjugating by

$$\begin{pmatrix} 1 & b_0/(1-d_0) & 0 \\ 0 & d_0 - 1 & w_0 \\ 0 & 0 & 1 \end{pmatrix},$$

we can assume that every element in N_2 has the form

$$h = \begin{pmatrix} 1 & 0 & v \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It follows that N_2 is a (finitely generated) finite-index abelian subgroup of G.

Case 7. $B = 0$, $W = 0$ and $V = 0$.

In this case, every element in G is of finite order, so G is finite. □

Proof of Theorem 1.1(iii).

If G contains the ellipto-parabolic element E_1 then Proposition 5.6 implies the group N_1 is a finite-index normal abelian subgroup, and no element in N_1 acts on ℓ as a parabolic element. If we assume that there exists $g \in G$ such that $\phi(g)$ acts on ℓ as a parabolic element then for some $n_0 \in \mathbb{N}$, we have that $\phi(g)^{n_0} \in \phi(N_1)$ acts as a parabolic element on ℓ, which is a contradiction. A similar argument, using Proposition 5.6, shows that G cannot contain the ellipto-parabolic element E_2.

The second part is analogous to the proof of Theorem 4.1 in [Fillmore and Scheuneman 1973], and we include it here for reader's convenience. By Proposition 5.8,

G is a finite extension of a discrete unipotent group N, and by Theorem 5.4.3 in [Corwin and Greenleaf 1990], there exists a unique Lie unipotent group H such that H/N is compact (in consequence, N is finitely generated). This group H is necessarily simply connected, and in fact it is a Euclidean space.

Since N acts properly discontinuously (and freely) on $\mathbb{C}^2$, it follows that the projective dimension of the integer group ring of N is less or equal to four (see [Cartan and Eilenberg 1999]). Moreover, the dimension of H is less or equal to four.

The nilpotent simply connected Lie groups of dimension four are $\mathbb{R}^4$ and $H_3 \times \mathbb{R}$, where H_3 denotes the real Heisenberg group. The nilpotent simply connected Lie groups of dimension three are H_3 and $\mathbb{R}^3$. Finally, those nilpotent simply connected Lie groups of dimension two and one are $\mathbb{R}^2$ and $\mathbb{R}$.

The discrete subgroups with compact quotient of $H_3 \times \mathbb{R}$ are of the form

$$\Delta_k = \langle A, B, C, D : C, D \text{ are central and } [A, B] = C^k \rangle,$$

where $k \in \mathbb{N}$. The discrete subgroups of H_3 with compact quotient are of the form

$$\Gamma_k = \langle A, B, C : C \text{ is central and } [A, B] = C^k \rangle,$$

where $k \in \mathbb{N}$. □

Acknowledgements. We would like to thank Professor John Parker for enlightening discussions during the process of this work and the kind hospitality received by the three authors at Durham University. Also, we would like to thank the referee for his helpful suggestions.

References

[Barrera Vargas et al. 2011] W. d. J. Barrera Vargas, A. C. Cordero, and J. P. N. Carrillo, "The limit set of discrete subgroups of PSL(3, $\mathbb{C}$)", *Math. Proc. Cambridge Philos. Soc.* **150**:1 (2011), 129–146. MR 2012b:32037 Zbl 1214.30032

[Cano and Seade 2010] A. Cano and J. Seade, "On the equicontinuity region of discrete subgroups of PU(1, n)", *J. Geom. Anal.* **20**:2 (2010), 291–305. MR 2011c:32045 Zbl 1218.37059

[Cano and Seade 2014] A. Cano and J. Seade, "On discrete groups of automorphisms of $\mathbb{P}^2_{\mathbb{C}}$", *Geom. Dedicata* **168** (2014), 9–60. MR 3158029 Zbl 06269858

[Cano et al. 2013] A. Cano, J. P. Navarrete, and J. Seade, *Complex Kleinian groups*, Progress in Mathematics **303**, Birkhäuser, Basel, 2013. MR 2985759 Zbl 1267.30001

[Cartan and Eilenberg 1999] H. Cartan and S. Eilenberg, *Homological algebra*, Princeton University Press, 1999. MR 2000h:18022 Zbl 0933.18001

[Corwin and Greenleaf 1990] L. J. Corwin and F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications, I: Basic theory and examples*, Cambridge Studies in Advanced Mathematics **18**, Cambridge University Press, 1990. MR 92b:22007 Zbl 0704.22007

[Dekimpe 1996] K. Dekimpe, *Almost–Bieberbach groups: Affine and polynomial structures*, Lecture Notes in Mathematics **1639**, Springer, Berlin, 1996. MR 2000b:20066 Zbl 0865.20001

[Fillmore and Scheuneman 1973] J. P. Fillmore and J. Scheuneman, "Fundamental groups of compact complete locally affine complex surfaces", *Pacific J. Math.* **44** (1973), 487–496. MR 48 #6474 Zbl 0259.32008

[Kulkarni 1978] R. S. Kulkarni, "Groups with domains of discontinuity", *Math. Ann.* **237**:3 (1978), 253–272. MR 81m:30046 Zbl 0369.20028

[Marden 2007] A. Marden, *Outer circles: An introduction to hyperbolic 3-manifolds*, Cambridge University Press, 2007. MR 2008i:57001 Zbl 1149.57030

[Maskit 1988] B. Maskit, *Kleinian groups*, Grundlehren der Mathematischen Wissenschaften **287**, Springer, Berlin, 1988. MR 90a:30132 Zbl 0627.30039

[Navarrete 2008] J.-P. Navarrete, "The trace function and complex Kleinian groups in $\mathbb{P}^2_{\mathbb{C}}$", *Internat. J. Math.* **19**:7 (2008), 865–890. MR 2009g:32056 Zbl 1167.30025

[Ratcliffe 1994] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Mathematics **149**, Springer, New York, 1994. MR 95j:57011 Zbl 0809.51001

[Scheuneman 1974] J. Scheuneman, "Fundamental groups of compact complete locally affine complex surfaces, II", *Pacific J. Math.* **52** (1974), 553–566. MR 54 #612 Zbl 0291.32029

[Suwa 1975] T. Suwa, "Compact quotient spaces of $\mathbf{C}^2$ by affine transformation groups", *J. Differential Geometry* **10** (1975), 239–252. MR 51 #5977 Zbl 0311.57025

Received September 26, 2013. Revised December 17, 2013.

WALDEMAR BARRERA
FACULTAD DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA DE YUCATÁN
ANILLO PERIFÉRICO NORTE TABLAJE CAT
13615 MÉRIDA, YUCATÁN
MÉXICO
bvargas@uady.mx

ANGEL CANO
INSTITUTO DE MATEMÁTICAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
AV. UNIVERSIDAD S/N
COL. LOMAS DE CHAMILPA
62210 CUERNAVACA, MORELOS
MÉXICO
angel@matcuer.unam.mx

JUAN PABLO NAVARRETE
FACULTAD DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA DE YUCATÁN
ANILLO PERIFÉRICO NORTE TABLAJE CAT
13615 MÉRIDA, YUCATÁN
MÉXICO
jp.navarrete@uady.mx

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.305

A NOTE ON FLUX INTEGRALS OVER SMOOTH REGULAR DOMAINS

IDO BRIGHT AND JOHN M. LEE

We provide new bounds on a flux integral over the portion of the boundary of one regular domain contained inside a second regular domain, based on properties of the second domain rather than the first one. This bound is amenable to numerical computation of a flux through the boundary of a domain, for example, when there is a large variation in the normal vector near a point. We present applications of this result to occupational measures and two-dimensional differential equations, including a new proof that all minimal invariant sets in the plane are trivial.

1. Introduction

A *regular domain* in $\mathbb{R}^d$ is a closed, embedded d-dimensional smooth submanifold with boundary, such as a closed ball or a closed half-space. (Throughout this paper, *smooth* means infinitely differentiable.) If $D \subset \mathbb{R}^d$ is a regular domain, its interior $\mathring{D}$ is an open subset of $\mathbb{R}^d$, and its boundary ∂D is a closed, embedded, codimension-1 smooth submanifold (without boundary) which is the common topological boundary of the open sets $\mathring{D}$ and $\mathbb{R}^d \setminus D$. For this reason, the boundary of a regular domain is often called a *space-separating hypersurface*. The Jordan–Brouwer separation theorem (see, for example, [Guillemin and Pollack 1974, p. 89]) shows that if $S \subset \mathbb{R}^d$ is any compact, connected, embedded hypersurface, then the complement of S has two connected components, one bounded (the *interior* of S) and another unbounded (the *exterior* of S), with S as their common boundary; thus $S \cup \operatorname{Int} S$ and $S \cup \operatorname{Ext} S$ are both regular domains. But in general, the boundary of a regular domain need not be connected (for example, an annulus in the plane).

Surface integrals computing the flux through boundaries of regular domains are ubiquitous in physics and engineering. We present two bounds for surface integrals on a portion of the boundary of one domain contained inside a second domain. The results are presented for regular domains in Euclidean space for simplicity, but Theorems 1.1 and 1.2 extend to regular domains in Riemannian manifolds. See Theorem 3.3. For more details about the notation in these theorems, see Section 2.

MSC2010: primary 53A05, 58C35; secondary 28A99.
Keywords: flux integral, smooth, regular domain, occupational measure.

Theorem 1.1. *Suppose D_1, $D_2 \subset \mathbb{R}^d$ are regular domains, such that $D_1 \cap D_2$ is compact and D_2 has finite volume and surface area. Suppose f is a smooth vector field defined on a neighborhood of D_2 such that $|f|$ and $|\nabla \cdot f|$ are bounded. Then the absolute value of the flux of f across the portion of ∂D_1 inside D_2 satisfies*

$$\left| \int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1} \, dA \right| \le \operatorname{Area}(\partial D_2) \| f \|_\infty + \operatorname{Vol}(D_2) \| \nabla \cdot f \|_\infty. \tag{1-1}$$

When the vector field is divergence-free, we have the following much better bound.

Theorem 1.2. *Suppose D_1, $D_2 \subset \mathbb{R}^d$ are regular domains with compact intersection and finite surface areas, and f is a smooth bounded vector field on $\mathbb{R}^d$ satisfying $\nabla \cdot f \equiv 0$. Then*

$$\left| \int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1} \, dA \right| \le \tfrac{1}{2} \operatorname{Area}(\partial D_2) \| f \|_\infty. \tag{1-2}$$

A surprising corollary to Theorem 1.2 bounds the integral of the normal vector of the portion of a hypersurface contained inside a second regular domain.

Corollary 1.3. *Suppose D_1, $D_2 \subset \mathbb{R}^d$ are regular domains with compact intersection and finite surface areas. The following inequality holds*:

$$\left| \int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1} \, dA \right| \le \tfrac{1}{2} \operatorname{Area}(\partial D_2). \tag{1-3}$$

When D_2 is convex we have the following alternative bound, which is an improvement in some cases.

Theorem 1.4. *Suppose D_1, $D_2 \subset \mathbb{R}^d$ are regular domains. If D_2 is compact and convex with diameter δ, then*

$$\left| \int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1} \, dA \right| \le \tfrac{1}{2} \operatorname{Vol}(B^{d-1}(\delta/2)), \tag{1-4}$$

where $B^{d-1}(\delta/2)$ denotes the ball in $\mathbb{R}^{d-1}$ of radius $\delta/2$.

The significance of these results is that, although the integration is with respect to the portion of ∂D_1 inside D_2, which might have arbitrarily large surface area (see Figure 1), the bound depends only on D_2. This is due to the cancellations of the normal vector that occur in hypersurfaces that bound regular domains, and would not hold for images of general immersions of codimension 1 (see Example 4.2).

Theorem 1.1 is applicable to the numerical computation of the flux on the surface of a regular domain when there is a large variation of the normal vector near a point, resulting in a large surface area contained in a region of small volume. Indeed, the flux over the problematic part can be estimated by finding a domain containing it, avoiding direct computation. We provide an application of Corollary 1.3 in

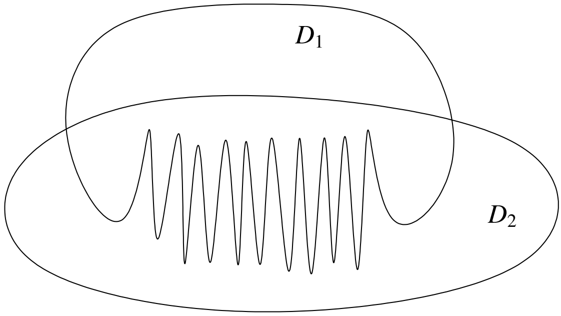


Figure 1. The setup for Theorems 1.1 and 1.2.

Section 5, for limits of sequences of regular domains with surface area increasing without bound; there we use the bound to show that in the limit, the average normal vector, say in a ball, is zero. Such a result is applied in the case $d = 2$, in [Artstein and Bright 2010], to obtain a new Poincaré–Bendixson type result for planar infinite-horizon optimal control.

Corollary 1.3 generalizes a previous result, for $d = 2$, established by Artstein and Bright [2010; 2013]. This topological result has proved fruitful in applications, providing new Poincaré–Bendixson type results, in an optimal-control setting [Artstein and Bright 2010; Bright 2012], and in the context of dynamics with no differentiability assumptions by Bright [2012]. The proofs of the planar result in [Artstein and Bright 2010; 2013] employ a dynamical argument, which is similar to the one used in the textbook proof of the Poincaré–Bendixson theorem. In this paper, we generalize the results to boundaries of open sets, restricting ourselves in this presentation to regular domains; however the results hold for more general sets and vector fields. The results in their fullest generality for nonsmooth domains and fluxes are presented in [Bright and Torres 2014].

Remark 1.5. The requirement that $D_1 \cap D_2$ be compact is essential, as it implies that $\partial D_1 \cap D_2$ is compact, so that the integrals in (1-1)–(1-3) are finite.

Remark 1.6. Theorem 1.1 can be extended, by replacing the smooth vector field f with a smooth matrix-valued function Π, using the induced norm.

Remark 1.7. For simplicity, Theorem 1.2 is stated under the assumption that f is defined on all of $\mathbb{R}^d$; but as the proof will show, if D_2 has finite volume it is only necessary that f be defined on some neighborhood of D_2.

The structure of this paper is as follows. Section 2 presents notation and lemmas used in the paper. In Section 3 we prove Theorems 1.1 and 1.2, and describe how our results extend to regular domains in a Riemannian manifold. In Section 4 we prove Corollary 1.3 and Theorem 1.4, and also provide examples showing the tightness of the bound. In the last section we provide three applications of Corollary 1.3: an application to limits of sequences of regular domains; an extension when $d = 2$; and a simplified proof of a theorem on invariant sets for dynamical systems.

2. Notation and lemmas

Throughout this paper, we denote the characteristic function of a set $A \subset \mathbb{R}^d$ by χ_A. The d-dimensional volume is denoted by $\mathrm{Vol}(A)$, and the $(d-1)$-dimensional surface area of its boundary by $\mathrm{Area}(\partial A)$. Given two submanifolds $S_1, S_2 \subset \mathbb{R}^d$, the notation $S_1 \pitchfork S_2$ means that S_1 and S_2 intersect transversally. The Euclidean norm on $\mathbb{R}^d$ is denoted by $|\cdot|$, and the supremum norm on functions by $\|\cdot\|_\infty$. The divergence of a smooth vector field $f = (f^1, f^2, \dots, f^d)$ at the point $x = (x^1, x^2, \dots, x^d) \in \mathbb{R}^d$ is denoted by

$$\nabla \cdot f(x) = \frac{\partial}{\partial x^1} f^1(x) + \frac{\partial}{\partial x^2} f^2(x) + \cdots + \frac{\partial}{\partial x^d} f^d(x).$$

The following is a simple lemma we need for the proof of the main theorems.

Lemma 2.1. *Suppose (X, μ) is a measure space, $U, V \subset X$, and U has finite measure. For every real-valued function $f \in L^\infty(X)$, we have*

$$\left| \int_{U \setminus V} f(x)\mu(dx) \right| \le \frac{1}{2}\left(\mu(U)\|f\|_\infty + \left| \int_U f(x)\mu(dx) \right| \right),$$
$$\left| \int_{U \cap V} f(x)\mu(dx) \right| \le \frac{1}{2}\left(\mu(U)\|f\|_\infty + \left| \int_U f(x)\mu(dx) \right| \right).$$

Proof. The first inequality follows from the triangle inequality:

$$\begin{aligned}
\int_U \left| f(x) \right| \mu(dx) &\ge \left| \int_{U \setminus V} f(x)\mu(dx) \right| + \left| \int_{U \cap V} f(x)\mu(dx) \right| \\
&= \left| \int_{U \setminus V} f(x)\mu(dx) \right| + \left| \int_{U \setminus V} f(x)\mu(dx) - \int_U f(x)\mu(dx) \right| \\
&\ge 2\left| \int_{U \setminus V} f(x)\mu(dx) \right| - \left| \int_U f(x)\mu(dx) \right|.
\end{aligned}$$

The second inequality follows by replacing V with $X \setminus V$. □

The proofs of the main theorems are based on the divergence theorem for certain domains in $\mathbb{R}^d$. Let us say a *regular domain with corners* in $\mathbb{R}^d$ is a closed subset $D \subset \mathbb{R}^d$ such that for each point $p \in D$, there exist an open set $U \subset \mathbb{R}^d$ containing p and a smooth coordinate chart $\varphi : U \to \mathbb{R}^d$ such that $\varphi(U \cap D)$ is the intersection of $\varphi(U)$ with $\overline{\mathbb{R}}^d_+ = \{x \in \mathbb{R}^d \mid x^1 \ge 0, \dots, x^d \ge 0\}$. Some typical examples are closed simplices and rectangular solids. Every regular domain is a regular domain with corners, and a regular domain with corners is a d-dimensional smooth manifold with corners in the sense defined in [Lee 2013].

Here is the version of the divergence theorem we will use.

Lemma 2.2. *Suppose $D \subset \mathbb{R}^d$ is a regular domain with corners, having finite volume and surface area. If f is a smooth vector field defined on D such that both $|f|$ and $|\nabla\cdot f|$ are bounded, then*

$$\int_{\partial D} f\cdot \boldsymbol{n}_{\partial D}\, dA = \int_D \nabla\cdot f\, dV.$$

Proof. If D is compact, or more generally if f is compactly supported, this follows immediately from Stokes's theorem applied to the $(d-1)$-form $f \lrcorner (dx^1 \wedge \cdots \wedge dx^d)$, where $\lrcorner$ denotes interior multiplication. (For Stokes's theorem on manifolds with corners, see, for example, [Lee 2013, Theorem 16.25, p. 419].) In the general case, we argue as follows. Let $\varphi : [0, \infty) \to [0, 1]$ be a smooth function that is equal to 1 on $\left[0, \frac{1}{2}\right]$ and supported in $[0, 1]$, and for each $r > 0$ let $\varphi_r(x) = \varphi(|x|^2/r^2)$. Then the vector field $\varphi_r f$ is compactly supported, so the divergence theorem implies

$$\int_{\partial D} \varphi_r f\cdot \boldsymbol{n}_{\partial D}\, dA = \int_D \nabla\cdot(\varphi_r f)\, dV. \tag{2-1}$$

As $r \to \infty$, the integral on the left-hand side of (2-1) converges to $\int_{\partial D} f\cdot \boldsymbol{n}_{\partial D}\, dA$ by the dominated convergence theorem. On the other hand, for each $r > 0$,

$$\begin{aligned} |\nabla\cdot(\varphi_r f)(x)| &= \left|\varphi_r(x)\nabla\cdot f(x) + \frac{2}{r^2}\sum_{i=1}^{d} \varphi'\left(\frac{|x|^2}{r^2}\right) x^i f^i(x)\right| \\ &\le \|\nabla\cdot f\|_\infty + \frac{2}{r}\|\varphi'\|_\infty \|f\|_\infty, \end{aligned}$$

because $|x| \le r$ on the support of $\varphi'(|x|^2/r^2)$. Since $\nabla\cdot(\varphi_r f)$ converges pointwise to $\nabla\cdot f$ and D has finite volume, it follows from the dominated convergence theorem that the right-hand side of (2-1) converges to $\int_D \nabla\cdot f\, dV$. □

The next proposition is used in the proof of the main theorems.

Proposition 2.3. *Suppose D_1 and D_2 are regular domains in $\mathbb{R}^d$, with $D_1 \cap D_2$ compact and with D_2 of finite volume and surface area. Suppose further that f is a smooth bounded vector field defined on a neighborhood of D_2. There exists a sequence of regular domains $D_{2,i}$ such that $\partial D_{2,i}$ is transverse to ∂D_1, and the following limits hold as $i \to \infty$:*

(a) $\mathrm{Vol}(D_{2,i}) \to \mathrm{Vol}(D_2)$;

(b) $\mathrm{Area}(\partial D_{2,i}) \to \mathrm{Area}(\partial D_2)$;

(c) $\int_{\partial D_{2,i}} f\cdot \boldsymbol{n}_{\partial D_{2,i}}\, dA \to \int_{\partial D_2} f\cdot \boldsymbol{n}_{\partial D_2}\, dA$.

The domains can be chosen so that $D_{2,i}$ is either a decreasing sequence of domains whose intersection is D_2, or an increasing sequence of domains whose union is $\mathring{D}_2$.

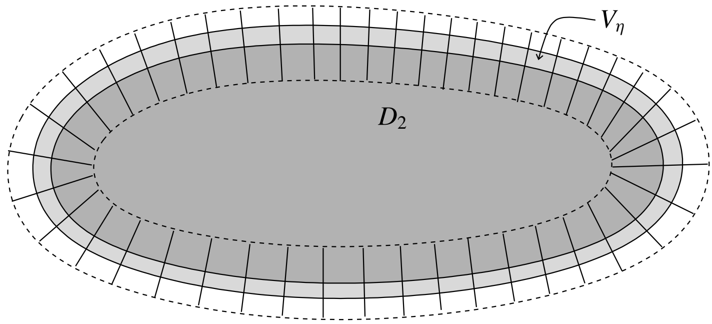


Figure 2. Defining a domain D_2^η containing D_2.

Proof. As a smooth embedded hypersurface, ∂D_2 has a tubular neighborhood N, and there exists a smooth embedding $E : \partial D_2 \times (-\delta, \delta) \to N$ such that $E(\,\cdot\,, 0)$ is the identity on ∂D_2. It can be chosen such that $E(\partial D_2 \times (0, \delta)) \cap D_2 = \varnothing$ and $E(\partial D_2 \times (-\delta, 0]) \subset D_2$.

Let $W \subset \mathbb{R}^d$ be a precompact neighborhood of $D_1 \cap D_2$ contained in the set on which f is defined, and let $\varphi : \mathbb{R}^d \to [0, 1]$ be a smooth compactly supported function that is equal to 1 on $\overline{W}$. For each η such that $\delta > \eta > 0$, define

$$V_\eta = \{E(x, s) : 0 \le s \le \eta\varphi(x)\},$$
$$D_2^\eta = D_2 \cup V_\eta.$$

(See Figure 2.) Then D_2^η is a regular domain containing D_2, which agrees with D_2 outside the support of φ. Its boundary ∂D_2^η is the image of the embedding $\iota_\eta : \partial D_2 \to \mathbb{R}^d$ given by $\iota_\eta(x) = E(x, \eta\varphi(x))$, which is equal to the inclusion map $\partial D_2 \hookrightarrow \mathbb{R}^d$ outside supp φ. The map E has full rank in $(\partial D_2 \cap W) \times (-\delta, \delta)$, and $\varphi \equiv 1$ there, so by the parametric transversality theorem (see, for example, [Lee 2013, Theorem 6.35, p. 145]), ∂D_2^η is transverse to ∂D_1 for almost every $\eta \in (-\delta, \delta)$.

Now let η_i be a sequence of positive numbers that decreases to zero, chosen so that $\partial D_2^{\eta_i}$ is transverse to ∂D_1 for each i, and set $D_{2,i} = D_2^{\eta_i}$. Then $D_{2,i}$ decreases to D_2 and $\mathrm{Vol}(D_{2,i})$ decreases to $\mathrm{Vol}(D_2)$. Moreover, because the embeddings ι_{η_i} converge uniformly with all derivatives to the inclusion map $\partial D_2 \hookrightarrow \mathbb{R}^d$, the surface area of $\partial D_{2,i}$ converges to that of ∂D_2. Further, the function $\boldsymbol{n}_{\partial D_{2,i}} \circ \iota_{\eta_i} : \partial D_2 \to \mathbb{R}^d$ converges to $\boldsymbol{n}_{\partial D_2}$. Combining these two arguments, we conclude that (c) is satisfied.

To obtain a sequence of domains that increase to $\mathring{D}_2$, we proceed instead as follows. For each η such that $-\delta < \eta < 0$, define

$$V_\eta = \{E(x, s) \mid \eta\varphi(x) < s \le 0\}, \qquad D_2^\eta = D_2 \smallsetminus V_\eta.$$

In this case, we can choose a sequence of negative numbers η_i increasing to zero such that $\partial D_2^{\eta_i}$ is transverse to ∂D_1. The rest of the proof proceeds as before. $\square$

3. Proof of Theorems 1.1 and 1.2

In this section, we prove Theorems 1.1 and 1.2. We start with a more general result that implies both theorems; first, we prove it when the boundaries of the domains intersect transversally, then, employing an approximation argument, we prove the general case.

Theorem 3.1. *Suppose D_1 and D_2 are two regular domains in $\mathbb{R}^d$, such that $D_1 \cap D_2$ is compact and D_2 has finite volume and surface area. Let f be a smooth vector field defined on a neighborhood of D_2 such that both $|f|$ and $|\nabla \cdot f|$ are bounded. The absolute value of the flux of f across the portion of ∂D_1 inside D_2 satisfies the following bound*:

$$\left|\int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA\right| \le \frac{1}{2}\left(\operatorname{Area}(\partial D_2)\|f\|_\infty + \left|\int_{\partial D_2} f \cdot \boldsymbol{n}_{\partial D_2}\, dA\right| + \operatorname{Vol}(D_2)\|\nabla \cdot f\|_\infty + \left|\int_{D_2} \nabla \cdot f\, dV\right|\right). \tag{3-1}$$

The same estimate holds when $\partial D_1 \cap D_2$ is replaced by $\partial D_1 \cap \mathring{D}_2$ on the left-hand side.

Proposition 3.2. *Theorem 3.1 holds when $\partial D_1 \pitchfork \partial D_2$.*

Proof. Note that $\partial(D_1 \cap D_2)$ is compact, and

$$\partial(D_1 \cap D_2) = (\partial D_1 \cap D_2) \cup (D_1 \cap \partial D_2). \tag{3-2}$$

Adding and subtracting $\int_{\partial D_2 \cap D_1} f \cdot \boldsymbol{n}_{\partial D_2}\, dA$, we obtain

$$\begin{aligned}\int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA &= \int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA + \int_{\partial D_2 \cap D_1} f \cdot \boldsymbol{n}_{\partial D_2}\, dA - \int_{\partial D_2 \cap D_1} f \cdot \boldsymbol{n}_{\partial D_2}\, dA \\ &= \int_{\partial (D_1 \cap D_2)} f \cdot \boldsymbol{n}_{\partial (D_1 \cap D_2)}\, dA - \int_{\partial D_2 \cap D_1} f \cdot \boldsymbol{n}_{\partial D_2}\, dA,\end{aligned}$$

since $\partial D_1 \cap \partial D_2$ is a smooth $(d-2)$-dimensional submanifold and thus has zero $(d-1)$-dimensional area.

The assumption $\partial D_1 \pitchfork \partial D_2$ implies that $D_1 \cap D_2$ is a smooth manifold with corners. To see this, we just need to show that each point is contained in the domain of an appropriate smooth coordinate chart. For points not in $\partial D_1 \cap \partial D_2$, this follows easily from the fact that D_1 and D_2 are regular domains. If $x \in \partial D_1 \cap \partial D_2$, we can find a local defining function u^1 for D_1 such that D_1 is locally given by the equation $u^1 \ge 0$; and similarly we can find a local defining function u^2 for D_2. The assumption $\partial D_1 \pitchfork \partial D_2$ ensures that du^1 and du^2 are linearly independent at x. Thus we can find smooth functions $u^3, \dots, u^d$ such that $(u^1, \dots, u^d)$ form the required local coordinates in a neighborhood of x.

Applying the divergence theorem, we get

$$\int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA = \int_{D_2 \cap D_1} \nabla \cdot f\, dV - \int_{\partial D_2 \cap D_1} f \cdot \boldsymbol{n}_{\partial D_2}\, dA.$$

Applying Lemma 2.1 to both terms on the right hand side completes the proof for $\partial D_1 \cap D_2$. The result for $\partial D_1 \cap \mathring{D}_2$ is immediate in this case, because $\partial D_1 \cap \partial D_2$ has zero surface area. □

Proof of Theorem 3.1. Let $D_{2,i}$ be a sequence of regular domains decreasing to D_2 and satisfying the conclusions of Proposition 2.3. By Proposition 3.2, for every i we have that $\left|\int_{\partial D_1 \cap D_{2,i}} f \cdot \boldsymbol{n}_{\partial D_1}\, dA\right|$ is bounded by

$$\frac{1}{2}\left(\operatorname{Area}(\partial D_{2,i})\|f\|_\infty + \left|\int_{\partial D_{2,i}} f \cdot \boldsymbol{n}_{\partial D_{2,i}}\, dA\right| + \operatorname{Vol}(D_{2,i})\|\nabla \cdot f\|_\infty + \left|\int_{D_{2,i}} \nabla \cdot f\, dV\right|\right).$$

Proposition 2.3 shows that the first three terms above converge to the first three terms on the right-hand side of (3-1). To complete the proof, we use the facts that the sets $D_{2,i}$ decrease to D_2 and the compact sets $\partial D_1 \cap D_{2,i}$ decrease to $\partial D_1 \cap D_2$ as i goes to infinity, and thus the Lebesgue dominated convergence theorem yields

$$\lim_{i\to\infty}\left|\int_{D_{2,i}} \nabla \cdot f\, dV\right| = \left|\int_{D_2} \nabla \cdot f\, dV\right|$$

and

$$\lim_{i\to\infty}\left|\int_{\partial D_1 \cap D_{2,i}} f \cdot \boldsymbol{n}_{\partial D_1}\, dA\right| = \left|\int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA\right|.$$

This completes the proof for $\partial D_1 \cap D_2$.

To prove the estimate for $\partial D_1 \cap \mathring{D}_2$, we use the same argument, but with $D_{2,i}$ chosen to increase to $\mathring{D}_2$. Because ∂D_2 has d-dimensional measure zero, we have $\int_{\mathring{D}_2} \nabla \cdot f\, dV = \int_{D_2} \nabla \cdot f\, dV$, and the result follows. □

Proof of Theorem 1.1. Inequality (1-1) follows immediately from (3-1) and obvious estimates for the integrals. □

Proof of Theorem 1.2. We first assume that $\operatorname{Vol}(D_2) < \infty$, so that (3-1) holds. In this case, the last two terms in (3-1) are zero because $\nabla \cdot f = 0$, and the second term is zero by the divergence theorem.

Now consider the case in which D_2 has infinite volume. Let D_2' denote the closure of $\mathbb{R}^d \setminus D_2$, which is a regular domain with interior $\mathring{D}_2' = \mathbb{R}^d \setminus D_2$. Because $\operatorname{Area}(\partial D_2') = \operatorname{Area}(\partial D_2) < \infty$, the isoperimetric inequality (see [De Giorgi 1953]) implies that D_2' has finite volume. If D_1 also has finite volume, the divergence theorem gives

$$\int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA + \int_{\partial D_1 \cap \mathring{D}_2'} f \cdot \boldsymbol{n}_{\partial D_1}\, dA = \int_{\partial D_1} f \cdot \boldsymbol{n}_{\partial D_1}\, dA = \int_{D_1} \nabla \cdot f\, dV = 0,$$

and (1-2) follows from Theorem 3.1 applied to the second term on the left-hand side above. On the other hand, if $\mathrm{Vol}(D_1) = \infty$, we let D_1' be the closure of $\mathbb{R}^d \setminus D_1$ (which has finite volume), and apply the above argument with D_1' in place of D_1. □

To conclude this section, we explain what modifications need to be made to Theorems 1.1 and 1.2 and their proofs to adapt them to the case of regular domains in Riemannian manifolds.

Suppose M is a d-dimensional smooth Riemannian manifold with Riemannian metric g and volume density dV_g. (If M is oriented, dV_g can be interpreted as a differential d-form; but otherwise it needs to be interpreted as a density. See [Lee 2013, pp. 427–434] for basic properties of densities.) A *regular domain* $D \subset M$ is defined just as in the case $M = \mathbb{R}^d$. If $D \subset M$ is a regular domain, it has a uniquely defined outward unit normal vector field $\boldsymbol{n}_{\partial D}$. For any such domain, we let $\tilde{g}$ denote the induced Riemannian metric on ∂D, and let $dA_{\tilde{g}}$ denote its volume density.

For any smooth vector field f defined on an open subset of M, the *divergence* of f, denoted by $\nabla\cdot f$, is defined as follows. If M is oriented, then $\nabla\cdot f$ is the unique vector field that satisfies $(\nabla\cdot f)\, dV_g = d(f \lrcorner\, dV_g)$. On a nonorientable manifold, we define it locally by choosing an orientation and using the same formula; because $\nabla\cdot f$ is unchanged when the orientation is reversed, it is globally defined. The divergence theorem then holds in exactly the same form for smooth d-dimensional submanifolds with corners in M. Moreover, any compact smooth embedded hypersurface in M has a tubular neighborhood in M. (See [Bredon 1993, Theorem 11.14, p. 100] for a proof. Although the proof there is for manifolds embedded in Euclidean space, it follows from the Whitney embedding theorem that it applies to all smooth manifolds.)

Using these facts, the proof of the following theorem is carried out exactly like the proofs of Theorems 1.1 and 1.2. To avoid complications, we restrict to the case in which D_2 is compact.

Theorem 3.3. *If D_1 and D_2 are regular domains in a Riemannian manifold (M, g) with D_2 compact, and f is a smooth vector field defined on a neighborhood of D_2, then the conclusions of Theorems 1.1 and 1.2 hold, namely,*

$$\left| \int_{\partial D_1 \cap D_2} \langle f, \boldsymbol{n}_{\partial D_1} \rangle_g \, dA_{\tilde{g}} \right| \le \mathrm{Area}(\partial D_2) \|f\|_\infty + \mathrm{Vol}(D_2) \|\nabla\cdot f\|_\infty,$$

and if $\nabla\cdot f \equiv 0$,

$$\left| \int_{\partial D_1 \cap D_2} \langle f, \boldsymbol{n}_{\partial D_1} \rangle_g \, dA_{\tilde{g}} \right| \le \tfrac{1}{2} \mathrm{Area}(\partial D_2) \|f\|_\infty.$$

4. Bounding integrals of normal fields

In this section, we prove Corollary 1.3 and Theorem 1.4. We also provide examples on the tightness of the bound.

Proof of Corollary 1.3. Let $v = \int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1}\, dA$. If $|v| = 0$ there is nothing to prove, so we assume that $|v| > 0$, and let $f : \mathbb{R}^d \to \mathbb{R}^d$ be the constant vector field $f \equiv v/|v|$. Clearly, $|v| = v \cdot v/|v| = \int_{\partial D_1 \cap D_2} f \cdot \boldsymbol{n}_{\partial D_1}\, dA$. Now, since $\nabla \cdot f \equiv 0$ and $\|f\|_\infty = 1$, the proof follows from Theorem 1.2. □

To prove Theorem 1.4, we begin with a lemma.

Lemma 4.1. *Suppose $D \subset \mathbb{R}^d$ is a compact convex regular domain with diameter δ and C is any measurable subset of ∂D. Then for any unit vector $v \in \mathbb{R}^d$, we have*

$$\int_C v \cdot \boldsymbol{n}_{\partial D}\, dA \le \tfrac{1}{2} \operatorname{Vol}(B^{d-1}(\delta/2)). \tag{4-1}$$

Proof. First consider the case $v = e_d = (0, \dots, 0, 1)$. After applying a translation, we can assume that D is contained in the set where $x^d \ge 0$. Its boundary is the union of the three subsets ∂D_+, ∂D_0, and ∂D_-, defined as the subsets of ∂D where $v \cdot \boldsymbol{n}_{\partial D}$ is positive, zero, or negative, respectively.

Now, let A be the following subset of $\mathbb{R}^d$:

$$A = \{(x^1, \dots, x^{d-1}, t x^d) \mid (x^1, \dots, x^d) \in D,\ 0 \le t \le 1\}.$$

Then A is a compact convex set, and its boundary is the union of the three subsets ∂A_+, ∂A_0, and ∂A_-, defined in the same way as above. (See Figure 3.)

The fact that D is convex ensures that $\partial A_+ = \partial D_+$, $\partial A_0 \supset \partial D_0$, and ∂A_- is contained in the hyperplane where $x^d = 0$. Moreover, A is a C^1 manifold with corners. (Its boundary might not be smooth at points where ∂A_0 meets $\overline{\partial A_+}$, but it is at least C^1 there.)

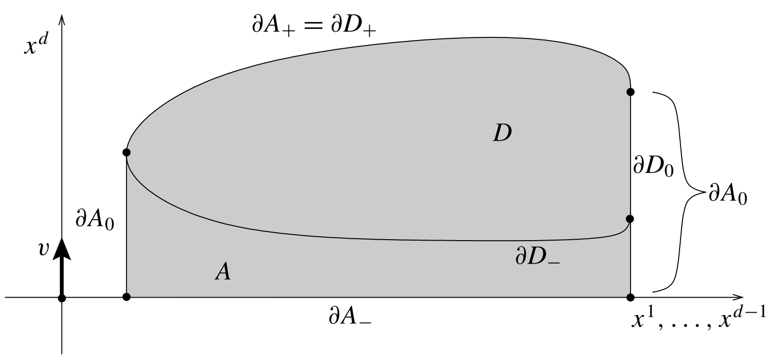


Figure 3. Proof of Lemma 4.1.

Using the fact that $v \cdot \boldsymbol{n}_{\partial D} < 0$ on ∂D_- and $v \cdot \boldsymbol{n}_{\partial D} = 0$ on ∂D_0, we compute

$$\begin{aligned}\int_C v \cdot \boldsymbol{n}_{\partial D}\, dA &= \int_{C \cap \partial D_+} v \cdot \boldsymbol{n}_{\partial D}\, dA + \int_{C \cap \partial D_0} v \cdot \boldsymbol{n}_{\partial D}\, dA + \int_{C \cap \partial D_-} v \cdot \boldsymbol{n}_{\partial D}\, dA \\ &\le \int_{C \cap \partial D_+} v \cdot \boldsymbol{n}_{\partial D}\, dA \le \int_{\partial D_+} v \cdot \boldsymbol{n}_{\partial D}\, dA \\ &= \int_{\partial A_+} v \cdot \boldsymbol{n}_{\partial A}\, dA = -\int_{\partial A_-} v \cdot \boldsymbol{n}_{\partial A}\, dA,\end{aligned}$$

where in the last line we have used the divergence theorem for the vector field $f \equiv v$ and the fact that $v \cdot \boldsymbol{n}_{\partial A} = 0$ on ∂A_0. Since $\boldsymbol{n}_{\partial A} = -v$ on ∂A_-, the last integral is equal to the area of ∂A_-. Since ∂A_- is contained in a $(d-1)$-dimensional ball of radius $\delta/2$, the result follows.

Finally, for the case of a general unit vector v, we just apply a rotation to D and apply the above argument. □

Proof of Theorem 1.4. Let D_1 and D_2 be as in the statement of the theorem. If $\int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1}\, dA = 0$, there is nothing to prove, so assume the integral is nonzero, and let v be the unit vector in the direction of $\int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1}\, dA$. Then

$$\begin{aligned}\left| \int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1}\, dA \right| &= v \cdot \int_{\partial D_1 \cap D_2} \boldsymbol{n}_{\partial D_1}\, dA \\ &= \int_{\partial D_1 \cap D_2} v \cdot \boldsymbol{n}_{\partial D_1}\, dA,\end{aligned}$$

and the result follows from Lemma 4.1. □

The following examples demonstrate the tightness of the bound for nonconvex sets, as well as the necessity of the condition that the hypersurface be the boundary of a regular domain.

Example 4.2. The main theorem explicitly uses the divergence theorem, which is applied to space-separating hypersurfaces. In fact, the bounds do not apply for images of general smooth immersions. To construct a counterexample in the plane (i.e., for $d = 2$), start with a smooth Jordan curve in the plane, then cover it m times, with small perturbations, making the integral on the left-hand side of (1-3) roughly m times as large, while the right-hand side is fixed because it depends only on ∂D_2. Clearly, whenever the left-hand side of (1-3) is not zero, we can choose m large enough that the inequality does not hold.

Example 4.3. To see that the bound obtained in Corollary 1.3 is tight, and cannot be replaced by a bound based only on the diameter of D_2 when D_2 is not convex, we consider comb-shaped subsets of $\mathbb{R}^d$, for $d \ge 2$, generated in the following manner. Fix $n > 2$, and let D_n be a closed nonsmooth comb-shaped set defined as

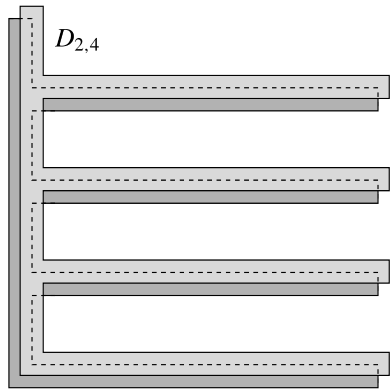


Figure 4. The domains of Example 4.3 in the case $n = 4$ (before smoothing).

the union of the following rectangles:

$$R_{i,n} = \{(x^1, \ldots, x^d) \in [0, 1]^d \mid i/n \le x^2 \le i/n + 1/n^2\}, \quad 0 \le i \le n-1;$$
$$R_{n,n} = \{(x^1, \ldots, x^d) \in [0, 1]^d \mid 0 \le x^1 \le 1/n^2\}.$$

Applying a small perturbation we then smooth its corners, and set $D_{1,n}$ accordingly. Let $D_{2,n}$ be the translation of $D_{1,n}$ by the vector $(1/(2n^2), 1/(2n^2), 0, \ldots, 0) \in \mathbb{R}^d$. (See Figure 4.) By our construction, the surface area of each set $\partial D_{1,n}$ or $\partial D_{2,n}$ is roughly $2n+2$, and the fraction of the boundary area where the normal vector of $\partial D_{1,n}$ is parallel to the x^2-axis is roughly $n/(n+1)$, approaching 1 when n is large. Notice that by the choice of $D_{2,n}$, when we integrate the normal vector in the portion of $\partial D_{1,n}$ inside $D_{2,n}$ we capture only the part pointing in the positive direction of the x^2-axis. This shows that the integral of the normal vector has magnitude of roughly n, approaching half the surface area when we take n to infinity.

5. Applications: limits of hypersurfaces & planar results

In this section we provide two applications of Corollary 1.3, extending previous planar results in [Artstein and Bright 2010; 2013]. The first is for limits of regular domains whose surface areas increase without bound. The second is an application in the planar case.

Corollary 1.3 bounds the normal vector of the boundary of a regular domain in a second regular domain, by the surface area of the boundary of the second domain, and completely disregarding the surface area of the original hypersurface. This is now applied to surfaces with increasing surface area, establishing a new result on the limit.

We will denote by $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ the unit $(d-1)$-sphere. For every hypersurface we define a corresponding probability measure using the following notation:

Definition 5.1. Suppose $S \subset \mathbb{R}^d$ is a smooth hypersurface endowed with a unit normal vector field $\boldsymbol{n}_S$. We define the *empirical measure* $\mu \in P(\mathbb{R}^d \times \mathbb{S}^{d-1})$ corresponding to S by

$$\mu(U \times V) = \frac{1}{\operatorname{Area}(S)} \int_{S \cap U} \chi_V(\boldsymbol{n}_S)\, dA,$$

for all open sets $U \subset \mathbb{R}^d$ and $V \subset \mathbb{S}^{d-1}$.

A useful property of empirical measures is that, if $f : \mathbb{R}^d \times \mathbb{S}^{d-1} \to \mathbb{R}$ is continuous, we have

$$\frac{1}{\operatorname{Area}(S)} \int_S f(x, \boldsymbol{n}_S)\, dA = \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} f(x, n) \mu(dx, dn).$$

We endow the set of probability measures $P(\mathbb{R}^d \times \mathbb{S}^{d-1})$ with the weak topology, namely, a sequence of measures $\mu_1, \mu_2, \ldots \in P(\mathbb{R}^d \times \mathbb{S}^{d-1})$ converges to a measure $\mu_0 \in P(\mathbb{R}^d \times \mathbb{S}^{d-1})$ if for every bounded continuous function $g(x, n)$,

$$\int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} g(x, n) \mu_0(dx, dn) = \lim_{i \to \infty} \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} g(x, n) \mu_i(dx, dn).$$

Another tool we need for the next theorem is disintegration of measures. Given a probability measure $\mu \in P(\mathbb{R}^d \times \mathbb{S}^{d-1})$, we define its marginal measure, $p(dx)$, as the projection on $\mathbb{R}^d$, namely, $p(A) = \mu(A \times \mathbb{S}^{d-1})$ for every measurable set $A \subset \mathbb{R}^d$. Also, we denote the measure valued function $\mu^x(dn)$, the disintegration of μ with respect to p, for p-almost every x. With this notation, for every pair of measurable sets $U \subset \mathbb{R}^d$ and $V \subset \mathbb{S}^{d-1}$, we have that

$$\mu(U \times V) = \int_U \mu^x(V) p(dx).$$

We now state the main result regarding the limits of regular domains.

Theorem 5.2. *Let $D_1, D_2, \ldots \subset \mathbb{R}^d$ be a sequence of compact regular domains, such that the surface areas of their boundaries increases to infinity. If the empirical measures $\mu_1, \mu_2, \ldots$, corresponding to the sequence $\partial D_1, \partial D_2, \ldots$, converge weakly to μ_0, then*

$$h(x) = \int_{\mathbb{S}^{d-1}} n\, \mu_0^x(x)(dn) = 0$$

for p_0-almost every x, where $\mu_0(dx, dn) = p_0(dx)\mu_0^x(dn)$ is the disintegration of μ_0 with respect to its projection p_0.

Proof. Let $B = B(x, r) \subset \mathbb{R}^d$ be a ball centered at x with radius $r > 0$. By the definition of the empirical measures and by Corollary 1.3,

$$\left| \int_{B \times \mathbb{S}^{d-1}} n\, d\mu_i(dx, dn) \right| = \left| \frac{1}{\operatorname{Area}(\partial D_i)} \int_{\partial D_i \cap B} \boldsymbol{n}_{\partial D_i}\, dA \right| \le \frac{\operatorname{Area}(\partial B)}{2 \operatorname{Area}(\partial D_i)}.$$

Weak convergence of measures and the dominated convergence theorem imply that

$$\left|\int_{B\times\mathbb{S}^{d-1}} n\, d\mu_0(dx,dn)\right| = \lim_{i\to\infty}\left|\int_{B\times\mathbb{S}^{d-1}} n\, d\mu_i(dx,dn)\right| \le \lim_{i\to\infty}\frac{\operatorname{Area}(\partial B)}{2\operatorname{Area}(\partial D_i)} = 0$$

for a set of values of $r > 0$ of full measure for which $\mu_i(\partial B(x,r)\times\mathbb{S}^{d-1}) = 0$, for all $i = 0, 1, 2, \ldots$. Using the disintegration notation we obtain that

$$\left|\int_{B\times\mathbb{S}^{d-1}} n\, d\mu_0(dx,dn)\right| = \left|\int_B\left(\int_{\mathbb{S}^{d-1}} n\,\mu_0^x(dn)\right)p_0(dx)\right| = \int_B h(x)\,p_0(dx) = 0$$

for almost every ball B. If the measure $p_0(dx)$ is Lebesgue measure, by the Lebesgue differentiation theorem we have $h(x) = 0$ almost everywhere. The Lebesgue–Besicovitch differentiation theorem extends this result to Radon measures (see, for example, [Evans and Gariepy 1992, p. 43]). □

Remark 5.3. Theorem 5.2 requires the convergence of the empirical measures. When the domains in the sequence are contained in some compact set K, the compactness of the space $K\times\mathbb{S}^{d-1}$ implies the compactness of $P(K\times\mathbb{S}^{d-1})$, and, therefore, the existence of a converging subsequence [Billingsley 1999, p. 72].

In two dimensions, our result extends as follows.

Corollary 5.4. *Suppose $x_1 : [0,\tau_1]\to\mathbb{R}^2$ is a parametrized smooth Jordan curve and $D_2\subset\mathbb{R}^2$ is a regular domain. If the length of ∂D_2 is L_2, then*

$$\left|\int_0^{\tau_1}\chi_{D_2}(x_1(t))\frac{d}{dt}x_1(t)\,dt\right| \le \frac{L_2}{2}.$$

Proof. Let T_1 and N_1 be the unit tangent and normal vectors of x_1. Using the arc-length parametrization, we have that

$$\left|\int_0^{\tau_1}\chi_{D_2}(x_1(t))\frac{d}{dt}x_1(t)\,dt\right| = \left|\int_0^{L_1}\chi_{D_2}(x_1(s))T_1(s)\,ds\right|,$$

where L_1 is the length of x_1. Expressing the tangent vector in terms of the normal vector, we reduce the previous expression to

$$\left|\int_0^{L_1}\chi_{D_2}(x_1(s))\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}N_1(s)\,ds\right| = \left|\int_0^{L_1}\chi_{D_2}(x_1(s))N_1(s)\,ds\right|,$$

as the rotation matrix is orthogonal. Applying Corollary 1.3 completes the proof. □

For our final application, we consider an ordinary differential equation in the plane defined by

$$\frac{dx}{dt} = f(x), \tag{5-1}$$

where $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a vector field (generally assumed at least Lipschitz continuous). An *invariant set* for f is a subset of $\mathbb{R}^2$ that is invariant under the forward flow of f and a *minimal set* is a nonempty closed invariant set that is minimal with respect to inclusions. A *trivial minimal set* is a set that is the image of either a stationary solution or a periodic solution.

We present a new short proof of the following well-known result.

Theorem 5.5. *Suppose f is a smooth vector field on $\mathbb{R}^2$. Then every minimal set for f is trivial.*

The textbook proof of this theorem (see [Verhulst 1996]) relies on the Poincaré–Bendixson theorem, and employs dynamical arguments. Here we present a simpler proof based on the divergence theorem, and specifically on Corollary 1.3. Note that the divergence theorem was used by Bendixson in the proof of the Bendixson criterion, which verifies that no periodic solutions exist.

Our proof uses the following well-known lemmas.

Lemma 5.6. *Suppose $\Omega \subset \mathbb{R}^2$ is a minimal set for (5-1) and $x^* : [0, \infty) \to \mathbb{R}^2$ is a solution to (5-1) with trajectory contained in Ω. For every $y_0 \in \Omega$, $s \in [0, \infty)$, and $\delta > 0$, there exists $t > s$ such that $|x^*(t) - y_0| < \delta$.*

Proof. Suppose the lemma does not hold for some y_0, s, and δ. Then the curve $y^*(t) = x^*(s + t)$ is a solution to (5-1) with trajectory contained in $\Omega \setminus B(y_0, \delta)$ for a suitable $\delta > 0$, in contradiction to the minimality of Ω. □

The next lemma follows easily from Sard's theorem.

Lemma 5.7. *Suppose $I \subset \mathbb{R}$ is a compact interval and $g : I \to \mathbb{R}$ is smooth. Then for almost every $r \in \mathbb{R}$, the set $g^{-1}(r) = \{t \in I \mid g(t) = r\}$ is finite.*

Proof of Theorem 5.5. Clearly, Ω is a singleton if and only if it contains a point $y \in \Omega$ such that $f(y) = 0$, so we may assume henceforth that f does not vanish in Ω and Ω contains more than one point. Choose $D > 0$ such that $\Omega \setminus B(x^*(0), 3D) \neq \varnothing$. We construct sequences of real numbers $\{\delta_i\}$ and $\{t_i\}$, and a sequence of simple closed curves $\{\gamma_i\}$, as follows. Set $\delta_0 = D$, and let t_0 be the first time where x^* meets $\partial B(x^*(0), \delta_0)$. For $i = 1, 2, \ldots$ do the following:

(a) Choose $\delta_i < \delta_{i-1}/2$ small enough that $|x^*(0) - x^*(t)| > \delta_i$ for all $t \in [t_0, t_{i-1}]$.

(b) Let t_i be the first time after t_0 where the curve x^* meets $\partial B(x^*(0), \delta_i)$. (Here we use Lemma 5.6.)

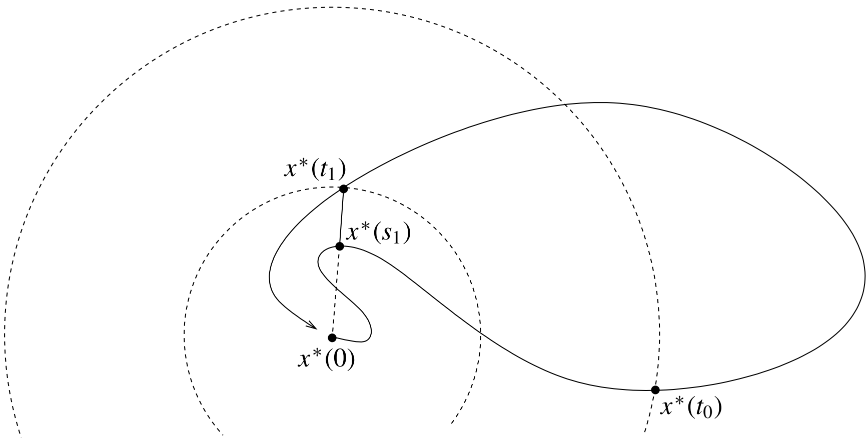


Figure 5. Proof of Theorem 5.5.

(c) Starting from $x^*(t_i)$, follow the line connecting it to $x^*(0)$, until that line first meets a point in $x^*([0, t_0])$. Let $x^*(s_i)$ be this point. (See Figure 5.)

(d) Let γ_i be the parametrized piecewise smooth curve obtained by following the curve x^* in the interval $[s_i, t_i]$, and then the line connecting its endpoints with unit speed.

Note that t_i is an increasing sequence and that the uniqueness of the solution with respect to the initial condition implies that every γ_i is a Jordan curve. Suppose first that the sequence $\{t_i\}$ is bounded above. Then $t_i \to t^* \in \mathbb{R}^+$ and $x^*(t_i) \to x^*(t^*)$. According to our construction, $|x^*(0) - x^*(t_i)| = \delta_i < 2^{-i} D$ for every i. Hence by continuity $x^*(t^*) = x^*(0)$, and x^* is periodic. By the minimality of Ω, the image of x^* is Ω.

The only remaining possibility is $t_i \nearrow \infty$. Fix $y_0 \in \Omega$ such that $|y_0 - x^*(0)| > 2D$. By Lemma 5.7, there exists arbitrarily small $r_0 < D$ such that the set

$$\{t \in [0, s] \mid |x^*(t) - y_0| = r_0\}$$

is finite for every $s > 0$. (This follows from the fact that $g(t) = |x^*(t) - y_0|^2$ is a smooth function of t.) Note that this implies that the portion of γ_i in $B_0 = B(y_0, r_0)$ is part of the trajectory x^*, and that for every i the Jordan curve γ_i intersects ∂B_0 at a finite number of points.

For every i, we let D_i denote the domain consisting of the Jordan curve γ_i together with its interior. Although D_i is not a regular domain, it is a regular domain with two corner points, which are outside of $\bar{B}_0$, and it is easy to see that Corollary 1.3 can be applied to $\partial D_i \cap B_0$. Thus by Corollary 5.4,

$$\left| \int_{\{t \le t_i \mid x^*(t) \in B_0\}} \frac{d}{dt} x^*(t)\, dt \right| = \left| \int_{\{t \le t_i \mid x^*(t) \in B_0\}} f(x^*(t))\, dt \right| \le \pi r_0.$$

Because Ω is minimal, Lemma 5.6 implies that the set $\{t \mid x^*(t) \in B_0\}$ has infinite measure. This implies that 0 is contained in the convex hull of the set

$$\{f(y) \mid y \in \bar{B}_0\}.$$

The radius r_0 can be chosen arbitrary small; therefore, the continuity of f implies that $f(y_0) = 0$, in contradiction. $\square$

Acknowledgements

The first author wishes to thank Vered Rom-Kedar for suggesting the idea of studying the integral of the normal vector, and Monica Torres for valuable remarks.

This work was supported by the Applied Mathematics Program within the Department of Energy (DOE) Office of Advanced Scientific Computing Research.

References

[Artstein and Bright 2010] Z. Artstein and I. Bright, "Periodic optimization suffices for infinite horizon planar optimal control", *SIAM J. Control Optim.* **48**:8 (2010), 4963–4986. MR 2011k:49008 Zbl 1208.49006

[Artstein and Bright 2013] Z. Artstein and I. Bright, "On the velocity of planar trajectories", *NoDEA Nonlinear Differential Equations Appl.* **20**:2 (2013), 177–185. MR 3046971 Zbl 06196734

[Billingsley 1999] P. Billingsley, *Convergence of probability measures*, 2nd ed., Wiley, New York, 1999. MR 2000e:60008 Zbl 0944.60003

[Bredon 1993] G. E. Bredon, *Topology and geometry*, Graduate Texts in Mathematics **139**, Springer, New York, 1993. MR 94d:55001 Zbl 0791.55001

[Bright 2012] I. Bright, "A reduction of topological infinite-horizon optimization to periodic optimization in a class of compact 2-manifolds", *J. Math. Anal. Appl.* **394**:1 (2012), 84–101. MR 2926206 Zbl 1244.49004

[Bright and Torres 2014] I. Bright and M. Torres, "The integral of the normal and fluxes over sets of finite perimeter", 2014. To appear in *Interfaces Free Bound.*

[De Giorgi 1953] E. De Giorgi, "Definizione ed espressione analitica del perimetro di un insieme", *Atti Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat.* (8) **14** (1953), 390–393. MR 15,20b Zbl 0051.29403

[Evans and Gariepy 1992] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, CRC, Boca Raton, FL, 1992. MR 93f:28001 Zbl 0804.28001

[Guillemin and Pollack 1974] V. Guillemin and A. Pollack, *Differential topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974. Reprinted by AMS Chelsea, Providence, RI, 2010. MR 50 #1276 Zbl 0361.57001

[Lee 2013] J. M. Lee, *Introduction to smooth manifolds*, 2nd ed., Graduate Texts in Mathematics **218**, Springer, New York, 2013. MR 2954043 Zbl 1258.53002

[Verhulst 1996] F. Verhulst, *Nonlinear differential equations and dynamical systems*, Springer, Berlin, 1996. MR 97g:34003 Zbl 0854.34002

Received October 13, 2013. Revised November 26, 2013.

IDO BRIGHT
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195-3925
UNITED STATES

ibright@uw.edu

JOHN M. LEE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195-4350
UNITED STATES

johnmlee@uw.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.323

ON STABLE COMMUTATOR LENGTH IN HYPERELLIPTIC MAPPING CLASS GROUPS

DANNY CALEGARI, NAOYUKI MONDEN AND MASATOSHI SATO

We give a new upper bound on the stable commutator length of Dehn twists in hyperelliptic mapping class groups and determine the stable commutator length of some elements. We also calculate values and the defects of homogeneous quasimorphisms derived from ω-signatures and show that they are linearly independent in the mapping class groups of pointed 2-spheres when the number of points is small.

1. Introduction

The aim of this paper is to investigate stable commutator length in hyperelliptic mapping class groups and in mapping class groups of pointed 2-spheres. Given a group G and an element $x \in [G, G]$, the commutator length of x, denoted by $\mathrm{cl}_G(x)$, is the smallest number of commutators in G whose product is x, and the stable commutator length of x is defined by the limit $\mathrm{scl}_G(x) := \lim_{n\to\infty} \mathrm{cl}_G(x^n)/n$ (see Definition 2.1 for details).

We investigate stable commutator length in two groups, $\mathcal{M}_0^m$ and $\mathcal{H}_g$. Let m be a positive integer greater than 3. Choose m distinct points $\{q_i\}_{i=1}^m$ in a 2-sphere S^2. Let $\mathrm{Diff}_+(S^2, \{q_i\}_{i=1}^m)$ denote the set of all orientation-preserving diffeomorphisms in S^2 which preserve $\{q_i\}_{i=1}^m$ setwise with the C^∞-topology. We define the mapping class group of the m-pointed 2-sphere by $\mathcal{M}_0^m = \pi_0 \,\mathrm{Diff}_+(S^2, \{q_i\}_{i=1}^m)$. Let Σ_g be a closed connected oriented surface of genus $g \geq 1$. An involution $\iota : \Sigma_g \to \Sigma_g$ defined as in Figure 1 is called the hyperelliptic involution.

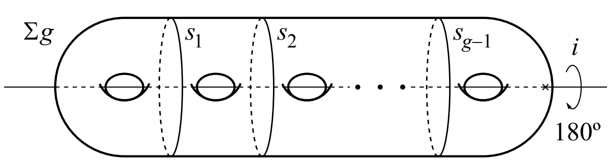


Figure 1. Hyperelliptic involution ι and the curves $s_1, \ldots, s_{g-1}$.

MSC2010: primary 57M07; secondary 20F12, 57N05.
Keywords: stable commutator length, mapping class groups.

Let $\mathcal{M}_g$ denote the mapping class group of Σ_g, that is, the group of isotopy classes of orientation-preserving diffeomorphisms of Σ_g, and let $\mathcal{H}_g$ be the centralizer of the isotopy class of a hyperelliptic involution in $\mathcal{M}_g$, which is called the hyperelliptic mapping class group of genus g. Note that $\mathcal{M}_g = \mathcal{H}_g$ when $g = 1, 2$. Since there exists a surjective homomorphism $\mathcal{P} : \mathcal{H}_g \to \mathcal{M}_0^{2g+2}$ with finite kernel (see Lemma 3.3 and the paragraph before Remark 3.7), these two groups have the same stable commutator length.

Let s_0 be a nonseparating curve on Σ_g satisfying $\iota(s_0) = s_0$, and let s_h be a separating curve in Figure 1 for $h = 1, \dots, g-1$. We denote by t_{s_j} the Dehn twist about s_j for $j = 0, 1, \dots, g-1$. In general, it is difficult to compute stable commutator length, but those of some mapping classes are known. In the mapping class group of a compact oriented surface with connected boundary, Baykur, Korkmaz and the second author [Baykur et al. 2013] determined the commutator length of the Dehn twist about a boundary curve. In the mapping class group of a closed oriented surface, interesting lower bounds on scl of Dehn twists are obtained using gauge theory. Endo and Kotschick [2001], and Korkmaz [2004] proved that $1/(18g-6) \le \mathrm{scl}_{\mathcal{M}_g}(t_{s_j})$ for $j = 0, 1, \dots, g-1$. For technical reasons, this result is stated in [Endo and Kotschick 2001] only for separating curves. This technical assumption is removed in [Korkmaz 2004]. The second author [Monden 2012] also showed that $1/(8g+4) \le \mathrm{scl}_{\mathcal{H}_g}(t_{s_0})$ and

$$\frac{h(g-h)}{g(2g+1)} \le \mathrm{scl}_{\mathcal{H}_g}(t_{s_h}) \quad \text{for } h = 1, \dots, g-1.$$

Stable commutator length on a group is closely related to functions on the group called homogeneous quasimorphisms through Bavard's duality theorem. Homogeneous quasimorphisms are homomorphisms up to bounded error called the defect (see Definition 2.2 for details). By Bavard's theorem, if we obtain a homogeneous quasimorphism on the group and calculate its defect, we also obtain a lower bound on stable commutator length. Actually, Bestvina and Fujiwara [2002, Theorem 12] proved that the spaces of homogeneous quasimorphisms on $\mathcal{M}_g$ and $\mathcal{M}_0^m$ are infinite-dimensional when $g \ge 2$ and $m \ge 5$, respectively. However it is hard to compute explicit values of these quasimorphisms and their defects. To compute stable commutator length, we consider computable quasimorphisms derived from ω-signature in [Gambaudo and Ghys 2005] on symmetric mapping class groups.

In Section 3, we review symmetric mapping class groups, which are defined by Birman and Hilden as generalizations of hyperelliptic mapping class groups. We reconsider cobounding functions of ω-signatures as a series of quasimorphisms $\phi_{m,j}$ on a symmetric mapping class group $\pi_0 C_g(t)$. Since there exists a surjective homomorphism $\mathcal{P} : \pi_0 C_g(t) \to \mathcal{M}_0^m$ with finite kernel, the homogenizations $\bar{\phi}_{m,j}$

induce homogeneous quasimorphisms on $\mathcal{M}_0^m$. Let $\sigma_i \in \mathcal{M}_0^m$ be a half twist which permutes the i-th point and the $(i+1)$-th point. We denote by $\tilde{\sigma}_i \in \pi_0 C_g(t)$ a lift of σ_i, which will be defined on page 333.

In Section 6, we calculate $\phi_{m,j}$ and their homogenizations $\bar{\phi}_{m,j}$.

Theorem 1.1. *Let r be an integer such that $2 \le r \le m$. Then:*

(i)
$$\phi_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{r-1}) = \frac{2(r-1)j(m-j)}{m(m-1)}.$$

(ii)
$$\bar{\phi}_{m,j}(\sigma_1 \cdots \sigma_{r-1}) = -\frac{2}{r}\left\{\frac{jr(m-j)(m-r)}{m^2(m-1)} + \left(\frac{rj}{m} - \left[\frac{rj}{m}\right] - \frac{1}{2}\right)^2 - \frac{1}{4}\right\},$$

where $[x]$ denotes the greatest integer $\le x$.

Since this requires straightforward and lengthy calculations to prove, we leave it until the last section. A computer calculation shows that the $([m/2]-1)\times([m/2]-1)$ matrix whose (i, j)-entry is $\bar{\phi}_{m,j+1}(\sigma_1 \cdots \sigma_i)$ is nonsingular when $4 \le m \le 30$. Thus we have:

Corollary 1.2. *The set $\{\bar{\phi}_{m,j}\}_{j=2}^{[m/2]}$ is linearly independent when $4 \le m \le 30$.*

In Section 4, we calculate the defects of the homogenizations of these quasimorphisms. In particular, we determine the defect of $\bar{\phi}_{m,m/2}$ when m is even. Actually $\bar{\phi}_{m,m/2}$ is the same as the homogenization of the Meyer function on the hyperelliptic mapping class group $\mathcal{H}_g$.

Theorem 1.3. *Let $D(\phi_{m,j})$ and $D(\bar{\phi}_{m,j})$ be the defects of the quasimorphisms $\phi_{m,j}$ and $\bar{\phi}_{m,j}$, respectively.*

(i) *For $j = 1, 2, \ldots, [m/2]$,*
$$D(\bar{\phi}_{m,j}) \le D(\phi_{m,j}) \le m-2.$$

(ii) *When m is even and $j = m/2$,*
$$D(\bar{\phi}_{m,m/2}) = m-2.$$

Remark 1.4. If $\phi : G \to \mathbb{R}$ is a quasimorphism and $\bar{\phi} : G \to \mathbb{R}$ is its homogenization, they satisfy
$$D(\bar{\phi}) \le 2D(\phi)$$
(see [Calegari 2009] Corollary 2.59). We will claim in Lemma 4.1 that, when ϕ is antisymmetric and a class function, they satisfy the sharper inequality
$$D(\bar{\phi}) \le D(\phi).$$

Note that when $g = 2$, the hyperelliptic mapping class group $\mathcal{H}_2$ coincides with $\mathcal{M}_2$. We may think of the lift of $\sigma_i \in \mathcal{M}_0^6$ for $i = 1, 2, 3, 4, 5$ to $\mathcal{M}_2$ as the Dehn twist t_{c_i} along the simple closed curve c_i in Figure 2 (see page 333). Similarly the Dehn twist $t_{s_1} \in \mathcal{M}_2$ can be considered as a lift of $(\sigma_1\sigma_2)^6 \in \mathcal{M}_0^6$ by the chain relation (see Lemma 2.8). Since Theorem 1.1(ii) implies $\bar{\phi}_{6,2}((\sigma_1\sigma_2)^6) = -8/5$ and Theorem 1.3(i) implies $D(\bar{\phi}_{6,2}) \leq 4$, by applying Bavard's duality theorem, we have:

Corollary 1.5.
$$\frac{1}{5} \leq \mathrm{scl}_{\mathcal{M}_2}(t_{s_1}).$$

To the best of our knowledge, for $g \geq 2$, there is not known an element x in $\mathcal{H}_g$ (or $\mathcal{M}_g$) such that $\mathrm{scl}(x)$ is nonzero and can be computed explicitly. By Theorem 1.3(ii), we can determine the stable commutator length of the following element in $\mathcal{H}_g$.

Theorem 1.6. *Let $d_2^+, d_2^-, \ldots, d_{g-1}^+, d_{g-1}^-$ be simple closed curves in Figure 7. Let c be a nonseparating simple closed curve satisfying $\iota(c) = c$ which is disjoint from d_i^+, d_i^- and s_h $(i = 1, \ldots, g, h = 1, \ldots, g-1)$. For $g \geq 2$,*
$$\mathrm{scl}_{\mathcal{H}_g}\bigl(t_c^{2g+8}(t_{d_2^+}t_{d_2^-}\cdots t_{d_{g-1}^+}t_{d_{g-1}^-})^2(t_{s_1}\cdots t_{s_{g-1}})^{-1}\bigr) = \frac{1}{2}.$$
In particular, if $g = 2$, then we have $\mathrm{scl}_{\mathcal{H}_2}(t_c^{12}t_{s_1}^{-1}) = 1/2$.

Next we consider upper bounds on stable commutator length. Korkmaz [2004] also gave the upper bound $\mathrm{scl}_{\mathcal{M}_g}(t_{s_0}) \leq 3/20$ for $g \geq 2$. In the case of $g = 2$, the second author [Monden 2012] showed $\mathrm{scl}_{\mathcal{M}_2}(t_{s_0}) < \mathrm{scl}_{\mathcal{M}_2}(t_{s_1})$. However these upper bounds do not depend on g. On the other hand, Kotschick [2008] proved that there is an estimate $\mathrm{scl}_{\mathcal{M}_g}(t_{s_0}) = O(1/g)$ by using the so-called "Munchhausen trick".

In Section 5, we give the following upper bounds.

Theorem 1.7. *Let s_0 be a nonseparating curve on Σ_g, and let G_g be either $\mathcal{M}_g$ or $\mathcal{H}_g$. For all $g \geq 1$, we have*
$$\mathrm{scl}_{G_g}(t_{s_0}) \leq \frac{1}{2\{2g+3+(1/g)\}}.$$

2. Preliminaries

Stable commutator lengths and quasimorphisms. Let G denote a group, and let $[G, G]$ denote the commutator subgroup, which is the subgroup of G generated by all commutators $[x, y] = xyx^{-1}y^{-1}$ for $x, y \in G$.

Definition 2.1. For $x \in [G, G]$, the commutator length $\mathrm{cl}_G(x)$ of x is the least number of commutators in G whose product is equal to x. The stable commutator length of x, denoted $\mathrm{scl}(x)$, is the limit
$$\mathrm{scl}_G(x) = \lim_{n\to\infty} \frac{\mathrm{cl}_G(x^n)}{n}.$$

For each fixed x, the function $n \mapsto \mathrm{cl}_G(x^n)$ is nonnegative and

$$\mathrm{cl}_G(x^{m+n}) \le \mathrm{cl}_G(x^m) + \mathrm{cl}_G(x^n).$$

Hence this limit exists. If x is not in $[G, G]$ but has a power x^m which is, define $\mathrm{scl}_G(x) = \mathrm{scl}_G(x^m)/m$. We also define $\mathrm{scl}_G(x) = \infty$ if no power of x is contained in $[G, G]$ (we refer the reader to [Calegari 2009] for the details of the theory of the stable commutator length).

Definition 2.2. A quasimorphism is a function $\phi : G \to \mathbb{R}$ for which there is a least constant $D(\phi) \ge 0$ such that

$$|\phi(xy) - \phi(x) - \phi(y)| \le D(\phi),$$

for all $x, y \in G$. We call $D(\phi)$ the defect of ϕ. A quasimorphism is homogeneous if it satisfies the additional property $\phi(x^n) = n\phi(x)$ for all $x \in G$ and $n \in \mathbb{Z}$.

We recall the following basic facts. Let ϕ be a quasimorphism on G. For each $x \in G$, define

$$\bar{\phi}(a) := \lim_{n\to\infty} \frac{\phi(x^n)}{n}.$$

The limit exists and defines a homogeneous quasimorphism. Homogeneous quasimorphisms have the following properties, shown for example in [Calegari 2009, Section 5.5.2] and [Kotschick 2008, Lemma 2.1(1)].

Lemma 2.3. *Let ϕ be a homogeneous quasimorphism on G. For all $x, y \in G$,*

(i) $\phi(x) = \phi(yxy^{-1})$,

(ii) $xy = yx \implies \phi(xy) = \phi(x) + \phi(y)$.

Theorem 2.4 (Bavard's duality theorem [1991]). *Let Q be the set of homogeneous quasimorphisms on G with positive defects. For any $x \in [G, G]$, we have*

$$\mathrm{scl}_G(x) = \sup_{\phi \in Q} \frac{|\phi(x)|}{2D(\phi)}.$$

Mapping class groups. For $g \ge 1$, the abelianizations of the mapping class group $\mathcal{M}_g$ of the surface Σ_g and its subgroup $\mathcal{H}_g$ are finite (see [Powell 1978]). Therefore all elements of $\mathcal{M}_g$ and $\mathcal{H}_g$ have powers that are products of commutators. Dehn showed that the mapping class group $\mathcal{M}_g$ is generated by Dehn twists along nonseparating simple closed curves. We review some relations between them. Hereafter we do not distinguish a simple closed curve in Σ_g and its isotopy class. The following relations are well known. See, for example, [Farb and Margalit 2012, Sections 3.3, 3.5.1, 5.1.4, and 4.4.1].

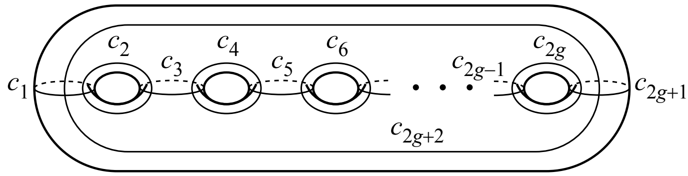


Figure 2. The curves $c_1, c_2, \ldots, c_{2g+2}$.

Lemma 2.5. *Let c be a simple closed curve in Σ_g and $f \in \mathcal{M}_g$. Then we have*

$$t_{f(c)} = f t_c f^{-1}.$$

From this lemma, the values of scl and homogeneous quasimorphisms on the Dehn twists about nonseparating simple closed curves are constant.

Lemma 2.6. *Let c and d be simple closed curves in Σ_g.*

(i) *If c is disjoint from d, then $t_c t_d = t_d t_c$.*

(ii) *If c intersects d in one point transversely, then $t_c t_d t_c = t_d t_c t_d$.*

Lemma 2.7 (hyperelliptic involution). *Let $c_1, \ldots, c_{2g+1}$ be nonseparating curves in Σ_g as in Figure 2. We call the product*

$$\iota := t_{c_{2g+1}} t_{c_{2g}} \cdots t_{c_2} t_{c_1} t_{c_1} t_{c_2} \cdots t_{c_{2g}} t_{c_{2g+1}}$$

the hyperelliptic involution. For $g = 1$, the hyperelliptic involution ι equals $t_{c_1} t_{c_2} t_{c_1} t_{c_1} t_{c_2} t_{c_1}$, where c_1 and c_2 are respectively the meridian and longitude of Σ_1.

Lemma 2.8 (chain relation). *For a positive integer n, let $a_1, a_2, \ldots, a_n$ be a sequence of simple closed curves in Σ_g such that a_i and a_j are disjoint if $|i - j| \geq 2$ and a_i and a_{i+1} intersect at one point.*

When n is odd, a regular neighborhood of $a_1 \cup a_2 \cup \cdots \cup a_n$ is a subsurface of genus $(n-1)/2$ with two boundary components, denoted by d_1 and d_2. Then

$$(t_{a_n} \cdots t_{a_2} t_{a_1})^{n+1} = t_{d_1} t_{d_2}.$$

When n is even, a regular neighborhood of $a_1 \cup a_2 \cup \cdots \cup a_n$ is a subsurface of genus $n/2$ with connected boundary, denoted by d. Then

$$(t_{a_n} \cdots t_{a_2} t_{a_1})^{2(n+1)} = t_d.$$

Meyer's signature cocycle. Let X be a compact oriented $(4n+2)$-manifold for nonnegative integer n, and let Γ be a local system on X such that $\Gamma(x)$ is a finite-dimensional real or complex vector space for $x \in X$. If we are given a regular antisymmetric (respectively, skew-hermitian) form $\Gamma \otimes \Gamma \to \mathbb{R}$ (respectively, $\Gamma \otimes \Gamma \to \mathbb{C}$), we have a symmetric (respectively, hermitian) form on $H_{2n+1}(X; \Gamma)$

as in [Meyer 1972, p. 12]. For simplicity, we only explain the complex case. It is defined by

$$\begin{aligned} H_{2n+1}(X;\Gamma)\otimes H_{2n+1}(X;\Gamma) &\cong H^{2n+1}(X,\partial X;\Gamma)\otimes H^{2n+1}(X,\partial X;\Gamma)\\ &\xrightarrow{\cup} H^{4n+2}(X,\partial X;\Gamma\otimes\Gamma)\\ &\to H^{4n+2}(X,\partial X;\mathbb{C})\\ &\xrightarrow{[X,\partial X]} \mathbb{C}, \end{aligned}$$

where the first row is defined by the Poincaré duality, the second row is defined by the cup product of the base space, the third row comes from the skew-hermitian form of Γ as above, and the fourth row is the evaluation by the fundamental class of X. Meyer showed additivity of signatures with respect to this hermitian form (more strongly, he showed Wall's nonadditivity formula for G-signatures of homology groups with local coefficients).

Theorem 2.9 [Meyer 1972, Satz I.3.2]. *Let X and Γ be as above. Assume that X is obtained by gluing two compact oriented $(4n+2)$-manifold X_- and X_+ along some boundary components.*

Then we have

$$\mathrm{Sign}(H_{2n+1}(X;\Gamma)) = \mathrm{Sign}(H_{2n+1}(X_-;\Gamma|_{X_-})) + \mathrm{Sign}(H_{2n+1}(X_+;\Gamma|_{X_+})).$$

Consider the case when X is a pair of pants, which we denote by P. Let α and β be loops in P as in Figure 3, left.

For $\varphi,\psi\in\mathcal{M}_g$, there exists a Σ_g-bundle $E_{\varphi,\psi}$ on P whose monodromies along α and β are φ and ψ, respectively. This is unique up to bundle isomorphism. In this setting, the intersection form on the local system $H_1(\Sigma_g;\mathbb{R})$ induces the symmetric form on $H_1(P;H_1(\Sigma_g;\mathbb{R}))$. Meyer showed that the signature of this symmetric form on $H_1(P;H_1(\Sigma_g;\mathbb{R}))$ coincides with that of $E_{\varphi,\psi}$. Moreover he explicitly described it in terms of the action of the mapping class group on $H_1(\Sigma_g;\mathbb{R})$ as

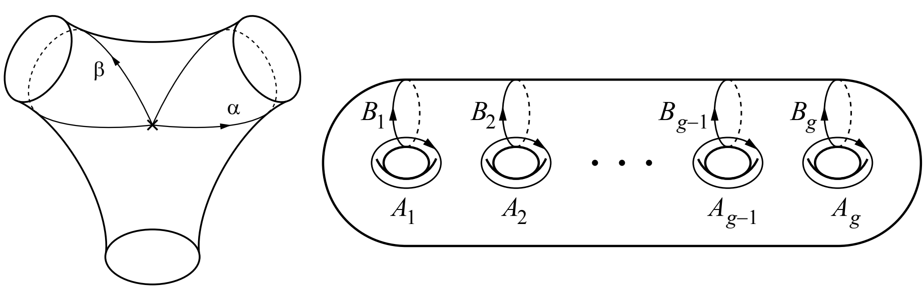


Figure 3. Left: loops in a pair of pants. Right: a symplectic basis of $H_1(\Sigma_g;\mathbb{Z})$.

follows. Fix the symplectic basis $\{A_i, B_i\}_{i=1}^g$ of $H_1(\Sigma_g; \mathbb{Z})$ as in Figure 3, right; then the action induces a homomorphism $\rho : \mathcal{M}_g \to \mathrm{Sp}(2g; \mathbb{Z})$. Let I denote the identity matrix of rank g and define

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

For symplectic matrices A and B of rank $2g$, define the vector space

$$V_{A,B} = \{(v, w) \in \mathbb{R}^{2g} \times \mathbb{R}^{2g} \mid (A^{-1} - I)v + (B - I)w = 0\}.$$

Consider the symmetric bilinear form

$$\langle\ ,\ \rangle_{A,B} : V_{A,B} \times V_{A,B} \to \mathbb{R}$$

on $V_{A,B}$ defined by

$$\langle (v_1, w_1), (v_2, w_2)\rangle_{A,B} := (v_1 + w_1)^T J (I - B) w_2.$$

Then, the space $V_{A,B}$ coincides with $H_1(P; H_1(\Sigma_g; \mathbb{R}))$, and the above form $\langle\ ,\ \rangle_{\rho(\varphi),\rho(\psi)}$ corresponds to the symmetric form on $H_1(P; H_1(\Sigma_g; \mathbb{R}))$.

Meyer's signature cocycle $\tau_g : \mathcal{M}_g \times \mathcal{M}_g \to \mathbb{Z}$ is the map defined by

$$(\varphi, \psi) \mapsto \mathrm{Sign}(\langle\ ,\ \rangle_{\rho(\varphi),\rho(\psi)}),$$

which is known to be a bounded 2-cocycle by Theorem 2.9. When we restrict it to the hyperelliptic mapping class group $\mathcal{H}_g$, it represents the trivial cohomology class in $H^2(\mathcal{H}_g; \mathbb{Q})$. Since the first homology $H_1(\mathcal{H}_g; \mathbb{Q})$ is trivial, the cobounding function $\phi_g : \mathcal{H}_g \to \mathbb{Q}$ of τ_g is unique. It is a quasimorphism, called the Meyer function. Endo [2000] computed it to investigate signatures of fibered 4-manifolds called hyperelliptic Lefschetz fibrations. Morifuji [2003] relates it to the eta invariants of mapping tori and the Casson invariants of integral homology 3-spheres.

3. Cobounding functions of the Meyer's signature cocycles on symmetric mapping class groups

As in the introduction, let m be a positive integer greater than 3 and $\{q_i\}_{i=1}^m$ be m distinct points in a 2-sphere S^2. Choose a base point $* \in S^2 - \{q_i\}_{i=1}^m$, and denote by $\alpha_i \in \pi_1(S^2 - \{q_i\}_{i=1}^m, *)$ a loop which rounds the point q_i clockwise as in Figure 4.

For an integer d such that $d \geq 2$ and $d \mid m$, define a homomorphism

$$\pi_1(S^2 - \{q_i\}_{i=1}^m) \to \mathbb{Z}/d\mathbb{Z}$$

by mapping each generator α_i to $1 \in \mathbb{Z}/d\mathbb{Z}$. This homomorphism induces a d-cyclic branched covering $p_d : \Sigma_h \to S^2$ with m branched points, where Σ_h is a closed oriented surface of genus h. Applying the Riemann–Hurwitz formula, we have

$h = (d-1)(m-2)/2$. We denote by $t : \Sigma_h \to \Sigma_h$ the deck transformation which corresponds to the generator $1 \in \mathbb{Z}/d\mathbb{Z}$.

Let η denote the d-th root of unity $\exp(2\pi \boldsymbol{i}/d)$, where $\boldsymbol{i}$ is a square root of -1. The first homology $H_1(\Sigma_h; \mathbb{C})$ decomposes into a direct sum $\bigoplus_{j=1}^{d-1} V^{\eta^j}$, where V^z is an eigenspace whose eigenvalue is $z \in \mathbb{C}$. Note that V^1 is trivial since the quotient space $\Sigma_g/\langle t\rangle$ is a 2-sphere, where $\langle t\rangle$ denotes the cyclic group generated by the deck transformation t. We also denote by $C_h(t)$ the centralizer of t in the diffeomorphism group $\mathrm{Diff}_+ \Sigma_h$. We call the path-connected component $\pi_0 C_h(t)$ the symmetric mapping class group of the covering p, which is defined by Birman and Hilden [1973].

In this section, we introduce 2-cocycles on the symmetric mapping class group $\pi_0 C_h(t)$, derived from the nonadditivity formula for signatures. These are almost the same as the ω-signatures defined in [Gambaudo and Ghys 2005].

Let us consider an oriented Σ_h-bundle $E_{\varphi,\psi}$ over P whose structure group is contained in $C_h(t)$, and monodromies along α and β are φ and ψ in $\pi_0 C_h(t)$, respectively. Since coordinate transformations commute with the deck transformation t, we can define a fiberwise $\mathbb{Z}/d\mathbb{Z}$-action on $E_{\varphi,\psi}$. Since the structure group is in $C_h(t)$, not only $H_1(\Sigma_h; \mathbb{C})$ but also each eigenspace V^{η^j} is a local system on P. We can extend the intersection form as a skew-hermitian form $H_1(\Sigma_h; \mathbb{C}) \otimes H_1(\Sigma_h; \mathbb{C}) \to \mathbb{C}$ defined by

$$(x_1 + x_2\boldsymbol{i}) \cdot (y_1 + y_2\boldsymbol{i}) = x_1 \cdot y_1 + x_2 \cdot y_2 + (x_1 \cdot y_2 - x_2 \cdot y_1)\boldsymbol{i}.$$

For $v \in V^{\eta^j}$ and $w \in V^{\eta^k}$ $(1 \le j \le d-1,\ 1 \le k \le d-1)$,

$$\begin{aligned}(tv) \cdot w &= (\omega^j v) \cdot w = \omega^{-j}(v \cdot w),\\ (tv) \cdot w &= v \cdot (t^{-1} w) = v \cdot (\omega^{-k} w) = \omega^{-k}(v \cdot w).\end{aligned}$$

Since ω^{-j} is not equal to ω^{-k}, we have $v \cdot w = 0$. Hence, $H_1(\Sigma_h; \mathbb{C})$ decomposes into an orthogonal sum of subspaces $\{V^{\omega^j}\}_{j=1}^{d-1}$. By restricting the intersection form on $H_1(\Sigma_h; \mathbb{C})$ to V^{η^j}, we can define a hermitian form on $H_1(P; V^{\eta^j})$. By Theorem 2.9, we have a 2-cocycle on $\pi_0 C_h(t)$ as follows.

Lemma 3.1. *Let j be an integer such that $1 \le j \le m-1$. The map*

$$\tau_{m,d,j} : \pi_0 C_h(t) \times \pi_0 C_h(t) \to \mathbb{Z}$$

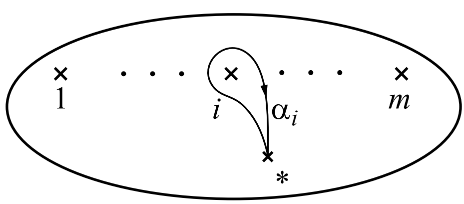


Figure 4. A loop α_i.

defined by

$$\tau_{m,d,j}(\varphi,\psi)=\mathrm{Sign}(H_1(P;V^{\eta^j}))$$

is a 2-cocycle, where V^{η^j} *is the local system on* P *induced from the oriented* Σ_h*-bundle* $E_{\varphi,\psi}\to P$.

Proof. The proof is the same as for [Meyer 1972, p. 43, Equation (0)]. Applying additivity of signatures to two oriented Σ_h-bundles on P, we can see that $\tau_{m,d,j}$ satisfies

$$\tau_{m,d,j}(\varphi_1,\varphi_2)+\tau_{m,d,j}(\varphi_1\varphi_2,\varphi_3)=\tau_{m,d,j}(\varphi_1,\varphi_2\varphi_3)+\tau_{m,d,j}(\varphi_2,\varphi_3),$$

for $\varphi_1,\varphi_2,\varphi_3\in\pi_0C_h(t)$. $\square$

Since the deck transformation t acts on $H^1(P,\partial P;V^{\eta^j})$ by multiplication of η^j, we can calculate $\mathbb{Z}/d\mathbb{Z}$-signature as

$$\mathrm{Sign}(H_1(P;V^{\eta^j}),t^k)=\eta^{kj}\,\mathrm{Sign}(H_1(P;V^{\eta^j}))=\eta^{kj}\tau_{m,d,j}(\varphi,\psi),$$

for $0\le k\le m-1$. Moreover Meyer [1972, Satz I.2.2] proved $\mathrm{Sign}(E_{\varphi,\psi},t^k)=\mathrm{Sign}(H_1(P;H^1(\Sigma_h;\mathbb{C})),t^k)$. Hence we have:

Lemma 3.2. *For* $0\le k\le m-1$,

$$\mathrm{Sign}(E_{\varphi,\psi},t^k)=\sum_{j=1}^{d-1}\eta^{kj}\tau_{m,d,j}(\varphi,\psi).$$

The symmetric mapping class groups. A diffeomorphism $f:\Sigma_h\to\Sigma_h$ in $C_h(t)$ induces a diffeomorphism $\bar f:S^2\to S^2$ which satisfies the commutative diagram

$$\begin{tikzcd}
\Sigma_h \arrow[r, "f"] \arrow[d, "p_d"'] & \Sigma_h \arrow[d, "p_d"'] \\
S^2 \arrow[r, "\bar f"] & S^2.
\end{tikzcd}$$

Moreover since $\bar f$ satisfies $p_d^{-1}(q)=p_d^{-1}(\bar f(q))$ for any $q\in S^2$, we have

$$\bar f\in\mathrm{Diff}_+(S^2,\{q_i\}_{i=1}^m).$$

Therefore we have a natural homomorphism $\mathcal{P}:\pi_0C_h(t)\to\mathcal{M}_0^m$ which maps $[f]$ to $[\bar f]$. By a method similar to [Birman and Hilden 1971, Theorem 1] (see also [Birman and Hilden 1973, Section 5]), we have:

Lemma 3.3. *Let* $m\ge 4$. *The sequence*

$$1\longrightarrow\mathbb{Z}/d\mathbb{Z}\longrightarrow\pi_0C_h(t)\xrightarrow{\mathcal{P}}\mathcal{M}_0^m\longrightarrow 1$$

is exact.

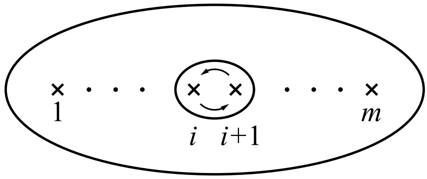


Figure 5. The diffeomorphism s_i.

Let $s_i : S^2 \to S^2$ be a half twist of the disk which exchanges the points q_i and q_{i+1} as in Figure 5.

We denote by $\sigma_i \in \mathcal{M}_0^m$ the mapping class represented by s_i. By lifting s_i, we have a unique diffeomorphism $\tilde{s}_i : \Sigma_h \to \Sigma_h$ which satisfies $\operatorname{supp} \tilde{s}_i = p_d^{-1}(\operatorname{supp} s_i)$. Let us denote the mapping class of $\tilde{s}_i$ by $\tilde{\sigma}_i \in \pi_0 C_h(t)$. Note that when $d = 2$, $\tilde{\sigma}_i$ is the Dehn twist along a nonseparating simple closed curve.

Lemma 3.4. *The set $\{\tilde{\sigma}_i\}_{i=1}^{m-1} \subset \pi_0 C_h(t)$ generates the group $\pi_0 C_h(t)$.*

Proof. Since $\{\sigma_i\}_{i=1}^{m-1}$ generates the group $\mathcal{M}_0^m$, it suffices to represent $[t] \in \pi_0 C_h(t)$ as a product of $\{\sigma_i\}_{i=1}^{m-1}$. Let $C_h^{(*)}(t)$ denote the subgroup of $C_h(t)$ defined by $C_h^{(*)}(t) = \{f \in C_h(t) \mid f(p_d^{-1}(*)) = p_d^{-1}(*)\}$. In this proof, by abuse of terminology, we use the term "Dehn twist" both for a diffeomorphism and for a mapping class. The diffeomorphism $s_1 \cdots s_{m-2} s_{m-1}^2 s_{m-2} \cdots s_1$ in $\operatorname{Diff}_+(S^2, \{q_i\}_{i=1}^m)$ is isotopic to the product of Dehn twists $t_c^{-1} t_{c'}$ in Figure 6, and it is also isotopic to the Dehn twist t_d^{-1}.

Therefore the lift $\tilde{s}_1 \cdots \tilde{s}_{m-2} \tilde{s}_{m-1}^2 \tilde{s}_{m-2} \cdots \tilde{s}_1$ is isotopic to some lift $\tilde{f}_1 : \Sigma_h \to \Sigma_h$ of t_d^{-1}. Since we can choose the isotopy in $\operatorname{Diff}_+(S^2, \{q_i\}_{i=1}^m)$ so that it does not move $*$, the lift $\tilde{f}_1$ fixes $p^{-1}(*)$ pointwise. Let D be the closed disk which is bounded by d and contains $*$, and let $\tilde{f}_2$ denote the lift of t_d which satisfies $\operatorname{supp} \tilde{f}_2 \subset p^{-1}(D)$. Since $f_1 f_2$ is a lift of the identity map of S^2, and the action of $\tilde{f}_2$ on $p^{-1}(*)$ coincides with that of t, we have $\tilde{f}_1 \tilde{f}_2 = t \in \operatorname{Diff}_+ \Sigma_h$. Since t_d is isotopic to the identity map in $\operatorname{Diff}_+ \Sigma_h$, we have $[\tilde{f}_2] = 1 \in \pi_0 C_h(t)$. Thus we obtain

$$\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-2} \tilde{\sigma}_{m-1}^2 \tilde{\sigma}_{m-2} \cdots \tilde{\sigma}_1 = [\tilde{f}_1] = [\tilde{f}_1 \tilde{f}_2] = [t] \in \pi_0 C_h(t). \qquad \square$$

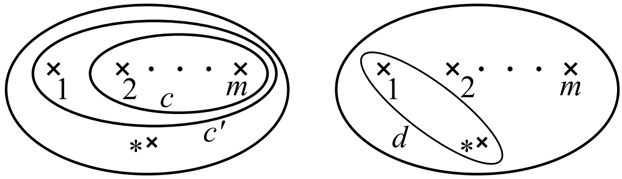


Figure 6. The curves c, c', d.

The cobounding function of the cocycles $\tau_{m,d,j}$. Recall that, for an integer d with $d \mid m$, we have a covering space $p_d : \Sigma_h \to S^2$. Let $g = (m-1)(m-2)/2$. If we consider the case when $d = m$, p_d is the m-cyclic covering on S^2 whose genus of the covering surface is g. Thus we identify it with the surface Σ_g, and denote the covering by $p : \Sigma_g \to S^2$.

Since the quotient space $\Sigma_g/\langle t^d\rangle$ is also a d-cyclic covering of S^2 with m branched points, we can identify $\Sigma_h \cong \Sigma_g/\langle t^d\rangle$. Since a diffeomorphism $f \in C_g(t)$ induces a diffeomorphism $\bar{f}$ on $\Sigma_g/\langle t^d\rangle$ which commutes with t, we have a natural homomorphism $\mathcal{P} : \pi_0 C_g(t) \to \pi_0 C_h(t)$ which maps $[f]$ to $[\bar{f}]$. Since $H^*(\pi_0 C_h(t); \mathbb{Q}) \cong H^*(\mathcal{M}_0^m; \mathbb{Q})$, and $H^*(\mathcal{M}_0^m; \mathbb{Q})$ is trivial (see [Cohen 1987] Corollary 2.2), there exists a unique cobounding function of $\tau_{m,d,j}$. Denote it by $\phi_{m,d,j} : \pi_0 C_h(t) \to \mathbb{Q}$. Since $\tau_{m,d,j}$ is bounded, the cobounding function $\phi_{m,d,j}$ is a quasimorphism.

Remark 3.5. Gambaudo and Ghys [2005] already constructed almost the same quasimorphisms on the mapping class groups of pointed disks, called ω-signatures. They calculated the value of their quasimorphisms for an element similar to

$$\tilde{\sigma}_1\tilde{\sigma}_2 \cdots \tilde{\sigma}_{r-1} \in \pi_0 C_h(t)$$

in [Gambaudo and Ghys 2005, Proposition 5.2].

Remark 3.6. This construction is also similar to higher-order signature cocycles in Cochran, Harvey and Horn's paper [Cochran et al. 2012]. They considered von Neumann signatures of surface bundles whose fibers are nonfinite regular coverings on a surface with boundary.

Let us recall a natural homomorphism $\pi_0 C_h(t) \to \mathcal{M}_h$ defined by forgetting symmetries of mapping classes. It maps a mapping class $[f] \in \pi_0 C_h(t)$ to $[f] \in \mathcal{M}_h$, and is injective as shown in Birman and Hilden [1973, Theorem 1]. In particular, if we consider the case when m is even and the double covering $p_2 : \Sigma_h \to S^2$, this homomorphism induces isomorphism between $\pi_0 C_h(t)$ and $\mathcal{H}_h$. In this case, the eigenspace V^{-1} coincides with $H_1(\Sigma_h; \mathbb{C})$. Thus we have:

Remark 3.7. When m is even, $\phi_{m,2,1} : \pi_0 C_h(t) \to \mathbb{Q}$ is equal to the Meyer function $\phi_h : \mathcal{H}_h \to \mathbb{Q}$ on the hyperelliptic mapping class group, under the natural isomorphism $\pi_0 C_h(t) \cong \mathcal{H}_h$.

Lemma 3.8. *For $1 \leq j \leq d-1$ and $\varphi \in \pi_0 C_g(t)$,*

$$\phi_{m,m,mj/d}(\varphi) = \phi_{m,d,j}(\mathcal{P}(\varphi)).$$

Proof. Since $H_1(\pi_0 C_g(t); \mathbb{Q})$ is trivial, it suffices to show that

$$\tau_{m,m,mj/d}(\varphi, \psi) = \tau_{m,d,j}(\mathcal{P}(\varphi), \mathcal{P}(\psi)) \quad \text{for } \varphi, \psi \in \pi_0 C_g(t).$$

If $f : E \to P$ is an oriented Σ_g-bundle with structure group $C_g(t)$, the induced map $\bar{f} : E/\langle t^d \rangle \to P$ is an oriented Σ_h-bundle with structure group $C_h(t)$. If we denote the monodromies of f along α and β by φ and ψ, the ones of $\bar{f}$ are $\mathcal{P}(\varphi)$ and $\mathcal{P}(\psi)$.

Let ω be the m-th root of unity $\exp(2\pi \boldsymbol{i}/m)$, and let $q_d : \Sigma_g \to \Sigma_g/\langle t^d \rangle$ denote the projection. To distinguish eigenspaces of $H_1(\Sigma_g; \mathbb{C})$ and $H_1(\Sigma_h; \mathbb{C})$ of the action by t, we denote them by $(V_g)^z$ and $(V_h)^z$ instead of V^z, respectively. The projection q_d induces the isomorphism $H_1(\Sigma_g; \mathbb{C})^{\langle t^d \rangle} \cong H_1(\Sigma_h; \mathbb{C})$. Moreover we have $(V_g)^{\omega^{mj/d}} \cong (V_h)^{\eta^j}$. Hence it also induces a natural isomorphism between $H_1(P; (V_g)^{\omega^{mj/d}})$ and $H_1(P; (V_h)^{\eta^j})$, where $(V_g)^{\omega^{mj/d}}$ and $(V_h)^{\eta^j}$ are local systems coming from f and $\bar{f}$.

Let $\tilde{a}$, $\tilde{b}$ be loops in $\Sigma_g - \{q_i\}_{i=1}^m$. We may assume that $q_d(\tilde{a}) \cup q_d(\tilde{b})$ has no triple point. Then the intersection number $[q_d(\tilde{a})] \cdot [q_d(\tilde{b})]$ in Σ_h coincides with $[q_d^{-1}(q_d(\tilde{a}))] \cdot [\tilde{b}]$ in Σ_g. Hence we have

$$\begin{aligned}\sum_{i=0}^{m/d-1} [(t^{di})_* \tilde{a}] \cdot \sum_{j=0}^{m/d-1} [(t^{dj})_* \tilde{b}] &= \sum_{i=0}^{m/d-1} \sum_{j=0}^{m/d-1} [(t^{di-dj})_* \tilde{a}] \cdot [\tilde{b}] \\ &= \frac{m}{d}[q_d^{-1}(q_d(\tilde{a}))] \cdot [\tilde{b}] = \frac{m}{d}[q_d(\tilde{a})] \cdot [q_d(\tilde{b})].\end{aligned}$$

Therefore the isomorphism $H_1(\Sigma_g; \mathbb{C})^{\langle t^d \rangle} \cong H_1(\Sigma_h; \mathbb{C})$ induced by the quotient map $q_d : \Sigma_g \to \Sigma_h$ preserves the intersection form up to constant multiple. Thus it also preserves the intersection forms on $H_1(P; (V_g)^{\omega^{mj/d}})$ and $H_1(P; (V_h)^{\eta^j})$, and we obtain

$$\begin{aligned}\tau_{m,m,mj/d}(\varphi, \psi) &= \mathrm{Sign}(H_1(P; (V_g)^{\omega^{mj/d}})) \\ &= \mathrm{Sign}(H_1(P; (V_h)^{\eta^j})) \\ &= \tau_{m,d,j}(\mathcal{P}(\varphi), \mathcal{P}(\psi)).\end{aligned}$$

□

By Lemma 3.8, it suffices to consider the case when $d = m$. We shorten $\tau_{m,m,j}$ and $\phi_{m,m,j}$ to $\tau_{m,j}$ and $\phi_{m,j}$.

Lemma 3.9. $$\phi_{m,j}(\varphi) = \phi_{m,m-j}(\varphi).$$

Proof. By taking complex conjugates, we have an isomorphism $i : V^{\omega^j} \cong V^{\omega^{m-j}}$. Moreover it induces the isomorphism $i_* : H_1(P; V^{\omega^j}) \cong H_1(P; V^{\omega^{m-j}})$.

Let us denote the hermitian form on $H_1(P; V^{\omega^j})$ by $\langle\ ,\ \rangle_j$. By the definition of the hermitian form, we have $\langle x, y \rangle_j = \overline{\langle i_* x, i_* y \rangle_{m-j}}$ for $x, y \in H_1(P; V^{\omega^j})$, where $\bar{z}$ is a complex conjugate of $z \in \mathbb{C}$. Thus the signatures of the hermitian forms $\langle\ ,\ \rangle_j$ and $\langle\ ,\ \rangle_{m-j}$ coincide, and the cobounding functions of $\tau_{m,j}$ and $\tau_{m,m-j}$ also coincide. □

4. Defects of homogeneous quasimorphisms

In this section, we will prove Theorems 1.3 and 1.6. On page 336, we give an inequality between the defects of a quasimorphism and its homogenization when the quasimorphism is antisymmetric and a class function (Lemma 4.1) and prove Theorem 1.3(i). On page 337, we prove Theorem 1.3(ii) by giving a lower bound on the defect of $\phi_{m,m/2} : \pi_0 C_g(t) \to \mathbb{R}$, which is the cobounding function of the 2-cocycle $\tau_{m,m/2}$. On page 337, we prove Theorem 1.6.

Proof of Theorem 1.3(i). Endo [2000, Proposition 3.1] showed that the Meyer function $\phi_g : \mathcal{H}_g \to \mathbb{Q}$ satisfies the conditions in Lemma 4.1. The quasimorphisms $\bar{\phi}_{m,j}$ also satisfy these conditions.

Turaev [1985] defined another 2-cocycle on the symplectic group. Endo and Nagami [2005, Proposition A.3] showed that Turaev's cocycle coincides with the Meyer cocycle up to sign. Since Turaev's cocycle is defined by the signature on a vector space of rank less than or equal to $m-2$. A similar argument shows $D(\phi_{m,j}) \le m-2$. Thus Theorem 1.3(i) follows from Lemma 4.1 below.

Lemma 4.1. *Let G be a group, and $\phi : G \to \mathbb{R}$ a quasimorphism satisfying*

$$\phi(xyx^{-1}) = \phi(y), \quad \phi(x^{-1}) = -\phi(x).$$

Then we have

$$D(\bar{\phi}) \le D(\phi),$$

where $\bar{\phi}$ is the homogenization of ϕ.

Proof of Lemma 4.1. Without loss of generality, we may assume that the quasimorphism $\phi : G \to \mathbb{R}$ is antisymmetric:

$$\phi(x^{-1}) = -\phi(x).$$

Otherwise pass to the antisymmetrization $\phi' : G \to \mathbb{R}$ defined by

$$\phi'(x) = \frac{\phi(x) - \phi(x^{-1})}{2},$$

which satisfies

$$D(\phi') \le D(\phi), \text{ and } \bar{\phi}' = \bar{\phi}.$$

For any $x, y \in G$, we have

$$\begin{aligned}\phi([x, y]) &= |\phi([x, y]) - \phi(y) + \phi(y)| \\ &= |\phi(xyx^{-1}y^{-1}) - \phi(xyx^{-1}) - \phi(y^{-1})| \le D(\phi).\end{aligned}$$

Thus for any $g \in [G, G]$,

$$|\phi(g)| \le (2\,\mathrm{cl}(g) - 1) D(\phi).$$

As observed by Bavard [1991, Lemma 3.6],

$$\mathrm{cl}(x^n y^n (xy)^{-n}) \leq \frac{n}{2},$$

for every $n \geq 0$. Therefore we have $|\phi(x^n y^n (xy)^{-n})| \leq (n-1)D(\phi)$. Hence

$$\begin{aligned}|\delta\bar{\phi}(x, y)| &= \lim_{n\to\infty}\left|\frac{\phi(x^n)+\phi(y^n)-\phi(x^n y^n)}{n}\right| \\ &= \lim_{n\to\infty}\left|\frac{\phi(x^n y^n (xy)^{-n})}{n}\right| \leq D(\phi).\end{aligned}$$

□

Proof of Theorem 1.3(ii). Let m be an even number greater than or equal to 4. By Remark 3.7, we consider the Meyer function ϕ_g on the hyperelliptic mapping class group $\mathcal{H}_g$ instead of $\phi_{m,m/2}$.

Lemma 4.2 [Barge and Ghys 1992, Proposition 3.5]. *For any* $A \in \mathrm{Sp}(2g; \mathbb{Z})$,

$$\mathrm{Sign}(\langle\ ,\ \rangle_{A^k, A}) = \mathrm{Sign}\biggl(-J\sum_{i=1}^{k}(A^i - A^{-i})\biggr).$$

Let c_i, d_i^+, and d_i^- denote the simple closed curves in Figure 7. For simplicity, we also denote by c_i, d_i^+, and d_i^- the Dehn twists along these curves.

To prove Theorem 1.3(ii), it suffices to show the following.

Lemma 4.3. $\qquad \delta\bar{\phi}_g(c_2^2 c_4^2 \cdots c_{2g}^2,\ d_1^+ d_1^- d_2^+ d_2^- \cdots d_g^+ d_g^-) = -2g.$

Proof of Lemma 4.3. Since the pairs (c_i, c_j), $(d_i^+ d_i^-, d_j^+ d_j^-)$, and $(c_i, d_j^+ d_j^-)$ mutually commute when $i \neq j$, the expression in the lemma equals

$$\begin{aligned}&\bar{\phi}_g(c_2^2 c_4^2 \cdots c_{2g}^2) + \bar{\phi}_g(d_1^+ d_1^- d_2^+ d_2^- \cdots d_g^+ d_g^-) - \bar{\phi}_g(c_2^2 d_1^+ d_1^- c_4^2 d_2^+ d_2^- \cdots c_{2g}^2 d_g^+ d_g^-) \\ &\qquad = \sum_{i=1}^{g}(\bar{\phi}_g(c_{2i}^2) + \bar{\phi}_g(d_i^+ d_i^-) - \bar{\phi}_g(c_{2i}^2 d_i^+ d_i^-)).\end{aligned}$$

Hence it suffices to prove $\bar{\phi}_g(c_{2i}^2)+\bar{\phi}_g(d_i^+ d_i^-)-\bar{\phi}_g(c_{2i}^2 d_i^+ d_i^-) = -2$ for $1 \leq i \leq g$.

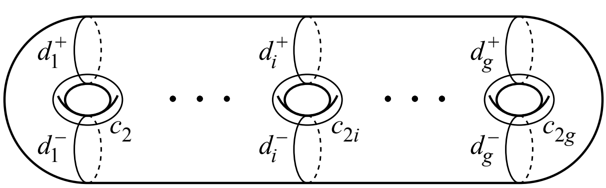


Figure 7. Curves in Σ_g.

Since $\rho(d_i^+) = \rho(d_i^-)$, we have

$$\begin{aligned}
&\bar{\phi}_g(c_{2i}^2) + \bar{\phi}_g(d_i^+ d_i^-) - \bar{\phi}_g(c_{2i}^2 d_i^+ d_i^-) \\
&\quad = -\lim_{n\to\infty} \frac{1}{n}\big\{\phi_g((c_{2i}^2 d_i^+ d_i^-)^n) - \phi_g((c_{2i}^2)^n) - \phi_g((d_i^+ d_i^-)^n)\big\} \\
&\quad = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n-1}\big\{\tau_g((c_{2i}^2 d_i^+ d_i^-)^k, c_{2i}^2 d_i^+ d_i^-) - \tau_g(c_{2i}^{2i}, c_{2i}^2) - \tau_g((d_i^+ d_i^-)^i, d_i^+ d_i^-)\big\} \\
&\qquad\qquad + \tau_g(c_{2i}^2, d_i^+ d_i^-) \\
&\quad = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n-1}\big\{\tau_g((c_{2i}^2 (d_i^+)^2)^k, c_{2i}^2 (d_i^+)^2)\tau_g(c_{2i}^{2i}, c_{2i}^2) - \tau_g((d_i^+)^{2i}, (d_i^+)^2)\big\} \\
&\qquad\qquad + \tau_g(c_{2i}^2, (d_i^+)^2).
\end{aligned}$$

There exists a mapping class ψ_i such that $\psi_i c_{2i}\psi_i^{-1} = c_2$ and $\psi_i d_i^+ \psi_i^{-1} = d_i^+$ for $i = 2, \ldots, g$. Since the Meyer cocycle satisfies the property

$$\tau_g(xyx^{-1}, xzx^{-1}) = \tau_g(y, z)$$

for $x, y, z \in \mathcal{M}_g$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{n-1} \frac{1}{n}\{\tau_g((c_{2i}^2 (d_i^+)^2)^k, c_{2i}^2 (d_i^+)^2) - \tau_g(c_{2i}^{2i}, c_{2i}^2) - \tau_g((d_i^+)^{2i}, (d_i^+)^2)\} \\
&\qquad\qquad + \tau_g(c_{2i}^2, (d_i^+)^2) \\
&\quad = \lim_{n\to\infty} \sum_{k=1}^{n-1} \frac{1}{n}\{\tau_g((c_2^2 (d_1^+)^2)^k, c_2^2 (d_1^+)^2) - \tau_g(c_2^{2i}, c_2^2) - \tau_g((d_1^+)^{2i}, (d_1^+)^2)\} \\
&\qquad\qquad + \tau_g(c_2^2, (d_1^+)^2).
\end{aligned}$$

Let us consider the case when $g = 1$. Since $\rho(c_2^2) = \left(\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right)$, $\rho((d_1^+)^2) = \left(\begin{smallmatrix}1 & 0\\ -2 & 1\end{smallmatrix}\right)$ and $\rho(c_2^2 (d_1^+)^2) = \left(\begin{smallmatrix}-3 & 2\\ -2 & 1\end{smallmatrix}\right)$, we have

$$\begin{aligned}
-J\sum_{k=1}^{n}(\rho(c_2^{2k}) - \rho(c_2^{-2k})) &= \begin{pmatrix}0 & 0\\ 0 & 2n(n+1)\end{pmatrix}, \\
-J\sum_{k=1}^{n}(\rho(d_1^+)^{2k} - \rho(d_1^+)^{-2k}) &= \begin{pmatrix}2n(n+1) & 0\\ 0 & 0\end{pmatrix}, \\
-J\sum_{k=1}^{n}(\rho((c_2^2 (d_1^+)^2)^k) - \rho((c_2^2 (d_1^+)^2)^{-k})) &= \sum_{k=1}^{n} 4k(-1)^k \left(\begin{smallmatrix}-1 & 1\\ 1 & -1\end{smallmatrix}\right) \\
&= \{(-1)^n(2n+1) - 1\}\left(\begin{smallmatrix}-1 & 1\\ 1 & -1\end{smallmatrix}\right).
\end{aligned}$$

By Lemma 4.2 we obtain

(1) $$\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{\tau_g(c_2^{2k},c_2^2)}{n}=\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{\tau_g((d_1^+)^{2k},(d_1^+)^2)}{n}=1,$$

(2) $$\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{\tau_g((c_2^2(d_1^+)^2)^k,c_2^2(d_1^+)^2)}{n}=0.$$

When $g \geq 2$, the same calculation also shows (1). It is an easy calculation to show that

$$\tau_g(c_2^2, d_1^+d_1^-)=0.$$

Therefore we obtain

$$\bar{\phi}_g(c_{2i}^2)+\bar{\phi}_g(d_i^+d_i^-)-\bar{\phi}_g(c_{2i}^2d_i^+d_i^-)=-2.$$ □

In the same way as for (1), we have $\tau_g(s_0^i, s_0)=1$. Hence we obtain

$$\bar{\phi}_g(s_0)=-\lim_{n\to\infty}\frac{\sum_{i=1}^{n-1}\tau_g(s_0^i,s_0)}{n}+\phi_g(s_0)=-1+\phi_g(s_0)$$

and

$$\bar{\phi}_g(s_h)=\phi_g(s_0).$$

By [Endo 2000, Lemmas 3.3 and 3.5], we have

$$\bar{\phi}_g(t_{s_0})=-\frac{g}{2g+1},\quad \text{and}\quad \bar{\phi}_g(t_{s_h})=-\frac{4h(g-h)}{2g+1}.$$

Remark 4.4. By Theorems 1.3 and 2.4, $\bar{\phi}_g$ gives the lower bounds for $\mathrm{scl}_{\mathcal{H}_g}(t_{s_h})$ $(j=0,\dots,g-1)$ corresponding to ones given in [Monden 2012].

Remark 4.5. By Theorems 1.7 and 2.4 and Remark 4.4, we have $\mathrm{scl}_{\mathcal{M}_1}(t_c)=\frac{1}{12}$. Let $\rho:\mathcal{M}_1\cong \mathrm{SL}(2,\mathbb{Z})\to \mathrm{PSL}(2,\mathbb{Z})$ be the natural quotient map. It is easily seen that for all $x\in\mathcal{M}_1$, $\mathrm{scl}_{\mathcal{M}_1}(x)=\mathrm{scl}_{\mathrm{PSL}(2,\mathbb{Z})}(\rho(x))$. Louwsma [2011] determined $\mathrm{scl}_{\mathrm{PSL}(2,\mathbb{Z})}(y)=\frac{1}{12}$ for $y=\rho(t_c)$.

Proof of Theorem 1.6. If $x\in\mathcal{H}_g$ satisfies $|\bar{\phi}_g(x)|=D(\bar{\phi}_g)$ and $|\phi(x)|\leq D(\phi)$ for any homogeneous quasimorphism $\phi:\mathcal{H}_g\to\mathbb{R}$, we obtain $\mathrm{scl}(x)=\frac{1}{2}$ by Bavard's duality theorem (Theorem 2.4). We will show that

$$x=c^{2g+8}(d_2^+d_2^-\cdots d_{g-1}^+d_{g-1}^-)^2(s_1\cdots s_{g-1})^{-1}$$

satisfies this property.

Firstly we will prove

$$\sum_{j=1}^{g-1}\phi(s_j)=\sum_{i=1}^{g}(\phi(c_{2i}^2d_i^-d_i^+)-\phi(d_i^-d_i^+))$$

for any homogeneous quasimorphism $\phi : \mathcal{H}_g \to \mathbb{R}$. By Lemma 2.8, we have

$$(d_1^+ c_2 d_1^-)^4 = s_1, \quad (d_g^+ c_{2g} d_g^-)^4 = s_{g-1},$$
$$(d_i^+ c_{2i} d_i^-)^4 = s_{i-1} s_i \quad (i = 2, \dots, g-1).$$

Since c_{2i} commutes with s_j, $(c_2 d_1^- d_1^+ c_2 d_1^- d_1^+)^2 = s_1$, $(c_{2i} d_i^- d_i^+ c_{2i} d_i^- d_i^+)^2 = s_{i-1} s_i$, and $(c_{2g} d_g^- d_g^+ c_{2g} d_g^- d_g^+)^2 = s_{g-1}$. By Lemma 2.6, $c_{2i} d_i^- d_i^+ c_{2i}$ commutes with $d_i^- d_i^+$ for $i = 1, \dots, g$, as is easy to check. Therefore $(c_2 d_1^- d_1^+ c_2)^2 = s_1 (d_1^- d_1^+)^{-2}$, $(c_{2i} d_i^- d_i^+ c_{2i})^2 = s_{i-1} s_i (d_i^- d_i^+)^{-2}$, and $(c_{2g} d_g^- d_g^+ c_{2g})^2 = s_{g-1} (d_g^- d_g^+)^{-2}$. These equations give

$$2\phi(c_2^2 d_1^- d_1^+) = \phi(s_1) - 2\phi(d_1^- d_1^+),$$
$$2\phi(c_{2i}^2 d_i^- d_i^+) = \phi(s_{i-1}) + \phi(s_i) - 2\phi(d_i^- d_i^+),$$
$$2\phi(c_{2g}^2 d_g^- d_g^+) = \phi(s_{g-1}) - 2\phi(d_g^- d_g^+).$$

Thus we obtain $\sum_{j=1}^{g-1} \phi(s_j) = \sum_{i=1}^{g} (\phi(c_{2i}^2 d_i^- d_i^+) - \phi(d_i^- d_i^+))$.

Secondly we will prove $\bar{\phi}_g(x) = D(\bar{\phi}_g)$. The curves $c, s_1, \dots, s_{g-1}, d_2^+, d_2^-, \dots, d_{g-1}^+, d_{g-1}^-$ are mutually disjoint, and c_i is conjugate to c. Hence, by Lemma 2.3(i) and (ii), we have

$$\begin{aligned}\phi(x) &= (g+4)\phi(c^2) + 2\sum_{i=2}^{g-1} \phi(d_i^+ d_i^-) - \sum_{j=1}^{g-1} \phi(s_i) \\ &= \sum_{i=1}^{g} (\phi(c_{2i}^2) + \phi(d_i^+ d_i^-) - \phi(c_{2i}^2 d_i^- d_i^+)).\end{aligned}$$

In the proof of Lemma 4.3, we showed

$$\sum_{i=1}^{g} (\bar{\phi}_g(c_{2i}^2) + \bar{\phi}_g(d_i^+ d_i^-) - \bar{\phi}_g(c_{2i}^2 d_i^- d_i^+)) = -2g = -D(\bar{\phi}_g).$$

Thus we obtain $|\bar{\phi}_g(x)| = D(\bar{\phi}_g)$.

Lastly we prove $\phi(x) \le D(\phi)$ for any homogeneous quasimorphism $\phi : \mathcal{H}_g \to \mathbb{R}$:

$$\begin{aligned}D(\phi) &\ge |\delta(c_2^2 \cdots c_{2g}^2, d_1^+ d_1^- \cdots d_g^+ d_g^-)| \\ &= \left|\phi(c_2^2 \cdots c_{2g}^2) + \phi(d_1^+ d_1^- \cdots d_g^+ d_g^-) - \phi(c_2^2 \cdots c_{2g}^2 d_1^+ d_1^- \cdots d_g^+ d_g^-)\right| \\ &= \left|\phi(c_2^2 \cdots c_{2g}^2) + \phi(d_1^+ d_1^- \cdots d_g^+ d_g^-) - \phi((c_2^2 d_1^+ d_1^-) \cdots (c_{2g}^2 d_g^+ d_g^-))\right| \\ &= \left|\sum_{i=1}^{g} (\phi(c_{2i}^2) + \phi(d_i^+ d_i^-) - \phi(c_{2i}^2 d_i^+ d_i^-))\right| = |\phi(x)|. \qquad \square\end{aligned}$$

5. Proof of Theorem 1.7

Let $c_1, \dots, c_{2g+2}$ be nonseparating simple closed curves on Σ_g as in Figure 2 and let ϕ be a homogeneous quasimorphism on $\mathcal{H}_g$. For simplicity of notation, we write t_i instead of t_{c_i} By $\iota = \iota^{-1}$, we have $t_{2g+1}^2 t_{2g} \cdots t_2 t_1^2 = (t_{2g} \cdots t_2)^{-1} \iota^{-1}$. Since each of the two boundary components of a regular neighborhood of $c_2 \cup c_3 \cup \cdots \cup c_{2g}$ is c_{2g+2} by Lemma 2.8, we have $(t_{2g} \cdots t_2)^{2g} = t_{2g+2}^2$. Note that this relation holds in $\mathcal{H}_g$. Therefore, by Definition 2.2 and Lemma 2.3, we have

$$\phi(t_{2g+1}^2 t_{2g} \cdots t_2 t_1^2) = -\phi(t_{2g} \cdots t_2) + \phi(\iota^{-1}) = -\frac{1}{g}\phi(t_{2g+2}). \tag{3}$$

Applying Lemma 2.3(i) and 2.6(i), we can move the factors with single and double underlines alternatively as follows.

$$\begin{aligned}
\phi(t_{2g+1}^2 t_{2g} \cdots t_3 t_2 \underline{t_1^2}) &= \phi(\underline{t_1^2} t_{2g+1}^2 t_{2g} \cdots t_3 t_2) && \text{(by Lemma 2.3)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_3 \underline{t_1^2} \underline{\underline{t_2}}) && \text{(by Lemma 2.6)}\\
&= \phi(\underline{\underline{t_2}} t_{2g+1}^2 t_{2g} \cdots t_4 t_3 t_1^2) && \text{(by Lemma 2.3)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_4 \underline{\underline{t_2}} \underline{t_3 t_1^2}) && \text{(by Lemma 2.6)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_6 t_5 \underline{t_3 t_1^2} \underline{\underline{t_4 t_2}}) && \text{(by Lemma 2.3 and 2.6)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_6 \underline{\underline{t_4 t_2}} \underline{t_5 t_3 t_1^2}) && \text{(by Lemmas 2.3 and 2.6)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_7 \underline{t_5 t_3 t_1^2} \underline{\underline{t_6 t_4 t_2}}) && \text{(by Lemmas 2.3 and 2.6)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_8 \underline{\underline{t_6 t_4 t_2}} \underline{t_7 t_5 t_3 t_1^2}) && \text{(by Lemmas 2.3 and 2.6)}\\
&= \phi(t_{2g+1}^2 t_{2g} \cdots t_9 \underline{t_7 t_5 t_3 t_1^2} \underline{\underline{t_8 t_6 t_4 t_2}}) && \text{(by Lemmas 2.3 and 2.6)}\\
&\;\;\vdots\\
&= \phi\big((t_{2g+1}^2 t_{2g-1} \cdots t_5 t_3 t_1^2)(t_{2g} t_{2g-4} \cdots t_4 t_2)\big).
\end{aligned}$$

From Definition 2.2 and Equation (3),

$$\begin{aligned}
D(\phi) &\geq |\phi((t_{2g+1}^2 \cdots t_3 t_1^2)(t_{2g} \cdots t_4 t_2)) - \phi(t_{2g+1}^2 \cdots t_3 t_1^2) - \phi(t_{2g} \cdots t_4 t_2)|\\
&= \left| -\frac{1}{g}\phi(t_{2g+2}) - \phi(t_{2g+1}^2 \cdots t_3 t_1^2) - \phi(t_{2g} \cdots t_4 t_2) \right|,
\end{aligned}$$

where $D(\phi)$ is the defect of ϕ. From Lemmas 2.3, 2.5 and 2.6 we have

$$D(\phi) \geq \left| \frac{1}{g}\phi(t_1) + (g+3)\phi(t_1) + g\phi(t_1) \right| = (2g + 3 + \frac{1}{g})|\phi(t_1)|.$$

By Theorem 2.4 we have $\mathrm{scl}_{\mathcal{H}_g}(t_1) \leq \dfrac{1}{(2(2g+3+1/g))}$. This completes the proof of Theorem 1.7. □

Remark 5.1. By a similar argument to the proof of Theorem 1.7, we can show that

$$\mathrm{scl}_{\mathcal{M}_0^m}(\sigma_1) = \frac{1}{2\{m+1+2/(m-2)\}} \quad \text{for all } m \geq 4.$$

6. Calculation of quasimorphisms

In this section, we prove Theorem 1.1. To prove it, we perform a straightforward and elementary calculation of the hermitian form $\langle\ ,\ \rangle_{\tilde{\sigma}^k,\tilde{\sigma}}$ on the eigenspace V^{ω^j}.

Let $p : \Sigma_g \to S^2$ be the regular branched m-cyclic covering on S^2 with m branched points as on page 333. Choose a point in $p^{-1}(*)$, and denote it by $\tilde{*} \in \Sigma_g$. We denote by $\tilde{\alpha}_i$ the lift of α_i which starts at $\tilde{*}$. Note that $\tilde{\alpha}_i t(\tilde{\alpha}_{i+1})^{-1}$ is a loop in Σ_g while $\tilde{\alpha}_i$ is an arc. We denote by $e_i(k) \in H_1(\Sigma_g;\mathbb{Z})$ the homology class represented by $t^k(\tilde{\alpha}_i t(\tilde{\alpha}_{i+1})^{-1})$.

Lemma 6.1. *The homology classes $\{e_i(k)\}_{\substack{1\leq i\leq m-2\\ 0\leq k\leq m-2}}$ form a basis of $H_1(\Sigma_g;\mathbb{Z})$.*

Proof. We use the Schreier method. Let T denote a Schreier transversal $T=\{\alpha_1^k\}_{i=0}^{m-1}$ and S a generating set $S=\{\alpha_i\}_{i=1}^{m-1}$ of $\pi_1(S^2-\{q_i\}_{i=1}^m)$. Then the subgroup $\pi_1(\Sigma_g - \{p^{-1}(q_i)\}_{i=1}^m)$ is generated by

$$\{(rs(\overline{rs})^{-1} \mid r\in T, s\in S\} = \{\alpha_1^k\alpha_i\alpha_1^{-k-1}\}_{\substack{2\leq i\leq m-1\\ 0\leq k\leq m-2}} \cup \{\alpha_1^{m-1}\alpha_i\}_{1\leq i\leq m-1}.$$

By van Kampen's theorem, the group $\pi_1(\Sigma_g)$ is obtained by adding the relation $\alpha_i^m=1$ to $\pi_1(\Sigma_g-\{p^{-1}(q_i)\}_{i=1}^m)$. Thus, the set $\{\alpha_1^k\alpha_i\alpha_{i+1}^{-1}\alpha_1^{-k}\}_{i,k}$, where from now through the end of the proof we have $1\leq i\leq m-2$ and $0\leq k\leq m-2$, generates the group $\pi_1(\Sigma_g)$. This implies that $\{e_i(k)\}_{i,k}$ is a generating set of $H_1(\Sigma_g;\mathbb{Z})$.

By the Riemann–Hurwitz formula, $H_1(\Sigma_g;\mathbb{Z})$ is a free module of rank equal to $2g=(m-1)(m-2)$, and this is equal to the order of the set $\{e_i(k)\}_{i,k}$. Therefore the set $\{e_i(k)\}_{i,k}$ is a basis of the free module $H_1(\Sigma_g;\mathbb{Z})$. □

The intersection form and the action of $\tilde{\sigma}_i$. Let j be an integer with $1\leq j\leq m-1$. Firstly we find a basis of $V^{\omega^j}\subset H_1(\Sigma_g;\mathbb{C})$ and calculate intersection numbers.

Lemma 6.2. *The intersection numbers of $\{e_i(k)\}_{\substack{1\leq i\leq m-2\\ 0\leq k\leq m-2}}$ are*

$$e_i(k)\cdot e_{i'}(k) = \begin{cases} -1 & \textit{if } i=i'-1,\\ 1 & \textit{if } i=i'+1,\\ 0 & \textit{otherwise},\end{cases} \qquad e_i(k)\cdot e_{i'}(k+1) = \begin{cases} -1 & \textit{if } i=i',\\ 1 & \textit{if } i=i'-1,\\ 0 & \textit{otherwise},\end{cases}$$

$$e_i(k)\cdot e_{i'}(k-1) = \begin{cases} -1 & \textit{if } i=i',\\ 1 & \textit{if } i=i'+1,\\ 0 & \textit{otherwise},\end{cases} \qquad e_i(k)\cdot e_{i'}(k')=0 \quad \textit{if } |k-k'|\geq 2.$$

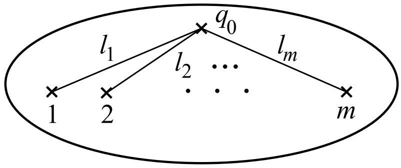


Figure 8. The paths $l_1, l_2, \dots, l_m$.

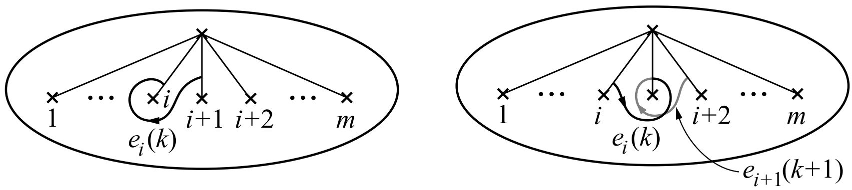


Figure 9. Left: the k-th copy. Right: the $(k+1)$-th copy.

Proof. We only prove the equality $e_i(k) \cdot e_{i+1}(k+1) = 1$ since the other cases are proved in the same way.

Let l_i be the paths as in Figure 8. Consider m copies of the 2-sphere cut along the l_i, and number them from 1 to m. (For convenience, copy 1 will also be called copy $m+1$.) Gluing the left-hand side of l_i in the k-th copy to the right-hand side of l_i in the $(k+1)$-th copy for $k = 1, 2, \dots, m$, we obtain a closed connected surface homeomorphic to Σ_g, and it is naturally a covering space on S^2. As in Figure 9, the loops representing $e_i(k)$ and $e_{i+1}(k+1)$ intersect once positively in the $(k+1)$-th copy. Hence we have $e_i(k) \cdot e_{i+1}(k+1) = 1$. □

For $1 \le i \le m-2$, we define $w_i = \sum_{k=0}^{m-1} \omega^{-jk} e_i(k)$. Since $te_i(k) = e_i(k+1)$ for $1 \le k \le m-2$ and $e_i(m-1) = -\sum_{k=0}^{m-2} e_i(k)$, we have $w_i \in V^{\omega^j}$, and the set $\{w_i\}_{i=1}^{m-2}$ is a basis of V^{ω^j}.

Lemma 6.3. *The intersection numbers of* $\{w_i\}_{1 \le i \le m-2}$ *are*

$$w_i \cdot w_{i'} = \begin{cases} d(1-\omega^j) & \text{if } i = i'+1, \\ d(-\omega^{-j}+\omega^j) & \text{if } i = i', \\ d(\omega^{-j}-1) & \text{if } i = i'-1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. By Lemma 6.2, we have

$$w_i \cdot w_i = \sum_{k=0}^{d-1} \sum_{l=0}^{d-1} \omega^{j(k-l)} e_i(k) \cdot e_i(l) = d(-\omega^{-j}+\omega^j),$$

$$w_i \cdot w_{i+1} = \sum_{k=0}^{d-1} \sum_{l=0}^{d-1} \omega^{j(k-l)} e_i(k) \cdot e_{i+1}(l) = d(\omega^{-j}-1),$$

and $w_i \cdot w_k = 0$ when $|i-k| \geq 2$. □

Let $\tilde{\sigma} = \tilde{\sigma}_1 \cdots \tilde{\sigma}_{r-1}$. We next find eigenvectors in V^{ω^j} relative to the action by $\tilde{\sigma}$.

Lemma 6.4. *Let i be an integer such that $1 \leq i \leq m-1$. Then we have*

$$(\tilde{\sigma}_i)_* e_l(k) = \begin{cases} e_l(k) + e_{l+1}(k) & \text{if } 2 \leq i \leq m-1, \text{ and } l = i-1, \\ -e_l(k-1) & \text{if } l = i, \\ e_{l-1}(k-1) + e_l(k) & \text{if } l = i+1, \\ e_l(k) & \text{if } l \neq i-1, i, i+1. \end{cases}$$

Proof. Recall that $e_i(k)$ is the homology class represented by the loop $\tilde{\alpha}_1^k \tilde{\alpha}_i \tilde{\alpha}_{i+1}^{-1} \tilde{\alpha}_1^{-k}$.

In the fundamental group $\pi_1(S^2 - \{q_i\}_{i=1}^m)$, we have

$$\begin{aligned}
(\sigma_i)_*(\alpha_{i-1}\alpha_i^{-1}) &= \alpha_{i-1}\alpha_{i+1}^{-1} = (\alpha_{i-1}\alpha_i^{-1})(\alpha_i\alpha_{i+1}^{-1}), \\
(\sigma_i)_*(\alpha_i\alpha_{i+1}^{-1}) &= \alpha_{i+1}(\alpha_{i+1}^{-1}\alpha_i\alpha_{i+1})^{-1} = \alpha_{i+1}^{-1}(\alpha_i\alpha_{i+1}^{-1})^{-1}\alpha_{i+1}, \\
(\sigma_i)_*(\alpha_{i+1}\alpha_{i+2}^{-1}) &= (\alpha_{i+1}^{-1}\alpha_i\alpha_{i+1})\alpha_{i+2}^{-1} = \alpha_{i+1}^{-1}(\alpha_i\alpha_{i+1}^{-1})\alpha_{i+1}(\alpha_{i+1}\alpha_{i+2}^{-1}).
\end{aligned}$$

By lifting these loops to the covering space Σ_g, we obtain what we want. □

By Lemma 6.4, the matrix representations of the actions of $\{\tilde{\sigma}_i\}_{i=1}^{m-1}$ on V^{ω^j} with respect to the basis $\{w_i\}_{1 \leq i \leq m-2}$ are calculated as

$$(\tilde{\sigma}_1)_* = \begin{pmatrix} -\omega^{-j} & \omega^{-j} & O \\ 0 & 1 & O \\ O & O & I_{m-4} \end{pmatrix}, \qquad (\tilde{\sigma}_i)_* = \begin{pmatrix} I_{i-1} & O & O \\ O & L & O \\ O & O & I_{m-i-4} \end{pmatrix},$$

$$(\tilde{\sigma}_{m-2})_* = \begin{pmatrix} I_{m-4} & O & O \\ O & 1 & O \\ O & 1 & -\omega^{-j} \end{pmatrix}, \quad (\tilde{\sigma}_{m-1})_* = \begin{pmatrix} I_{m-3} & v \\ O & -1 + \sum_{k=1}^{m-2} \omega^{-jk} \end{pmatrix},$$

where

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -\omega^{-j} & \omega^{-j} \\ 0 & 0 & 1 \end{pmatrix}, \qquad v = \left(1, 1 + \omega^{-j}, \ldots, \sum_{k=0}^{m-3} \omega^{-jk}\right)^T.$$

Let r be an integer with $2 \leq r \leq m$, and put

$$e'_r(k) = [\tilde{a}_1^k \tilde{a}_r (\tilde{a}_1 \tilde{a}_2 \cdots \tilde{a}_r)^{-1} \tilde{a}_1^{-1} (\tilde{a}_1 \tilde{a}_2 \cdots \tilde{a}_r) \tilde{a}_1^{-k}].$$

By Lemma 6.4, we have

$$\begin{aligned}
\tilde{\sigma}_* e_i(k) &= e_{i+1}(k), \qquad \text{when } 1 \leq i \leq r-2, \\
\tilde{\sigma}_* e_r(k) &= -e'_r(k) + e_r(k), \\
\tilde{\sigma}_* e_{r-1}(k) &= e'_r(k), \\
\tilde{\sigma}_* e'_r(k) &= e_1(k - r + 1).
\end{aligned}$$

The sum $w'_r := \sum_{k=0}^{m-1} \omega^{-jk} e'_r(k)$ is contained in V^{ω^j}. For $i = 1, 2, \dots, r-2$, we have

$$\tilde{\sigma}_* w_i = \tilde{\sigma}_* \sum_{k=0}^{m-1} \omega^{-jk} e_i(k) = \sum_{k=0}^{m-1} \omega^{-jk} e_{i+1}(k) = w_{i+1},$$

$$\tilde{\sigma}_* w_{r-1} = \sum_{k=0}^{m-1} \omega^{-jk} (e_{r-1}(k) + e'_r(k)) = w_{r-1} + w'_r,$$

$$\tilde{\sigma}_* w'_r = \sum_{k=0}^{m-1} \omega^{-jk} e_1(k-r+1) = \sum_{k=0}^{m-1} \omega^{-j(k+r-1)} e_1(k) = \omega^{-(r-1)j} w_1.$$

Let $\zeta = \exp(2\pi \boldsymbol{i}/r)$ and $v_i = \sum_{k=1}^{r-1} \omega^{(k-1)j} \zeta^{-(k-1)i} w_k + \omega^{(r-1)j} \zeta^{-(r-1)i} w'_r$. Then

$$\begin{aligned}
\tilde{\sigma}_* v_i &= \sum_{k=1}^{r-1} \omega^{(k-1)j} \zeta^{-(k-1)i} (\tilde{\sigma})_* w_k + \omega^{(r-1)j} \zeta^{-(r-1)i} (\tilde{\sigma})_* w'_r \\
&= \sum_{k=1}^{r-2} \omega^{(k-1)j} \zeta^{-(k-1)i} w_{k+1} + \omega^{(r-2)j} \zeta^{-(r-2)i} w'_r + \omega^{(r-1)j} \zeta^{-(r-1)i} \omega^{-pj} w_1 \\
&= \omega^{-j} \zeta^i \left(\sum_{k=1}^{r-1} \omega^{(k-1)j} \zeta^{-(k-1)i} w_k + \omega^{(r-1)j} \zeta^{-(r-1)i} w'_r \right) \\
&= (\omega^{-j} \zeta^i) v_i.
\end{aligned}$$

Hence v_i is an eigenvector with eigenvalue $\omega^{-j}\zeta^i$ with respect to the action by $\tilde{\sigma}$. Note that the subspace generated by $\{w_i\}_{i=1}^{r-1}$ coincides with one generated by $\{v_i\}_{i=1}^{r-1}$. Since $\tilde{\sigma}$ acts trivially on $\{w_i\}_{i=r+1}^{m-1}$, they are also eigenvectors with eigenvalue 0. Moreover the set $\{v_i\}_{i=1}^{r-1} \cup \{w_i\}_{i=r+1}^{m-2}$ is linearly independent.

Lemma 6.5. *Let i, i' be integers such that $1 \le i \le r-1$ and $1 \le i' \le r-1$. Then we have*

$$v_i \cdot v_{i'} = \begin{cases} 8rd\boldsymbol{i} \sin \dfrac{\pi i}{r} \sin \dfrac{\pi j}{m} \sin \pi \left(\dfrac{i}{r} - \dfrac{j}{m} \right) & \textit{if } i = i', \\ 0 & \textit{otherwise}. \end{cases}$$

Proof. Since the action of the mapping class group $\pi_0 C_g(t)$ preserves the intersection form,

$$\begin{aligned}
v_i \cdot v_i &= \sum_{k=0}^{r-1} \sum_{l=0}^{r-1} \omega^{(l-k)j} \zeta^{-(l-k)i} (\tilde{\sigma}_*^k w_1 \cdot \tilde{\sigma}_*^l w_1) \\
&= \sum_{k=0}^{r-1} \sum_{l=0}^{r-1} \omega^{(l-k)j} \zeta^{-(l-k)i} (w_2 \cdot \tilde{\sigma}_*^{l-k+1} w_1).
\end{aligned}$$

Thus Lemma 6.3 implies

$$\begin{aligned} v_i \cdot v_i &= \omega^{(r-1)j}\zeta^{-(r-1)i}(w_2 \cdot \tilde{\sigma}_*^r w_1) + \omega^j \zeta^{-i}(r-1)(w_2 \cdot \tilde{\sigma}_*^2 w_1) + r(w_2 \cdot \tilde{\sigma}_* w_1) \\ &\qquad + \omega^{-j}\zeta^i(r-1)(w_2 \cdot w_1) + \omega^{-(r-1)j}\zeta^{(r-1)i}(w_2 \cdot \tilde{\sigma}_*^{-r+2} w_1) \\ &= r\{(\omega^{-j}\zeta^i) w_2 \cdot w_1 + (\omega^j \zeta^{-i}) w_2 \cdot w_3 + w_2 \cdot w_2\} \\ &= 8rd\boldsymbol{i} \sin \frac{\pi i}{r} \sin \frac{\pi j}{m} \sin \pi \Big(\frac{i}{r} - \frac{j}{m}\Big). \qquad \square \end{aligned}$$

Calculation of ω-signatures and the cobounding functions $\phi_{m,j}$. Lastly we will calculate the hermitian form $\langle\ ,\ \rangle_{\tilde{\sigma}^k,\tilde{\sigma}}$ and the ω-signature. We have already found the set of eigenvectors $\{v_i\}_{i=1}^{r-1} \cup \{w_i\}_{i=r+1}^{m-2}$ with respect to the action by $\tilde{\sigma}$ which is linearly independent. Since $\dim V^{\omega^j} = m-2$, we need to find another eigenvector.

Lemma 6.6. $$\sum_{k=1}^{rm} \tau(\tilde{\sigma}^k, \tilde{\sigma}) = rm - 2|mi - rj|.$$

Proof. We first consider the case when rj/m is not an integer. Put

$$\beta = \sum_{i=1}^r w_i - \frac{1}{r}\sum_{k=1}^r \frac{1}{1-\omega^j\zeta^{-k}} v_k.$$

The subspace generated by $\{v_i\}_{i=1}^{r-1}$ and that generated by $\{w_i\}_{i=1}^{r-1}$ coincide. Thus the set $\{v_i\}_{i=1}^{r-1}$, β, $\{w_i\}_{i=r+1}^{m-2}$ forms a basis of V^{ω^j} when $1 \le r \le m-2$, and the set $\{v_i\}_{i=1}^{m-2}$ forms a basis of V^{ω^j} when $r = m-1$. We have

$$\begin{aligned} \tilde{\sigma}_*\beta &= \sum_{i=2}^r w_i - \frac{1}{r}\sum_{k=1}^r \frac{\omega^j\zeta^{-k}}{1-\omega^j\zeta^{-k}} v_k \\ &= \sum_{i=2}^r w_i + \frac{1}{r}\sum_{k=1}^r v_k - \frac{1}{r}\sum_{k=1}^r \frac{1}{1-\omega^j\zeta^{-k}} v_k \\ &= \sum_{i=1}^r w_i - \frac{1}{r}\sum_{k=1}^r \frac{1}{1-\omega^j\zeta^{-k}} v_k = \beta. \end{aligned}$$

Note that β and $\{w_i\}_{i=r+1}^{m-2}$ are in the annihilator of the hermitian form $\langle\ ,\ \rangle_{\tilde{\sigma}^k,\tilde{\sigma}}$ since they have eigenvalue 1 with respect to the action by $\tilde{\sigma}$.

By Lemma 4.2, we have

$$\begin{aligned} \tau(\tilde{\sigma}^k, \tilde{\sigma}) &= \sum_{i=1}^r \operatorname{sign}\langle v_i, v_i\rangle_{\tilde{\sigma}^k,\tilde{\sigma}} = -\sum_{i=1}^r \operatorname{sign}\Big((v_i\cdot v_i)\sum_{l=1}^k ((\omega^{-j}\zeta^i)^l - (\omega^j\zeta^{-i})^l)\Big) \\ &= -\sum_{i=1}^r \operatorname{sign}\Big((v_i\cdot v_i)(1-\omega^j\zeta^{-i})(1-\omega^{-j}\zeta^i)\sum_{l=1}^k ((\omega^{-j}\zeta^i)^l - (\omega^j\zeta^{-i})^l)\Big). \end{aligned}$$

By the equation

$$(1-\omega^j\zeta^{-i})(1-\omega^{-j}\zeta^i)\sum_{l=1}^{k}((\omega^{-j}\zeta^i)^l-(\omega^j\zeta^{-i})^l)$$
$$=8\boldsymbol{i}\sin\Bigl(-\frac{\pi(k+1)j}{m}+\frac{\pi(k+1)i}{r}\Bigr)\sin\Bigl(-\frac{\pi kj}{m}+\frac{\pi ki}{r}\Bigr)\sin\Bigl(-\frac{\pi j}{m}+\frac{\pi i}{r}\Bigr)$$

and Lemma 6.5, we have

$$\tau((\tilde{\sigma})^k,\tilde{\sigma})=\sum_{i=1}^{r-1}\operatorname{sign}\Bigl(\sin k\pi\Bigl(\frac{i}{r}-\frac{j}{m}\Bigr)\sin(k+1)\pi\Bigl(\frac{i}{r}-\frac{j}{m}\Bigr)\Bigr).$$

Since rj/m is not an integer, $i/r-j/m$ is not zero. Thus we obtain

$$\sum_{k=1}^{rm}\tau((\tilde{\sigma})^k,\tilde{\sigma})=\sum_{i=1}^{r-1}\sum_{k=1}^{rm}\operatorname{sign}\Bigl(\sin k\pi\Bigl(\frac{i}{r}-\frac{j}{m}\Bigr)\sin(k+1)\pi\Bigl(\frac{i}{r}-\frac{j}{m}\Bigr)\Bigr)$$
$$=\sum_{i=1}^{r-1}(rm-2|mi-rj|).$$

Next consider the case when rj/m is an integer and $1\le r\le m-1$. Denote this integer rj/m by i_0. Then, the eigenvalue of v_{i_0} is 1, and v_{i_0} and $\{w_i\}_{i=r+1}^{m-2}$ are in the annihilator of $\langle\ ,\ \rangle_{\tilde{\sigma}^k,\tilde{\sigma}}$. If we put

$$\beta'=\sum_{i=1}^{r}w_i-\frac{1}{r}\sum_{\substack{1\le k\le r\\ k\ne i_0}}\frac{1}{1-\omega^j\zeta^{-k}}v_k,$$

the set of the homology classes $\{v_i\}_{i=1}^{r-1}$, β', $\{w_i\}_{i=r+1}^{m-2}$ forms a basis of V^{ω^j}. We have

$$\tilde{\sigma}\beta'=\sum_{i=2}^{r}w_i-\frac{1}{r}\sum_{\substack{1\le k\le r\\ k\ne i_0}}\frac{\omega^j\zeta^{-k}}{1-\omega^j\zeta^{-k}}v_k$$
$$=\sum_{i=2}^{r}w_i+\frac{1}{r}\sum_{\substack{1\le k\le r\\ k\ne i_0}}v_k-\frac{1}{r}\sum_{\substack{1\le k\le r\\ k\ne i_0}}\frac{1}{1-\omega^j\zeta^{-k}}v_k$$
$$=\sum_{i=1}^{r}w_i-\frac{1}{r}v_{i_0}-\frac{1}{r}\sum_{\substack{1\le k\le r\\ k\ne i_0}}\frac{1}{1-\omega^j\zeta^{-k}}v_k=\beta'-\frac{1}{r}v_{i_0}.$$

By Lemma 4.2,

$$\langle\beta',\beta'\rangle_{\tilde{\sigma}^k,\tilde{\sigma}}=\beta'\cdot\frac{1}{r}\sum_{i=1}^{k}2i\,v_{i_0}=\frac{k(k+1)}{r}\sum_{i=1}^{r}w_i\cdot v_{i_0}.$$

Since the eigenvalues of $\{v_i\}_{i=1}^{r-1}$ are different from 1, the intersection $v_i \cdot v_{i_0} = 0$ for $1 \le i \le r-1$. Since the subspace generated by $\{w_i\}_{i=1}^{r-1}$ and that generated by $\{v_i\}_{i=1}^{r-1}$ coincide, we also have $w_i \cdot v_{i_0} = 0$. Thus we have

$$\begin{aligned}\frac{r}{k(k+1)}\langle \beta', \beta'\rangle_{\tilde{\sigma}^k,\tilde{\sigma}} &= w_r \cdot v_{i_0} = w_r \cdot (w_{r-1} + w_r') \\ &= w_r \cdot \left(w_{r-1} - \sum_{k=0}^{r-1} \omega^{(k-r)j} w_k \right) \\ &= (1-\omega^{-j}) w_r \cdot w_{r-1} = (1-\omega^{-j})(1-\omega^{j}) > 0.\end{aligned}$$

Moreover since $v_i \cdot v_{i_0} = 0$, Lemma 4.2 implies $\langle v_i, \beta'\rangle_{\tilde{\sigma}^k,\tilde{\sigma}} = 0$ for $1 \le i \le r-1$. Therefore we have

$$\begin{aligned}\sum_{k=1}^{rm} \tau(\tilde{\sigma}^k, \tilde{\sigma}) &= \sum_{k=1}^{rm}\left(\sum_{i=1}^{k} \operatorname{sign}(\langle v_i, v_i\rangle_{\tilde{\sigma}^k,\tilde{\sigma}}) + \operatorname{sign}(\langle \beta', \beta'\rangle_{\tilde{\sigma}^k,\tilde{\sigma}}) \right) \\ &= \sum_{k=1}^{rm}\left(\sum_{\substack{1\le i\le r-1 \\ i\ne i_0}} \operatorname{sign}\left(\sin k\pi\left(\frac{i}{r}-\frac{j}{m}\right) \sin(k+1)\pi\left(\frac{i}{r}-\frac{j}{m}\right)\right) + 1\right) \\ &= \sum_{\substack{1\le i\le r-1 \\ i\ne i_0}} (rm - 2|mi - rj|) + rm = \sum_{i=1}^{r-1} (rm - 2|mi-rj|).\end{aligned}$$

In the case when $r = m$, the set $\{v_i\}_{i=1}^{r-2}$ forms a basis of V^{ω^j}. By a similar calculation, we can also prove what we want. $\square$

Lemma 6.7. *For $r = 2, 3, \ldots, m$,*

$$\phi_{m,j}(\tilde{\sigma}) - \bar{\phi}_{m,j}(\tilde{\sigma}) = \frac{2}{r}\left\{ \left(\frac{rj}{m} - \left[\frac{rj}{m}\right] - \frac{1}{2}\right)^2 - \frac{r^2 j(m-j)}{m^2} - \frac{1}{4}\right\}.$$

Proof.
$$\tau(\tilde{\sigma}^k, \tilde{\sigma}) = \sum_{i=1}^{r-1} \operatorname{sign}\left(\sin k\pi\left(\frac{i}{r}-\frac{j}{m}\right) \sin(k+1)\pi\left(\frac{i}{r}-\frac{j}{m}\right)\right).$$

Since we have $\tau(\tilde{\sigma}^{k+rm}, \tilde{\sigma}) = \tau(\tilde{\sigma}^k, \tilde{\sigma})$,

$$\begin{aligned}\phi_{m,j}(\tilde{\sigma}) - \bar{\phi}_{m,j}(\tilde{\sigma}) &= \frac{1}{rm}\sum_{k=1}^{rm} \tau(\tilde{\sigma}^k, \tilde{\sigma}) = \frac{1}{rm}\sum_{i=1}^{r-1}(rm - 2|mi - rj|) \\ &= r - 1 - \frac{2}{rm}\left(\sum_{i=1}^{[\frac{rj}{m}]} (rj - mi) + \sum_{[\frac{rj}{m}]+1}^{r-1} (mi - rj)\right) \\ &= \frac{2}{r}\left\{ \left(\frac{rj}{m} - \left[\frac{rj}{m}\right] - \frac{1}{2}\right)^2 + \frac{r^2 j(m-j)}{m^2} - \frac{1}{4}\right\}.\end{aligned}$$ $\square$

Proof of Theorem 1.1. Applying Lemma 6.7 to the case when $r = m$, we have

$$\phi_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}) - \bar{\phi}_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}) = \frac{2j(m-j)}{m}.$$

Since

$$\bar{\phi}_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}) = \frac{1}{m}\bar{\phi}_{m,j}((\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1})^m) = 0,$$

we have

$$\phi_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}) = \frac{2j(m-j)}{m}.$$

Put $\varphi = \tilde{\sigma}_1\tilde{\sigma}_3 \cdots \tilde{\sigma}_{m-1}$, $\psi = \tilde{\sigma}_2\tilde{\sigma}_4 \cdots \tilde{\sigma}_{m-2}$ when m is even, and $\varphi = \tilde{\sigma}_1\tilde{\sigma}_3 \cdots \tilde{\sigma}_{m-2}$, $\psi = \tilde{\sigma}_2\tilde{\sigma}_4 \cdots \tilde{\sigma}_{m-1}$, when m is odd. As we saw in Section 5, $\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}$ is conjugate to $\varphi\psi$. By direct computation, if $(\varphi_*^{-1} - I_{2g})x + (\psi_* - I_{2g})y = 0$ for $x, y \in V^{\omega^j}$, we have $(\varphi_*^{-1} - I_{2g})x = (\psi_* - I_{2g})y = 0$. Hence we have $\tau_g(\varphi, \psi) = 0$.

In the same way, for $i = 1, 2, \ldots, [(m-1)/2]$, we have

$$\tau_g(\tilde{\sigma}_1\tilde{\sigma}_3 \cdots \tilde{\sigma}_{2i+1}, \tilde{\sigma}_2\tilde{\sigma}_4 \cdots \tilde{\sigma}_{2i}) = \tau_g(\tilde{\sigma}_1\tilde{\sigma}_3 \cdots \tilde{\sigma}_{2i+1}, \tilde{\sigma}_2\tilde{\sigma}_4 \cdots \tilde{\sigma}_{2i+2}) = 0,$$
$$\tau_g(\tilde{\sigma}_1\tilde{\sigma}_3 \cdots \tilde{\sigma}_{2i-1}, \tilde{\sigma}_{2i+1}) = \tau_g(\tilde{\sigma}_2\tilde{\sigma}_4 \cdots \tilde{\sigma}_{2i}, \tilde{\sigma}_{2i+2}) = 0.$$

Thus

$$\phi_{m,j}(\tilde{\sigma}) = (r-1)\phi_{m,j}(\tilde{\sigma}_1) = \frac{r-1}{m-1}\phi_{m,j}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{m-1}) = \frac{2(r-1)j(m-j)}{m(m-1)}.$$

Hence we obtain

$$\begin{aligned}
\bar{\phi}_{m,j}(\sigma_1 \cdots \sigma_{r-1}) &= \bar{\phi}_{m,j}(\tilde{\sigma}) \\
&= \phi_{m,j}(\tilde{\sigma}) - (\phi_{m,j}(\tilde{\sigma}) - \bar{\phi}_{m,j}(\tilde{\sigma})) \\
&= -\frac{2}{r}\left\{\frac{jr(m-j)(m-r)}{m^2(m-1)} + \left(\frac{rj}{m} - \left[\frac{rj}{m}\right] - \frac{1}{2}\right)^2 - \frac{1}{4}\right\}. \qquad \square
\end{aligned}$$

By the values of $\bar{\phi}_{m,1}$, we see:

Remark 6.8. Let r be an integer such that $2 \le r \le m$. Then

$$\bar{\phi}_{m,1}(\tilde{\sigma}_1 \cdots \tilde{\sigma}_{r-1}) = 0.$$

However we do not know whether the quasimorphism $\bar{\phi}_{m,1}$ is trivial or not.

References

[Barge and Ghys 1992] J. Barge and É. Ghys, "Cocycles d'Euler et de Maslov", *Math. Ann.* **294**:2 (1992), 235–265. MR 95b:55021 Zbl 0894.55006

[Bavard 1991] C. Bavard, "Longueur stable des commutateurs", *Enseign. Math.* (2) **37**:1-2 (1991), 109–150. MR 92g:20051 Zbl 0810.20026

[Baykur et al. 2013] R. İ. Baykur, M. Korkmaz, and N. Monden, "Sections of surface bundles and Lefschetz fibrations", *Trans. Amer. Math. Soc.* **365**:11 (2013), 5999–6016. MR 3091273 Zbl 06208247

[Bestvina and Fujiwara 2002] M. Bestvina and K. Fujiwara, "Bounded cohomology of subgroups of mapping class groups", *Geom. Topol.* **6** (2002), 69–89. MR 2003f:57003 Zbl 1021.57001

[Birman and Hilden 1971] J. S. Birman and H. M. Hilden, "On the mapping class groups of closed surfaces as covering spaces", pp. 81–115 in *Advances in the theory of Riemann surfaces* (Stony Brook, NY, 1969), edited by L. V. Ahlfors et al., Ann. of Math. Studies **66**, Princeton University Press, 1971. MR 45 #1169 Zbl 0217.48602

[Birman and Hilden 1973] J. S. Birman and H. M. Hilden, "On isotopies of homeomorphisms of Riemann surfaces", *Ann. of Math.* (2) **97** (1973), 424–439. MR 48 #4305 Zbl 0237.57001

[Calegari 2009] D. Calegari, *scl*, MSJ Memoirs **20**, Math. Soc. Japan, Tokyo, 2009. MR 2011b:57003 Zbl 1187.20035

[Cochran et al. 2012] T. D. Cochran, S. Harvey, and P. D. Horn, "Higher-order signature cocycles for subgroups of mapping class groups and homology cylinders", *Int. Math. Res. Not.* **2012**:14 (2012), 3311–3373. MR 2946227 Zbl 1257.57014

[Cohen 1987] F. R. Cohen, "Homology of mapping class groups for surfaces of low genus", pp. 21–30 in *The Lefschetz Centennial Conference* (Mexico City, 1984), edited by S. Gitler, Contemp. Math. **58**:2, Amer. Math. Soc., Providence, RI, 1987. MR 88j:57046 Zbl 0655.57006

[Endo 2000] H. Endo, "Meyer's signature cocycle and hyperelliptic fibrations", *Math. Ann.* **316**:2 (2000), 237–257. MR 2001b:57047 Zbl 0948.57013

[Endo and Kotschick 2001] H. Endo and D. Kotschick, "Bounded cohomology and non-uniform perfection of mapping class groups", *Invent. Math.* **144**:1 (2001), 169–175. MR 2001m:57046 Zbl 0987.57004

[Endo and Nagami 2005] H. Endo and S. Nagami, "Signature of relations in mapping class groups and non-holomorphic Lefschetz fibrations", *Trans. Amer. Math. Soc.* **357**:8 (2005), 3179–3199. MR 2006g:57051 Zbl 1081.57023

[Farb and Margalit 2012] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series **49**, Princeton University Press, 2012. MR 2012h:57032 Zbl 1245.57002

[Gambaudo and Ghys 2005] J.-M. Gambaudo and É. Ghys, "Braids and signatures", *Bull. Soc. Math. France* **133**:4 (2005), 541–579. MR 2007d:57014 Zbl 1103.57001

[Korkmaz 2004] M. Korkmaz, "Stable commutator length of a Dehn twist", *Michigan Math. J.* **52**:1 (2004), 23–31. MR 2005b:57003 Zbl 1061.57022

[Kotschick 2008] D. Kotschick, "Stable length in stable groups", pp. 401–413 in *Groups of diffeomorphisms*, edited by R. Penner et al., Adv. Stud. Pure Math. **52**, Math. Soc. Japan, Tokyo, 2008. MR 2011j:57003 Zbl 1188.20028

[Louwsma 2011] J. R. Louwsma, *Extremality of the rotation quasimorphism on the modular group*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2011, Available at http://search.proquest.com/docview/1022051368. MR 3047198

[Meyer 1972] W. Meyer, "Die Signatur von lokalen Koeffizientensystemen und Faserbündeln", *Bonn. Math. Schr.* **53** (1972). MR 46 #4532 Zbl 0243.58004

[Monden 2012] N. Monden, "On upper bounds on stable commutator lengths in mapping class groups", *Topology Appl.* **159**:4 (2012), 1085–1091. MR 2876715 Zbl 1239.57035

[Morifuji 2003] T. Morifuji, "On Meyer's function of hyperelliptic mapping class groups", *J. Math. Soc. Japan* **55**:1 (2003), 117–129. MR 2003m:57054 Zbl 1031.57017

[Powell 1978] J. Powell, "Two theorems on the mapping class group of a surface", *Proc. Amer. Math. Soc.* **68**:3 (1978), 347–350. MR 58 #13045 Zbl 0391.57009

[Turaev 1985] V. G. Turaev, "Первый симплектический класс Черна и индексы Маслова", *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **143** (1985), 110–129. Translated as "First symplectic Chern class and Maslov indices" in *J. Sov. Math.* **37**:3 (1987), 1115–1127. MR 86m:58059 Zbl 0577.55012

Received August 23, 2013. Revised February 28, 2014.

DANNY CALEGARI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CHICAGO
5734 SOUTH UNIVERSITY AVENUE
CHICAGO, IL 60637
UNITED STATES
dannyc@math.uchicago.edu

NAOYUKI MONDEN
DEPARTMENT OF ENGINEERING SCIENCE
OSAKA ELECTRO-COMMUNICATION UNIVERSITY
HATSU-CHO 18-8
NEYAGAWA 572-8530
JAPAN
n-monden@math.kyoto-u.ac.jp

MASATOSHI SATO
DEPARTMENT OF MATHEMATICS EDUCATION
GIFU UNIVERSITY
YANAGITO 1-1
GIFU 501-1193
JAPAN
msato@gifu-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.353

COMPOSITIONS OF BIRATIONAL ENDOMORPHISMS OF THE AFFINE PLANE

PIERRETTE CASSOU-NOGUÈS AND DANIEL DAIGLE

Besides contributing several new results in the general theory of birational endomorphisms of $\mathbb{A}^2$, this article describes certain classes of birational endomorphisms $f : \mathbb{A}^2 \to \mathbb{A}^2$ defined by requiring that the missing curves or contracting curves of f are lines. The last part of the article is concerned with the monoid structure of the set of birational endomorphisms of $\mathbb{A}^2$.

Let $\boldsymbol{k}$ be an algebraically closed field and let $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$ be the affine plane over $\boldsymbol{k}$. A *birational endomorphism of* $\mathbb{A}^2$ is a morphism of algebraic varieties, $f : \mathbb{A}^2 \to \mathbb{A}^2$, which restricts to an isomorphism $U \to V$ for some nonempty Zariski-open subsets U and V of $\mathbb{A}^2$. The set $\text{Bir}(\mathbb{A}^2)$ of birational endomorphisms of $\mathbb{A}^2$ is a monoid under composition of morphisms, and the group of invertible elements of this monoid is the automorphism group $\text{Aut}(\mathbb{A}^2)$. An element f of $\text{Bir}(\mathbb{A}^2)$ is *irreducible* if it is not invertible and if, for every factorization $f = h \circ g$ with $g, h \in \text{Bir}(\mathbb{A}^2)$, one of g, h is invertible. Elements $f, g \in \text{Bir}(\mathbb{A}^2)$ are *equivalent* (denoted $f \sim g$) if there exist $u, v \in \text{Aut}(\mathbb{A}^2)$ satisfying $u \circ f \circ v = g$. The elements of $\text{Bir}(\mathbb{A}^2)$ which are equivalent to the birational morphism $c : \mathbb{A}^2 \to \mathbb{A}^2$, $c(x, y) = (x, xy)$ are called *simple affine contractions* (SAC) and are the simplest examples of noninvertible elements of $\text{Bir}(\mathbb{A}^2)$. It was once an open question whether $\text{Aut}(\mathbb{A}^2) \cup \{c\}$ generated $\text{Bir}(\mathbb{A}^2)$ as a monoid (the question arose in Abhyankar's seminar at Purdue in the early 1970s); P. Russell showed that the answer was negative by giving an example of an irreducible element of $\text{Bir}(\mathbb{A}^2)$ which is not a SAC (this example appeared later in [Daigle 1991a, 4.7]). This was the first indication that $\text{Bir}(\mathbb{A}^2)$ could be a complicated object.

The papers [Daigle 1991a; 1991b] seem to be the first publications that study birational endomorphisms of $\mathbb{A}^2$ in a systematic way (these are our main references). We know of two more contributions to the subject: a certain family of elements of $\text{Bir}(\mathbb{A}^2)$ is described explicitly in [Cassou-Noguès and Russell 2007], and [Shpilrain and Yu 2004] gives an algorithm for deciding whether a given element of $\text{Bir}(\mathbb{A}^2)$ is in the submonoid generated by $\text{Aut}(\mathbb{A}^2) \cup \{c\}$.

The research of Cassou-Noguès was partially supported by Spanish grants MTM2010-21740-C02-01 and MTM2010-21740-C02-02. The research of Daigle was supported by grant RGPIN/104976-2010 from NSERC Canada.

MSC2010: 14R10, 14H50.

Keywords: affine plane, birational morphism, plane curve.

The list of references is much longer if we include another aspect of the problem. Indeed, there is a long history of studying polynomials $F \in \boldsymbol{k}[X, Y]$ which appear as components of birational endomorphisms of $\mathbb{A}^2$. To explain this, we recall some definitions. A polynomial $F \in \boldsymbol{k}[X, Y]$ is called a *field generator* if there exists $G \in \boldsymbol{k}(X, Y)$ satisfying $\boldsymbol{k}(F, G) = \boldsymbol{k}(X, Y)$; in the special case where G can be chosen in $\boldsymbol{k}[X, Y]$, one says that F is a *good* field generator. So a good field generator is just the same thing as a component of a birational endomorphism of $\mathbb{A}^2$. By a *generally rational polynomial*[1] we mean an $F \in \boldsymbol{k}[X, Y]$ such that, for almost all $\lambda \in \boldsymbol{k}$, $F - \lambda$ is an irreducible polynomial whose zero set in $\mathbb{A}^2$ is a rational curve (where "almost all" means "all but possibly finitely many"). If $\operatorname{char} \boldsymbol{k} = 0$, then field generators and generally rational polynomials are one and the same thing (this is noted in [Miyanishi and Sugie 1980]; see [Daigle 2013] for the positive characteristic case). The study of these polynomials is a classical subject, as is clear from considering the following (incomplete) list of references: [Nishino 1968; 1969; 1970; Saitō 1972; 1977; Jan 1974; Russell 1975; 1977; Miyanishi and Sugie 1980; Kaliman 1992; Artal Bartolo and Cassou-Noguès 1996; Neumann and Norbury 2002; Cassou-Noguès 2005; Sasao 2006; Daigle 2013; Cassou-Noguès and Daigle 2013].

This paper is a contribution to the theory of birational endomorphisms of $\mathbb{A}^2$. Our methods are those of [Daigle 1991a; 1991b], and we place ourselves at the same level of generality as in those papers: the base field $\boldsymbol{k}$ is algebraically closed but otherwise arbitrary.

In [Daigle 1991a; 1991b] and [Cassou-Noguès and Russell 2007], there is a tendency to restrict attention to irreducible elements of $\operatorname{Bir}(\mathbb{A}^2)$. Going in an orthogonal direction, the present paper is mainly concerned with *compositions* of birational endomorphisms. This choice is motivated by many reasons. First, the examples given in the three papers above show that $\operatorname{Bir}(\mathbb{A}^2)$ contains a great diversity of irreducible elements of arbitrarily high complexity; since the task of finding all irreducible elements is probably hopeless, it seems to us that finding more examples of them might be less relevant than, say, trying to understand the monoid structure of $\operatorname{Bir}(\mathbb{A}^2)$. Also, a significant portion of this paper is geared towards proving Theorem 3.15, which we need in the forthcoming [Cassou-Noguès and Daigle $\geq$ 2014] to prove the following fact: *Let $\boldsymbol{k}$ be an arbitrary field and $A_0 \supset A_1 \supset \cdots$ an infinite, strictly descending sequence of rings such that* (i) *each A_i is a polynomial ring in 2 variables over $\boldsymbol{k}$*; (ii) *all A_i have the same field of fractions*; *and* (iii) *the ring $R = \bigcap_i A_i$ is not equal to $\boldsymbol{k}$. Then $R = \boldsymbol{k}[F]$ for some F, where F is a good field generator of A_0 and a variable of A_i for $i \gg 0$.* Moreover, if

[1] Generally rational polynomials are sometimes called "rational polynomials" or "generically rational polynomials".

one is interested in field generators and generally rational polynomials, one should not restrict one's attention to irreducible endomorphisms. In this respect we point out that the components of the morphisms described by Theorem 3.15 are precisely the "rational polynomials of simple type" listed in [Neumann and Norbury 2002].

The paper is organized as follows.

Section 1 contains some preliminary observations on "admissible" and "weakly admissible" configurations of curves in $\mathbb{A}^2$.

Section 2 gives several new results in the general theory of birational endomorphisms of $\mathbb{A}^2$ (in particular Proposition 2.9, Proposition 2.13, Corollary 2.14 and Lemma 2.17, but also several useful lemmas).

Given $f \in \text{Bir}(\mathbb{A}^2)$, let $\text{Miss}(f)$ (resp. $\text{Cont}(f)$) be the set of missing curves (resp. contracting curves) of f; see 2.2 for definitions. Section 3 studies those $f \in \text{Bir}(\mathbb{A}^2)$ satisfying the condition that $\text{Miss}(f)$ is a weakly admissible configuration or the stronger condition that $\text{Miss}(f)$ is an admissible configuration or the even stronger condition that both $\text{Miss}(f)$ and $\text{Cont}(f)$ are admissible configurations. The main result of Section 3, Theorem 3.15, gives a complete description of these three subsets of $\text{Bir}(\mathbb{A}^2)$.

While Sections 2 and 3 are mainly concerned with individual endomorphisms, Section 4 considers the monoid structure of $\text{Bir}(\mathbb{A}^2)$. The first part of that section shows, in particular, that if S is a subset of $\text{Bir}(\mathbb{A}^2)$ such that $\text{Aut}(\mathbb{A}^2) \cup S$ generates $\text{Bir}(\mathbb{A}^2)$ as a monoid, then $\{\deg f \mid f \in S\}$ is not bounded and $|S| = |\boldsymbol{k}|$ (giving a very strong negative answer to the already mentioned question of Abhyankar). The second part shows that the submonoid $\mathcal{A}$ of $\text{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms is "factorially closed" in $\text{Bir}(\mathbb{A}^2)$; i.e., if $f, g \in \text{Bir}(\mathbb{A}^2)$ are such that $g \circ f \in \mathcal{A}$, then $f, g \in \mathcal{A}$.

Conventions. All algebraic varieties (in particular all curves and surfaces) are irreducible and reduced. All varieties and morphisms are over an algebraically closed field $\boldsymbol{k}$ of arbitrary characteristic ($\boldsymbol{k}$ is assumed to be algebraically closed from Definition 1.3 until the end of the paper). The word "point" means "closed point", unless otherwise specified.

All rings are commutative and have a unity. The symbol A^* denotes the set of units of a ring A. If A is a subring of a ring B and $n \in \mathbb{N}$, the notation $B = A^{[n]}$ means that B is isomorphic (as an A-algebra) to the polynomial ring in n variables over A. We adopt the conventions that $0 \in \mathbb{N}$, that "$\subset$" means strict inclusion and that "$\setminus$" denotes set difference.

1. Admissible configurations of curves in $\mathbb{A}^2$

Recall the following terminology. Let $\boldsymbol{k}$ be a field, $A = \boldsymbol{k}^{[2]}$, and $\mathbb{A}^2_{\boldsymbol{k}} = \text{Spec}\, A$. We abbreviate $\mathbb{A}^2_{\boldsymbol{k}}$ to $\mathbb{A}^2$. By a *coordinate system of* A, we mean an ordered pair (F, G)

of elements of A satisfying $A=\boldsymbol{k}[F,G]$. A *variable* of A is an element $F \in A$ for which there exists G satisfying $\boldsymbol{k}[F,G]=A$.

Let $F \in A$ be an irreducible element and let $C \subset \mathbb{A}^2$ be the zero set of F (i.e., the set of prime ideals $\mathfrak{p} \in \operatorname{Spec} A=\mathbb{A}^2$ satisfying $F \in \mathfrak{p}$); we call C a *line* if $A/FA=\boldsymbol{k}^{[1]}$ and a *coordinate line* if F is a variable of A. Note that C is a line if and only if $C \cong \mathbb{A}^1$; given a coordinate system (X,Y) of A, C is a coordinate line if and only if some automorphism of $\mathbb{A}^2$ maps C onto the zero set of X. It is clear that all coordinate lines are lines, and the epimorphism theorem [Abhyankar and Moh 1975; Suzuki 1974] states that the converse is true if $\operatorname{char}\boldsymbol{k}=0$. It is known that not all lines are coordinate lines if $\operatorname{char}\boldsymbol{k}\neq 0$ (on the subject of lines which are not coordinate lines, see, e.g., [Ganong 2011] for a recent survey). Coordinate lines are sometimes called rectifiable lines.

By a *coordinate system of* $\mathbb{A}^2=\operatorname{Spec} A$, we mean a coordinate system of A. That is, a coordinate system of $\mathbb{A}^2$ is a pair $(X,Y)\in A\times A$ satisfying $A=\boldsymbol{k}[X,Y]$.

We adopt the viewpoint that A (or $\mathbb{A}^2$) does not come equipped with a preferred coordinate system, i.e., no coordinate system is better than the others. This may be confusing to some readers, especially those who like to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$, because any such identification inevitably depends on the choice of a coordinate system. So perhaps the following trivial remarks deserve to be made:

1.1. Let $\mathfrak{C}$ denote the set of coordinate systems of $\mathbb{A}^2$ (or A).

(a) Given $\mathfrak{c}=(X,Y)\in\mathfrak{C}$ and an element $F\in A$, one can consider the map $\boldsymbol{k}^2\to\boldsymbol{k}$, $(x,y)\mapsto F(x,y)$, defined by first writing $F=\sum_{i,j}a_{ij}X^iY^j$ with $a_{ij}\in\boldsymbol{k}$ (recall that $A=\boldsymbol{k}[X,Y]$) and then setting $F(x,y)=\sum_{i,j}a_{ij}x^iy^j$ for $(x,y)\in\boldsymbol{k}^2$. One can then consider the zero set $\boldsymbol{Z}(F)\subseteq\boldsymbol{k}^2$ of that map F. We stress that the map $(x,y)\mapsto F(x,y)$ and the set $\boldsymbol{Z}(F)$ depend on *both* F *and* $\mathfrak{c}$; we should write $F_{\mathfrak{c}}(x,y)$ and $\boldsymbol{Z}_{\mathfrak{c}}(F)$, but we omit the $\mathfrak{c}$.

(b) Let $P,Q\in\boldsymbol{k}[T_1,T_2]$, where $P=\sum_{i,j}a_{ij}T_1^iT_2^j$ and $Q=\sum_{i,j}b_{ij}T_1^iT_2^j$ with $a_{ij},b_{ij}\in\boldsymbol{k}$.

  (i) The pair (P,Q) determines the map $\boldsymbol{k}^2\to\boldsymbol{k}^2$, $(x,y)\mapsto(P(x,y),Q(x,y))$, where we define $P(x,y)=\sum_{i,j}a_{ij}x^iy^j$ and $Q(x,y)=\sum_{i,j}b_{ij}x^iy^j$.

  (ii) Choose $\mathfrak{c}=(X,Y)\in\mathfrak{C}$. Then $(P,Q,\mathfrak{c})$ determines the morphism of schemes $f:\mathbb{A}^2\to\mathbb{A}^2$ defined by stipulating that f corresponds to the $\boldsymbol{k}$-homomorphism $A\to A$ which maps X to $P(X,Y)=\sum_{i,j}a_{ij}X^iY^j$ and Y to $Q(X,Y)=\sum_{i,j}b_{ij}X^iY^j$ (where $P(X,Y),Q(X,Y)\in A=\boldsymbol{k}[X,Y]$).

(c) Suppose that $\mathfrak{c}=(X,Y)\in\mathfrak{C}$ has been chosen. Then it is convenient to define morphisms of schemes $\mathbb{A}^2\to\mathbb{A}^2$ simply by giving the corresponding polynomial maps $\boldsymbol{k}^2\to\boldsymbol{k}^2$ (this will be done repeatedly in Section 3). To do so, we abuse language as follows: given $P, Q \in \boldsymbol{k}[T_1, T_2]$, the sentence

$$\text{“}f : \mathbb{A}^2 \to \mathbb{A}^2 \textit{ is defined by } f(x, y) = (P(x, y), Q(x, y))\text{”}$$

means that $f : \mathbb{A}^2 \to \mathbb{A}^2$ is the morphism of schemes determined by $(P, Q, \mathfrak{c})$ as in remark (b). For instance one can define $f : \mathbb{A}^2 \to \mathbb{A}^2$ by $f(x, y) = (x, xy)$, always keeping in mind that this f depends on $\mathfrak{c}$.

1.2. Lemma. *Let $F, G \in A = \boldsymbol{k}^{[2]}$, where $\boldsymbol{k}$ is any field, and suppose that each of F, G is a variable of A. Consider the ideal (F, G) of A generated by F and G.*

(a) *If $(F, G) = A$, then $G = aF + b$ for some $a, b \in \boldsymbol{k}^*$.*

(b) *If $A/(F, G) = \boldsymbol{k}$, then $A = \boldsymbol{k}[F, G]$.*

Proof. Choose Y such that $A = \boldsymbol{k}[F, Y]$ and define $X = F$. Then $A = \boldsymbol{k}[X, Y]$ and we may write G as a polynomial in X, Y, say $G = P(X, Y)$.

Suppose that $(F, G) = A$. Then $1 \in (F, G) = (X, P(X, Y)) = (X, P(0, Y))$ implies $P(0, Y) \in \boldsymbol{k}^*$. Writing $P(0, Y) = b \in \boldsymbol{k}^*$, we obtain that $G - b = P(X, Y) - P(0, Y)$ is divisible by X; as $G - b$ is irreducible, $G - b = aX = aF$ with $a \in \boldsymbol{k}^*$, and (a) is proved.

To prove (b), we first observe that since $G = P(X, Y)$ is a variable of $\boldsymbol{k}[X, Y]$ and $P(X, Y) \notin \boldsymbol{k}[X]$, we have that P is “almost monic” in Y, i.e.,

$$P(X, Y) = aY^d + p_1(X)Y^{d-1} + \cdots + p_d(X) \tag{1}$$

with $d \geq 1$, $a \in \boldsymbol{k}^*$ and $p_i(X) \in \boldsymbol{k}[X]$ for $i = 1, \ldots, d$. Now

$$\boldsymbol{k} = A/(F, G) = \boldsymbol{k}[X, Y]/(X, P(X, Y)) = \boldsymbol{k}[X, Y]/(X, P(0, Y))$$

implies that $\deg P(0, Y) = 1$, so $d = 1$ in (1). Then $G = aY + p_1(X)$ and $\boldsymbol{k}[F, G] = \boldsymbol{k}[X, aY + p_1(X)] = \boldsymbol{k}[X, Y] = A$. $\square$

Until the end of this paper, we assume that $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic. Consider $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$.

1.3. Definition. Let $C_1, \ldots, C_n$ $(n \geq 0)$ be distinct curves in $\mathbb{A}^2$, and consider the set $S = \{C_1, \ldots, C_n\}$. We say that S is a *weakly admissible configuration* if

(a) each C_i is a coordinate line;

(b) for every choice of $i \neq j$ such that $C_i \cap C_j$ is nonempty, $C_i \cap C_j$ is one point and the local intersection number of C_i and C_j at that point is equal to 1.

1.4. Lemma. *Given distinct curves $C_1, \ldots, C_n$ $(n \geq 0)$ in $\mathbb{A}^2$, the following are equivalent*:

(a) *$\{C_1, \ldots, C_n\}$ is a weakly admissible configuration.*

(b) *There exists a coordinate system of $\mathbb{A}^2$ with respect to which all C_i have degree* 1.

Proof. We show that (a) implies (b), the converse being trivial. Suppose that (a) holds. Write $\mathbb{A}^2 = \operatorname{Spec} A$, where $A = \boldsymbol{k}^{[2]}$. We may assume that $n \geq 2$, otherwise the assertion is trivial. Let $F_1, \dots, F_n \in A$ be variables of A whose zero sets are $C_1, \dots, C_n$ respectively. Condition (a) implies that, whenever $i \neq j$, we have either $(F_i, F_j) = A$ or $A/(F_i, F_j) = \boldsymbol{k}$. Consider the graph G whose vertex set is $\{F_1, \dots, F_n\}$ and in which distinct vertices F_i, F_j are joined by an edge if and only if $A/(F_i, F_j) = \boldsymbol{k}$.

In the case where G is discrete, Lemma 1.2 implies that $F_i = a_i F_1 + b_i$ with $a_i, b_i \in \boldsymbol{k}^*$ for $i = 2, \dots, n$. Let $X = F_1$ and let Y be such that $A = \boldsymbol{k}[X, Y]$. Then all F_i have degree 1 with respect to the coordinate system (X, Y).

From now on, assume that G is not discrete. Then we may assume that F_1, F_2 are joined by an edge. By Lemma 1.2, $\boldsymbol{k}[F_1, F_2] = A$. Let $X_1 = F_1$ and $X_2 = F_2$; then $A = \boldsymbol{k}[X_1, X_2]$ and for each $i \in \{3, \dots, n\}$ we have:

- If F_i, F_1 are not joined by an edge, then Lemma 1.2 implies that $F_i = a_i X_1 + b_i$ for some $a_i, b_i \in \boldsymbol{k}^*$, so F_i has degree 1 with respect to (X_1, X_2).
- If F_i, F_2 are not joined by an edge, then $F_i = a_i X_2 + b_i$ for some $a_i, b_i \in \boldsymbol{k}^*$, so F_i has degree 1 with respect to (X_1, X_2).
- If F_i is linked to each of F_1, F_2 by edges, then $\boldsymbol{k}[F_i, F_1] = A = \boldsymbol{k}[F_i, F_2]$, so $F_2 = aF_1 + \beta(F_i)$ for some $a \in \boldsymbol{k}^*$ and $\beta(T) \in \boldsymbol{k}[T]$; then $\beta(F_i) = X_2 - aX_1$ has degree 1 with respect to (X_1, X_2) and consequently F_i has degree 1.

So all F_i have degree 1 with respect to the coordinate system (X_1, X_2). □

1.5. Let $C_1, \dots, C_n$ ($n \geq 0$) be distinct curves in a nonsingular surface W. We say that $\sum_{i=1}^n C_i$ is an *SNC-divisor* of W if

- each C_i is a nonsingular curve;
- for every choice of $i \neq j$ such that $C_i \cap C_j$ is nonempty, $C_i \cap C_j$ is one point and the local intersection number of C_i and C_j at that point is equal to 1;
- if i, j, k are distinct then $C_i \cap C_j \cap C_k = \varnothing$.

If $D = \sum_{i=1}^n C_i$ is an SNC-divisor of W, we write $\mathcal{G}(D, W)$ for the graph whose vertex set is $\{C_1, \dots, C_n\}$ and in which distinct vertices C_i, C_j are joined by an edge if and only if $C_i \cap C_j \neq \varnothing$.

1.6. Definition. Let $C_1, \dots, C_n$ ($n \geq 0$) be distinct curves in $\mathbb{A}^2$. We say that the set $\{C_1, \dots, C_n\}$ is an *admissible configuration* if

(a) each C_i is a coordinate line;

(b) $D = \sum_{i=1}^n C_i$ is an SNC-divisor of $\mathbb{A}^2$;

(c) the graph $\mathcal{G}(D, \mathbb{A}^2)$ defined in 1.5 is a forest.

1.7. Proposition. *Given distinct curves $C_1, \dots, C_n$ ($n \geq 0$) in $\mathbb{A}^2$, the following are equivalent*:

(a) $\{C_1, \dots, C_n\}$ *is an admissible configuration.*

(b) *There exists a coordinate system (X, Y) of $\mathbb{A}^2$ such that $\bigcup_{i=1}^n C_i$ is the zero set of $\varphi(X)Y^j$ for some $j \in \{0, 1\}$ and some $\varphi(X) \in \boldsymbol{k}[X] \setminus \{0\}$.*

Proof. It's enough to show that (a) implies (b), as the converse is trivial. Assume that (a) holds. By Lemma 1.4, we may choose a coordinate system which respect to which all C_i have degree 1. As $D = \sum_{i=1}^n C_i$ is an SNC-divisor and $\mathcal{G}(D, \mathbb{A}^2)$ is a forest, $\bigcup_{i=1}^n C_i$ must be either

- a union of n parallel lines, or
- a union of $n-1$ parallel lines with another line.

Indeed, any other configuration of lines would either contain three concurrent lines or produce a circuit in the graph. Now it is clear that (b) is satisfied. □

2. Birational morphisms $f : X \to Y$ of nonsingular surfaces with special emphasis on the case $X = Y = \mathbb{A}^2$

Throughout this section, $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic and we abbreviate $\mathbb{A}^2_{\boldsymbol{k}}$ to $\mathbb{A}^2$. We consider birational morphisms $f : X \to Y$, where X and Y are nonsingular algebraic surfaces over $\boldsymbol{k}$ (a morphism $f : X \to Y$ is *birational* if there exist nonempty Zariski-open subsets $U \subseteq X$ and $V \subseteq Y$ such that f restricts to an isomorphism $U \to V$). We are particularly interested in the case $X = \mathbb{A}^2 = Y$.

Essentially all the material given in 2.1–2.8 can be found in [Daigle 1991a]. From 2.9 to the end of the section, the material appears to be new (except 2.13(a)).

2.1. Let $f : X \to Y$ and $f' : X' \to Y'$ be birational morphisms of nonsingular surfaces. We say that f, f' are *equivalent* ($f \sim f'$) if there exist isomorphisms $x : X \to X'$ and $y : Y \to Y'$ such that $y \circ f = f' \circ x$.

2.2 [Daigle 1991a, 1.2]. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. A *missing curve* of f is a curve $C \subset Y$ whose intersection with the image of f is a finite set of points. We write $\mathrm{Miss}(f)$ for the set of missing curves of f, $q(f)$ for the cardinality of $\mathrm{Miss}(f)$ and $q_0(f)$ for the number of missing curves of f which are disjoint from $f(X)$. A *contracting curve* of f is a curve $C \subset X$ such that $f(C)$ is a point. The set of contracting curves of f is denoted $\mathrm{Cont}(f)$, and $c(f)$ denotes the cardinality of $\mathrm{Cont}(f)$. The natural numbers $q(f)$, $q_0(f)$ and $c(f)$ are invariant with respect to equivalence of birational morphisms (2.1), i.e., $f \sim f' \Longrightarrow c(f) = c(f')$ and similarly for q and q_0. Call a point of Y a *fundamental point* of f if it is $f(C)$ for some contracting curve C of f. The set of fundamental points of f is also called the *center* of f, denoted $\mathrm{cent}(f)$. The *exceptional locus* of f is defined to be $\operatorname{exc}(f) = f^{-1}(\operatorname{cent}(f))$ and is equal to the union of the contracting curves of f.

2.3 [Daigle 1991a, 1.1 and 1.2]. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. There exists a commutative diagram

(2)
$$\begin{tikzcd} Y_n \arrow[r, "\pi_n"] & \cdots \arrow[r, "\pi_1"] & Y_0 \arrow[d, equal] \\ X \arrow[u, hook] \arrow[rr, "f"'] & & Y \end{tikzcd}$$

where "$\hookrightarrow$" denotes an open immersion and, for each i, $\pi_i : Y_i \to Y_{i-1}$ is the blowing-up of the nonsingular surface Y_{i-1} at a point $P_i \in Y_{i-1}$.

Let $n(f)$ be the least natural number n for which there exists a diagram (2). Note that $n(f)$ is invariant with respect to equivalence of birational morphisms.

For each $i \in \{1, \dots, n\}$, consider the exceptional curve $E_i = \pi_i^{-1}(P_i) \subset Y_i$, and let the same symbol E_i also denote the strict transform of E_i in Y_n. It is clear that the union of the contracting curves of f is the intersection of $E_1 \cup \cdots \cup E_n \subset Y_n$ with the open subset X of Y_n; thus

(3) $$c(f) \le n(f);$$

(4) *each contracting curve is nonsingular and rational,* $D = \sum_{C \in \operatorname{Cont}(f)} C$ *is an SNC-divisor of* X *and the graph* $\mathcal{G}(D, X)$ *is a forest.*[2]

Given $i \in \{1, \dots, n\}$, consider the complete curve $E_i \subset Y_n$. Note that if S is a projective nonsingular surface and $\mu : Y_n \hookrightarrow S$ is an open immersion, the self-intersection number of $\mu(E_i)$ in S is independent of the choice of (S, μ); we denote this number by $(E_i^2)_{Y_n}$. Then the following holds by [ibid., 1.2(c)]:

(5) *Diagram* (2) *satisfies* $n = n(f)$ *if and only if* $(E_i^2)_{Y_n} \le -2$ *holds for all* $i \in \{1, \dots, n\}$ *such that* $E_i \subseteq Y_n \setminus X$.

In particular, if Diagram (2) satisfies $n = n(f)$, then

(6) $$\operatorname{cent}(f) = \{(\pi_1 \circ \cdots \circ \pi_{i-1})(P_i) \mid 1 \le i \le n\}.$$

The following remarks are trivial consequences of 2.3, but are very useful:

2.4. Remarks. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces.

(a) For each $C \in \operatorname{Miss}(f)$, we have $C \cap f(X) \subseteq \operatorname{cent}(f)$. In particular, the condition $q_0(f) = 0$ is equivalent to "every missing curve contains a fundamental point".

(b) Let $C \subset Y$ be a curve. Then there exists at most one curve $C' \subset X$ such that $f(C')$ is a dense subset of C. Moreover, C' exists if and only if $C \notin \operatorname{Miss}(f)$.

[2] Note that contracting curves are not necessarily isomorphic to $\mathbb{A}^1$. So, in the case $X = \mathbb{A}^2 = Y$, $\operatorname{Cont}(f)$ is not necessarily an admissible configuration in the sense of Definition 1.6.

2.5. Lemma. *If* $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ *are birational morphisms then*

$$n(g \circ f) = n(g) + n(f).$$

Proof. Follows from [ibid., 2.12]. □

2.6. Lemma. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism.*

(a) $q(f) = c(f) \leq n(f)$.

(b) f *is an isomorphism* $\iff n(f) = 0 \iff c(f) = 0 \iff q(f) = 0$.

(c) *Each missing curve of* f *is rational with one place at infinity.*

(d) *Each fundamental point belongs to some missing curve*; *each missing curve contains some fundamental point.*

(e) *If a point* P *is a singular point of some missing curve of* f *or a common point of two missing curves*, *then* P *is a fundamental point of* f.

Proof. The equality $q(f) = c(f)$ in (a) follows from [ibid., 4.3(a)], and $c(f) \leq n(f)$ was noted in (3). Assertion (b) follows from the observation that if $n(f) = 0$ or $c(f) = 0$ then f is an open immersion $\mathbb{A}^2 \hookrightarrow \mathbb{A}^2$ and hence an automorphism. Assertion (c) follows from [ibid., 4.3(c)]. The first (resp. the second) assertion of (d) follows from [ibid., 2.1] (resp. from the claim that $q_0(f) = 0$, in [ibid., 4.3(a)]). Refer to [Daigle 2010, 1.6] for a proof of assertion (e). □

Several of the above facts are stated in greater generality in [ibid.]. For instance, if $X \xrightarrow{f} Y \xrightarrow{g} Z$ are birational morphisms of nonsingular surfaces and $q_0(f) = 0$, then (by [ibid., 1.3]) $n(g \circ f) = n(g) + n(f)$. Also, if X, Y are nonsingular surfaces satisfying $\Gamma(X, \mathcal{O}_X)^* = \boldsymbol{k}^*$ and $\mathrm{Cl}(Y) = 0$, then (by [ibid., 2.11]) every birational morphism $f : X \to Y$ satisfies $q_0(f) = 0$. The following generalization of Lemma 2.6(a) also deserves to be noted:

2.7. Lemma. *Let* $f : X \to X$ *be a birational morphism*, *where* X *is a nonsingular surface. Then* $c(f) = q(f)$.

Proof. Follows from Remark 2.13 of [ibid.]. □

2.8. Lemma. *Let* $f : X \to Y$ *be a birational morphism of nonsingular surfaces and* Γ_f *the union of the missing curves of* f. *If* X *is affine then the following hold*:

(a) $\mathrm{cent}(f) \subseteq \Gamma_f$.

(b) $Y \setminus \Gamma_f$ *is the interior of* $f(X)$ *and* $f^{-1}(Y \setminus \Gamma_f) = X \setminus \mathrm{exc}(f)$.

(c) *The surfaces* $X \setminus \mathrm{exc}(f)$ *and* $Y \setminus \Gamma_f$ *are affine*, *and* f *restricts to an isomorphism* $X \setminus \mathrm{exc}(f) \to Y \setminus \Gamma_f$.

Proof. Follows from [ibid., Prop. 2.1] and its proof. □

2.9. Proposition. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism. If* P *is a singular point of a missing curve of* f, *then* P *belongs to at least two missing curves of* f.

Proof. By Lemma 2.6(e), P is a fundamental point of f; so it suffices to show that if a fundamental point P belongs to only one missing curve C, then the multiplicity $\mu(P, C)$ of C at P is equal to 1. So assume that P is a fundamental point which belongs to only one missing curve C. Choose a diagram (2) satisfying $n = n(f)$, and let the notation P_i, E_i, etc., be as in 2.3. In fact let us choose diagram (2) in such a way that $P_1 = P$ and, for some $s \in \{1, \dots, n\}$,

(7) $$P_2, \dots, P_s \text{ are infinitely near } P_1, \text{ but } P_{s+1}, \dots, P_n \text{ are not.}$$

Let us label the missing curves as $C_1, \dots, C_q$, where

(8) $$P_1 \in C_j \iff j = 1.$$

The diagram (2) together with the ordering $(C_1, \dots, C_q)$ of the set of missing curves constitutes a "minimal decomposition" of f, in the terminology of [ibid., 1.2(h)]. This minimal decomposition $\mathcal{D}$ determines matrices $\mu_{\mathcal{D}}$, $\mathcal{E}_{\mathcal{D}}$, $\varepsilon_{\mathcal{D}}$ and $\varepsilon'_{\mathcal{D}}$, defined in [ibid., 2.8]. These are matrices with entries in $\mathbb{N}$, and the result [ibid., 4.3(b)] asserts that the product $\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}}$ is a square matrix of determinant ± 1. We shall now argue that the condition $\det(\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}}) = \pm 1$ implies that $\mu(P_1, C_1) = 1$ (this will complete the proof of the proposition). By (7), the $n \times n$ matrix $\mathcal{E}_{\mathcal{D}}$ has the shape

$$\mathcal{E}_{\mathcal{D}} = (e_{ij}) = \left(\begin{array}{c|c} \mathcal{E}_0 & 0 \\ \hline 0 & \mathcal{E}_1 \end{array}\right) \quad (\text{with } e_{ij} \in \mathbb{N} \text{ for all } i, j),$$

where $\mathcal{E}_0$ is an $s \times s$ lower-triangular matrix with zero diagonal, and where

$$\text{the first row is the only zero row of } \mathcal{E}_0.$$

Consider the $n \times n$ matrix $\varepsilon_{\mathcal{D}}$, determined by $\mathcal{E}_{\mathcal{D}}$ as explained in [ibid., 2.7]. The already mentioned properties of $\mathcal{E}_{\mathcal{D}}$ imply that $\varepsilon_{\mathcal{D}}$ is as follows:

$$\varepsilon_{\mathcal{D}} = (\varepsilon_{ij}) = \left(\begin{array}{c|c} \varepsilon_0 & 0 \\ \hline 0 & \varepsilon_1 \end{array}\right) \quad (\text{with } \varepsilon_{ij} \in \mathbb{N} \text{ for all } i, j),$$

where ε_0 is an $s \times s$ lower-triangular matrix with diagonal entries equal to 1 and where

$$\text{all entries in the first column of } \varepsilon_0 \text{ are positive.}$$

Next, $\varepsilon'_{\mathcal{D}}$ is the submatrix of $\varepsilon_{\mathcal{D}}$ obtained by deleting the i-th row for each $i \in J$, where $J = \{\, i \mid 1 \le i \le n,\ E_i \cap X = \varnothing \text{ in } Y_n \}$ in the notation of 2.3 (J is defined in [ibid., 1.2(h)]). Let $j_0 = |J \cap \{1, \dots, s\}|$; then the $(n - |J|) \times n$ matrix $\varepsilon'_{\mathcal{D}}$ has the form

$$\varepsilon'_{\mathcal{D}} = \left(\begin{array}{c|c} \varepsilon'_0 & 0 \\ \hline 0 & \varepsilon'_1 \end{array}\right),$$

where ε'_0 is an $(s - j_0) \times s$ matrix with entries in $\mathbb{N}$ and

(9) $$\text{all entries in the first column of } \varepsilon'_0 \text{ are positive.}$$

Finally, consider the $n \times q$ matrix $\mu_{\mathfrak{D}}$; by (8),

$$\mu_{\mathfrak{D}} = \left(\begin{array}{c|c} F & 0 \\ \hline G & H \end{array}\right), \quad \text{where } F = \begin{pmatrix} \mu(P_1, C_1) \\ \vdots \\ \mu(P_s, C_1) \end{pmatrix} \text{ is } s \times 1.$$

We have

$$\varepsilon'_{\mathfrak{D}}\mu_{\mathfrak{D}} = \left(\begin{array}{c|c} \varepsilon'_0 & 0 \\ \hline 0 & \varepsilon'_1 \end{array}\right)\left(\begin{array}{c|c} F & 0 \\ \hline G & H \end{array}\right) = \left(\begin{array}{c|c} \varepsilon'_0 F & 0 \\ \hline \varepsilon'_1 G & \varepsilon'_1 H \end{array}\right), \tag{10}$$

where the block $\varepsilon'_0 F$ has size $(s - j_0) \times 1$. By (7), $(E_s^2)_{Y_n} = -1$; so $E_s \nsubseteq Y_n \setminus X$ by (5) and hence $s \notin J$ by definition of J. It follows that $s - j_0 \geq 1$. In view of (10), the fact that $\det(\varepsilon'_{\mathfrak{D}}\mu_{\mathfrak{D}}) = \pm 1$ implies that $s - j_0 = 1$ and that the unique entry of $\varepsilon'_0 F$ is ± 1. We have $\{1, \dots, s\} \setminus J = \{s\}$, so $\varepsilon'_0 = (\varepsilon_{s1} \dots \varepsilon_{ss})$ and $\sum_{j=1}^{s} \varepsilon_{sj}\mu(P_j, C_1) = \pm 1$. Since $\varepsilon_{sj} \in \mathbb{N}$ for all j and (by (9)) $\varepsilon_{s1} \geq 1$, we get $1 \leq \mu(P_1, C_1) \leq \sum_{j=1}^{s} \varepsilon_{sj}\mu(P_j, C_1) = \pm 1$, so $\mu(P_1, C_1) = 1$. □

2.10. Remark. Let $\varphi : X \to Y$ be a dominant morphism of nonsingular surfaces. By a *deficient curve* of φ, we mean a curve $C \subset Y$ satisfying

$$\text{for almost all points } P \in C, \quad |f^{-1}(P)| < [\boldsymbol{k}(X) : \boldsymbol{k}(Y)]_s$$

where "almost all" means "all except possibly finitely many", "$|\cdot|$" denotes cardinality, $\boldsymbol{k}(X)$ and $\boldsymbol{k}(Y)$ are the function fields of X and Y and $[\boldsymbol{k}(X) : \boldsymbol{k}(Y)]_s$ is the separable degree of the field extension $\boldsymbol{k}(X)/\boldsymbol{k}(Y)$. Note that φ has finitely many deficient curves and that if φ is birational then the deficient curves are precisely the missing curves.

Then it is interesting to note that Proposition 2.9 does not generalize to all dominant morphisms $\mathbb{A}^2 \to \mathbb{A}^2$; i.e., *it is not the case that each singular point of a deficient curve is a common point of at least two deficient curves.* This is shown by the following example, in which we assume that $\operatorname{char} \boldsymbol{k} = 0$.

Choose a coordinate system of $\mathbb{A}^2$ and define morphisms $f, g, h : \mathbb{A}^2 \to \mathbb{A}^2$ by

$$f(x, y) = (x, xy), \quad g(x, y) = (x + y(y^2 - 1), y), \quad h(x, y) = (x, y^2).$$

Note that f is a SAC and $g \in \operatorname{Aut}(\mathbb{A}^2)$. Define $\varphi = h \circ g \circ f : \mathbb{A}^2 \to \mathbb{A}^2$. Then φ has two deficient curves C_1 and C_2, where

- C_1 is "$y = 0$" (the deficient curve of h);
- C_2 is "$x^2 - y(y-1)^2 = 0$" (the image by $h \circ g$ of the missing curve of f).

Moreover, $(0, 1)$ is a singular point of C_2 which is not on C_1.

2.11. Lemma. *Let $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ be birational morphisms. Then*

$$\operatorname{cent}(g \circ f) = \operatorname{cent}(g) \cup g(\operatorname{cent}(f)).$$

In particular, every fundamental point of g is a fundamental point of $g \circ f$.

Proof. Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces. The reader may easily verify that $\mathrm{cent}(g \circ f) \subseteq \mathrm{cent}(g) \cup g(\mathrm{cent}(f))$ and $g(\mathrm{cent}(f)) \subseteq \mathrm{cent}(g \circ f)$. In order to obtain the desired equality, it remains to show that

$$\mathrm{cent}(g) \subseteq \mathrm{cent}(g \circ f). \tag{11}$$

We claim that (11) is true whenever $q_0(f) = 0$. Indeed, consider $P \in \mathrm{cent}(g)$. Then there exists a curve $C \subset Y$ such that $g(C) = \{P\}$. If $C \notin \mathrm{Miss}(f)$, then (by 2.4(b)) there exists a curve $C' \subset X$ such that $f(C')$ is a dense subset of C; in particular, $(g \circ f)(C') = \{P\}$ and hence $P \in \mathrm{cent}(g \circ f)$. If $C \in \mathrm{Miss}(f)$, then, since $q_0(f) = 0$, 2.4(a) implies that some fundamental point Q of f lies on C; then there exists a curve $C' \subset X$ such that $f(C') = \{Q\}$. Then $(g \circ f)(C') = \{P\}$ and hence $P \in \mathrm{cent}(g \circ f)$.

By [Daigle 1991a, 2.11], the condition $q_0(f) = 0$ is satisfied whenever $\mathrm{Cl}(Y) = 0$ and $\Gamma(X, \mathcal{O}_X)^* = \boldsymbol{k}^*$. In particular, if $X = \mathbb{A}^2 = Y$, then $q_0(f) = 0$, so (11) holds and consequently $\mathrm{cent}(g \circ f) = \mathrm{cent}(g) \cup g(\mathrm{cent}(f))$. □

2.12. Lemma. *Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces and let Γ_f (resp. Γ_g, $\Gamma_{g \circ f}$) be the union of all missing curves of f (resp. of g, $g \circ f$).*

(a) *$\Gamma_{g \circ f}$ is equal to the union of all 1-dimensional components of $\Gamma_g \cup \overline{g(\Gamma_f)}$, where $\overline{g(\Gamma_f)}$ denotes the closure of $g(\Gamma_f)$ in Z.*

(b) *If Y is affine, then $\Gamma_{g \circ f} = \Gamma_g \cup \overline{g(\Gamma_f)}$; in particular, each missing curve of f is included in $g^{-1}(\Gamma_{g \circ f})$.*

Proof. To prove (a), it's enough to show that a curve in Z is not included in $\Gamma_{g \circ f}$ if and only if it is not included in $\Gamma_g \cup \overline{g(\Gamma_f)}$. Let $C \subset Z$ be a curve such that $C \nsubseteq \Gamma_{g \circ f}$. Then there exists a curve $C_0 \subset X$ such that $g(f(C_0))$ is a dense subset of C; consequently, the set $C_1 = \overline{f(C_0)}$ is a curve in Y and $g(C_1)$ is dense in C, so C is not a missing curve of g and hence $C \nsubseteq \Gamma_g$. If $C \subseteq \overline{g(\Gamma_f)}$, then there exists a missing curve C_1' of f such that $\overline{g(C_1')} = C$; however, C_1 is the only curve in Y whose image by g is a dense subset of C, and C_1 is not a missing curve of f; so $C \nsubseteq \overline{g(\Gamma_f)}$ and hence $C \nsubseteq \Gamma_g \cup \overline{g(\Gamma_f)}$.

Conversely, let $C \subset Z$ be a curve such that $C \nsubseteq \Gamma_g \cup \overline{g(\Gamma_f)}$. Then $C \nsubseteq \Gamma_g$, so there exists a curve $C_1 \subset Y$ such that $g(C_1)$ is a dense subset of C. Note that C_1 is not a missing curve of f because $C \nsubseteq \overline{g(\Gamma_f)}$; so there exists a curve $C_0 \subset X$ such that $f(C_0)$ is a dense subset of C_1. Then $(g \circ f)(C_0)$ is a dense subset of C and consequently $C \nsubseteq \Gamma_{g \circ f}$. This proves (a).

To prove (b), suppose that Y is affine. If a point $P \in Z$ is an irreducible component of $\Gamma_g \cup \overline{g(\Gamma_f)}$, then $\{P\} = g(C)$, where C is a missing curve of f, so P is a fundamental point of g; since Y is affine, Lemma 2.8 implies that $\mathrm{cent}(g) \subseteq \Gamma_g$, so $P \in \Gamma_g$, which contradicts the hypothesis that P is an irreducible component of $\Gamma_g \cup \overline{g(\Gamma_f)}$. This shows that $\Gamma_g \cup \overline{g(\Gamma_f)}$ is a union of curves, so the equality $\Gamma_{g\circ f} = \Gamma_g \cup \overline{g(\Gamma_f)}$ follows from (a). □

Results 2.13 and 2.14 are valid in all characteristics but are particularly interesting when char $\boldsymbol{k} > 0$ (recall that not all lines are coordinate lines when char $\boldsymbol{k} > 0$).

2.13. Proposition. *Let $f : \mathbb{A}^2 \to \mathbb{A}^2$ be a birational morphism.*

(a) *If a missing curve of f is nonsingular, then it is a coordinate line.*

(b) *If a contracting curve of f has one place at infinity, then it is a coordinate line.*

Proof. Assertion (a) follows from [Daigle 1991a, 4.6]. To prove (b), consider a contracting curve C of f such that C has one place at infinity. We noted in (4) that C is a nonsingular rational curve, so $C \cong \mathbb{A}^1$ is clear.

Let us embed $\operatorname{dom}(f) = \mathbb{A}^2$ in $X \cong \mathbb{P}^2$, and let L be the function field of X and V the prime divisor of $L/\boldsymbol{k}$ whose center on X is the closure of C in X (i.e., V is the DVR $\mathcal{O}_{X,\xi}$ where $\xi \in X$ is the generic point of C). Also embed $\operatorname{codom}(f) = \mathbb{A}^2$ in $Y \cong \mathbb{P}^2$, and note that the center of V on Y is zero-dimensional, since $C \in \operatorname{Cont}(f)$.

Consider the Kodaira dimension $\kappa(V)$ as defined in the introduction of Section 2 of [Ganong 1985]. Then $\kappa(V) < 0$ by [ibid., 2.1] and the fact that the center of V on Y is zero-dimensional; so C is a coordinate line by [ibid., 2.4]. □

2.14. Corollary. *Let C, C' be curves in $\mathbb{A}^2$ such that $C \cong \mathbb{A}^1 \cong C'$, and suppose that there exists a birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$ such that $f(C)$ is a dense subset of C'. Then $f(C) = C'$. Moreover, if one of C, C' is a coordinate line, then both are coordinate lines.*

Proof. It is a simple fact that every dominant morphism $\mathbb{A}^1 \to \mathbb{A}^1$ is finite, hence surjective; so $f(C) = C'$.

If C is a coordinate line then there exists a birational morphism $g : \mathbb{A}^2 \to \mathbb{A}^2$ such that $C \in \operatorname{Miss}(g)$ (choose a coordinate system (X, Y) such that $C = \boldsymbol{Z}(X)$, and take $g(x, y) = (x, xy)$); then $C' \in \operatorname{Miss}(f \circ g)$ by Lemma 2.12, so Proposition 2.13(a) implies that C' is a coordinate line.

If C' is a coordinate line, then there exists a birational morphism $g : \mathbb{A}^2 \to \mathbb{A}^2$ such that $C' \in \operatorname{Cont}(g)$ (choose (X, Y) such that $C' = \boldsymbol{Z}(X)$ and take $g(x, y) = (x, xy)$); then $C \in \operatorname{Cont}(g \circ f)$, so Proposition 2.13(b) implies that C is a coordinate line. □

2.15. Lemma. *Let $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ be birational morphisms.*

(a) *If* $\operatorname{Miss}(f) \subseteq \operatorname{Cont}(g)$, *then* $\operatorname{Miss}(f)$ *is admissible.*

(b) *If* $\operatorname{Cont}(g) \subseteq \operatorname{Miss}(f)$, *then* $\operatorname{Cont}(g)$ *is admissible.*

Proof. Applying statement (4) in 2.3 to the morphism g gives

(12) $D' = \sum_{C \in \mathrm{Cont}(g)} C$ *is an SNC-divisor of* $\mathbb{A}^2$ *and the graph* $\mathcal{G}(D', \mathbb{A}^2)$ *is a forest.*

If $\mathrm{Cont}(g) \subseteq \mathrm{Miss}(f)$ then each element of $\mathrm{Cont}(g)$ is a *nonsingular* missing curve of f, and hence is a coordinate line by 2.13(a); then (12) implies that $\mathrm{Cont}(g)$ is admissible, so (b) is proved.

If $\mathrm{Miss}(f) \subseteq \mathrm{Cont}(g)$ then, by (12), $D = \sum_{C \in \mathrm{Miss}(f)} C$ is an SNC-divisor of $\mathbb{A}^2$ and the graph $\mathcal{G}(D, \mathbb{A}^2)$ is a forest; in particular each missing curve of f is nonsingular and hence is a coordinate line by 2.13(a); so $\mathrm{Miss}(f)$ is admissible and (a) is proved. □

2.16. Lemma. *Let* $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ *be birational morphisms.*

(a) *If* $\mathrm{Miss}(f) \nsubseteq \mathrm{Cont}(g)$, *then* $q(g \circ f) > q(g)$.

(b) *If* $\mathrm{Cont}(g) \nsubseteq \mathrm{Miss}(f)$, *then* $c(g \circ f) > c(f)$.

Proof. (a) It is clear that $\mathrm{Miss}(g) \subseteq \mathrm{Miss}(g \circ f)$. If C is a missing curve of f which is not contracted by g, then the closure $\overline{g(C)}$ of $g(C)$ is a missing curve of $g \circ f$ but not a missing curve of g, so $\mathrm{Miss}(g) \subset \mathrm{Miss}(g \circ f)$ and hence $q(g) < q(g \circ f)$.

(b) We have $\mathrm{Cont}(f) \subseteq \mathrm{Cont}(g \circ f)$. If C is a contracting curve of g which is not a missing curve of f, then there exists a curve $C' \subset \mathbb{A}^2$ such that $f(C')$ is a dense subset of C. Then C' is a contracting curve of $g \circ f$ but not one of f, so $\mathrm{Cont}(f) \subset \mathrm{Cont}(g \circ f)$ and hence $c(f) < c(g \circ f)$. □

In Lemma 2.17 and Proposition 2.18, we write $\#\Gamma$ for the number of irreducible components of a closed set Γ and $\Gamma_\tau = \bigcup_{C \in \mathrm{Miss}(\tau)} C$ for any birational morphism τ of nonsingular surfaces.

2.17. Lemma. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism and* Γ *a union of missing curves of* f *such that*

(13) *each missing curve of* f *is either included in* Γ *or disjoint from* Γ.

Then $\#f^{-1}(\Gamma) = \#\Gamma$ *and* f *can be factored as* $\mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2 \xrightarrow{h} \mathbb{A}^2$, *where* g, h *are birational morphisms*, $\Gamma_h = \Gamma$ *and* $\Gamma_g \cap \mathrm{exc}(h) = \varnothing$.

Lemma 2.17 is an immediate consequence of the next result:

2.18. Proposition. *Let* $f : X \to Y$ *be a birational morphism where* X, Y *are nonsingular affine surfaces, and let* $\Gamma \subset Y$ *be a union of missing curves of* f *satisfying* (13). *Then the following hold*:

(a) f *can be factored as* $X \xrightarrow{g} W \xrightarrow{h} Y$, *where* W *is a nonsingular affine surface*, g, h *are birational morphisms*, $\Gamma_h = \Gamma$, $c(h) = \#f^{-1}(\Gamma)$ *and* $\Gamma_g \cap \mathrm{exc}(h) = \varnothing$.

(b) *If* X, Y *are factorial with trivial units, then* $\#f^{-1}(\Gamma) = \#\Gamma$ *and, in part* (a), W *can be chosen to be factorial with trivial units.*

(c) *If* $X = \mathbb{A}^2$ *and* Y *is factorial, then* $Y = \mathbb{A}^2$ *and we can choose* $W = \mathbb{A}^2$ *in part* (a).

Proof. (a) We may choose a commutative diagram (2) satisfying $n = n(f)$ and in which the blowings-up $\pi_1, \dots, \pi_n$ are ordered in such a way that the points over Γ are blown-up first; i.e., there exists $m \in \{0, \dots, n\}$ such that

$$\{ i \in \{1, \dots, n\} \mid P_i \in \Gamma \text{ or } P_i \text{ is infinitely near a point of } \Gamma\} = \{1, \dots, m\}.$$

Refer to 2.3 for the notation. If $0 \le j \le k \le n$ and $D \subset Y_j$ is a curve or a union of curves, we write $\widetilde{D}^{Y_k}$ for the strict transform of D on Y_k. Let J be the set of $j \in \{1, \dots, m\}$ such that $\widetilde{E}_j^{Y_n} \cap X = \varnothing$ (recall that X is an open subset of Y_n), and define

$$W = Y_m \setminus \left(\widetilde{\Gamma}^{Y_m} \cup \bigcup_{j \in J} \widetilde{E}_j^{Y_m} \right). \tag{14}$$

Then W is a nonsingular surface and f factors as $X \xrightarrow{g} W \xrightarrow{h} Y$, where g, h are birational morphisms, $\Gamma_h = \Gamma$ and $\mathrm{Cont}(h) = \{\widetilde{E}_i^{Y_m} \cap W \mid i \in \{1, \dots, m\} \setminus J\}$; thus

$$q(h) = \#\Gamma \quad \text{and} \quad c(h) = \#f^{-1}(\Gamma). \tag{15}$$

Let $\Gamma' = C_1 \cup \dots \cup C_s$, where $C_1, \dots, C_s \subset Y$ are the missing curves of f not included in Γ; then $\Gamma_f = \Gamma \cup \Gamma'$ and $\Gamma \cap \Gamma' = \varnothing$. Moreover,

$$\mathrm{Miss}(g) = \{\widetilde{C}_i^{Y_m} \cap W \mid i = 1, \dots, s\}. \tag{16}$$

Indeed, consider $C \in \mathrm{Miss}(g)$. If $h(C)$ is a point, then $C = \widetilde{E}_j^{Y_m} \cap W$ for some $j \in \{1, \dots, m\}$, and, in fact, $\widetilde{E}_j^{Y_n} \cap X = \varnothing$ (so $j \in J$) otherwise C would not be a missing curve of g; then (14) implies that $\widetilde{E}_j^{Y_m} \cap W = \varnothing$, which is absurd. So $h(C)$ is a dense subset of a curve $C_* \subset Y$. Now $C_* \in \mathrm{Miss}(f)$ by Lemma 2.12, and (14) implies that $C \nsubseteq \widetilde{\Gamma}^{Y_m}$, hence $C_* \nsubseteq \Gamma$; so $C_* \subseteq \Gamma'$ and consequently $C = \widetilde{C}_i^{Y_m} \cap W$ for some $i \in \{1, \dots, s\}$. This proves "$\subseteq$" in (16), and "$\supseteq$" easily follows from Lemma 2.12.

From (16), we deduce that $\Gamma_g \cap \mathrm{exc}(h) \subseteq h^{-1}(\Gamma') \cap h^{-1}(\Gamma)$, so

$$\Gamma_g \cap \mathrm{exc}(h) = \varnothing. \tag{17}$$

To prove (a), it only remains to show that W is affine. Since X is affine, Lemma 2.8 implies that $W \setminus \Gamma_g$ is affine; as (by (17)) $\mathrm{exc}(h) \subset W \setminus \Gamma_g$,

(18) no contracting curve of h is a complete curve.

Embed Y_0 in a nonsingular projective surface $\overline{Y}_0$ and enlarge diagram (2) as follows:

$$\begin{tikzcd}
\overline{Y}_n \arrow[r, "\overline{\pi}_n"] & \cdots \arrow[r, "\overline{\pi}_{m+1}"] & \overline{Y}_m \arrow[r, "\overline{\pi}_m"] & \cdots \arrow[r, "\overline{\pi}_1"] & \overline{Y}_0 \\
Y_n \arrow[u, hook] \arrow[r, "\pi_n"] & \cdots \arrow[r, "\pi_{m+1}"] & Y_m \arrow[u, hook] \arrow[r, "\pi_m"] & \cdots \arrow[r, "\pi_1"] & Y_0 \arrow[u, hook] \\
X \arrow[u, hook] \arrow[rr, "g"] & & W \arrow[u, hook] \arrow[rr, "h"] & & Y \arrow[u, equal]
\end{tikzcd}$$

Let $D_i = \overline{Y}_i \setminus Y_i$ $(0 \le i \le n)$. Since $Y = Y_0$ is affine, D_0 is a nonempty connected union of curves and each irreducible component of $\overline{\Gamma}$ meets D_0 (where $\overline{\Gamma}$ denotes the closure of Γ in $\overline{Y}_0$). It follows that D_m is a nonempty connected union of curves and each irreducible component of $\overline{\widetilde{\Gamma}^{Y_m}}$ meets D_m. Let us argue that

(19) $$W \text{ is connected at infinity.}$$

Suppose that (19) is false; then $\overline{Y}_m \setminus W$ is not connected, so some connected component $\mathcal{C}$ of $\overline{Y}_m \setminus W$ is disjoint from D_m. Then $\mathcal{C}$ does not contain any irreducible component of $\widetilde{\Gamma}^{Y_m}$. By (14), it follows that $\mathcal{C} \subseteq \bigcup_{j \in J} \widetilde{E}_j^{Y_m}$.

We have $\overline{Y}_m \setminus (W \setminus \Gamma_g) = \overline{\widetilde{C}_1^{Y_m}} \cup \cdots \cup \overline{\widetilde{C}_s^{Y_m}} \cup (\overline{Y}_m \setminus W)$ by (16); since $W \setminus \Gamma_g$ is affine,

$$\overline{\widetilde{C}_1^{Y_m}} \cup \cdots \cup \overline{\widetilde{C}_s^{Y_m}} \cup (\overline{Y}_m \setminus W) \text{ is connected.}$$

As $\overline{Y}_m \setminus W$ is not connected and $\mathcal{C}$ is a connected component of it, some $\overline{\widetilde{C}_i^{Y_m}}$ must meet $\mathcal{C}$. So there exist $i \in \{1, \ldots, s\}$ and $j \in J$ such that $\overline{\widetilde{C}_i^{Y_m}} \cap \widetilde{E}_j^{Y_m} \neq \varnothing$. As $C_i \subseteq \Gamma'$, this implies that $P_j \in \Gamma'$ or P_j is i.n. a point of Γ'; since $j \le m$, we also have $P_j \in \Gamma$ or P_j is i.n. a point of Γ; so $\Gamma \cap \Gamma' \neq \varnothing$, a contradiction. Thus (19) is true.

In view of (18), (19) and the fact that Y is affine, applying [Daigle 1991a, 2.2] to $h : W \to Y$ shows that W is affine and concludes the proof of (a).

(b) Assume that X, Y are factorial and have trivial units; then [ibid., 3.4] gives $q(h) \le c(h)$, so $\#\Gamma \le \#f^{-1}(\Gamma)$ by (15). Since Γ' also satisfies (13), it follows that $\#\Gamma' \le \#f^{-1}(\Gamma')$.

By Lemma 2.8 we have $\mathrm{cent}(f) \subseteq \Gamma_f = \Gamma \cup \Gamma'$, so $f^{-1}(\Gamma) \cup f^{-1}(\Gamma')$ is exactly the union of all contracting curves of f; as $f^{-1}(\Gamma) \cap f^{-1}(\Gamma') = \varnothing$, we get $\#f^{-1}(\Gamma) + \#f^{-1}(\Gamma') = c(f)$. We have $c(f) = q(f)$ by [ibid., 2.9], and it is clear that $q(f) = \#\Gamma + \#\Gamma'$, so

$$\#\Gamma \le \#f^{-1}(\Gamma), \quad \#\Gamma' \le \#f^{-1}(\Gamma') \quad \text{and} \quad \#\Gamma + \#\Gamma' = \#f^{-1}(\Gamma) + \#f^{-1}(\Gamma');$$

consequently,

(20) $$\#\Gamma = \#f^{-1}(\Gamma),$$

(21) $$q(h) = c(h),$$

where (21) follows from (20) and (15). By (21), (19) and [ibid., 3.4], we get that W is factorial and has trivial units, which proves (b).

If $X = \mathbb{A}^2$ and Y is factorial then, by (b), W may be chosen to be factorial; then [ibid., 4.2] implies that W and Y are isomorphic to $\mathbb{A}^2$, which proves (c) and completes the proof of the proposition. □

2.19. Definition. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. Consider a diagram (2) satisfying $n = n(f)$ and with notation as in 2.3 (for each $i \in \{1, \dots, n\}$, let $\pi_i : Y_i \to Y_{i-1}$ be the blowing-up of Y_{i-1} at the point $P_i \in Y_{i-1}$).

(a) Let C be a missing curve of f. For each $i = 0, 1, \dots, n$, let $C^{Y_i} \subset Y_i$ denote the strict transform of C on Y_i ($C^{Y_0} = C$). Then we define the natural number

$$n(f, C) = \text{cardinality of the set } \{ i \mid 1 \le i \le n,\ P_i \in C^{Y_{i-1}} \}$$

and note that $n(f, C)$ depends only on (f, C), i.e., is independent of the choice of diagram (2). To indicate that $n(f, C) = k$, we say that “C is blown-up k times”.

(b) For $i = 1, \dots, n$, let $\bar{P}_i \in Y_0$ be the image of P_i by $\pi_1 \circ \cdots \circ \pi_{i-1} : Y_{i-1} \to Y_0$. For each $P \in Y$, define the natural number

$$n(f, P) = \text{cardinality of the set } \{ i \mid 1 \le i \le n,\ \bar{P}_i = P \}$$

and note that $n(f, P)$ depends only on (f, P), i.e., is independent of the choice of the diagram (2) used for defining it.

2.20. Remarks. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces.

(a) Let $C \in \operatorname{Miss}(f)$. Then $n(f, C) = 0 \iff C \cap f(X) = \varnothing$, and $n(f, C) = 1$ implies that there exists exactly one fundamental point of f lying on C. Note that if $X = \mathbb{A}^2 = Y$, then each missing curve contains at least one fundamental point (Lemma 2.6(d)), so each missing curve is blown-up at least once.

(b) Let $P \in Y$. Then $n(f, P) > 0 \iff P \in \operatorname{cent}(f)$, where “$\Leftarrow$” is obvious and “$\Rightarrow$” follows from (6). It is also clear that $n(f) = \sum_{P \in Y} n(f, P)$.

2.21. Lemma. *Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces, and assume that $n(g \circ f) = n(g) + n(f)$ or $X = Y = Z = \mathbb{A}^2$.*

(a) *Let $D \in \operatorname{Miss}(g)$; then $D \in \operatorname{Miss}(g \circ f)$ and $n(g \circ f, D) = n(g, D)$.*

(b) *Let $C \in \operatorname{Miss}(f) \setminus \operatorname{Cont}(g)$ and let D be the closure of $g(C)$ in Z. Then:*

- *$D \in \operatorname{Miss}(g \circ f)$ and $n(f, C) \le n(g \circ f, D)$;*
- *$n(f, C) = n(g \circ f, D) \implies C \cap \operatorname{exc}(g) = \varnothing$;*
- *if $g(C) = D$ or $C \cong \mathbb{A}^1$, then*

$$n(f, C) = n(g \circ f, D) \iff C \cap \operatorname{exc}(g) = \varnothing.$$

(c) *For each* $P \in Z$, *we have* $n(g \circ f, P) = n(g, P) + \sum_{Q \in g^{-1}(P)} n(f, Q)$.

Proof. If $X = Y = Z = \mathbb{A}^2$, then $n(g \circ f) = n(g) + n(f)$ by Lemma 2.5; so $n(g \circ f) = n(g) + n(f)$ holds in all cases. Let $m = n(f)$ and $n = n(g)$. Choose commutative diagrams (I) and (II) as follows:

$$\begin{tikzcd} Y_m \arrow[r, "\pi_m"] & \cdots \arrow[r, "\pi_1"] & Y_0 \arrow[d, equal] \\ X \arrow[u, hook] \arrow[rr, "f"'] \arrow[urr, phantom, "\text{(I)}"] & & Y \end{tikzcd} \qquad \begin{tikzcd} Z_n \arrow[r, "\rho_n"] & \cdots \arrow[r, "\rho_1"] & Z_0 \arrow[d, equal] \\ Y \arrow[u, hook] \arrow[rr, "g"'] \arrow[urr, phantom, "\text{(II)}"] & & Z \end{tikzcd}$$

and use them to build the commutative diagram

(III)
$$\begin{tikzcd} Z_{n+m} \arrow[r, "\rho_{n+m}"] & \cdots \arrow[r, "\rho_{n+2}"] & Z_{n+1} \arrow[r, "\rho_{n+1}"] & Z_n \arrow[r, "\rho_n"] & \cdots \arrow[r, "\rho_1"] & Z_0 \arrow[dd, equal] \\ Y_m \arrow[u, hook] \arrow[r, "\pi_m"] & \cdots \arrow[r, "\pi_2"] & Y_1 \arrow[u, hook] \arrow[r, "\pi_1"] & Y_0 \arrow[u, hook] \arrow[d, equal] & & \\ X \arrow[u, hook] \arrow[rrr, "f"] & & & Y \arrow[rr, "g"] & & Z \end{tikzcd}$$

In the three diagrams, "$\hookrightarrow$" are open immersions, Y_i, Z_i are nonsingular surfaces, $Y_i \xrightarrow{\pi_i} Y_{i-1}$ is the blowing-up of Y_{i-1} at a point $P_i \in Y_{i-1}$ and $Z_i \xrightarrow{\rho_i} Z_{i-1}$ is the blowing-up of Z_{i-1} at a point $Q_i \in Z_{i-1}$. Moreover, $Y_{i-1} \hookrightarrow Z_{n+i-1}$ maps P_i on Q_{n+i} (let us simply write $P_i = Q_{n+i}$). Diagrams (I) and (II) are minimal in the sense that $n(f) = m$ and $n(g) = n$; since $n(g \circ f) = n(f) + n(g) = m + n$, it follows that (III) is also minimal.

(a) Let $D \in \mathrm{Miss}(g)$; then $D \in \mathrm{Miss}(g \circ f)$ by Lemma 2.12(a). Let $D^{Z_i} \subset Z_i$ be the strict transform of $D \subset Z_0$ on Z_i. As $D^{Z_n} \subseteq Z_n \setminus Y_0$ and $\mathrm{cent}(\rho_{n+1} \circ \cdots \circ \rho_{n+m}) = \mathrm{cent}(\pi_1 \circ \cdots \circ \pi_m) \subset Y_0$, we see that

$$\{ i \mid 1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \} = \{ i \mid 1 \le i \le n,\ Q_i \in D^{Z_{i-1}} \}. \tag{22}$$

Since $n(g \circ f, D)$ (resp. $n(g, D)$) is by definition the cardinality of the set in the left-hand side (resp. right-hand side) of (22), we have $n(g \circ f, D) = n(g, D)$.

(b) Let $C \in \mathrm{Miss}(f) \setminus \mathrm{Cont}(g)$ and let D be the closure of $g(C)$ in Z. Then $D \in \mathrm{Miss}(g \circ f)$ by Lemma 2.12(a). Define $D^{Z_i} \subset Z_i$ as before; then

$$\{ i \mid n+1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \} \subseteq \{ i \mid 1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \}. \tag{23}$$

Since $n(f, C)$ (resp. $n(g \circ f, D)$) is the cardinality of the set in the left-hand side (resp. right-hand side) of (23), we have $n(f, C) \le n(g \circ f, D)$, and, moreover,

$$n(f, C) \ne n(g \circ f, D) \tag{24}$$

is equivalent to

$$\{ i \mid 1 \le i \le n,\ \ Q_i \in D^{Z_{i-1}} \} \neq \varnothing. \tag{25}$$

By the minimality of diagram (II) together with (6), (25) is equivalent to

$$D \cap \operatorname{cent}(g) \neq \varnothing. \tag{26}$$

Now

$$C \cap \operatorname{exc}(g) \neq \varnothing \tag{27}$$

implies (26) and, if $g(C) = D$, the converse is true. So we have shown that

$$n(f, C) = n(g \circ f, D) \implies C \cap \operatorname{exc}(g) = \varnothing, \tag{28}$$

and that the converse holds whenever $g(C) = D$. Finally, we observe that if $C \cong \mathbb{A}^1$, then the dominant morphism $C \xrightarrow{g} D$ is necessarily finite, hence surjective, so the converse of (28) is true whenever $C \cong \mathbb{A}^1$. This proves (b).

(c) Define $\bar{Q}_i = (\rho_1 \circ \cdots \circ \rho_{i-1})(Q_i) \in Z_0$ for all $i = 1, \ldots, m+n$ and observe that the trivial equality (for any $P \in Z$)

$$\bigl|\{ i \mid \bar{Q}_i = P \}\bigr| = \bigl|\{ i \mid i \le n \ \text{ and } \ \bar{Q}_i = P \}\bigr| + \bigl|\{ i \mid i > n \ \text{ and } \ \bar{Q}_i = P \}\bigr|$$

is the desired equality. □

3. Compositions of simple affine contractions

Let $\boldsymbol{k}$ be an algebraically closed field of arbitrary characteristic, and let $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$. As in the introduction, we write $\operatorname{Bir}(\mathbb{A}^2)$ for the monoid of birational endomorphisms $f : \mathbb{A}^2 \to \mathbb{A}^2$, and we declare that $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ are equivalent ($f \sim g$) if $u \circ f \circ v = g$ for some $u, v \in \operatorname{Aut}(\mathbb{A}^2)$. The equivalence class of $f \in \operatorname{Bir}(\mathbb{A}^2)$ is denoted $[f]$. Note that the conditions $f \sim f'$ and $g \sim g'$ do *not* imply that $f \circ g \sim f' \circ g'$.

The aim of this section is to describe the subsets $S_{\mathrm{w}} \supset S_{\mathrm{a}} \supset S_{\mathrm{aa}}$ of $\operatorname{Bir}(\mathbb{A}^2)$ defined by

$$\begin{aligned} S_{\mathrm{w}} &= \{ f \in \operatorname{Bir}(\mathbb{A}^2) \mid \operatorname{Miss}(f) \text{ is weakly admissible} \}, \\ S_{\mathrm{a}} &= \{ f \in \operatorname{Bir}(\mathbb{A}^2) \mid \operatorname{Miss}(f) \text{ is admissible} \}, \\ S_{\mathrm{aa}} &= \{ f \in \operatorname{Bir}(\mathbb{A}^2) \mid \text{both } \operatorname{Miss}(f) \text{ and } \operatorname{Cont}(f) \text{ are admissible} \} \end{aligned}$$

(refer to Definition 1.3, Lemma 1.4, Definition 1.6 and Proposition 1.7); note that these sets are not closed under composition of morphisms. We learn at a relatively early stage (see Proposition 3.6(c)) that each element of S_{w} is a composition of simple affine contractions (SACs are defined in the introduction and again in Definition 3.2). However, an arbitrary composition of SACs does not necessarily belong to S_{w} (resp. S_{a}, S_{aa}), so in each of the three cases one has to determine which compositions of SACs give the desired type of endomorphism. The answer is given in Theorem 3.15, which is the main result of this section.

The material of 3.1–3.5(a) can be found in [Daigle 1991a; 1991b]; everything else appears to be new.

As before, we have $\mathbb{A}^2 = \operatorname{Spec} A$, where $A = \boldsymbol{k}^{[2]}$ is fixed throughout, and by a coordinate system of $\mathbb{A}^2$ we mean a pair $(X, Y) \in A \times A$ satisfying $A = \boldsymbol{k}[X, Y]$ (see the introduction of Section 1).

3.1. Let $\mathfrak{C}$ temporarily denote the set of coordinate systems of $\mathbb{A}^2$. Given an element $\mathfrak{c} = (X, Y)$ of $\mathfrak{C}$, consider the $\boldsymbol{k}$-homomorphism $A \to A$ given by $X \mapsto X$ and $Y \mapsto XY$; this homomorphism determines a morphism $\mathbb{A}^2 \to \mathbb{A}^2$, which we denote $\alpha_{\mathfrak{c}}$; clearly, $\alpha_{\mathfrak{c}} \in \operatorname{Bir}(\mathbb{A}^2)$. Note that if $\mathfrak{c}_1, \mathfrak{c}_2 \in \mathfrak{C}$, then $\alpha_{\mathfrak{c}_1}^m \sim \alpha_{\mathfrak{c}_2}^m$ for every $m \geq 1$. So, for each $m \geq 1$, the equivalence class $[\alpha_{\mathfrak{c}}^m]$ of $\alpha_{\mathfrak{c}}^m$ is independent of the choice of $\mathfrak{c} \in \mathfrak{C}$.

3.2. Definition. A birational morphism $\mathbb{A}^2 \to \mathbb{A}^2$ is called a *simple affine contraction (SAC)* if it is equivalent to $\alpha_{\mathfrak{c}}$ for some (hence for every) coordinate system $\mathfrak{c}$ of $\mathbb{A}^2$.

Note that if f is a SAC and $\mathfrak{c} \in \mathfrak{C}$, then $f \sim \alpha_{\mathfrak{c}}$, but f^2 need not be equivalent to $\alpha_{\mathfrak{c}}^2$.

For readers who like to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$, we note that $\alpha_{\mathfrak{c}}$ corresponds to the map $\boldsymbol{k}^2 \to \boldsymbol{k}^2$, $(x, y) \mapsto (x, xy)$ and that the SACs are obtained by composing this map on both sides with automorphisms. See 1.1.

3.3. Lemma. (a) *A birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC if and only if $n(f) = 1$.*

(b) *If $f : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC, then f has one missing curve L and one fundamental point P; moreover, L is a coordinate line and $P \in L$.*

(c) *Let $L \subset \mathbb{A}^2$ be a coordinate line and $P \in L$ a point. Let $X \xrightarrow{\pi} \mathbb{A}^2$ be the blowing-up of $\mathbb{A}^2$ at P and $U \subset X$ the complement of the strict transform of L. Then $U \cong \mathbb{A}^2$ and the composition $\mathbb{A}^2 \xrightarrow{\cong} U \hookrightarrow X \xrightarrow{\pi} \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P.*

(d) *If $f, g : \mathbb{A}^2 \to \mathbb{A}^2$ are two SAC with the same missing curve and the same fundamental point then there exists an automorphism $\theta : \mathbb{A}^2 \to \mathbb{A}^2$ such that $g = f \circ \theta$:*

$$\begin{tikzcd} \mathbb{A}^2 \arrow[rr, "\cong", "\theta"'] \arrow[dr, "g"'] & & \mathbb{A}^2 \arrow[dl, "f"] \\ & \mathbb{A}^2 & \end{tikzcd}$$

(e) *Let $\mathfrak{c}$ be a coordinate system of $\mathbb{A}^2$ and $\alpha_{\mathfrak{c}} \in \operatorname{Bir}(\mathbb{A}^2)$ as in 3.1. For $f \in \operatorname{Bir}(\mathbb{A}^2)$,*

$$q(f) = 1 \iff c(f) = 1 \iff f \sim \alpha_{\mathfrak{c}}^{n(f)}.$$

Proof. Assertion (a) is [Daigle 1991a, 4.10] or [Daigle 1991b, 4.1]; assertions (b)–(d) are trivial; (e) is [Daigle 1991a, 4.11] together with the fact that $q(f) = c(f)$ by Lemma 2.6(a). □

Remark. Assertion 3.3(e) can be phrased as follows: given an integer $m \geq 1$ and a coordinate system $\mathfrak{c}$ of $\mathbb{A}^2$, the equivalence class $[\alpha_{\mathfrak{c}}^m]$ of $\alpha_{\mathfrak{c}}^m$ is

$$\begin{aligned}[\alpha_{\mathfrak{c}}^m] &= \{f \in \mathrm{Bir}(\mathbb{A}^2) \mid q(f) = 1 \text{ and } n(f) = m\} \\ &= \{f \in \mathrm{Bir}(\mathbb{A}^2) \mid c(f) = 1 \text{ and } n(f) = m\}.\end{aligned}$$

3.4. Corollary. *If $f \in \mathrm{Bir}(\mathbb{A}^2)$ has a unique missing curve C, then C is a coordinate line.*

Proof. This follows from Lemma 3.3(e). It also follows from Proposition 2.13(a) and Proposition 2.9. □

See Definition 2.19 for the definition of the phrase "L is blown-up only once".

3.5. Lemma. *Let $f \in \mathrm{Bir}(\mathbb{A}^2)$.*

(a) *Suppose that some missing curve L of f is blown-up only once. Then L is a coordinate line. Moreover, if $P \in L$ is the unique fundamental point of f which is on L and $\gamma \in \mathrm{Bir}(\mathbb{A}^2)$ is a SAC with missing curve L and fundamental point P, then f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$ with $f' \in \mathrm{Bir}(\mathbb{A}^2)$.*

(b) *Suppose that f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$ with $f', \gamma \in \mathrm{Bir}(\mathbb{A}^2)$, where γ is a SAC. Let L be the missing curve of γ. Then L is a missing curve of f which is blown-up only once.*

Proof. Part (a) is an improvement of [Daigle 1991b, 4.6]. The proof of [ibid., 4.6] shows that P is a nonsingular point of L; then [Daigle 1991a, 4.6] implies that L is a coordinate line. Choose a diagram (2) for f such that $n = n(f)$ and $P_1 = P$. Let $L^{Y_1} \subset Y_1$ denote the strict transform of L on Y_1 and define $W = Y_1 \setminus L^{Y_1} \subset Y_1$. As L is a missing curve of f and is blown-up only once, the image of $\mathbb{A}^2 \hookrightarrow Y_n \xrightarrow{\pi_2 \circ \cdots \circ \pi_n} Y_1$ is included in W; so f factors as $\mathbb{A}^2 \xrightarrow{g'} W \xrightarrow{h'} \mathbb{A}^2$, where g', h' are birational morphisms and h' is the composition $W \hookrightarrow Y_1 \xrightarrow{\pi_1} Y_0 = \mathbb{A}^2$. By Lemma 3.3(c), $W \cong \mathbb{A}^2$ and the composition $\mathbb{A}^2 \cong W \xrightarrow{h'} \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P; so f factors as $\mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2 \xrightarrow{h} \mathbb{A}^2$, where $g, h \in \mathrm{Bir}(\mathbb{A}^2)$ and h is a SAC with missing curve L and fundamental point P. By Lemma 3.3(d), $h = \gamma \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$. Then $f = \gamma \circ \theta \circ g$ and we are done.

For (b), we know that by Lemma 2.21(a), L is in $\mathrm{Miss}(f)$ and $n(f, L) = n(\gamma \circ f', L) = n(\gamma, L) = 1$. □

3.6. Proposition. *Let $f : \mathbb{A}^2 \to \mathbb{A}^2$ be a birational morphism such that*

(i) *f is not an isomorphism*;

(ii) *there exists a coordinate system of $\mathbb{A}^2$ with respect to which all missing curves of f have degree* 1.

Then there exists a missing curve of f which is blown-up only once. Moreover, *if L is such a curve and $P \in L$ is the unique fundamental point of f which is on L*, *then the following holds*:

(a) *There exists a coordinate system (X, Y) of $\mathbb{A}^2$ such that $L = \boldsymbol{Z}(X)$ and $P = (0, 0)$, and such that the union of the missing curves of f is equal to the zero set of one of the following polynomials in $\boldsymbol{k}[X, Y]$*:

   (i) *$XY^m \prod_{i=1}^{n}(X - a_i)$ for some $m \in \{0, 1\}$, $n \geq 0$ and $a_1, \ldots, a_n \in \boldsymbol{k}$*;

   (ii) *$X(X-1)^m \prod_{i=1}^{n}(Y - b_i X)$ for some $m \in \{0, 1\}$, $n \geq 0$ and $b_1, \ldots, b_n \in \boldsymbol{k}$.*

(b) *If $\gamma : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P*, *then f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$*, *where f' is a birational morphism such that* $\operatorname{Miss}(f')$ *is admissible.*

(c) *f is a composition of SACs.*

Proof. By [Daigle 1991b, 4.7], $f = h \circ g$, where $g, h \in \operatorname{Bir}(\mathbb{A}^2)$ and h is a SAC; then Lemma 3.5(b) implies that some missing curve of f is blown-up only once.

Let L be a missing curve of f which is blown-up only once, and let $P \in L$ be the unique fundamental point of f which is on L. Choose a coordinate system (X, Y) of $\mathbb{A}^2$ such that $L = \boldsymbol{Z}(X)$ and $P = (0, 0)$ and with respect to which all missing curves of f have degree 1. Define $\gamma_0 : \mathbb{A}^2 \to \mathbb{A}^2$ by $\gamma_0(x, y) = (x, xy)$. As γ_0 is a SAC with missing curve L and fundamental point P, Lemma 3.5(a) implies that f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma_0} \mathbb{A}^2$ for some $f' \in \operatorname{Bir}(\mathbb{A}^2)$. Let Γ (resp. Γ') be the union of the missing curves of f (resp. of f'). Then

$$\Gamma = \boldsymbol{Z}(X) \cup \overline{\gamma_0(\Gamma')}, \tag{29}$$

by Lemma 2.12. In particular, if C is a missing curve of f', then $\gamma_0(C)$ is included into a line of degree 1; from $\gamma_0(x, y) = (x, xy)$, it easily follows that C is either a vertical line $\boldsymbol{Z}(X - a)$ (for some $a \in \boldsymbol{k}$), a horizontal line $\boldsymbol{Z}(Y - b)$ (for some $b \in \boldsymbol{k}$) or a hyperbola $\boldsymbol{Z}(X(\alpha + \beta Y) - 1)$ (for some $\alpha, \beta \in \boldsymbol{k}$, $\beta \neq 0$), where, in fact, the last case cannot occur because C has one place at infinity by Lemma 2.6(c). So

(30) *each missing curve of f' is either a vertical line or a horizontal line.*

In particular, all missing curves of f' have degree 1. It follows from the first part of the proof that

(31) *if f' is not an isomorphism, some missing curve of f' is blown-up only once.*

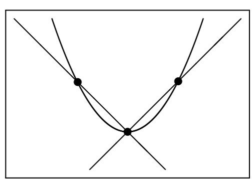
Let h (resp. v) be the number of missing curves of f' which are horizontal (resp. vertical) lines. Then $\min(h, v) \leq 1$, otherwise (by Lemma 2.6(e)) every missing curve of f' would contain at least two fundamental points of f', and hence would be blown-up at least twice, contradicting (31). Statement (30), together with $\min(h, v) \leq 1$, implies that $\mathrm{Miss}(f')$ is admissible, which proves the special case "$\gamma = \gamma_0$" of assertion (b); in view of Lemma 3.3(d), it follows that (b) is true.

If $h \leq 1$, then Γ' is the zero set of $(Y-b)^h \prod_{i=1}^{v}(X-a_i)$ for some $b \in \boldsymbol{k}$ and $a_1, \dots, a_v \in \boldsymbol{k}$. Then, by (29), Γ is the zero set of $X(Y-bX)^h \prod_{i=1}^{v}(X-a_i)$. Replacing the coordinate system (X, Y) by $(X, Y-bX)$, we see that (a-i) is satisfied.

If $v \leq 1$, then Γ' is the zero set of $(X-a)^v \prod_{i=1}^{h}(Y-b_i)$ for some $a \in \boldsymbol{k}$ and $b_1, \dots, b_h \in \boldsymbol{k}$. Then Γ is the zero set of $X(X-a)^v \prod_{i=1}^{h}(Y-b_iX)$. If $a = 0$ or $v = 0$, then Γ is the zero set of $X(X-1)^0 \prod_{i=1}^{h}(Y-b_iX)$, so (a-ii) holds; if $a \neq 0$ and $v \neq 0$, then Γ is the zero set of $X(X-a)\prod_{i=1}^{h}(Y-b_iX)$, so (a-ii) holds after replacing (X, Y) by $(a^{-1}X, Y)$. So assertion (a) is true.

To prove assertion (c), consider the factorization $f = \gamma \circ f'$ given by (b). Since $n(\gamma) = 1$ by Lemma 3.3(a), we have $n(f') = n(f) - 1$ by Lemma 2.5. Moreover, the fact that $\mathrm{Miss}(f')$ is admissible implies, by Proposition 1.7, that there exists a coordinate system of $\mathbb{A}^2$ with respect to which all missing curves of f' have degree 1. It is clear that (c) follows by induction on $n(f)$. $\square$

3.7. Remark. We stress that assumption (ii) of Proposition 3.6 is strictly stronger than "all missing curves are coordinate lines". Indeed, there exists an irreducible element $f \in \mathrm{Bir}(\mathbb{A}^2)$ with three missing curves, these being the lines $\boldsymbol{Z}(X+Y)$ and $\boldsymbol{Z}(X-Y)$ and the parabola $\boldsymbol{Z}(Y-X^2)$:

$\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 =$ 

(an example of such an f, due to Russell, appeared in [Daigle 1991a, 4.7]). Here, each missing curve is a coordinate line and hence has degree 1 with respect to a suitable choice of coordinate system. However, these three lines are not *simultaneously* rectifiable, so f does not satisfy assumption (ii) of Proposition 3.6 (it does not satisfy the conclusion either: since f is not a SAC and is irreducible, it is not a composition of SACs).

Also note that by Lemma 1.4, assumption (ii) of Proposition 3.6 is equivalent to "$\mathrm{Miss}(f)$ is weakly admissible".

3.8. Until the end of Section 3,

$$\text{we fix a coordinate system } \mathfrak{c} = (X, Y) \text{ of } \mathbb{A}^2.$$

This allows us to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$. See 1.1 for the notation $\boldsymbol{Z}(F)$ and for our convention regarding the definition of morphisms using coordinates.

3.9. In 3.9.1–3.9.3 below, we define three submonoids of $\mathrm{Bir}(\mathbb{A}^2)$, denoted $\mathcal{H}_\mathfrak{c}$, $\mathcal{G}_\mathfrak{c}$ and $\mathcal{V}_\mathfrak{c}$, respectively. The subscript $\mathfrak{c}$ reminds us that these sets depend on the choice of $\mathfrak{c}$ made in 3.8. Since $\mathfrak{c}$ is fixed until the end of this section, there is no harm in omitting it and writing simply $\mathcal{H}$, $\mathcal{G}$ and $\mathcal{V}$. It is clear from the definitions below that these three monoids are included in the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.

3.9.1. Given $m \in \mathbb{N}$ and $p \in \boldsymbol{k}[Y]$ such that[3] $\deg p < m$, define $h_{m,p} \in \mathrm{Bir}(\mathbb{A}^2)$ by $h_{m,p}(x, y) = (xy^m + p(y), y)$. Observe that $h_{m,p}$ is equivalent to γ^m, where γ is the SAC given by $(x, y) \mapsto (xy, y)$; consequently, $n(h_{m,p}) = n(\gamma^m) = mn(\gamma)$, i.e.,

$$n(h_{m,p}) = m.$$

Define $\mathcal{H} = \mathcal{H}_\mathfrak{c} = \{h_{m,p} \mid m \in \mathbb{N},\ p \in \boldsymbol{k}[Y] \text{ and } \deg p < m\}$. It is easily verified that $\mathcal{H}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.9.2. Let $\boldsymbol{M}$ be the multiplicative monoid whose elements are the 2×2 matrices $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right)$ with $i, j, k, \ell \in \mathbb{N}$ and $i\ell - jk = \pm 1$. It is easily verified that $\boldsymbol{M}$ is generated by $G_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $G_2 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Given $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$, define the morphism $\gamma_M : \mathbb{A}^2 \to \mathbb{A}^2$ by $(x, y) \mapsto (x^i y^j, x^k y^\ell)$. Note that $\gamma_{M_1} \circ \gamma_{M_2} = \gamma_{M_1 M_2}$ for all $M_1, M_2 \in \boldsymbol{M}$, so the set

$$\mathcal{G} = \mathcal{G}_\mathfrak{c} = \{\gamma_M \mid M \in \boldsymbol{M}\}$$

is a monoid (under composition) generated by $\{\gamma_{G_1}, \gamma_{G_2}\}$. As γ_{G_1} is a SAC and γ_{G_2} is an automorphism, it follows that $\mathcal{G}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.9.3. Given a polynomial $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$, define $v_\varphi \in \mathrm{Bir}(\mathbb{A}^2)$ by $v_\varphi(x, y) = (x, \varphi(x)y)$. Then let

$$\mathcal{V} = \mathcal{V}_\mathfrak{c} = \{v_\varphi \mid \varphi \in \boldsymbol{k}[X] \setminus \{0\}\}.$$

Note that $v_\varphi \circ v_\psi = v_{\varphi \cdot \psi} = v_\psi \circ v_\varphi$ for any $\varphi, \psi \in \boldsymbol{k}[X] \setminus \{0\}$, so $\mathcal{V}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.10. Lemma. *For a birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$, the following are equivalent*:

(a) *The union of the missing curves of f is included in $\boldsymbol{Z}(Y)$.*

(b) *There exists $(h, \theta) \in \mathcal{H} \times \mathrm{Aut}(\mathbb{A}^2)$ such that $f = h \circ \theta$:*

[3] We adopt the convention that the zero polynomial has degree $-\infty$; consequently, the condition $\deg p < 0$ is equivalent to p being the zero polynomial (so $h_{0,p} = h_{0,0}$ is the identity map).

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "\theta"'] \arrow[rr, bend left, "f"] & \mathbb{A}^2 \arrow[r, "h"'] & \mathbb{A}^2 \end{tikzcd}$$

Moreover, if these conditions are satisfied, then the pair (h, θ) *in* (b) *is unique.*

Proof. We leave it to the reader to verify that (b) implies (a) and that (h, θ) in (b) is unique. By induction on $n(f)$, we show that (a) implies (b).

If $n(f) = 0$, then (b) holds with $\theta = f$ and $h = h_{0,0}$.

If $n(f) > 0$, then f is not an isomorphism, and hence has at least one missing curve, so $\boldsymbol{Z}(Y)$ is the unique missing curve of f; by Proposition 3.6, this missing curve is blown-up only once. This missing curve must contain a fundamental point $(c, 0)$ of f; as $h_{1,c} \in \mathcal{H}$ is a SAC with missing curve $\boldsymbol{Z}(Y)$ and fundamental point $(c, 0)$, Proposition 3.6 implies that $f = h_{1,c} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$. It is immediate that $h_{1,c}^{-1}(\Gamma) = \Gamma$, where $\Gamma = \boldsymbol{Z}(Y)$ is the missing curve of f; so Lemma 2.12 implies that the union of the missing curves of f' is included in $\boldsymbol{Z}(Y)$. As $n(f') = n(f) - 1$, we may assume by induction that $f' = h' \circ \theta$ for some $h' \in \mathcal{H}$ and $\theta \in \mathrm{Aut}(\mathbb{A}^2)$. Then $f = h_{1,c} \circ h' \circ \theta$ is the desired factorization, where we note that $h_{1,c} \circ h' \in \mathcal{H}$. ☐

3.11. Lemma. *For a birational morphism* $f : \mathbb{A}^2 \to \mathbb{A}^2$*, the following are equivalent*:

(a) *The union of the missing curves of* f *is included in* $\boldsymbol{Z}(XY)$.

(b) *There exists* $(M, h, \theta) \in \boldsymbol{M} \times \mathcal{H} \times \mathrm{Aut}(\mathbb{A}^2)$ *such that* $f = \gamma_M \circ h \circ \theta$:

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "\theta"'] \arrow[rrr, bend left, "f"] & \mathbb{A}^2 \arrow[r, "h"'] & \mathbb{A}^2 \arrow[r, "\gamma_M"'] & \mathbb{A}^2 \end{tikzcd}$$

Proof. It is easily verified that (b) implies (a). We prove that (a) implies (b) by induction on $n(f)$. Assume that f satisfies (a).

If $n(f) = 0$, then f is an isomorphism, so (b) holds with $\theta = f$, $h = h_{0,0}$ and $M = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Let $n > 0$ and assume that the result is true whenever $n(f) < n$. Now consider f satisfying (a) and such that $n(f) = n$.

If $q(f) = 1$, then the missing curve Γ of f is $\boldsymbol{Z}(X)$ or $\boldsymbol{Z}(Y)$. Define $M = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (resp. $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$) if $\Gamma = \boldsymbol{Z}(Y)$ (resp. $\Gamma = \boldsymbol{Z}(X)$). Then $\gamma_M \circ f$ has a unique missing curve, and this curve is $\boldsymbol{Z}(Y)$. Applying Lemma 3.10 to $\gamma_M \circ f$ gives $\gamma_M \circ f = h \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ and $h \in \mathcal{H}$. Noting that $\gamma_M \circ \gamma_M$ is the identity, we get $f = \gamma_M \circ h \circ \theta$.

From now on, assume that $q(f) = 2$. Let Γ be the union of the missing curves of f; i.e., $\Gamma = \boldsymbol{Z}(XY)$. By Proposition 3.6, some element L of $\mathrm{Miss}(f) = \{\boldsymbol{Z}(X), \boldsymbol{Z}(Y)\}$ is blown-up only once. As $(0, 0)$ is a common point of the two missing curves, it must be a fundamental point of f. For a suitable choice of

$M_1 \in \{\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)\}$, γ_{M_1} is a SAC with missing curve L and fundamental point $(0,0)$. Then Proposition 3.6 implies that $f = \gamma_{M_1} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$. By Lemma 2.12, the union of the missing curves of f' is included in $\gamma_{M_1}^{-1}(\Gamma) = \Gamma$, so f' satisfies (a). As $n(f') = n(f) - 1$, the inductive hypothesis implies that $f' = \gamma_{M_2} \circ h \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$, $h \in \mathcal{H}$ and $M_2 \in \boldsymbol{M}$. So $f = \gamma_{M_1} \circ \gamma_{M_2} \circ h \circ \theta$, and since $\gamma_{M_1} \circ \gamma_{M_2} = \gamma_{M_1 M_2}$, we are done. □

3.12. Let $\Delta = \Delta_{\mathfrak{c}}$ be the subgroup of $\mathrm{Aut}(\mathbb{A}^2)$ whose elements are of the form $\delta(x, y) = (x, y + q(x))$ with $q \in \boldsymbol{k}[X]$.

3.13. Lemma. *Let $\Gamma = \boldsymbol{Z}\big(\prod_{i=1}^{s}(X - c_i)\big)$, where $c_1, \dots, c_s$ $(s \geq 0)$ are distinct elements of $\boldsymbol{k}$. For a birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$, the following are equivalent*:

(a) *The union of the missing curves of f is included in Γ.*

(b) *There exists a commutative diagram*

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "f"] \arrow[d, "\theta"'] & \mathbb{A}^2 \arrow[d, "\delta"] \\ \mathbb{A}^2 \arrow[r, "v_\varphi"'] & \mathbb{A}^2 \end{tikzcd}$$

where $\delta \in \Delta$, $\theta \in \mathrm{Aut}(\mathbb{A}^2)$, $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$ and where the set of roots of φ is included in $\{c_1, \dots, c_s\}$.

Proof. That (b) implies (a) is left to the reader. Suppose that f satisfies (a). We prove (b) by induction on $n(f)$.

If $n(f) = 0$ then f is an isomorphism, so (b) holds with $\theta = f$, $\varphi = 1$ and $\delta = \mathrm{id}$.

Let $n > 0$ be such that the result is true whenever $n(f) < n$. Consider f satisfying (a) and such that $n(f) = n$. Then f is not an isomorphism, and hence has at least one missing curve (so $s > 0$). By Proposition 3.6, one of the missing curves (say $L = \boldsymbol{Z}(X - c_j)$) of f is blown-up only once. We know that L contains a fundamental point (c_j, d) of f; let $\delta_1 \in \Delta$ be defined by $\delta_1(x, y) = (x, y - d)$ and let $f_1 = \delta_1 \circ f$. Since L is a missing curve of f which is blown-up only once and $(c_j, d) \in L$ is a fundamental point of f, it follows that $\delta_1(L) = L$ is a missing curve of f_1 which is blown-up only once and that $\delta_1(c_j, d) = (c_j, 0) \in L$ is a fundamental point of f_1. As $v_{(X-c_j)}$ is a SAC with missing curve L and fundamental point $(c_j, 0)$, Proposition 3.6 implies that f_1 factors through $v_{(X-c_j)}$. Thus $\delta_1 \circ f = v_{(X-c_j)} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$:

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "f"] \arrow[d, "f'"'] & \mathbb{A}^2 \arrow[d, "\delta_1"] \\ \mathbb{A}^2 \arrow[r, "v_{(X-c_j)}"'] & \mathbb{A}^2 \end{tikzcd}$$

Since δ_1 maps each vertical line onto itself, the union of all missing curves of f_1 is Γ; so, by Lemma 2.12, the union of the missing curves of f' is included in $v_{(X-c_j)}^{-1}(\Gamma)=\Gamma$, so f' satisfies (a). As $n(f')=n(f)-1$, the inductive hypothesis implies that there exists a commutative diagram (ignore the dotted arrows for now)

(32)
$$\begin{tikzcd}
& \mathbb{A}^2 \arrow[r, "f"] \arrow[dl, "\theta"'] \arrow[d, "f'"'] & \mathbb{A}^2 \arrow[d, "\delta_1"] \\
\mathbb{A}^2 \arrow[d, "v_{\varphi'}"'] & \mathbb{A}^2 \arrow[r, "v_{(X-c_j)}"'] \arrow[dl, "\delta_2"] & \mathbb{A}^2 \arrow[dl, dotted, "\delta_3"] \\
\mathbb{A}^2 \arrow[r, dotted, "v_{(X-c_j)}"'] & \mathbb{A}^2 &
\end{tikzcd}$$

with $\varphi'\in \boldsymbol{k}[X]\setminus\{0\}$ (and all roots of φ' are in $\{c_1,\dots,c_s\}$), $\theta\in\operatorname{Aut}(\mathbb{A}^2)$ and $\delta_2\in\Delta$ defined by $\delta_2(x,y)=(x,y+q(x))$ (for some $q\in\boldsymbol{k}[X]$). Now if we define $\delta_3\in\Delta$ by $\delta_3(x,y)=(x,y+(x-c_j)q(x))$, then

$$\delta_3\circ v_{(X-c_j)}=v_{(X-c_j)}\circ\delta_2.$$

So diagram (32), including the dotted arrows, is commutative. Let $\delta=\delta_3\circ\delta_1\in\Delta$ and $\varphi=(X-c_j)\varphi'(X)$ (so $v_{(X-c_j)}\circ v_{\varphi'}=v_\varphi$); then δ,θ,v_φ give the commutative diagram displayed in the statement of assertion (b). $\square$

3.14. Lemma. *Let $\Gamma=\boldsymbol{Z}\big(Y\prod_{i=1}^{s}(X-c_i)\big)$, where $s\geq 1$ and $c_1,\dots,c_s$ are distinct elements of $\boldsymbol{k}$. Let $f:\mathbb{A}^2\to\mathbb{A}^2$ be a birational morphism such that*

(33) *the union of the missing curves of f is equal to Γ.*

Then there exists a commutative diagram

$$\begin{tikzcd}
\mathbb{A}^2 \arrow[rrr, "f"] \arrow[d, "\theta"'] & & & \mathbb{A}^2 \arrow[d, "T"] \\
\mathbb{A}^2 \arrow[r, "h"'] & \mathbb{A}^2 \arrow[r, "\gamma_M"'] & \mathbb{A}^2 \arrow[r, "v_\varphi"'] & \mathbb{A}^2
\end{tikzcd}$$

where $T\in\operatorname{Aut}(\mathbb{A}^2)$ is of the form $T(x,y)=(x-c,y)$ with $c\in\boldsymbol{k}$, θ is an arbitrary element of $\operatorname{Aut}(\mathbb{A}^2)$ and $(\varphi,M,h)\in(\boldsymbol{k}[X]\setminus\{0\})\times\boldsymbol{M}\times\mathfrak{H}$.

Proof. We first settle the case $s=1$. Define $T\in\operatorname{Aut}(\mathbb{A}^2)$ by $T(x,y)=(x-c_1,y)$. Then the union of the missing curves of $T\circ f$ is $\boldsymbol{Z}(XY)$, so Lemma 3.11 implies that there exists $(M,h,\theta)\in\boldsymbol{M}\times\mathfrak{H}\times\operatorname{Aut}(\mathbb{A}^2)$ such that $T\circ f=\gamma_M\circ h\circ\theta=v_\varphi\circ\gamma_M\circ h\circ\theta$ with $\varphi=1$. Thus the result is true when $s=1$.

We proceed by induction on $n(f)$. For f satisfying (33), we have $q(f)=s+1\geq 2$, so the least possible value for $n(f)$ is 2. If $n(f)=2$, then $q(f)\leq n(f)=2$, so $s=1$ and the result is true in this case.

Let $n>2$ be such that the result is true whenever $n(f)<n$. Consider f satisfying (33) and such that $n(f)=n$.

By the first paragraph, we may assume that $s>1$. By Proposition 3.6, one of the missing curves (say L) of f is blown-up only once; we choose such an L. By Lemma 2.6(e), the points $(c_i,0)$, $1\leq i\leq s$, are fundamental points of f; so $\boldsymbol{Z}(Y)$ is blown-up at least $s\geq 2$ times and hence $L=\boldsymbol{Z}(X-c_j)$ for some $j\in\{1,\dots,s\}$. As $v_{(X-c_j)}$ is a SAC with missing curve L and fundamental point $(c_j,0)$, Proposition 3.6 implies that $f=v_{(X-c_j)}\circ f'$ for some birational morphism $f':\mathbb{A}^2\to\mathbb{A}^2$. Let $\Gamma'=\bigcup_{C\in\mathrm{Miss}(f')}C$. By Lemma 2.12, $\Gamma'\subseteq v_{(X-c_j)}^{-1}(\Gamma)=\Gamma$; in fact, it is easy to see (again by Lemma 2.12) that

$$\boldsymbol{Z}\Big(Y\prod_{i\in I}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big),$$

where $I=\{1,\dots,s\}\setminus\{j\}$, so f' satisfies the hypothesis of the lemma. As $n(f')=n(f)-1$, the inductive hypothesis implies that $T\circ f'=v_\psi\circ\gamma_M\circ h\circ\theta$ for some $(\psi,M,h)\in(\boldsymbol{k}[X]\setminus\{0\})\times\boldsymbol{M}\times\mathcal{H}$ and $\theta,T\in\mathrm{Aut}(\mathbb{A}^2)$, where T is of the form $T(x,y)=(x-c,y)$ for some $c\in\boldsymbol{k}$. Noting that $T\circ v_{(X-c_j)}=v_{(X+c-c_j)}\circ T$, we get

$$\begin{aligned}T\circ f&=T\circ v_{(X-c_j)}\circ f'=v_{(X+c-c_j)}\circ T\circ f'=v_{(X+c-c_j)}\circ v_\psi\circ\gamma_M\circ h\circ\theta\\&=v_\varphi\circ\gamma_M\circ h\circ\theta,\end{aligned}$$

where $\varphi(X)=(X+c-c_j)\psi(X)\in\boldsymbol{k}[X]\setminus\{0\}$, as desired. $\square$

Before stating the main theorem of this section, let us recall the assumptions under which it is valid. Our base field $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic, and $\mathbb{A}^2$ is the affine plane over $\boldsymbol{k}$. We fix a coordinate system $\mathfrak{c}=(X,Y)$ of $\mathbb{A}^2$; this allows us to use coordinates for defining morphisms $\mathbb{A}^2\to\mathbb{A}^2$ (see Section 1). The choice of $\mathfrak{c}$ also determines the submonoids $\mathcal{V}=\mathcal{V}_\mathfrak{c}$, $\mathcal{G}=\mathcal{G}_\mathfrak{c}$ and $\mathcal{H}=\mathcal{H}_\mathfrak{c}$ of $\mathrm{Bir}(\mathbb{A}^2)$ (see 3.9). Then we have the following result:

3.15. Theorem. *Let $f:\mathbb{A}^2\to\mathbb{A}^2$ be a birational morphism.*

(a) *The following conditions are equivalent*:

  (i) $\mathrm{Miss}(f)$ *is weakly admissible.*

  (ii) *f is equivalent to one of the following elements of* $\mathrm{Bir}(\mathbb{A}^2)$:

  - $\alpha_i^m\circ v\circ\gamma\circ h$ *for some* $(v,\gamma,h)\in\mathcal{V}\times\mathcal{G}\times\mathcal{H}$, $m\in\{0,1\}$ *and* $i\in\{1,2\}$, *where* $\alpha_1,\alpha_2\in\mathrm{Bir}(\mathbb{A}^2)$ *are the SACs defined by* $\alpha_1(x,y)=(xy,y)$ *and* $\alpha_2(x,y)=(x(1-y),1-y)$;

- *the morphism* $(x, y) \mapsto \big(x(p(x)y + q(x)), p(x)y + q(x)\big)$, *for some* $p, q \in \boldsymbol{k}[X]$ *with* $p \neq 0$.

(b) *The following conditions are equivalent*:

(i) $\mathrm{Miss}(f)$ *is admissible.*

(ii) f *is equivalent to* $v \circ \gamma \circ h$ *for some* $(v, \gamma, h) \in \mathcal{V} \times \mathcal{G} \times \mathcal{H}$.

(c) *The following conditions are equivalent*:

(i) *Both* $\mathrm{Miss}(f)$ *and* $\mathrm{Cont}(f)$ *are admissible.*

(ii) f *is equivalent to an element of* $\mathcal{V} \cup \mathcal{G}$.

Proof. For each of (a), (b) and (c), we show that (i) implies (ii) and leave the converse to the reader. We begin with (b).

Suppose that f satisfies (b-i). Let $\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C$. By Proposition 1.7, there exists $\omega \in \mathrm{Aut}(\mathbb{A}^2)$ such that $\omega(\Gamma) = \boldsymbol{Z}\big(Y^d \prod_{i=1}^{s}(X - c_i)\big)$, where

$$d \in \{0, 1\},\ s \geq 0 \text{ and } c_1, \ldots, c_s \text{ are distinct elements of } \boldsymbol{k}.$$

Note that the union of the missing curves of $f_1 = \omega \circ f$ equals $\boldsymbol{Z}\big(Y^d \prod_{i=1}^{s}(X - c_i)\big)$; since $f_1 \sim f$, it is enough to prove that f_1 is equivalent to $v \circ \gamma \circ h$ for some $(v, \gamma, h) \in \mathcal{V} \times \mathcal{G} \times \mathcal{H}$. So we may as well replace f by f_1 throughout; so from now on we assume that

$$\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C = \boldsymbol{Z}\bigg(Y^d \prod_{i=1}^{s}(X - c_i)\bigg).$$

If $d = 0$ (resp. $s = 0$), then the desired conclusion follows from Lemma 3.13 (resp. from Lemma 3.10). So we may assume that $d = 1$ and $s \geq 1$. Then Lemma 3.14 gives the desired conclusion; i.e., we showed that (b-i) implies (b-ii).

Suppose that (a-i) holds. Let $\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C$. By Lemma 1.4, f satisfies the hypotheses of Proposition 3.6. To prove (a-ii), we may assume that $\mathrm{Miss}(f)$ is not admissible (otherwise (a-ii) follows from (b)). Then Proposition 3.6 implies that there exists $\omega \in \mathrm{Aut}(\mathbb{A}^2)$ such that $\omega(\Gamma) = \boldsymbol{Z}(F)$ with

$$F = Y \prod_{i=1}^{s}(X - c_i Y) \quad \text{or} \quad F = Y(Y-1) \prod_{i=1}^{s}(X - c_i Y), \tag{34}$$

where $s \geq 2$ and $c_1, \ldots, c_s \in \boldsymbol{k}$ are distinct. We know, also by Proposition 3.6, that some missing curve of f (say $C_0 \in \mathrm{Miss}(f)$) is blown-up only once. In the second case of (34), $\omega(C_0)$ is necessarily equal to $\boldsymbol{Z}(Y)$; in the first case, we may choose ω in such a way that $\omega(C_0) = \boldsymbol{Z}(Y)$.

It is clear that we may replace f by $\omega \circ f$ throughout. Then we have $\Gamma = \boldsymbol{Z}(F)$, $\boldsymbol{Z}(Y)$ is a missing curve of f which is blown-up only once and $(0, 0)$ is the unique fundamental point of f which lies on $\boldsymbol{Z}(Y)$. If F is as in the first (resp. the second) case of (34), let $\alpha = \alpha_1$ (resp. $\alpha = \alpha_2$), where $\alpha_1, \alpha_2 \in \mathrm{Bir}(\mathbb{A}^2)$ are defined in the statement of (a-ii); then α is a SAC with missing curve $\boldsymbol{Z}(Y)$ and fundamental point $(0,0)$. By Proposition 3.6, it follows that $f=\alpha\circ f'$ for some $f'\in \operatorname{Bir}(\mathbb{A}^2)$. Let Γ' be the union of the missing curves of f'. Using Lemma 2.12, we find

$$\text{in the first case of (34),}\quad \boldsymbol{Z}\Big(\prod_{i=1}^{s}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big);$$

$$\text{in the second case of (34),}\quad \boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y(Y-1)\prod_{i=1}^{s}(X-c_i)\Big).$$

In particular, f' satisfies the hypothesis of Proposition 3.6; by that result, some missing curve of f' is blown-up only once, and thus Γ' cannot be equal to $\boldsymbol{Z}\big(Y(Y-1)\prod_{i=1}^{s}(X-c_i)\big)$. It follows that $\Gamma'=\boldsymbol{Z}(G)$, where

$$G=\prod_{i=1}^{s}(X-c_i)\quad\text{or}\quad G=Y\prod_{i=1}^{s}(X-c_i). \tag{35}$$

First consider the case $G=\prod_{i=1}^{s}(X-c_i)$; then $\alpha=\alpha_1$ because the first case of (35) can only happen in the first case of (34). By Lemma 3.13, there is a commutative diagram

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "f'"] \arrow[d, "\theta"'] & \mathbb{A}^2 \arrow[d, "\delta"] \\ \mathbb{A}^2 \arrow[r, "v"'] & \mathbb{A}^2 \end{tikzcd}$$

where $v\in\mathcal{V}$, $\delta,\theta\in\operatorname{Aut}(\mathbb{A}^2)$ and δ is of the form $\delta(x,y)=(x,y-q(x))$ for some $q\in\boldsymbol{k}[X]$. Then $f=\alpha_1\circ f'=\alpha_1\circ\delta^{-1}\circ v\circ\theta\sim\alpha_1\circ\delta^{-1}\circ v$. Let $p\in\boldsymbol{k}[X]\setminus\{0\}$ be such that $v(x,y)=(x,p(x)y)$; then

$$(\alpha_1\circ\delta^{-1}\circ v)(x,y)=\big(x(p(x)y+q(x)),\,p(x)y+q(x)\big),$$

which shows that (a-ii) holds in this case.

Consider the second case, $G=Y\prod_{i=1}^{s}(X-c_i)$. Here, α may be either one of α_1,α_2. By Lemma 3.14, there is a commutative diagram

$$\begin{tikzcd} \mathbb{A}^2 \arrow[r, "f'"] \arrow[d, "\theta"'] & \mathbb{A}^2 \arrow[d, "T"] \\ \mathbb{A}^2 \arrow[r, "v\circ\gamma\circ h"'] & \mathbb{A}^2 \end{tikzcd}$$

where $(v,\gamma,h)\in\mathcal{V}\times\mathcal{G}\times\mathcal{H}$, $\theta\in\operatorname{Aut}(\mathbb{A}^2)$, and $T\in\operatorname{Aut}(\mathbb{A}^2)$ is of the form $T(x,y)=(x-c,y)$, with $c\in\boldsymbol{k}$. Now $\alpha\circ T^{-1}=\nu\circ\alpha$, where $\nu\in\operatorname{Aut}(\mathbb{A}^2)$ is given by $\nu(x, y) = (x + cy, y)$. Thus

$$f = \alpha \circ f' = \alpha \circ T^{-1} \circ v \circ \gamma \circ h \circ \theta = \nu \circ \alpha \circ v \circ \gamma \circ h \circ \theta \sim \alpha \circ v \circ \gamma \circ h,$$

showing that (a-ii) holds in this case as well. So (a-i) implies (a-ii).

Let $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$, $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$ and $h_{m,p}(x, y) = (xy^m + p(y), y) \in \mathcal{H}$, where $m \in \mathbb{N}$ and $p(Y) \in \boldsymbol{k}[Y]$ are such that $\deg p < m$. As a preparation for the proof that (c-i) implies (c-ii), we first show:

(36) *If* $\operatorname{Cont}(\gamma_M \circ h_{m,p})$ *is admissible, then* $\gamma_M \circ h_{m,p} \sim \gamma$ *for some* $\gamma \in \mathcal{G}$.

(37) *If* $\operatorname{Cont}(v_\varphi \circ h_{m,p})$ *is admissible, then* $v_\varphi \circ h_{m,p}$ *is equivalent to an element of* $\mathcal{V} \cup \mathcal{G}$.

Observe that

$$(\gamma_M \circ h_{m,p})(x, y) = \big((xy^m + p(y))^i y^j,\ (xy^m + p(y))^k y^\ell\big). \tag{38}$$

To prove (36), first consider the case $ik \neq 0$; then (38) implies that $\boldsymbol{Z}(XY^m + p(Y))$ is a contracting curve (or a union of contracting curves) of $\gamma_M \circ h_{m,p}$. So, by the hypothesis of (36), each irreducible component of $\boldsymbol{Z}(XY^m + p(Y))$ has one place at infinity. The only way to achieve this is to have $p = 0$, in which case we have $h_{m,p} = \gamma_{M'}$ with $M' = \left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$. Then $\gamma_M \circ h_{m,p} = \gamma_N \in \mathcal{G}$ with $N = MM'$, so (36) is true in this case.

Consider next the case where $ik = 0$. Then $M \in \left\{\left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)\right\}$ for some $j, \ell \in \mathbb{N}$. If $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right)$, then

$$(\gamma_M \circ h_{m,p})(x, y) = \big(y,\ (xy^m + p(y))y^\ell\big) = \big(y,\ xy^{m+\ell} + y^\ell p(y)\big),$$

which is equivalent to the birational morphism $(x, y) \mapsto (y, xy^{m+\ell})$, that is, $\gamma_M \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 0 & 1 \\ 1 & m+\ell \end{smallmatrix}\right)$. Similarly, if $M = \left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)$ then $\gamma_M \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m+j \\ 0 & 1 \end{smallmatrix}\right)$. This completes the proof of (36).

To prove (37), we first note that if $m = 0$, then $v_\varphi \circ h_{m,p} = v_\varphi \circ \mathrm{id} = v_\varphi \in \mathcal{V}$. Likewise, if $\varphi \in \boldsymbol{k}^*$, then v_φ is an isomorphism, so $v_\varphi \circ h_{m,p} \sim h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$. So we may assume from now on that $m > 0$ and that φ has at least one root.

If $c \in \boldsymbol{k}$ is a root of φ, then $\boldsymbol{Z}(XY^m + p(Y) - c)$ is a union of contracting curves of $v_\varphi \circ h_{m,p}$. Therefore, by the hypothesis of (37), each irreducible component of $\boldsymbol{Z}(XY^m + p(Y) - c)$ has one place at infinity. As $m > 0$, this implies that $p(Y) - c$ is the zero polynomial, and this is true for each root c of φ. So $\varphi = a(X - c)^n$ for some $a \in \boldsymbol{k}^*$ and $n \geq 1$, and $h_{m,p}(x, y) = (xy^m + c, y)$. Then $(v_\varphi \circ h_{m,p})(x, y) = (xy^m + c, a(xy^m)^n y)$, which is equivalent to $(x, y) \mapsto (xy^m, (xy^m)^n y)$; that is, $v_\varphi \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m \\ n & mn+1 \end{smallmatrix}\right)$. This proves (37).

To prove that (c-i) implies (c-ii), we consider $f = v_\varphi \circ \gamma_M \circ h$ for some $(\varphi, M, h) \in (\boldsymbol{k}[X] \setminus \{0\}) \times \boldsymbol{M} \times \mathcal{H}$; we assume that $\operatorname{Cont}(f)$ is admissible, and we have to prove (c-ii). We use the notation $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$ and $h(x, y) = (xy^m + p(y), y)$, where $m \in \mathbb{N}$ and $p(Y) \in \boldsymbol{k}[Y]$ are such that $\deg p < m$.

The assumption that $\operatorname{Cont}(f)$ is admissible implies, in particular, that

(39) each contracting curve of $v_\varphi \circ \gamma_M$ has one place at infinity.

Indeed, suppose that $C \in \operatorname{Cont}(v_\varphi \circ \gamma_M)$ has more than one place at infinity; then, by Lemma 2.6(c), C is not a missing curve of h and consequently there exists a curve $C' \subset \mathbb{A}^2$ such that $h(C')$ is a dense subset of C. Then C' is a contracting curve of $f = v_\varphi \circ \gamma_M \circ h$ but has more than one place at infinity (because it dominates a curve with more than one place at infinity). This contradicts the assumption that $\operatorname{Cont}(f)$ is admissible, so (39) is proved.

We claim

(40) *$ij = 0$ or $\varphi(X) = aX^n$, for some $a \in \boldsymbol{k}^*$ and $n \in \mathbb{N}$.*

Indeed, suppose that φ is not of the form aX^n with $a \in \boldsymbol{k}^*$ and $n \in \mathbb{N}$; then there exists $c \in \boldsymbol{k}^*$ such that $\varphi(c) = 0$. Then $\boldsymbol{Z}(x^i y^j - c)$ is a contracting curve of $v_\varphi \circ \gamma_M$ and, if $ij \neq 0$, this curve has more than one place at infinity, contradicting (39). So (40) is proved.

Consider the case where $\varphi(X) = aX^n$. Then $v_\varphi = \theta \circ \gamma_{M_1}$, where $\theta \in \operatorname{Aut}(\mathbb{A}^2)$ and $M_1 = \left(\begin{smallmatrix} 1 & 0 \\ n & 1 \end{smallmatrix}\right) \in \boldsymbol{M}$. Then $f = v_\varphi \circ \gamma_M \circ h = \theta \circ \gamma_{M_1} \circ \gamma_M \circ h \sim \gamma_{M_1 M} \circ h$, so (36) implies that $f \sim \gamma$ for some $\gamma \in \mathcal{G}$, and we are done in this case.

There remains the case $ij = 0$; here we have $M \in \left\{ \left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ k & 1 \end{smallmatrix}\right) \right\}$ for some $k, \ell \in \mathbb{N}$.

If $M = \left(\begin{smallmatrix} 1 & 0 \\ k & 1 \end{smallmatrix}\right)$, then $\gamma_M = v_{(X^k)}$. So $f = v_\varphi \circ v_{(X^k)} \circ h_{m,p} = v_{\varphi_1} \circ h_{m,p}$, where $\varphi_1 = X^k \varphi(X)$, and (37) implies that f is equivalent to an element of $\mathcal{V} \cup \mathcal{G}$ (so we are done).

If $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right)$, then $M = M_1 M_2$, where $M_1 = \left(\begin{smallmatrix} 1 & 0 \\ \ell & 1 \end{smallmatrix}\right)$ and $M_2 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Now $\gamma_{M_2} = \tau$, where $\tau \in \operatorname{Aut}(\mathbb{A}^2)$ is defined by $\tau(x, y) = (y, x)$, and $\gamma_{M_1} = v_{(X^\ell)}$. So we have

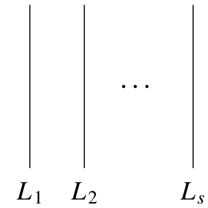
$$f \sim f \circ \tau = v_\varphi \circ \gamma_{M_1} \circ \gamma_{M_2} \circ h_{m,p} \circ \tau = (v_\varphi \circ v_{(X^\ell)}) \circ (\tau \circ h_{m,p} \circ \tau) = v_{\varphi_1} \circ (\tau \circ h_{m,p} \circ \tau),$$

where $\varphi_1 = X^\ell \varphi(X)$. We have $(\tau \circ h_{m,p} \circ \tau)(x, y) = (x, yx^m + p(x))$, so

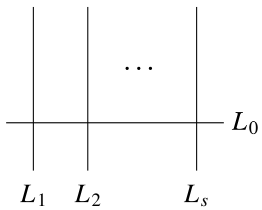
$$(v_{\varphi_1} \circ (\tau \circ h_{m,p} \circ \tau))(x, y) = \big(x, \varphi_1(x)(yx^m + p(x))\big) = (x, x^m \varphi_1(x) y + \varphi_1(x) p(x)),$$

which is equivalent to the birational morphism $(x, y) \mapsto (x, x^m \varphi_1(x) y) = v_\psi(x, y)$ with $\psi = X^m \varphi_1$. So $f \sim v_\psi \in \mathcal{V}$ and we have shown that (c-i) implies (c-ii). □

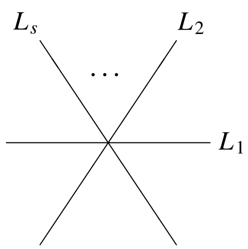
3.16. Corollary. *Let $f \in \operatorname{Bir} \mathbb{A}^2$. Suppose that all missing curves of f are lines and that these are simultaneously rectifiable. Then there exists a coordinate system of $\mathbb{A}^2$ with respect to which the configuration of missing curves is one of the following*:

(a) 

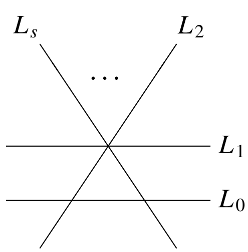
Parallel lines $L_1, \dots, L_s$ $(s \geq 0)$.

(b) 

Parallel lines $L_1, \dots, L_s$ $(s \geq 1)$, *plus one line* L_0 *not parallel to* $L_1, \dots, L_s$.

(c) 

Concurrent lines $L_1, \dots, L_s$ $(s \geq 3)$.

(d) 

Concurrent lines $L_1, \dots, L_s$ $(s \geq 3)$, *plus one line* L_0, *where* L_0 *is parallel to one of the concurrent lines.*

Conversely, each of the above configurations of lines occurs as the configuration of missing curves of some $f \in \text{Bir}(\mathbb{A}^2)$.

The proof below gives in each of the cases (a)–(d) an example of an $f \in \text{Bir}(\mathbb{A}^2)$ having the desired configuration of missing curves.

Proof of Corollary 3.16. The hypothesis on f is that $\text{Miss}(f)$ is weakly admissible, so f is described by part (a-ii) of Theorem 3.15; it follows that $\text{Miss}(f)$ must be one of the configurations (a)–(d). Note that $\text{Miss}(f)$ is admissible in cases (a) and (b). In cases (c) and (d), $\text{Miss}(f)$ is weakly admissible but not admissible.

Conversely, consider the configurations of lines (a)–(d). In each of the four cases we may choose a coordinate system $\mathfrak{c} = (X, Y)$ of $\mathbb{A}^2$ with respect to which the configuration of lines is $\boldsymbol{Z}(F)$, where

$$F = \begin{cases} \prod_{i=1}^{s}(X - c_i) & \text{in case (a),} \\ Y\prod_{i=1}^{s}(X - c_i) & \text{in case (b),} \\ Y\prod_{i=1}^{s-1}(X - c_i Y) & \text{in case (c),} \\ Y(Y-1)\prod_{i=1}^{s-1}(X - c_i Y) & \text{in case (d),} \end{cases}$$

where $c_1, \dots, c_s$ (resp. $c_1, \dots, c_{s-1}$) are distinct elements of $\boldsymbol{k}$ in cases (a) and (b) (resp. in cases (c) and (d)). Let us exhibit in each case an $f \in \mathrm{Bir}(\mathbb{A}^2)$ such that the union of all missing curves of f is $\boldsymbol{Z}(F)$. In cases (a) and (b), choose a univariate polynomial $\varphi \in \boldsymbol{k}[t]$ whose roots are exactly $c_1, \dots, c_s$, and define $f \in \mathrm{Bir}(\mathbb{A}^2)$ by

$$f(x, y) = \begin{cases} (x, \varphi(x)y) & \text{in case (a),} \\ (xy, \varphi(xy)y) & \text{in case (b).} \end{cases}$$

Then the union of the missing curves of f is $\boldsymbol{Z}(F)$, as desired. In cases (c) and (d), first choose $g \in \mathrm{Bir}(\mathbb{A}^2)$ such that the union of the missing curves of g is $\boldsymbol{Z}(G)$, where

$$G = \begin{cases} \prod_{i=1}^{s-1}(X - c_i) & \text{in case (c),} \\ Y\prod_{i=1}^{s-1}(X - c_i) & \text{in case (d)} \end{cases}$$

(we know that g exists by cases (a) and (b)). Then define

$$f = \begin{cases} \alpha_1 \circ g & \text{in case (c),} \\ \alpha_2 \circ g & \text{in case (d),} \end{cases}$$

where α_1 and α_2 are defined in the statement of Theorem 3.15. It follows from Lemma 2.12(b) that the union of the missing curves of f is $\boldsymbol{Z}(F)$. □

4. Some aspects of the monoid Bir($\mathbb{A}^2$)

Let $\boldsymbol{k}$ be an algebraically closed field and $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$, and consider the noncommutative monoid $\mathrm{Bir}(\mathbb{A}^2)$ defined in the introduction. Note that this is a cancellative monoid since it is included in the group of birational automorphisms of $\mathbb{P}^2$.

In view of Lemma 2.5 and Lemma 2.6(b), it is clear that each noninvertible element of $\mathrm{Bir}(\mathbb{A}^2)$ is a composition of finitely many irreducible elements. In other words,

the monoid $\mathrm{Bir}(\mathbb{A}^2)$ *has factorizations into irreducibles.*

Essentially nothing is known regarding uniqueness of factorizations.[4]

[4] We do know that $\mathrm{Bir}(\mathbb{A}^2)$ is not a "unique factorization monoid" in the sense of [Johnson 1971], but this by no means settles the question of uniqueness of factorizations in $\mathrm{Bir}(\mathbb{A}^2)$. Indeed, there are several nonequivalent definitions of what one might mean by "uniqueness of factorization" in noncommutative monoids, and the one used in [Johnson 1971] seems to be particularly inadequate in the case of $\mathrm{Bir}(\mathbb{A}^2)$.

It is natural to ask whether one can find all irreducible elements of $\mathrm{Bir}(\mathbb{A}^2)$ up to equivalence. However, considering the examples given in [Daigle 1991a; 1991b] and [Cassou-Noguès and Russell 2007] and certain facts such as [Daigle 1991a, 4.12], one gets the impression that the irreducible endomorphisms might be too numerous and too diverse to be listed. The first part of the present section gives some simple observations (4.1–4.5) that strengthen that impression.

Given $f, g \in \mathrm{Bir}(\mathbb{A}^2)$, let us write $f \mid g$ if there exist $u, v \in \mathrm{Bir}(\mathbb{A}^2)$ such that $u \circ f \circ v = g$. By a *prime element* of $\mathrm{Bir}(\mathbb{A}^2)$, we mean a noninvertible element p satisfying

$$\text{for all } f, g \in \mathrm{Bir}(\mathbb{A}^2), \quad p \mid (g \circ f) \Rightarrow p \mid f \text{ or } p \mid g.$$

It follows from Lemmas 2.5 and 2.6(b) that every prime element of $\mathrm{Bir}(\mathbb{A}^2)$ is irreducible. It is natural to ask whether the converse is true, and, in particular, whether SACs are prime (SACs are certainly irreducible). These questions are open; we don't even know if there exists a prime element in $\mathrm{Bir}(\mathbb{A}^2)$.

We say that a submonoid $\mathcal{M}$ of $\mathrm{Bir}(\mathbb{A}^2)$ is *factorially closed in* $\mathrm{Bir}(\mathbb{A}^2)$ if the conditions $f, g \in \mathrm{Bir}(\mathbb{A}^2)$ and $g \circ f \in \mathcal{M}$ imply $f, g \in \mathcal{M}$. It is natural to ask whether $\mathcal{A}$ is factorially closed in $\mathrm{Bir}(\mathbb{A}^2)$, where $\mathcal{A}$ is the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.[5] The main result of this section, Theorem 4.8, states that $\mathcal{A}$ is indeed factorially closed in $\mathrm{Bir}(\mathbb{A}^2)$.

Remark. It is obvious that the only irreducible elements of $\mathcal{A}$ are the SACs, that each noninvertible element of $\mathcal{A}$ is a composition of irreducible elements and that $\mathcal{A}$ has the following "unique factorization" property: if $x_1, \dots, x_m, y_1, \dots, y_n$ are irreducible elements of $\mathcal{A}$ such that $x_1 \circ \cdots \circ x_m = y_1 \circ \cdots \circ y_n$, then $m = n$ and for each $i = 1, \dots, n$, we have $x_i = u_i \circ y_i \circ v_i$ for some invertible elements $u_i, v_i \in \mathcal{A}$. (However, it is easy to see that $\mathcal{A}$ is not a unique factorization monoid in the sense defined in [Johnson 1971].)

Irreducible elements and generating sets

We write $[f]$ for the equivalence class of an element f of $\mathrm{Bir}(\mathbb{A}^2)$.

4.1. Lemma. $\big|\{[f] \mid f \textit{ is an irreducible element of } \mathrm{Bir}(\mathbb{A}^2)\}\big| = |\boldsymbol{k}|$.

Proof. Fix a coordinate system (X, Y) of $\mathbb{A}^2$. For each $a \in \boldsymbol{k}^*$, let $C_a \subset \mathbb{A}^2$ be the zero set of $aY^2(Y-1) + X \in \boldsymbol{k}[X, Y]$.

[5] The question is natural in view of the question of whether SACs are prime and in view of the following trivial fact: let P be a set of prime elements in a commutative and cancellative monoid $\mathcal{N}$, and let $\overline{P}$ be the submonoid of $\mathcal{N}$ generated by P and all invertible elements of $\mathcal{N}$; then $\overline{P}$ is factorially closed in $\mathcal{N}$.

Define $U = \{(a_1, a_2, a_3) \in \boldsymbol{k}^3 \mid a_1, a_2, a_3 \text{ are distinct and nonzero}\}$. Define an equivalence relation $\approx$ on the set U by declaring that $(a_1, a_2, a_3) \approx (b_1, b_2, b_3)$ if and only if there exists $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ satisfying $\theta(C_{a_1} \cup C_{a_2} \cup C_{a_3}) = C_{b_1} \cup C_{b_2} \cup C_{b_3}$. The reader may check[6] that the set $U/{\approx}$ of equivalence classes has cardinality $|\boldsymbol{k}|$.

Given $q \geq 2$ and distinct elements $a_1, \dots, a_q \in \boldsymbol{k}^*$, there exists an irreducible element $f \in \mathrm{Bir}(\mathbb{A}^2)$ such that $\mathrm{Miss}(f) = \{C_{a_1}, \dots, C_{a_q}\}$ and $n(f) = q + 2$ (to see this, set $m = 3$ and $\delta_1 = \cdots = \delta_{q-1} = 0$ in [Daigle 1991a, 4.13][7]). In particular, for each $\boldsymbol{a} = (a_1, a_2, a_3) \in U$ there exists an irreducible $f_{\boldsymbol{a}} \in \mathrm{Bir}(\mathbb{A}^2)$ such that $\mathrm{Miss}(f_{\boldsymbol{a}}) = \{C_{a_1}, C_{a_2}, C_{a_3}\}$. If $\boldsymbol{a}, \boldsymbol{b} \in U$ are such that $f_{\boldsymbol{a}} \sim f_{\boldsymbol{b}}$ then there exist $\theta, \theta' \in \mathrm{Aut}(\mathbb{A}^2)$ satisfying $\theta \circ f_{\boldsymbol{a}} = f_{\boldsymbol{b}} \circ \theta'$; then $\theta(C_{a_1} \cup C_{a_2} \cup C_{a_3}) = C_{b_1} \cup C_{b_2} \cup C_{b_3}$, so $\boldsymbol{a} \approx \boldsymbol{b}$. By the preceding paragraph we get $\big|\{[f_{\boldsymbol{a}}] \mid \boldsymbol{a} \in U\}\big| = |\boldsymbol{k}|$, from which the desired conclusion follows. □

4.2. Lemma. *For any subset S of* $\mathrm{Bir}(\mathbb{A}^2)$, *the following are equivalent*:

(i) $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$.

(ii) *For each irreducible $f \in \mathrm{Bir}(\mathbb{A}^2)$, we have $[f] \cap S \neq \varnothing$.*

Proof. Suppose that S satisfies (i) and consider an irreducible $f \in \mathrm{Bir}(\mathbb{A}^2)$. By (i),

$$f = g_1 \circ \cdots \circ g_n \quad \text{for some finite subset } \{g_1, \dots, g_n\} \text{ of } \mathrm{Aut}(\mathbb{A}^2) \cup S.$$

By irreducibility of f, exactly one element g_i of $\{g_1, \dots, g_n\}$ is not in $\mathrm{Aut}(\mathbb{A}^2)$ (consequently, $g_i \in S$). So $f \sim g_i \in S$, which proves that S satisfies (ii).

Conversely, suppose that (ii) holds and consider $h \in \mathrm{Bir}(\mathbb{A}^2)$; we claim that

$$h = g_1 \circ \cdots \circ g_N \quad \text{for some finite subset } \{g_1, \dots, g_N\} \text{ of } \mathrm{Aut}(\mathbb{A}^2) \cup S.$$

This is clear if $h \in \mathrm{Aut}(\mathbb{A}^2)$, so assume that $h \notin \mathrm{Aut}(\mathbb{A}^2)$. Then $h = f_1 \circ \ldots \circ f_n$ for some finite collection $\{f_1, \dots, f_n\}$ of irreducible elements of $\mathrm{Bir}(\mathbb{A}^2)$ (existence of a factorization into irreducibles is a consequence of Lemma 2.5). For each $i \in \{1, \dots, n\}$, we have $[f_i] \cap S \neq \varnothing$, so $f_i = u_i \circ s_i \circ v_i$ for some $s_i \in S$ and $u_i, v_i \in \mathrm{Aut}(\mathbb{A}^2)$. Then

$$h = (u_1 \circ s_1 \circ v_1) \circ \cdots \circ (u_n \circ s_n \circ v_n) = g_1 \circ \cdots \circ g_N,$$

where $\{g_1, \dots, g_N\} \subset \mathrm{Aut}(\mathbb{A}^2) \cup S$. This proves (i). □

4.3. Corollary. *Let S be a subset of* $\mathrm{Bir}(\mathbb{A}^2)$ *such that* $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$. *Then $|S| = |\boldsymbol{k}|$.*

Proof. Follows from Lemmas 4.1 and 4.2. □

[6]This is a tedious exercise. We leave it to the reader because it is completely elementary and has nothing to do with the subject matter of this paper.

[7]Note that in Example 4.13 of [Daigle 1991a] one has $\mathrm{Miss}(f) = \{C_1, \dots, C_q\}$. This doesn't seem to be stated explicitly, but it is clear if one reads the construction.

4.4. Remark. Let $f \in \mathrm{Bir}(\mathbb{A}^2)$ and let $\gamma = (X, Y)$ be a coordinate system of $\mathbb{A}^2$. Then $f : \mathbb{A}^2 \to \mathbb{A}^2$ is given by $f(x, y) = (u(x, y), v(x, y))$ for some polynomials $u, v \in \boldsymbol{k}[X, Y]$. We define $\deg_\gamma f = \max(\deg_\gamma u, \deg_\gamma v)$. We may also define $\deg f$ to be the minimum of $\deg_\gamma f$ for γ ranging over the set of coordinate systems of $\mathbb{A}^2$. Then

$$\deg f \geq \frac{c(f)+2}{2}. \tag{41}$$

Indeed, if $F_1, \dots, F_c \in \boldsymbol{k}[X, Y]$ are irreducible polynomials whose zero sets are the contracting curves of f (so $c(f) = c$), then the jacobian determinant of (u, v) with respect to (X, Y) is divisible by $\prod_{i=1}^{c} F_i$. This implies that $\deg_\gamma f \geq (c+2)/2$, where the right-hand side is independent of γ. Statement (41) follows.

4.5. Corollary.[8] *Let S be a subset of* $\mathrm{Bir}(\mathbb{A}^2)$ *such that* $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$. *Then* $\{\deg f \mid f \in S\}$ *is not bounded.*

Proof. Let $n \in \mathbb{N}$. By [Daigle 1991a, 4.13], there exists an irreducible element $g \in \mathrm{Bir}(\mathbb{A}^2)$ satisfying $c(g) \geq 2n$. By 4.2, there exists $f \in S$ satisfying $f \sim g$; then $c(f) = c(g) \geq 2n$, so $\deg f > n$ by (41). □

Factorial closedness of $\mathcal{A}$ in $\mathrm{Bir}(\mathbb{A}^2)$

Let $\mathcal{A}$ be the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.

See Definition 2.19 for the definition of $n(f, C)$, where $f \in \mathrm{Bir}(\mathbb{A}^2)$ and $C \in \mathrm{Miss}(f)$.

4.6. Lemma. *Consider* $\mathbb{A}^2 \xrightarrow{\alpha} \mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2$, *where* $\alpha, f \in \mathrm{Bir}(\mathbb{A}^2)$ *and α is a SAC. Assume that the missing curve C of α is disjoint from* $\mathrm{exc}(f)$ *and let D be the closure of $f(C)$ in $\mathbb{A}^2$. Then there exist a SAC α' and some $f' \in$* $\mathrm{Bir}(\mathbb{A}^2)$ *satisfying* $f \circ \alpha = \alpha' \circ f'$ *and* $\mathrm{Miss}(\alpha') = \{D\}$. *Moreover, if f is a SAC then so is f'.*

Proof. By Lemma 2.21(b), we have $D \in \mathrm{Miss}(f \circ \alpha)$ and $n(f \circ \alpha, D) = n(\alpha, C) = 1$ (because $C \cap \mathrm{exc}(f) = \varnothing$ and $C \cong \mathbb{A}^1$). Let P be the unique fundamental point of $f \circ \alpha$ which lies on D and let α' be a SAC with missing curve D and fundamental point P. Then Lemma 3.5(a) implies that $f \circ \alpha = \alpha' \circ f'$ for some $f' \in \mathrm{Bir}(\mathbb{A}^2)$. Then $n(f') = n(f)$, so if f is a SAC, then so is f'. □

4.7. Definition. Let $h \in \mathrm{Bir}(\mathbb{A}^2)$ be such that $h \notin \mathrm{Aut}(\mathbb{A}^2)$. Let $C \in \mathrm{Miss}(h)$.

(a) A *factorization* of h is a tuple $\mathfrak{f} = (h_1, \dots, h_n)$ of elements of $\mathrm{Bir}(\mathbb{A}^2)$ satisfying $h = h_1 \circ \cdots \circ h_n$ (where $n \geq 1$). If $h_1, \dots, h_n$ are SACs, we say that $\mathfrak{f}$ is a *factorization of h into SACs*.

[8] This result answers a question posed by Patrick Popescu-Pampu.

(b) Given a factorization $\mathfrak{f} = (h_1, \dots, h_n)$ of h, we define $\operatorname{depth}_{\mathfrak{f}}(h, C)$ to be the unique $i \in \{1, \dots, n\}$ satisfying

there exists a missing curve of h_i whose image by $h_1 \circ \cdots \circ h_{i-1}$ is a dense subset of C.

Observe that $\operatorname{depth}_{\mathfrak{f}}(h, C) \geq 1$ and that $\operatorname{depth}_{\mathfrak{f}}(h, C) = 1 \iff C \in \operatorname{Miss}(h_1)$.

(c) If $h \in \mathcal{A}$, then we define

$$\operatorname{depth}(h, C) = \min\{\operatorname{depth}_{\mathfrak{f}}(h, C) \mid \mathfrak{f} \text{ is a factorization of } h \text{ into SACs}\}.$$

Note that $\operatorname{depth}(h, C) \geq 1$ and that $\operatorname{depth}(h, C) = 1$ is equivalent to the existence of SACs $\alpha_1, \dots, \alpha_n$ satisfying

$$h = \alpha_1 \circ \cdots \circ \alpha_n \quad \text{and} \quad \operatorname{Miss}(\alpha_1) = \{C\}.$$

4.8. Theorem. *If $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ satisfy $g \circ f \in \mathcal{A}$, then $f, g \in \mathcal{A}$.*

Proof. We proceed by induction on $n(g \circ f)$, the result being trivial for $n(g \circ f) \leq 2$. Let $n \geq 3$ be such that

$$(*) \qquad \text{for all } f, g \in \operatorname{Bir}(\mathbb{A}^2), \quad g \circ f \in \mathcal{A} \quad \text{and} \quad n(g \circ f) < n \implies f, g \in \mathcal{A}.$$

Consider $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ such that $g \circ f \in \mathcal{A}$ and $n(g \circ f) = n$; we have to show that $f, g \in \mathcal{A}$. Since $g \circ f \in \mathcal{A}$, the number $\operatorname{depth}(g \circ f, C)$ is defined for every $C \in \operatorname{Miss}(g \circ f)$. Observe that

(42) there exists $C \in \operatorname{Miss}(g \circ f)$ satisfying $\operatorname{depth}(g \circ f, C) = 1$, and any such C satisfies $n(g \circ f, C) = 1$.

Indeed, for any factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs, the missing curve C of α_1 satisfies $C \in \operatorname{Miss}(g \circ f)$ and $\operatorname{depth}(g \circ f, C) = 1$, so C exists. Given any $C \in \operatorname{Miss}(g \circ f)$ satisfying $\operatorname{depth}(g \circ f, C) = 1$, there exists a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\operatorname{Miss}(\alpha_1) = \{C\}$; then $n(g \circ f, C) = n(\alpha_1 \circ \cdots \circ \alpha_n, C) = n(\alpha_1, C) = 1$, where the second equality follows from Lemma 2.21(a). This proves (42).

We now proceed to prove that $f, g \in \mathcal{A}$. We first do so in two special cases (numbered 1 and 2) and then in the general case.

Case 1: there exists $C \in \operatorname{Miss}(g)$ such that $\operatorname{depth}(g \circ f, C) = 1$.
Then there exist SACs $\alpha_1, \dots, \alpha_n$ satisfying $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ and $\operatorname{Miss}(\alpha_1) = \{C\}$. We note that $n(g, C) = n(g \circ f, C) = 1$, where the first equality follows from Lemma 2.21(a) and the second from (42), and where the assumption $C \in \operatorname{Miss}(g)$ is needed for the first equality. As $n(g, C) = 1$, there is a unique fundamental point P of g lying on C. Consider the fundamental point P_1 of α_1; then Lemma 2.11 implies that P and P_1 are fundamental points of $g \circ f$ (lying on C); as $n(g \circ f, C) = 1$, we have $P = P_1$, so α_1 is a SAC with missing curve C and fundamental point P. By

Lemma 3.5(a), there exists $g' \in \mathrm{Bir}(\mathbb{A}^2)$ such that $g = \alpha_1 \circ g'$. Then $\alpha_1 \circ g' \circ f = g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$; canceling α_1 yields $g' \circ f = \alpha_2 \circ \cdots \circ \alpha_n \in \mathcal{A}$. As $n(g' \circ f) = n-1$, we get $g', f \in \mathcal{A}$ by $(*)$. Then $g = \alpha_1 \circ g' \in \mathcal{A}$ as well, so we are done in Case 1.

Case 2: $n(f) = 1$.
Note that f is a SAC; let C be its missing curve. By (42), we may consider $D_1 \in \mathrm{Miss}(g \circ f)$ satisfying $\mathrm{depth}(g \circ f, D_1) = 1$ and $n(g \circ f, D_1) = 1$.

By Case 1, we may assume that $D_1 \notin \mathrm{Miss}(g)$. Then (by Lemma 2.12) D_1 is the closure of $g(C)$; since $C \cong \mathbb{A}^1$, we have, in fact, $g(C) = D_1$ (every dominant morphism $\mathbb{A}^1 \to C$ is surjective). Let P be the unique fundamental point of $g \circ f$ on D_1 and let $Q \in C$ be the fundamental point of f; then $g(Q) \in D_1$ is a fundamental point of $g \circ f$ by Lemma 2.11, so $g(Q) = P$.

Since $n(g \circ f, D_1) = 1 = n(f, C)$, Lemma 2.21(b) implies that $C \cap \mathrm{exc}(g) = \varnothing$. By Lemma 2.8, g restricts to an isomorphism $\mathbb{A}^2 \setminus \mathrm{exc}(g) \to \mathbb{A}^2 \setminus \Gamma_g$, where Γ_g is the union of all missing curves of g. Since $C \subset \mathbb{A}^2 \setminus \mathrm{exc}(g)$ and $D_1 = g(C)$, it follows that $D_1 \subset \mathbb{A}^2 \setminus \Gamma_g$. Since $P \in D_1 \subset \mathbb{A}^2 \setminus \Gamma_g$ and $\mathrm{cent}(g) \subseteq \Gamma_g$, we have $P \notin \mathrm{cent}(g)$ and hence $n(g, P) = 0$; so Lemma 2.21 gives $n(g \circ f, P) = n(g, P) + \sum_{P' \in \{Q\}} n(f, P') = 1$ and we have shown

$$D_1 \cap \Gamma_g = \varnothing \quad \text{and} \quad n(g \circ f, P) = 1. \tag{43}$$

Since $\mathrm{depth}(g \circ f, D_1) = 1$, we may choose a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\mathrm{Miss}(\alpha_1) = \{D_1\}$. We have $\mathrm{cent}(\alpha_1) = \{P\}$ because the fundamental point of α_1 is a fundamental point of $g \circ f$ lying on D_1. Write $\mathrm{Cont}(\alpha_1) = \{E_1\}$; then by Lemma 2.21,

$$n(\alpha_1 \circ (\alpha_2 \circ \cdots \circ \alpha_n), P) = n(\alpha_1, P) + \sum_{P' \in E_1} n(\alpha_2 \circ \cdots \circ \alpha_n, P'),$$

where the left-hand side is equal to $n(g \circ f, P) = 1$ by (43). As $n(\alpha_1, P) = 1$, we have $n(\alpha_2 \circ \cdots \circ \alpha_n, P') = 0$ for all $P' \in E_1$, so $\mathrm{cent}(\alpha_2 \circ \cdots \circ \alpha_n) \cap E_1 = \varnothing$ and, in particular, $\mathrm{cent}(\alpha_2) \cap E_1 = \varnothing$. It follows that the missing curve C_2 of α_2 is not equal to E_1 (because $\mathrm{cent}(\alpha_2) \subset C_2$). So the closure of $\alpha_1(C_2)$ in $\mathbb{A}^2$ is a curve D_2 such that

$$D_2 \in \mathrm{Miss}(g \circ f) \setminus \{D_1\} = \mathrm{Miss}(g).$$

Then $D_2 \subseteq \Gamma_g$, so $D_2 \cap D_1 = \varnothing$ by (43). If $C_2 \cap E_1 \neq \varnothing$, then $\alpha_1(C_2) \cap \alpha_1(E_1) \neq \varnothing$, so $P \in D_2$, contradicting $D_2 \cap D_1 = \varnothing$; thus

$$C_2 \cap E_1 = \varnothing.$$

This allows us to use Lemma 4.6. By that result, there exist SACs α_1', α_2' such that $\alpha_1 \circ \alpha_2 = \alpha_1' \circ \alpha_2'$ and $\mathrm{Miss}(\alpha_1') = \{D_2\}$. Since $g \circ f = \alpha_1' \circ \alpha_2' \circ \alpha_3 \circ \cdots \circ \alpha_n$ is a factorization of $g \circ f$ into SACs satisfying $\mathrm{Miss}(\alpha_1') = \{D_2\}$, we have $\mathrm{depth}(g \circ f, D_2) = 1$. Since $D_2 \in \mathrm{Miss}(g)$, Case 1 implies that $f, g \in \mathcal{A}$.

General case. The result is trivial if $n(f) = 0$ and follows from Case 2 if $n(f) = 1$. So we may assume that $n(f) \geq 2$. Consequently, $n(g) \leq n - 2$.

By (42), we may pick $D \in \operatorname{Miss}(g \circ f)$ satisfying $\operatorname{depth}(g \circ f, D) = 1$ and $n(g \circ f, D) = 1$. By Case 1, we may assume that $D \notin \operatorname{Miss}(g)$. Then D is the closure of $g(C)$ for some $C \in \operatorname{Miss}(f)$. We have $1 \leq n(f, C) \leq n(g \circ f, D) = 1$, so $n(f, C) = 1$. Then Lemma 3.5(a) implies that there exist an SAC α and some $f' \in \operatorname{Bir}(\mathbb{A}^2)$ such that $f = \alpha \circ f'$ and $\operatorname{Miss}(\alpha) = \{C\}$. On the other hand, the fact that $\operatorname{depth}(g \circ f, D) = 1$ allows us to choose a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\operatorname{Miss}(\alpha_1) = \{D\}$. We have $D \in \operatorname{Miss}(g \circ \alpha)$ and

$$n(g \circ \alpha, D) \overset{2.21(a)}{=} n(g \circ \alpha \circ f', D) = n(g \circ f, D) = 1.$$

Let P be the unique fundamental point of $g \circ \alpha$ lying on D; then P is a fundamental point of $g \circ f$ and hence is the unique fundamental point of $g \circ f$ lying on D. As the fundamental point of α_1 is a fundamental point of $g \circ f$ lying on D, it follows that α_1 is a SAC with missing curve D and fundamental point P. Then Lemma 3.5(a) implies that there exists $g' \in \operatorname{Bir}(\mathbb{A}^2)$ satisfying $g \circ \alpha = \alpha_1 \circ g'$.

$$\begin{tikzcd}
\mathbb{A}^2 \arrow[r, "f"] \arrow[dr, "f'"'] & \mathbb{A}^2 \arrow[r, "g"] & \mathbb{A}^2 \\
& \mathbb{A}^2 \arrow[u, "\alpha"'] \arrow[r, "g'"'] & \mathbb{A}^2 \arrow[u, "\alpha_1"']
\end{tikzcd}$$

Since $\alpha_1 \circ g' \circ f' = g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$, canceling α_1 gives $g' \circ f' = \alpha_2 \circ \cdots \circ \alpha_n \in \mathcal{A}$. By $(*)$, we obtain $f', g' \in \mathcal{A}$.

Since $f' \in \mathcal{A}$, it follows that $f = \alpha \circ f' \in \mathcal{A}$.

Since $g' \in \mathcal{A}$, we get $g \circ \alpha = \alpha_1 \circ g' \in \mathcal{A}$; we also have $n(g \circ \alpha) < n$ because $n(g) \leq n - 2$; so $g \in \mathcal{A}$ by $(*)$.

So $f, g \in \mathcal{A}$. $\square$

Acknowledgements

Daigle would like to thank his teacher, Peter Russell, for introducing him to birational endomorphisms of $\mathbb{A}^2$. He would also like to thank the faculty and staff at the Université de Bordeaux 1 for their hospitality. The research leading to this paper was initiated when Daigle spent one month at that institution as a *professeur invité*.

References

[Abhyankar and Moh 1975] S. S. Abhyankar and T. T. Moh, "Embeddings of the line in the plane", *J. Reine Angew. Math.* **276** (1975), 148–166. MR 52 #407 Zbl 0332.14004

[Artal Bartolo and Cassou-Noguès 1996] E. Artal Bartolo and P. Cassou-Noguès, "One remark on polynomials in two variables", *Pacific J. Math.* **176**:2 (1996), 297–309. MR 97m:32044 Zbl 0878.32003

[Cassou-Noguès 2005] P. Cassou-Noguès, "Bad field generators", pp. 77–83 in *Affine algebraic geometry*, edited by J. Gutierrez et al., Contemp. Math. **369**, Amer. Math. Soc., Providence, RI, 2005. MR 2005k:14127 Zbl 1076.14084

[Cassou-Noguès and Daigle 2013] P. Cassou-Noguès and D. Daigle, "Very good and very bad field generators", preprint, 2013. To appear in *Kyoto J. Math.* arXiv 1308.0154

[Cassou-Noguès and Daigle $\geq$ 2014] P. Cassou-Noguès and D. Daigle, "Lean factorizations of polynomial morphisms". In preparation.

[Cassou-Noguès and Russell 2007] P. Cassou-Noguès and P. Russell, "Birational morphisms $\mathbb{C}^2 \to \mathbb{C}^2$ and affine ruled surfaces", pp. 57–105 in *Affine algebraic geometry*, edited by T. Hibi, Osaka Univ. Press, Osaka, 2007. MR 2008d:14098 Zbl 1222.14027

[Daigle 1991a] D. Daigle, "Birational endomorphisms of the affine plane", *J. Math. Kyoto Univ.* **31**:2 (1991), 329–358. MR 92k:14012 Zbl 0766.14006

[Daigle 1991b] D. Daigle, "Local trees in the theory of affine plane curves", *J. Math. Kyoto Univ.* **31**:3 (1991), 593–634. MR 92h:14023 Zbl 0758.14018

[Daigle 2010] D. Daigle, "Triangular derivations of $\mathbf{k}[X, Y, Z]$", *J. Pure Appl. Algebra* **214**:7 (2010), 1173–1180. MR 2011g:13056 Zbl 1196.14055

[Daigle 2013] D. Daigle, "Generally rational polynomials in two variables", preprint, 2013. To appear in *Osaka J. Math.* **52**:1. arXiv 1307.3752

[Ganong 1985] R. Ganong, "Kodaira dimension of embeddings of the line in the plane", *J. Math. Kyoto Univ.* **25**:4 (1985), 649–657. MR 87c:14013 Zbl 0596.14018

[Ganong 2011] R. Ganong, "The pencil of translates of a line in the plane", pp. 57–71 in *Affine algebraic geometry*, edited by D. Daigle et al., CRM Proc. Lecture Notes **54**, Amer. Math. Soc., Providence, RI, 2011. MR 2768634 Zbl 1263.14057

[Jan 1974] C. J. Jan, *On polynomial generators of* $\boldsymbol{k}(x, y)$, Ph.D. thesis, Purdue University, 1974.

[Johnson 1971] R. E. Johnson, "Unique factorization monoids and domains", *Proc. Amer. Math. Soc.* **28** (1971), 397–404. MR 43 #3186 Zbl 0219.20046

[Kaliman 1992] S. Kaliman, "Two remarks on polynomials in two variables", *Pacific J. Math.* **154**:2 (1992), 285–295. MR 93b:32041 Zbl 0723.32001

[Miyanishi and Sugie 1980] M. Miyanishi and T. Sugie, "Generically rational polynomials", *Osaka J. Math.* **17**:2 (1980), 339–362. MR 82e:14019 Zbl 0457.13006

[Neumann and Norbury 2002] W. D. Neumann and P. Norbury, "Rational polynomials of simple type", *Pacific J. Math.* **204**:1 (2002), 177–207. MR 2003f:14010 Zbl 1055.14062

[Nishino 1968] T. Nishino, "Nouvelles recherches sur les fonctions entiéres de plusieurs variables complexes, I", *J. Math. Kyoto Univ.* **8** (1968), 49–100. MR 38 #3469 Zbl 0165.40302

[Nishino 1969] T. Nishino, "Nouvelles recherches sur les fonctions entières de plusieurs variables complexes, II: Fonctions entières qui se réduisent à celles d'une variable", *J. Math. Kyoto Univ.* **9** (1969), 221–274. MR 41 #502 Zbl 0192.43703

[Nishino 1970] T. Nishino, "Nouvelles recherches sur les fonctions entières de plusiers variables complexes, III: Sur quelques propriétés topologiques des surfaces premières", *J. Math. Kyoto Univ.* **10** (1970), 245–271. MR 42 #533 Zbl 0204.09403

[Russell 1975] P. Russell, "Field generators in two variables", *J. Math. Kyoto Univ.* **15**:3 (1975), 555–571. MR 53 #2946 Zbl 0341.14008

[Russell 1977] P. Russell, "Good and bad field generators", *J. Math. Kyoto Univ.* **17**:2 (1977), 319–331. MR 56 #2977 Zbl 0367.12013

[Saitō 1972] H. Saitō, "Fonctions entières qui se réduisent à certains polynomes, I", *Osaka J. Math.* **9** (1972), 293–332. MR 49 #3187 Zbl 0248.32005

[Saito 1977] H. Saito, "Fonctions entières qui se rédiusent à certains polynômes, II", *Osaka J. Math.* **14**:3 (1977), 649–674. MR 57 #679 Zbl 0372.32004

[Sasao 2006] I. Sasao, "Generically rational polynomials of quasi-simple type", *J. Algebra* **298**:1 (2006), 58–104. MR 2007b:14131 Zbl 1115.14056

[Shpilrain and Yu 2004] V. Shpilrain and J.-T. Yu, "Birational morphisms of the plane", *Proc. Amer. Math. Soc.* **132**:9 (2004), 2511–2515. MR 2005c:14086 Zbl 1060.14086

[Suzuki 1974] M. Suzuki, "Propriétés topologiques des polynômes de deux variables complexes, et automorphismes algébriques de l'espace $\mathbf{C}^2$", *J. Math. Soc. Japan* **26** (1974), 241–257. MR 49 #3188 Zbl 0276.14001

Received July 17, 2013.

PIERRETTE CASSOU-NOGUÈS
INSTITUT DE MATHÉMATIQUES DE BORDEAUX
UNIVERSITÉ DE BORDEAUX
351, COURS DE LA LIBÉRATION
33405 TALENCE CEDEX
FRANCE
pierrette.cassou-nogues@math.u-bordeaux1.fr

DANIEL DAIGLE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
OTTAWA K1N 6N5
CANADA
ddaigle@uottawa.ca

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.395

THE BOCHNER FORMULA FOR ISOMETRIC IMMERSIONS

ALESSANDRO SAVO

We study the Bochner formula for a manifold isometrically immersed into another and obtain a Gauss-type splitting of its curvature term. In fact, we prove that the curvature term in the Bochner formula is an operator that can be explicitly expressed in terms of the curvature operator of the ambient manifold and the extrinsic geometry (second fundamental form) of the immersion. Several applications of this splitting are given, namely, eigenvalue estimates for the Hodge Laplacian, vanishing results for the de Rham cohomology and rigidity of immersions of Kähler manifolds into negatively curved spaces.

1. Introduction

Let Σ^n be a Riemannian manifold of dimension n (all manifolds in this paper are connected, orientable and without boundary) and let ω be a differential p-form on Σ. The Bochner formula states

$$\Delta\omega = \nabla^\star\nabla\omega + \mathcal{B}^{[p]}\omega. \tag{1}$$

Here $\Delta = d\delta + \delta d$ is the Hodge Laplacian (δ being the adjoint of d), $\nabla^\star\nabla$ is the connection Laplacian, and

$$\mathcal{B}^{[p]} : \Lambda^p(\Sigma) \to \Lambda^p(\Sigma)$$

is a certain symmetric endomorphism of the bundle of p-forms. We call $\mathcal{B}^{[p]}$ the *Bochner curvature term* or simply the *Bochner operator*.

When Σ is compact, knowing that $\mathcal{B}^{[p]}$ is positive at each point implies that any harmonic p-form must vanish; then, by the Hodge–de Rham theorem, the de Rham cohomology $H^p(\Sigma, \mathbb{R})$ must vanish. More generally, a positive lower bound of the eigenvalues of $\mathcal{B}^{[p]}$ implies a positive lower bound of the first eigenvalue of the Hodge Laplacian acting on p-forms. All these facts are consequences of the well-known *Bochner method*, and will be recalled in Proposition 3.

It is then important to look for estimates of the eigenvalues of $\mathcal{B}^{[p]}$. It turns out that $\mathcal{B}^{[1]}$ is simply the Ricci tensor (acting on 1-forms); but for degrees $2 \le p \le n-2$

MSC2010: 58J50, 35P15.
Keywords: differential forms, Bochner formula, isometric immersions, vanishing theorems.

the operator $\mathcal{B}^{[p]}$ is more complicated and difficult to control. A breakthrough was obtained by Gallot and Meyer [1975]. They proved that, for all $p = 1, \dots, n-1$,

$$\mathcal{B}^{[p]} \geq p(n-p)\gamma_\Sigma, \tag{2}$$

where γ_Σ is a lower bound of the eigenvalues of the curvature operator of Σ.

In this paper, we study the Bochner curvature term of a manifold Σ^n isometrically immersed in a larger manifold M^{n+q} by a smooth map $f : \Sigma^n \to M^{n+q}$. The natural problem we address is to give a new expression of the Bochner curvature in terms of the extrinsic geometry of the immersion, which would eventually improve the estimate (2) in this important situation. It turns out that this is in fact possible. In Theorem 1 we prove a Gauss-type formula and show that $\mathcal{B}^{[p]}$ splits into the sum of two operators acting on $\Lambda^p(\Sigma)$:

$$\mathcal{B}^{[p]} = \mathcal{B}^{[p]}_{\text{res}} + \mathcal{B}^{[p]}_{\text{ext}}.$$

We then prove that $\mathcal{B}^{[p]}_{\text{res}}$, which depends on the geometry of the ambient manifold M, is bounded below by the lowest eigenvalue of the curvature operator of M:

$$\mathcal{B}^{[p]}_{\text{res}} \geq p(n-p)\gamma_M,$$

while the extrinsic part $\mathcal{B}^{[p]}_{\text{ext}}$ is explicitly described in terms of the second fundamental form of the immersion. For example, if Σ has codimension one and S is the shape operator relative to any of the two choices of the unit normal vector field, then

$$\mathcal{B}^{[p]}_{\text{ext}} = \operatorname{tr} S \cdot S^{[p]} - S^{[p]} \circ S^{[p]}, \tag{3}$$

where $S^{[p]} : \Lambda^p(\Sigma) \to \Lambda^p(\Sigma)$ is the self-adjoint extension of S, acting on the form ω by

$$S^{[p]}\omega(X_1, X_2, \dots, X_p) = \sum_{j=1}^{p} \omega(X_1, \dots, S(X_j), \dots, X_p) \tag{4}$$

for all tangent vectors $X_1, \dots, X_p$. In higher codimensions, one simply sums the expression (3) over the shape operators of an orthonormal basis of the normal bundle. We refer to Theorem 1 for the precise statement.

From (3) and (4) one sees that the eigenvalues of $\mathcal{B}^{[p]}_{\text{ext}}$ can be estimated in terms of the eigenvalues of S (principal curvatures) and explicit bounds of the extrinsic part will follow.

Thus, one can bound the Bochner curvature in terms of the curvature operator of the ambient manifold M and the second fundamental form of the immersion. This can be done in different ways (see Section 4) and leads to a number of applications. In particular:

- We prove an extrinsic, sharp lower bound of the first eigenvalue of the Hodge Laplacian of Σ, assuming that Σ has codimension one in M (Theorem 7).
- We prove a vanishing theorem for the de Rham cohomology of Σ, assuming that the norm of the traceless second fundamental form of Σ is bounded above by a suitable function of the mean curvature (Theorem 8). The condition is sharp when the ambient manifold is a sphere, in which case we get a rigidity result for Clifford tori (Theorem 9).
- We prove a lower bound of the first eigenvalue of the Hodge Laplacian of Σ, assuming that the norm of its second fundamental form is bounded above by a suitable constant and that the ambient manifold is positively curved (Theorem 10). When the ambient manifold is a sphere this reproves a vanishing theorem for the de Rham cohomology of Σ due to Lawson and Simons; moreover the limit case leads to a rigidity result for a certain Clifford torus (Theorem 11).
- We prove that if Σ supports a nontrivial parallel p-form and admits an isometric immersion into a negatively curved space, then its mean curvature vector has norm bounded below by an explicit positive constant (Theorem 12). This has applications to immersions of Kähler manifolds (Corollary 13).
- We classify the compact hypersurfaces of $\mathbb{S}^{n+1}$ which support a parallel p-form (Theorem 4): if $n \geq 3$ they are in fact products of spheres (Clifford tori). This result is perhaps of independent interest, and is needed to prove the rigidity theorems above.

In conclusion, we hope that this new representation of the Bochner formula will be useful, and that it will lead to other interesting applications in submanifold geometry.

The paper is organized as follows. In Section 2 we state the main theorem; in Section 3 we state the main applications; these will be proved in Section 4, along with the explicit bounds on the Bochner operator. In Section 5 we prove the main theorem and in Section 6 we prove Theorem 4.

2. The Bochner curvature term

Let $f : \Sigma^n \to M^{n+q}$ be an isometric immersion with codimension $q \geq 1$. The second fundamental form L of f is defined by the relation

$$\nabla^M_X Y = \nabla_X Y + L(X, Y),$$

where $X, Y \in T\Sigma$ and ∇^M, ∇ denote the Levi-Civita connections in M and Σ, respectively. Note that $L(X, Y)$ is a vector normal to Σ, so that L takes values in the normal bundle of the immersion. If $\nu \in T^\perp\Sigma$ is a normal vector, we denote its associated shape operator by S_ν. It is the self-adjoint endomorphism of $T\Sigma$ defined on $X, Y \in T\Sigma$ by

$$\langle S_\nu(X), Y\rangle = \langle L(X, Y), \nu\rangle.$$

The mean curvature vector H is defined by $H = (1/n)\,\mathrm{tr}\, L$; so, for any orthonormal basis $(e_1, \dots, e_n)$ of $T\Sigma$, one has $nH = \sum_{j=1}^n L(e_j, e_j)$. The squared norm of the second fundamental form is denoted by $|S|^2$. Note that if $(\nu_1, \dots, \nu_q)$ is an orthonormal basis in the normal bundle, then

$$n^2|H|^2 = \sum_{i=1}^q (\mathrm{tr}\, S_{\nu_i})^2 \quad \text{and} \quad |S|^2 = \sum_{i=1}^q |S_{\nu_i}|^2.$$

Now extend S_ν to a self-adjoint operator acting on p-forms, $S_\nu^{[p]} : \Lambda^p(\Sigma) \to \Lambda^p(\Sigma)$, as in (4), and let

$$T_\nu^{[p]} \doteq (\mathrm{tr}\, S_\nu) S_\nu^{[p]} - S_\nu^{[p]} \circ S_\nu^{[p]}.$$

Introduce the self-adjoint endomorphism of $\Lambda^p(\Sigma)$

$$\mathcal{B}_{\mathrm{ext}}^{[p]} = \sum_{j=1}^q T_{\nu_j}^{[p]} = \sum_{j=1}^q \big((\mathrm{tr}\, S_{\nu_j}) S_{\nu_j}^{[p]} - S_{\nu_j}^{[p]} \circ S_{\nu_j}^{[p]}\big). \tag{5}$$

Then, we have:

Theorem 1. *Let $f : \Sigma^n \to M^{n+q}$ be an isometric immersion. The Bochner operator acting on p-forms of Σ splits as*

$$\mathcal{B}^{[p]} = \mathcal{B}_{\mathrm{res}}^{[p]} + \mathcal{B}_{\mathrm{ext}}^{[p]},$$

where the operator $\mathcal{B}_{\mathrm{ext}}^{[p]}$ is defined by (5), and the operator $\mathcal{B}_{\mathrm{res}}^{[p]}$ satisfies the bounds

$$p(n-p)\gamma_M \le \mathcal{B}_{\mathrm{res}}^{[p]} \le p(n-p)\Gamma_M,$$

where γ_M and Γ_M are respectively a lower and an upper bound on the curvature operator of M. If M has constant sectional curvature γ, then

$$\mathcal{B}_{\mathrm{res}}^{[p]} = p(n-p)\gamma.$$

For the definition of $\mathcal{B}_{\mathrm{res}}^{[p]}$ and the proof of Theorem 1, see Section 5. The notation "res" is chosen because $\mathcal{B}_{\mathrm{res}}^{[p]}$ depends on the restriction of the curvature operator of M to the submanifold Σ.

Let us give a more explicit expression of $\mathcal{B}_{\mathrm{ext}}^{[p]}$. In what follows, $S = S_\nu$ will be the shape operator associated to a given unit normal vector ν. We denote by I_p the set of p-multi-indices

$$I_p = \{\{j_1, \dots, j_p\} : 1 \le j_1 < \dots < j_p \le n\}.$$

Let $(e_1, \dots, e_n)$ be an orthonormal basis of $T\Sigma$ which diagonalizes S; then

$$S(e_j) = k_j e_j,$$

where $k_1, \dots, k_n$ are the principal curvatures. If $\alpha = \{j_1, \dots, j_p\} \in I_p$, we define

$$K_\alpha \doteq k_{j_1} + \cdots + k_{j_p} \tag{6}$$

and call it a *p-curvature* of S. Let $(\theta_1, \dots, \theta_n)$ be the dual basis of $(e_1, \dots, e_n)$; if $\alpha = \{j_1, \dots, j_p\} \in I_p$, consider the p-form

$$\Theta_\alpha \doteq \theta_{j_1} \wedge \cdots \wedge \theta_{j_p}.$$

Then $\{\Theta_\alpha\}_{\alpha \in I_p}$ is an orthonormal basis of $\Lambda^p(\Sigma)$. From the definition (4) one sees that S is extended as a derivation of $\Lambda^\star(\Sigma)$, hence Θ_α is an eigenform of $S^{[p]}$ associated to the eigenvalue K_α:

$$S^{[p]}\Theta_\alpha = K_\alpha \Theta_\alpha.$$

In turn, if $\alpha = \{j_1, \dots, j_p\} \in I_p$, let $\star\alpha \in I_{n-p}$ be the multi-index given by the complement of α in $\{1, \dots, n\}$:

$$\star\alpha \doteq \{1, \dots, n\} \setminus \{j_1, \dots, j_p\}.$$

Let $T^{[p]} = (\operatorname{tr} S)S^{[p]} - S^{[p]} \circ S^{[p]}$. Since $K_\alpha + K_{\star\alpha} = k_1 + \cdots + k_n$, we have

$$T^{[p]}\Theta_\alpha = \big((k_1 + \cdots + k_n)K_\alpha - K_\alpha^2\big)\Theta_\alpha = K_\alpha K_{\star\alpha}\Theta_\alpha$$

and then Θ_α is also an eigenform of $T^{[p]}$ associated to the eigenvalue $K_\alpha K_{\star\alpha}$. In conclusion, we have:

Lemma 2. *Let $S = S_\nu$ be the shape operator relative to a given unit normal vector $\nu \in T^\perp\Sigma$. Let $T^{[p]} : \Lambda^p(\Sigma) \to \Lambda^p(\Sigma)$ be the operator*

$$T^{[p]} = (\operatorname{tr} S)S^{[p]} - S^{[p]} \circ S^{[p]}.$$

Then the eigenvalues of $T^{[p]}$ are given by the $\binom{n}{p}$ numbers $K_\alpha K_{\star\alpha}$, where $\alpha \in I_p$ runs over the set of p-multi-indices and K_α are the p-curvatures defined in (6).

In particular, if Σ is a hypersurface of M^{n+1}, and S is the shape operator of Σ relative to any of the two choices of the unit normal vector, then $T^{[p]} = \mathcal{B}^{[p]}_{\text{ext}}$ and

$$\min_{\alpha \in I_p} K_\alpha K_{\star\alpha} \le \mathcal{B}^{[p]}_{\text{ext}} \le \max_{\alpha \in I_p} K_\alpha K_{\star\alpha}. \tag{7}$$

In higher codimensions, in order to estimate $\mathcal{B}^{[p]}_{\text{ext}}$ it is enough to estimate the p-curvatures of the shape operators S_{ν_j} for an orthonormal frame $(\nu_1, \dots, \nu_q)$ in the normal bundle of the immersion. It will then be possible to bound $\mathcal{B}^{[p]}_{\text{ext}}$ in terms of $|S|^2$ and $|H|^2$ (see Section 3).

Let us briefly explain why Theorem 1 improves the bound (2). For simplicity, assume that Σ^n is a hypersurface of $\mathbb{R}^{n+1}$, with principal curvatures $k_1, \dots, k_n$. It is known that the curvature operator of Σ has eigenvalues $\{k_i k_j : i \neq j\}$ (see also Section 5.2). Hence

$$\gamma_\Sigma = \inf\{k_i k_j : i \neq j\}.$$

Now observe that since $\mathcal{B}^{[p]}_{\text{res}} = 0$, we have $\mathcal{B}^{[p]} = \mathcal{B}^{[p]}_{\text{ext}}$; if $\alpha \in I_p$ is a multi-index, then $K_\alpha K_{\star\alpha}$ is a sum of $p(n-p)$ products of type $k_i k_j$ with $i \neq j$. Then (7) gives

$$\mathcal{B}^{[p]} \geq \min_{\alpha \in I_p} K_\alpha K_{\star\alpha} \geq p(n-p)\gamma_\Sigma,$$

the expression on the right being the lower bound in (2). Numerical examples show that the lower bound in (7) is often much better than (2).

2.1. ***The Bochner method.*** When the manifold Σ is compact, lower bounds of the Bochner curvature lead to lower bounds of the Hodge–Laplace spectrum. Here we recall the main facts. Let $\lambda_{1,p}(\Sigma)$ be the lowest eigenvalue of the Hodge–Laplace operator acting on p-forms of Σ. It is well known that

$$\lambda_{1,p}(\Sigma) = \inf\left\{ \frac{\int_\Sigma |d\omega|^2 + |\delta\omega|^2}{\int_\Sigma |\omega|^2} : \omega \in \Lambda^p(\Sigma) \setminus \{0\} \right\}. \tag{8}$$

As Σ is orientable and the quadratic form in (8) is invariant under the Poincaré duality induced by the Hodge $\star$-operator, we have $\lambda_{1,p} = \lambda_{1,n-p}$ and so we can assume that $p \leq n/2$. Clearly $\lambda_{1,p} = 0$ if and only if Σ supports a nontrivial harmonic p-form, in which case $H^p(\Sigma, \mathbb{R}) = H^{n-p}(\Sigma, \mathbb{R}) \neq 0$ (the Hodge–de Rham theorem).

The next proposition is well known and is often called the *Bochner method* (the estimate (iii) follows from a lower estimate of the energy of a form due to Gallot and Meyer [1975]). We state it in the following form for future reference.

Proposition 3. *Let Σ^n be a compact, orientable manifold, $1 \leq p \leq n/2$ and $\mathcal{B}^{[p]}$ the curvature term in the Bochner formula (1).*

(i) *If $\mathcal{B}^{[p]} \geq 0$ and the strict inequality holds at some point, then we have $H^p(\Sigma, \mathbb{R}) = H^{n-p}(\Sigma, \mathbb{R}) = 0$.*

(ii) *If $\mathcal{B}^{[p]} \geq 0$ and $H^p(\Sigma, \mathbb{R}) \neq 0$, then any harmonic p-form is parallel. In particular, Σ^n supports a parallel p-form.*

(iii) *If $\mathcal{B}^{[p]} \geq p(n-p)\Lambda$ for some $\Lambda > 0$, then*

$$\lambda_{1,p}(\Sigma) \geq p(n-p+1)\Lambda.$$

Proof. Let ω be a p-form. Taking the scalar product with ω on both sides of (1), we obtain

$$\langle \Delta\omega, \omega \rangle = |\nabla\omega|^2 + \langle \mathcal{B}^{[p]}\omega, \omega \rangle + \tfrac{1}{2}\Delta|\omega|^2. \tag{9}$$

Integrating on Σ (with respect to the canonical Riemannian measure), we get

$$\int_\Sigma \langle \Delta\omega, \omega\rangle = \int_\Sigma \big(|\nabla\omega|^2 + \langle \mathcal{B}^{[p]}\omega, \omega\rangle\big). \tag{10}$$

If we assume that $\mathcal{B}^{[p]} \geq 0$ and that ω is harmonic, we get

$$0 = \int_\Sigma \big(|\nabla\omega|^2 + \langle \mathcal{B}^{[p]}\omega, \omega\rangle\big) \geq 0,$$

which implies $|\nabla\omega| = 0$ and $\langle \mathcal{B}^{[p]}\omega, \omega\rangle = 0$ everywhere. It is well known that a harmonic form cannot vanish on an open set unless it is zero everywhere. If, at some point $x_0 \in \Sigma$, the strict inequality $\mathcal{B}^{[p]} > 0$ holds, we see that ω must vanish in a neighborhood of x_0, hence $\omega = 0$ everywhere. This proves (i). Assertion (ii) is immediate.

We now prove (iii). Let ω be a p-eigenform, so that $\Delta\omega = \lambda_{1,p}\omega$. By an inequality of Gallot and Meyer [1975] we have

$$|\nabla\omega|^2 \geq \frac{|d\omega|^2}{p+1} + \frac{|\delta\omega|^2}{n-p+1}.$$

Since $p \leq n/2$ we see that $p+1 \leq n-p+1$, hence

$$\int_\Sigma |\nabla\omega|^2 \geq \int_\Sigma \frac{|d\omega|^2 + |\delta\omega|^2}{n-p+1} = \frac{\lambda_{1,p}}{n-p+1}\int_\Sigma |\omega|^2.$$

Inserting this in (10) and using the lower bound on $\mathcal{B}^{[p]}$ we arrive easily at

$$\lambda_{1,p}\int_\Sigma |\omega|^2 \geq p(n-p+1)\Lambda \int_\Sigma |\omega|^2,$$

which gives the assertion. □

We already remarked that for $p = 1$, the Bochner operator $\mathcal{B}^1$ is simply given by the Ricci tensor acting on 1-forms. In particular, when $n = 2$, $\mathcal{B}^{[1]}$ is a scalar operator and is given by multiplication by the Gaussian curvature K_Σ of Σ: $\mathcal{B}^{[1]}\omega = K_\Sigma\omega$.

2.2. ***Rigidity of Clifford tori.*** For $p = 1, \ldots, n-1$ and $r \in (0, 1)$, consider the manifold (Clifford torus)

$$T_{p,r} \doteq \mathbb{S}^p(r) \times \mathbb{S}^{n-p}\big(\sqrt{1-r^2}\big) \tag{11}$$

which is naturally isometrically embedded as a hypersurface in $\mathbb{S}^{n+1}$. Note that $T_{p,r}$ supports a parallel p-form, which is the pullback of the volume form of $\mathbb{S}^p(r)$ by the projection onto the first factor. We have the following rigidity theorem (when Σ is minimal, it reduces to [Colbois and Savo 2012, Theorem 10]).

Theorem 4. *Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$ supporting a (nontrivial) parallel p-form for some $p = 1, \dots, n-1$.*

(a) *If $n = 2$ and $p = 1$, then Σ is a flat 2-torus.*

(b) *If $n \geq 3$, then Σ is isometric to a Clifford torus $T_{p,r}$ for some $r \in (0, 1)$.*

For the proof, see Section 6. The interest in the theorem lies in the case $n \geq 3$. In fact, (a) holds for any compact, orientable surface supporting a parallel 1-form. We remark that there exist flat 2-tori in $\mathbb{S}^3$ which are not isometric to any Clifford torus $T_{1,r}$ (see [Weiner 1991] for a classification result).

Combining Theorem 4 with Proposition 3(ii), we obtain:

Corollary 5. *Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$ such that $H^p(\Sigma, \mathbb{R}) \neq 0$ and $\mathcal{B}^{[p]} \geq 0$ for some $p = 1, \dots, n-1$. Then Σ is a flat torus if $n = 2$ and, if $n \geq 3$, it is isometric to a Clifford torus $T_{p,r}$ for some $r \in (0, 1)$.*

We also record the following consequence.

Corollary 6. *Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$ having nonnegative sectional curvature. If $n \geq 3$, then either Σ is a homology sphere (i.e., $H^p(\Sigma, \mathbb{R}) = 0$ for all $p = 1, \dots, n-1$) or Σ is isometric to a Clifford torus.*

Proof. Let $(e_1, \dots, e_n)$ be an orthonormal basis of principal directions and let $k_1, \dots, k_n$ be the associated principal curvatures of Σ. Let $i \neq j$. By the Gauss formula $R(e_i, e_j, e_i, e_j) = 1 + k_i k_j$ and then, by our assumptions,

$$k_i k_j \geq -1.$$

If $\alpha \in I_p$ is any multi-index, we observe that $K_\alpha K_{\star\alpha}$ is a sum of $p(n-p)$ products of type $k_i k_j$ with $i \neq j$. Hence

$$K_\alpha K_{\star\alpha} \geq -p(n-p),$$

and by (7) we obtain $\mathcal{B}^{[p]}_{\text{ext}} \geq -p(n-p)$. As $\mathcal{B}^{[p]}_{\text{res}} = p(n-p)$ we conclude by Theorem 1 that $\mathcal{B}^{[p]} \geq 0$ for all $p = 1, \dots, n-1$. If $H^p(\Sigma, \mathbb{R}) \neq 0$ we see that Σ is a Clifford torus by Corollary 5. This completes the proof. □

3. Applications

3.1. ***Applications in codimension one.*** Let Σ^n be a hypersurface of M^{n+1} and S its shape operator. Fix a point $x \in \Sigma$ and let $(k_1, \dots, k_n)$ be the principal curvatures of Σ at x. The (scalar) mean curvature is denoted by

$$H = \frac{1}{n}(k_1 + \cdots + k_n).$$

For each multi-index $\alpha = \{j_1, \ldots, j_p\} \in I_p$, let $K_\alpha(x)$ be the corresponding p-curvature, as defined in (6). Set

$$\begin{cases} \beta_p(x) = \dfrac{1}{p(n-p)} \displaystyle\inf_{\alpha \in I_p} K_\alpha(x) K_{\star\alpha}(x), \\ \beta_p(\Sigma) = \displaystyle\inf_{x \in \Sigma} \beta_p(x). \end{cases} \tag{12}$$

From (7) we then see that at all points of Σ one has $\mathcal{B}^{[p]}_{\text{ext}} \geq p(n-p)\beta_p(\Sigma)$. If γ_M is a lower bound of the curvature operator of M^{n+1} then $\mathcal{B}^{[p]}_{\text{res}} \geq p(n-p)\gamma_M$, and then by Theorem 1

$$\mathcal{B}^{[p]} \geq p(n-p)(\gamma_M + \beta_p(\Sigma)). \tag{13}$$

From (13) and Proposition 3(iii) we immediately get the following estimate.

Theorem 7. *Let Σ^n be a compact hypersurface of M^{n+1}, a manifold with curvature operator bounded below by $\gamma_M \in \mathbb{R}$. Let $1 \leq p \leq n/2$. Then*

$$\lambda_{1,p}(\Sigma) \geq p(n-p+1)(\gamma_M + \beta_p(\Sigma)), \tag{14}$$

where $\beta_p(\Sigma)$ is defined by (12). If Σ is a geodesic sphere in a simply connected manifold of constant curvature γ_M, then equality holds.

To verify that the inequality is sharp, let $M^{n+1}(c)$ be the simply connected manifold of constant curvature $c = \gamma_M$. Then $M^{n+1}(c)$ is, respectively, the Euclidean space $\mathbb{R}^{n+1}$ when $c = 0$, the unit sphere $\mathbb{S}^{n+1}$ when $c = 1$ and the hyperbolic space $\mathbb{H}^{n+1}$ when $c = -1$. If Σ is a geodesic sphere in $M^{n+1}(c)$ then Σ is totally umbilical: $K_\alpha(x) = pH$, $K_{\star\alpha}(x) = (n-p)H$, where H is the mean curvature of Σ. Hence $\beta_p(x) = H^2$ and (14) becomes

$$\lambda_{1,p}(\Sigma) \geq p(n-p+1)(c + H^2).$$

On the other hand, Σ is known to be isometric to $\mathbb{S}^n(r)$, with $r = (c + H^2)^{-\frac{1}{2}}$. Therefore, by a well-known calculation in [Gallot and Meyer 1975],

$$\lambda_{1,p}(\Sigma) = p(n-p+1)(c + H^2).$$

This shows that Theorem 7 is sharp. Note also that the condition $\beta_p(\Sigma) > -\gamma_M$ implies $H^p(\Sigma, \mathbb{R}) = H^{n-p}(\Sigma, \mathbb{R}) = 0$. For an upper bound of $\lambda_{1,p}(\Sigma)$ when M is a sphere, see [Savo 2005].

We now compare this estimate with the results in [Raulot and Savo 2011]. Let us say that Σ is *p-convex* if we can choose an orientation of Σ so that all of its p-curvatures are nonnegative at every point: $K_\alpha(x) \geq 0$ for all $\alpha \in I_p$ and $x \in \Sigma$. This is equivalent to asking that the operator $S^{[p]}$ is nonnegative at every point.

Clearly, if Σ is p-convex, then it is q-convex for all $q \geq p$. In that case, let

$$\mathcal{K}_p = \frac{1}{p} \inf_{x\in\Sigma} \inf_{\alpha\in I_p} K_\alpha(x) \geq 0$$

be a lower bound of the *mean p-curvatures* (note that $\mathcal{K}_1$ is a lower bound of the principal curvatures, while $\mathcal{K}_n$ is a lower bound of the mean curvature). It is easy to prove that $\mathcal{K}_p \leq \mathcal{K}_q$ whenever $p \leq q$. Using a Reilly-type formula for differential forms, it is proved in [Raulot and Savo 2011] that if Σ is the boundary of a domain in a manifold M^{n+1} with nonnegative curvature operator ($\gamma_M \geq 0$), and if Σ is p-convex (for the orientation given by the inner unit normal), then

$$\lambda_{1,p}(\Sigma) \geq p(n-p+1)\mathcal{K}_p\mathcal{K}_{n-p+1} \tag{15}$$

with equality if and only if Σ is a sphere in $\mathbb{R}^{n+1}$. Numerical examples show that, in that situation, (14) is often better than (15). Moreover, (14) applies also in negative curvature, and for immersions which are not necessarily embeddings. For example, if M^{n+1} is the hyperbolic space $\mathbb{H}^{n+1}$, it is easy to see that the inequality gives a positive lower bound whenever $\mathcal{K}_p > 1$.

The next application is inspired by the following result of [Alencar and do Carmo 1994]. Assume that Σ^n is a compact hypersurface of $\mathbb{S}^{n+1}$ with constant mean curvature H. Let

$$\Phi = S - HI$$

be the *traceless* second fundamental form of Σ, and let $R(1, H)$ be the positive root of the polynomial

$$F_1(x) = x^2 + \frac{n(n-2)}{\sqrt{n(n-1)}}|H|x - n(H^2+1). \tag{16}$$

Alencar and do Carmo [1994] proved that if $|\Phi| \leq R(1, H)$, then Σ is either totally umbilical ($|\Phi| = 0$) or a Clifford torus

$$T_{1,r} = \mathbb{S}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1-r^2})$$

with $r \geq \sqrt{1/n}$ (in which case $|\Phi| = R(1, H)$). To the best of our knowledge, there is still no similar characterization of the other Clifford tori $T_{p,r}$ for $2 \leq p \leq n-2$ among constant mean curvature hypersurfaces of the sphere.

We prove a vanishing result for the de Rham cohomology in degree p assuming that the norm of the traceless second fundamental form is bounded above by a suitable function of the mean curvature. When the ambient manifold is the sphere, this will lead to a topological version of the Alencar–do Carmo result, in which the assumption of constant mean curvature is replaced by an assumption of nontrivial cohomology in degree p.

More precisely, fix real numbers γ, H and an integer $1 \le p \le n/2$. Consider the polynomial function

$$F_p(x) = x^2 + \frac{n|n-2p|}{\sqrt{pn(n-p)}}|H|x - n(H^2+\gamma). \tag{17}$$

Note that $F_p(x)$ reduces to (16) when $p = \gamma = 1$. If $\gamma < 0$, we assume that $H^2 + \gamma \ge 0$. Then, the largest root of $F_p(x)$ is nonnegative, and will be denoted by $R(p, H)$. It is easy to check that for fixed H and γ one has

$$0 \le R(1, H) \le R(2, H) \le \cdots \le R\Big(\Big[\frac{n}{2}\Big], H\Big)$$

where $[n/2]$ is the largest integer less than or equal to $n/2$. If n is even, one has

$$R\Big(\frac{n}{2}, H\Big) = \sqrt{n(H^2+\gamma)}.$$

We then have the following result.

Theorem 8. *Let Σ^n be a compact hypersurface of M^{n+1}, a manifold with curvature operator bounded below by $\gamma \in \mathbb{R}$. We suppose $p \le n/2$ and $H^2 + \gamma \ge 0$. If*

$$|\Phi| \le R(p, H)$$

everywhere on Σ and strict inequality holds somewhere, then

$$H^k(\Sigma, \mathbb{R}) = 0 \quad \textit{for all } k = p, \dots, n-p.$$

Here Φ is the traceless second fundamental form of Σ, and $R(p, H) \ge 0$ is the largest root of (17). In particular, if $|\Phi| \le R(1, H)$, with strict inequality somewhere, then Σ is a homology sphere.

The proof is given in Section 4.1. Now let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$; we take $\gamma = 1$ and then consider the polynomial

$$F_p(x) = x^2 + \frac{n|n-2p|}{\sqrt{pn(n-p)}}|H|x - n(H^2+1). \tag{18}$$

We have the following rigidity result, which shows that the conditions in the previous theorem are sharp.

Theorem 9. *Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$, take $1 \le p \le n/2$ and denote by $R(p, H)$ the positive root of (18). We also assume $n \ge 3$.*

(a) *Let $1 \le p < n/2$. If*

$$|\Phi| \le R(p, H) \quad \textit{and} \quad H^p(\Sigma, \mathbb{R}) \ne 0,$$

then Σ is isometric to a Clifford torus $T_{p,r} = \mathbb{S}^p(r) \times \mathbb{S}^{n-p}\big(\sqrt{1-r^2}\big)$ for some $r \ge \sqrt{p/n}$.

(b) *Let n be even and $p = n/2$. If we assume that*

$$|\Phi|^2 \le n(1+H^2) \quad \textit{and} \quad H^p(\Sigma,\mathbb{R}) \neq 0,$$

then Σ is isometric to a Clifford torus $\mathbb{S}^p(r)\times\mathbb{S}^p\big(\sqrt{1-r^2}\big)$ for some $r \in (0,1)$.

For the proof, see Section 4.1.

3.2. ***Applications in arbitrary codimension.*** Lawson and Simons [1973] proved the following vanishing result. If Σ^n is a compact, immersed submanifold of a sphere $\mathbb{S}^{n+q}$ and if at all points of Σ one has

$$|S|^2 < \min\big\{p(n-p),\ 2\sqrt{p(n-p)}\big\}$$

for some $p = 1,\dots,n-1$, then $H^p(\Sigma,\mathbb{Z}) = H^{n-p}(\Sigma,\mathbb{Z}) = 0$. The proof depends on deep results of geometric measure theory, and used the fact that any integral homology class is represented by a stable current. Taking variations induced by suitable vector fields (namely, orthogonal projections to Σ of parallel vector fields on $\mathbb{R}^{n+2}$), one gets the stated result.

If we limit ourselves to real cohomology theory, we have another proof of this result by a completely different method (the Bochner method) as follows. Note that this also gives an explicit lower bound of the spectrum of the Hodge Laplacian and an associated rigidity result (Theorem 11).

Theorem 10. *Let Σ^n be a compact submanifold of M^{n+q}, a manifold with curvature operator bounded below by $\gamma_M > 0$, and let $1 \le p \le n/2$. If*

$$|S|^2 \le 2\gamma_M\sqrt{p(n-p)}$$

and strict inequality holds somewhere, then $H^k(\Sigma,\mathbb{R}) = 0$ for all $k = p,\dots,n-p$.

More generally, if

$$|S|^2 \le 2\gamma_M\sqrt{p(n-p)}(1-\Lambda)$$

for some $\Lambda \in (0,1]$, then $\lambda_{1,p}(\Sigma) \ge p(n-p+1)\gamma_M\Lambda$.

For the proof, see Section 4.3.

In codimension one the condition is sharp, and our approach gives the following rigidity result.

Theorem 11. *Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$, $n \ge 2$, such that*

$$|S|^2 \le 2\sqrt{p(n-p)} \quad \textit{and} \quad H^p(\Sigma,\mathbb{R}) \neq 0$$

for some $1 \le p \le n/2$. Then Σ is isometric to the Clifford torus $T_{p,r} = \mathbb{S}^p(r)\times\mathbb{S}^{n-p}\big(\sqrt{1-r^2}\big)$ for

$$r = \left(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{n-p}}\right)^{\frac{1}{2}}.$$

In fact, one can check that the Clifford torus of Theorem 11 minimizes $|S|^2$ among the family of Clifford tori $\{T_{p,r} : r \in (0, 1)\}$. The proof is given in Section 4.4.

3.3. ***Submanifolds with a parallel p-form.*** In our final application, we study immersions of a manifold supporting a (nontrivial) parallel p-form (our estimates are local, and so we do not assume compactness). Noteworthy examples of such manifolds are given by:

- Riemannian products $N_1 \times N_2$ (having parallel forms in degrees $p = \dim N_1$, $\dim N_2$).
- Kähler manifolds.

In fact, the Kähler 2-form Ω is parallel. As all powers of Ω are nontrivial (and parallel), we see that a Kähler manifold supports nontrivial parallel forms in all even degrees.

It is well known that a Kähler manifold does not admit any minimal immersion into a hyperbolic space (see [Dajczer and Rodríguez 1986; El Soufi and Petit 2000]). More generally, in [Grosjean 2004], it is proved that a manifold Σ with a parallel p-form does not admit any minimal immersion into a manifold M if certain curvature conditions on Σ and M are met. We also refer to [Grosjean 2004] for other rigidity results on minimal immersions.

Our point of view is to observe that if ω is a parallel p-form, then $\langle \mathcal{B}^{[p]}\omega, \omega\rangle = 0$ everywhere on Σ. More generally, if ω is a harmonic p-form with constant length, then, from the Bochner formula (9),

$$\langle \mathcal{B}^{[p]}\omega, \omega\rangle = -|\nabla\omega|^2 \leq 0.$$

Using the pointwise bounds on the eigenvalues of $\mathcal{B}^{[p]}$ derived in Section 4, we then obtain pointwise bounds for the extrinsic geometry of Σ. Precisely:

Theorem 12. *Let Σ^n be an immersed submanifold of M^{n+q} and let $p = 1, \ldots, n-1$.*

(a) *If Σ supports a parallel p-form and M has curvature operator bounded above by $\Gamma_M < 0$ then*

$$|H|^2 \geq \frac{4p(n-p)}{n^2}|\Gamma_M| \quad \text{and} \quad |S|^2 \geq 2|\Gamma_M|\sqrt{p(n-p)}$$

at all points of Σ.

(b) *If Σ supports a harmonic p-form of constant length (in particular, a parallel p-form) and M has curvature operator bounded below by $\gamma_M > 0$ then*

$$|S|^2 \geq 2\gamma_M\sqrt{p(n-p)}$$

at all points of Σ.

For the proof, see Section 4.5. Assertion (a) is sharp when $p = n/2$ (see below) and (b) is sharp for the Clifford torus of Theorem 11.

Now let Σ^{2m} be a Kähler manifold of complex dimension $m \geq 2$. Since Σ supports nontrivial parallel forms in all even degrees, we see that it supports a parallel form of degree m or $m-1$ depending on whether m is even or odd. Applying the previous theorem we immediately get the following estimates.

Corollary 13. *Let Σ^{2m} be a Kähler manifold of complex dimension $m \geq 2$ isometrically immersed in the Riemannian manifold M^{2m+q}.*

(a) *If M has curvature operator bounded above by $\Gamma_M < 0$, then at all points of Σ*

$$|H|^2 \geq \begin{cases} |\Gamma_M| & \text{if } m \text{ is even,} \\ \dfrac{m^2-1}{m^2}|\Gamma_M| & \text{if } m \text{ is odd.} \end{cases}$$

(b) *If M has curvature operator bounded below by $\gamma_M > 0$, then*

$$|S^2| \geq \begin{cases} 2m\gamma_M & \text{if } m \text{ is even,} \\ 2\gamma_M\sqrt{m^2-1} & \text{if } m \text{ is odd.} \end{cases}$$

We remark that if m is even and the ambient space is the hyperbolic space $\mathbb{H}^{2m+q}$ then (a) gives

$$|H|^2 \geq 1,$$

which is an equality when $\Sigma^{2m} = \mathbb{R}^{2m}$, embedded in $\mathbb{H}^{2m+1}$ as a horosphere.

4. Estimates of the Bochner operator

In this section we first estimate the extrinsic part of the Bochner operator, thanks to Lemma 2 and some elementary algebra. We then apply these estimates to prove the theorems of Section 3. Let Σ^n be a submanifold of the Riemannian manifold M^{n+q}. We start from the following algebraic lemma.

Lemma 14. *Let $S = S_\nu$ be the shape operator of Σ relative to a unit normal vector $\nu \in T^\perp\Sigma$, and let $T^{[p]} = (\operatorname{tr} S)S^{[p]} - S^{[p]} \circ S^{[p]}$. If $k_1, \dots, k_n$ are the eigenvalues of S, set*

$$nH = \sum_{j=1}^n k_j, \quad |S|^2 = \sum_{j=1}^n k_j^2, \quad |\Phi|^2 = \sum_{j=1}^n (k_j - H)^2. \tag{19}$$

Then the following inequalities for $T^{[p]}$ hold. Recall that if Σ has codimension one, then $T^{[p]} = \mathcal{B}^{[p]}_{\text{ext}}$.

(a) $-\dfrac{\sqrt{p(n-p)}}{2}|S|^2 \leq T^{[p]} \leq \dfrac{\sqrt{p(n-p)}}{2}|S|^2.$

(b) *If* $H = 0$, *then*

$$-\frac{p(n-p)}{n}|S|^2 \leq T^{[p]} \leq 0.$$

(c) $\dfrac{n^2|H|^2}{4} - \dfrac{1}{4}\left(|n-2p||H| + \sqrt{\dfrac{4p(n-p)}{n}}|\Phi|\right)^2 \leq T^{[p]} \leq \dfrac{n^2|H|^2}{4}.$

(d) *If n is even and $p = n/2$, then*

$$T^{[p]} \geq \tfrac{1}{4}n^2|H|^2 - \tfrac{1}{4}n|\Phi|^2.$$

Proof. We know from Lemma 2 that the eigenvalues of $T^{[p]}$ are given by $\lambda_\alpha = K_\alpha K_{\star\alpha}$, where α runs over the set of p-multi-indices. Then, it is enough to show the inequalities for any such eigenvalue. Fix a multi-index α. After reordering, we can assume that

$$\lambda_\alpha = (k_1 + \cdots + k_p)(k_{p+1} + \cdots + k_n). \tag{20}$$

In conclusion, it is enough to show the given bounds for the product (20), given any set of real numbers $k_1, \ldots, k_n$ satisfying (19).

Proof of (a). We use the Schwarz inequality and the inequality $\sqrt{ab} \leq (a+b)/2$ applied to $a = k_1^2 + \cdots + k_p^2$, $b = k_{p+1}^2 + \cdots + k_n^2$. We obtain

$$\begin{aligned}|\lambda_\alpha| &= |k_1 + \cdots + k_p||k_{p+1} + \cdots + k_n| \\ &\leq \sqrt{p(n-p)}\sqrt{k_1^2 + \cdots + k_p^2}\sqrt{k_{p+1}^2 + \cdots + k_n^2} \leq \tfrac{1}{2}(\sqrt{p(n-p)})|S|^2,\end{aligned}$$

and (a) follows.

Proof of (b). Since $k_{p+1} + \cdots + k_n = -(k_1 + \cdots + k_p)$, the Schwarz inequality yields

$$\begin{aligned}\lambda_\alpha &= -(k_1 + \cdots + k_p)^2 \geq -p\big(k_1^2 + \cdots + k_p^2\big), \\ \lambda_\alpha &= -(k_{p+1} + \ldots + k_n)^2 \geq -(n-p)\big(k_{p+1}^2 + \cdots + k_n^2\big).\end{aligned}$$

Summing the two inequalities, we get

$$\frac{\lambda_\alpha}{p} + \frac{\lambda_\alpha}{n-p} \geq -|S|^2,$$

from which the lower bound follows. The upper bound is obvious.

Proof of (c). As $\sum_{j=1}^{n}(k_j - H) = 0$, we see that, by the lower bound in (b),

$$((k_1 - H) + \cdots + (k_p - H))((k_{p+1} - H) + \cdots + (k_n - H)) \geq -\frac{p(n-p)}{n}|\Phi|^2.$$

Hence

$$
\begin{aligned}
(21)\quad & -\frac{p(n-p)}{n}|\Phi|^2 \\
&\le (k_1+\cdots+k_p-pH)(k_{p+1}+\cdots+k_n-(n-p)H) \\
&= \lambda_\alpha + p(n-p)H^2 - pH(k_{p+1}+\cdots+k_n) - (n-p)H(k_1+\cdots+k_p).
\end{aligned}
$$

Substituting $k_{p+1}+\cdots+k_n = nH-(k_1+\cdots+k_p)$ in (21), we have

$$
(22)\qquad -\frac{p(n-p)}{n}|\Phi|^2 \le \lambda_\alpha - p^2H^2 - (n-2p)H(k_1+\cdots+k_p).
$$

Substituting $k_1+\cdots+k_p = nH-(k_{p+1}+\cdots+k_n)$ in (21), we also have

$$
(23)\qquad -\frac{p(n-p)}{n}|\Phi|^2 \le \lambda_\alpha - (n-p)^2H^2 + (n-2p)H(k_{p+1}+\cdots+k_n).
$$

We now sum (22) and (23) to obtain

$$
\begin{aligned}
(24)\quad & 2\lambda_\alpha - \big(p^2+(n-p)^2\big)H^2 + \frac{2p(n-p)}{n}|\Phi|^2 \\
&\ge (n-2p)H((k_1+\cdots+k_p)-(k_{p+1}+\cdots+k_n)) \\
&\ge -|n-2p||H||(k_1+\cdots+k_p)-(k_{p+1}+\cdots+k_n)|.
\end{aligned}
$$

Set $a=k_1+\cdots+k_p$, $b=k_{p+1}+\cdots+k_n$. As $|a-b|^2=(a+b)^2-4ab$, we see that

$$
n^2H^2-4\lambda_\alpha \ge 0,
$$

which is the upper bound in (c), and $|a-b|=\sqrt{n^2H^2-4\lambda_\alpha}$. Substituting in (24),

$$
(25)\quad 2\lambda_\alpha - \big(p^2+(n-p)^2\big)H^2 + \frac{2p(n-p)}{n}|\Phi|^2 \ge -|n-2p||H|\sqrt{n^2H^2-4\lambda_\alpha}.
$$

If we set $\delta=\sqrt{n^2H^2-4\lambda_\alpha}\ge 0$, then (25) takes the form

$$
\delta^2 - 2|n-2p||H|\delta + (n-2p)^2H^2 - \frac{4p(n-p)}{n}|\Phi|^2 \le 0,
$$

which implies

$$
\delta \le |n-2p||H| + \sqrt{\frac{4p(n-p)}{n}}|\Phi|.
$$

Recalling the definition of δ, one concludes that

$$
4\lambda_\alpha \ge n^2H^2 - \left(|n-2p||H| + \sqrt{\frac{4p(n-p)}{n}}|\Phi|\right)^2,
$$

which is the lower bound in (c). Finally, (d) is a particular case of (c). □

4.1. *Proofs of Theorems 8 and 9.* Let Σ^n be a hypersurface of M^{n+1}, a manifold with curvature operator bounded below by $\gamma \in \mathbb{R}$. Let H be the mean curvature of Σ and Φ its traceless second fundamental form. Recall that

$$F_p(x) = x^2 + \frac{n|n-2p|}{\sqrt{pn(n-p)}}|H|x - n(H^2+\gamma).$$

The theorems are a result of the following bound for $\mathcal{B}^{[p]}$.

Proposition 15. *In the above notation, one has*

$$\mathcal{B}^{[p]} \geq -\frac{p(n-p)}{n} F_p(|\Phi|).$$

Proof. By Lemma 14(c) we have

$$\mathcal{B}^{[p]}_{\text{ext}} = T^{[p]} \geq \frac{n^2|H|^2}{4} - \frac{1}{4}\left(|n-2p||H| + \sqrt{\frac{4p(n-p)}{n}}|\Phi|\right)^2.$$

Then, as $\mathcal{B}^{[p]}_{\text{res}} \geq p(n-p)\gamma$ and $\mathcal{B}^{[p]} = \mathcal{B}^{[p]}_{\text{res}} + \mathcal{B}^{[p]}_{\text{ext}}$,

$$\begin{aligned}
4\mathcal{B}^{[p]} &\geq 4p(n-p)\gamma + 4\mathcal{B}^{[p]}_{\text{ext}} \\
&\geq 4p(n-p)\gamma + n^2|H|^2 - \left(|n-2p||H| + \sqrt{\frac{4p(n-p)}{n}}|\Phi|\right)^2 \\
&= 4p(n-p)\gamma + 4p(n-p)|H|^2 - \frac{4p(n-p)}{n}|\Phi|^2 \\
&\qquad\qquad\qquad - 2|n-2p|\sqrt{\frac{4p(n-p)}{n}}|\Phi||H| \\
&= \frac{4p(n-p)}{n}\left(n\big(\gamma + |H|^2\big) - |\Phi|^2 - \frac{n|n-2p|}{\sqrt{pn(n-p)}}|\Phi||H|\right) \\
&= -\frac{4p(n-p)}{n} F_p(|\Phi|)
\end{aligned}$$

and the assertion follows. □

Proof of Theorem 8. By assumption, $|\Phi| \leq R(p, H)$, hence $F_p(|\Phi|) \leq 0$ by the definition of $R(p, H)$. By Proposition 15 we see that

$$\mathcal{B}^{[p]} \geq 0.$$

As the inequality is strict somewhere, we can apply the Bochner method and conclude that $H^p(\Sigma, \mathbb{R}) = 0$. By Poincaré duality, one has also $H^{n-p}(\Sigma, \mathbb{R}) = 0$. Since $R(p, H)$ is nondecreasing in p, we see that $|\Phi| \leq R(k, H)$ for all $p \leq k \leq [n/2]$ and the conclusion follows. □

Proof of Theorem 9. Under the given assumptions, one has $\mathcal{B}^{[p]} \geq 0$ and $H^p(\Sigma, \mathbb{R}) \neq 0$. Then, Σ is a Clifford torus by Corollary 5. Conversely, the Clifford torus $T_{p,r}$ obviously satisfies $H^p(T_{p,r}) \neq 0$. Moreover, $T_{p,r}$ is known to have two distinct principal curvatures given (up to sign) by

$$\begin{cases} \lambda = \dfrac{\sqrt{1-r^2}}{r} & \text{with multiplicity } p, \\ \mu = -\dfrac{r}{\sqrt{1-r^2}} & \text{with multiplicity } n-p. \end{cases} \tag{26}$$

Therefore

$$|H| = \frac{|nr^2 - p|}{nr\sqrt{1-r^2}} \quad \text{and} \quad |\Phi| = \sqrt{\frac{p(n-p)}{n}} \frac{1}{r\sqrt{1-r^2}}.$$

A straightforward calculation shows that if $p < n/2$ and $r^2 \geq p/n$, then $F_p(|\Phi|) = 0$, that is, $|\Phi| = R(p, H)$. If $p = n/2$ then $|\Phi| = R(p, H) = \sqrt{n(1+H^2)}$ for all $r \in (0, 1)$. The proof is complete. $\square$

4.2. *An estimate in higher codimensions.*

Proposition 16. *Let Σ^n be a submanifold of the manifold M^{n+q} having curvature operator bounded below by γ_M. Then*

$$\mathcal{B}^{[p]} \geq p(n-p)\left(\gamma_M - \frac{|S|^2}{2\sqrt{p(n-p)}}\right).$$

If M has curvature operator bounded above by Γ_M, then

$$\mathcal{B}^{[p]} \leq p(n-p)\left(\Gamma_M + \frac{|S|^2}{2\sqrt{p(n-p)}}\right) \quad \text{and} \quad \mathcal{B}^{[p]} \leq p(n-p)\Gamma_M + \frac{n^2|H|^2}{4}.$$

Proof. Let $(\nu_1, \ldots, \nu_q)$ be an orthonormal frame in the normal bundle. We know from the main theorem that

$$\mathcal{B}^{[p]}_{\text{ext}} = \sum_{j=1}^{q} T^{[p]}_{\nu_j}, \quad \text{where } T^{[p]}_{\nu_j} = (\operatorname{tr} S_{\nu_j}) S^{[p]}_{\nu_j} - S^{[p]}_{\nu_j} \circ S^{[p]}_{\nu_j}.$$

If $\lambda_1\big(T^{[p]}_{\nu_j}\big)$ denotes the lowest eigenvalue of $T^{[p]}_{\nu_j}$, we see that $\mathcal{B}^{[p]}_{\text{ext}} \geq \sum_{j=1}^{q} \lambda_1(T^{[p]}_{\nu_j})$. From Lemma 14(a) applied to $S = S_{\nu_j}$, we obtain

$$\lambda_1(T^{[p]}_{\nu_j}) \geq -\frac{\sqrt{p(n-p)}}{2}|S_{\nu_j}|^2.$$

Summing over j, we get

$$\mathcal{B}^{[p]}_{\text{ext}} \geq -\frac{\sqrt{p(n-p)}}{2}|S|^2.$$

From Theorem 1 and the above,

$$\begin{aligned}\mathcal{B}^{[p]} &= \mathcal{B}^{[p]}_{\mathrm{res}} + \mathcal{B}^{[p]}_{\mathrm{ext}} \\ &\geq p(n-p)\gamma_M - \frac{\sqrt{p(n-p)}}{2}|S|^2 \\ &= p(n-p)\left(\gamma_M - \frac{|S|^2}{2\sqrt{p(n-p)}}\right),\end{aligned}$$

as asserted. The other inequalities are proved similarly, using Lemma 14. □

4.3. *Proof of Theorem 10.* If

$$|S|^2 \leq 2\gamma_M\sqrt{p(n-p)}\,(1-\Lambda),$$

for some $\Lambda \in [0, 1]$, we see from the first estimate of Proposition 16 that

$$\mathcal{B}^{[p]} \geq p(n-p)\Lambda\gamma_M.$$

The assertions follow immediately from the Bochner method (Proposition 3). □

4.4. *Proof of Theorem 11.* Let $n \geq 3$. Together with Proposition 16, the assumptions give $\mathcal{B}^{[p]} \geq 0$; as $H^p(\Sigma, \mathbb{R}) \neq 0$, we get immediately that Σ must be a Clifford torus $T_{p,r}$ by Corollary 5. On the other hand, it is seen from (26) that the only Clifford torus satisfying $|S|^2 \leq 2\sqrt{p(n-p)}$ is the one corresponding to the stated value of r.

Now assume $n = 2$ and $p = 1$, so that $|S|^2 \leq 2$. We know that $\mathcal{B}^{[1]}$ is multiplication by the Gaussian curvature K_Σ of Σ. From the formula $4H^2 = |S|^2 + 2K_\Sigma - 2$ we obtain $K_\Sigma \geq 2H^2 \geq 0$. The assumption $H^1(\Sigma, \mathbb{R}) \neq 0$ and the Bochner formula force $K_\Sigma = 0$, hence Σ is a minimal flat torus. As such, it is isometric with $\mathbb{S}^1(1/\sqrt{2}) \times \mathbb{S}^1(1/\sqrt{2})$ and the assertion follows. □

4.5. *Proof of Theorem 12.* Assume that ω is a parallel p-form. Then $\mathcal{B}^{[p]}\omega = 0$ identically and, by the last upper bound in Proposition 16,

$$0 = \langle \mathcal{B}^{[p]}\omega, \omega\rangle \leq \left(p(n-p)\Gamma_M + \frac{n^2|H|^2}{4}\right)|\omega|^2.$$

As $|\omega|$ is a positive constant, the assertion follows.

Now assume that ω is a harmonic p-form with constant length. From the Bochner formula (9) we see that $\langle \mathcal{B}^{[p]}\omega, \omega\rangle = -|\nabla\omega|^2 \leq 0$ at every point. Hence, applying Proposition 16,

$$0 \geq \langle \mathcal{B}^{[p]}\omega, \omega\rangle \geq p(n-p)\left(\gamma_M - \frac{|S|^2}{2\sqrt{p(n-p)}}\right)|\omega|^2,$$

which implies $|S|^2 \geq 2\gamma_M\sqrt{p(n-p)}$ everywhere, as asserted. □

5. Proof of the main theorem

Let Σ^n be a Riemannian manifold and R its curvature tensor, defined on tangent vectors X, Y by

$$R(X,Y) = -\nabla_X\nabla_Y + \nabla_Y\nabla_X + \nabla_{[X,Y]}. \tag{27}$$

The Bochner curvature term acting on p-forms is given by

$$\mathcal{B}^{[p]}\omega(X_1,\dots,X_p) = \sum_{j=1}^{n}\sum_{k=1}^{p}(-1)^k (R(e_j,X_k)\omega)(e_j,X_1,\dots,\hat{X}_k,\dots,X_p)$$

where $(e_1,\dots,e_n)$ is an orthonormal frame in $T\Sigma$ and $X_1,\dots,X_p$ are arbitrary tangent vectors. However, in our proof of Theorem 1, we follow the approach of [Petersen 1998], which uses the formalism of Clifford multiplication, and allows to express $\mathcal{B}^{[p]}$ directly in terms of the curvature operator (see Theorem 17 below). Our Theorem 1 follows from Theorem 17 and the splitting of the curvature operator induced by the Gauss formula. The next section relies on the exposition in [Petersen 1998, Section 7.4], which we follow closely; the only difference is the sign of the Riemann tensor.

5.1. ***The Bochner operator in the Clifford formalism.*** Let $\Lambda^\star(\Sigma)$ be the algebra of forms on Σ. Given $\theta \in \Lambda^1$ and $\omega \in \Lambda^p$, define their Clifford multiplication by

$$\begin{cases} \theta\cdot\omega = \theta\wedge\omega - i_{\theta^\#}\omega, \\ \omega\cdot\theta = (-1)^p(\theta\wedge\omega + i_{\theta^\#}\omega), \end{cases}$$

where $i_{\theta^\#}$ denotes interior multiplication of a form by $\theta^\#$, the dual vector field of θ. Note that by demanding that the product be bilinear and associative, the preceding equalities extend uniquely to define the Clifford multiplication of a p-form by a q-form. For 1-forms,

$$\begin{cases} \theta\cdot\theta = -|\theta|^2, \\ \theta_1\cdot\theta_2 + \theta_2\cdot\theta_1 = -2\langle\theta_1,\theta_2\rangle, \end{cases} \tag{28}$$

hence orthogonal 1-forms anticommute; moreover, any two orthogonal forms satisfy

$$\omega_1\cdot\omega_2 = \omega_1\wedge\omega_2. \tag{29}$$

Define the bracket as usual: $[\omega_1,\omega_2] = \omega_1\cdot\omega_2 - \omega_2\cdot\omega_1$. If θ is a 1-form and ψ is a 2-form, one checks that for all forms ω_1,ω_2

$$\begin{cases} \langle\theta\cdot\omega_1,\omega_2\rangle = -\langle\omega_1,\theta\cdot\omega_2\rangle, \\ \langle[\psi,\omega_1],\omega_2\rangle = -\langle\omega_1,[\psi,\omega_2]\rangle. \end{cases} \tag{30}$$

Fix an orthonormal frame $(e_1, \dots, e_n)$ with dual coframe $(\theta_1, \dots, \theta_n)$, and define the Dirac operator on forms $D : \Lambda^\star(\Sigma) \to \Lambda^\star(\Sigma)$ by

$$D\omega = \sum_{j=1}^{n} \theta_j \cdot \nabla_{e_j} \omega.$$

As $d\omega = \sum_{j=1}^{n} \theta_j \wedge \nabla_{e_j} \omega$ and $\delta\omega = -\sum_{j=1}^{n} i_{\theta_j^\#} \nabla_{e_j} \omega$, one sees that $D = d + \delta$, hence

$$D^2 = \Delta,$$

where Δ is the Laplacian on forms. Theorem 4.5 of [Petersen 1998] proves that

$$D^2\omega = \nabla^\star\nabla\omega + \frac{1}{2} \sum_{i,j=1}^{n} R(e_i, e_j)\omega \cdot \theta_i \cdot \theta_j,$$

$$D^2\omega = \nabla^\star\nabla\omega - \frac{1}{2} \sum_{i,j=1}^{n} \theta_i \cdot \theta_j \cdot R(e_i, e_j)\omega.$$

(As already observed, the change of sign in our formula is due to the opposite sign convention for the Riemann tensor adopted by Petersen.)

As $D^2\omega = \Delta\omega$, summing the two relations and dividing by 2 we then see

$$\mathscr{B}^{[p]}\omega = \frac{1}{4} \sum_{i,j=1}^{n} [R(e_i, e_j)\omega, \theta_i \cdot \theta_j]. \tag{31}$$

Now recall that the curvature operator $\mathfrak{R} : \Lambda_2(\Sigma) \to \Lambda_2(\Sigma)$ is the self-adjoint operator uniquely determined by the formula

$$\langle \mathfrak{R}(X \wedge Y), Z \wedge T \rangle = R(X, Y, Z, T) \doteq \langle R(X, Y)Z, T \rangle. \tag{32}$$

for all tangent vectors X, Y, Z, T. Then, we arrive at the following description of $\mathscr{B}^{[p]}$ in terms of $\mathfrak{R}$.

Theorem 17. *Let Σ be a manifold, and let $\mathscr{B}^{[p]}$ be the Bochner operator acting on p-forms of Σ. At any point of Σ, fix any orthonormal basis $\{\xi_r\}$ of $\Lambda_2(\Sigma)$ (here $r = 1, \dots, \binom{n}{2}$) and let $\{\hat{\xi}_r\}$ be its dual basis. Then*

$$\langle \mathscr{B}^{[p]}\omega, \phi \rangle = \frac{1}{4} \sum_{r,s} \langle \mathfrak{R}\xi_r, \xi_s \rangle \langle [\hat{\xi}_r, \omega], [\hat{\xi}_s, \phi] \rangle,$$

where the bracket is relative to Clifford multiplication and $\mathfrak{R}$ is the curvature operator of Σ.

For the proof, we start from [Petersen 1998, Lemma 4.7], which gives

$$R(e_i, e_j)\omega = \frac{1}{4} \sum_{h,k=1}^{n} \langle R(e_h, e_k)e_i, e_j \rangle [\theta_h \cdot \theta_k, \omega].$$

By (31),

$$\mathcal{B}^{[p]}\omega = \frac{1}{4}\sum_{\substack{i<j\\ h<k}} \langle \mathcal{R}(e_h \wedge e_k), e_i \wedge e_j\rangle [[\theta_h \cdot \theta_k, \omega], \theta_i \cdot \theta_j]. \tag{33}$$

By the adjointness property (30), we see that if ϕ is another p-form,

$$\langle [[\theta_h \cdot \theta_k, \omega], \theta_i \cdot \theta_j],\ \phi\rangle = -\langle [\theta_i \cdot \theta_j, [\theta_h \cdot \theta_k, \omega]],\ \phi\rangle = \langle [\theta_h \cdot \theta_k, \omega],\ [\theta_i \cdot \theta_j, \phi]\rangle,$$

and then

$$\langle \mathcal{B}^{[p]}\omega, \phi\rangle = \frac{1}{4}\sum_{\substack{i<j\\ h<k}} \langle \mathcal{R}(e_h \wedge e_k), e_i \wedge e_j\rangle \langle [\theta_h \cdot \theta_k, \omega],\ [\theta_i \cdot \theta_j, \phi]\rangle.$$

This is the expression in the orthonormal basis $\{\xi_r\} = \{e_i \wedge e_j\}_{i<j}$ of $\Lambda_2(\Sigma)$. Obviously, the choice of the orthonormal basis is not important and the theorem follows. □

5.2. ***A splitting of the curvature operator.*** Assume that Σ^n is a submanifold of M^{n+q}. Let R be the Riemann tensor of Σ and R^M that of M. For $X, Y, Z, T \in T\Sigma$, the Gauss formula gives

$$R(X, Y, Z, T) = R^M(X, Y, Z, T) + R_{\text{ext}}(X, Y, Z, T),$$

where

$$R_{\text{ext}}(X, Y, Z, T) = \langle L(X, Z), L(Y, T)\rangle - \langle L(X, T), L(Y, Z)\rangle$$

and L is the second fundamental form. Accordingly, we can split the curvature operator $\mathcal{R}$ of Σ as the sum of two self-adjoint operators acting on $\Lambda_2(\Sigma)$,

$$\mathcal{R} = \mathcal{R}_{\text{res}} + \mathcal{R}_{\text{ext}},$$

which are respectively defined on decomposable elements $X \wedge Y$, $Z \wedge T$ in $\Lambda_2(\Sigma)$ by

$$\begin{cases} \langle \mathcal{R}_{\text{res}}(X \wedge Y), Z \wedge T\rangle = \langle \mathcal{R}^M(X \wedge Y), Z \wedge T\rangle, \\ \langle \mathcal{R}_{\text{ext}}(X \wedge Y), Z \wedge T\rangle = \langle L(X, Z), L(Y, T)\rangle - \langle L(X, T), L(Y, Z)\rangle. \end{cases} \tag{34}$$

Let γ_M be the lowest eigenvalue of $\mathcal{R}^M$. As $\langle \mathcal{R}_{\text{res}}\xi, \xi\rangle = \langle \mathcal{R}^M\xi, \xi\rangle \geq \gamma_M|\xi|^2$ for all $\xi \in \Lambda_2(\Sigma)$, we see that $\mathcal{R}_{\text{res}} \geq \gamma_M$. The same remark applies to the largest eigenvalue Γ_M of $\mathcal{R}^M$. Hence

$$\gamma_M \leq \mathcal{R}_{\text{res}} \leq \Gamma_M. \tag{35}$$

Now let $(\nu_1, \dots, \nu_q)$ be an orthonormal frame in the normal bundle of Σ. By the definition of S_ν, we can write $\mathcal{R}_{\text{ext}} = \sum_{j=1}^q \mathcal{R}_{\text{ext}}^{(j)}$, where

$$\langle \mathcal{R}_{\text{ext}}^{(j)}(X \wedge Y), Z \wedge T\rangle = \langle S_{\nu_j}(X), Z\rangle \langle S_{\nu_j}(Y), T\rangle - \langle S_{\nu_j}(X), T\rangle \langle S_{\nu_j}(Y), Z\rangle. \tag{36}$$

In conclusion, one has the splitting

$$\mathcal{R} = \mathcal{R}_{\mathrm{res}} + \sum_{j=1}^{q} \mathcal{R}_{\mathrm{ext}}^{(j)} \tag{37}$$

with $\mathcal{R}_{\mathrm{res}}$ and $\mathcal{R}_{\mathrm{ext}}^{(j)}$ respectively given by (34) and (36).

5.3. *Algebraic lemma.* The proof of Theorem 1 depends on the following algebraic fact. Let S be a self-adjoint endomorphism of $T\Sigma$ and consider the associated "curvature operator" $\mathcal{R}_S : \Lambda_2(\Sigma) \to \Lambda_2(\Sigma)$ uniquely determined by the formula

$$\langle \mathcal{R}_S(X \wedge Y), Z \wedge T\rangle = \langle S(X), Z\rangle\langle S(Y), T\rangle - \langle S(X), T\rangle\langle S(Y), Z\rangle \tag{38}$$

for all $X, Y, Z, T \in T\Sigma$. Clearly, $\mathcal{R}_S$ is self-adjoint. Introduce the self-adjoint operator $T_S^{[p]} : \Lambda^p(\Sigma) \to \Lambda^p(\Sigma)$ such that, on any pair of p-forms ω, ϕ,

$$\langle T_S^{[p]}\omega, \phi\rangle = \frac{1}{4}\sum_{r,s}\langle \mathcal{R}_S\xi_r, \xi_s\rangle\langle[\hat{\xi}_r, \omega], [\hat{\xi}_s, \phi]\rangle, \tag{39}$$

where $\{\xi_r\}$ is any fixed orthonormal basis of $\Lambda_2(\Sigma)$ and $\{\hat{\xi}_r\}$ is its dual basis.

Lemma 18. *In the above notation, the operator $T_S^{[p]}$ can be written as*

$$T_S^{[p]} = (\operatorname{tr} S)S^{[p]} - S^{[p]} \circ S^{[p]},$$

where $S^{[p]}$ is the self-adjoint extension of S to $\Lambda^p(\Sigma)$.

Proof. Let $(e_1, \dots, e_n)$ be an orthonormal basis which diagonalizes S, so that $S(e_j) = k_j e_j$ for all $j = 1, \dots, n$ and k_j are the associated eigenvalues. Let $(\theta_1, \dots, \theta_n)$ be its dual basis. We refer to the notation in the proof of Lemma 2. Denote by I_p the set of multi-indices $\{j_1, \dots, j_p\}$ with $j_1 < \cdots < j_p$. If $\alpha = \{j_1, \dots, j_p\}$, let

$$\Theta_\alpha = \theta_{j_1} \wedge \cdots \wedge \theta_{j_p} = \theta_{j_1} \cdots \theta_{j_p}$$

where the dots in the last term indicate Clifford multiplication. The set $\{\Theta_\alpha : \alpha \in I_p\}$ is then an orthonormal basis of $\Lambda^p(\Sigma)$. It is enough to show that

$$\begin{cases} \langle T^{[p]}\Theta_\alpha, \Theta_\beta\rangle = 0 & \text{if } \alpha \neq \beta, \\ \langle T^{[p]}\Theta_\alpha, \Theta_\alpha\rangle = K_\alpha K_{\star\alpha}. \end{cases}$$

In fact, in that case, each Θ_α is an eigenform of $T_S^{[p]}$ associated to the eigenvalue $K_\alpha K_{\star\alpha}$, and it is readily seen from the discussion in Lemma 2 that the operator $(\operatorname{tr} S)S^{[p]} - S^{[p]} \circ S^{[p]}$ is the only one having that property.

Observe from (38) that the 2-vector $e_i \wedge e_j$ with $i < j$ is an eigenvector of $\mathcal{R}_S$ with associated eigenvalue $k_i k_j$. The set $\{\xi_r\} = \{e_i \wedge e_j\}_{i<j}$ forms an orthonormal

basis of $\Lambda_2(\Sigma)$. Therefore, by the definition in (39),

$$\langle T_S^{[p]}\omega, \phi\rangle = \frac{1}{4}\sum_{i<j} k_i k_j \langle [\theta_i\cdot\theta_j, \omega], [\theta_i\cdot\theta_j, \phi]\rangle. \tag{40}$$

A straightforward calculation using (28) shows that, for any $\alpha \in I_p$,

$$[\theta_i\cdot\theta_j, \Theta_\alpha] = \begin{cases} 0 & \text{if } i, j \in \alpha, \\ 0 & \text{if } i, j \in \star\alpha, \\ 2\theta_i\cdot\theta_j\cdot\Theta_\alpha & \text{otherwise.} \end{cases} \tag{41}$$

Then $\langle [\theta_i\cdot\theta_j, \Theta_\alpha], [\theta_i\cdot\theta_j, \Theta_\beta]\rangle$ is either zero or is equal to

$$4\langle \theta_i\cdot\theta_j\cdot\Theta_\alpha, \theta_i\cdot\theta_j\cdot\Theta_\beta\rangle = 4\langle\Theta_\alpha, \Theta_\beta\rangle$$

because $\langle\theta\cdot\omega_1, \theta\cdot\omega_2\rangle = \langle\omega_1, \omega_2\rangle$ for any 1-form θ and p-forms ω_1, ω_2. In particular $\langle T^{[p]}\Theta_\alpha, \Theta_\beta\rangle = 0$ when $\alpha \neq \beta$. It remains to show that $\langle T^{[p]}\Theta_\alpha, \Theta_\alpha\rangle = K_\alpha K_{\star\alpha}$. After renumbering, we can assume that $\alpha = \{1, \dots, p\}$ so that $\star\alpha = \{p+1, \dots, n\}$. Then

$$|[\theta_i\cdot\theta_j, \Theta_\alpha]|^2 = \begin{cases} 4 & \text{if } i \le p,\ j \ge p+1, \\ 0 & \text{otherwise,} \end{cases}$$

so that, by (40),

$$\begin{aligned}\langle T^{[p]}\Theta_\alpha, \Theta_\alpha\rangle &= \frac{1}{4}\sum_{i<j} k_i k_j |[\theta_i\cdot\theta_j, \Theta_\alpha]|^2 \\ &= \sum_{\substack{i\le p\\ j\ge p+1}} k_i k_j = (k_1+\cdots+k_p)(k_{p+1}+\cdots+k_n) = K_\alpha K_{\star\alpha}\end{aligned}$$

as asserted. □

5.4. *Proof of Theorem 1.* Let $\{\xi_r\}$ be an orthonormal basis of $\Lambda_2(\Sigma)$ with dual basis $\{\hat{\xi}_r\}$, where r is an index running from 1 to $\binom{n}{2}$. By Theorem 17 and the splitting given in (37), we have

$$\mathcal{B}^{[p]} = \mathcal{B}^{[p]}_{\text{res}} + \sum_{j=1}^{q} T^{[p]}_{\nu_j},$$

where, on given p-forms ω and ϕ,

$$\langle \mathcal{B}^{[p]}_{\text{res}}\omega, \phi\rangle = \frac{1}{4}\sum_{r,s}\langle \mathcal{R}_{\text{res}}\xi_r, \xi_s\rangle\langle[\hat{\xi}_r, \omega], [\hat{\xi}_s, \phi]\rangle,$$

$$\langle T^{[p]}_{\nu_j}\omega, \phi\rangle = \frac{1}{4}\sum_{r,s}\langle \mathcal{R}^{(j)}_{\text{ext}}\xi_r, \xi_s\rangle\langle[\hat{\xi}_r, \omega], [\hat{\xi}_s, \omega]\rangle.$$

From Lemma 18 applied to $S = S_{\nu_j}$, we get directly that

$$T^{[p]}_{\nu_j} = (\operatorname{tr} S_{\nu_j}) S^{[p]}_{\nu_j} - S^{[p]}_{\nu_j} \circ S^{[p]}_{\nu_j}.$$

It remains to show the bounds on $\mathcal{B}^{[p]}_{\text{res}}$. Choose $\{\xi_r\}$ to be an orthonormal basis of eigenvectors of $\mathcal{R}_{\text{res}}$. We know from (35) that $\langle \mathcal{R}_{\text{res}}\xi_r, \xi_r\rangle \geq \gamma_M |\xi_r|^2 = \gamma_M$ for all r. Then, for any p-form ω,

$$\langle \mathcal{B}^{[p]}_{\text{res}}\omega, \omega\rangle = \frac{1}{4}\sum_r \langle \mathcal{R}_{\text{res}}\xi_r, \xi_r\rangle |[\hat{\xi}_r, \omega]|^2 \geq \frac{\gamma_M}{4}\sum_r |[\hat{\xi}_r, \omega]|^2.$$

Now, the right-hand side does not depend on the choice of the basis $\{\xi_r\}$ of $\Lambda_2(\Sigma)$; choosing the basis $\{\theta_i \cdot \theta_j\}_{i<j}$ relative to an orthonormal coframe $(\theta_1, \dots, \theta_n)$ of $T\Sigma$, we have

$$\frac{1}{4}\sum_r |[\hat{\xi}_r, \omega]|^2 = \frac{1}{4}\sum_{i<j} |[\theta_i \cdot \theta_j, \omega]|^2 = p(n-p)|\omega|^2,$$

which follows from Lemma 18 applied to $S = Id$, with eigenvalues k_j all equal to 1. Then $\mathcal{B}^{[p]}_{\text{res}} \geq p(n-p)\gamma_M$. The upper bound $\mathcal{B}^{[p]}_{\text{res}} \leq p(n-p)\Gamma_M$ is proved similarly. □

6. Proof of Theorem 4

Let Σ^n be a compact hypersurface of $\mathbb{S}^{n+1}$ and let ω be a nontrivial parallel p-form on Σ, for some $p = 1, \dots, n-1$.

We first take care of the case $n = 2$, $p = 1$. As $\mathcal{B}^{[1]}\omega = K_\Sigma \omega$, we see immediately that $K_\Sigma = 0$, hence Σ is flat. As Σ is compact and orientable, Σ must be a flat torus.

We then assume $n \geq 3$. Let $(e_1, \dots, e_n)$ be a local orthonormal frame of principal directions associated to the principal curvatures $k_1, \dots, k_n$ on an open set U and let $(\theta_1, \dots, \theta_n)$ be its dual basis. The following facts have been proved in [Colbois and Savo 2012, Theorem 9]. As a consequence of the identity $R(e_i, e_j)\omega = 0$ one obtains, for all $i \neq j$,

$$(1 + k_i k_j)\Phi_{ij} = 0, \tag{42}$$

where Φ_{ij} is the p-form $\Phi_{ij} = \theta_j \wedge i_{e_i}\omega - \theta_i \wedge i_{e_j}\omega$. As ω is parallel, it never vanishes; as it is nontrivial we can assume, after renumbering the basis, that $\omega(e_1, \dots, e_p) \neq 0$ on U. This implies that for all $i \leq p$ and $j \geq p+1$ the form Φ_{ij} is nonzero, which forces $1 + k_i k_j = 0$. One quickly concludes that at each point, there are two principal curvatures, λ (with multiplicity p) and μ (with multiplicity $n-p$); that is,

$$\begin{cases} S(e_i) = \lambda e_i & \text{for } i = 1, \dots, p, \\ S(e_j) = \mu e_j & \text{for } j = p+1, \dots, n. \end{cases}$$

Moreover, one has, on U,

$$\lambda\mu = -1.$$

Now λ, μ are smooth functions on U. To prove Theorem 4 it is enough to show that $\nabla\lambda = \nabla\mu = 0$ on U. In fact, as U is arbitrary and Σ is connected, λ and μ will be constant on Σ. Therefore Σ is a compact isoparametric hypersurface with two principal curvatures, and by a well-known classification result it is isometric to a Clifford torus.

A result by T. Otsuki [1970] states that if Σ is a hypersurface in $\mathbb{S}^{n+1}$ such that the multiplicities of its principal curvatures are constant, then the distribution $D_\lambda = \{v \in T\Sigma : S(v) = \lambda v\}$ relative to a principal curvature λ is completely integrable. Moreover, if the multiplicity of λ is greater than one, then λ is constant on each of the integral leaves of the corresponding distribution. When there are only two principal curvatures (which is our case) this fact was also proved in [Ryan 1969, Proposition 2.3].

We first assume that $2 \le p \le n-2$. By what we have just said, on U we have

$$\begin{cases} \langle \nabla\lambda, e_i\rangle = 0 & \text{for } i = 1, \dots, p, \\ \langle \nabla\mu, e_j\rangle = 0 & \text{for } j = p+1, \dots, n. \end{cases} \tag{43}$$

Differentiating $\lambda\mu = -1$, we see that, on U,

$$\mu\nabla\lambda + \lambda\nabla\mu = 0. \tag{44}$$

Fix $i = 1, \dots, p$. As $\langle \nabla\lambda, e_i\rangle = 0$, we obtain from (44) that $\langle \lambda\nabla\mu, e_i\rangle = 0$; as $\lambda \neq 0$ we then have

$$\nabla\mu(e_i) = 0 \quad \text{for all } i = 1, \dots, p.$$

By (43) we see that $\nabla\mu = 0$ (hence $\nabla\lambda = 0$) on U.

We now assume that $p = 1$. Therefore

$$S(e_1) = \lambda e_1, \qquad S(e_j) = \mu e_j \quad \text{for } j = 2, \dots, n.$$

As $n \ge 3$, the multiplicity of μ is greater than one, and we have

$$\langle \nabla\mu, e_j\rangle = 0 \quad \text{for } j = 2, \dots, n.$$

By (44) we also have

$$\langle \nabla\lambda, e_j\rangle = 0 \quad \text{for } j = 2, \dots, n.$$

To prove the theorem, it then remains to show that $\langle \nabla\lambda, e_1\rangle = 0$.

From (42), we see that $(1 + k_i k_j)\Phi_{ij} = 0$, where $\Phi_{ij} = \omega(e_i)\theta_j - \omega(e_j)\theta_i$. Take $i, j \ge 2$ with $i \neq j$. As $1 + k_i k_j = 1 + \mu^2 \neq 0$, we must have $\Phi_{ij} = 0$. Hence

$$0 = \Phi_{ij}(e_j) = \omega(e_i) \quad \text{for all } i = 2, \dots, n.$$

As $\omega(e_1) \neq 0$, this gives

$$S^{[1]}\omega(e_1) = \lambda\omega(e_1) \quad \text{and} \quad S^{[1]}\omega(e_j) = 0 \quad \text{for all } j \geq 2.$$

This means that $S^{[1]}\omega = \lambda\omega$, and that the dual vector field X of ω is parallel and is a principal direction associated to λ:

$$S(X) = \lambda X.$$

As X has constant length, we can normalize so that $X = e_1$. We now compute $\operatorname{div}(S(X))$ in two ways. Since X is parallel, one has $\nabla_{e_j}(S(X)) = \nabla_{e_j} S(X)$. Then, by the Codazzi formula,

$$\operatorname{div}(S(X)) = \langle n\nabla H, X\rangle. \tag{45}$$

On the other hand, since $\operatorname{div} X = 0$,

$$\operatorname{div}(S(X)) = \operatorname{div}(\lambda X) = \langle \nabla\lambda, X\rangle. \tag{46}$$

Therefore $\langle n\nabla H - \nabla\lambda, X\rangle = 0$. Differentiating the identity $nH = \lambda - (n-1)/\lambda$, we obtain

$$\frac{n-1}{\lambda^2}\langle \nabla\lambda, X\rangle = 0 \quad \text{and} \quad \langle \nabla\lambda, e_1\rangle = \langle \nabla\lambda, X\rangle = 0.$$

The proof is complete. □

References

[Alencar and do Carmo 1994] H. Alencar and M. do Carmo, "Hypersurfaces with constant mean curvature in spheres", *Proc. Amer. Math. Soc.* **120**:4 (1994), 1223–1229. MR 94f:53108 Zbl 0802.53017

[Colbois and Savo 2012] B. Colbois and A. Savo, "Involutive isometries, eigenvalue bounds and a spectral property of Clifford tori", *Indiana Univ. Math. J.* **61**:1 (2012), 337–357. MR 3029401 Zbl 1269.58012

[Dajczer and Rodríguez 1986] M. Dajczer and L. Rodríguez, "Rigidity of real Kaehler submanifolds", *Duke Math. J.* **53**:1 (1986), 211–220. MR 87g:53089 Zbl 0599.53005

[El Soufi and Petit 2000] A. El Soufi and R. Petit, "Immersions minimales et immersions pluriharmoniques entre variétés Riemanniennes: résultants de non existence et de rigidité", *Ann. Inst. Fourier* (*Grenoble*) **50**:1 (2000), 235–256. MR 2001h:53085 Zbl 0945.53038

[Gallot and Meyer 1975] S. Gallot and D. Meyer, "Opérateur de courbure et Laplacien des formes différentielles d'une variété Riemannienne", *J. Math. Pures Appl.* (9) **54**:3 (1975), 259–284. MR 56 #13128 Zbl 0316.53036

[Grosjean 2004] J.-F. Grosjean, "Minimal submanifolds with a parallel or a harmonic p-form", *J. Geom. Phys.* **51**:2 (2004), 211–228. MR 2005h:53103 Zbl 1074.53050

[Lawson and Simons 1973] H. B. Lawson, Jr. and J. Simons, "On stable currents and their application to global problems in real and complex geometry", *Ann. of Math.* (2) **98** (1973), 427–450. MR 48 #2881 Zbl 0283.53049

[Otsuki 1970] T. Ôtsuki, "Minimal hypersurfaces in a Riemannian manifold of constant curvature", *Amer. J. Math.* **92** (1970), 145–173. MR 41 #9157 Zbl 0196.25102

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

[Raulot and Savo 2011] S. Raulot and A. Savo, "A Reilly formula and eigenvalue estimates for differential forms", *J. Geom. Anal.* **21**:3 (2011), 620–640. MR 2012j:58043 Zbl 1228.58018

[Ryan 1969] P. J. Ryan, "Homogeneity and some curvature conditions for hypersurfaces", *Tôhoku Math. J.* (2) **21** (1969), 363–388. MR 40 #6458 Zbl 0185.49904

[Savo 2005] A. Savo, "On the first Hodge eigenvalue of isometric immersions", *Proc. Amer. Math. Soc.* **133**:2 (2005), 587–594. MR 2005h:58053 Zbl 1054.58023

[Weiner 1991] J. L. Weiner, "Flat tori in S^3 and their Gauss maps", *Proc. London Math. Soc.* (3) **62**:1 (1991), 54–76. MR 92d:53057 Zbl 0724.53036

Received September 19, 2013. Revised April 7, 2014.

ALESSANDRO SAVO
DIPARTIMENTO SBAI, SEZIONE DI MATEMATICA
SAPIENZA UNIVERSITÀ DI ROMA
VIA ANTONIO SCARPA 16
I-00161 ROMA
ITALY
alessandro.savo@sbai.uniroma1.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.423

ON SOLUTIONS TO COURNOT–NASH EQUILIBRIA EQUATIONS ON THE SPHERE

MICAH WARREN

We discuss equations associated to Cournot–Nash Equilibria as put forward recently by Blanchet and Carlier. These equations are related to an optimal transport problem in which the source measure is known, but the target measure is part of the problem. The resulting equation is of Monge–Ampère type with possible nonlocal terms. If the cost function is of a particular form, the equation is vulnerable to standard optimal transportation PDE techniques, with some modifications to deal with the new terms. We give some sufficient conditions for the problem on the sphere from which we can conclude that solutions are smooth.

1. Introduction

In this note, we discuss equations associated to Cournot–Nash equilibria as put forward in [Blanchet and Carlier 2012], a reference we henceforth abbreviate as [BC]. These equations are related to an optimal transport problem in which the source measure is known but the target measure is to be determined. A Cournot–Nash equilibrium (CNE) is a special type of optimal transport: Each individual x is transported to a point $T(x)$ in a way that not only minimizes the total cost of transportation, but minimizes a cost to the individual x (transportation plus other). This latter cost may depend on the target distribution, and may involve congestion, isolation and geographical terms.

Blanchet and Carlier demonstrated how CNE are related to nonlinear elliptic PDEs, explicitly deriving a Euclidean version of the equation [BC, (4.6)] and showing that this problem has some very nice properties [BC, Theorem 3.8]. The fully nonlinear Monge–Ampère equation differs from "standard" optimal transport equations in that the potential itself occurs on the right-hand side, along with possibly some nonlocal terms. Here we study the problem on the sphere. Immediately one can conclude from [BC, Theorem 3.8] and [Loeper 2009] that optimal maps are

I am supported in part by NSF DMS-0901644 and NSF DMS-1161498. This work was completed while I was at Princeton University.
MSC2010: 49Q20, 91A10, 35J96.
Keywords: optimal transportation, Cournot–Nash equilibrium.

continuous with control on the Hölder norm. We move this a step further and show that all derivative norms can be controlled in terms of the data, when the solution is smooth. When the solution is known to be differentiable enough, one can easily adapt the estimates of Ma, Trudinger and Wang [Ma et al. 2005]. To make the conclusion a priori, we must use the continuity method. Closedness follows from the same estimates, but openness is not immediate and requires some conditions. In Theorem 6 we give some conditions on the data so that the problem can be solved smoothly.

2. Background and setup

In this section we briefly recap the setup in [BC]. Given a space of player types X endowed with a probability measure μ, an action space Y, and a cost function

$$\Phi : X \times Y \times \mathcal{P}(Y) \to \mathbb{R},$$

assume x-type agents pay cost $\Phi(x, y, \nu)$ to take action y. Here $\nu \in \mathcal{P}(Y)$ is the probability measure in the action space which is the push forward of μ by the map of actions from X to Y. Supposing that x-type agents know the distribution ν, they can choose the best action y. A Cournot–Nash equilibrium is a joint probability distribution measure $\gamma \in \mathcal{P}(X \times Y)$ with first marginal μ such that

$$\gamma\Big\{(x, y) \in X \times Y : \Phi(x, y, \nu) = \min_{z \in Y} \Phi(x, z, \nu)\Big\} = 1, \tag{2-1}$$

where ν is the second marginal.

We will be interested in a particular type of cost,

$$\Phi(x, y, \nu) = c(x, y) + \mathcal{V}[\nu](y),$$

where c is the transportation cost. Lemma 2.2 of [BC] shows that a CNE will necessarily be an optimal transport pairing for the cost c between the measures μ and ν. They further show that, if $\mathcal{V}[\nu]$ is the differential of a functional $\mathcal{E}[\nu]$, then, at a minimizer for $\mathcal{E}[\nu]+\mathcal{W}_c(\mu, \nu)$, the optimal transport will necessarily be a CNE (here $\mathcal{W}_c(\mu, \nu)$ is the Wasserstein distance). In particular, if the cost $\mathcal{V}_m[\nu]$ is of the form

$$\mathcal{V}_m[\nu](y) = f\left(\frac{d\nu}{dm}(y)\right) + \int \phi(y, z)\, d\nu(z) + V(y), \tag{2-2}$$

where m is a "background" measure and the function $\phi(y, z)$ is symmetric on $Y \times Y$, then $\mathcal{V}_m$ is a differential, and a solution to the optimal transport is a CNE. (We will be licentious with notation, letting ν denote not only the measure, but also the density with respect to the background m.) From here on we suppose we are working with a solution to an optimal transport with cost c between measures

μ and ν which is also a CNE for a total cost Φ. We also assume that the manifolds X and Y are compact without boundary.

One can consider the pair (u, u^*) which maximizes the Kantorovich functional

$$J(u, v) = \int -u \, d\mu + \int v \, d\nu$$

over all $-u(x) + v(y) \le \Phi(x, y)$. The pair (u, u^*) will satisfy

$$-u(x) + u^*(y) = \Phi(x, y) \tag{2-3}$$

γ-almost everywhere, where γ is the optimal measure for the Kantorovich problem. If the cost satisfies the standard Spence–Mirrlees condition (in the mathematics literature, the "twist", or [Ma et al. 2005, Section 2, condition (A1)]) we have, μ-almost everywhere,

$$-u(x) + u^*(T(x)) = \Phi(x, T(x)). \tag{2-4}$$

The twist condition says that $T(x)$ is uniquely determined by

$$T(x) = \{y : Du(x) + Dc(x, y) = 0\}, \tag{2-5}$$

which gives the identity

$$Du(x) + Dc(x, T(x)) = 0. \tag{2-6}$$

Note that, fixing an x, the quantity

$$\Phi(x, y) - u^*(y)$$

must have a minimum at $T(x)$; we conclude that

$$D_y\Phi(x, T(x)) = Du^*(T(x)).$$

Then by condition (2-1), for fixed x,

$$\Phi(x, T(x)) \le \Phi(x, y),$$

which implies that

$$D_y\Phi(x, T(x)) = 0,$$

from which we conclude that

$$Du^*(y) \equiv 0.$$

Now the pair (u, u^*) is determined up to a constant. One can choose the constant in u or u^* but not both. At this point we simply choose $u^* = 0$. Having fixed this choice, we obtain information about u and the measure ν, using (2-2) and (2-4):

$$-u(x) = c(x, T(x)) + f\big(\nu(T(x))\big) + \int \phi(T(x), z) \, d\nu(z) + V(T(x)).$$

In particular, the density $\nu(y)$ must be determined by

$$\nu(T(x)) = f^{-1}\Big(-u(x) - c(x,T(x)) - \int \phi(T(x),T(z))\,d\mu(z) - V(T(x))\Big), \tag{2-7}$$

having used the change of integration variables T between μ and ν. The optimal transportation equation (see [Ma et al. 2005]) becomes

$$\frac{\det\big(u_{ij}(x)+c_{ij}(x,T(x))\big)}{\det\big(-c_{is}(x,T(x))\big)} = \frac{\mu(x)}{f^{-1}(Q(x,u))}. \tag{2-8}$$

Here and in the sequel, we use i, j, k to denote derivatives in the source X, and p, s, t to denote derivatives in the target Y. It will be convenient to assume that c_{is} is negative definite, which follows if we are assuming condition (A2) of [Ma et al. 2005] and have chosen an appropriate coordinate system. We will use $b_{is}(x) = -c_{is}(x,T(x))$. Also (to keep equations within one line) we abbreviate

$$Q(x,u) = -u(x) - c(x,T(x)) - \int \phi(T(x),T(z))\,d\mu(z) - V(T(x)),$$

with $T(x)$ being determined by (2-5).

Before we say how this fully nonlinear equation is vulnerable, we mention the "Inada-like" conditions [BC, Section 3.3]

$$\begin{gathered} \lim_{\nu\to 0^+} f(\nu) = -\infty \quad \text{and} \quad \lim_{\nu\to+\infty} f(\nu) = +\infty, \\ f' > 0 \quad \text{and} \quad f \in C^2(\mathbb{R}^+). \end{gathered} \tag{2-9}$$

If f satisfies these conditions, then several observations are in order. First, as noted in [BC, Theorem 3.8], on a compact manifold we get bounds away from zero and infinity for the density ν. In the spherical distance-squared transportation cost case, this immediately gives C^α-continuity of the map, by results of Loeper. Secondly, the right-hand side of (2-8) is strictly monotone in the zeroth-order term — this is crucial in obtaining existence and uniqueness results, as it will allow us to invert the linearized operator. Finally, as we will show below, the first derivatives of this density will be bounded in terms of an a priori constant (depending on the smoothness of f) and the second derivatives will be bounded by a constant times second derivatives of u. These estimates will allow us to take advantage of the Ma–Trudinger–Wang estimates.

We will show an estimate on smooth solutions: If a solution to (2-8) is C^4, then it enjoys estimates of all orders subject to universal bounds. In order to show that arbitrary solutions are C^4 and hence smooth, we must use a continuity method. This method relies on a linearization which requires some discussion, given the integral terms in the equation.

The problem here, on a compact manifold with cost function satisfying the Ma–Trudinger–Wang condition, is quite simpler than the delicate boundary value problem in [BC]. With or without the nonlocal terms, such a problem may be approached as in [Liu and Trudinger 2010]. We leave this problem aside for now.

3. Linearization

We take the natural log of (2-8) and then consider the functional

$$F(x,u,Du,D^2u) = \ln\det\big(u_{ij}(x)+c_{ij}(x,T(x))\big) - \ln\det\big(b_{is}(x,T(x))\big) - \ln\mu(x) + \ln f^{-1}(Q(x,u)); \tag{3-1}$$

the equation we want to solve is

$$F(x,u,Du,D^2u) = 0. \tag{3-2}$$

Preparing for linearization, consider (2-6) applied to $u+tv$:

$$Du(x) + t\,D\eta(x) + Dc(x,T_t(x)) = 0.$$

Differentiate with respect to t to get

$$D\eta(x) = b_{is}(x,T(x))\frac{dT^s}{dt}.$$

Linearizing, we obtain

$$L\eta = \frac{d}{dt}F(u+t\eta) = L^0\eta + L^1\eta,$$

where

$$L^0\eta = w^{ij}\eta_{ij} + w^{ij}c_{ijs}b^{sk}\eta_k + b^{is}c_{isp}b^{pk}\eta_k, \tag{3-3}$$

$$\begin{aligned} L^1\eta = \frac{(f^{-1}(Q))'}{f^{-1}(Q)}\Big(&-c_s(x,T(x))b^{sk}\eta_k - \eta - V_s b^{sk}\eta_k \\ &- b^{sk}\eta_k(x)\int\phi_s(T(x),T(z))\,d\mu(z) \\ &- \int\phi_{\bar{s}}(T(x),T(z))b^{sk}(z)\,\eta_k(z)\,d\mu(z)\Big). \end{aligned} \tag{3-4}$$

Here we are using

$$w_{ij}(x) = u_{ij}(x) + c_{ij}(x,T(x)).$$

We note also that differentiating (2-6) shows

$$T_i^s(x,T(x)) = \frac{\partial T^s}{\partial x_i} = b^{sk}(x,T(x))\,w_{ki}(x,T(x)). \tag{3-5}$$

We take $g_{ij}(x) = w_{ij}(x)$ to define a metric (one can check that it transforms as such), then write

$$d\mu(x) = e^{-a(x)}\,dV_g(x), \tag{3-6}$$

where

$$-a(x) = \ln \mu(x) - \tfrac{1}{2} \ln \det w_{ij}(x).$$

From the definition of F in (3-1) we have

$$-a(x) = \tfrac{1}{2} \ln \det w_{ij} - \ln \det b + \ln \nu - F,$$

having introduced

$$\nu(x) = \ln f^{-1}(Q(x, u)).$$

First, we compute the weighted Laplace

$$\triangle_a \eta = \triangle_g \eta - \nabla a \cdot \nabla \eta.$$

We begin with $\triangle_g \eta$, differentiating in some coordinate system (see (4-1) for very similar computations):

$$\begin{aligned}
\frac{\left(\sqrt{\det w}\, w^{ij} \eta_j\right)_i}{\sqrt{\det w}} &= w^{ij}\eta_{ij} + \tfrac{1}{2} w^{ab} \partial_i w_{ab} w^{ij} \eta_j - w^{ia} w^{bj} \partial_i w_{ab} \eta_j \\
&= w^{ij}\eta_{ij} + w^{ab} w^{ij} (\partial_i w_{ab} - \partial_b w_{ia}) \eta_j - \tfrac{1}{2} w^{ab} \partial_i w_{ab} w^{ij} \eta_j \\
&= w^{ij}\eta_{ij} + (w^{ba} c_{abs} b^{sj} - w^{ij} c_{isk} b^{sk}) \eta_j - \tfrac{1}{2} w^{ij} (\ln \det w)_i \eta_j \\
&= L^0 \eta - b^{is} c_{isp} b^{pk} \eta_k - w^{ij} c_{kis} b^{sk} \eta_j - \tfrac{1}{2} w^{ij} (\ln \det w)_i \eta_j.
\end{aligned}$$

Thus

$$\begin{aligned}
\triangle_a \eta = L^0 \eta - b^{is} c_{isp} b^{pk} \eta_k - w^{ij} c_{kis} b^{sk} \eta_j - \tfrac{1}{2} w^{ij} (\ln \det w)_i \eta_j \\
+ \tfrac{1}{2} w^{ij} (\ln \det w)_i \eta_j - w^{ij} (\ln \det b)_i \eta_j + (\ln \nu)_i w^{ij} \eta_j - F_i w^{ij} \eta_j \\
= L^0 v + (\ln \nu)_i w^{ij} \eta_j - F_i w^{ij} \eta_j,
\end{aligned}$$

and hence

$$L\eta = \triangle_a \eta + L^1 \eta - (\ln \nu)_i w^{ij} \eta_j + F_i w^{ij} \eta_j.$$

Next, we compute

$$\begin{aligned}
(\ln \nu)_i = \frac{(f^{-1}(Q))'}{f^{-1}(Q)} \Big(-u_i(x) - c_i(x, T(x)) - c_s(x, T(x)) b^{sk} w_{ki} \\
- b^{sk} w_{ki} \int \phi_s(T(x), T(z))\, d\mu(z) - V_s b^{sk} w_{ki} \Big).
\end{aligned}$$

Noting that $-u_i(x) - c_i(x, T(x))$ vanishes, and (3-4), we have

$$L^1 \eta - (\ln \nu)_i w^{ij} \eta_j = \frac{(f^{-1}(Q))'}{f^{-1}(Q)} \Big(-\eta - \int \phi_s(T(x), T(z)) b^{sk}(z) \eta_k(z)\, d\mu(z) \Big).$$

Next, we compute the integral term in the previous expression. Notice

$$\begin{aligned}\int \langle \nabla\phi(y, T(z)), \nabla\eta\rangle e^{-a(z)}\, dV_g(z) &= \int \phi_s(y, T(z)) b^{sk} w_{ki}\eta_j w^{ij} e^{-a(x)}\, dV_g \\ &= \int \phi_s(T(x), T(z)) b^{sk}\eta_k(z)\, d\mu(z).\end{aligned}$$

Now, integrating by parts, we have

$$-\int \phi_s(T(x), T(z)) b^{sk}\eta_k(z)\, d\mu(z) = \int \phi(T(x), T(z)) \bigtriangleup_a \eta(z) e^{-a(z)}\, dV_g(z).$$

Combining, we have

$$L\eta = \bigtriangleup_a\eta - h(x)\eta(x) - h(x)\int \phi(T(x), T(z)) \bigtriangleup_a \eta(z)\, d\mu(z) + \langle\nabla F, \nabla\eta\rangle, \tag{3-7}$$

using the shorthand

$$h(x) = \frac{(f^{-1}(Q))'}{f^{-1}(Q)},$$

which represents a positive differentiable quantity if f satisfies (2-9). In particular, if $f(\tau) = \ln\tau$ then h will be identically 1. When $F \equiv 0$ we have the following.

Proposition 1. *At a solution of (3-2), the linearized operator takes the form*

$$L\eta = \bigtriangleup_a\eta - h(x)\eta(x) - h(x)\int \phi(T(x), T(z)) \bigtriangleup_a \eta(z)\, d\mu(z). \tag{3-8}$$

Lemma 2. *Suppose that*

$$\max_{(x,y)\in X\times Y} h(x)|\phi(x, y)| < 1. \tag{3-9}$$

Then the operator (3-8) has trivial kernel.

Proof. To make use of some functional analytic formality, we define operators A, J, h, and I on the space $\mathcal{B} = L^2(X, d\mu)$ by

$$\begin{aligned}[A\eta](x) &= \bigtriangleup_a\eta(x), \\ [J\eta](x) &= \int \phi(T(x), T(z))\eta(z)\, d\mu(z), \\ [h\eta](x) &= h(x)\eta(x), \\ [I\eta](x) &= \eta(x).\end{aligned}$$

Then $L = A - h - hJA = (I - hJ)A - h = (I - hJ)\big(A - (I - hJ)^{-1}h\big)$.

First, we have the pointwise estimate

$$\begin{aligned}[hJ\eta](x) &= \int h(x)\phi(T(x), T(y))\eta(y)\, d\mu(y) \\ &\le \Big\| \int h(x)\phi(T(x), T(y))\, d\mu(x)\Big\|_{L^2}^{1/2} \|\eta\|_{L^2}^{1/2} \\ &\le \big(\max_{(x,y)\in X\times Y} h(x)|\phi(x, y)|\big)^{1/2} < \|\eta\|_{L^2}^{1/2},\end{aligned}$$

using (3-9). Integrating this quantity over μ yields

$$\|hJ\| < 1$$

as an operator on $\mathcal{B}$, so $(I - hJ)$ is invertible. Thus we have

$$\operatorname{Ker} L = \operatorname{Ker}(A - (I - hJ)^{-1}h).$$

Now suppose, for purposes of contradiction, that we have nontrivial $\eta \in \operatorname{Ker} L$. Then

$$A\eta = (I - hJ)^{-1}h\eta,$$

thus

$$\langle (I - hJ)^{-1}h\eta, \eta\rangle = \langle A\eta, \eta\rangle = -\int |\nabla\eta|^2 \, d\mu < 0.$$

But, as $(I - hJ)$ is invertible, we can let

$$(I - hJ)\omega = h\eta.$$

Then

$$\langle \omega, h^{-1}(I - hJ)\omega\rangle = \langle (I - hJ)^{-1}h\eta, \eta\rangle < 0,$$

that is,

$$0 > \left\langle \omega, \frac{1}{h}\omega \right\rangle - \langle \omega, J\omega\rangle \geq \frac{1}{\max h}\|\omega\|^2 - \|J\|\,\|\omega\|^2 = \left(\frac{1}{\max h} - \|J\|\right)\|\omega\|^2,$$

which is clearly a contradiction if $1 > \max h \|J\|$. □

4. Estimates on the sphere

From here on we specialize to the round unit sphere, with cost function half of distance squared. Note that this sphere has Riemannian volume $n\omega_n$.

Oscillation estimates. The following estimates are a version of [BC, Lemma 3.7]. On a compact manifold, the cost function will be bounded. Since the solution u is c-convex, at its maximum point $x_{\max}$, u is supported below by the cost support function $c(x, T(x_0)) + \lambda$. Hence, at the minimum point $x_{\min}$, we have that $u(x_{\min}) \geq c(x_{\min}, T(x_{\max})) + \lambda$, which in turn tells us that

$$\operatorname{osc} u \leq \operatorname{osc} c = \tfrac{1}{2}\pi^2.$$

Next we observe that, because the integration of the density ν against m gives a probability measure, the density ν must be larger than $1/(n\omega_n)$ at some point y_0. Using (2-7), it follows that, at the point $x_0 = T^{-1}(y_0)$,

$$-c(x_0, y_0) - u(x_0) - \int \phi(y_0, T(z))\, d\mu(z) - V(y_0) \geq f\left(\frac{1}{n\omega_n}\right),$$

and similarly, at the point x_1 where the density ν is smallest,

$$-c(x_1, y_1) - u(x_1) - \int \phi(y_1, T(z))\, d\mu(z) - V(y_1) = f(\nu(x_1)).$$

Hence,

$$-c(x_0, y_0) + c(x_1, y_1) - u(x_0) + u(x_1) - \int \big(\phi(y_0, T(z)) + \phi(y_1, T(z))\big)\, d\mu(z) \\ -V(y_0) + V(y_1) \geq f\Big(\frac{1}{n\omega_n}\Big) - f(\nu(x_1)),$$

that is,

$$f(\nu(x_1)) \geq f\Big(\frac{1}{n\omega_n}\Big) - 2 \operatorname{osc} c - 2 \operatorname{osc} \phi - \operatorname{osc} V > -\infty.$$

By Inada's conditions,

$$\nu \geq f^{-1}\Big(f\Big(\frac{1}{n\omega_n}\Big) - \pi^2 - 2 \operatorname{osc} \phi - \operatorname{osc} V\Big) > 0.$$

Similarly, an upper bound can be derived:

$$\nu \leq f^{-1}\Big(f\Big(\frac{1}{n\omega_n}\Big) + \pi^2 + 2 \operatorname{osc} \phi + \operatorname{osc} V\Big) < \infty.$$

4.1. ***Stayaway.*** Now that ν is under control, it follows from the stayaway estimates of [Delanoë and Loeper 2006] that the map $T(x)$ must satisfy

$$\operatorname{dist}_{\mathbb{S}^n}(x, T(x)) \leq \pi - \epsilon(f, \mu, V, \phi).$$

In particular, the map stays clear of the cut locus. All derivatives of the cost function are now controlled.

MTW estimates.

Lemma 3. *If the map T is differentiable and locally invertible, then the target measure density*

$$\nu(T(x)) = f^{-1}\Big(-c(x, T(x)) - u(x) - \int \phi(T(x), T(z))\, d\mu(z) - V(T(x))\Big)$$

has first derivatives bounded by a universal constant and has second derivatives

$$\nu_{sr} = C_1 + C_{2k}(T^{-1})^k_r,$$

where the bounding constants are within a controlled range.

Proof. Differentiate in the x_k direction:

$$\nu_s T^s_k(x) = (f^{-1})'\Big(-c_k(x, T(x)) - c_s(x, T(x)) T^s_k - u_k \\ - T^s_k \int \phi_s(T(x), T(z))\, d\mu(z) - V_s T^s_k\Big) \\ = (f^{-1})' T^s_k(x)\Big(-c_s(x, T(x)) - \int \phi_s(T(x), T(z))\, d\mu(z) - V_s(T(x))\Big).$$

As this is true for all k, and DT is invertible, we can conclude that

$$\nu_s(T(x)) = (f^{-1})'\Big(-c_s(x,T(x)) - \int \phi_s(T(x),T(z))\,d\mu(z) - V_s(T(x))\Big)$$

is a bounded quantity. For second derivatives, differentiate this equation in x again:

$$\begin{aligned}\nu_{sp}T_k^p = (f^{-1})''T_k^p(x)\Big(-c_s(x,T(x)) - \int \phi_s(T(x),T(z))\,d\mu(z) - V_s(T(x))\Big)&\\ \times\Big(-c_p(x,T(x)) - \int \phi_p(T(x),T(z))\,d\mu(z) - V_p(T(x))\Big)&\\ +(f^{-1})'\Big(-c_{sk}(x,T(x)) - c_{sp}(x,T(x))T_k^p(x)&\\ - T_k^p(x)\int \phi_{ps}(T(x),T(z))\,d\mu(z) - T_k^p(x)V_{sp}(T(x))\Big)&,\end{aligned}$$

that is,

$$\begin{aligned}\nu_{sr} = (f^{-1})''\Big(-c_s(x,T(x)) - \int \phi_s(T(x),T(z))\,d\mu(z) - V_s(T(x))\Big)&\\ \times\Big(-c_p(x,T(x)) - \int \phi_p(T(x),T(z))\,d\mu(z) - V_p(T(x))\Big)&\\ +(f^{-1})'\Big(-c_{sk}(x,T(x))(T^{-1})_r^k - c_{sp}(x,T(x))&\\ - \int \phi_{ps}(T(x),T(z))\,d\mu(z) - V_{sp}(T(x))\Big)&.\end{aligned}$$

Now all the terms, with the exception of the $(T^{-1})_r^k$ term, are given by controlled constants, independent of u. We are done. □

Before we state the main a priori estimate, we recall the Ma–Trudinger–Wang (MTW) tensor [Ma et al. 2005, p. 154]. For each y in the target, one can define the MTW tensor as a (2, 2)-tensor on T_xM via

$$\mathsf{MTW}_{ij}^{kl}(x,y) = \{(-c_{ijpr} + c_{ijs}c^{sm}c_{mrp})c^{pk}c^{rl}\}(x,y).$$

It is by now a well-known fact that, on the sphere,

$$\mathsf{MTW}_{ij}^{kl}\xi_k\xi_l\tau^i\tau^j \geq \delta_n\|\xi\|^2\|\tau\|^2$$

for a positive δ_n and all vector–covector pairs such that

$$\xi(\tau) = 0.$$

(For more discussion of the geometry of this tensor, see [Kim and McCann 2010].)

Given a solution, we define an operator on (2, 0)-tensors as follows. Let h be a (2, 0)-tensor. Given vector fields X_1, X_2, we define

$$(L_w h)(X_1, X_2) = \frac{1}{\sqrt{\det w}}\nabla_j\big(\sqrt{\det w}\,w^{ij}\nabla_i h\big) - w^{ij}\nabla_j a\nabla_i h(X_1, X_2),$$

where

$$-a(x) = \tfrac{1}{2}\ln\det w(x) - \ln\det b(x) + \ln\nu(x,T(x))$$

and covariant differentiation is taken with respect to the round metric.

Proposition 4. *Let u be a solution of (2-8). If e is a unit direction in a local chart on S^n, then*

$$L_w w(e,e) \\ \geq w^{ij}(-c_{ijpr}+c_{ijs}c_{krp}c^{sk})c^{pm}c^{rl}w_{me}w_{le}-C\big(1+\sum w^{ii}\sum w_{jj}+\sum w^{ii}+\sum w_{ii}^2\big)$$

Proof. This was proven in the case where densities are known ahead of time by Ma et al. [2005]. Adapting their proof requires only a small modification somewhere in the middle, but for completeness (and mostly for fun), we will present the calculation.

First, we note that

$$\begin{aligned}
\text{(4-1)}\quad & \frac{\partial_j\big(\sqrt{\det w}\,w^{ij}\big)}{\sqrt{\det w_{ij}}}-w^{ij}a_j \\
&=\partial_j w^{ij}+\tfrac{1}{2}w^{ij}(\ln\det w)_j+w^{ij}\tfrac{1}{2}(\ln\det w)_j-w^{ij}(\ln\det b)_j+w^{ij}(\ln\nu)_s T^s_j \\
&=-w^{ia}w^{bj}\partial_j w_{ab}+w^{ij}(\ln\det w)_j-w^{ij}\big(b^{sk}b_{skj}+b^{sk}b_{skt}T^t_j\big)+b^{si}(\ln\nu)_s \\
&=-w^{ia}w^{bj}(\partial_j w_{ab}-\partial_a w_{bj})-w^{ia}w^{bj}\partial_a w_{bj}+w^{ij}(\ln\det w)_j \\
&\qquad\qquad -w^{ij}b^{sk}b_{skj}-b^{ti}b^{sk}b_{skt}+b^{si}(\ln\nu)_s \\
&=-w^{ia}w^{bj}(c_{abs}T^s_j-c_{bjs}T^s_a)-w^{ij}b^{sk}b_{skj}-b^{ti}b^{sk}b_{skt}+b^{si}(\ln\nu)_s \\
&=b^{si}w^{bj}c_{bjs}-b^{ti}b^{sk}b_{skt}+b^{si}(\ln\nu)_s
\end{aligned}$$

using (among others) the relations

$$\text{(4-2)}\qquad \partial_j w_{ab}-\partial_a w_{bj}=c_{abs}T^s_j-c_{bjs}T^s_a,\qquad w^{bj}T^s_j=b^{sj}.$$

Now

$$\begin{aligned}
L_w & w(e_1,e_1) \\
&=\frac{1}{\sqrt{\det w}}\nabla_j(\sqrt{\det w}\,w^{ij}\nabla_i w)(e_1,e_1)-w^{ij}\nabla_j a\nabla_i w(e_1,e_1) \\
&=w^{ij}\nabla_j\nabla_i w(e_1,e_1)+(b^{si}w^{bj}c_{bjs}-b^{ti}b^{sk}b_{skt}+b^{si}(\ln\nu)_s)\nabla_i w(e_1,e_1) \\
&=w^{ij}\,\big(\partial_i\partial_j w(e_1,e_1)-\nabla_j\partial_i w(e_1,e_1)+2w(\nabla_{\nabla_j\partial_i}e_1,e_1)-2\partial_i w(\nabla_j e_1,e_1) \\
&\qquad\qquad -2\partial_j w(\nabla_i e_1,e_1)+2w(\nabla_j\nabla_i e_1,e_1)+2w(\nabla_i e_1,\nabla_j e_1)\big) \\
&\qquad +(b^{si}w^{bj}c_{bjs}-b^{ti}b^{sk}b_{skt}+b^{si}(\ln\nu)_s)(\partial_i w(e_1,e_1)-2w(\nabla_i e_1,e_1)).
\end{aligned}$$

At this point, we choose a normal coordinate system (in the round metric), then

$$\begin{aligned}L_w w(e_1,e_1) &= (b^{si}w^{bj}c_{bjs} - b^{ti}b^{sk}b_{skt} + b^{si}(\ln\nu)_s)\partial_i w(e_1,e_1)\\ &\qquad + w^{ij}(\partial_i\partial_j w(e_1,e_1) + 2w(\nabla_j\nabla_i e_1,e_1))\\ &= (b^{is}w^{bj}c_{bjs} - b^{it}b^{sk}b_{skt} + b^{is}(\ln\nu)_s)\partial_i w_{11}\\ &\qquad + w^{ij}(\partial_i\partial_j w_{11} - \partial_1\partial_1 w_{ij}) + w^{ij}(\partial_1\partial_1 w_{ij} + 2w(\nabla_j\nabla_i e_1,e_1)).\end{aligned}$$

Again harking back to [Ma et al. 2005], we let

$$K = C\sum w^{ii}\sum w_{jj} + C\sum w^{ii} + C\sum w_{ii}^2 + C$$

and note that terms of the following form are K:

$$K = w^{ij}T_b^s, \quad K = (\partial_j w_{ik} - \partial_k w_{ij}), \quad K = w^{ij}2w(\nabla_j\nabla_i e_1,e_1), \quad K = w^{ij}w_{kl};$$

so that

$$\begin{aligned}L_w w(e_1,e_1) = -K + (b^{si}w^{bj}c_{bjs} - b^{ti}b^{sk}b_{skt} + b^{si}(\ln\nu)_s)\partial_i w_{11}\\ + w^{ij}(\partial_i\partial_j w_{11} - \partial_1\partial_1 w_{ij}) + w^{ij}\partial_1\partial_1 w_{ij}.\end{aligned}$$

Now, differentiating

$$\ln\det w_{ij} = \ln\det b_{is} + \ln\mu - \ln\nu, \tag{4-3}$$

we have

$$w^{ij}\partial_1 w_{ij} = b^{si}(b_{is1} + b_{ist}T_1^t) + (\ln\mu)_1 - (\ln\nu)_s T_1^s \tag{4-4}$$

and again

$$w^{ij}\partial_{11}w_{ij} + \partial_1 w^{ij}\partial_1 w_{ij} = K + b^{si}b_{ist}T_{11}^t + (\ln\nu)_{sr}T_1^r T_1^s - (\ln\nu)_s T_{11}^s.$$

Now recall Lemma 3, which gives

$$(\ln\nu)_{sr}T_1^r T_1^s = \frac{C_{1sr} + C_{2sk}(T^{-1})_r^k}{\nu}T_1^r T_1^s - (\ln\nu)_s(\ln\nu)_r T_1^r T_1^s = K;$$

thus

$$w^{ij}\partial_{11}w_{ij} = w^{ia}w^{bj}\partial_1 w_{ab}\partial_1 w_{ij} + K + b^{si}b_{ist}T_{11}^t - (\ln\nu)_s T_{11}^s. \tag{4-5}$$

Note that differentiating $T_i^s = b^{sk}w_{ki}$ yields

$$T_{ij}^s = b^{sk}\partial_j w_{ki} - b^{sa}b^{pk}w_{ki}(b_{apj} + b_{apq}T_j^q), \tag{4-6}$$

in particular

$$T_{11}^s = b^{sk}\partial_1 w_{k1} - b^{sa}b^{pk}w_{k1}(b_{ap1} + b_{apq}T_1^q).$$

Now it follows that

$$T_{11}^s - b^{sk}\partial_k w_{11} = b^{sk}(\partial_1 w_{k1} - \partial_k w_{11}) - b^{sa}b^{pk}w_{k1}(b_{ap1} + b_{apq}T_1^q) \tag{4-7}$$

$$= K. \tag{4-8}$$

Bringing in the concavity of the Monge–Ampère equation (4-5) and (4-8), we can eliminate some terms to see

$$L_w w(e_1, e_1) \geq -K + b^{is} w^{bj} c_{bjs} \partial_i w_{11} + w^{ij} (\partial_i \partial_j w_{11} - \partial_1 \partial_1 w_{ij}).$$

Then, using

$$\begin{aligned}
\partial_1 \partial_1 w_{ij} &= u_{ij11} + c_{ij11} + 2c_{ijs1} T_1^s + c_{ijs} T_{11}^s + c_{ijpr} T_1^p T_1^r, \\
\partial_i \partial_j w_{11} &= u_{11ij} + c_{11ij} + c_{11si} T_j^s + c_{11sj} T_i^s + c_{11s} T_{ij}^s + c_{11pr} T_i^p T_j^r,
\end{aligned}$$

we have

$$\begin{aligned}
&L_w w(e_1, e_1) \\
&\quad \geq -K + (b^{is} w^{bj} c_{bjs}) \partial_i w_{11} + w^{ij} (c_{11s} T_{ij}^s + c_{11pr} T_i^p T_j^r - c_{ijs} T_{11}^s - c_{ijpr} T_1^p T_1^r).
\end{aligned}$$

From (4-6),

$$\begin{aligned}
w^{ij} T_{ij}^s &= w^{ij} (b^{sk} \partial_j w_{ki} - b^{sa} b^{pk} w_{ki} (b_{apj} + b_{apq} T_j^q)) \\
&= w^{ij} b^{sk} (\partial_j w_{ki} - \partial_k w_{ij} + \partial_k w_{ij}) - b^{sa} b^{pj} (b_{apj} + b_{apq} T_j^q) \\
&= K + b^{sk} \partial_k (\ln \det w) \\
&= K
\end{aligned}$$

by (4-4). Using (4-7) we conclude

$$L_w w(e_1, e_1) \geq -K - w^{bj} c_{bjs} b^{sa} b^{pk} w_{k1} b_{apq} T_1^q - w^{ij} c_{ijpr} T_1^p T_1^r,$$

which is the desired result after reindexing. □

Corollary 5. *Second derivatives of u are uniformly bounded.*

Proof. Given the maximum principle estimate, this proof is standard, following [Ma et al. 2005]. For some more details in the setting of Riemannian manifolds see [Kim et al. 2012, Theorem 3.5]. □

5. Main theorem

In order to make a precise statement, we define

$$\begin{aligned}
\nu_{\text{lower}} &= f^{-1}\left(f\left(\frac{1}{n\omega_n} \right) - 2 \operatorname{osc} c - 2\|\phi\|_\infty - \operatorname{osc} V \right) \\
\nu_{\text{upper}} &= f^{-1}\left(f\left(\frac{1}{n\omega_n} \right) + 2 \operatorname{osc} c + 2\|\phi\|_\infty + \operatorname{osc} V \right).
\end{aligned}$$

Similarly, an upper bound can be defined by

$$h_{\max} = \sup_{Q \in [\nu_{\text{lower}}, \nu_{\text{upper}}]} \frac{(f^{-1}(Q))'}{f^{-1}(Q)}.$$

Theorem 6. *Suppose that f satisfies the Inada-like conditions (2-9), μ and m are smooth, and ϕ and V are Lipschitz. If*

$$\max_{x,y\in M} |\phi(x,y)| < \frac{1}{h_{\max}}, \tag{5-1}$$

then there exists a smooth solution to (3-2).

Proof. For existence, we proceed by continuity [Gilbarg and Trudinger 2001, Theorem 17.6] on (3-2), letting

$$\begin{aligned} F(t, x, u, Du, D^2u) &\\ = \ln\det\bigl(D^2u + D^2c(x, T(x))\bigr) &- \ln\det\bigl(-D\bar{D}c(x, T(x))\bigr)\\ &- \ln(t\mu(x) + (1-t)m(x)) + \ln f^{-1}\bigl(Q(t, x, T(x))\bigr), \end{aligned} \tag{5-2}$$

where

$$Q(t, x, T(x)) = -u(x) - c(x, T(x)) - t\int \phi(T(x), T(z))\, d\mu(z) - tV(T(x)).$$

At time $t = 0$, a solution is given by $u \equiv 0$: this maps the measure m to itself via the identity mapping. Thus the interval $\mathcal{I}$ of t for which a solution exists is nonempty. Notice that the form of (5-2) is the same as of (3-2) up to a scale of the functions ϕ and V and a change of measure, so the estimates from the previous section all hold. From the theory of Krylov and Evans one can obtain $C^{2,\alpha}$ estimates. Thus $\mathcal{I}$ is closed. Lemma 2 with these conditions gives openness, noting that on the sphere a Laplacian has index zero, and that the linearized operator which has the same principal symbol has index zero as well. □

Remark. For uniqueness, the standard PDE trick does not work immediately, even under assumptions such as those in the theorem. One may be tempted to use the standard argument [Gilbarg and Trudinger 2001, Theorem 17.1] to obtain a contradiction. However, the intermediate linearized operator will have the additional ∇F term that arises in (3-7) because combinations of u and v are not solutions. Our proof of invertibility fails for these, so we have no reason to expect that the proof would remain valid after being integrated. Uniqueness may be more easily obtained from geometric consideration as in [BC, Section 4]; see also [Villani 2009, Chapters 15 and 16].

However, if the integral term is not present, we can use the argument [Gilbarg and Trudinger 2001, Theorem 17.1], making the important note that on the sphere the set of c-convex functions is convex [Figalli et al. 2011, Theorem 3.2]. In this case, invertibility of the linearized operator follows easily from standard maximum principle arguments.

References

[Blanchet and Carlier 2012] A. Blanchet and G. Carlier, "Optimal transport and Cournot–Nash equilibria", preprint, 2012. arXiv 1206.6571

[Delanoë and Loeper 2006] P. Delanoë and G. Loeper, "Gradient estimates for potentials of invertible gradient-mappings on the sphere", *Calc. Var. Partial Differential Equations* **26**:3 (2006), 297–311. MR 2007b:35108 Zbl 1136.35358

[Figalli et al. 2011] A. Figalli, Y.-H. Kim, and R. J. McCann, "When is multidimensional screening a convex program?", *J. Econom. Theory* **146**:2 (2011), 454–478. MR 2888826 Zbl 05882918

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, Berlin, 2001. MR 2001k:35004 Zbl 1042.35002

[Kim and McCann 2010] Y.-H. Kim and R. J. McCann, "Continuity, curvature, and the general covariance of optimal transportation", *J. Eur. Math. Soc.* (*JEMS*) **12**:4 (2010), 1009–1040. MR 2011f:49071 Zbl 1191.49046

[Kim et al. 2012] Y.-H. Kim, J. Streets, and M. Warren, "Parabolic optimal transport equations on manifolds", *Int. Math. Res. Not.* **2012**:19 (2012), 4325–4350. MR 2981711 Zbl 1253.49036

[Liu and Trudinger 2010] J. Liu and N. S. Trudinger, "On Pogorelov estimates for Monge–Ampère type equations", *Discrete Contin. Dyn. Syst.* **28** (2010), 1121–1135. MR 2011e:35103 Zbl 05760053

[Loeper 2009] G. Loeper, "On the regularity of solutions of optimal transportation problems", *Acta Math.* **202**:2 (2009), 241–283. MR 2010c:49084 Zbl 1219.49038

[Ma et al. 2005] X.-N. Ma, N. S. Trudinger, and X.-J. Wang, "Regularity of potential functions of the optimal transportation problem", *Arch. Ration. Mech. Anal.* **177** (2005), 151–183. MR 2006m:35105 Zbl 1072.49035

[Villani 2009] C. Villani, *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften **338**, Springer, Berlin, 2009. MR 2010f:49001 Zbl 1156.53003

Received October 4, 2013.

MICAH WARREN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OR 97403-1222
UNITED STATES

micahw@uoregon.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.439

DOUBLE AND TRIPLE GIVENTAL'S J-FUNCTIONS FOR STABLE QUOTIENTS INVARIANTS

ALEKSEY ZINGER

We use mirror formulas for the stable quotients analogue of Givental's J-function for twisted projective invariants obtained in a previous paper to obtain mirror formulas for the analogues of the double and triple Givental's J-functions (with descendants at all marked points) in this setting. We then observe that the genus-0 stable quotients invariants need not satisfy the divisor, string, or dilaton relations of the Gromov–Witten theory, but they do possess the integrality properties of the genus-0 three-point Gromov–Witten invariants of Calabi–Yau manifolds. We also relate the stable quotients invariants to the BPS counts arising in Gromov–Witten theory and obtain mirror formulas for certain twisted Hurwitz numbers.

1. Introduction

Gromov–Witten invariants of projective varieties are counts of curves that are conjectured (and known in some cases) to possess a rich structure. In particular, so-called mirror formulas relate these symplectic invariants of a nonsingular variety X to complex-geometric invariants of the mirror family of X. In genus 0, this relation is often described by assembling two-point Gromov–Witten invariants (but

MSC2010: 14N35, 53D45.
Keywords: stable quotients, mirror symmetry.

without constraints on the second marked point) into a generating function, known as Givental's J-function, and expressing it in terms of an explicit hypergeometric series. The genus-0 Gromov–Witten invariants of a projective complete intersection X are equal to the twisted Gromov–Witten invariants of the ambient space associated to the direct sum of positive line bundles corresponding to X. The genus-0 mirror formula in Gromov–Witten theory extends to the twisted Gromov–Witten invariants associated with direct sums of line bundles over projective spaces; see [Elezi 2003; Givental 1996; Lian et al. 1999]. By [Cooper and Zinger 2014], the analogue of Givental's J-function for the twisted stable quotients invariants defined in [Marian et al. 2011] satisfies a simpler version of the mirror formula from Gromov–Witten theory. In this paper, we obtain mirror formulas for the stable quotients analogues of the double and triple Givental's J-functions for direct sums of line bundles. We use them to test the stable quotients invariants for the analogues of the standard properties satisfied by Gromov–Witten invariants. In the future, we intend to apply the methods of this paper to show that the stable quotients and Gromov–Witten invariants of projective complete intersections are related by a simple mirror transform, in all genera, but with at least one marked point.

1A. *Stable quotients.* The moduli spaces of stable quotients, $\overline{Q}_{g,m}(X,d)$, constructed in [Marian et al. 2011] and generalized in [Ciocan-Fontanine et al. 2014], provide an alternative to the moduli spaces of stable maps, $\overline{\mathfrak{M}}_{g,m}(X,d)$, for compactifying spaces of degree-d morphisms from genus-g nonsingular curves with m marked points to a projective variety X (with a choice of polarization). A stable tuple of quotients is a tuple

$$(\mathcal{C}, y_1, \dots, y_m; S_1 \subset \mathbb{C}^{n_1} \otimes \mathcal{O}_{\mathcal{C}}, \dots, S_p \subset \mathbb{C}^{n_p} \otimes \mathcal{O}_{\mathcal{C}}), \tag{1-1}$$

where $\mathcal{C}$ is a connected (at worst) nodal curve, $y_1, \dots, y_m \in \mathcal{C}^*$ are distinct smooth points, and

$$S_1 \subset \mathbb{C}^{n_1} \otimes \mathcal{O}_{\mathcal{C}}, \dots, S_p \subset \mathbb{C}^{n_p} \otimes \mathcal{O}_{\mathcal{C}}$$

are subsheaves such that the quotients $\mathbb{C}^{n_1} \otimes \mathcal{O}_{\mathcal{C}}/S_1, \dots, \mathbb{C}^{n_p} \otimes \mathcal{O}_{\mathcal{C}}/S_p$ are locally free at the nodes of $\mathcal{C}$ and the marked points $y_1, \dots, y_m$ and the $\mathbb{Q}$-line bundle

$$\omega_{\mathcal{C}}(y_1 + \dots + y_m) \otimes (\Lambda^{\mathrm{top}} S_1^*)^{\epsilon} \otimes \dots \otimes (\Lambda^{\mathrm{top}} S_p^*)^{\epsilon} \to \mathcal{C}$$

is ample for all $\epsilon \in \mathbb{Q}^+$; this implies that $2g+m \geq 2$.

In this paper, we are concerned only with the case $g=0$. For $m, d_1, \dots, d_p \in \mathbb{Z}^{\geq 0}$ and $n_1, \dots, n_p \in \mathbb{Z}^+$, the moduli space

$$\overline{Q}_{0,m}(\mathbb{P}^{n_1-1} \times \dots \times \mathbb{P}^{n_p-1}, (d_1, \dots, d_p)) \tag{1-2}$$

parameterizing the stable tuples of quotients as in (1-1) with $h^1(\mathcal{C}, \mathcal{O}_{\mathcal{C}}) = 0$, that is, $\mathcal{C}$ is a rational curve, $\mathrm{rk}(S_i) = 1$, and $\deg(S_i) = -d_i$, is a nonsingular irreducible

Deligne–Mumford stack and

$$\dim \overline{Q}_{0,m}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(d_1,\dots,d_p)) \\ =(d_1+1)n_1+\cdots+(d_p+1)n_p-p-3+m;$$

see [Cooper and Zinger 2014, Propositions 2.1, 2.2].

As in the case of stable maps, there are evaluation morphisms,

$$\mathrm{ev}_i:\overline{Q}_{0,m}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(d_1,\dots,d_p))\to\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},$$

$i=1,2,\dots,m$, corresponding to each marked point. There is also a universal curve

$$\pi:\mathcal{U}\to\overline{Q}_{0,m}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(d_1,\dots,d_p))$$

with m sections $\sigma_1,\dots,\sigma_m$ (given by the marked points) and p universal rank-1 subsheaves

$$\mathcal{S}_i\subset\mathbb{C}^{n_i}\otimes\mathcal{O}_{\mathcal{U}}.$$

In the case $p=1$, we will denote $\mathcal{S}_1$ by $\mathcal{S}$. For each $i=1,2,\dots,m$, let

$$\psi_i=-\pi_*(\sigma_i^2)\in H^2\big(\overline{Q}_{0,m}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(d_1,\dots,d_p))\big)$$

be the first chern class of the universal cotangent line bundle as usual.

The twisted invariants of projective spaces that we study in this paper are indexed by tuples $\mathbf{a}=(a_1,\dots,a_l)\in(\mathbb{Z}^*)^l$ of nonzero integers, with $l\in\mathbb{Z}^{\geq 0}$. For each such tuple $\mathbf{a}$, let

$$|\mathbf{a}|=\sum_{k=1}^{l}|a_k|,\quad \langle\mathbf{a}\rangle=\prod_{a_k>0}a_k\Big/\prod_{a_k<0}a_k,\quad \mathbf{a}!=\prod_{a_k>0}a_k!,\quad \mathbf{a}^{\mathbf{a}}=\prod_{k=1}^{l}a_k^{|a_k|},$$

$$\nu_n(\mathbf{a})=n-|\mathbf{a}|,\quad \ell^{\pm}(\mathbf{a})=|\{k:(\pm1)a_k>0\}|,\quad \ell(\mathbf{a})=\ell^+(\mathbf{a})-\ell^-(\mathbf{a}).$$

If in addition $n,d\in\mathbb{Z}^+$, let

$$\mathcal{V}_{n;\mathbf{a}}^{(d)}=\bigoplus_{a_k>0}R^0\pi_*(\mathcal{S}^{*a_k})\oplus\bigoplus_{a_k<0}R^1\pi_*(\mathcal{S}^{*a_k})\to\overline{Q}_{0,m}(\mathbb{P}^{n-1},d), \tag{1-3}$$

where $\pi:\mathcal{U}\to\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)$ is the universal curve and $m\geq 2$; these sheaves are locally free.

By [Ciocan-Fontanine et al. 2014, Theorem 4.5.2 and Proposition 6.2.3],

$$\mathrm{SQ}_{n;\mathbf{a}}^{d}(c_1,\dots,c_m)\equiv\int_{\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)}\mathbf{e}(\mathcal{V}_{n;\mathbf{a}}^{(d)})\prod_{i=1}^{m}\mathrm{ev}_i^*x^{c_i},\quad m\geq 2,d\in\mathbb{Z}^+,c_i\in\mathbb{Z}^{\geq 0},$$

where $x \in H^2(\mathbb{P}^{n-1})$ is the hyperplane class, are invariants of the total space $X_{n;\mathbf{a}}$ of the vector bundle

$$\bigoplus_{a_k<0} \mathcal{O}_{\mathbb{P}^{n-1}}(a_k)\big|_{X_{n;(a_k)_{a_k>0}}} \to X_{n;(a_k)_{a_k>0}}, \tag{1-4}$$

where $X_{n;(a_k)_{a_k>0}} \subset \mathbb{P}^{n-1}$ is a nonsingular complete intersection of multidegree $(a_k)_{a_k>0}$. If $\nu_n(\mathbf{a}) = 0$, that is, $X_{n;\mathbf{a}}$ is a Calabi–Yau complete intersection, let

$$\mathrm{GW}^{c_1,\dots,c_m}_{n;\mathbf{a}}(q) = \sum_{d=0}^{\infty} q^d \, \mathrm{SQ}^d_{n;\mathbf{a}}(c_1,\dots,c_m),$$

with $\mathrm{GW}^0_{n;\mathbf{a}}(\mathbf{c}) \equiv \langle \mathbf{a}\rangle$ if $|\mathbf{c}| = n-4-\ell(\mathbf{a})+m$ and 0 otherwise.

1B. ***SQ invariants and GW invariants.*** In Gromov–Witten theory, there is a natural evaluation morphism $\mathrm{ev}: \mathcal{U} \to \mathbb{P}^{n-1}$ from the universal curve

$$\pi: \mathcal{U} \to \overline{\mathfrak{M}}_{0,m}(\mathbb{P}^{n-1}, d).$$

If $n, d \in \mathbb{Z}^+$, the sheaf

$$\mathcal{V}^{(d)}_{n;\mathbf{a}} = \bigoplus_{a_k>0} R^0\pi_*\mathrm{ev}^*\mathcal{O}_{\mathbb{P}^{n-1}}(a_k) \oplus \bigoplus_{a_k<0} R^1\pi_*\mathrm{ev}^*\mathcal{O}_{\mathbb{P}^{n-1}}(a_k) \to \overline{\mathfrak{M}}_{0,m}(\mathbb{P}^{n-1}, d), \tag{1-5}$$

is locally free. It is well known that

$$\mathrm{GW}^d_{n;\mathbf{a}}(c_1,\dots,c_m) \equiv \int_{\overline{\mathfrak{M}}_{0,m}(\mathbb{P}^{n-1},d)} \mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}}) \prod_{i=1}^m \mathrm{ev}_i^* x^{c_i}, \qquad m, c_i \in \mathbb{Z}^{\ge 0}, d \in \mathbb{Z}^+,$$

are also invariants of $X_{n;\mathbf{a}}$. If $\nu_n(\mathbf{a}) = 0$ and $m \ge 2$, let

$$\mathrm{GW}^{c_1,\dots,c_m}_{n;\mathbf{a}}(Q) = \sum_{d=0}^{\infty} Q^d \, \mathrm{GW}^d_{n;\mathbf{a}}(c_1,\dots,c_m),$$

with $\mathrm{GW}^0_{n;\mathbf{a}}(\mathbf{c}) \equiv \langle \mathbf{a}\rangle$ if $|\mathbf{c}| = n-4-\ell(\mathbf{a})+m$ and 0 otherwise.

Stable quotients invariants and Gromov–Witten invariants are equal in many cases, but differ for many Calabi–Yau targets, as we now describe. Let

$$\begin{aligned} \dot{F}_{n;\mathbf{a}}(w,q) \\ \equiv \sum_{d=0}^{\infty} q^d w^{\nu_n(\mathbf{a})d} \frac{\prod_{a_k>0}\prod_{r=1}^{a_k d}(a_k w + r)\prod_{a_k<0}\prod_{r=0}^{-a_k d-1}(a_k w - r)}{\prod_{r=1}^{d}((w+r)^n - w^n)} \in \mathbb{Q}(w)[\![q]\!], \end{aligned} \tag{1-6}$$

$$\dot{I}_0(q) = \dot{F}_{n;\mathbf{a}}(0,q), \qquad J_{n;\mathbf{a}}(q) = \frac{1}{\dot{I}_0(q)} \frac{\partial \dot{F}_{n;\mathbf{a}}}{\partial w}\bigg|_{(0,q)}. \tag{1-7}$$

The term w^n in the denominator in (1-6) above is irrelevant for the purposes of the main formulas of Sections 1–3. Its introduction is related to the expansion (4-9), which is used in an essential way in the proof of (3-14) in Section 10.

Theorem 1. *Let $n, l \in \mathbb{Z}^+$ and $\mathbf{a} \in (\mathbb{Z}^*)^l$ be such that $\nu_n(\mathbf{a}) = 0$. If $m = 2, 3$ and $\mathbf{c} \in (\mathbb{Z}^{\geq 0})^m$, then*

$$d^{3-m}\mathrm{SQ}^d_{n;\mathbf{a}}(\mathbf{c}) \in \mathbb{Z} \;\text{ for all } d \in \mathbb{Z}, \tag{1-8}$$

$$\mathrm{GW}^{\mathbf{c}}_{n;\mathbf{a}}(Q) = \dot{I}_0(q)^{m-2}\,\mathrm{SQ}^{\mathbf{c}}_{n;\mathbf{a}}(q) - \delta_{m,2}\langle\mathbf{a}\rangle J_{n;\mathbf{a}}(q), \tag{1-9}$$

where $\delta_{m,2}$ is the Kronecker delta function and $Q = q\mathrm{e}^{J_{n;\mathbf{a}}(q)}$ is the mirror map. *Furthermore, the genus-0 three-marked stable quotients invariants of $X_{n;\mathbf{a}}$ satisfy the analogue of the dilaton equation of Gromov–Witten theory if and only if $\ell^-(\mathbf{a}) > 0$, and of the divisor and string relations if and only if $\ell^-(\mathbf{a}) > 1$.*

The relation (1-9) follows from the explicit mirror formulas for the stable quotients analogues of the double and triple Givental's J-functions provided by Theorem 2 in Section 2 and similar results in Gromov–Witten theory [Popa 2012; Zinger 2014]; see Section 2 for more details. By [Ciocan-Fontanine and Kim 2013, Theorem 1.2.2 and Corollaries 1.4.1, 1.4.2], (1-9) holds for $m > 3$ as well. As the mirror formulas of Theorem 2 relate SQ invariants to the hypergeometric series arising in the B-model of the mirror family without a change of variables, (1-9) illustrates the principle that the mirror map relating Gromov–Witten theory to the B-model reflects the choice of the curve-counting theory in the A-model and is not intrinsic to mirror symmetry itself.

The analogue of (1-8) for GW invariants is well known. By [McDuff and Salamon 2004, Proposition 7.3.2], the genus-0 GW invariants of a Calabi–Yau manifold with 3+ marked points are integer. The $m = 2$ case of (1-8) for GW invariants is implied by the $m = 3$ case and the divisor relation. The $m = 2, 3$ cases of (1-8) for SQ invariants follow from the $m = 2, 3$ cases of (1-8) for GW invariants and from (1-9), since $\dot{I}_0(q), Q(q) \in \mathbb{Z}[\![q]\!]$; the integrality of the coefficients of $Q(q)$ whenever $\ell^-(\mathbf{a}) = 0$ is a special case of [Krattenthaler and Rivoal 2010, Theorem 1].[1] Since (1-9) extends to $m > 3$ by [Ciocan-Fontanine and Kim 2013], so does (1-8), but without the d^{3-m} factors.

Since $\dot{I}_0(q) = 1$ if $\ell^-(\mathbf{a}) = 0$ and $J_{n;\mathbf{a}}(q) = 0$ if $\ell^-(\mathbf{a}) = 0, 1$, (1-9) gives the following corollary.

Corollary 1.1. *Let $n, l \in \mathbb{Z}^+$ and $\mathbf{a} \in (\mathbb{Z}^*)^l$ be such that $\nu_n(\mathbf{a}) = 0$ and $a_{k_1}, a_{k_2} < 0$ for some $k_1 \neq k_2$. If $m = 2, 3$ and $\mathbf{c} \in (\mathbb{Z}^{\geq 0})^m$, then*

$$\mathrm{GW}^{\mathbf{c}}_{n;\mathbf{a}}(Q) = \mathrm{SQ}^{\mathbf{c}}_{n;\mathbf{a}}(q).$$

[1] The integrality of the coefficients of $\dot{I}_0(q)$ and of $Q(q)$ in the cases $\ell^-(\mathbf{a}) > 0$ is immediate from their definitions.

d	$d\mathrm{GW}^d_{n;\mathbf{a}}(1,1)$	$d\mathrm{SQ}^d_{n;\mathbf{a}}(1,1)$
1	2875	6725
2	4876875	16482625
3	8564575000	44704818125
4	15517926796875	126533974065625
5	28663236110956000	366622331794131725
6	53621944306062201000	1078002594137326617625
7	101216230345800061125625	3201813567943782886368125
8	192323666400003538944396875	9579628267176528143932815625
9	367299732093982242625847031250	28820906443427523291443507328125
10	704288164978454714776724365580000	87086311562396929291553775833982625

Table 1. Some genus-0 GW and SQ invariants of a quintic threefold $X_{5;(5)}$.

*Furthermore, the genus-*0 *three-marked stable quotients invariants of* $X_{n;\mathbf{a}}$ *satisfy the analogue of the divisor, dilaton, and string equations of Gromov–Witten theory.*

By Theorem 2, the conclusions of Corollary 1.1 also apply to the descendant invariants. Stable quotients replacements for the divisor, string, or dilaton relations [Hori et al. 2003, Section 26.3] for an arbitrary Calabi–Yau complete intersection $X_{n;\mathbf{a}}$ are provided by (2-23), (2-24), and (2-25), respectively. For the sake of comparison, we list a few genus-0 SQ and GW invariants of the quintic threefold $X_{5,(5)} \subset \mathbb{P}^4$ in Table 1; these are obtained from (2-33) and (2-34), respectively.

1C. ***SQ invariants and BPS states.*** Using (1-9), the genus-0 two- and three-marked SQ invariants of a Calabi–Yau complete intersection threefold $X_{n;\mathbf{a}}$ can be expressed in terms of the BPS counts of GW theory. For example, by the $m=2$ case of (1-9),

$$\mathrm{SQ}^{1,1}_{n;\mathbf{a}}(q) = \langle\mathbf{a}\rangle J_{n;\mathbf{a}}(q) - \sum_{d=1}^{\infty} \mathrm{BPS}^d_{n;\mathbf{a}}(1,1)\ln\bigl(1-q^d e^{dJ_{n;\mathbf{a}}(q)}\bigr), \tag{1-10}$$

where $\mathrm{BPS}^d_{n;\mathbf{a}}(1,1)$ are the genus-0 two-marked BPS counts for $X_{n;\mathbf{a}}$ defined by

$$\mathrm{GW}^{1,1}_{n;\mathbf{a}}(Q) = -\sum_{d=1}^{\infty} \mathrm{BPS}^d_{n;\mathbf{a}}(1,1)\ln(1-Q^d).$$

If all genus-0 curves in $X_{n;\mathbf{a}}$ of degree at most d were smooth and had normal bundles $\mathcal{O}(-1)\oplus\mathcal{O}(-1)$, the number of degree-$d$ genus-0 curves in $X_{n;\mathbf{a}}$ would be $\mathrm{BPS}^d_{n;\mathbf{a}}(1,1)$; see [Voisin 1996, Section 1].

Under the regularity assumption of the previous paragraph, the moduli space

$$\overline{Q}^{1,1}_{0,2}(X_{n;\mathbf{a}}, d) \equiv \{u \in \overline{Q}_{0,2}(X_{n;\mathbf{a}}, d) : \mathrm{ev}_1(u) \in H_1, \mathrm{ev}_2(u) \in H_2\},$$

where $H_1, H_2 \subset \mathbb{P}^{n-1}$ are generic hyperplanes, would split into the topological components:

- $\mathcal{Z}^{1,1}_0(d)$ of stable quotients with torsion of degree d and thus corresponding to a constant map to $H_1 \cap H_2$;
- $\mathcal{Z}^{1,1}_C(d)$ of stable quotients with image in a genus-0 curve $C \subset X_{n;\mathbf{a}}$ of degree $d_C \le d$.

For $C \subset X_{n;\mathbf{a}}$ as above, $\mathcal{Z}^{1,1}_C(d)$ consists of the closed subspaces $\mathcal{Z}^{1,1}_{C;r}(d)$, with $r \in \mathbb{Z}^+$ and $d_C r \le d$, whose generic element has torsion of degree $d - d_C r$. We note that

$$\dim \mathcal{Z}^{1,1}_{C;r}(d) = 2r - 2 + d - d_C r + 2 = d - (d_C - 2)r,$$

which implies that each $\mathcal{Z}^{1,1}_{C;r}(d)$ is an irreducible component if $d_C > 1$. When $d_C = 1$, $\mathcal{Z}^{1,1}_{C;r}(d)$ is contained in $\mathcal{Z}^{1,1}_{C;d}(d)$, but still gives rise to a separate contribution to $\mathrm{SQ}^d_{n;\mathbf{a}}(1, 1)$, according to (1-10).

The number $\mathrm{SQ}^d_{n;\mathbf{a}}(2, 0)$, which arises from the constrained moduli space

$$\overline{Q}^{2,0}_{0,2}(X_{n;\mathbf{a}}, d) = \mathcal{Z}^{2,0}_0(d) = \mathcal{Z}^{1,1}_0(d),$$

is $\langle\mathbf{a}\rangle$ times the coefficient $[\![J_{n;\mathbf{a}}(q)]\!]_d$ of q^d in $J_{n;\mathbf{a}}(q)$; see [Cooper and Zinger 2014, Theorem 1]. The contribution of $\mathcal{Z}^{1,1}_0(d)$ to $\mathrm{SQ}^d_{n;\mathbf{a}}(2, 0)$ is the same; this explains the first term on the right-hand side of (1-10). Under the above regularity assumption, (1-10) can be rewritten as

$$\mathrm{SQ}^d_{n;\mathbf{a}}(1, 1) = \langle\mathbf{a}\rangle[\![J_{n;\mathbf{a}}(q)]\!]_d + \sum_C \sum_{r=1}^{\infty} \frac{1}{r}[\![\mathrm{e}^{J_{n;\mathbf{a}}(q)}]\!]_{d-rd_C}, \tag{1-11}$$

where the outer sum is taken over all genus-0 curves $C \subset X_{n;\mathbf{a}}$. This suggests that the contribution of $\mathcal{Z}^{1,1}_{C;r}(d)$ to $\mathrm{SQ}^d_{n;\mathbf{a}}(1, 1)$ is $\frac{1}{r}[\![\mathrm{e}^{J_{n;\mathbf{a}}(q)}]\!]_{d-rd_C}$. This contribution depends on the embedding into $\mathbb{P}^{n-1}$, which is as expected, given the nature of SQ invariants.

Since the embedding $C \to \mathbb{P}^{n-1}$ corresponds to an inclusion $\mathcal{O}_{\mathbb{P}^1}(-d_C) \to \mathbb{C}^n \otimes \mathcal{O}_{\mathbb{P}^1}$, each element of $\mathcal{Z}^{1,1}_{C;r}(d)$ corresponds to a tuple

$$(\mathcal{C}, y_1, y_2; S \subset S'^{\otimes d_c}, S' \subset \mathbb{C}^2 \otimes \mathcal{O}_{\mathcal{C}}), \quad \text{where}$$
$$(\mathcal{C}, y_1, y_2; S \subset \mathbb{C}^2 \otimes \mathcal{O}_{\mathcal{C}}) \in \overline{Q}_{0,2}(\mathbb{P}^1, d), \quad (\mathcal{C}, y_1, y_2; S' \subset \mathbb{C}^2 \otimes \mathcal{O}_{\mathcal{C}}) \in \overline{Q}_{0,2}(\mathbb{P}^1, r).$$

This modular style definition readily extends to arbitrary genus, number of marked points, and dimension of projective space. The arising deformation-obstruction theory can be studied as in [Marian et al. 2011, Section 6].

1D. *Outline of the paper.* Theorem 1 is a direct consequence of Theorem 2 in Section 2, which in turn is the nonequivariant specialization of Theorem 3 in Theorem 3. We adapt the approaches of [Zinger 2009; Popa and Zinger 2014; Popa 2012] from Gromov–Witten theory, outlined in Sections 5 and 6, to show that certain equivariant two-point generating functions, including the stable quotients analogue of the double Givental's J-function, satisfy certain good properties which guarantee uniqueness. The proof that these generating functions satisfy the required properties follows principles similar to the proof of the analogous statements in [Zinger 2009; Popa and Zinger 2014; Popa 2012] and uses the localization theorem of [Atiyah and Bott 1984]; it is carried out in Sections 7 and 8.

This approach also implies that certain equivariant three-point generating functions, including the stable quotients analogue of the triple Givental's J-function, are determined by three-point primary (without ψ-classes) SQ invariants. In the Fano cases, that is, $\nu_n(\mathbf{a}) > 0$, enough of these invariants are essentially trivial for dimensional reasons to confirm Proposition 3.1 in these cases; see Corollary 9.1. However, there is no dimensional reason for the vanishing of these invariants to extend to the Calabi–Yau cases, that is, $\nu_n(\mathbf{a}) = 0$; thus, a different argument is needed in these cases. We employ the same kind of trick as used in [Cooper and Zinger 2014] to confirm mirror symmetry for the stable quotients analogue of Givental's J-function and essentially deduce the Calabi–Yau cases from the Fano cases. Specifically, we show that the equivariant three-point mirror formula of Proposition 3.1 is equivalent to the closed formula for twisted three-point Hurwitz numbers of Proposition 4.1, whenever $|\mathbf{a}| \le n$. In Section 9, we show that the validity of the latter does not depend n; since it holds whenever $|\mathbf{a}| < n$, it follows that it holds for all $\mathbf{a}$, and so the equivariant three-point mirror formula of Proposition 3.1 holds whenever $|\mathbf{a}| \le n$. Along with [Zinger 2014], Proposition 3.1 finally leads to the mirror formula for the stable quotients analogue of the triple Givental's J-function in Theorem 3; see Section 10.

The closed formulas for twisted Hurwitz numbers of Propositions 4.1 and 4.2 are among the key ingredients in computing the genus-1 twisted stable quotients invariants with 1 marked point. At the same time, this paper and [Zinger 2014] provide an approach to comparing the (equivariant) genus-g m-fold Givental's J-functions,

$$\sum_{d=0}^{\infty} q^d \{\mathrm{ev}_1 \times \cdots \times \mathrm{ev}_m\}_* \left[\frac{e(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})}{(\hbar_1 - \psi_1)\cdots(\hbar_m - \psi_m)} \right] \in H^*(\mathbb{P}^{n-1})[\hbar_1^{-1}, \ldots, \hbar_m^{-1}][\![q]\!] \tag{1-12}$$

in the SQ and GW theories for all $g \ge 0$ and $m \ge 1$ with $2g + m \ge 2$. By Proposition 6.3 and Lemmas 6.5 and 6.6, in the genus-0 case the restrictions of these generating functions to insertions at only one marked point agree whenever $\nu_{\mathbf{a}} > 1$. In all cases, the approach of [Zinger 2014] can be adapted to show that (1-12) is a sum over (at least) trivalent m-marked graphs with coefficients that involve equivariant m'-pointed Hurwitz numbers with $m' \le m$, which are conversely completely determined by the stable quotients analogue of the m'-pointed Givental's J-function with insertions at only one marked point through relations that do not involve n. Since these relations hold whenever $\nu_n(\mathbf{a}) > 0$, they hold for all $\mathbf{a}$. We intend to clarify these points in a future paper.

The Gromov–Witten analogues of Theorem 2 and its equivariant counterpart, Theorem 3 in Section 3, extend to the so-called concavex vector bundles over products of projective spaces, that is, vector bundles of the form

$$\bigoplus_{k=1}^{l} \mathcal{O}_{\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1}}(a_{k;1},\dots,a_{k;p}) \to \mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},$$

where for each given $k=1,2,\dots,l$ either $a_{k;1},\dots,a_{k;p}\in\mathbb{Z}^{\ge 0}$, with $a_{k;i}\neq 0$ for some i, or $a_{k;1},\dots,a_{k;p}\in\mathbb{Z}^{-}$. The stable quotients analogue of these bundles are the sheaves

$$\bigoplus_{k=1}^{l} \mathcal{S}_1^{*a_{k;1}}\otimes\cdots\otimes\mathcal{S}_p^{*a_{k;p}} \to \mathcal{U} \to \overline{Q}_{0,2}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(d_1,\dots,d_p)) \tag{1-13}$$

with the same condition on $a_{k;i}$, where $\mathcal{S}_i\to\mathcal{U}$ is the universal subsheaf corresponding to the i-th factor. We will comment on the necessary modifications at each step of the proof.

2. Main theorem

We arrange stable quotients invariants with two and three marked points into generating functions in Section 2A and give explicit closed formulas for them in Section 2B. In Section 2C, we use these formulas to relate SQ and GW invariants, with descendants, and obtain replacements for the divisor, string, and dilaton relations for SQ invariants.

2A. ***Givental's J-functions.*** For computational purposes, it is convenient to define variations of the bundle (1-3) by

$$\begin{aligned}\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)} &= \bigoplus_{a_k>0} R^0\pi_*(\mathcal{S}^{*a_k}(-\sigma_1)) \oplus \bigoplus_{a_k<0} R^1\pi_*(\mathcal{S}^{*a_k}(-\sigma_1)) \to \overline{Q}_{0,m}(\mathbb{P}^{n-1},d),\\ \ddot{\mathcal{V}}_{n;\mathbf{a}}^{(d)} &= \bigoplus_{a_k>0} R^0\pi_*(\mathcal{S}^{*a_k}(-\sigma_2)) \oplus \bigoplus_{a_k<0} R^1\pi_*(\mathcal{S}^{*a_k}(-\sigma_2)) \to \overline{Q}_{0,m}(\mathbb{P}^{n-1},d),\end{aligned} \tag{2-1}$$

where $n, d \in \mathbb{Z}^+$, $m \geq 2$, and $\pi : \mathcal{U} \to \overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$ is the universal curve; these sheaves are also locally free. Whenever $\nu_n(\mathbf{a}) \geq 0$, [Cooper and Zinger 2014, Theorem 1] provides an explicit closed formula for the stable quotients analogue of Givental's J-function: the power series

$$\dot{Z}_{n;\mathbf{a}}(x, \hbar, q) \equiv 1 + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{e(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})}{\hbar - \psi_1}\right] \in H^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!], \tag{2-2}$$

where $\mathrm{ev}_1 : \overline{Q}_{0,2}(\mathbb{P}^{n-1}, d) \to \mathbb{P}^{n-1}$ is as before and $x \in H^2(\mathbb{P}^{n-1})$ is the hyperplane class. In this paper, we obtain a closed formula for the power series

$$\ddot{Z}_{n;\mathbf{a}}(x, \hbar, q) \equiv 1 + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{e(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})}{\hbar - \psi_1}\right] \in H^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!]; \tag{2-3}$$

see (2-26).

We also give explicit formulas for the stable quotients analogues of the double and triple Givental's J-functions, the power series

$$\begin{aligned} &\dot{Z}^*_{n;\mathbf{a}}(x_1, x_2, \hbar_1, \hbar_2, q) \\ &\quad \equiv \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2\}_*\left[\frac{e(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})}{(\hbar_1 - \psi_1)(\hbar_2 - \psi_2)}\right] \in H^*(\mathbb{P}^{n-1})[\hbar_1^{-1}, \hbar_2^{-1}][\![q]\!], \end{aligned} \tag{2-4}$$

$$\begin{aligned} &\dot{Z}^*_{n;\mathbf{a}}(x_1, x_2, x_3, \hbar_1, \hbar_2, \hbar_3, q) \\ &\quad \equiv \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2 \times \mathrm{ev}_3\}_*\left[\frac{e(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})}{(\hbar_1 - \psi_1)(\hbar_2 - \psi_2)(\hbar_3 - \psi_3)}\right], \end{aligned} \tag{2-5}$$

where $x_i = \pi_i^* x$ is the pullback of the hyperplane class in $\mathbb{P}^{n-1}$ by the i-th projection map and

$$\begin{aligned} \mathrm{ev}_1 \times \mathrm{ev}_2 &: \overline{Q}_{0,2}(\mathbb{P}^{n-1}, d) \to \mathbb{P}^{n-1} \times \mathbb{P}^{n-1}, \\ \mathrm{ev}_1 \times \mathrm{ev}_2 \times \mathrm{ev}_3 &: \overline{Q}_{0,3}(\mathbb{P}^{n-1}, d) \to \mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \end{aligned} \tag{2-6}$$

are the total evaluation maps. Let

$$\begin{aligned} &\dot{Z}_{n;\mathbf{a}}(x_1, x_2, \hbar_1, \hbar_2, q) \\ &\quad = \left(\frac{1}{\hbar_1 + \hbar_2} \sum_{\substack{s_1, s_2 \geq 0 \\ s_1 + s_2 = n-1}} x_1^{s_1} x_2^{s_2}\right) + \dot{Z}^*_{n;\mathbf{a}}(x_1, x_2, \hbar_1, \hbar_2, q), \\ &\dot{Z}_{n;\mathbf{a}}(x_1, x_2, x_3, \hbar_1, \hbar_2, \hbar_3, q) \\ &\quad = \left(\frac{1}{\hbar_1 \hbar_2 \hbar_3} \sum_{\substack{s_1, s_2, s_3 \geq 0 \\ s_1, s_2, s_3 \leq n-1 \\ s_1 + s_2 + s_3 = 2n-2}} x_1^{s_1} x_2^{s_2} x_3^{s_3}\right) + \dot{Z}^*_{n;\mathbf{a}}(x_1, x_2, x_3, \hbar_1, \hbar_2, \hbar_3, q). \end{aligned} \tag{2-7}$$

For each $s \in \mathbb{Z}^{\geq 0}$, define

$$\begin{aligned} \dot{Z}^{(s)}_{n;\mathbf{a}}(x,\hbar,q) &\equiv x^s + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{e(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^* x^s}{\hbar-\psi_1}\right] \in H^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!], \\ \ddot{Z}^{(s)}_{n;\mathbf{a}}(x,\hbar,q) &\equiv x^s + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{e(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^* x^s}{\hbar-\psi_1}\right] \in H^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!], \end{aligned} \tag{2-8}$$

where $\mathrm{ev}_1, \mathrm{ev}_2 : \overline{Q}_{0,2}(\mathbb{P}^{n-1}, d) \to \mathbb{P}^{n-1}$. Thus, $\dot{Z}^{(0)}_{n;\mathbf{a}} = \dot{Z}_{n;\mathbf{a}}$, $\ddot{Z}^{(0)}_{n;\mathbf{a}} = \ddot{Z}_{n;\mathbf{a}}$, and

$$x^{\ell^{+}_{-}(\mathbf{a})}\dot{Z}^{(\ell^{-}_{+}(\mathbf{a})+s)}_{n;\mathbf{a}}(x,\hbar,q) = x^{\ell^{-}_{+}(\mathbf{a})}\ddot{Z}^{(\ell^{+}_{-}(\mathbf{a})+s)}_{n;\mathbf{a}}(x,\hbar,q)$$

for all $s \geq 0$, where

$$\ell^{+}_{-}(\mathbf{a}) = \max(\ell(\mathbf{a}), 0), \quad \ell^{-}_{+}(\mathbf{a}) = \max(-\ell(\mathbf{a}), 0).$$

By Theorem 2 below, $\dot{Z}^{(s)}_{n;\mathbf{a}}$, $\ddot{Z}^{(s)}_{n;\mathbf{a}}$, and the stable quotients analogues of the double and triple Givental's J-functions, (2-4) and (2-5), are explicit transforms of Givental's J-function $\dot{Z}_{n;\mathbf{a}}$ and its "reflection" $\ddot{Z}_{n;\mathbf{a}}$; this transform depends only on $\mathbf{a}$ (and s in the first two cases).

2B. *Mirror symmetry.* Givental's J-function $\dot{Z}_{n;\mathbf{a}}$ and its "reflection" $\ddot{Z}_{n;\mathbf{a}}$ in the Gromov–Witten and stable quotients theories are described by the hypergeometric series (1-6) and

$$\ddot{F}_{n;\mathbf{a}}(w,q) \equiv \sum_{d=0}^{\infty} q^d w^{\nu_n(\mathbf{a})d} \frac{\prod_{a_k>0}\prod_{r=0}^{a_kd-1}(a_kw+r)\prod_{a_k<0}\prod_{r=1}^{-a_kd}(a_kw-r)}{\prod_{r=1}^{d}((w+r)^n-w^n)} \in \mathbb{Q}(w)[\![q]\!]. \tag{2-9}$$

These are power series in q with constant term 1 whose coefficients are rational functions in w which are regular at $w=0$. We denote the subgroup of all such power series by $\mathcal{P}$ and define

$$\begin{aligned} &\mathbf{D} : \mathbb{Q}(w)[\![q]\!] \to \mathbb{Q}(w)[\![q]\!], && \mathbf{M} : \mathcal{P} \to \mathcal{P} \quad \text{by} \\ &\mathbf{D}H(w,q) \equiv \left\{1+\frac{q}{w}\frac{\mathrm{d}}{\mathrm{d}q}\right\}H(w,q), && \mathbf{M}H(w,q) \equiv \mathbf{D}\left(\frac{H(w,q)}{H(0,q)}\right); \end{aligned} \tag{2-10}$$

the operator $\mathbf{D}$ multiplies the coefficient of q^d by $(w+d)/w$. If $\nu_n(\mathbf{a}) = 0$ and $s \in \mathbb{Z}^{\geq 0}$, let

$$\dot{I}_s(q) \equiv \mathbf{M}^s \dot{F}_{n;\mathbf{a}}(0,q), \quad \ddot{I}_s(q) \equiv \mathbf{M}^s \ddot{F}_{n;\mathbf{a}}(0,q). \tag{2-11}$$

For example, $\dot{I}_s(q) = 1$ if $s < \ell^-(\mathbf{a})$, $\ddot{I}_s(q) = 1$ if $s < \ell^+(\mathbf{a})$,

$$\dot{I}_{\ell^-(\mathbf{a})}(q) = \ddot{I}_{\ell^+(\mathbf{a})}(q) = \sum_{d=0}^{\infty} q^d \frac{\prod_{a_k>0}(a_k d)! \prod_{a_k<0}(-1)^{a_k d}(-a_k d)!}{(d!)^n} \quad \text{if } \nu_n(\mathbf{a}) = 0,$$

and more generally $\dot{I}_{s+\ell^-_+(\mathbf{a})}(q) = \ddot{I}_{s+\ell^+_-(\mathbf{a})}(q)$ for all $s \geq 0$. If $\nu_n(\mathbf{a}) > 0$, we set $\dot{I}_s(q), \ddot{I}_s(q) = 1$.

It is also convenient to introduce

$$\text{(2-12)} \quad F_{n;\mathbf{a}}(w,q) \equiv \sum_{d=0}^{\infty} q^d w^{\nu_n(\mathbf{a})d} \frac{\prod_{a_k>0}\prod_{r=1}^{a_k d}(a_k w + r) \prod_{a_k<0}\prod_{r=1}^{-a_k d}(a_k w - r)}{\prod_{r=1}^{d}(w+r)^n} \in \mathbb{Q}(w)[\![q]\!]$$

and the associated power series $I_s(q) = \mathbf{M}^s F_{n;\mathbf{a}}(0,q)$ in the $\nu_n(\mathbf{a}) = 0$ case. In the case $0 < \nu_n(\mathbf{a}) < n$, we define $\mathrm{c}^{(d)}_{s,s'} \in \mathbb{Q}$ with $d, s, s' \geq 0$ by

$$\text{(2-13)} \quad \begin{aligned} \sum_{d=0}^{\infty}\sum_{s'=0}^{\infty} \mathrm{c}^{(d)}_{s,s'} w^{s'} q^d &\equiv w^s \mathbf{D}^s F_{n;\mathbf{a}}(w, q/w^{\nu_n(\mathbf{a})}) \\ &= w^s \mathbf{D}^{s+\ell^-(\mathbf{a})} \dot{F}_{n;\mathbf{a}}(w, q/w^{\nu_n(\mathbf{a})}) \\ &= w^s \mathbf{D}^{s+\ell^+(\mathbf{a})} \ddot{F}_{n;\mathbf{a}}(w, q/w^{\nu_n(\mathbf{a})}). \end{aligned}$$

Since $\mathrm{c}^{(0)}_{s,s'} = \delta_{s,s'}$, the relations

$$\text{(2-14)} \quad \sum_{\substack{d_1,d_2\geq 0\\ d_1+d_2=d}} \sum_{t=0}^{s-\nu_n(\mathbf{a})d_1} \tilde{\mathrm{c}}^{(d_1)}_{s,t} \mathrm{c}^{(d_2)}_{t,s'} = \delta_{d,0}\delta_{s,s'} \quad \text{for all } d, s' \in \mathbb{Z}^{\geq 0},\ s' \leq s - \nu_n(\mathbf{a})d,$$

inductively define $\tilde{\mathrm{c}}^{(d)}_{s,s'} \in \mathbb{Q}$ in terms of the numbers $\tilde{\mathrm{c}}^{(d_1)}_{s,t}$ with $d_1 < d$. For example, $\tilde{\mathrm{c}}^{(0)}_{s,s'} = \delta_{s,s'}$ and

$$\sum_{s'=0}^{s-\nu_n(\mathbf{a})} \tilde{\mathrm{c}}^{(1)}_{s,s'} w^{s'} + \prod_{k=1}^{l} a_k \frac{\prod_{a_k>0}\prod_{r=1}^{a_k-1}(a_k w + r) \prod_{a_k<0}\prod_{r=1}^{-a_k-1}(a_k w - r)}{(w+1)^{n-\ell^+(\mathbf{a})-\ell^-(\mathbf{a})-s}} \in w^{s-\nu_n(\mathbf{a})+1}\mathbb{Q}[\![w]\!].$$

If $s' < 0$ or $\nu_n(\mathbf{a}) = 0, n$, we set $\tilde{\mathrm{c}}^{(d)}_{s,s'} = \delta_{d,0}\delta_{s,s'}$. The coefficients $\tilde{\mathrm{c}}^{(d)}_{s,s'}$ are used to express the power series (2-7) and (2-8) in terms of derivatives of the power series (2-2) and (2-3); see Theorem 2.

For $s_1, s_2, s_3, d \in \mathbb{Z}^{\ge 0}$ with $s_1, s_2, s_3 \le n-1$, let

$$\tilde{\mathrm{c}}^{(d)}_{s_1,s_2,s_3} = \begin{cases} \Big[\!\Big[\big((1-\mathbf{a}^{\mathbf{a}}q)\dot{\mathbb{I}}^c_{s_1}(q)\ddot{\mathbb{I}}^c_{s_2}(q)\ddot{\mathbb{I}}^c_{s_3}(q)\big)^{-1}\Big]\!\Big]_d & \text{if } \nu_n(\mathbf{a})=0; \\ \displaystyle\sum_{\substack{d_0,d_1,d_2,d_3\ge 0\\ d_0+d_1+d_2+d_3=d}} (\mathbf{a}^{\mathbf{a}})^{d_0} \prod_{t=1}^{3} \tilde{\mathrm{c}}^{(d_t)}_{\hat{s}_t-\ell_t(\mathbf{a}),\hat{s}_t-\nu_n(\mathbf{a})d_t-\ell_t(\mathbf{a})} & \text{if } \nu_n(\mathbf{a})>0; \end{cases} \tag{2-15}$$

where

$$\dot{\mathbb{I}}^c_s = \prod_{t=s+1}^{n-\ell^+(\mathbf{a})} \dot{I}_t, \quad \ddot{\mathbb{I}}^c_s = \prod_{t=s+1}^{n-\ell^-(\mathbf{a})} \ddot{I}_t, \quad \hat{s}_t = n-1-s_t, \tag{2-16}$$
$$\ell_t(\mathbf{a}) = \begin{cases} \ell^+(\mathbf{a}) & \text{if } t=1; \\ \ell^-(\mathbf{a}) & \text{if } t=2,3; \end{cases}$$

and $[\![f(q)]\!]_d$ is the coefficient of q^d of $f(q) \in \mathbb{Q}[\![q]\!]$. In particular, $\dot{\mathbb{I}}^c_s = 1$ if $s \ge n-\ell^+(\mathbf{a})$ and $\ddot{\mathbb{I}}^c_s = 1$ if $s \ge n-\ell^-(\mathbf{a})$. Since $I_t = \dot{I}_{t+\ell^-(\mathbf{a})} = \ddot{I}_{t+\ell^+(\mathbf{a})}$, we find that

$$\dot{\mathbb{I}}^c_s(q) = (1-\mathbf{a}^{\mathbf{a}}q)^{-1} \text{if } s<\ell^-(\mathbf{a}), \quad \ddot{\mathbb{I}}^c_s(q) = (1-\mathbf{a}^{\mathbf{a}}q)^{-1} \text{if } s<\ell^+(\mathbf{a});$$

see [Zinger 2014, Proposition 4.4]. This implies that

$$\sum_{d=0}^{\infty} \tilde{\mathrm{c}}^{(d)}_{s_1,s_2,s_3} q^d = 1 \ \text{ if } \nu_n(\mathbf{a})=0, \quad s_1+s_2+s_3=2n-2, \tag{2-17}$$
$$\min(s_1,s_2,s_3) < \ell^-(\mathbf{a}).$$

We use this observation in Section 2C. Since $\tilde{\mathrm{c}}^{(0)}_{s,s'} = \delta_{s,s'}$, $\tilde{\mathrm{c}}^{(0)}_{s_1,s_2,s_3} = 1$.

Finally, we define $\mathfrak{D}^s \dot{Z}_{n;\mathbf{a}}(x,\hbar,q)$, $\mathfrak{D}^s \ddot{Z}_{n;\mathbf{a}}(x,\hbar,q) \in H^*(\mathbb{P}^{n-1})[\hbar][\![q]\!]$ for each $s \in \mathbb{Z}^+$ inductively by

$$\begin{aligned} \mathfrak{D}^0 \dot{Z}_{n;\mathbf{a}}(x,\hbar,q) &= \dot{Z}_{n;\mathbf{a}}(x,\hbar,q), \\ \mathfrak{D}^s \dot{Z}_{n;\mathbf{a}}(x,\hbar,q) &= \frac{1}{\dot{I}_s(q)}\Big\{x+\hbar\, q\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\dot{Z}_{n;\mathbf{a}}(x,\hbar,q), \\ \mathfrak{D}^0 \ddot{Z}_{n;\mathbf{a}}(x,\hbar,q) &= \ddot{Z}_{n;\mathbf{a}}(x,\hbar,q), \\ \mathfrak{D}^s \ddot{Z}_{n;\mathbf{a}}(x,\hbar,q) &= \frac{1}{\ddot{I}_s(q)}\Big\{x+\hbar\, q\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\ddot{Z}_{n;\mathbf{a}}(x,\hbar,q). \end{aligned} \tag{2-18}$$

The operator $\mathfrak{D}$ first multiplies the coefficient of q^d by $x+d\hbar$ and then renormalizes the power series so that the coefficient of x^s becomes 1 in the Calabi–Yau cases (there is no renormalization in the Fano cases).

Theorem 2. *If $l \in \mathbb{Z}^{\geq 0}$, $n \in \mathbb{Z}^{+}$, and $\mathbf{a} \in (\mathbb{Z}^*)^l$, the stable quotients analogue of the double Givental's J-function satisfies*

$$\dot{Z}_{n;\mathbf{a}}(x_1, x_2, \hbar_1, \hbar_2, q) = \frac{1}{\hbar_1 + \hbar_2} \sum_{\substack{s_1, s_2 \geq 0 \\ s_1 + s_2 = n-1}} \dot{Z}^{(s_1)}_{n;\mathbf{a}}(x_1, \hbar_1, q) \ddot{Z}^{(s_2)}_{n;\mathbf{a}}(x_2, \hbar_2, q). \tag{2-19}$$

If in addition $\nu_n(\mathbf{a}) \geq 0$,

$$\begin{aligned} \dot{Z}_{n;\mathbf{a}}&(x_1, x_2, x_3, \hbar_1, \hbar_2, \hbar_3, q) \\ &= \frac{1}{\hbar_1 \hbar_2 \hbar_3} \sum_{\substack{d, s_1, s_2, s_3 \geq 0 \\ s_1, s_2, s_3 \leq n-1 \\ s_1 + s_2 + s_3 + \nu_n(\mathbf{a}) d = 2n-2}} \tilde{\mathrm{c}}^{(d)}_{s_1, s_2, s_3} q^d \dot{Z}^{(s_1)}_{n;\mathbf{a}}(x_1, \hbar_1, q) \prod_{t=2}^{3} \ddot{Z}^{(s_t)}_{n;\mathbf{a}}(x_t, \hbar_t, q), \end{aligned} \tag{2-20}$$

$$\check{Z}^{(s)}_{n;\mathbf{a}}(x, \hbar, q) = \sum_{d=0}^{\infty} \sum_{s'=0}^{s - \nu_n(\mathbf{a}) d} \tilde{\mathrm{c}}^{(d)}_{s - \ell^*(\mathbf{a}), s' - \ell^*(\mathbf{a})} q^d \hbar^{s - \nu_n(\mathbf{a}) d - s'} \mathfrak{D}^{s'} \check{Z}_{n;\mathbf{a}}(x, \hbar, q), \tag{2-21}$$

where $(\check{Z}, \ell^*) = (\dot{Z}, \ell^-), (\ddot{Z}, \ell^+)$.

2C. ***Some computations.*** The first identity in Theorem 2 also holds for the Gromov–Witten analogues of the generating series $\dot{Z}^*_{n;\mathbf{a}}$, $\dot{Z}^{(s)}_{n;\mathbf{a}}$, and $\ddot{Z}^{(s)}_{n;\mathbf{a}}$; see [Popa 2012, Theorem 1.2] for the general (toric) case. If $\nu_n(\mathbf{a}) \geq 2 - \ell^-(\mathbf{a})$, the analogues of (2-20), (2-21), (2-26), and (2-27) hold in Gromov–Witten theory as well. Thus, in this case the double Givental's J-functions in Gromov–Witten and stable quotients theories agree. If $\nu_n(\mathbf{a}) = 1$ and $\ell^-(\mathbf{a}) = 0$, the analogue of (2-21) in Gromov–Witten theory holds with $\{x + \hbar q\, \mathrm{d}/\mathrm{d}q\}$ replaced by $\{\mathbf{a}! q + x + \hbar q\, \mathrm{d}/\mathrm{d}q\}$ in (2-18). Finally, if $\nu_n(\mathbf{a}) = 0$ and $\ell^-(\mathbf{a}) \leq 1$, the analogue of (2-21) in Gromov–Witten theory holds with

$$\mathfrak{D}^s \dot{Z}_{n;\mathbf{a}}(x, \hbar, Q) = \frac{\dot{I}_1(q)}{\dot{I}_s(q)} \Big\{ x + \hbar Q \frac{\mathrm{d}}{\mathrm{d}Q} \Big\} \mathfrak{D}^{s-1} \dot{Z}_{n;\mathbf{a}}(x, \hbar, Q)$$

$$\mathfrak{D}^s \ddot{Z}_{n;\mathbf{a}}(x, \hbar, Q) = \frac{\dot{I}_1(q)}{\ddot{I}_s(q)} \Big\{ x + \hbar Q \frac{\mathrm{d}}{\mathrm{d}Q} \Big\} \mathfrak{D}^{s-1} \ddot{Z}_{n;\mathbf{a}}(x, \hbar, Q)$$

for all $s \in \mathbb{Z}^+$, where $Q = q \mathrm{e}^{J_{n;\mathbf{a}}(q)}$. The same comparison applies to the equivariant version of Theorem 2, Theorem 3 in Section 3, and its Gromov–Witten analogue; see [Popa 2012, Theorem 4.1] for the general toric case. Thus, just as is the case for the standard Givental's J-function, the mirror formulas for the double Givental's J-function in the stable quotients theory are simpler versions of the mirror formulas for the double Givental's J-function in the Gromov–Witten theory. Furthermore, just as in Gromov–Witten theory, the generating functions $\dot{Z}^{(s)}_{n;\mathbf{a}}$, $\ddot{Z}^{(s)}_{n;\mathbf{a}}$, and $\dot{Z}^*_{n;\mathbf{a}}$ above do not change when the tuple $(a_1, \ldots, a_l)$ is replaced by $(a_1, \ldots, a_l, 1)$; this is consistent with [Ciocan-Fontanine et al. 2014, Proposition 4.6.1].

Comparing Theorem 2 and [Cooper and Zinger 2014, Equation (1.7)] with [Popa 2012, Theorem 1.2] and the $m=3$ case of [Zinger 2014, Theorem A], we find that

$$\begin{aligned}(2\text{-}22)\quad & \dot{Z}^{\mathrm{GW}}_{n;\mathbf{a}}(x_1,\dots,x_m,\hbar_1,\dots,\hbar_m,Q)\\ &= \dot{I}_0(q)^{m-2}\mathrm{e}^{-J_{n;\mathbf{a}}(q)(x_1/\hbar_1+\dots+x_m/\hbar_m)}\dot{Z}_{n;\mathbf{a}}(x_1,\dots,x_m,\hbar_1,\dots,\hbar_m,q)\end{aligned}$$

with $Q=q\mathrm{e}^{J_{n;\mathbf{a}}(q)}$ as before and $m=2,3$; we intend to extend this comparison to $m>3$ in a future paper. The same relations hold between the generating series $Z_{n;\mathbf{a}}$ described below. For $m=2,3$ and $b_1,b_2,b_3,c_1,c_2,c_3\ge 0$, let

$$\mathrm{SQ}^0_{n;\mathbf{a}}(\tau_{b_1}c_1,\tau_{b_2}c_2,\tau_{b_3}c_3)=\begin{cases}\langle\mathbf{a}\rangle & \text{if } b_1,b_2,b_3=0,\ c_1+c_2+c_3=n-1-\ell(\mathbf{a});\\ 0 & \text{otherwise;}\end{cases}$$

$$\mathrm{SQ}^0_{n;\mathbf{a}}(\tau_{b_1}c_1,\tau_{b_2}c_2)=\begin{cases}\langle\mathbf{a}\rangle & \text{if } b_1,b_2=0,\ c_1+c_2=n-2-\ell(\mathbf{a});\\ \dfrac{\langle\mathbf{a}\rangle}{2} & \text{if } \{b_1,b_2\}=\{0,-1\},\ c_1+c_2=n-1-\ell(\mathbf{a});\\ 0 & \text{otherwise;}\end{cases}$$

$$\mathrm{SQ}^d_{n;\mathbf{a}}(\tau_{b_1}c_1,\dots,\tau_{b_m}c_m)=\int_{\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)}\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}})\prod_{i=1}^{m}(\psi_i^{b_i}\mathrm{ev}_i^*x^{c_i})\quad\text{for all } d\in\mathbb{Z}^+,$$

$$\mathrm{SQ}^{c_1,\dots,c_m}_{n;\mathbf{a}}(q)_{b_1,\dots,b_m}=\sum_{d=0}^{\infty}q^d\,\mathrm{SQ}^d_{n;\mathbf{a}}(\tau_{b_1}c_1,\dots,\tau_{b_m}c_m).$$

For $m=3$, the degree-0 terms are as expected; for $m=2$, the degree-0 terms are chosen to insure the necessary recursivity and polynomiality properties, as outlined in Section 5. Since GW invariants satisfy the divisor, string, and dilaton relations, (2-22) leads to modified versions of these relations for SQ invariants:

$$\begin{aligned}(2\text{-}23)\quad & \dot{I}_0(q)\dot{I}_1(q)\mathrm{SQ}^{c_1,c_2,1}_{n;\mathbf{a}}(q)_{b_1,b_2,0}=\\ &\qquad q\frac{\mathrm{d}}{\mathrm{d}q}\mathrm{SQ}^{c_1,c_2}_{n;\mathbf{a}}(q)_{b_1,b_2}+\mathrm{SQ}^{c_1+1,c_2}_{n;\mathbf{a}}(q)_{b_1-1,b_2}+\mathrm{SQ}^{c_1,c_2+1}_{n;\mathbf{a}}(q)_{b_1,b_2-1},\end{aligned}$$

$$(2\text{-}24)\qquad \dot{I}_0\mathrm{SQ}^{c_1,c_2,0}_{n;\mathbf{a}}(q)_{b_1,b_2,0}=\mathrm{SQ}^{c_1,c_2}_{n;\mathbf{a}}(q)_{b_1-1,b_2}+\mathrm{SQ}^{c_1,c_2}_{n;\mathbf{a}}(q)_{b_1,b_2-1},$$

$$(2\text{-}25)\qquad \mathrm{SQ}^{c_1,c_2,0}_{n;\mathbf{a}}(q)_{b_1,b_2,1}=-J_{n;\mathbf{a}}(q)\mathrm{SQ}^{c_1,c_2,1}_{n;\mathbf{a}}(q)_{b_1,b_2,0}.$$

The discrepancy from the corresponding relations of GW invariants is exhibited by the power series $\dot{I}_0$ and $\dot{I}_1$ (or equivalently $J_{n;\mathbf{a}}(q)$).

By (3-12), (3-9), (1-6), and (2-9)

(2-26)

$$\dot{Z}_{n;\mathbf{a}}(x,\hbar,q)=\frac{\dot{F}_{n;\mathbf{a}}(x/h,q/x^{\nu_n(\mathbf{a})})}{\dot{I}_0(q)},\qquad \ddot{Z}_{n;\mathbf{a}}(x,\hbar,q)=\frac{\ddot{F}_{n;\mathbf{a}}(x/h,q/x^{\nu_n(\mathbf{a})})}{\ddot{I}_0(q)},$$

if $\nu_n(a) \geq 0$.[2] These two formulas explicitly determine the basic stable quotients invariants appearing in (2-2) and (2-3). For $s \in \mathbb{Z}^+$, define

$$\mathfrak{D}^0 \dot{F}_{n;\mathbf{a}}(w,q) = \frac{\dot{F}_{n;\mathbf{a}}(w,q)}{\dot{I}_0(q)}, \quad \mathfrak{D}^s \dot{F}_{n;\mathbf{a}}(w,q) = \frac{1}{\dot{I}_s(q)} \Big\{ 1 + \frac{q}{w}\frac{\mathrm{d}}{\mathrm{d}q} \Big\} \mathfrak{D}^{s-1} \dot{F}_{n;\mathbf{a}}(w,q),$$
$$\mathfrak{D}^0 \ddot{F}_{n;\mathbf{a}}(w,q) = \frac{\ddot{F}_{n;\mathbf{a}}(w,q)}{\ddot{I}_0(q)}, \quad \mathfrak{D}^s \ddot{F}_{n;\mathbf{a}}(w,q) = \frac{1}{\ddot{I}_s(q)} \Big\{ 1 + \frac{q}{w}\frac{\mathrm{d}}{\mathrm{d}q} \Big\} \mathfrak{D}^{s-1} \ddot{F}_{n;\mathbf{a}}(w,q).$$

Combining (2-26) with (2-19) and (2-21), we find that

$$\begin{aligned} \text{(2-27)} \quad & \dot{Z}_{n;\mathbf{a}}(x_1,x_2,\hbar_1,\hbar_2,q) \\ & = \frac{1}{\hbar_1+\hbar_2} \sum_{\substack{s_1,s_2 \geq 0 \\ s_1+s_2=n-1}} x_1^{s_1} \dot{F}_{n;\mathbf{a}}^{(s_1)}\Big(\frac{x_1}{\hbar_1}, \frac{q}{x_1^{\nu_n(\mathbf{a})}}\Big) \cdot x_2^{s_2} \ddot{F}_{n;\mathbf{a}}^{(s_2)}\Big(\frac{x_2}{\hbar_2}, \frac{q}{x_2^{\nu_n(\mathbf{a})}}\Big), \end{aligned}$$

where

$$\text{(2-28)} \qquad \check{F}_{n;\mathbf{a}}^{(s)}(w,q) = \sum_{d=0}^{\infty} \sum_{s'=0}^{s-\nu_n(\mathbf{a})d} \frac{\tilde{\mathrm{c}}^{(d)}_{s-\ell^*(\mathbf{a}),s'-\ell^*(\mathbf{a})} q^d}{w^{s-\nu_n(\mathbf{a})d-s'}} \mathfrak{D}^{s'} \check{F}_{n;\mathbf{a}}(w,q),$$

with $(\check{F}, \ell^*) = (\dot{F}, \ell^-), (\ddot{F}, \ell^+)$.[3] Thus, (2-26) and Theorem 2 provide closed formulas for the twisted genus-0 two-point and three-point SQ invariants of projective spaces.

The equivariant versions of the generating functions $\dot{Z}_{n;\mathbf{a}}$ defined in (2-7) are ideally suited for further computations, such as of genus-0 invariants with more marked points and of positive-genus twisted invariants with at least one marked point. However, for the purposes of computing the genus-0 two-point and three-point invariants, it is more natural to consider the generating functions

(2-29)

$$Z^*_{n;\mathbf{a}}(x_1,x_2,\hbar_1,\hbar_2,q) \equiv \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2\}_* \Big[\frac{e(\mathcal{V}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)} \Big],$$
$$\begin{aligned} & Z^*_{n;\mathbf{a}}(x_1,x_2,x_3,\hbar_1,\hbar_2,\hbar_3,q) \equiv \\ & \qquad \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2 \times \mathrm{ev}_3\}_* \Big[\frac{e(\mathcal{V}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)(\hbar_3-\psi_3)} \Big], \end{aligned}$$

[2]The right-hand sides of these expressions should be first simplified in $\mathbb{Q}(x,\hbar)[[q]]$, eliminating division by x, and then projected to $H^*(\mathbb{P}^{n-1})[\hbar][[q]]$.

[3]The right-hand side of (2-27) should be first simplified in $\mathbb{Q}(x_1,x_2,\hbar_1,\hbar_2)[[q]]$, eliminating division by x_1 and x_2, and then projected to $H^*(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1})[\hbar_1,\hbar_2][[q]]$.

where $\mathcal{V}_{n;\mathbf{a}}^{(d)}$ is given by (1-3) and the evaluation maps are as in (2-6). In the case $\ell(\mathbf{a})\ge 0$, (2-27) immediately gives

$$Z^*_{n;\mathbf{a}}(x_1,x_2,\hbar_1,\hbar_2,q) = \frac{\langle\mathbf{a}\rangle x_1^{\ell(\mathbf{a})}}{\hbar_1+\hbar_2}\sum_{\substack{s_1,s_2\ge 0\\ s_1+s_2=n-1}}\left(-x_1^{s_1}x_2^{s_2}+x_1^{s_1}x_2^{s_2}\dot{F}^{(s_1)}_{n;\mathbf{a}}\left(\frac{x_1}{\hbar_1},\frac{q}{x_1^{\nu_n(\mathbf{a})}}\right)\cdot\ddot{F}^{(s_2)}_{n;\mathbf{a}}\left(\frac{x_2}{\hbar_2},\frac{q}{x_2^{\nu_n(\mathbf{a})}}\right)\right) \tag{2-30}$$

and similarly for the three-point generating function in (2-29). In general, (3-28), (3-30), (3-15), the second identity in (3-12), (3-31), the middle identity in (3-13), and (2-28) give

$$\begin{aligned} Z^*_{n;\mathbf{a}}(x_1,x_2,\hbar_1,\hbar_2,q) = \frac{\langle\mathbf{a}\rangle}{\hbar_1+\hbar_2}\sum_{\substack{s_1,s_2\ge 0\\ s_1+s_2=n-1}}&\left(x_1^{s_1}x_2^{s_2+\ell(\mathbf{a})}\dot{F}^{(s_2)*}_{n;\mathbf{a}}\left(\frac{x_2}{\hbar_2},\frac{q}{x_2^{\nu_n(\mathbf{a})}}\right)\right.\\ &\left.+x_1^{s_1+\ell(\mathbf{a})}x_2^{s_2}\dot{F}^{(s_1)*}_{n;\mathbf{a}}\left(\frac{x_1}{\hbar_1},\frac{q}{x_1^{\nu_n(\mathbf{a})}}\right)\ddot{F}^{(s_2)}_{n;\mathbf{a}}\left(\frac{x_2}{\hbar_2},\frac{q}{x_2^{\nu_n(\mathbf{a})}}\right)\right), \end{aligned} \tag{2-31}$$

where $\dot{F}^{(s)*}_{n;\mathbf{a}}(w,q)\equiv\dot{F}^{(s)}_{n;\mathbf{a}}(w,q)-1$.[4]

An analogue of (2-31) for the three-point generating function in (2-29) can be similarly obtained from (3-29), the last identity in (3-13), and (2-17). In particular, in the Calabi–Yau case, $\nu_n(\mathbf{a})=0$,

$$\begin{aligned} &Z^*_{n;\mathbf{a}}(x_1,x_2,x_3,\hbar_1,\hbar_2,\hbar_3,q)\\ &=\frac{\langle\mathbf{a}\rangle}{\hbar_1\hbar_2\hbar_3}\left\{\sum_{\substack{s_1,s_2,s_3\ge 0\\ s_1,s_2,s_3\le n-1\\ s_1+s_2+s_3=2n-2}}\left(x_1^{s_1}x_2^{s_2}x_3^{s_3+\ell(\mathbf{a})}\dot{F}^{(s_3)*}_{n;\mathbf{a}}\left(\frac{x_3}{\hbar_3},q\right)\right.\right.\\ &\qquad+x_1^{s_1}x_2^{s_2+\ell(\mathbf{a})}x_3^{s_3}\dot{F}^{(s_2)*}_{n;\mathbf{a}}\left(\frac{x_2}{\hbar_2},q\right)\ddot{F}^{(s_3)}_{n;\mathbf{a}}\left(\frac{x_3}{\hbar_3},q\right)\\ &\qquad\left.+x_1^{s_1+\ell(\mathbf{a})}x_2^{s_2}x_3^{s_3}\tilde{\mathrm{c}}_{s_1,s_2,s_3}(q)\dot{F}^{(s_1)*}_{n;\mathbf{a}}\left(\frac{x_1}{\hbar_1},q\right)\prod_{t=2}^{3}\ddot{F}^{(s_t)}_{n;\mathbf{a}}\left(\frac{x_t}{\hbar_t},q\right)\right)\\ &\qquad\left.+\sum_{\substack{s_1\ge\ell^-(\mathbf{a}),\,s_2,s_3\ge 0\\ s_1,s_2,s_3\le n-1\\ s_1+s_2+s_3=2n-2}}x_1^{s_1+\ell(\mathbf{a})}x_2^{s_2}x_3^{s_3}\tilde{\mathrm{c}}^*_{s_1,s_2,s_3}(q)\prod_{t=2}^{3}\ddot{F}^{(s_t)}_{n;\mathbf{a}}\left(\frac{x_t}{\hbar_t},q\right)\right\}, \end{aligned} \tag{2-32}$$

where

$$\tilde{\mathrm{c}}_{s_1,s_2,s_3}(q)\equiv 1+\tilde{\mathrm{c}}^*_{s_1,s_2,s_3}(q)=\frac{1}{(1-\mathbf{a}^{\mathbf{a}}q)\dot{\mathbb{I}}^c_{s_1}(q)\ddot{\mathbb{I}}^c_{s_2}(q)\ddot{\mathbb{I}}^c_{s_3}(q)}.$$

[4]The right-hand side of (2-31) should be first simplified in $\mathbb{Q}(x_1,x_2,\hbar_1,\hbar_2)[\![q]\!]$, eliminating division by x_1 and x_2, and then projected to $H^*(\mathbb{P}^{n-1}\times\mathbb{P}^{n-1})[\hbar_1,\hbar_2][\![q]\!]$.

This presentation of the three-point formula eliminates division by x_1, even if $\ell(\mathbf{a})<0$, since $\dot{F}^{(s)*}_{n;\mathbf{a}}(w,q)$ is divisible by $w^{\ell^-(\mathbf{a})-s}$ for $s\le\ell^-(\mathbf{a})$.

In the Calabi–Yau case, that is, $\nu_n(\mathbf{a})=0$, we find that

(2-33)
$$\langle\mathbf{a}\rangle+q\frac{\mathrm{d}}{\mathrm{d}q}\mathrm{SQ}^{c_1,c_2}_{n;\mathbf{a}}(q)=\langle\mathbf{a}\rangle\dot{I}_{c_1+1}(q),\qquad \mathrm{SQ}^{c_1,c_2,c_3}_{n;\mathbf{a}}(q)=\frac{\langle\mathbf{a}\rangle}{(1-\mathbf{a}^{\mathbf{a}}q)\prod_{t=1}^{t=3}\prod_{c=0}^{c=c_t}\dot{I}_c(q)},$$

whenever $c_1,c_2,c_3\in\mathbb{Z}^{\ge 0}$, $c_1+c_2=n-2-\ell(\mathbf{a})$ in the first equation, and $c_1+c_2+c_3=n-1-\ell(\mathbf{a})$ in the second equation. The $c_1=0$ case of (2-33) agrees with the $W\mathbin{/\!\!/}G=X_{n;\mathbf{a}}$ case of [Ciocan-Fontanine and Kim 2013, Corollary 5.5.4(bc)]. By (2-33),

$$\max(c_1,c_2)\ge n-\ell^+(\mathbf{a})\quad\Rightarrow\quad \mathrm{SQ}^d_{n;\mathbf{a}}(c_1,c_2)(q)=0\ \text{ for all } d\in\mathbb{Z}^+,$$
$$\max(c_1,c_2,c_3)\ge n-\ell^+(\mathbf{a})\quad\Rightarrow\quad \mathrm{SQ}^d_{n;\mathbf{a}}(c_1,c_2,c_3)(q)=0\ \text{ for all } d\in\mathbb{Z}^+,$$

as the case should be for intrinsic invariants of $X_{n;\mathbf{a}}$. On the other hand,

(2-34)
$$\langle\mathbf{a}\rangle+Q\frac{\mathrm{d}}{\mathrm{d}Q}\mathrm{GW}^{c_1,c_2}_{n;\mathbf{a}}(Q)=\langle\mathbf{a}\rangle\frac{\dot{I}_{c_1+1}(q)}{\dot{I}_1(q)},$$
$$\mathrm{GW}^{c_1,c_2,c_3}_{n;\mathbf{a}}(Q)=\frac{\langle\mathbf{a}\rangle\dot{I}_0(q)}{(1-\mathbf{a}^{\mathbf{a}}q)\prod_{t=1}^{t=3}\prod_{c=0}^{c=c_t}\dot{I}_c(q)},$$

with the same assumptions on c_1,c_2,c_3 as in (2-33) and $Q=q\mathrm{e}^{J_{n;\mathbf{a}}(q)}$, as before; see [Popa and Zinger 2014, Equation (1.5)] and [Zinger 2014, Equation (1.7)], respectively.

In the case of products of projective spaces and concavex sheaves (1-13), the analogues of the above mirror formulas relate power series:

(2-35) $$\check{F}_{n_1,\dots,n_p;\mathbf{a}}\in\mathbb{Q}(w)[\![q_1,\dots,q_p]\!],$$

(2-36) $$\check{Z}^{(s_1,\dots,s_p)}_{n_1,\dots,n_p;\mathbf{a}}\in H^*(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})[\hbar^{-1}][\![q_1,\dots,q_p]\!],$$

(2-37) $$\check{Z}^{*}_{n_1,\dots,n_p;\mathbf{a}}\in H^*((\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})^m)[\hbar_1^{-1},\dots,\hbar_m^{-1}][\![q_1,\dots,q_p]\!],$$

with $\check{F}$ and $\check{Z}$ denoting F, $\dot{F}$, $\ddot{F}$, Z, $\dot{Z}$, or $\ddot{Z}$ and $m=2,3$. The coefficients of $q_1^{d_1}\cdots q_p^{d_p}$ in (2-36) and (2-37) are defined by the same pushforwards as in (2-4), (2-5), (2-8), and (2-29), with the degree d of the stable quotients replaced by $(d_1,\dots,d_p)$ and x^s by $x_1^{s_1}\cdots x_p^{s_p}$. The coefficients of $q_1^{d_1}\cdots q_p^{d_p}$ in (2-35) are obtained from the coefficients in (1-6), (2-9), and (2-12) by replacing a_kd and a_kx by $a_{k;1}d_1+\cdots+a_{k;p}d_p$ and $a_{k;1}x_1+\cdots+a_{k;p}x_p$ in the numerator and taking the product of the denominators with $(n,x,d)=(n_i,x_i,d_i)$ for each $i=1,\dots,p$;

$$x_1,\dots,x_p\in H^*(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})$$

now correspond to the pullbacks of the hyperplane classes by the projection maps. If $\ell^-(\mathbf{a})=0$, the analogue of (2-30) with $\langle\mathbf{a}\rangle x^{\ell(\mathbf{a})}$ replaced by the products of $a_{k;1}x_{1;1}+\cdots+a_{k;p}x_{1;p}$ and sums over pairs of p-tuples $(s_{1;1},\dots,s_{1;p})$ and $(s_{2;1},\dots,s_{2;p})$ with $s_{1;i}+s_{2;i}=n_i-1$ provides a closed formula for $Z^*_{n_1,\dots,n_p;\mathbf{a}}$. In general, the relation (2-31) extends to this case by replacing $\langle\mathbf{a}\rangle x_i^{\ell(\mathbf{a})}$ by the products and ratios of the terms $a_{k;1}x_{i;1}+\cdots+a_{k;p}x_{i;p}$.

3. Equivariant mirror formulas

We begin this section by reviewing the equivariant setup used in [Zinger 2009; Popa and Zinger 2014; Cooper and Zinger 2014], closely following [Cooper and Zinger 2014, Section 3]. After defining equivariant versions of the generating functions $\dot{Z}^{(s)}_{n;\mathbf{a}}$, $\ddot{Z}^{(s)}_{n;\mathbf{a}}$, $\dot{Z}^*_{n;\mathbf{a}}$, and $Z^*_{n;\mathbf{a}}$ and of the hypergeometric series $\dot{F}_{n;\mathbf{a}}$ and $\ddot{F}_{n;\mathbf{a}}$, we state an equivariant version of Theorem 2; see Theorem 3 below. This theorem immediately implies Theorem 2. The proof of the two-point mirror formulas in Theorem 3 is outlined in Sections 5 and 6 and completed in Sections 7 and 8. We conclude this section with a specialization of the three-point formula of Theorem 3 in Proposition 3.1, which is proved in Section 9 and is a key step in the proof of the full three-point formula of Theorem 3 in Section 10.

3A. ***Equivariant setup.*** The quotient of the classifying space for the n-torus $\mathbb{T}$ is $B\mathbb{T}\equiv(\mathbb{P}^\infty)^n$. Thus, the group cohomology of $\mathbb{T}$ is

$$H^*_{\mathbb{T}}\equiv H^*(B\mathbb{T})=\mathbb{Q}[\alpha_1,\dots,\alpha_n],$$

where $\alpha_i\equiv\pi_i^*c_1(\gamma^*)$, $\gamma\to\mathbb{P}^\infty$ is the tautological line bundle, and $\pi_i:(\mathbb{P}^\infty)^n\to\mathbb{P}^\infty$ is the projection to the i-th component. The field of fractions of $H^*_{\mathbb{T}}$ will be denoted by

$$\mathbb{Q}_\alpha\equiv\mathbb{Q}(\alpha_1,\dots,\alpha_n).$$

We denote the equivariant $\mathbb{Q}$-cohomology of a topological space M with a $\mathbb{T}$-action by $H^*_{\mathbb{T}}(M)$. If the $\mathbb{T}$-action on M lifts to an action on a complex vector bundle $V\to M$, let $\mathbf{e}(V)\in H^*_{\mathbb{T}}(M)$ denote the equivariant Euler class of V. A continuous $\mathbb{T}$-equivariant map $f:M\to M'$ between two compact oriented manifolds induces a pushforward homomorphism

$$f_*:H^*_{\mathbb{T}}(M)\to H^*_{\mathbb{T}}(M').$$

The standard action of $\mathbb{T}$ on $\mathbb{C}^n$,

$$(\mathrm{e}^{\mathrm{i}\theta_1},\dots,\mathrm{e}^{\mathrm{i}\theta_n})\cdot(z_1,\dots,z_n)\equiv(\mathrm{e}^{\mathrm{i}\theta_1}z_1,\dots,\mathrm{e}^{\mathrm{i}\theta_n}z_n),$$

descends to a $\mathbb{T}$-action on $\mathbb{P}^{n-1}$, which has n fixed points:

$$P_1=[1,0,\dots,0],\quad P_2=[0,1,0,\dots,0],\quad \dots,\quad P_n=[0,\dots,0,1]. \tag{3-1}$$

This standard $\mathbb{T}$-action on $\mathbb{P}^{n-1}$ lifts to a natural $\mathbb{T}$-action on the tautological line bundle $\gamma \to \mathbb{P}^{n-1}$, since $\gamma \subset \mathbb{P}^{n-1} \times \mathbb{C}^n$ is preserved by the diagonal $\mathbb{T}$-action. With

$$\mathbf{x} \equiv \mathbf{e}(\gamma^*) \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})$$

denoting the equivariant hyperplane class, the equivariant cohomology of $\mathbb{P}^{n-1}$ is given by

$$H^*_{\mathbb{T}}(\mathbb{P}^{n-1}) = \mathbb{Q}[\mathbf{x}, \alpha_1, \dots, \alpha_n]\big/(\mathbf{x}-\alpha_1)\cdots(\mathbf{x}-\alpha_n). \tag{3-2}$$

Let $\mathbf{x}_t \in H^*_{\mathbb{T}}((\mathbb{P}^{n-1})^m)$ be the pullback of $\mathbf{x}$ by the t-th projection map.

The standard $\mathbb{T}$-representation on $\mathbb{C}^n$ (as well as any other representation) induces $\mathbb{T}$-actions on $\bar{Q}_{0,m}(\mathbb{P}^{n-1}, d)$, $\mathcal{U}$, $\mathcal{V}^{(d)}_{n;\mathbf{a}}$, $\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}$, and $\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}$; see (1-3) and (2-1) for the notation. Thus, $\mathcal{V}^{(d)}_{n;\mathbf{a}}$, $\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}$, and $\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}$ have well-defined equivariant Euler classes

$$\mathbf{e}\big(\mathcal{V}^{(d)}_{n;\mathbf{a}}\big), \mathbf{e}\big(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}\big), \mathbf{e}\big(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}}\big) \in H^*_{\mathbb{T}}(\bar{Q}_{0,m}(\mathbb{P}^{n-1}, d)).$$

The universal cotangent line bundle for the i-th marked point also has a well-defined equivariant Euler class, which will still be denoted by ψ_i.

Similarly to (2-2) and (2-3), let

$$\begin{aligned}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) &\equiv 1 + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{\hbar-\psi_1}\right] \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1}, q]\!],\\ \ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) &\equiv 1 + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{\hbar-\psi_1}\right] \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1}, q]\!].\end{aligned} \tag{3-3}$$

For each $s \in \mathbb{Z}^{\ge 0}$, let

$$\begin{aligned}\dot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) &\equiv \mathbf{x}^s + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^*\mathbf{x}^s}{\hbar-\psi_1}\right] \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1}, q]\!],\\ \ddot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) &\equiv \mathbf{x}^s + \sum_{d=1}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^*\mathbf{x}^s}{\hbar-\psi_1}\right] \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1}, q]\!].\end{aligned} \tag{3-4}$$

Similarly to (2-4) and (2-5), we define

$$\begin{aligned}&\dot{\mathcal{Z}}^*_{n;\mathbf{a}}(\mathbf{x}_1, \mathbf{x}_2, \hbar_1, \hbar_2, q)\\ &\quad\equiv \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2\}_*\left[\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)}\right] \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar_1^{-1}, \hbar_2^{-1}, q]\!],\end{aligned} \tag{3-5}$$

$$\begin{aligned}&\dot{\mathcal{Z}}^*_{n;\mathbf{a}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \hbar_1, \hbar_2, \hbar_3, q)\\ &\quad\equiv \sum_{d=1}^{\infty} q^d \{\mathrm{ev}_1 \times \mathrm{ev}_2 \times \mathrm{ev}_3\}_*\left[\frac{e(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)(\hbar_3-\psi_3)}\right],\end{aligned} \tag{3-6}$$

with the total pushforwards by the total evaluation maps taken in equivariant cohomology. Similarly to (2-7), let

$$\text{(3-7)}\qquad \dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)=\frac{\mathbf{PD}(\Delta^{(2)}_{\mathbb{P}^{n-1}})}{\hbar_1+\hbar_2}+\dot{\mathcal{Z}}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q),$$

$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\hbar_1,\hbar_2,\hbar_3,q)=\frac{\mathbf{PD}(\Delta^{(3)}_{\mathbb{P}^{n-1}})}{\hbar_1\hbar_2\hbar_3}+\dot{\mathcal{Z}}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\hbar_1,\hbar_2,\hbar_3,q),$$

where $\mathbf{PD}(\Delta^{(2)}_{\mathbb{P}^{n-1}})$ and $\mathbf{PD}(\Delta^{(3)}_{\mathbb{P}^{n-1}})$ are the equivariant Poincaré duals of the (small) diagonals in $\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}$ and $\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}$, respectively.

The above Poincaré duals can be written as

(3-8)

$$\mathbf{PD}(\Delta^{(2)}_{\mathbb{P}^{n-1}})=\sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}}(-1)^r\mathbf{s}_r\mathbf{x}_1^{s_1}\mathbf{x}_2^{s_2},$$

$$\begin{aligned}\mathbf{PD}(\Delta^{(3)}_{\mathbb{P}^{n-1}})&=\sum_{\substack{s_1,s_2,s_3,r\ge 0\\ s_1+s_2+s_3+r=2n-2}}\mathbf{s}^{(2)}_r\mathbf{x}_1^{s_1}\mathbf{x}_2^{s_2}\mathbf{x}_3^{s_3}\\ &=\sum_{\substack{s_1,s_2,s_3,r\ge 0\\ s_1,s_2,s_3\le n-1\\ s_1+s_2+s_3+r=2n-2}}\ \sum_{\substack{r_0,r_1,r_2,r_3\ge 0\\ r_1\le\hat{s}_1,r_2\le\hat{s}_2,r_3\le\hat{s}_3\\ r_0+r_1+r_2+r_3=r}}(-1)^{r_1+r_2+r_3}\eta_{r_0}\mathbf{s}_{r_1}\mathbf{s}_{r_2}\mathbf{s}_{r_3}\mathbf{x}_1^{s_1}\mathbf{x}_2^{s_2}\mathbf{x}_3^{s_3},\end{aligned}$$

where $\mathbf{s}_r,\eta_r,\mathbf{s}^{(2)}_r\in\mathbb{Q}[\alpha_1,\dots,\alpha_n]$ are the r-th elementary symmetric polynomial in $\alpha_1,\dots,\alpha_n$, the sum of all degree-r monomials in $\alpha_1,\dots,\alpha_n$, and the degree-r term in $(1-\mathbf{s}_1+\mathbf{s}_2-\cdots)^2$, respectively. All three expressions for the Poincaré duals can be confirmed by pairing them with $\mathbf{x}_1^{t_1}\mathbf{x}_2^{t_2}$ and $\mathbf{x}_1^{t_1}\mathbf{x}_2^{t_2}\mathbf{x}_3^{t_3}$, with $t_1,t_2,t_3\le n-1$, and using the localization theorem of [Atiyah and Bott 1984] on $(\mathbb{P}^{n-1})^m$ and the residue theorem on S^2 to reduce the equivariant integrals of $\mathbf{x}^{s+t}$ on $\mathbb{P}^{n-1}$ to the polynomials η_r; these are the homogeneous polynomials in the power series expansion of $1/(1-\alpha_1)(1-\alpha_2)\cdots$. The coefficient of $\mathbf{x}_1^{s_1}\mathbf{x}_2^{s_2}\mathbf{x}_3^{s_3}$ in the second expression for $\mathbf{PD}(\Delta^{(3)}_{\mathbb{P}^{n-1}})$ is precisely $\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(0)$, with $\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}$ as in Theorem 3; see the end of Section 3B. This provides a direct check of the degree-0 term in (3-14).

3B. ***Equivariant mirror symmetry.*** The equivariant analogues of the power series in (1-6) and (2-9) are given by

$$\text{(3-9)}\quad \dot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\equiv\sum_{d=0}^{\infty}q^d\frac{\prod_{a_k>0}\prod_{r=1}^{a_kd}(a_k\mathbf{x}+r\hbar)\prod_{a_k<0}\prod_{r=0}^{-a_kd-1}(a_k\mathbf{x}-r\hbar)}{\prod_{r=1}^{d}\Big(\prod_{k=1}^{n}(\mathbf{x}-\alpha_k+r\hbar)-\prod_{k=1}^{n}(\mathbf{x}-\alpha_k)\Big)}\in\mathbb{Q}[\alpha_1,\dots,\alpha_n,\mathbf{x}][\![\hbar^{-1},q]\!],$$

$$\ddot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q) \equiv \sum_{d=0}^{\infty} q^d \frac{\prod_{a_k>0}\prod_{r=0}^{a_kd-1}(a_k\mathbf{x}+r\hbar)\prod_{a_k<0}\prod_{r=1}^{-a_kd}(a_k\mathbf{x}-r\hbar)}{\prod_{r=1}^{d}\Big(\prod_{k=1}^{n}(\mathbf{x}-\alpha_k+r\hbar)-\prod_{k=1}^{n}(\mathbf{x}-\alpha_k)\Big)} \in \mathbb{Q}[\alpha_1,\dots,\alpha_n,\mathbf{x}][\![\hbar^{-1},q]\!].$$

The second products in the denominators above are irrelevant for the statements in this section, but are material to (4-9) and thus to the proof of (3-14) in this paper.

For each $s\in\mathbb{Z}^+$, we define $\mathfrak{D}^s\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q), \mathfrak{D}^s\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1},q]\!]$ inductively by

$$\begin{aligned}
&\mathfrak{D}^0\,\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q),\\
&\mathfrak{D}^s\,\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{1}{\dot{I}_s(q)}\Big\{\mathbf{x}+\hbar q\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\,\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q),\\
&\mathfrak{D}^0\,\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q),\\
&\mathfrak{D}^s\,\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{1}{\ddot{I}_s(q)}\Big\{\mathbf{x}+\hbar q\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\,\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q).
\end{aligned}\tag{3-10}$$

The next theorem is the equivariant analogue of Theorem 2. It expresses the equivariant stable quotient invariants in (3-5) and (3-6) in terms of the basic equivariant stable quotient invariants in (3-3).

Theorem 3. *If $l\in\mathbb{Z}^{\ge 0}$, $n\in\mathbb{Z}^+$, and $\mathbf{a}\in(\mathbb{Z}^*)^l$, then*

$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q) = \frac{1}{\hbar_1+\hbar_2}\sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}}(-1)^r\mathbf{s}_r\,\dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\mathbf{x}_1,\hbar_1,q)\,\ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar_2,q),\tag{3-11}$$

where $\mathbf{s}_r\in\mathbb{Q}_\alpha$ is the r-th elementary symmetric polynomial in $\alpha_1,\dots,\alpha_n$. If in addition $\nu_n(\mathbf{a})\ge 0$,

$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{\dot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)}{\dot{I}_0(q)},\qquad \ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{\ddot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)}{\ddot{I}_0(q)},\tag{3-12}$$

and there exist $\widetilde{\mathcal{C}}^{(r)}_{s,s'}, \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}\in\mathbb{Q}[\alpha_1,\dots,\alpha_n][\![q]\!]$ with $s,s',s_1,s_2,s_3,r\in\mathbb{Z}^{\ge 0}$ such that

$$\widetilde{\mathcal{C}}^{(r)}_{s,s'}(0)=\delta_{0,r}\delta_{s,s'},\tag{3-13}$$

$$[\![\widetilde{\mathcal{C}}^{(\nu_n(\mathbf{a})d)}_{s,s'}(q)\big|_{\alpha=0}]\!]_d=\widetilde{\mathrm{c}}^{(d)}_{s,s'},\qquad [\![\widetilde{\mathcal{C}}^{(\nu_n(\mathbf{a})d)}_{s_1,s_2,s_3}(q)\big|_{\alpha=0}]\!]_d=\widetilde{\mathrm{c}}^{(d)}_{s_1,s_2,s_3},$$

the coefficients of q^d *in* $\widetilde{\mathcal{C}}^{(r)}_{s,s'}(q)$ *and* $\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(q)$ *are homogeneous symmetric polynomials in* $\alpha_1, \alpha_2, \dots, \alpha_n$ *of degree* $r - \nu_n(\mathbf{a})d$, *and*

$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \hbar_1, \hbar_2, \hbar_3, q) = \frac{1}{\hbar_1\hbar_2\hbar_3} \sum_{\substack{r,s_1,s_2,s_3\ge 0\\ s_1,s_2,s_3\le n-1\\ s_1+s_2+s_3+r=2n-2}} \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(q)\dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\mathbf{x}_1, \hbar_1, q) \prod_{t=2}^{3} \ddot{\mathcal{Z}}^{(s_t)}_{n;\mathbf{a}}(\mathbf{x}_t, \hbar_t, q), \tag{3-14}$$

$$\check{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) = \sum_{r=0}^{s}\sum_{s'=0}^{s-r} \widetilde{\mathcal{C}}^{(r)}_{s-\ell^*(\mathbf{a}),s'-\ell^*(\mathbf{a})}(q)\, \hbar^{s-r-s'} \mathfrak{D}^{s'} \check{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q), \tag{3-15}$$

where $(\check{\mathcal{Z}}, \ell^*) = (\dot{\mathcal{Z}}, \ell^-), (\ddot{\mathcal{Z}}, \ell^+)$.

Setting $\alpha = 0$ in (3-11), (3-12), (3-14), and (3-15), we obtain (2-19), (2-26), (2-20) and (2-21), respectively.

We now completely describe the power series $\widetilde{\mathcal{C}}^{(r)}_{s,s'}$ of Theorem 3; it will be shown in Section 5 that they indeed satisfy (3-15). Let

$$\mathfrak{D}^0\mathcal{Y}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) = \frac{1}{I_0(q)} \sum_{d=0}^{\infty} q^d \frac{\prod_{a_k>0}\prod_{r=1}^{a_k d}(a_k\mathbf{x} + r\hbar)\prod_{a_k<0}\prod_{r=1}^{-a_k d}(a_k\mathbf{x} - r\hbar)}{\prod_{r=1}^{d}\prod_{k=1}^{n}(\mathbf{x} - \alpha_k + r\hbar)}. \tag{3-16}$$

For $s \in \mathbb{Z}^+$, let

$$\mathfrak{D}^s\mathcal{Y}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) = \frac{1}{I_s(q)}\left\{\mathbf{x} + \hbar\, q\frac{\mathrm{d}}{\mathrm{d}q}\right\}\mathfrak{D}^{s-1}\mathcal{Y}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \in \mathbf{x}^s + q\cdot\mathbb{Q}[\alpha_1, \dots, \alpha_n, \mathbf{x}][\hbar][\![\hbar^{-1}, q]\!]. \tag{3-17}$$

Comparing with (2-12), we find that

$$\mathfrak{D}^s\mathcal{Y}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)\big|_{\alpha=0} = \mathbf{x}^s\mathfrak{D}^s F_{n;\mathbf{a}}(\mathbf{x}/\hbar, q/\mathbf{x}^{\nu_n(\mathbf{a})}), \quad \text{where} \tag{3-18}$$

$$\mathfrak{D}^0 F_{n;\mathbf{a}}(w, q) = \frac{F_{n;\mathbf{a}}(w, q)}{I_0(q)}, \quad \mathfrak{D}^s F_{n;\mathbf{a}}(w, q) = \frac{1}{I_s(q)}\left\{1 + \frac{q}{w}\frac{\mathrm{d}}{\mathrm{d}q}\right\}\mathfrak{D}^{s-1}F_{n;\mathbf{a}}(w, q)$$

for all $s \in \mathbb{Z}^+$. For $r, s, s' \ge 0$, define $\mathcal{C}^{(r)}_{s,s'} \in \mathbb{Q}[\alpha_1, \dots, \alpha_n][\![q]\!]$ by

$$\hbar^s \sum_{s'=0}^{\infty}\sum_{r=0}^{s'} \mathcal{C}^{(r)}_{s,s'}(q)\mathbf{x}^{s'-r}\hbar^{-s'} = \mathfrak{D}^s\mathcal{Y}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q). \tag{3-19}$$

By (3-16), (3-17), and (3-19), the coefficient of q^d in $\mathcal{C}^{(r)}_{s,s'}$ is a degree-$r - \nu_n(\mathbf{a})d$ homogeneous symmetric polynomial in α. By (3-17) and (3-18),

$$\mathcal{C}^{(0)}_{s,s}(q) = 1, \quad \mathcal{C}^{(0)}_{s,s'}(q) = 0 \ \text{ for } s > s', \quad \mathcal{C}^{(r)}_{s,s'}(0) = \delta_{r,0}\delta_{s,s'}. \tag{3-20}$$

By the first two statements above, the relations

$$\sum_{\substack{r_1,r_2\ge 0\\ r_1+r_2=r}}\sum_{t=0}^{s-r_1}\widetilde{\mathcal{C}}^{(r_1)}_{s,t}(q)\mathcal{C}^{(r_2)}_{t,s'-r_1}(q)=\delta_{r,0}\delta_{s,s'}\quad\text{for } r,s'\in\mathbb{Z}^{\ge 0},\ r\le s'\le s, \tag{3-21}$$

inductively define $\widetilde{\mathcal{C}}^{(r)}_{s,s'-r}\in\mathbb{Q}[\alpha_1,\dots,\alpha_n][\![q]\!]$ with $r\le s'\le s$ in terms of the power series $\widetilde{\mathcal{C}}^{(r_1)}_{s,t}$ with $r_1<r$ or $r_1=r$ and $t<s'-r$. By (3-20) and (3-21),

$$\widetilde{\mathcal{C}}^{(0)}_{s,s'}=\delta_{s,s'},\qquad \widetilde{\mathcal{C}}^{(r)}_{s,s'}(0)=\delta_{r,0}\delta_{s,s'},$$

and the coefficient of q^d in $\widetilde{\mathcal{C}}^{(r)}_{s,s'}$ is a degree-$r-\nu_n(\mathbf{a})d$ homogeneous symmetric polynomial in α. If $s'<0$, we set $\widetilde{\mathcal{C}}^{(r)}_{s,s'}=\delta_{r,0}\delta_{s,s'}$. If $\nu_n(\mathbf{a})>0$,

$$\mathcal{C}^{(\nu_n(\mathbf{a})d)}_{s,s'}\big|_{\alpha=0}=\mathrm{c}^{(d)}_{s,s'-\nu_n(\mathbf{a})d}q^d\quad\text{for all } s'\ge\nu_n(\mathbf{a})d$$

by (3-19), (3-18), and (2-13). Thus, setting $\alpha=0$ in (3-21) and comparing with (2-14) with s' replaced by $s'-\nu_n(\mathbf{a})d$, we obtain the second identity in (3-13).

We next completely describe the power series $\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}$ of Theorem 3; it will be shown in Section 10 that they indeed satisfy (3-14). For each $r\in\mathbb{Z}^{\ge 0}$, let $p_r,\mathcal{H}^{(r)}\in\mathbb{Q}[z_1,z_2,\dots]$ be such that

$$p_r(\alpha_1,\alpha_2,\dots)=\alpha_1^r+\alpha_2^r+\cdots=\mathcal{H}^{(r)}(\mathbf{s}_1,\mathbf{s}_2,\dots). \tag{3-22}$$

For $r,\nu\in\mathbb{Z}^{\ge 0}$, we define $\mathcal{H}^{(r)}_\nu\in\mathbb{Q}[\mathbf{s}_1,\mathbf{s}_2,\dots][\![z]\!]$ by

$$\mathcal{H}^{(r)}_\nu(z)=\begin{cases}(1-z)^{-1} & \text{if } \nu=0,\ r=0;\\ \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{H}^{(r)}\big((1-z)^{-1}\mathbf{s}_1,(1-z)^{-1}\mathbf{s}_2,\dots\big) & \text{if } \nu=0,\ r\ge 1;\\ \frac{1}{r+\nu}\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{H}^{(r+\nu)}(\mathbf{s}_1,\dots,\mathbf{s}_{\nu-1},\mathbf{s}_\nu-(-1)^\nu z,\mathbf{s}_{\nu+1},\dots) & \text{if } \nu>0.\end{cases} \tag{3-23}$$

In particular, the coefficient of z^d in $\mathcal{H}^{(r)}_\nu(z)$ is a degree-$r-\nu d$ homogeneous symmetric polynomial in α,

$$\mathcal{H}^{(r)}_\nu(0)=\eta_r,\qquad [\![\mathcal{H}^{(\nu d)}_\nu(z)\big|_{\alpha=0}]\!]_d=1. \tag{3-24}$$

The second identity above follows from [Zinger 2014, Lemma B.3]. Using induction via Newton's identity [Artin 1991, page 577], the first identity in (3-24) can be reduced to

$$\sum_{t=0}^{r}(-1)^t\eta_{r-t}\mathbf{s}_t=0,\qquad \sum_{t=0}^{r}(-1)^t(r-t)\eta_{r-t}\mathbf{s}_t=p_r\quad\text{for all } r\in\mathbb{Z}^+;$$

these two identities are equivalent to

$$\frac{(1-\alpha_1 u)(1-\alpha_2 u)\cdots}{(1-\alpha_1 u)(1-\alpha_2 u)\cdots}=1,$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\frac{(1-\alpha_1 u)(1-\alpha_2 u)\cdots}{(1-\alpha_1 uz)(1-\alpha_2 uz)\cdots}\bigg|_{z=0}=\frac{\alpha_1 u}{1-\alpha_1 u}+\frac{\alpha_2 u}{1-\alpha_2 u}+\cdots$$

Let

$$\begin{aligned}\text{(3-25)}\quad &\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(q)\\ &=\sum_{\substack{r_0,r_1,r_2,r_3\ge 0\\ r_1\le\hat{s}_1,r_2\le\hat{s}_2,r_3\le\hat{s}_3\\ r_0+r_1+r_2+r_3=r}}\frac{\mathcal{H}^{(r_0)}_{\nu_n(\mathbf{a})}(\mathbf{a}^{\mathbf{a}}q)}{\ddot{\mathbb{I}}^c_{s_1+r_1}(q)\ddot{\mathbb{I}}^c_{s_2+r_2}(q)\ddot{\mathbb{I}}^c_{s_3+r_3}(q)}\ddot{\mathcal{C}}^{(r_1)}_{\hat{s}_1}(q)\dot{\mathcal{C}}^{(r_2)}_{\hat{s}_2}(q)\dot{\mathcal{C}}^{(r_3)}_{\hat{s}_3}(q),\end{aligned}$$

where

$$\text{(3-26)}\qquad \check{\mathcal{C}}^{(r)}_s(q)=\sum_{\substack{r',r''\ge 0\\ r'+r''=r}}(-1)^{r'}\mathbf{s}_{r'}\check{\mathcal{C}}^{(r'')}_{s-r'-\ell^*(\mathbf{a}),s-r-\ell^*(\mathbf{a})}(q)$$

with $(\check{\mathcal{C}},\ell^*)=(\dot{\mathcal{C}},\ell^-)$, $(\ddot{\mathcal{C}},\ell^+)$. Since the coefficients of q^d in $\mathcal{H}^{(r)}_\nu$ and in $\widetilde{\mathcal{C}}^{(r)}_{s,s'}$ are degree-$r-\nu_n(\mathbf{a})d$ homogeneous symmetric polynomials in α, the coefficient of q^d in $\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}$ is also a degree-$r-\nu_n(\mathbf{a})d$ homogeneous symmetric polynomial in α. The last identity in (3-13) follows from (3-25), the second identity in (3-24), the middle identity in (3-13), and (2-15).

3C. ***Related mirror formulas.*** Similarly to (2-29), we define

$$\mathcal{Z}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)\equiv\sum_{d=1}^{\infty}q^d\{\mathrm{ev}_1\times\mathrm{ev}_2\}_*\left[\frac{\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)}\right],$$

$$\begin{aligned}\text{(3-27)}\quad &\mathcal{Z}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\hbar_1,\hbar_2,\hbar_3,q)\\ &\equiv\sum_{d=1}^{\infty}q^d\{\mathrm{ev}_1\times\mathrm{ev}_2\times\mathrm{ev}_3\}_*\left[\frac{e(\mathcal{V}^{(d)}_{n;\mathbf{a}})}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)(\hbar_3-\psi_3)}\right],\end{aligned}$$

with the evaluation maps as in (2-6). For each $s\in\mathbb{Z}^{\ge 0}$, let

$$\mathcal{Z}^{(s)*}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\equiv\sum_{d=1}^{\infty}q^d\mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^*\mathbf{x}^s}{\hbar-\psi_1}\right]\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1},q]\!],$$

$$\ddot{\mathcal{Z}}^{(s)*}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\equiv\sum_{d=1}^{\infty}q^d\mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_2^*\mathbf{x}^s}{\hbar-\psi_1}\right]\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1},q]\!].$$

Since $\mathbf{x}_1, \mathbf{x}_2 \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}) \otimes_{\mathbb{Q}[\alpha_1,\dots,\alpha_n]} \mathbb{Q}_\alpha$ are invertible, the first equation in (3-8) gives

$$\begin{aligned}
\langle \mathbf{a}\rangle \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \mathbf{x}_1^{s_1} \ddot{\mathcal{Z}}^{(s_2)*}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar,q)
&= \sum_{d=1}^{\infty} q^d \{\mathrm{id}\times \mathrm{ev}_1\}_* \left[\frac{\pi_2^*\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}})\,\{\mathrm{id}\times\mathrm{ev}_2\}^*(\mathbf{PD}(\Delta_{\mathbb{P}^{n-1}})\mathbf{x}_2^{-\ell(\mathbf{a})})}{\hbar-\psi_1}\right]\\
&= \sum_{d=1}^{\infty} q^d \{\mathrm{id}\times \mathrm{ev}_1\}_* \left[\frac{\pi_2^*\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a}})\,\{\mathrm{id}\times\mathrm{ev}_2\}^*(\mathbf{PD}(\Delta_{\mathbb{P}^{n-1}})\mathbf{x}_1^{-\ell(\mathbf{a})})}{\hbar-\psi_1}\right]\\
&= \mathbf{x}_1^{-\ell(\mathbf{a})} \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \mathbf{x}_1^{s_1} \mathcal{Z}^{(s_2)*}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar,q),
\end{aligned}$$

where $\pi_2 : \mathbb{P}^{n-1}\times \overline{Q}_{0,2}(\mathbb{P}^{n-1},d) \to \overline{Q}_{0,2}(\mathbb{P}^{n-1},d)$ is the projection map. Combining the last identity with (3-11), we obtain

$$\begin{aligned}
\text{(3-28)}\quad \mathcal{Z}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)
&= \frac{1}{\hbar_1+\hbar_2} \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \Big(\mathbf{x}_1^{s_1} \mathcal{Z}^{(s_2)*}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar_2,q)\\
&\qquad\qquad + \mathcal{Z}^{(s_1)*}_{n;\mathbf{a}}(\mathbf{x}_1,\hbar_1,q)\ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar_2,q)\Big).
\end{aligned}$$

Similar reasoning gives

$$\begin{aligned}
\text{(3-29)}\quad &\mathcal{Z}^*_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\hbar_1,\hbar_2,\hbar_3,q)\\
&= \frac{1}{\hbar_1\hbar_2\hbar_3} \sum_{\substack{r,s_1,s_2,s_3\ge 0\\ s_1,s_2,s_3\le n-1\\ s_1+s_2+s_3+r=2n-2}} \Big(\widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(0)\mathbf{x}_1^{s_1}\mathbf{x}_2^{s_2}\mathcal{Z}^{(s_3)*}_{n;\mathbf{a}}(\mathbf{x}_3,\hbar_3,q)\\
&\qquad + \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(0)\mathbf{x}_1^{s_1}\mathcal{Z}^{(s_2)*}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar_2,q)\ddot{\mathcal{Z}}^{(s_3)}_{n;\mathbf{a}}(\mathbf{x}_3,\hbar_3,q)\\
&\qquad + \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(0)\mathcal{Z}^{(s_1)*}_{n;\mathbf{a}}(\mathbf{x}_1,\hbar_1,q)\prod_{t=2}^{3}\ddot{\mathcal{Z}}^{(s_t)}_{n;\mathbf{a}}(\mathbf{x}_t,\hbar_t,q)\\
&\qquad + \langle\mathbf{a}\rangle\mathbf{x}_1^{\ell^-(\mathbf{a})}\widetilde{\mathcal{C}}^{(r)*}_{s_1,s_2,s_3}(q)\dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\mathbf{x}_1,\hbar_1,q)\prod_{t=2}^{3}\ddot{\mathcal{Z}}^{(s_t)}_{n;\mathbf{a}}(\mathbf{x}_t,\hbar_t,q)\Big),
\end{aligned}$$

where $\widetilde{\mathcal{C}}^{(r)*}_{s_1,s_2,s_3}(q) = \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(q) - \widetilde{\mathcal{C}}^{(r)}_{s_1,s_2,s_3}(0)$.

On the other hand, by (3-15) and the first identity in (3-12),

$$\begin{aligned}
\text{(3-30)}\quad &\mathcal{Z}^{(s)*}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\\
&= -\langle\mathbf{a}\rangle\mathbf{x}^{\ell(\mathbf{a})+s} + \langle\mathbf{a}\rangle\mathbf{x}^{\ell(\mathbf{a})}\sum_{r=0}^{s}\sum_{s'=0}^{s-r}\widetilde{\mathcal{C}}^{(r)}_{s-\ell^-(\mathbf{a}),s'-\ell^-(\mathbf{a})}(q)\hbar^{s-r-s'}\mathfrak{D}^{s'}\dot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q),
\end{aligned}$$

where

$$\mathfrak{D}^0\check{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{\check{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)}{\check{I}_0(q)},$$
$$\mathfrak{D}^s\check{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)=\frac{1}{\check{I}_s(q)}\Big\{\mathbf{x}+\hbar\,q\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\check{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$$

for all $s\in\mathbb{Z}^+$ and $(\check{\mathcal{Y}},\check{I})=(\dot{\mathcal{Y}},\dot{I}),(\ddot{\mathcal{Y}},\ddot{I})$. By (3-9), (1-6), and (2-9),

$$\mathfrak{D}^s\check{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\big|_{\alpha=0}=\mathbf{x}^s\mathfrak{D}^s\check{F}_{n;\mathbf{a}}(\mathbf{x}/\hbar,q/\mathbf{x}^{\nu_n(\mathbf{a})}),\quad\text{where}\tag{3-31}$$

$$\mathfrak{D}^0\check{F}_{n;\mathbf{a}}(w,q)=\frac{\check{F}_{n;\mathbf{a}}(w,q)}{\check{I}_0(q)},\quad \mathfrak{D}^s\check{F}_{n;\mathbf{a}}(w,q)=\frac{1}{\check{I}_s(q)}\Big\{1+\frac{q}{w}\frac{\mathrm{d}}{\mathrm{d}q}\Big\}\mathfrak{D}^{s-1}\check{F}_{n;\mathbf{a}}(w,q)$$

for all $s\in\mathbb{Z}^+$, with $(\check{\mathcal{Y}},\check{F},\check{I})=(\dot{\mathcal{Y}},\dot{F},\dot{I}),(\ddot{\mathcal{Y}},\ddot{F},\ddot{I})$. Simplifying the right-hand side of (3-30) in $\mathbb{Q}_\alpha(\mathbf{x},\hbar)[\![\hbar^{-1},q]\!]$ to eliminate division by $\mathbf{x}$ and setting $\alpha=0$, we obtain (2-31).

3D. *Other three-point generating functions.* The main step in the proof of the mirror formula (3-14) for the stable quotients analogue of the triple Givental's J-function involves determining a mirror formula for the generating function

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x},\hbar,q)\equiv 1+\sum_{d=1}^{\infty}q^d\mathrm{ev}_{1*}\Bigg[\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{\hbar-\psi_1}\Bigg]\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1},q]\!],\tag{3-32}$$

where $\mathrm{ev}_1:\overline{Q}_{0,3}(\mathbb{P}^{n-1},d)\to\mathbb{P}^{n-1}$ is the evaluation map at the first marked point; the meaning of the superscript $(\mathbf{0},\mathbf{1})$ is explained in (6-7). By (3-33), the SQ invariants do *not* satisfy the string relation [Hori et al. 2003, Section 26.3] in the pure Calabi–Yau cases, $\nu_n(\mathbf{a})=0$ and $\ell^-(\mathbf{a})=0$ (when $\dot{I}_0(q)\neq 1$), even though the relevant forgetful morphism, $f_{2,3}$ below, is defined. Since in these cases the twisted invariants of $\mathbb{P}^{n-1}$ are intrinsic invariants of the corresponding complete intersection $X_{n;\mathbf{a}}$, this implies that the construction of virtual fundamental class in [Ciocan-Fontanine et al. 2014] does not respect the forgetful morphism

$$f_{2,3}:\overline{Q}_{0,3}(X_{n;\mathbf{a}},d)\to\overline{Q}_{0,2}(X_{n;\mathbf{a}},d),$$

at least in the Calabi–Yau cases.

Proposition 3.1. *If $l\in\mathbb{Z}^{\geq 0}$, $n\in\mathbb{Z}^+$, and $\mathbf{a}\in(\mathbb{Z}^*)^l$ are such that $\nu_n(\mathbf{a})\geq 0$, then*

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x},\hbar,q)=\hbar^{-1}\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)}{\dot{I}_0(q)}.\tag{3-33}$$

In principle, this proposition is contained in [Ciocan-Fontanine and Kim 2013, Corollary 1.4.1]. We give a direct proof, along the lines of [Cooper and Zinger 2014]. In the process of proving this proposition, we establish the mirror formula for equivariant Hurwitz numbers in Proposition 4.1. This in turn allows us to derive (3-14) from (3-11) and (3-15) following the approach of [Zinger 2014]; see Section 10.

Similarly to (3-32), let

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\mathbf{x},\hbar,q)\equiv 1+\sum_{d=1}^{\infty}q^d\,\mathrm{ev}_{1*}\left[\frac{f_{2,3}^*\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})}{\hbar-\psi_1}\right]\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1},q]\!], \tag{3-34}$$

where $\mathrm{ev}_1:\overline{Q}_{0,3}(\mathbb{P}^{n-1},d)\to\mathbb{P}^{n-1}$ is the evaluation map at the first marked point and

$$f_{2,3}:\overline{Q}_{0,3}(\mathbb{P}^{n-1},d)\to\overline{Q}_{0,2}(\mathbb{P}^{n-1},d)$$

is the forgetful morphism. By the proof of the string relation [Hori et al. 2003, Section 26.3],

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\mathbf{x},\hbar,q)=\hbar^{-1}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q). \tag{3-35}$$

We use this identity to establish the mirror formula for Hurwitz numbers in Proposition 4.2.

As stated in Section 1, Theorem 3 generalizes to products of projective spaces and concavex sheaves (1-13). The relevant torus action is then the product of the actions on the components described above. If its weights are denoted by $\alpha_{i;j}$, with $i=1,\dots,p$ and $j=1,\dots,n_i$, the analogues of the above mirror formulas relate power series

$$\check{\mathcal{Y}}_{n_1,\dots,n_p;\mathbf{a}}\in\mathbb{Q}[\alpha_{1;1},\dots,\alpha_{p;n_p},\mathbf{x}_1,\dots,\mathbf{x}_p][\![\hbar^{-1},q_1,\dots,q_p]\!], \tag{3-36}$$

$$\check{\mathcal{Z}}^{(s_1,\dots,s_p)}_{n_1,\dots,n_p;\mathbf{a}}\in H^*_{\mathbb{T}}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})[\![\hbar^{-1},q_1,\dots,q_p]\!], \tag{3-37}$$

$$\check{\mathcal{Z}}^*_{n_1,\dots,n_p;\mathbf{a}}\in H^*_{\mathbb{T}}((\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})^m)[\![\hbar_1^{-1},\dots,\hbar_M^{-1},q_1,\dots,q_p]\!], \tag{3-38}$$

with $\check{\mathcal{Y}}$ and $\check{\mathcal{Z}}$ denoting $\mathcal{Y}$, $\dot{\mathcal{Y}}$, $\ddot{\mathcal{Y}}$, $\mathcal{Z}$, $\dot{\mathcal{Z}}$, or $\ddot{\mathcal{Z}}$ and $m=2,3$. The coefficients of $q_1^{d_1}\dots q_p^{d_p}$ in (3-37) and (3-38) are defined by the same pushforwards as in (3-4), (3-5), (3-6), and (3-27) with the degree d of the stable quotients replaced by $(d_1,\dots,d_p)$ and $\mathbf{x}^s$ by $\mathbf{x}_1^{s_1}\cdots\mathbf{x}_p^{s_p}$. The coefficients of $q_1^{d_1}\cdots q_p^{d_p}$ in (3-36) are obtained from the coefficients in (3-9) and (3-16) by replacing a_kd and $a_k\mathbf{x}$ by $a_{k;1}d_1+\cdots+a_{k;p}d_p$ and $a_{k;1}\mathbf{x}_1+\cdots+a_{k;p}\mathbf{x}_p$ in the numerator and taking the product of the denominators with $(n,\mathbf{x},d)=(n_i,\mathbf{x}_i,d_i)$ for each $s=1,\dots,p$; in the i-th factor, α_k is also replaced by $\alpha_{i;k}$;

$$\mathbf{x}_1,\dots,\mathbf{x}_p\in H^*_{\mathbb{T}}(\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1})$$

now correspond to the pullbacks of the equivariant hyperplane classes by the projection maps. The statements of Theorem 3, (3-28), and (3-29) extend by replacing the symmetric polynomials by products of symmetric polynomials in the

p different sets of variables and $\langle \mathbf{a} \rangle \mathbf{x}^{\ell(\mathbf{a})}$ by the products and ratios of the terms $a_{k;1}\mathbf{x}_1 + \cdots + a_{k;p}\mathbf{x}_p$; our proofs extend directly to this situation.

4. Equivariant twisted Hurwitz numbers

The fixed loci of the $\mathbb{T}$-action on $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$ involve moduli spaces of weighted curves and certain vector bundles, which we describe in this section. As a corollary of the proof of Theorem 3, we obtain closed formulas for Euler classes of these vector bundles in some cases. These formulas, described in Propositions 4.1 and 4.2 below, are a key ingredient in computing the genus-1 stable quotients invariants.

A stable d-tuple of flecks on a quasistable m-marked curve is a tuple

$$(\mathcal{C}, y_1, \ldots, y_m; \hat{y}_1, \ldots, \hat{y}_d), \tag{4-1}$$

where $\mathcal{C}$ is a connected (at worst) nodal curve, $y_1, \ldots, y_m \in \mathcal{C}^*$ are distinct smooth points, and $\hat{y}_1, \ldots, \hat{y}_d \in \mathcal{C}^* - \{y_1, \ldots, y_m\}$, such that the $\mathbb{Q}$-line bundle

$$\omega_{\mathcal{C}}(y_1 + \cdots + y_m + \epsilon(\hat{y}_1 + \cdots + \hat{y}_d)) \to \mathcal{C}$$

is ample for all $\epsilon \in \mathbb{Q}^+$; this again implies that $2g + m \geq 2$. An isomorphism

$$\phi : (\mathcal{C}, y_1, \ldots, y_m; \hat{y}_1, \ldots, \hat{y}_d) \to (\mathcal{C}', y_1', \ldots, y_m'; \hat{y}_1', \ldots, \hat{y}_d')$$

between curves with m marked points and d flecks is an isomorphism $\phi : \mathcal{C} \to \mathcal{C}'$ such that

$$\phi(y_i) = y_i' \text{ for all } i = 1, \ldots, m, \quad \phi(\hat{y}_j) = \hat{y}_j' \text{ for all } j = 1, \ldots, d.$$

The automorphism group of any stable curve with m marked points and d flecks is finite. For $g, m, d \in \mathbb{Z}^{\geq 0}$, the moduli space $\overline{\mathcal{M}}_{g,m|d}$ parameterizing the stable d-tuples of flecks as in (4-1) with $h^1(\mathcal{C}, \mathcal{O}_{\mathcal{C}}) = g$ is a nonsingular irreducible proper Deligne–Mumford stack; see [Cooper and Zinger 2014, Proposition 2.3]. If $m \geq m' \geq 2$, let

$$f_{m',m} : \overline{\mathcal{M}}_{0,m|d} \to \overline{\mathcal{M}}_{0,m'|d+m-m'},$$
$$(\mathcal{C}, y_1, \ldots, y_m; \hat{y}_1, \ldots, \hat{y}_d) \mapsto (\mathcal{C}', y_1, \ldots, y_{m'}; \hat{y}_1, \ldots, \hat{y}_d, y_{m'+1}, \ldots, y_m),$$

be the morphism converting the last $m - m'$ marked points into the last $m - m'$ flecks and contracting components of $\mathcal{C}$ if necessary.

Any tuple as in (4-1) induces a quasistable quotient

$$\mathcal{O}_{\mathcal{C}}(-\hat{y}_1 - \cdots - \hat{y}_d) \subset \mathcal{O}_{\mathcal{C}} \equiv \mathbb{C}^1 \otimes \mathcal{O}_{\mathcal{C}}.$$

For any ordered partition $d = d_1 + \cdots + d_p$ with $d_1, \ldots, d_p \in \mathbb{Z}^{\geq 0}$, this correspondence gives rise to a morphism

$$\overline{\mathcal{M}}_{g,m|d} \to \overline{Q}_{g,m}(\mathbb{P}^0 \times \cdots \times \mathbb{P}^0, (d_1, \ldots, d_p)).$$

In turn, this morphism induces an isomorphism

(4-2) $$\phi : \overline{\mathcal{M}}_{g,m|d} \big/ \mathbb{S}_{d_1} \times \cdots \times \mathbb{S}_{d_p} \xrightarrow{\sim} \overline{Q}_{g,m}(\mathbb{P}^0 \times \cdots \times \mathbb{P}^0, (d_1, \ldots, d_p)),$$

with the symmetric group $\mathbb{S}_{d_1}$ acting on $\overline{\mathcal{M}}_{g,m|d}$ by permuting the points $\hat{y}_1, \ldots, \hat{y}_{d_1}$, $\mathbb{S}_{d_2}$ acting on $\overline{\mathcal{M}}_{g,m|d}$ by permuting the points $\hat{y}_{d_1+1}, \ldots, \hat{y}_{d_1+d_2}$, etc.

There is again a universal curve

$$\pi : \mathcal{U} \to \overline{\mathcal{M}}_{g,m|d}$$

with sections $\sigma_1, \ldots, \sigma_m$ and $\hat{\sigma}_1, \ldots \hat{\sigma}_d$. Let

$$\psi_i = -\pi_*(\sigma_i^2),\ \hat{\psi}_i = -\pi_*(\hat{\sigma}_i^2) \in H^2(\overline{\mathcal{M}}_{g,m|d})$$

be the first chern classes of the universal cotangent line bundles. For $m \geq 2$, $d', d \in \mathbb{Z}^+$ with $d' \leq d$, and $\mathbf{r} \equiv (r_1, \ldots, r_{d'}) \in (\mathbb{Z}^{\geq 0})^{d'}$, let

$$\mathcal{S}_{\mathbf{r}} = \mathcal{O}(-\hat{\sigma}_1 - \cdots - \hat{\sigma}_{d-d'} - r_1\hat{\sigma}_{d-d'+1} - \cdots - r_{d'}\hat{\sigma}_d) \to \mathcal{U} \to \overline{\mathcal{M}}_{0,m|d}.$$

If $\beta \in H^2_{\mathbb{T}}$, denote by

(4-3) $$\mathcal{S}^*_{\mathbf{r}}(\beta) \to \mathcal{U} \to \overline{\mathcal{M}}_{0,m|d}$$

the sheaf $\mathcal{S}^*_{\mathbf{r}}$ with the $\mathbb{T}$-action so that

$$\mathbf{e}(\mathcal{S}^*_{\mathbf{r}}(\beta)) = \beta \times 1 + 1 \times e(\mathcal{S}^*_{\mathbf{r}}) \in H^*_{\mathbb{T}}(\mathcal{U}) = H^*_{\mathbb{T}} \otimes H^*(\mathcal{U}).$$

Similarly to (2-1), let

(4-4)
$$\dot{\mathcal{V}}'^{(d)}_{\mathbf{a};\mathbf{r}}(\beta) = \bigoplus_{a_k>0} R^0\pi_*\big(\mathcal{S}^*_{\mathbf{r}}(\beta)^{a_k}(-\sigma_1)\big) \oplus \bigoplus_{a_k<0} R^1\pi_*\big(\mathcal{S}^*_{\mathbf{r}}(\beta)^{a_k}(-\sigma_1)\big) \to \overline{\mathcal{M}}_{0,m|d},$$
$$\ddot{\mathcal{V}}'^{(d)}_{\mathbf{a};\mathbf{r}}(\beta) = \bigoplus_{a_k>0} R^0\pi_*\big(\mathcal{S}^*_{\mathbf{r}}(\beta)^{a_k}(-\sigma_2)\big) \oplus \bigoplus_{a_k<0} R^1\pi_*\big(\mathcal{S}^*_{\mathbf{r}}(\beta)^{a_k}(-\sigma_2)\big) \to \overline{\mathcal{M}}_{0,m|d},$$

where $\pi : \mathcal{U} \to \overline{\mathcal{M}}_{0,m|d}$ is the projection as before; these sheaves are locally free. If $m' \in \mathbb{Z}^+$, $2 \leq m' \leq m$, and $\mathbf{r} \in (\mathbb{Z}^{\geq 0})^{m-m'}$, let

(4-5) $$\dot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta) = f^*_{m',m}\dot{\mathcal{V}}'^{(d)}_{\mathbf{a};\mathbf{r}}(\beta),\ \ddot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta) = f^*_{m',m}\ddot{\mathcal{V}}'^{(d)}_{\mathbf{a};\mathbf{r}}(\beta) \to \overline{\mathcal{M}}_{0,m|d}.$$

In the case $m' = m$, we will denote the bundles $\dot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta)$ and $\ddot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta)$ by $\dot{\mathcal{V}}^{(d)}_{\mathbf{a}}(\beta)$ and $\ddot{\mathcal{V}}^{(d)}_{\mathbf{a}}(\beta)$, respectively.

The equivariant Euler classes of the bundles $\dot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta)$ and $\ddot{\mathcal{V}}^{(d)}_{\mathbf{a};\mathbf{r}}(\beta)$ enter into the localization computations in Sections 7–9. As a corollary of these computations, we obtain closed formulas for the Euler classes of these bundles in the case $m = 3$; see Propositions 4.1 and 4.2 below. These formulas are a key ingredient in computing the genus-0 three-point and genus-1 SQ invariants.

If $f \in \mathbb{Q}_\alpha[\![q]\!]$ and $d \in \mathbb{Z}^{\geq 0}$, let $[\![f]\!]_{q;d} \in \mathbb{Q}_\alpha$ denote the coefficient of q^d in f. If $f = f(z)$ is a rational function in z and possibly some other variables, for any $z_0 \in \mathbb{P}^1 \supset \mathbb{C}$ let

(4-6) $$\underset{z=z_0}{\mathfrak{R}} f(z) \equiv \frac{1}{2\pi \mathrm{i}} \oint f(z)\,\mathrm{d}z,$$

where the integral is taken over a positively oriented loop around $z = z_0$ with no other singular points of $f\,\mathrm{d}z$, denote the residue of the 1-form $f\,\mathrm{d}z$. If $z_1, \ldots, z_k \in \mathbb{P}^1$ is any collection of points, let

(4-7) $$\underset{z=z_1,\ldots,z_k}{\mathfrak{R}} f(z) \equiv \sum_{i=1}^{i=k} \underset{z=z_i}{\mathfrak{R}} f(z)$$

be the sum of the corresponding residues.

For any variable $\mathbf{y}$ and $r \in \mathbb{Z}^{\geq 0}$, let $\mathbf{s}_r(\mathbf{y})$ denote the r-th elementary symmetric polynomial in $\{\mathbf{y} - \alpha_k\}$. We define power series $L_{n;\mathbf{a}}, \xi_{n;\mathbf{a}} \in \mathbb{Q}_\alpha[\mathbf{x}][\![q]\!]$ by

(4-8) $$\begin{aligned} &L_{n;\mathbf{a}} \in \mathbf{x} + q\mathbb{Q}_\alpha[\mathbf{x}][\![q]\!], \quad && \mathbf{s}_n(L_{n;\mathbf{a}}(\mathbf{x},q)) - q\mathbf{a}^{\mathbf{a}} L_{n;\mathbf{a}}(\mathbf{x},q)^{|\mathbf{a}|} = \mathbf{s}_n(\mathbf{x}),\\ &\xi_{n;\mathbf{a}} \in q\mathbb{Q}_\alpha[\mathbf{x}][\![q]\!], && \mathbf{x} + q\frac{\mathrm{d}}{\mathrm{d}q}\xi_{n;\mathbf{a}}(\mathbf{x},q) = L_{n;\mathbf{a}}(\mathbf{x},q). \end{aligned}$$

By [Zinger 2014, Remark 4.5], the coefficients of the power series

$$\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\, \dot{\mathcal{Y}}_{n;\mathbf{a}}(\alpha_i,\hbar,q) \in \mathbb{Q}_\alpha[\hbar][\![q]\!]$$

are regular at $\hbar = 0$. Thus, there is an expansion

(4-9) $$\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\, \dot{\mathcal{Y}}_{n;\mathbf{a}}(\alpha_i,\hbar,q) = \sum_{r=0}^{\infty} \dot{\Phi}^{(r)}_{n;\mathbf{a}}(\alpha_i,q)\hbar^r,$$

with $\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\mathbf{x},q) - 1$, $\dot{\Phi}^{(1)}_{n;\mathbf{a}}(\mathbf{x},q)$, $\dot{\Phi}^{(2)}_{n;\mathbf{a}}(\mathbf{x},q), \cdots \in q\mathbb{Q}_\alpha[\mathbf{x}][\![q]\!]$. Furthermore,

(4-10) $$\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\mathbf{x},q) = \left(\frac{\mathbf{x}\cdot\mathbf{s}_{n-1}(\mathbf{x})}{L_{n;\mathbf{a}}(\mathbf{x},q)\,\mathbf{s}_{n-1}(L_{n;\mathbf{a}}(\mathbf{x},q)) - |\mathbf{a}|q\mathbf{a}^{\mathbf{a}}L_{n;\mathbf{a}}(\mathbf{x},q)^{|\mathbf{a}|}}\right)^{\frac{1}{2}} \left(\frac{L_{n;\mathbf{a}}(\mathbf{x},q)}{\mathbf{x}}\right)^{\frac{\ell(\mathbf{a})+1}{2}}.$$ [5]

Proposition 4.1. *If $l \in \mathbb{Z}^{\geq 0}$, $n \in \mathbb{Z}^+$, and $\mathbf{a} \in (\mathbb{Z}^*)^l$, then for every $i = 1, \ldots, n$*

$$\sum_{d=0}^{\infty} \frac{q^d}{d!} \int_{\overline{\mathcal{M}}_{0,3|d}} \frac{\mathbf{e}\big(\dot{\mathcal{V}}^{(d)}_{\mathbf{a}}(\alpha_i)\big)}{\prod_{k\neq i} \mathbf{e}\big(\dot{\mathcal{V}}^{(d)}_1(\alpha_i-\alpha_k)\big)\,(\hbar_1-\psi_1)(\hbar_2-\psi_2)(\hbar_3-\psi_3)}$$
$$= \frac{\mathrm{e}^{\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar_1 + \xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar_2 + \xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar_3}}{\hbar_1\hbar_2\hbar_3\, \dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)} \in \mathbb{Q}_\alpha[\![\hbar_1^{-1},\hbar_2^{-1},\hbar_3^{-1},q]\!].$$

[5] Only the case $\ell^-(\mathbf{a}) = 0$ is explicitly considered in [Zinger 2014], but the argument is the same in all cases.

Proposition 4.2. *If* $l \in \mathbb{Z}^{\ge 0}$, $n \in \mathbb{Z}^+$, *and* $\mathbf{a} \in (\mathbb{Z}^*)^l$, *then for every* $i = 1, \dots, n$

$$\sum_{b=0}^{\infty}\sum_{r=0}^{\infty}\sum_{d=0}^{\infty}\frac{q^d}{d!}\int_{\overline{\mathcal{M}}_{0,3|d}}\frac{\mathbf{e}\big(\dot{\mathcal{V}}_{\mathbf{a};r}^{(d)}(\alpha_i)\big)\psi_3^b\,\mathfrak{R}_{\hbar=0}\frac{(-1)^b}{\hbar^{b+1}}[\![\dot{\mathcal{Y}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;r}q^r}{\prod_{k\neq i}\mathbf{e}\big(\dot{\mathcal{V}}_1^{(d)}(\alpha_i-\alpha_k)\big)\,(\hbar_1-\psi_1)(\hbar_2-\psi_2)}$$
$$= \frac{\mathrm{e}^{\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar_1+\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar_2}}{\hbar_1\hbar_2} \in \mathbb{Q}_\alpha[\![\hbar_1^{-1},\hbar_2^{-1},q]\!].$$

5. Outline of the proof of Theorem 3

The first identity in (3-12) is the subject of [Cooper and Zinger 2014, Theorem 3]. The proof of the remaining statements of Theorem 3 follows the same principle as the proof of [Popa and Zinger 2014, Theorem 4]; it is outlined below. However, its adaptation to the present situation requires a number of modifications. In particular, the twisted stable quotients invariants are not known to satisfy the analogue of the string relation of Gromov–Witten theory (in fact, by Proposition 3.1, in general they do not). This requires a direct proof of the key properties for the stable quotients analogue of double Givental's J-function described in Lemmas 6.5 and 6.6 below; in Gromov–Witten theory, these properties are deduced from the analogous properties for three-point invariants, which simplifies the argument. We thus describe the argument in detail.

Let $\mathbb{Q}_\alpha\lceil\!\lceil\hbar\rceil\!\rceil \equiv \mathbb{Q}_\alpha[\![\hbar^{-1}]\!]+\mathbb{Q}_\alpha[\hbar]$ denote the $\mathbb{Q}_\alpha$-algebra of Laurent series in $\hbar^{-1}$ (with finite principal part). We will view the $\mathbb{Q}_\alpha$-algebra $\mathbb{Q}_\alpha(\hbar)$ of rational functions in $\hbar$ with coefficients in $\mathbb{Q}_\alpha$ as a subalgebra of $\mathbb{Q}_\alpha\lceil\!\lceil\hbar\rceil\!\rceil$ via the embedding given by taking the Laurent series of rational functions at $\hbar^{-1}=0$. If

$$\mathcal{F}(\hbar,q)=\sum_{d=0}^{\infty}\sum_{r=-N_d}^{\infty}\mathcal{F}^{(r)}(d)\hbar^{-r}q^d \in \mathbb{Q}_\alpha\lceil\!\lceil\hbar\rceil\!\rceil[\![q]\!]$$

for some $N_d\in\mathbb{Z}$ and $\mathcal{F}^{(r)}(d)\in\mathbb{Q}_\alpha$, we define

$$\mathcal{F}(\hbar,q)\cong\sum_{d=0}^{\infty}\sum_{r=-N_d}^{p-1}\mathcal{F}^{(r)}(d)\hbar^{-r}\pmod{\hbar^{-p}},$$

that is we drop $\hbar^{-p}$ and higher powers of $\hbar^{-1}$, instead of higher powers of $\hbar$.

For $1\le i,j\le n$ with $i\neq j$ and $d\in\mathbb{Z}^+$, let

(5-1)
$$\dot{\mathfrak{C}}_i^j(d)\equiv\frac{\prod_{a_k>0}\prod_{r=1}^{a_kd}\big(a_k\alpha_i+r\,\frac{\alpha_j-\alpha_i}{d}\big)\,\prod_{a_k<0}\prod_{r=0}^{-a_kd-1}\big(a_k\alpha_i-r\,\frac{\alpha_j-\alpha_i}{d}\big)}{d\prod_{r=1}^{d}\prod_{k=1}^{n}{}_{(r,k)\neq(d,j)}\big(\alpha_i-\alpha_k+r\,\frac{\alpha_j-\alpha_i}{d}\big)}\in\mathbb{Q}_\alpha,$$

$$\ddot{\mathfrak{C}}_i^j(d)\equiv\frac{\prod_{a_k>0}\prod_{r=0}^{a_kd-1}\big(a_k\alpha_i+r\,\frac{\alpha_j-\alpha_i}{d}\big)\,\prod_{a_k<0}\prod_{r=1}^{-a_kd}\big(a_k\alpha_i-r\,\frac{\alpha_j-\alpha_i}{d}\big)}{d\prod_{r=1}^{d}\prod_{k=1}^{n}{}_{(r,k)\neq(d,j)}\big(\alpha_i-\alpha_k+r\,\frac{\alpha_j-\alpha_i}{d}\big)}\in\mathbb{Q}_\alpha.$$

We will follow the five steps in [Zinger 2009, Section 1.3] to verify (3-11), the second statement in (3-12), and (3-15):

(Ma) if $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})⟦\hbar⟧⟦q⟧$,

$$\mathcal{F}(\mathbf{x}=\alpha_i, \hbar, q) \in \mathbb{Q}_\alpha(\hbar)⟦q⟧ \subset \mathbb{Q}_\alpha⟦\hbar⟧⟦q⟧ \text{ for all } i = 1, 2, \dots, n,$$

$\mathcal{F}'$ is recursive in the sense of Definition 6.1, and $\mathcal{F}$ and $\mathcal{F}'$ satisfy a mutual polynomiality condition (MPC) of Definition 6.2, then the transforms of $\mathcal{F}'$ of Lemma 6.4 are also recursive and satisfy the same MPC with respect to $\mathcal{F}$;

(Mb) if $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})⟦\hbar⟧⟦q⟧$,

$$\mathcal{F}(\mathbf{x}=\alpha_i, \hbar, q) \in \mathbb{Q}^*_\alpha + q \cdot \mathbb{Q}_\alpha(\hbar)⟦q⟧ \subset \mathbb{Q}_\alpha⟦\hbar⟧⟦q⟧ \text{ for all } i = 1, 2, \dots, n,$$

$\mathcal{F}'$ is recursive in the sense of Definition 6.1, and $\mathcal{F}$ and $\mathcal{F}'$ satisfy a fixed MPC, then $\mathcal{F}'$ is determined by its "mod $\hbar^{-1}$ part";

(Mc) the two sides of the second identity in (3-12) and the $\ddot{\mathcal{Z}}$ case in (3-15) are $\ddot{\mathfrak{C}}$-recursive in the sense of Definition 6.1 with $\ddot{\mathfrak{C}}$ as in (5-1), while the two sides of the $\dot{\mathcal{Z}}$ case in (3-15) are $\dot{\mathfrak{C}}$-recursive in the sense of Definition 6.1 with $\dot{\mathfrak{C}}$ as in (5-1);

(Md) the two sides of each of the equations in (3-12) and (3-15) satisfy the same η-MPC (dependent on the equation) with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)$;

(Me) the two sides of each of the four equations in (3-12) and (3-15), viewed as elements of $H^*_{\mathbb{T}}(\mathbb{P}^{n-1})⟦\hbar⟧⟦q⟧$, agree mod $\hbar^{-1}$.

The first two claims, (Ma) and (Mb), sum up Lemma 6.4 and Proposition 6.3, respectively. By Lemmas 6.5 and 6.6, the stable quotients generating functions $\dot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}$ and $\ddot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}$ are $\dot{\mathfrak{C}}$-recursive and $\ddot{\mathfrak{C}}$-recursive and satisfy MPCs with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)$. Along with the first identity in (3-12), the latter implies that they satisfy MPCs with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}$. It is immediate from (3-4) that

$$\dot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q), \ddot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \cong \mathbf{x}^s \pmod{\hbar^{-1}} \text{ for all } s \in \mathbb{Z}^{\geq 0}. \tag{5-2}$$

By the proof of the first identity in (3-12), as well as of its Gromov–Witten analogue, the power series $\dot{\mathcal{Y}}_{n;\mathbf{a}}$ is $\dot{\mathfrak{C}}$-recursive and satisfies the same MPC with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}$ as $\dot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}$; see [Cooper and Zinger 2014, Lemma 5.4]. A nearly identical argument shows that the power series $\ddot{\mathcal{Y}}_{n;\mathbf{a}}$ is $\ddot{\mathfrak{C}}$-recursive and satisfies the same MPC with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}$ as $\ddot{\mathcal{Z}}^{(s)}_{n;\mathbf{a}}$; see [Popa and Zinger 2014, Section 4.3] for the $\ell^-(\mathbf{a}) = 0$ case. Since

$$\ddot{\mathcal{Y}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \cong 1 \pmod{\hbar^{-1}},$$

this establishes the second identity in (3-12). Along with (3-12), the admissibility of transforms (i) and (ii) in Lemma 6.4 implies that both sides of the $\dot{\mathcal{Z}}$ equation in

(3-15) are $\dot{\mathfrak{C}}$-recursive and satisfy the same MPC with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}$, no matter what the coefficients $\widetilde{\mathcal{C}}^{(r)}_{s,s'}$ are. Similarly, both sides of the $\ddot{\mathcal{Z}}$ equation in (3-15) are $\ddot{\mathfrak{C}}$-recursive and satisfy the same MPC with respect to $\dot{\mathcal{Y}}_{n;\mathbf{a}}$. By (3-10), (3-12), (3-9), (3-21), (3-19), (3-17), and (3-16),

$$\begin{aligned}\sum_{r=0}^{s}\sum_{s'=0}^{s-r}\widetilde{\mathcal{C}}^{(r)}_{s-\ell^-(\mathbf{a}),s'-\ell^-(\mathbf{a})}(q)\,\hbar^{s-r-s'}\mathfrak{D}^{s'}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\cong\mathbf{x}^s \pmod{\hbar^{-1}},\\ \sum_{r=0}^{s}\sum_{s'=0}^{s-r}\widetilde{\mathcal{C}}^{(r)}_{s-\ell^+(\mathbf{a}),s'-\ell^+(\mathbf{a})}(q)\,\hbar^{s-r-s'}\mathfrak{D}^{s'}\ddot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\cong\mathbf{x}^s \pmod{\hbar^{-1}}.\end{aligned}\tag{5-3}$$ [6]

Thus, (3-15) follows from (Mb).

The proof of (3-11) follows the same principle, which we apply to a multiple of (3-11). For each $i=1,2,\dots,n$, let

$$\phi_i\equiv\prod_{k\neq i}(\mathbf{x}-\alpha_k)\in H^*_{\mathbb{T}}(\mathbb{P}^{n-1}).\tag{5-4}$$

By [Atiyah and Bott 1984, localization theorem], ϕ_i is the equivariant Poincaré dual of the fixed point $P_i\in\mathbb{P}^{n-1}$; see [Zinger 2009, Section 3.1]. Since $\mathbf{x}|_{P_i}=\alpha_i$,

$$\begin{aligned}&\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\alpha_j,\hbar_1,\hbar_2,q)\\ &\quad=\int_{P_i\times P_j}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)\\ &\quad=\int_{\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)\phi_i\times\phi_j\\ &\quad=\frac{1}{\hbar_1+\hbar_2}\prod_{k\neq i}(\alpha_j-\alpha_k)+\sum_{d=1}^{\infty}q^d\int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\,\mathrm{ev}_1^*\phi_i\,\mathrm{ev}_2^*\phi_j}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)};\end{aligned}\tag{5-5}$$

the last equality holds by the defining property of the cohomology push-forward [Zinger 2009, Equation (3.11)]. By Lemmas 6.5 and 6.6, $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q)$ is $\dot{\mathfrak{C}}$-recursive and satisfies the same MPC as $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ for $(\mathbf{x},\hbar)=(\mathbf{x}_1,\hbar_1)$ and $\mathbf{x}_2=\alpha_j$ fixed.[7] It is also $\ddot{\mathfrak{C}}$-recursive and satisfies the same MPC as $\ddot{\mathcal{Z}}_{n;\mathbf{a}}$ with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ for $(\mathbf{x},\hbar)=(\mathbf{x}_2,\hbar_2)$ and $\mathbf{x}_1=\alpha_i$ fixed.

[6] The left-hand side of (3-21) with s replaced by $s-\ell^-(\mathbf{a})$ is the coefficient of $\hbar^s\mathbf{x}^{-r}(\mathbf{x}/\hbar)^{s'+\ell^-(\mathbf{a})}$ in the first identity in (5-3) if $s\ge\ell^-(\mathbf{a})$; The left-hand side of (3-21) with s replaced by $s-\ell^+(\mathbf{a})$ is the coefficient of $\hbar^s\mathbf{x}^{-r}(\mathbf{x}/\hbar)^{s'+\ell^+(\mathbf{a})}$ in the second identity in (5-3) if $s\ge\ell^+(\mathbf{a})$.

[7] In other words, the coefficient of every power of $\hbar_2^{-1}$ in $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\alpha_j,\hbar,\hbar_2,q)$ is $\dot{\mathfrak{C}}$-recursive and satisfies the same MPC as $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$.

By (Ma) and (Mb), it is thus sufficient to compare

$$(\hbar_1+\hbar_2)\,\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}_1,\mathbf{x}_2,\hbar_1,\hbar_2,q) \quad \text{and} \quad \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r\, \dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\mathbf{x}_1,\hbar_1,q)\, \ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\mathbf{x}_2,\hbar_2,q) \tag{5-6}$$

for all $\mathbf{x}_1=\alpha_i$ and $\mathbf{x}_2=\alpha_j$ with $i, j=1,2,\dots,n$ modulo $\hbar_1^{-1}$:

$$(\hbar_1+\hbar_2)\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\alpha_j,\hbar_1,\hbar_2,q) \cong \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\alpha_j^{s_2} + \sum_{d=1}^{\infty} q^d \int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},d)} \frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_1^*\phi_i \mathrm{ev}_2^*\phi_j}{\hbar_2-\psi_2};$$

$$\sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\alpha_i,\hbar_1,q)\ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\alpha_j,\hbar_2,q) \cong \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\, \ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\alpha_j,\hbar_2,q).$$

In order to see that the two right-hand side power series are the same, it is sufficient to compare them modulo $\hbar_2^{-1}$:

$$\sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\alpha_j^{s_2} + \sum_{d=1}^{\infty} q^d \int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},d)} \frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a}})\mathrm{ev}_1^*\phi_i \mathrm{ev}_2^*\phi_j}{\hbar_2-\psi_2} \cong \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\alpha_j^{s_2};$$

$$\sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\, \ddot{\mathcal{Z}}^{(s_2)}_{n;\mathbf{a}}(\alpha_j,\hbar_2,q) \cong \sum_{\substack{s_1,s_2,r\ge 0\\ s_1+s_2+r=n-1}} (-1)^r \mathbf{s}_r \alpha_i^{s_1}\alpha_j^{s_2}.$$

From this we conclude that the two expressions in (5-6) are the same; this proves (3-11).

By Proposition 6.3 and Lemmas 6.5 and 6.6, the stable quotients analogue of triple Givental's J-function is determined by the primary three-point SQ invariants. Since all such invariants are related to the corresponding GW invariants by [Ciocan-Fontanine and Kim 2013, Theorem 1.2.2 and Corollaries 1.4.1, 1.4.2], a version of (3-14) can be proved by comparing it to its GW analogue provided by [Zinger 2014, Theorem B]. We instead prove (3-14) directly in Section 10 by reducing the computation to the two-point formulas of Theorem 3 and the mirror formula for Hurwitz numbers in Propositions 4.1. In the process, we obtain a precise description of the equivariant structure coefficients appearing in (3-14), which is not done in [Zinger 2014].

6. Recursivity, polynomiality, and admissible transforms

This section describes the algebraic observations used in the proof of Theorem 3. It is based on [Zinger 2009, Sections 2.1, 2.2] and [Popa and Zinger 2014, Section 4.1]. Let

$$[n]=\{1,2,\dots,n\}.$$

Definition 6.1. Let $C\equiv(C_i^j(d))_{d,i,j\in\mathbb{Z}^+}$ be any collection of elements of $\mathbb{Q}_\alpha$. A power series $\mathcal{F}\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})\lceil\hbar\rceil[\![q]\!]$ is C-recursive if the following holds: if $d^*\in\mathbb{Z}^{\ge0}$ is such that

$$[\![\mathcal{F}(\mathbf{x}=\alpha_i,\hbar,q)]\!]_{q;d^*-d}\in\mathbb{Q}_\alpha(\hbar)\subset\mathbb{Q}_\alpha\lceil\hbar\rceil\ \text{ for all } d\in[d^*],\ i\in[n],$$

and $[\![\mathcal{F}(\alpha_i,\hbar,q)]\!]_{q;d}$ is regular at $\hbar=(\alpha_i-\alpha_j)/d$ for all $d<d^*$ and $i\neq j$, then

$$[\![\mathcal{F}(\alpha_i,\hbar,q)]\!]_{q;d^*}-\sum_{d=1}^{d^*}\sum_{j\neq i}\frac{C_i^j(d)}{\hbar-(\alpha_j-\alpha_i)/d}[\![\mathcal{F}(\alpha_j,z,q)]\!]_{q;d^*-d}\Big|_{z=(\alpha_j-\alpha_i)/d} \tag{6-1}$$
$$\in\mathbb{Q}_\alpha[\hbar,\hbar^{-1}]\subset\mathbb{Q}_\alpha\lceil\hbar\rceil.$$

Thus, if $\mathcal{F}\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})\lceil\hbar\rceil[\![q]\!]$ is C-recursive, for any collection C, then

$$\mathcal{F}(\mathbf{x}=\alpha_i,\hbar,q)\in\mathbb{Q}_\alpha(\hbar)[\![q]\!]\subset\mathbb{Q}_\alpha\lceil\hbar\rceil[\![q]\!]\ \text{ for all } i\in[n],$$

as can be seen by induction on d, and

$$\begin{aligned}&\mathcal{F}(\alpha_i,\hbar,q)\\&\quad=\sum_{d=0}^{\infty}\sum_{r=-N_d}^{N_d}\mathcal{F}_i^r(d)\hbar^rq^d+\sum_{d=1}^{\infty}\sum_{j\neq i}\frac{C_i^j(d)q^d}{\hbar-(\alpha_j-\alpha_i)/d}\mathcal{F}(\alpha_j,(\alpha_j-\alpha_i)/d,q)\end{aligned} \tag{6-2}$$

for all $i\in[n]$, for some $\mathcal{F}_i^r(d)\in\mathbb{Q}_\alpha$. The nominal issue with defining C-recursivity by (6-2), as is normally done, is that a priori the evaluation of $\mathcal{F}(\alpha_j,\hbar,q)$ at $\hbar=(\alpha_j-\alpha_i)/d$ need not be well defined, since $\mathcal{F}(\alpha_j,\hbar,q)$ is a power series with coefficients in $\mathbb{Q}_\alpha\lceil\hbar^{-1}\rceil$; a priori they may not converge anywhere. However, taking the coefficient of each power of q in (6-2) shows by induction on the degree d that this evaluation does make sense; this is the substance of Definition 6.1.

Definition 6.2. Let $\eta\in\mathbb{Q}_\alpha(\mathbf{x})$ be such that $\eta(\mathbf{x}=\alpha_i)\in\mathbb{Q}_\alpha$ is well defined and nonzero for every $i\in[n]$. For any $\mathcal{F}\equiv\mathcal{F}(\mathbf{x},\hbar,q),\mathcal{F}'\equiv\mathcal{F}'(\mathbf{x},\hbar,q)\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})\lceil\hbar\rceil[\![q]\!]$, let

$$\begin{aligned}&\Phi^\eta_{\mathcal{F},\mathcal{F}'}(\hbar,z,q)\\&\quad\equiv\sum_{i=1}^{n}\frac{\eta(\alpha_i)\mathrm{e}^{\alpha_iz}}{\prod_{k\neq i}(\alpha_i-\alpha_k)}\mathcal{F}(\alpha_i,\hbar,q\mathrm{e}^{\hbar z})\mathcal{F}'(\alpha_i,-\hbar,q)\in\mathbb{Q}_\alpha\lceil\hbar\rceil[\![z,q]\!].\end{aligned} \tag{6-3}$$

If $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar]\!][\![q]\!]$, the pair $(\mathcal{F}, \mathcal{F}')$ satisfies the η mutual polynomiality condition (η-MPC) if $\Phi^{\eta}_{\mathcal{F},\mathcal{F}'} \in \mathbb{Q}_\alpha[\hbar][\![z, q]\!]$.

If $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar]\!][\![q]\!]$ and

$$\mathcal{F}(\mathbf{x}=\alpha_i, \hbar, q), \mathcal{F}'(\mathbf{x}=\alpha_i, \hbar, q) \in \mathbb{Q}_\alpha(\hbar)[\![q]\!] \quad \text{for all } i \in [n], \tag{6-4}$$

then the pair $(\mathcal{F}, \mathcal{F}')$ satisfies the η-MPC if and only if the pair $(\mathcal{F}', \mathcal{F})$ does; see [Zinger 2009, Lemma 2.2] for the $\eta = 1$, $\ell^+(\mathbf{a}) = 1$, $\ell^-(\mathbf{a}) = 0$ case (the proof readily carries over to the general case). Thus, if (6-4) holds, the statement that $\mathcal{F}$ and $\mathcal{F}'$ satisfy the MPC is unambiguous.

Proposition 6.3. *Let $\eta \in \mathbb{Q}_\alpha(\mathbf{x})$ be such that $\eta(\mathbf{x}=\alpha_i) \in \mathbb{Q}_\alpha$ is well defined and nonzero for every $i \in [n]$. If $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar]\!][\![q]\!]$,*

$$\mathcal{F}(\mathbf{x}=\alpha_i, \hbar, q) \in \mathbb{Q}_\alpha^* + q \cdot \mathbb{Q}_\alpha(\hbar)[\![q]\!] \subset \mathbb{Q}_\alpha[\![\hbar]\!][\![q]\!] \quad \text{for all } i \in [n],$$

$\mathcal{F}'$ is recursive, and $\mathcal{F}$ and $\mathcal{F}'$ satisfy the η-MPC, then $\mathcal{F}' \cong 0 \pmod{\hbar^{-1}}$ if and only if $\mathcal{F}' = 0$.

This is essentially [Zinger 2009, Proposition 2.1], with the assumptions corrected in [Popa and Zinger 2014, Footnote 3]. The proof in [Zinger 2009], which treats the $\eta = 1$ case, readily extends to the general case; see also the paragraph following [Popa and Zinger 2014, Proposition 4.3].

Lemma 6.4. *Let $C \equiv (C_i^j(d))_{d,i,j \in \mathbb{Z}^+}$ be any collection of elements of $\mathbb{Q}_\alpha$ and $\eta \in \mathbb{Q}_\alpha(\mathbf{x})$ be such that $\eta(\mathbf{x}=\alpha_i) \in \mathbb{Q}_\alpha$ is well defined and nonzero for every $i \in [n]$. If $\mathcal{F}, \mathcal{F}' \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar]\!][\![q]\!]$,*

$$\mathcal{F}(\mathbf{x}=\alpha_i, \hbar, q) \in \mathbb{Q}_\alpha(\hbar)[\![q]\!] \subset \mathbb{Q}_\alpha[\![\hbar]\!][\![q]\!] \quad \text{for all } i \in [n],$$

$\mathcal{F}'$ is C-recursive (and satisfies the η-MPC with respect to $\mathcal{F}$), then

(i) *$\{\mathbf{x} + \hbar\, q\, \mathrm{d}/\mathrm{d}q\}\mathcal{F}'$ is C-recursive (and satisfies the η-MPC with respect to $\mathcal{F}$);*

(ii) *if $f \in \mathbb{Q}_\alpha[\hbar][\![q]\!]$, then $f\mathcal{F}'$ is C-recursive (and satisfies the η-MPC with respect to $\mathcal{F}$).*

This lemma is essentially contained in [Zinger 2009, Lemma 2.3]. The proof in [Zinger 2009], which treats the $\eta = 1$ case, readily extends to the general case; see also the paragraph following [Popa and Zinger 2014, Lemma 4.4].

The next two sections establish Lemmas 6.5 and 6.6 below, the $m = 2$ cases of which complete the proofs of (3-11), the second statement in (3-12), and (3-15). The $m = 3$ cases of these lemmas are used in the proof of Proposition 3.1 and 4.1 in Section 9. If $m \geq m' \geq 2$, let

$$f_{m',m} : \overline{Q}_{0,m}(\mathbb{P}^{n-1}, d) \to \overline{Q}_{0,m'}(\mathbb{P}^{n-1}, d) \tag{6-5}$$

denote the forgetful morphism dropping the last $m-m'$ points; this morphism is defined if $m'>2$ or $d>0$. With the bundles

$$\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)}, \ddot{\mathcal{V}}_{n;\mathbf{a}}^{(d)} \to \overline{Q}_{0,m'}(\mathbb{P}^{n-1},d)$$

defined by (2-1), let

$$\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(d)} = f_{m',m}^*\dot{\mathcal{V}}_{n;\mathbf{a}}^{(d)},\ \ddot{\mathcal{V}}_{n;\mathbf{a};m'}^{(d)} = f_{m',m}^*\ddot{\mathcal{V}}_{n;\mathbf{a}}^{(d)} \to \overline{Q}_{0,m}(\mathbb{P}^{n-1},d). \tag{6-6}$$

For $\mathbf{b}\equiv(b_2,\dots,b_m)\in(\mathbb{Z}^{\ge 0})^{m-1}$ and $\varpi\equiv(\varpi_2,\dots,\varpi_m)\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})^{m-1}$, let

(6-7)

$$\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\mathbf{x},\hbar,q) \equiv \sum_{d=0}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(d)})}{\hbar-\psi_1}\prod_{j=2}^{j=m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)\right] \in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!],$$

$$\ddot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\mathbf{x},\hbar,q) \equiv \sum_{d=0}^{\infty} q^d \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a};m'}^{(d)})}{\hbar-\psi_1}\prod_{j=2}^{j=m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)\right] \in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})[\hbar^{-1}][\![q]\!],$$

where $\mathrm{ev}_j:\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)\to\mathbb{P}^{n-1}$ is the evaluation map at the j-th marked point and the degree-0 terms in the $m'=2$ case are defined by

$$\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};2}^{(0)}),\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a};2}^{(0)}) = 1 \qquad \text{if } m\ge 3,$$

$$\mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};2}^{(0)})}{\hbar-\psi_1}(\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2)\right], \mathrm{ev}_{1*}\left[\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a};2}^{(0)})}{\hbar-\psi_1}(\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2)\right] = (-\hbar)^{b_2}\varpi_2 \quad \text{if } m=2.$$

Lemma 6.5. *Let $l\in\mathbb{Z}^{\ge 0}$, $m,m',n\in\mathbb{Z}^+$ with $m\ge m'\ge 2$, and $\mathbf{a}\in(\mathbb{Z}^*)^l$. For all $\mathbf{b}\in(\mathbb{Z}^{\ge 0})^{m-1}$ and $\varpi\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})^{m-1}$, the power series $\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}$ and $\ddot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}$ defined by (6-7) are $\dot{\mathfrak{C}}$ and $\ddot{\mathfrak{C}}$-recursive, respectively.*

Lemma 6.6. *Let $l\in\mathbb{Z}^{\ge 0}$, $m,m',n\in\mathbb{Z}^+$ with $m\ge m'\ge 2$, $\mathbf{a}\in(\mathbb{Z}^*)^l$,*

$$\dot{\eta}(\mathbf{x}) = \langle\mathbf{a}\rangle\mathbf{x}^{\ell(\mathbf{a})}, \quad \ddot{\eta}(\mathbf{x}) = 1.$$

For all $\mathbf{b}\in(\mathbb{Z}^{\ge 0})^{m-1}$ and $\varpi\in H_{\mathbb{T}}^*(\mathbb{P}^{n-1})^{m-1}$, the power series

$$\hbar^{m-2}\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\mathbf{x},\hbar,q) \quad \text{and} \quad \hbar^{m-2}\ddot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\mathbf{x},\hbar,q)$$

satisfy the $\dot{\eta}$ and $\ddot{\eta}$-MPC, respectively, with respect to the power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ defined by (3-3).

By Lemma 6.5, the power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}^{(s)}$ and $\ddot{\mathcal{Z}}_{n;\mathbf{a}}^{(s)}$ defined by (3-4) are $\dot{\mathfrak{C}}$- and $\ddot{\mathfrak{C}}$-recursive, respectively. Furthermore, the power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ defined by (3-7) is $\dot{\mathfrak{C}}$-recursive for $(\mathbf{x},\hbar)=(\mathbf{x}_1,\hbar_1)$ and $\mathbf{x}_2=\alpha_j$ fixed and is $\ddot{\mathfrak{C}}$-recursive for $(\mathbf{x},\hbar)=(\mathbf{x}_2,\hbar_2)$ and $\mathbf{x}_1=\alpha_j$ fixed. By Lemma 6.6, $\dot{\mathcal{Z}}_{n;\mathbf{a}}^{(s)}$ and $\ddot{\mathcal{Z}}_{n;\mathbf{a}}^{(s)}$ satisfy the $\dot{\eta}$- and $\ddot{\eta}$-MPC, respectively, with respect to the power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ defined by (3-3).

Furthermore, the power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ defined by (3-7) satisfies the $\dot{\eta}$-MPC with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ for $(\mathbf{x},\hbar)=(\mathbf{x}_1,\hbar_1)$ and $\mathbf{x}_2=\alpha_j$ fixed and the $\ddot{\eta}$-MPC with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)$ for $(\mathbf{x},\hbar)=(\mathbf{x}_2,\hbar_2)$ and $\mathbf{x}_1=\alpha_j$ fixed.

In the case of products of projective spaces and concavex sheaves (1-13), the above Definition 6.1 becomes inductive on the total degree $d_1+\cdots+d_p$ of $q_1^{d_1}\cdots q_p^{d_p}$. The power series $\mathcal{F}$ is evaluated at $(\mathbf{x}_1,\dots,\mathbf{x}_p)=(\alpha_{1;i_1},\dots,\alpha_{p;i_p})$ for the purposes of the C-recursivity condition (6-1) and (6-2). The relevant structure coefficients, extending (5-1), are given by

$$\dot{\mathfrak{C}}^j_{i_1\dots i_p}(s;d)\equiv\frac{\displaystyle\prod_{a_{k;1}\ge0}\prod_{r=1}^{a_{k;s}d}\Big(\sum_{t=1}^{p}a_{k;t}\alpha_{t;i_t}+r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)\prod_{a_{k;1}<0}\prod_{r=0}^{-a_{k;s}d-1}\Big(\sum_{t=1}^{p}a_{k;t}\alpha_{t;i_t}-r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)}{d\displaystyle\prod_{\substack{r=1\\(r,k)\neq(d,j)}}^{d}\prod_{k=1}^{n_s}\Big(\alpha_{s;i_s}-\alpha_{s;k}+r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)},$$

$$\ddot{\mathfrak{C}}^j_{i_1\dots i_p}(s;d)\equiv\frac{\displaystyle\prod_{a_{k;1}\ge0}\prod_{r=0}^{a_{k;s}d-1}\Big(\sum_{t=1}^{p}a_{k;t}\alpha_{t;i_t}+r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)\prod_{a_{k;1}<0}\prod_{r=1}^{-a_{k;s}d}\Big(\sum_{t=1}^{p}a_{k;t}\alpha_{t;i_t}-r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)}{d\displaystyle\prod_{\substack{r=1\\(r,k)\neq(d,j)}}^{d}\prod_{k=1}^{n_s}\Big(\alpha_{s;i_s}-\alpha_{s;k}+r\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}\Big)},$$

with $s\in[p]$ and $j\neq i_s$. The double sums in these equations are then replaced by triple sums over $s\in[p]$, $j\in[n_s]-i_s$, and $d\in\mathbb{Z}^+$, and with $\mathcal{F}$ evaluated at

$$\mathbf{x}_t=\begin{cases}\alpha_{s;j} & \text{if } t=s;\\ \alpha_{t;i_t} & \text{if } t\neq s;\end{cases}\qquad z=\frac{\alpha_{s;j}-\alpha_{s;i_s}}{d}.$$

The secondary coefficients $\mathcal{F}_i^r(d)$ in (6-2) now become $\mathcal{F}^r_{i_1\dots i_p}(d_1,\dots,d_p)$, with $i_s\in[n_s]$ and $d_s\in\mathbb{Z}^{\ge0}$. In the analogue of Definition 6.2, $\eta\in R(\mathbf{x}_1,\dots,\mathbf{x}_p)$ is such that the evaluation of η at $(\alpha_{1;i_1},\dots,\alpha_{p;i_p})$ for all elements $(i_1,\dots,i_p)$ of $[n_1]\times\cdots\times[n_p]$ is well defined and not zero, $\Phi_{\mathcal{F}}$ is a power series in $z_1,\dots,z_p$ and $q_1,\dots,q_p$, the sum is taken over all elements $(i_1,\dots,i_p)$ of $[n_1]\times\cdots\times[n_p]$, the leading fraction is replaced by

$$\frac{\eta(\alpha_{1;i_1},\dots,\alpha_{p;i_p})\mathrm{e}^{\alpha_{1;i_1}z_1+\cdots+\alpha_{p;i_p}z_p}}{\prod_{s=1}^{p}\prod_{k\neq i_s}(\alpha_{s;i_s}-\alpha_{s;k})},$$

and the $q\mathrm{e}^{\hbar z}$-insertion in the first power series is replaced by $q_1\mathrm{e}^{\hbar z_1},\dots,q_p\mathrm{e}^{\hbar z_p}$. Lemma 6.6 holds with

$$\dot{\eta}(\mathbf{x}_1,\dots,\mathbf{x}_p)=\frac{\prod_{a_{k;1}\ge0}\sum_{s=1}^{p}a_{k;s}\mathbf{x}_s}{\prod_{a_{k;1}<0}\sum_{s=1}^{p}a_{k;s}\mathbf{x}_s}.$$

7. Recursivity for stable quotients

In this section, we use the classical localization theorem [Atiyah and Bott 1984] to show that the generating functions $\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ and $\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ defined in (6-7) are recursive. The argument is similar to the proof in [Cooper and Zinger 2014, Section 6] of recursivity for the generating function $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ defined by (3-3), but requires some modifications.

If $\mathbb{T}$ acts smoothly on a smooth compact oriented manifold M, there is a well-defined integration-along-the-fiber homomorphism

$$\int_M : H^*_{\mathbb{T}}(M) \to H^*_{\mathbb{T}}$$

for the fiber bundle $BM \to B\mathbb{T}$. The classical localization theorem of [Atiyah and Bott 1984] relates it to integration along the fixed locus of the $\mathbb{T}$-action. The latter is a union of smooth compact orientable manifolds F; $\mathbb{T}$ acts on the normal bundle $\mathcal{N}F$ of each F. Once an orientation of F is chosen, there is a well-defined integration-along-the-fiber homomorphism

$$\int_F : H^*_{\mathbb{T}}(F) \to H^*_{\mathbb{T}}.$$

The localization theorem states that

$$\int_M \eta = \sum_F \int_F \frac{\eta|_F}{\mathbf{e}(\mathcal{N}F)} \in \mathbb{Q}_\alpha \quad \text{for all } \eta \in H^*_{\mathbb{T}}(M), \tag{7-1}$$

where the sum is taken over all components F of the fixed locus of $\mathbb{T}$. Part of the statement of (7-1) is that $\mathbf{e}(\mathcal{N}F)$ is invertible in $H^*_{\mathbb{T}}(F) \otimes_{\mathbb{Q}[\alpha_1,\dots,\alpha_n]} \mathbb{Q}_\alpha$. In the case of the standard action of $\mathbb{T}$ on $\mathbb{P}^{n-1}$, (7-1) implies that

$$\eta|_{P_i} = \int_{\mathbb{P}^{n-1}} \eta\phi_i \in \mathbb{Q}_\alpha \tag{7-2}$$

for all $\eta \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})$, $i = 1, 2, \dots, n$, with ϕ_i as in (5-4).

7A. ***Fixed locus data.*** The proof of Lemma 6.5 involves a localization computation on $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$. Thus, we need to describe the fixed loci of the $\mathbb{T}$-action on $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$, their normal bundles, and the restrictions of the relevant cohomology classes to these fixed loci.

As in the case of stable maps described in [Hori et al. 2003, Section 27.3], the fixed loci of the $\mathbb{T}$-action on $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$ are indexed by decorated graphs,

$$\Gamma = (\mathrm{Ver}, \mathrm{Edg}; \mu, \mathfrak{d}, \vartheta), \tag{7-3}$$

where (Ver, Edg) is a connected graph that has no loops, with Ver and Edg denoting its sets of vertices and edges, and

$$\mu : \mathrm{Ver} \to [n], \quad \mathfrak{d} : \mathrm{Ver} \sqcup \mathrm{Edg} \to \mathbb{Z}^{\geq 0}, \quad \text{and} \quad \vartheta : [m] \to \mathrm{Ver}$$

are maps such that

$$\mu(v_1) \neq \mu(v_2) \ \text{ if } \{v_1, v_2\} \in \mathrm{Edg}, \quad \mathfrak{d}(e) \neq 0 \ \text{ for all } e \in \mathrm{Edg},$$
$$\mathrm{val}(v) \equiv |\vartheta^{-1}(v)| + |\{e \in \mathrm{Edg} : v \in e\}| + \mathfrak{d}(v) \geq 2 \ \text{ for all } v \in \mathrm{Ver}. \tag{7-4}$$

In Figure 1, the vertices of a decorated graph Γ are indicated by dots. The values of the map $(\mu, \mathfrak{d})$ on some of the vertices are indicated next to those vertices. Similarly, the values of the map $\mathfrak{d}$ on some of the edges are indicated next to them. The elements of the sets $[m]$ are shown in bold face; they are linked by line segments to their images under ϑ. By (7-4), no two consecutive vertices have the same first label and thus $j \neq i$.

With Γ as in (7-3), let

$$|\Gamma| \equiv \sum_{v \in \mathrm{Ver}} \mathfrak{d}(v) + \sum_{e \in \mathrm{Edg}} \mathfrak{d}(e)$$

be the degree of Γ. For each $v \in \mathrm{Ver}$, let

$$\mathrm{E}_v = \{e \in \mathrm{Edg} : v \in e\}$$

be the set of edges leaving from v. There is a unique partial order $\prec$ on Ver that has a unique minimal element $v_{\min}$ such that $v_{\min} = \vartheta(1)$ and $v \prec w$ if there exist distinct vertices $v_1, \dots, v_k \in \mathrm{Ver}$ such that

$$v \in \{v_{\min}, v_1, \dots, v_{k-1}\}, \quad w = v_k, \quad \text{and}$$
$$\{v_{\min}, v_1\}, \{v_1, v_2\}, \dots, \{v_{k-1}, v_k\} \in \mathrm{Edg},$$

in other words, v lies between $v_{\min}$ and w in (Ver, Edg). If $e = \{v_1, v_2\} \in \mathrm{Edg}$ is any edge in Γ with $v_1 \prec v_2$, let

$$\Gamma_e \equiv \big(\{v_1, v_2\}, \{e\}; \mu_e, \mathfrak{d}_e, \vartheta_e\big),$$

be the decorated graph as in (7-3) given by

$$\mu_e = \mu|_e, \quad \mathfrak{d}_e(e) = \mathfrak{d}(e), \mathfrak{d}_e|_e = 0, \quad \vartheta_e : \{1, 2\} \to e, \quad \vartheta_e(1) = v_1, \quad \vartheta_e(2) = v_2;$$

see Figure 2.

With $m' \leq m$ as in Lemmas 6.5 and 6.6, let

$$\mathrm{Ver}_{m'} = \big\{v \in \mathrm{Ver} : v \preceq \vartheta(i) \text{ for some } i \in [m']\big\},$$
$$\mathrm{Edg}_{m'} = \big\{\{v_1, v_2\} \in \mathrm{Edg} : v_1, v_2 \in \mathrm{Ver}_{m'}\big\}.$$

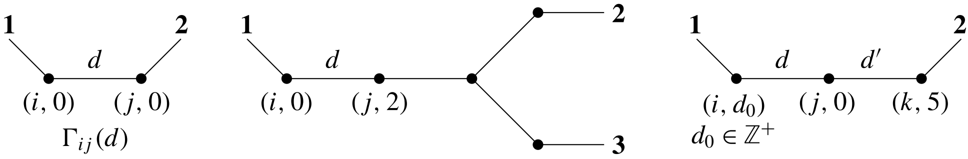


Figure 1. Two trees with $\mathrm{val}(v_{\min}) = 2$ and a tree with $\mathrm{val}(v_{\min}) \geq 3$.

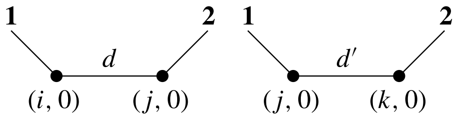


Figure 2. The subtrees corresponding to the edges of the last graph in Figure 1.

In particular, the graph $(\mathrm{Ver}_{m'}, \mathrm{Edg}_{m'})$ is a tree; it is obtained from the original graph $(\mathrm{Ver}, \mathrm{Edg})$ by discarding the branches that do not end at a vertex with a marked point labeled by $i \leq m'$. For each $v \in \mathrm{Ver}_{m'}$, define

$$r_{m';v} : \mathrm{E}_v - \mathrm{Edg}_{m'} \to \mathbb{Z}^+ \quad \text{by} \quad r_{m';v}(\{v, v'\}) = \sum_{\substack{\{v_1,v_2\}\in\mathrm{Edg}\\ v'\preceq v_2}} \mathfrak{d}(\{v_1, v_2\}) + \sum_{\substack{w\in\mathrm{Ver}\\ v'\preceq w}} \mathfrak{d}(w),$$

$$\mathfrak{d}_{m'}(v) = \mathfrak{d}(v) + \sum_{e\in\mathrm{E}_v-\mathrm{Edg}_{m'}} r_{m;'v}(e).$$

This construction increases the degree $\mathfrak{d}(v)$ of a vertex $v \in \mathrm{Ver}_{m'}$ by the total degree of all branches of Γ cut off at v to form the graph $(\mathrm{Ver}_{m'}, \mathrm{Edg}_{m'})$. The motivation for this construction is described at the end of the next paragraph.

As is described in [Marian et al. 2011, Section 7.3], the fixed locus Q_Γ of $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, |\Gamma|)$ corresponding to a decorated graph Γ consists of the stable quotients

$$(\mathcal{C}, y_1, \dots, y_m; S \subset \mathbb{C}^n \otimes \mathcal{O}_{\mathcal{C}})$$

over quasistable rational m-marked curves that satisfy the following conditions. The components of $\mathcal{C}$ on which the corresponding quotient is torsion free are rational and correspond to the edges of Γ; the restriction of S to any such component corresponds to a morphism to $\mathbb{P}^{n-1}$ of the opposite degree to that of the subsheaf. Furthermore, if $e = \{v_1, v_2\}$ is an edge, the corresponding morphism f_e is a degree-$\mathfrak{d}(e)$ cover of the line

$$\mathbb{P}^1_{\mu(v_1),\mu(v_2)} \subset \mathbb{P}^{n-1}$$

passing through the fixed points $P_{\mu(v_1)}$ and $P_{\mu(v_2)}$; it is ramified only over $P_{\mu(v_1)}$ and $P_{\mu(v_2)}$. In particular, f_e is unique up to isomorphism. The remaining components of $\mathcal{C}$ are indexed by the vertices $v \in \mathrm{Ver}$ of valence $\mathrm{val}(v) \geq 3$. The restriction of S to such a component $\mathcal{C}_v$ of $\mathcal{C}$ (or possibly a connected union of irreducible components) is a subsheaf of the trivial subsheaf $P_{\mu(v)} \subset \mathbb{C}^n \otimes \mathcal{O}_{\mathcal{C}_v}$ of degree $-\mathfrak{d}(v)$; thus, the induced morphism takes $\mathcal{C}_v$ to the fixed point $P_{\mu(v)} \in \mathbb{P}^{n-1}$. Each such component $\mathcal{C}_v$ also carries $|\vartheta^{-1}(v)| + |\mathrm{E}_v|$ marked points, corresponding to the marked points and/or the nodes of $\mathcal{C}$; we index these points by the set $\vartheta^{-1}(v) \sqcup \mathrm{E}_v$ in the canonical way. Thus, as stacks,

$$\begin{aligned} Q_\Gamma &\approx \prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)\geq 3}} \overline{Q}_{0,|\vartheta^{-1}(v)|+|\mathrm{E}_v|}(\mathbb{P}^0, \mathfrak{d}(v)) \times \prod_{e\in\mathrm{Edg}} Q_{\Gamma_e} \\ &\approx \prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)\geq 3}} \overline{\mathcal{M}}_{0,|\vartheta^{-1}(v)|+|\mathrm{E}_v||\mathfrak{d}(v)}/\mathbb{S}_{\mathfrak{d}(v)} \times \prod_{e\in\mathrm{Edg}} Q_{\Gamma_e} \\ &\approx \Bigg(\prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)\geq 3}} \overline{\mathcal{M}}_{0,|\vartheta^{-1}(v)|+|\mathrm{E}_v||\mathfrak{d}(v)}/\mathbb{S}_{\mathfrak{d}(v)}\Bigg) \Big/ \prod_{e\in\mathrm{Edg}} \mathbb{Z}_{\mathfrak{d}(e)}, \end{aligned} \tag{7-5}$$

with each cyclic group $\mathbb{Z}_{\mathfrak{d}(e)}$ acting trivially. For example, in the case of the last diagram in Figure 1,

$$Q_\Gamma \approx \big(\overline{\mathcal{M}}_{0,2|d_0}/\mathbb{S}_{d_0} \times \overline{\mathcal{M}}_{0,2|5}/\mathbb{S}_5\big)\big/\mathbb{Z}_d \times \mathbb{Z}_{d'}$$

is a fixed locus in $\overline{Q}_{0,2}(\mathbb{P}^{n-1}, d_0+5+d+d')$. If $m' \leq m$ is as in Lemmas 6.5 and 6.6, the morphism $f_{m',m}$ in (6-5) sends the locus Q_Γ of $\overline{Q}_{0,m}(\mathbb{P}^{n-1}, d)$ to (a subset of) the locus $Q_{\Gamma_{m'}}$ of $\overline{Q}_{0,m'}(\mathbb{P}^{n-1}, d)$, where

$$\Gamma_{m'} = \big(\mathrm{Ver}_{m'}, \mathrm{Edg}_{m'}; \mu|_{\mathrm{Ver}_{m'}}, \mathfrak{d}_{m'}, \vartheta|_{[m']}\big),$$

as $f_{m',m}$ contracts the ends of the elements of $\overline{Q}_{0,m'}(\mathbb{P}^{n-1}, d)$ that do not carry any of the marked points indexed by the set $[m']$.

If $v \in \mathrm{Ver}$ and $\mathrm{val}(v) \geq 3$, for the purposes of definitions (4-4) and (4-5) we identify $[|\vartheta^{-1}(v)|+|\mathrm{E}_v|]$ with the set $\vartheta^{-1}(v) \sqcup \mathrm{E}_v$ indexing the marked points on $\mathcal{C}_v$ so that the element 1 in the former is identified with $1 \in [m]$ if $\vartheta(1) = v$ and with the unique edge $e_v^- = \{v_-, v\}$ with $v^- \prec v$ separating v from the marked point 1 otherwise. Similarly, if $v \preceq \vartheta(2)$, we associate the element 2 of $[|\vartheta^{-1}(v)|+|\mathrm{E}_v|]$ with $2 \in [m]$ if $\vartheta(2) = v$ and with the unique edge $e_v^+ = \{v, v_+\}$ with $v_+ \preceq \vartheta(2)$ separating v from the marked point 2 otherwise. Finally, if $m' \leq m$ is as in Lemmas 6.5 and 6.6 and $v \in \mathrm{Ver}_{m'}$, we associate the $|\mathrm{E}_v - \mathrm{Edg}_{m'}|$ largest elements of $[|\vartheta^{-1}(v)|+|\mathrm{E}_v|]$ with the subset $\mathrm{E}_v - \mathrm{Edg}_{m'}$ of $\vartheta^{-1}(v) \sqcup \mathrm{E}_v$.

If Γ is a decorated graph as above and $e = \{v_1, v_2\} \in \mathrm{Edg}$ with $v_1 \prec v_2$, let

$$\pi_e : Q_\Gamma \to Q_{\Gamma_e} \subset \overline{Q}_{0,2}(\mathbb{P}^{n-1}, \mathfrak{d}(e))$$

be the projection in the decomposition (7-5) and

$$\omega_{e;v_1} = -\pi_e^*\psi_1, \omega_{e;v_2} = -\pi_e^*\psi_2 \in H^2(Q_\Gamma).$$

Similarly, for each $v\in\mathrm{Ver}$ such that $\mathrm{val}(v)\ge 3$, let

$$\pi_v: Q_\Gamma \to \overline{\mathcal{M}}_{0,|\vartheta^{-1}(v)|+|\mathrm{E}_v||\mathfrak{d}(v)}/\mathbb{S}_{\mathfrak{d}(v)}$$

be the corresponding projection and

$$\psi_{v;e} = \pi_v^*\psi_e \in H^2(Q_\Gamma) \quad \text{for all } v\in\mathrm{E}_v.$$

By [Hori et al. 2003, Section 27.2],

$$\omega_{e;v_i} = \frac{\alpha_{\mu(v_i)}-\alpha_{\mu(v_{3-i})}}{\mathfrak{d}(e)} \qquad i=1,2. \tag{7-6}$$

By [Marian et al. 2011, Section 7.4], the Euler class of the normal bundle of Q_Γ in $\overline{Q}_{0,m}(\mathbb{P}^{n-1},|\Gamma|)$ is described by

$$\begin{aligned}\frac{\mathbf{e}(\mathcal{N}Q_\Gamma)}{\mathbf{e}(T_{\mu(v_{\min})}\mathbb{P}^{n-1})} &= \prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)\ge 3}}\prod_{k\neq\mu(v)}\pi_v^*\mathbf{e}\big(\dot{\mathcal{V}}_1^{(\mathfrak{d}(v))}(\alpha_{\mu(v)}-\alpha_k)\big)\prod_{e\in\mathrm{Edg}}\pi_e^*\mathbf{e}\big(H^0(f_e^*T\mathbb{P}^n\otimes\mathcal{O}(-y_1))/\mathbb{C}\big)\\ &\quad\times\prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)=2,\vartheta^{-1}(v)=\varnothing}}\Big(\sum_{e\in\mathrm{E}_v}\omega_{e;v}\Big)\prod_{\substack{v\in\mathrm{Ver}\\ \mathrm{val}(v)\ge 3}}\Big(\prod_{e\in\mathrm{E}_v}(\omega_{e;v}-\psi_{v;e})\Big),\end{aligned} \tag{7-7}$$

where $\mathbb{C}\subset H^0(f_e^*T\mathbb{P}^n\otimes\mathcal{O}(-y_1))$ denotes the trivial $\mathbb{T}$-representation. The terms on the first line correspond to the deformations of the sheaf without changing the domain, while the terms on the second line correspond to the deformations of the domain. By (6-6), (2-1), (4-4), and (4-5),

$$\begin{aligned}\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(|\Gamma|)})\big|_{Q_\Gamma} &= \prod_{\substack{v\in\mathrm{Ver}_{m'}\\ \mathrm{val}(v)\ge 3}}\pi_v^*\mathbf{e}\big(\dot{\mathcal{V}}_{\mathbf{a};r_{m';v}}^{(\mathfrak{d}(v))}(\alpha_{\mu(v)})\big)\cdot\prod_{e\in\mathrm{Edg}_{m'}}\pi_e^*\mathbf{e}\big(\dot{\mathcal{V}}_{n;\mathbf{a}}^{\mathfrak{d}(e)}\big),\\ \mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a};m'}^{(|\Gamma|)})\big|_{Q_\Gamma} &= \prod_{\substack{v\in\mathrm{Ver}_2\\ \mathrm{val}(v)\ge 3}}\pi_v^*\mathbf{e}\big(\ddot{\mathcal{V}}_{\mathbf{a};r_{m';v}}^{(\mathfrak{d}(v))}(\alpha_{\mu(v)})\big)\cdot\prod_{e\in\mathrm{Edg}_2}\pi_e^*\mathbf{e}\big(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{\mathfrak{d}(e)}\big)\\ &\quad\times\prod_{\substack{v\in\mathrm{Ver}_{m'}-\mathrm{Ver}_2\\ \mathrm{val}(v)\ge 3}}\pi_v^*\mathbf{e}\big(\dot{\mathcal{V}}_{\mathbf{a};r_{m';v}}^{(\mathfrak{d}(v))}(\alpha_{\mu(v)})\big)\cdot\prod_{e\in\mathrm{Edg}_{m'}-\mathrm{Edg}_2}\pi_e^*\mathbf{e}\big(\dot{\mathcal{V}}_{n;\mathbf{a}}^{\mathfrak{d}(e)}\big).\end{aligned} \tag{7-8}$$

By [Hori et al. 2003, Section 27.2], for all $e=\{v_1,v_2\}$ with $v_1\prec v_2$

$$\begin{aligned}\int_{Q_{\Gamma_e}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a}}^{\mathfrak{d}(e)})}{\mathbf{e}\big(H^0(f_e^*T\mathbb{P}^n\otimes\mathcal{O}(-y_1))/\mathbb{C}\big)} &= \dot{\mathfrak{C}}_{\mu(v_1)}^{\mu(v_2)}(\mathfrak{d}(e)),\\ \int_{Q_{\Gamma_e}}\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{\mathfrak{d}(e)})}{\mathbf{e}\big(H^0(f_e^*T\mathbb{P}^n\otimes\mathcal{O}(-y_1))/\mathbb{C}\big)} &= \ddot{\mathfrak{C}}_{\mu(v_1)}^{\mu(v_2)}(\mathfrak{d}(e)),\end{aligned} \tag{7-9}$$

with $\dot{\mathfrak{C}}^{\mu(v_2)}_{\mu(v_1)}(\mathfrak{d}(e))$ and $\ddot{\mathfrak{C}}^{\mu(v_2)}_{\mu(v_1)}(\mathfrak{d}(e))$ given by (5-1).

7B. ***Proof of Lemma 6.5.*** We apply the localization theorem to

$$\begin{aligned}\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}(\alpha_i,\hbar,q)&=\sum_{d=0}^{\infty}q^d\int_{\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_i}{\hbar-\psi_1}\prod_{j=2}^{m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j),\\ \ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}(\alpha_i,\hbar,q)&=\sum_{d=0}^{\infty}q^d\int_{\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_i}{\hbar-\psi_1}\prod_{j=2}^{m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j),\end{aligned}\tag{7-10}$$

where ϕ_i is the equivariant Poincaré dual of the fixed point $P_i\in\mathbb{P}^{n-1}$, as in (5-4), and the degree-0 terms in the $m=2$ case are defined by

$$\int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},0)}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_i}{\hbar-\psi_1}(\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2)\equiv(-\hbar)^{b_2}\varpi_2|_{P_i},$$
$$\int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},0)}\frac{\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_i}{\hbar-\psi_1}(\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2)\equiv(-\hbar)^{b_2}\varpi_2|_{P_i}.$$

Since $\phi_i|_{P_j}=0$ unless $j=i$, a decorated graph as in (7-3) contributes to the two expressions in (7-10) only if the first marked point is attached to a vertex labeled i, that is, $\mu(v_{\min})=i$ for the smallest element $v_{\min}\in\mathrm{Ver}$. We show that, just as for Givental's J-function, the (d,j)-summand in (6-2) with $C=\dot{\mathfrak{C}},\ddot{\mathfrak{C}}$ and $\mathcal{F}=\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'},\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$, that is,

$$\begin{aligned}&\frac{\dot{\mathfrak{C}}_i^j(d)q^d}{\hbar-(\alpha_j-\alpha_i/d)}\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}(\alpha_j,(\alpha_j-\alpha_i)/d,q)\quad\text{and}\\ &\frac{\ddot{\mathfrak{C}}_i^j(d)q^d}{\hbar-(\alpha_j-\alpha_i/d)}\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}(\alpha_j,(\alpha_j-\alpha_i)/d,q),\end{aligned}\tag{7-11}$$

respectively, is the sum over all graphs such that $\mu(v_{\min})=i$, that is, the first marked point is mapped to the fixed point $P_i\in\mathbb{P}^{n-1}$, $v_{\min}$ is a bivalent vertex, that is, $\mathfrak{d}(v_{\min})=0$, $\vartheta^{-1}(v_{\min})=\{1\}$, the only edge leaving this vertex is labeled d, and the other vertex of this edge is labeled j. We also show that the first sum on the right-hand side of (6-2) is the sum over all graphs such that $\mu(v_{\min})=i$ and $\mathrm{val}(v_{\min})\geq 3$.

If Γ is a decorated graph with $\mu(v_{\min})=i$ as above,

$$\mathrm{ev}_1^*\phi_i\big|_{Q_\Gamma}=\prod_{k\neq i}(\alpha_i-\alpha_k)=\mathbf{e}(T_{\mu(v_{\min})}\mathbb{P}^{n-1}).\tag{7-12}$$

Suppose in addition that $\mathrm{val}(v_{\min})=2$ and $|\mathrm{E}_{v_{\min}}|=1$. Let $v_1\equiv(v_{\min})_+$ be the immediate successor of $v_{\min}$ in Γ and $e_1=\{v_{\min},v_1\}$ be the edge leaving $v_{\min}$. If

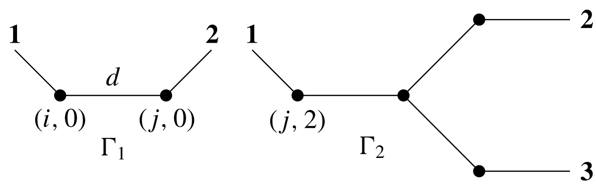


Figure 3. The two subgraphs of the second graph in Figure 1.

$|\mathrm{Edg}|>1$ or $\mathrm{val}(v_1)>2$, that is, Γ is not as in the first diagram in Figure 1, we break Γ at v_1 into two "subgraphs":

(i) $\Gamma_1=\Gamma_{e_1}$ consisting of the vertices $v_{\min}\prec v_1$, the edge $\{v_{\min},v_1\}$, with the $\mathfrak{d}$-value of 0 at both vertices, and a marked point at v and v_1;

(ii) Γ_2 consisting of all vertices, edges, and marked points of Γ, other than the vertex $v_{\min}$ and the edge $\{v_{\min},v_1\}$, and with the marked point 1 attached at v_1;

see Figure 3. By (7-5),

$$Q_\Gamma\approx Q_{\Gamma_1}\times Q_{\Gamma_2}. \tag{7-13}$$

Let $\pi_1,\pi_2:Q_\Gamma\to Q_{\Gamma_1},Q_{\Gamma_2}$ be the component projection maps. By (7-7) and (7-8),

$$\frac{\mathbf{e}(\mathcal{N}Q_\Gamma)}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}=\pi_1^*\bigg(\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_1})}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}\bigg)\cdot\pi_2^*\bigg(\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_2})}{\mathbf{e}(T_{P_{\mu(v_1)}}\mathbb{P}^{n-1})}\bigg)\cdot(\omega_{e;v_1}-\pi_2^*\psi_1),$$
$$\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};m'}\big)\big|_{Q_\Gamma}=\pi_1^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_1|)}_{n;\mathbf{a}}\big)\cdot\pi_2^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}\big),$$
$$\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};m'}\big)\big|_{Q_\Gamma}=\pi_1^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma_1|)}_{n;\mathbf{a}}\big)\cdot\pi_2^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}\big).$$

Combining the above splittings with (7-6), (7-9), and (7-12), we find that

$$\begin{aligned}&q^{|\Gamma|}\int_{Q_\Gamma}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_i}{\hbar-\psi_1}\prod_{j=2}^{j=m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)\Big|_{Q_\Gamma}\frac{1}{\mathbf{e}(\mathcal{N}Q_\Gamma)}\\&=\frac{\dot{\mathfrak{C}}_i^{\mu(v_1)}(\mathfrak{d}(e_1))q^{\mathfrak{d}(e_1)}}{\hbar-(\alpha_{\mu(v_1)}-\alpha_i)/\mathfrak{d}(e_1)}\\&\qquad\times\bigg(q^{|\Gamma_2|}\bigg\{\int_{Q_{\Gamma_2}}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'})\mathrm{ev}_1^*\phi_{\mu(v_1)}}{\hbar-\psi_1}\prod_{j=2}^{j=m}(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)\frac{1}{\mathbf{e}(\mathcal{N}Q_{\Gamma_2})}\bigg\}\bigg|_{\hbar=\frac{\alpha_{\mu(v_1)}-\alpha_i}{\mathfrak{d}(e_1)}}\bigg);\end{aligned}$$

the same identity holds with $\dot{\mathcal{V}}$ replaced by $\ddot{\mathcal{V}}$ and $\dot{\mathfrak{C}}_i^{\mu(v_1)}(\mathfrak{d}(e_1))$ by $\ddot{\mathfrak{C}}_i^{\mu(v_1)}(\mathfrak{d}(e_1))$. By the first equation in (7-10) with i replaced by $\mu(v_1)$ and the localization formula (7-1), the sum of the last factor above over all possibilities for Γ_2, with Γ_1 held fixed, is

$$\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}\big(\alpha_{\mu(v_1)}, (\alpha_{\mu(v_1)}-\alpha_i)/\mathfrak{d}(e_1), q\big) - \delta_{m,2}\bigg(\frac{\alpha_i-\alpha_{\mu(v_1)}}{\mathfrak{d}(e_1)}\bigg)^{b_2}\varpi_2|_{P_{\mu(v_1)}};$$

if $\dot{\mathcal{V}}$ is replaced by $\ddot{\mathcal{V}}$, then the sum becomes

$$\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}\big(\alpha_{\mu(v_1)}, (\alpha_{\mu(v_1)}-\alpha_i)/\mathfrak{d}(e_1), q\big) - \delta_{m,2}\bigg(\frac{\alpha_i-\alpha_{\mu(v_1)}}{\mathfrak{d}(e_1)}\bigg)^{b_2}\varpi_2|_{P_{\mu(v_1)}}.$$

In the $m=2$ case, the contributions of the one-edge graph $\Gamma_{i\mu(v_1)}(\mathfrak{d}(e_1))$ such as $\mathfrak{d}(v_1)=0$, as in the first diagram in Figure 1, to the two expressions in (7-10) are

$$\frac{\dot{\mathfrak{C}}^{\mu(v_1)}_i(\mathfrak{d}(e_1))q^{\mathfrak{d}(e_1)}}{\hbar_1-(\alpha_{\mu(v_1)}-\alpha_i)/\mathfrak{d}(e_1)}\bigg(\frac{\alpha_i-\alpha_{\mu(v_1)}}{\mathfrak{d}(e_1)}\bigg)^{b_2}\varpi_2|_{P_{\mu(v_1)}} \quad \text{and}$$
$$\frac{\ddot{\mathfrak{C}}^{\mu(v_1)}_i(\mathfrak{d}(e_1))q^{\mathfrak{d}(e_1)}}{\hbar_1-(\alpha_{\mu(v_1)}-\alpha_i)/\mathfrak{d}(e_1)}\bigg(\frac{\alpha_i-\alpha_{\mu(v_1)}}{\mathfrak{d}(e_1)}\bigg)^{b_2}\varpi_2|_{P_{\mu(v_1)}},$$

respectively. Thus, the contributions to the two expressions in (7-10) from all graphs Γ such that $\mathfrak{d}(v_{\min})=0$, $\mu(v_1)=j$, and $\mathfrak{d}(e_1)=d$ are given by (7-11), that is, they are the (d,j)-summands in the recursions (6-2) for $\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ and $\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$.

Suppose next that Γ is a graph such that $\mu(v_{\min})=i$ and $\mathrm{val}(v_{\min})\geq 3$. If $|\mathrm{Ver}|>1$, that is, Γ is not as in the first diagram in Figure 4, we break Γ at $v_{\min}$ into "subgraphs":

(i) Γ_0 consisting of the vertex $\{v_{\min}\}$ only, with the same μ and $\mathfrak{d}$-values as in Γ, with the same marked points as before, along with a marked point e for each edge $e\in \mathrm{E}_{v_{\min}}$ from $v_{\min}$;

(ii) for each $e\in\mathrm{E}_{v_{\min}}$, $\Gamma_{c;e}$ consisting of the branch of Γ beginning with the edge e at $v_{\min}$, with the $\mathfrak{d}$-value of $v_{\min}$ replaced by 0, and with one marked point at $v_{\min}$;

see Figure 4 and 8. By (7-5),

$$Q_\Gamma \approx Q_{\Gamma_0}\times\prod_{e\in\mathrm{E}_{v_{\min}}}Q_{\Gamma_{c;e}} = (\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}/\mathbb{S}_{\mathfrak{d}(v_{\min})})\times\prod_{e\in\mathrm{E}_{v_{\min}}}Q_{\Gamma_{c;e}}, \tag{7-14}$$
$$\text{where}\quad m_0=|\vartheta^{-1}(v_{\min})|+|\mathrm{E}_{v_{\min}}|.$$

Let $\pi_0,\pi_{c;e}$ be the component projection maps in (7-14). Since $\psi_1|_{Q_\Gamma}=\pi_0^*\psi_1$, $\mathbb{T}$ acts trivially on $\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}$,

$$\psi_1 = 1\times\psi_1\in H^*_{\mathbb{T}}(\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}) = H^*_{\mathbb{T}}\otimes H^*(\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}),$$

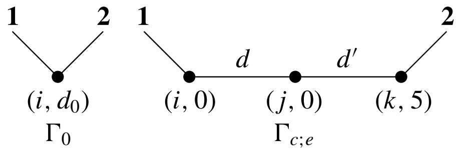


Figure 4. The two subgraphs of the last graph in Figure 1.

that is, $\mathbb{T}$ acts trivially on the universal cotangent line bundle for the first marked point on $\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}$, and the dimension of $\overline{\mathcal{M}}_{0,m_0|\mathfrak{d}(v_{\min})}$ is $m_0+\mathfrak{d}(v_{\min})-3$,

$$\frac{1}{\hbar-\psi_1}\Big|_{Q_\Gamma} = \sum_{r=0}^{m_0+\mathfrak{d}(v_{\min})-3} \hbar^{-(r+1)}\pi_0^*\psi_1^r.$$

Since $m_0+\mathfrak{d}(v_{\min}) \le m+|\Gamma|$ and Γ contributes to the coefficient of $q^{|\Gamma|}$ in (7-10), it follows that (6-2) holds with $\mathcal{F}$ replaced by $\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ and $\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ with $N_d = m+d-2$, $C_i^j(d) = \dot{\mathfrak{C}}_i^j(d)$ in the first case, and $C_i^j(d) = \ddot{\mathfrak{C}}_i^j(d)$ in the second case.

The argument in this section extends to products of projective spaces and concavex sheaves (1-13) as described in [Cooper and Zinger 2014, Section 6].

8. Polynomiality for stable quotients

In this section, we use the classical localization theorem [Atiyah and Bott 1984] to show that the generating functions $\hbar^{m-2}\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ and $\hbar^{m-2}\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ defined in (6-7) satisfy specific mutual polynomiality conditions of Definition 6.2 with respect to the generating function $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ defined in (3-3). The argument is similar to the proof in [Cooper and Zinger 2014, Section 7] of self-polynomiality for the generating function $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ defined in (3-3), but requires some modifications.

8A. ***Proof of Lemma 6.6.*** The proof involves applying the classical localization theorem [Atiyah and Bott 1984] with $(n+1)$-torus

$$\widetilde{\mathbb{T}} \equiv \mathbb{C}^* \times \mathbb{T},$$

where $\mathbb{T} = (\mathbb{C}^*)^n$ as before. We denote the weight of the standard action of the one-torus $\mathbb{C}^*$ on $\mathbb{C}$ by $\hbar$. Thus, by Section 3A,

$$H^*_{\mathbb{C}^*} \approx \mathbb{Q}[\hbar], \qquad H^*_{\widetilde{\mathbb{T}}} \approx \mathbb{Q}[\hbar, \alpha_1, \ldots, \alpha_n].$$

Throughout this section, $V = \mathbb{C}\oplus\mathbb{C}$ denotes the representation of $\mathbb{C}^*$ with the weights 0 and $-\hbar$. The induced action on $\mathbb{P}V$ has two fixed points:

$$q_1 \equiv [1, 0], \qquad q_2 \equiv [0, 1].$$

With $\gamma_1 \to \mathbb{P}V$ denoting the tautological line bundle,

$$\text{(8-1)} \qquad \mathbf{e}(\gamma_1^*)\big|_{q_1} = 0, \quad \mathbf{e}(\gamma_1^*)\big|_{q_2} = -\hbar, \quad \mathbf{e}(T_{q_1}\mathbb{P}V) = \hbar, \quad \mathbf{e}(T_{q_2}\mathbb{P}V) = -\hbar;$$

this follows from our definition of the weights in [Cooper and Zinger 2014, Section 3].

For each $d \in \mathbb{Z}^{\geq 0}$, the action of $\widetilde{\mathbb{T}}$ on $\mathbb{C}^n \otimes \mathrm{Sym}^d V^*$ induces an action on

$$\overline{\mathfrak{X}}_d \equiv \mathbb{P}(\mathbb{C}^n \otimes \mathrm{Sym}^d V^*).$$

It has $(d+1)n$ fixed points:

$$P_i(r) \equiv \big[\tilde{P}_i \otimes u^{d-r} v^r\big], \quad i \in [n], r \in \{0\} \cup [d],$$

if (u, v) are the standard coordinates on V and $\tilde{P}_i \in \mathbb{C}^n$ is the i-th coordinate vector (so that $[\tilde{P}_i] = P_i \in \mathbb{P}^{n-1}$). Let

$$\Omega \equiv \mathbf{e}(\gamma^*) \in H^*_{\widetilde{\mathbb{T}}}(\overline{\mathfrak{X}}_d)$$

denote the equivariant hyperplane class.

For all $i \in [n]$ and $r \in \{0\} \cup [d]$,

$$\Omega|_{P_i(r)} = \alpha_i + r\hbar, \quad \mathbf{e}(T_{P_i(r)}\overline{\mathfrak{X}}_d) = \bigg\{ \prod_{\substack{s=0 \\ (s,k)\neq(r,i)}}^{d} \prod_{k=1}^{n} (\Omega - \alpha_k - s\hbar) \bigg\}\bigg|_{\Omega=\alpha_i+r\hbar}. \tag{8-2}$$

Since

$$B\overline{\mathfrak{X}}_d = \mathbb{P}(B(\mathbb{C}^n \otimes \mathrm{Sym}^d V^*)) \to B\widetilde{\mathbb{T}} \quad \text{and}$$

$$c(B(\mathbb{C}^n \otimes \mathrm{Sym}^d V^*)) = \prod_{s=0}^{d} \prod_{k=1}^{n} (1 - (\alpha_k + s\hbar)) \in H^*(B\widetilde{\mathbb{T}}),$$

the $\widetilde{\mathbb{T}}$-equivariant cohomology of $\overline{\mathfrak{X}}_d$ is given by

$$\begin{aligned} H^*_{\widetilde{\mathbb{T}}}(\overline{\mathfrak{X}}_d) \equiv H^*(B\overline{\mathfrak{X}}_d) &= H^*(B\widetilde{\mathbb{T}})[\Omega] \Big/ \prod_{s=0}^{d} \prod_{k=1}^{n} (\Omega - (\alpha_k + s\hbar)) \\ &\approx \mathbb{Q}[\Omega, \hbar, \alpha_1, \dots, \alpha_n] \Big/ \prod_{s=0}^{d} \prod_{k=1}^{n} (\Omega - \alpha_k - s\hbar) \\ &\subset \mathbb{Q}_\alpha[\hbar, \Omega] \Big/ \prod_{s=0}^{d} \prod_{k=1}^{n} (\Omega - \alpha_k - s\hbar). \end{aligned}$$

In particular, every element of $H^*_{\widetilde{\mathbb{T}}}(\overline{\mathfrak{X}}_d)$ is a polynomial in Ω with coefficients in $\mathbb{Q}_\alpha[\hbar]$ of degree at most $(d+1)n - 1$.

For each $d \in \mathbb{Z}^{\geq 0}$, let

$$\mathfrak{X}'_d = \big\{ b \in \overline{Q}_{0,m}(\mathbb{P}V \times \mathbb{P}^{n-1}, (1, d)) : \mathrm{ev}_1(b) \in q_1 \times \mathbb{P}^{n-1}, \mathrm{ev}_2(b) \in q_2 \times \mathbb{P}^{n-1} \big\}. \tag{8-3}$$

A general element of b of $\mathfrak{X}'_d$ determines a morphism

$$(f,g):\mathbb{P}^1\to(\mathbb{P}V,\mathbb{P}^{n-1}),$$

up to an automorphism of the domain $\mathbb{P}^1$. Thus, the morphism

$$g\circ f^{-1}:\mathbb{P}V\to\mathbb{P}^{n-1}$$

is well defined and determines an element $\theta(b)\in\overline{\mathfrak{X}}_d$. By [Cooper and Zinger 2014, Section 7], this morphism extends to a $\widetilde{\mathbb{T}}$-equivariant morphism

$$\theta=\theta_d:\mathfrak{X}'_d\to\overline{\mathfrak{X}}_d.^{8}$$

If $d\in\mathbb{Z}^+$, there is also a natural forgetful morphism

$$F:\mathfrak{X}'_d\to\overline{Q}_{0,m}(\mathbb{P}^{n-1},d),$$

which drops the first sheaf in the pair and contracts one component of the domain if necessary. If in addition $m\ge m'\ge 2$, $f_{m',m}$ is as in (6-5), and $\mathcal{V}^{(d)}_{n;\mathbf{a}}$ is as in (1-3), let

$$\mathcal{V}^{(d)}_{n;\mathbf{a};m'}=f^*_{m',m}\mathcal{V}^{(d)}_{n;\mathbf{a}}\to\overline{Q}_{0,m}(\mathbb{P}^{n-1},d).$$

From the usual short exact sequence for the restriction along σ_1, we find that

$$\mathbf{e}(\mathcal{V}^{(d)}_{n;\mathbf{a};m'})=\langle\mathbf{a}\rangle\mathrm{ev}_1^*\mathbf{x}^{\ell(\mathbf{a})}\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\in H^*_{\mathbb{T}}(\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)).\tag{8-4}$$

In the case $d=0$, we set

$$F^*\mathbf{e}(\mathcal{V}^{(0)}_{n;\mathbf{a};m'})=\langle\mathbf{a}\rangle\mathrm{ev}_1^*(1\times\mathbf{x}^{\ell(\mathbf{a})})\in H^*\big(\overline{Q}_{0,m}(\mathbb{P}V\times\mathbb{P}^{n-1},(1,0))\big),$$
$$F^*\mathbf{e}(\ddot{\mathcal{V}}^{(0)}_{n;\mathbf{a};m'})=1\in H^*\big(\overline{Q}_{0,m}(\mathbb{P}V\times\mathbb{P}^{n-1},(1,0))\big).$$

Lemma 8.1. *Let $l\in\mathbb{Z}^{\ge0}$, $m,m',n\in\mathbb{Z}^+$ with $m\ge m'\ge2$, and $\mathbf{a}\in(\mathbb{Z}^*)^l$. With $\dot{\mathcal{Z}}_{n;\mathbf{a}}$, $\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$, $\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ as in (3-3) and (6-7),*

$$(-\hbar)^{m-2}\Phi^{\check{\eta}}_{\check{\mathcal{Z}}_{n;\mathbf{a};m'},\check{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}}(\hbar,z,q)=\sum_{d=0}^{\infty}q^d\int_{\mathfrak{X}'_d}\mathrm{e}^{(\theta^*\Omega)z}F^*\mathbf{e}(\check{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\,\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2\prod_{j=3}^{m}\psi_j^{b_j}\mathrm{ev}_j^*(\mathbf{e}(\gamma_1^*)\varpi_j).\tag{8-5}$$

with $(\check{\mathcal{Z}},\check{\mathcal{V}},\check{\eta})=(\dot{\mathcal{Z}},\mathcal{V},\dot{\eta}),(\ddot{\mathcal{Z}},\ddot{\mathcal{V}},\ddot{\eta})$.

Since the right-hand sides of the above expressions lie in $H^*_{\widetilde{\mathbb{T}}}[\![z,q]\!]\subset\mathbb{Q}_\alpha[\hbar][\![z,q]\!]$, this lemma is a more precise version of Lemma 6.6.

---

[8] This morphism is the composition of the morphism θ_d defined in [Cooper and Zinger 2014] in the $m=2$ case with the forgetful morphism

$$\overline{Q}_{0,m}(\mathbb{P}V\times\mathbb{P}^{n-1},(1,d))\to\overline{Q}_{0,2}(\mathbb{P}V\times\mathbb{P}^{n-1},(1,d)).$$

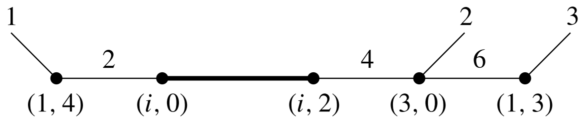


Figure 5. A graph representing a fixed locus in $\mathfrak{X}'_d$; $i \neq 1, 3$.

8B. ***Proof of Lemma 8.1.*** We apply the localization theorem of [Atiyah and Bott 1984] to the $\widetilde{\mathbb{T}}$-action on $\mathfrak{X}'_d$. We show that each fixed locus of the $\widetilde{\mathbb{T}}$-action on $\mathfrak{X}'_d$ contributing to the right-hand sides in (8-5) corresponds to a pair (Γ_1, Γ_2) of decorated graphs as in (7-3), with Γ_1 and Γ_2 contributing to the two generating functions in the subscript of the corresponding correlator Φ evaluated at $x = \alpha_i$ for some $i \in [n]$.

Similarly to Section 7, the fixed loci of the $\widetilde{\mathbb{T}}$-action on $\overline{Q}_{0,m}(\mathbb{P}V \times \mathbb{P}^{n-1}, (d', d))$ correspond to decorated graphs Γ with m marked points distributed between the ends of Γ. The map $\mathfrak{d}$ should now take values in pairs of nonnegative integers, indicating the degrees of the two subsheaves. The map μ should similarly take values in the pairs (i, j) with $i \in [2]$ and $j \in [n]$, indicating the fixed point (q_i, P_j) to which the vertex is mapped. The μ-values on consecutive vertices must differ by precisely one of the two components.

The situation for the $\widetilde{\mathbb{T}}$-action on

$$\mathfrak{X}'_d \subset \overline{Q}_{0,m}(\mathbb{P}V \times \mathbb{P}^{n-1}, (1, d))$$

is simpler, however. There is a unique edge of positive $\mathbb{P}V$-degree; we draw it as a thick line in Figure 5. The first component of the value of $\mathfrak{d}$ on all other edges and on all vertices must be 0; so we drop it. The first component of the value of μ on the vertices changes only when the thick edge is crossed. Thus, we drop the first components of the vertex labels as well, with the convention that these components are 1 on the left side of the thick edge and 2 on the right. In particular, the vertices to the left of the thick edge (including the left endpoint) lie in $q_1 \times \mathbb{P}^{n-1}$ and the vertices to its right lie in $q_2 \times \mathbb{P}^{n-1}$. Thus, by (8-3), the marked point 1 is attached to a vertex to the left of the thick edge and the marked point 2 is attached to a vertex to the right. By the localization formula (7-1) and the first equation in (8-1), Γ does not contribute to the right-hand sides in (8-5) unless the marked points indexed by $j \geq 3$ are also attached to vertices to the right of the thick edge. Finally, the remaining, second component of μ takes the same value $i \in [n]$ on the two vertices of the thick edge.

Let $\mathcal{A}_i$ denote the set of graphs as above so that the μ-value on the two endpoints of the thick edge is labeled i; see Figure 5. We break each graph $\Gamma \in \mathcal{A}_i$ into three subgraphs:

(i) Γ_1 consisting of all vertices of Γ to the left of the thick edge, including its left vertex v_1 with its $\mathfrak{d}$-value, and a new marked point attached to v_1;

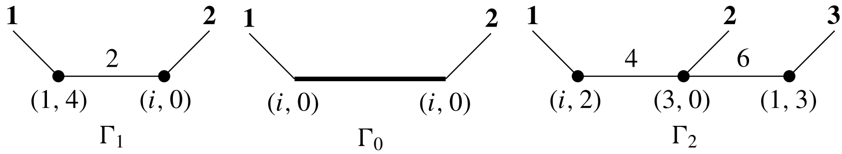


Figure 6. The three subgraphs of the graph in Figure 5.

(ii) Γ_0 consisting of the thick edge e_0, its two vertices v_1 and v_2, with $\mathfrak{d}$-values set to 0, and new marked points 1 and 2 attached to v_1 and v_2, respectively;

(iii) Γ_2 consisting of all vertices to the right of the thick edge, including its right vertex v_2 with its $\mathfrak{d}$-value, and a new marked point attached to v_2;

see Figure 6. From (7-5), we then obtain a splitting of the fixed locus in $\mathfrak{X}'_d$ corresponding to Γ:

(8-6)
$$Q_\Gamma \approx Q_{\Gamma_1} \times Q_{\Gamma_0} \times Q_{\Gamma_2} \subset \overline{Q}_{0,2}(\mathbb{P}^{n-1}, |\Gamma_1|) \times \overline{Q}_{0,2}(\mathbb{P}V, 1) \times \overline{Q}_{0,m}(\mathbb{P}^{n-1}, |\Gamma_2|).$$

The exceptional cases are $|\Gamma_1| = 0$ and $m = 2$, $|\Gamma_2| = 0$; the above isomorphism then holds with the corresponding component replaced by a point.

Let π_1, π_0, and π_2 denote the three component projection maps in (8-6). By (7-7),

(8-7)
$$\frac{\mathbf{e}(\mathcal{N}Q_\Gamma)}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})} = \pi_1^*\left(\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_1})}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}\right) \cdot \pi_2^*\left(\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_2})}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}\right) \cdot (\omega_{e_0;v_1} - \pi_1^*\psi_2)(\omega_{e_0;v_2} - \pi_2^*\psi_1).$$

Since for every $j = m'+1, \ldots, m$ the closest vertex of $\mathrm{Ver}_{m'}$ lies to the right of the thick edge, by (7-8) and (8-4),

(8-8)
$$\begin{aligned} F^*\mathbf{e}\big(\mathcal{V}^{(|\Gamma|)}_{n;\mathbf{a};m'}\big)\big|_{Q_\Gamma} &= \dot{\eta}(\alpha_i)\pi_1^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma_1|)}_{n;\mathbf{a}}\big)\pi_2^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}\big), \\ F^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};m'}\big)\big|_{Q_\Gamma} &= \ddot{\eta}(\alpha_i)\pi_1^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma_1|)}_{n;\mathbf{a}}\big)\pi_2^*\mathbf{e}\big(\ddot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}\big). \end{aligned}$$

Since Q_{Γ_0} consists of a degree-1 map, by the last two identities in (8-1)

(8-9)
$$\omega_{e_0;v_1} = \hbar, \quad \omega_{e_0;v_2} = -\hbar.$$

The morphism θ takes the locus Q_Γ to a fixed point $P_k(r) \in \overline{\mathfrak{X}}_d$. It is immediate that $k = i$. By continuity considerations, $r = |\Gamma_1|$. Thus, by the first identity in (8-2),

(8-10)
$$\theta^*\Omega\big|_{Q_\Gamma} = \alpha_i + |\Gamma_1|\hbar.$$

Combining (8-7)–(8-10) and the second equation in (8-1), we obtain

$$q^{|\Gamma|}\int_{Q_\Gamma}\frac{\mathrm{e}^{(\theta^*\Omega)z}F^*\mathbf{e}(\mathcal{V}_{n;\mathbf{a};m'}^{(|\Gamma|)})\,\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2\prod_{j=3}^m\psi_j^{b_j}\mathrm{ev}_j^*(\mathbf{e}(\gamma_1^*)\varpi_j)}{\mathbf{e}(\mathcal{N}Q_\Gamma)} \tag{8-11}$$
$$=\frac{(-\hbar)^{m-2}\dot\eta(\alpha_i)\mathrm{e}^{\alpha_i z}}{\prod_{k\neq i}(\alpha_i-\alpha_k)}\left\{\mathrm{e}^{|\Gamma_1|\hbar z}q^{|\Gamma_1|}\int_{Q_{\Gamma_1}}\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma_1|)})\mathrm{ev}_2^*\phi_i}{\hbar-\psi_2}\Big|_{Q_{\Gamma_1}}\frac{1}{\mathbf{e}(\mathcal{N}Q_{\Gamma_1})}\right\}$$
$$\times\left\{q^{|\Gamma_2|}\int_{Q_{\Gamma_2}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(|\Gamma_2|)})\mathrm{ev}_1^*\phi_i\prod_{j=2}^m(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)}{(-\hbar)-\psi_1}\Big|_{Q_{\Gamma_2}}\frac{1}{\mathbf{e}(\mathcal{N}Q_{\Gamma_2})}\right\}.$$

This identity remains valid with $|\Gamma_1|=0$ and/or $m=2$, $|\Gamma_2|=0$ if we set the corresponding integral to 1 or to $\hbar^{b_2}\varpi_2|_{P_i}$, respectively.

We now sum up the last identity over all $\Gamma\in\mathcal{A}_i$. This is the same as summing over all pairs (Γ_1,Γ_2) of decorated graphs such that

(1) Γ_1 is a 2-pointed graph of degree $d_1\geq 0$ such that the marked point 2 is attached to a vertex labeled i;

(2) Γ_2 is an m-pointed graph of degree $d_2\geq 0$ such that the marked point 1 is attached to a vertex labeled i.

By the localization formula (7-1) and symmetry,

$$1+\sum_{\Gamma_1}(q\mathrm{e}^{\hbar z})^{|\Gamma_1|}\left\{\int_{Q_{\Gamma_1}}\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma_1|)})\mathrm{ev}_2^*\phi_i}{(\hbar-\psi_2)\mathbf{e}(\mathcal{N}Q_{\Gamma_1})}\right\}$$
$$=1+\sum_{d=1}^\infty(q\mathrm{e}^{\hbar z})^d\int_{\overline{Q}_{0,2}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\ddot{\mathcal{V}}_{n;\mathbf{a}}^{(d)})\mathrm{ev}_2^*\phi_i}{\hbar-\psi_2}=\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q\mathrm{e}^{\hbar z});$$

$$\delta_{m,2}\hbar^{b_2}\varpi_2|_{P_i}+\sum_{\Gamma_2}q^{|\Gamma_2|}\left\{\int_{Q_{\Gamma_2}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(|\Gamma_2|)})\mathrm{ev}_1^*\phi_i\prod_{j=2}^m(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)}{\mathbf{e}(\mathcal{N}Q_{\Gamma_2})(-\hbar-\psi_1)}\right\}$$
$$=\delta_{m,2}\hbar^{b_2}\varpi_2|_{P_i}+\sum_{d=\max(3-m,0)}^\infty q^d\int_{\overline{Q}_{0,m}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a};m'}^{(|\Gamma_2|)})\mathrm{ev}_1^*\phi_i\prod_{j=2}^m(\psi_j^{b_j}\mathrm{ev}_j^*\varpi_j)}{(-\hbar-\psi_1)}$$
$$=\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\alpha_i,-\hbar,q).$$

Combining with this with (7-1), we obtain

$$\sum_{d=0}^\infty q^d\int_{\mathfrak{X}'_d}\mathrm{e}^{(\theta^*\Omega)z}F^*\mathbf{e}(\mathcal{V}_{n;\mathbf{a};m'}^{(d)})\,\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2\prod_{j=3}^m\psi_j^{b_j}\mathrm{ev}_j^*(\mathbf{e}(\gamma_1^*)\varpi_j)$$
$$=(-\hbar)^{m-2}\sum_{i=1}^n\frac{\dot\eta(\alpha_i)\mathrm{e}^{\alpha_i z}}{\prod_{k\neq i}(\alpha_i-\alpha_k)}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q\mathrm{e}^{\hbar z})\,\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}(\alpha_i,-\hbar,q)$$
$$=(-\hbar)^{m-2}\Phi^{\dot\eta}_{\dot{\mathcal{Z}}_{n;\mathbf{a}},\dot{\mathcal{Z}}_{n;\mathbf{a};m'}^{(\mathbf{b},\varpi)}}(\hbar,z,q),$$

as claimed in the $\dot{\mathcal{Z}}$ identity in (8-5).

From (8-7)–(8-10), we also find that (8-11) holds with $\mathcal{V}$ and $\dot{\mathcal{V}}$ replaced by $\ddot{\mathcal{V}}$ and $\dot{\eta}$ by $\ddot{\eta}$, with the same conventions in the $|\Gamma_1|=0$ and $m=2$, $|\Gamma_2|=0$ cases. We then sum up the resulting identity over all pairs (Γ_1,Γ_2) of decorated graphs as in the previous paragraph. The sum of the terms in the first curly brackets over all possibilities for Γ_1 is exactly the same as before, while the sum of the terms in the second curly brackets over all possibilities for Γ_2 is described by the same expression as before with $\dot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}$ and $\dot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$ replaced by $\ddot{\mathcal{V}}^{(|\Gamma_2|)}_{n;\mathbf{a};m'}$ and $\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}$, respectively. Thus,

$$\begin{aligned}\sum_{d=0}^{\infty} q^d \int_{\mathfrak{X}'_d} \mathrm{e}^{(\theta^*\Omega)z} F^*\mathbf{e}(\ddot{\mathcal{V}}^{(d)}_{n;\mathbf{a};m'})\,\psi_2^{b_2}\mathrm{ev}_2^*\varpi_2 \prod_{j=3}^{m} \psi_j^{b_j}\mathrm{ev}_j^*(\mathbf{e}(\gamma_1^*)\varpi_j)\\ = (-\hbar)^{m-2}\sum_{i=1}^{n} \frac{\ddot{\eta}(\alpha_i)\mathrm{e}^{\alpha_i z}}{\prod_{k\neq i}(\alpha_i-\alpha_k)} \dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q\mathrm{e}^{\hbar z})\,\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}(\alpha_i,-\hbar,q)\\ = (-\hbar)^{m-2}\Phi^{\ddot{\eta}}_{\dot{\mathcal{Z}}_{n;\mathbf{a}},\ddot{\mathcal{Z}}^{(\mathbf{b},\varpi)}_{n;\mathbf{a};m'}}(\hbar,z,q),\end{aligned}$$

as claimed in the $\ddot{\mathcal{Z}}$ identity in (8-5).

In the case of products of projective spaces and concavex sheaves (1-13), the spaces

$$\overline{Q}_{0,m}(\mathbb{P}V\times\mathbb{P}^{n-1},(1,d)) \quad\text{and}\quad \overline{\mathfrak{X}}_d=\mathbb{P}(\mathbb{C}^n\otimes\mathrm{Sym}^d V^*)$$

are replaced by

$$\begin{gathered}\overline{Q}_{0,m}(\mathbb{P}V\times\mathbb{P}^{n_1-1}\times\cdots\times\mathbb{P}^{n_p-1},(1,d_1,\dots,d_p)) \quad\text{and}\\ \mathbb{P}(\mathbb{C}^{n_1}\otimes\mathrm{Sym}^{d_1}V^*)\times\cdots\times\mathbb{P}(\mathbb{C}^{n_p}\otimes\mathrm{Sym}^{d_p}V^*),\end{gathered}$$

respectively. Lemma 8.1 extends to this situation by replacing z and q in (8-5) with $z_1,\dots,z_p$ and $q_1,\dots,q_p$, q^d with $q_1^{d_1}\cdots q_p^{d_p}$, $\mathfrak{X}'_d$ with $\mathfrak{X}'_{d_1,\dots,d_p}$, $\mathrm{e}^{(\theta^*\Omega z)}$ with $\mathrm{e}^{(\theta^*\Omega_1)z_1+\cdots+(\theta^*\Omega_p)z_p}$, and the indices d and n on the bundles $\mathcal{V}$, $\ddot{\mathcal{V}}$ with $(d_1,\dots,d_p)$ and $(n_1,\dots,n_p)$, and summing over $d_1,\dots,d_p\ge 0$ instead of $d\ge 0$. The vertices of the thick edge in Figure 5 are now labeled by a tuple $(i_1,\dots,i_p)$ with $i_s\in[n_s]$, as needed for the extension of Definition 6.2 described at the end of Section 6. The relation (8-10) becomes

$$\theta^*\Omega_s\big|_{Q_\Gamma}=\alpha_{s;i_s}+|\Gamma_1|_s\hbar,$$

where $|\Gamma_1|_s$ is the sum of the s-th components of the values of $\mathfrak{d}$ on the vertices and edges of Γ_1 (corresponding to the degree of the maps to $\mathbb{P}^{n_s-1}$). Otherwise, the proof is identical.

9. Stable quotients vs. Hurwitz numbers

Our proof of Propositions 4.1 and 4.2 that describe twisted Hurwitz numbers on $\overline{\mathcal{M}}_{0,3|d}$ is analogous to the proof of [Cooper and Zinger 2014, Theorem 4], which describes similar integrals on $\overline{\mathcal{M}}_{0,2|d}$. In particular, we show that it is sufficient to verify the statements of Propositions 4.1 and 4.2 for each fixed $\mathbf{a}$ and for all n sufficiently large (compared to $|\mathbf{a}|$). For $\nu_n(\mathbf{a}) > 0$, we obtain the statements of Propositions 4.1 and 4.2 by analyzing the secondary (middle) terms in the recursion (6-2) for the three-point generating functions $\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}$ and $\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}$ defined in (3-32) and (3-34), respectively. We also use (3-35) and (3-33). The latter is the string equation for stable quotients invariants; in Proposition 9.3, we show that it is equivalent to Proposition 4.2 whenever $\nu_n(\mathbf{a}) \geq 0$. In Proposition 9.2, we show that (3-33) is equivalent to Proposition 4.1 whenever $\nu_n(\mathbf{a}) \geq 0$. We confirm Proposition 4.1 whenever $\nu_n(\mathbf{a}) > 0$ using Proposition 6.3; see Corollary 9.1. Since it is sufficient to verify the statement of Proposition 4.1 with $\nu_n(\mathbf{a}) > 0$, the $\nu_n(\mathbf{a}) = 0$ case of Proposition 4.1 then concludes the proof of (3-33).

9A. ***Proof of Propositions 3.1, 4.1, and 4.2.*** With n and $\mathbf{a}$ as in Propositions 4.1 and 4.2 and $b_1, b_2, b_3, r \in \mathbb{Z}^{\geq 0}$, let

$$\mathcal{F}^{(b_1,b_2,b_3)}_{n;\mathbf{a}}(\alpha_i, q) = \sum_{d=0}^{\infty} \frac{q^d}{d!} \int_{\overline{\mathcal{M}}_{0,3|d}} \frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{\mathbf{a}}(\alpha_i))\psi_1^{b_1}\psi_2^{b_2}\psi_3^{b_3}}{\prod_{k\neq i} \mathbf{e}(\dot{\mathcal{V}}^{(d)}_1(\alpha_i - \alpha_k))},$$

$$\mathcal{F}^{(b_1,b_2,b_3)}_{n;\mathbf{a};r}(\alpha_i, q) = \sum_{d=0}^{\infty} \frac{q^d}{d!} \int_{\overline{\mathcal{M}}_{0,3|d}} \frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{\mathbf{a};r}(\alpha_i))\psi_1^{b_1}\psi_2^{b_2}\psi_3^{b_3}}{\prod_{k\neq i} \mathbf{e}(\dot{\mathcal{V}}^{(d)}_1(\alpha_i - \alpha_k))}.$$

By [Cooper and Zinger 2014, Remark 8.5],

$$\mathcal{F}^{(b_1,b_2,b_3)}_{n;\mathbf{a}}(\alpha_i, q) = \frac{\xi_{n;\mathbf{a}}(\alpha_i, q)^{b_1+b_2+b_3}}{b_1!\,b_2!\,b_3!} \mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i, q); \tag{9-1}$$

thus, it is sufficient to show that

$$\mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i, q) = \frac{1}{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i, q)}. \tag{9-2}$$

By the same reasoning as in [Cooper and Zinger 2014, Remarks 8.4, 8.5],

$$\mathcal{F}^{(b_1,b_2,b_3)}_{n;\mathbf{a};r}(\alpha_i, q) = \frac{\xi_{n;\mathbf{a}}(\alpha_i, q)^{b_1+b_2}}{b_1!\,b_2!} \mathcal{F}^{(0,0,b_3)}_{n;\mathbf{a};r}(\alpha_i, q);$$

thus, it is sufficient to show that

$$\sum_{b=0}^{\infty}\sum_{r=0}^{\infty} \mathcal{F}^{(0,0,b)}_{n;\mathbf{a};r}(\alpha_i, q) \mathop{\mathfrak{R}}_{\hbar=0}\left\{ \frac{(-1)^b}{\hbar^{b+1}} [\![\dot{\mathcal{Y}}_{n;\varnothing}(\alpha_i, \hbar, q)]\!]_{q;r} q^r \right\} = 1. \tag{9-3}$$

Corollary 9.1. *Let* $l \in \mathbb{Z}^{\geq 0}$, $n \in \mathbb{Z}^+$, *and* $\mathbf{a} \in (\mathbb{Z}^*)^l$. *If* $\nu_n(\mathbf{a}) > 0$,

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x}, \hbar, q) = \hbar^{-1}\, \dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \in H^*_{\mathbb{T}}(\mathbb{P}^{n-1})[\![\hbar^{-1}, q]\!].$$

Proof. By Lemma 6.4(ii) and Lemmas 6.5 and 6.6, the series $\hbar\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x}, \hbar, q)$ and $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)$ are $\mathfrak{C}$-recursive and satisfy the $\dot{\eta}$-MPC with respect to $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)$, no matter what n and $\mathbf{a}$ are. It is immediate that

$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \cong 1 \pmod{\hbar^{-1}}.$$

If $\nu_n(\mathbf{a}) > 0$ and $d \in \mathbb{Z}^+$,

$$\dim \overline{Q}_{0,3}(\mathbb{P}^{n-1}, d) - \mathrm{rk}\dot{\mathcal{V}}^{(d)}_{n;\mathbf{a};3} = \nu_n(\mathbf{a})d + (n-1) > n-1 = \dim \mathbb{P}^{n-1}.$$

Thus,

$$\hbar\, \dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x}, \hbar, q) \cong 1 \pmod{\hbar^{-1}},$$

whenever $\nu_n(\mathbf{a}) > 0$. The claim now follows from Proposition 6.3. □

Proposition 9.2. *If* $l \in \mathbb{Z}^{\geq 0}$, $n \in \mathbb{Z}^+$, *and* $\mathbf{a} \in (\mathbb{Z}^*)^l$ *are such that* $\nu_n(\mathbf{a}) \geq 0$, *then*

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x}, \hbar, q) = \hbar^{-1}\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q)}{\dot{I}_0(q)} \in (H^*_{\mathbb{T}}(\mathbb{P}^{n-1}))[\![\hbar^{-1}, q]\!] \tag{9-4}$$

if and only if (9-2) *holds for all* $i \in [n]$.

Proposition 9.3. *If* $l \in \mathbb{Z}^{\geq 0}$, $n \in \mathbb{Z}^+$, *and* $\mathbf{a} \in (\mathbb{Z}^*)^l$ *are such that* $\nu_n(\mathbf{a}) \geq n$, *then*

$$\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\mathbf{x}, \hbar, q) = \hbar^{-1}\, \dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x}, \hbar, q) \in (H^*_{\mathbb{T}}(\mathbb{P}^{n-1}))[\![\hbar^{-1}, q]\!] \tag{9-5}$$

if and only if (9-3) *holds for all* $i \in [n]$.

For any $t, t' \in [d]$ with $t \neq t'$, let $\Delta_{tt'} \in H^2(\overline{\mathcal{M}}_{0,m|d})$ denote the class of the diagonal divisor

$$\big\{[\mathcal{C}, y_1, \ldots, y_m; \hat{y}_1, \ldots, \hat{y}_d] \in \overline{\mathcal{M}}_{g,m|d} : \hat{y}_t = \hat{y}_{t'}\big\}.$$

For any $t \in [d]$, let

$$\Delta_t = \sum_{t'>t} \Delta_{tt'}.$$

We denote by $\mathfrak{s}_1, \mathfrak{s}_2, \ldots$ the elementary symmetric polynomials in

$$\{\beta_k\} = \big\{(\alpha_i - \alpha_k)^{-1} : k \neq i\big\}$$

for any given number of formal variables β_k. Let

$$A_{\mathbf{a}}(\alpha_i) = \prod_{a_k>0}(a_k^{a_k}\alpha_i^{a_k}) \prod_{a_k<0}(a_k^{-a_k}\alpha_i^{-a_k}), \qquad A_{n;\mathbf{a}}(\alpha_i) = \frac{A_{\mathbf{a}}(\alpha_i)}{\prod_{k\neq i}(\alpha_i - \alpha_k)}.$$

Proof of (9-2). By (1) in the proof of [Cooper and Zinger 2014, Proposition 8.3],

$$\begin{aligned}
\text{(9-6)}\quad \frac{\llbracket \mathcal{F}_{n;\mathbf{a}}^{(0,0,0)}(\alpha_i,q)\rrbracket_{q;d}}{A_{n;\mathbf{a}}^d(\alpha_i)}
&= \int_{\overline{\mathcal{M}}_{0,3|d}} \frac{\prod\limits_{a_k>0}\prod\limits_{t=1}^{d}\prod\limits_{\lambda=1}^{a_k}\Big(1-\frac{\lambda\hat\psi_t}{a_k\alpha_i}+\frac{\Delta_t}{\alpha_i}\Big)\prod\limits_{a_k<0}\prod\limits_{t=1}^{d}\prod\limits_{\lambda=0}^{-a_k-1}\Big(1+\frac{\lambda\hat\psi_t}{a_k\alpha_i}+\frac{\Delta_t}{\alpha_i}\Big)}{\prod\limits_{k\neq i}\prod\limits_{t=1}^{d}\Big(1-\frac{\hat\psi_t}{\alpha_i-\alpha_k}+\frac{\Delta_t}{\alpha_i-\alpha_k}\Big)}\\
&= \mathcal{H}_{\mathbf{a};d}(\alpha_i^{-1},\mathfrak{s}_1,\dots,\mathfrak{s}_d)
\end{aligned}$$

for some $\mathcal{H}_{\mathbf{a};d}\in\mathbb{Q}[y,\mathfrak{s}_1,\dots,\mathfrak{s}_d]$ dependent only on $\mathbf{a}$ and d, but not on n.[9] Similarly, for any $d,d'\in\mathbb{Z}^{\geq 0}$ there exists $\dot{\mathcal{Y}}_{\mathbf{a};d,d'}\in\mathbb{Q}[y,\mathfrak{s}_1,\dots,\mathfrak{s}_{d'}]$, independent of n, such that

$$\text{(9-7)}\qquad \llbracket \hbar^d\llbracket\dot{\mathcal{Y}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\rrbracket_{q;d}\rrbracket_{\hbar;d'} = A_{n;\mathbf{a}}^d(\alpha_i)\,\dot{\mathcal{Y}}_{\mathbf{a};d,d'}(y,\mathfrak{s}_1,\dots,\mathfrak{s}_{d'}).$$

Thus, by (4-9), there exist $\xi_{\mathbf{a};d},\dot\Phi_{\mathbf{a};d}^{(0)}\in\mathbb{Q}[y,\mathfrak{s}_1,\dots,\mathfrak{s}_d]$, independent of n, such that

$$\begin{aligned}
\llbracket\xi_{n;\mathbf{a}}(\alpha_i,q)\rrbracket_{q;d} &\equiv \mathfrak{R}_{\hbar=0}\llbracket\log\dot{\mathcal{Y}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\rrbracket_{q;d} = A_{n;\mathbf{a}}^d(\alpha_i)\xi_{\mathbf{a};d}(\alpha_i^{-1},\mathfrak{s}_1,\dots,\mathfrak{s}_{d-1}),\\
\llbracket\dot\Phi_{n;\mathbf{a}}^{(0)}(\alpha_i,q)\rrbracket_{q;d} &\equiv \mathfrak{R}_{\hbar=0}\frac{1}{\hbar}\llbracket \mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\dot{\mathcal{Y}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\rrbracket_{q;d}\\
&= A_{n;\mathbf{a}}^d(\alpha_i)\dot\Phi_{\mathbf{a};d}^{(0)}(\alpha_i^{-1},\mathfrak{s}_1,\dots,\mathfrak{s}_d).
\end{aligned}$$

We conclude that (9-2) is equivalent to

$$\sum_{\substack{d_1,d_2\geq 0\\ d_1+d_2=d}}\mathcal{H}_{\mathbf{a};d_1}\dot\Phi_{\mathbf{a};d_2}^{(0)}=\delta_{d,0}\quad\text{for all } d\in\mathbb{Z}^{\geq 0}.$$

By Corollary 9.1 and Proposition 9.2, these relations hold whenever $\nu_n(\mathbf{a})>0$; since they do not involve n, they thus hold for all pairs $(n,\mathbf{a})$. □

Proof of (9-3). For $t\in[d+1]$ and $r\in\mathbb{Z}^{\geq 0}$, we define $\hat\psi_t',\Delta_{t;r}'\in H^2(\overline{\mathfrak{M}}_{0,3|d})$ by

$$\hat\psi_t'=f_{2;3}^*\hat\psi_t,\qquad \Delta_{t;r}=f_{2;3}^*\Delta_t+\begin{cases}(r-1)f_{2;3}^*\Delta_{t,d+1} & \text{if } t\leq d;\\ 0 & \text{if } t=d+1.\end{cases}$$

[9] Whatever polynomial works for $n>d$ works for all n; this can be seen by setting the extra β_k's to 0.

Similarly to (1) in the proof of [Cooper and Zinger 2014, Proposition 8.3],

$$a_k>0 \;\Rightarrow\; \mathbf{e}(\dot{\mathcal{V}}^{(d)}_{a_k;r}(\alpha_i))=\prod_{t=1}^{d}\prod_{\lambda=1}^{a_k}(a_k\alpha_i-\lambda\hat{\psi}'_t+a_k\Delta'_{t;r})\cdot\prod_{\lambda=1}^{ra_k}(a_k\alpha_i-\lambda\hat{\psi}'_{d+1});$$
$$a_k<0 \;\Rightarrow\; \mathbf{e}(\dot{\mathcal{V}}^{(d)}_{a_k;r}(\alpha_i))=\prod_{t=1}^{d}\prod_{\lambda=0}^{-a_k-1}(a_k\alpha_i+\lambda\hat{\psi}'_t+a_k\Delta'_{t;r})\cdot\prod_{\lambda=0}^{-ra_k-1}(a_k\alpha_i+\lambda\hat{\psi}'_{d+1}).$$

Thus, similarly to (9-6),

$$\frac{[\![\mathcal{F}^{(0,0,b)}_{n;\mathbf{a};r}(\alpha_i,q)]\!]_{q;d}}{A_{n;\mathbf{a}}(\alpha_i)^d A_{\mathbf{a}}(\alpha_i)^r}=\mathcal{H}^{(b)}_{\mathbf{a};r;d}(\alpha_i^{-1},\mathfrak{s}_1,\dots,\mathfrak{s}_d)$$

for some $\mathcal{H}^{(b)}_{\mathbf{a};r;d}\in\mathbb{Q}[y,\mathfrak{s}_1,\dots,\mathfrak{s}_d]$ dependent only on $\mathbf{a}$, r, b, and d, but not on n. Thus, by (9-7) with $\mathbf{a}=\varnothing$, (9-3) is equivalent to

$$\sum_{\substack{d_1,d_2\ge 0\\ d_1+d_2=d}}\sum_{b=0}^{\infty}(-1)^b\mathcal{H}^{(b)}_{\mathbf{a};d_2;d_1}\dot{\mathcal{Y}}_{\varnothing;d_2,d_2+b}=\delta_{d,0}\quad\text{for all } d\in\mathbb{Z}^{\ge 0}.$$

By (3-35) and Proposition 9.3, these relations hold whenever $\nu_n(\mathbf{a})\ge 0$; since they do not involve n, they thus hold for all pairs $(n,\mathbf{a})$. □

9B. ***Proof of Proposition 9.2.*** We study the secondary (middle) terms in the recursions (6-2) for

$$\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\equiv\hbar^{-1}\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)}{\dot{I}_0(q)}\quad\text{and}\quad\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\mathbf{x},\hbar,q).$$

We show that (9-4) implies (9-2) by considering the $r=-1$ coefficients in these recursions. Conversely, if (9-2) holds, we show that the $r=-1$ coefficients in these recursions are described in the same degree-recursive way in terms of the corresponding power series; Proposition 6.3 and Lemma 6.5 then imply that $\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}=\widetilde{\mathcal{Z}}_{n;\mathbf{a}}$.[10]

By Lemmas 6.4 and 6.5,

(9-8)
$$\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)$$
$$=\sum_{d=0}^{\infty}\sum_{r=0}^{N_d-1}\{\dot{\mathcal{Z}}_{n;\mathbf{a}}\}^r_i(d)\hbar^{-r}q^d+\sum_{d=1}^{\infty}\sum_{j\neq i}\frac{\dot{\mathfrak{C}}^j_i(d)q^d}{\hbar-(\alpha_j-\alpha_i)/d}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_j,(\alpha_j-\alpha_i)/d,q),$$

[10] The same argument, with slightly more notation, can be used to show that all secondary coefficients are described in the same degree-recursive way, thus bypassing Proposition 6.3 and Lemma 6.5.

$$\begin{aligned}
&\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\\
&\quad=\sum_{d=0}^{\infty}\sum_{r=1}^{N_d}\{\widetilde{\mathcal{Z}}_{n;\mathbf{a}}\}_i^r(d)\hbar^{-r}q^d+\sum_{d=1}^{\infty}\sum_{j\neq i}\frac{\dot{\mathfrak{C}}_i^j(d)q^d}{\hbar-(\alpha_j-\alpha_i)/d}\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_j,(\alpha_j-\alpha_i)/d,q),
\end{aligned}$$

for some $N_d\in\mathbb{Z}^+$ and $\{\dot{\mathcal{Z}}_{n;\mathbf{a}}\}_i^r(d)$, $\{\widetilde{\mathcal{Z}}_{n;\mathbf{a}}\}_i^r(d)\in\mathbb{Q}_\alpha$. It is immediate that

$$\begin{aligned}
\dot{I}_0(q)\sum_{d=0}^{\infty}\{\widetilde{\mathcal{Z}}_{n;\mathbf{a}}\}_i^1(d)q^d-\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}_{n;\mathbf{a}}\}_i^0(d)q^d&\\
&=-\sum_{d=1}^{\infty}\sum_{j\neq i}\frac{\dot{\mathfrak{C}}_i^j(d)q^d}{(\alpha_j-\alpha_i)/d}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_j,(\alpha_j-\alpha_i)/d,q)\\
&=-\sum_{d=1}^{\infty}\sum_{j\neq i}\underset{\hbar=(\alpha_j-\alpha_i)/d}{\mathfrak{R}}\{\hbar^{-1}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\}\\
&=\underset{\hbar=0,\infty}{\mathfrak{R}}\{\hbar^{-1}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\}=\underset{\hbar=0}{\mathfrak{R}}\{\hbar^{-1}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\}-1;
\end{aligned}$$

the first and second equalities above follow from the first equation in (9-8), while the third from the residue theorem on $\mathbb{P}^1$ and (9-8) again, which implies that the coefficients of q^d in $\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)$ are regular in $\hbar$ away from $\hbar=(\alpha_j-\alpha_i)/d$ with $d\in\mathbb{Z}^+$ and $j\neq i$ and $\hbar=0,\infty$. Combining the last identity with the first statement in (3-12), and (4-9), we obtain

$$\sum_{d=0}^{\infty}\{\widetilde{\mathcal{Z}}_{n;\mathbf{a}}\}_i^1(d)q^d=\frac{\dot{\Phi}_{n;\mathbf{a}}^{(0)}(q)}{\dot{I}_0(q)^2}-\sum_{b=1}^{\infty}\frac{\xi_{n;\mathbf{a}}(q)^b}{b!}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{(-1)^b}{\hbar^b}\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\right\}. \tag{9-9}$$

By Lemma 6.5,

$$\begin{aligned}
&\dot{\mathcal{Z}}_{n;\mathbf{a};3}^{(\mathbf{0},\mathbf{1})}(\alpha_i,\hbar,q)\\
&\quad=\sum_{d=0}^{\infty}\sum_{r=1}^{N_d}\{\dot{\mathcal{Z}}_{n;\mathbf{a};3}^{(\mathbf{0},\mathbf{1})}\}_i^r(d)\hbar^{-r}q^d+\sum_{d=1}^{\infty}\sum_{j\neq i}\frac{\dot{\mathfrak{C}}_i^j(d)q^d}{\hbar-(\alpha_j-\alpha_i)/d}\dot{\mathcal{Z}}_{n;\mathbf{a};3}^{(\mathbf{0},\mathbf{1})}(\alpha_j,(\alpha_j-\alpha_i)/d,q),
\end{aligned}$$

for some $N_d\in\mathbb{Z}^+$ and $\{\dot{\mathcal{Z}}_{n;\mathbf{a};3}^{(\mathbf{0},\mathbf{1})}\}_i^r(d)\in\mathbb{Q}_\alpha$. By Section 7B, the secondary coefficients $\{\dot{\mathcal{Z}}_{n;\mathbf{a};3}^{(\mathbf{0},\mathbf{1})}\}_i^r(d)$ arise from the contributions of decorated graphs Γ as in (7-3) such that the vertex $v_{\min}$ to which the first marked point is attached is of valence 3 or higher. In this case, there are four types of such graphs, as shown in Figure 7:

(i) single-vertex graphs;

(ii) graphs with either marked point 2 or 3, but not both, attached to $v_{\min}$, that is, $|\vartheta^{-1}(v_{\min})|=2$;

(iii) graphs with two edges leaving $v_{\min}$, that is, $|\mathrm{E}_{v_{\min}}|=2$;

(iv) graphs with $|\vartheta^{-1}(v_{\min})|$, $|\mathrm{E}_{v_{\min}}|=1$, but $\mathfrak{d}(v_{\min})>0$.

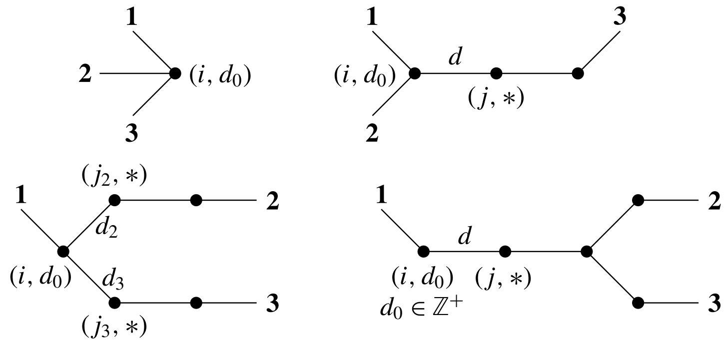


Figure 7. The four types of graphs determining the secondary coefficients $\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}^r_i(d)$.

By (7-7), (7-8), and (7-12), the contribution of the graphs of type (i) to the sum $\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}^1_i(d)q^d$ is

$$\sum_{d=0}^{\infty}\frac{q^d}{d!}\int_{\overline{\mathcal{M}}_{0,3|d}}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{\mathbf{a}}(\alpha_i))}{\prod_{k\neq i}\mathbf{e}(\dot{\mathcal{V}}^{(d)}_1(\alpha_i-\alpha_k))}=\mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i,q). \tag{9-10}$$

In the three remaining cases, we split each decorated graph Γ into subgraphs as on page 485; see Figure 8. Let π_0, $\pi_{c;e}$ denote the projection maps in the decomposition (7-14). By (7-7) and (7-8),

(9-11)

$$\frac{\mathbf{e}(\mathcal{N}Q_\Gamma)}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}=\prod_{k\neq i}\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_1(\alpha_i-\alpha_k)\big)\cdot\prod_{e\in\mathrm{E}_{v_{\min}}}\bigg(\pi^*_{c;e}\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_{c;e}})}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}(\omega_{e;v_{\min}}-\pi_0^*\psi_e)\bigg),$$

$$\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a}}\big)\big|_{Q_\Gamma}=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_{\mathbf{a}}(\alpha_i)\big)\cdot\prod_{e\in\mathrm{E}_{v_{\min}}}\pi^*_{c;e}\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_{c;e}|)}_{n;\mathbf{a}}\big).$$

Thus, the contribution of Γ to $\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}^1_i(d)q^d$ is

$$\begin{aligned} q^{|\Gamma|}\int_{Q_\Gamma}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a}})\mathrm{ev}_1^*\phi_i|_{Q_\Gamma}}{\mathbf{e}(\mathcal{N}Q_\Gamma)} &= \sum_{\mathbf{b}\in(\mathbb{Z}^{\geq 0})^{\mathrm{E}_{v_{\min}}}}\bigg(\frac{q^{d_0}}{d_0!}\int_{\overline{\mathcal{M}}_{0,m_0|d_0}}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d_0)}_{\mathbf{a}}(\alpha_i))\prod_{e\in\mathrm{E}_{v_{\min}}}\psi_e^{b_e}}{\prod_{k\neq i}\mathbf{e}(\dot{\mathcal{V}}^{(d_0)}_1(\alpha_i-\alpha_k))} \\ &\qquad\times\prod_{e\in\mathrm{E}_{v_{\min}}}q^{|\Gamma_{c;e}|}\omega_{e;v_{\min}}^{-(b_e+1)}\int_{Q_{\Gamma_{c;e}}}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(|\Gamma_{c;e}|)}_{n;\mathbf{a}})\mathrm{ev}_1^*\phi_i}{\mathbf{e}(\mathcal{N}Q_{\Gamma_{c;e}})}\bigg),\end{aligned} \tag{9-12}$$

where $m_0=|\vartheta^{-1}(v_{\min})|+|\mathrm{E}_{v_{\min}}|$ (which equals 3 if Γ is of type (ii) or (iii), or 2 if Γ is of type (iv) above) and $d_0=\mathfrak{d}(v_{\min})$.

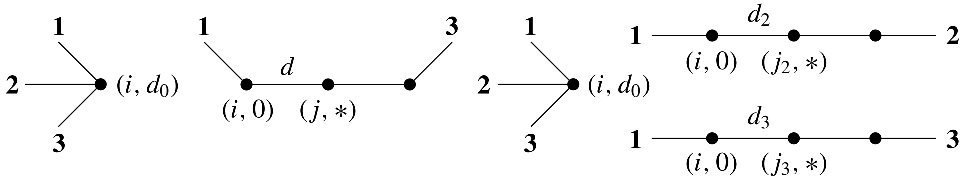


Figure 8. The subgraphs of the second and third graphs in Figure 7.

We now sum up (9-12) over all possibilities for Γ of each of the three types. For each $e \in \mathrm{E}_{v_{\min}}$, let $v_e \in \mathrm{Ver}$ denote the vertex of e other than $v_{\min}$. By (7-6) and Section 7B, the sum of the factor corresponding to $e \in \mathrm{E}_{v_{\min}}$ over all possibilities for Γ_e with $\mathfrak{d}(e) = d_e$ and $\mu(v_e) = j_e$ fixed is

$$(-1)^{b_e+1} \mathop{\mathfrak{R}}_{\hbar=(\alpha_{j_e}-\alpha_i/d_e)} \left\{ \hbar^{-(b_e+1)} \dot{\mathcal{Z}}(\alpha_i, \hbar, q) \right\},$$

where $\dot{\mathcal{Z}} = \dot{\mathcal{Z}}_{n;\mathbf{a}}$ in cases (ii) and (iii) and $\dot{\mathcal{Z}} = \dot{\mathcal{Z}}^{(\mathbf{1},\mathbf{0})}_{n;\mathbf{a};3}$ in case (iv). Thus, by the residue theorem on $\mathbb{P}^1$ and Lemma 6.5, the sum of the factors corresponding to $e \in \mathrm{E}_{v_{\min}}$ over all possibilities for Γ_e is

(9-13)
$$(-1)^{b_e} \mathop{\mathfrak{R}}_{\hbar=0,\infty} \left\{ \frac{\dot{\mathcal{Z}}(\alpha_i, \hbar, q)}{\hbar^{b_e+1}} \right\} = (-1)^{b_e} \mathop{\mathfrak{R}}_{\hbar=0} \left\{ \frac{\dot{\mathcal{Z}}(\alpha_i, \hbar, q)}{\hbar^{b_e+1}} \right\} - \begin{cases} \delta_{b_e,0} & \text{in cases (ii), (iii);} \\ 0 & \text{in case (iv).} \end{cases}$$

Combining (9-12) and (9-13) with (9-1), the first equation in (3-12), and (4-9), we find that the contribution to $\sum_{d=0}^{\infty} \{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}_i^1(d) q^d$ from all graphs Γ of types (ii) and (iii) above is given by

$$\begin{aligned} \mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i, q) \sum_{\mathbf{b} \in (\mathbb{Z}^{\ge 0})^{\mathrm{E}_{v_{\min}}}} \frac{(-\xi_{n;\mathbf{a}}(\alpha_i, q))^{|\mathbf{b}|}}{\mathbf{b}!} \prod_{e \in \mathrm{E}_{v_{\min}}} \left(\mathop{\mathfrak{R}}_{\hbar=0} \left\{ \frac{1}{\hbar^{b_e+1}} \dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i, \hbar, q) \right\} - \delta_{b_e,0} \right) \\ = \mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i, q) \left(\mathop{\mathfrak{R}}_{\hbar=0} \left\{ \frac{1}{\hbar} \mathrm{e}^{-\frac{\xi_{n;\mathbf{a}}(\alpha_i,q)}{\hbar}} \dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i, \hbar, q) \right\} - 1 \right)^{|\mathrm{E}_{v_{\min}}|} \\ = \mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i, q) \left(\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i, q)}{\dot{I}_0(q)} - 1 \right)^{|\mathrm{E}_{v_{\min}}|}, \end{aligned}$$

with $|\mathrm{E}_{v_{\min}}| = 1$ in (ii) and $|\mathrm{E}_{v_{\min}}| = 2$ in (iii). Using [Cooper and Zinger 2014, Theorem 4] instead of (9-1), we find that the contribution to $\sum_{d=0}^{\infty} \{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}_i^1(d) q^d$ from all graphs Γ of type (iv) above is given by

$$\begin{aligned} -\sum_{b=0}^{\infty} \frac{(-\xi_{n;\mathbf{a}}(\alpha_i, q))^{b+1}}{(b+1)!} \mathop{\mathfrak{R}}_{\hbar=0} \left\{ \frac{1}{\hbar^{b+1}} \dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\alpha_i, \hbar, q) \right\} \\ = -\sum_{b=1}^{\infty} \frac{\xi_{n;\mathbf{a}}(\alpha_i, q)^b}{b!} \mathop{\mathfrak{R}}_{\hbar=1} \left\{ \frac{(-1)^b}{\hbar^b} \dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\alpha_i, \hbar, q) \right\}. \end{aligned}$$

Putting this all together and taking into account that there are two flavors of type (ii) graphs, we conclude that

$$\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}^1_i(d)q^d=\mathcal{F}^{(0,0,0)}_{n;\mathbf{a}}(\alpha_i,q)\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)^2}{\dot{I}_0(q)^2} -\sum_{b=1}^{\infty}\frac{\xi_{n;\mathbf{a}}(\alpha_i,q)^b}{b!}\mathop{\mathfrak{R}}_{\hbar=0}\left\{\frac{(-1)^b}{\hbar^b}\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}(\alpha_i,\hbar,q)\right\}. \tag{9-14}$$

This is the same degree-recursive relation as (9-9) if and only if (9-2) holds.

9C. ***Proof of Proposition 9.3.*** We next apply the same argument to the power series

$$\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\equiv\hbar^{-1}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\mathbf{x},\hbar,q)\quad\text{and}\quad\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\mathbf{x},\hbar,q).$$

In this case, (9-9) becomes

$$\sum_{d=0}^{\infty}\{\widetilde{\mathcal{Z}}_{n;\mathbf{a}}\}^1_i(d)q^d=\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(q)}{\dot{I}_0(q)}-\sum_{b=1}^{\infty}\frac{\xi_{n;\mathbf{a}}(q)^b}{b!}\mathop{\mathfrak{R}}_{\hbar=0}\left\{\frac{(-1)^b}{\hbar^b}\widetilde{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)\right\}. \tag{9-15}$$

The graphs contributing to $\{\dot{\mathcal{Z}}_{n;\mathbf{a}}\}^r_i(d)$ are the same as before, as are the decomposition (7-14) and the first splitting in (9-11). However, the second splitting in (9-11) changes. For graphs Γ of type (i) and (ii) with $\vartheta(3)=v_{\min}$, it becomes

$$\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};2}\big)\big|_{Q_\Gamma}=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_{\mathbf{a};0}(\alpha_i)\big)\cdot\pi_{c;e}^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_{c;e}|)}_{n;\mathbf{a}}\big)$$

with the second factor being 1 for the graphs of type (i) and $e\in\mathrm{E}_{v_{\min}}$ denoting the unique element for the graphs of type (ii). For graphs of type (ii) with $\vartheta(2)=v_{\min}$, graphs of type (iii), and graphs of type (iv), it becomes

$$\begin{aligned}
\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};2}\big)\big|_{Q_\Gamma}&=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_{\mathbf{a};|\Gamma_{c;e}|}(\alpha_i)\big),\\
\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};2}\big)\big|_{Q_\Gamma}&=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_{\mathbf{a};|\Gamma_{c;e_3}|}(\alpha_i)\big)\cdot\pi_{c;e_2}^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_{c;e_2}|)}_{n;\mathbf{a}}\big),\\
\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma|)}_{n;\mathbf{a};2}\big)\big|_{Q_\Gamma}&=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_0|)}_{\mathbf{a}}(\alpha_i)\big)\cdot\pi_{c;e}^*\mathbf{e}\big(\dot{\mathcal{V}}^{(|\Gamma_{c;e}|)}_{n;\mathbf{a};2}\big),
\end{aligned}$$

respectively.

Thus, like (9-10), the contribution of the graphs of type (i) to $\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}\}^1_i(d)q^d$ is

$$\sum_{d=0}^{\infty}\frac{q^d}{d!}\int_{\overline{\mathcal{M}}_{0,3|d}}\frac{\mathbf{e}(\dot{\mathcal{V}}^{(d)}_{\mathbf{a};0}(\alpha_i))}{\prod_{k\neq i}\mathbf{e}(\dot{\mathcal{V}}^{(d)}_1(\alpha_i-\alpha_k))} =\sum_{b=0}^{\infty}\mathcal{F}^{(0,0,b)}_{n;\mathbf{a};0}(\alpha_i,q)\mathop{\mathfrak{R}}_{\hbar=0}\left\{\frac{(-1)^b}{\hbar^{b+1}}[\![\dot{\mathcal{Z}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;0}q^0\right\}.$$

Similarly to (9-13), the sum of the factor corresponding to an edge $e \in \mathrm{E}_{v_{\min}}$ in the analogue of (9-12) over all possibilities for Γ_e is

$$(-1)^{b_e}\begin{cases} \mathfrak{R}_{\hbar=0}\left\{\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)}{\hbar^{b_e+1}}\right\} - \delta_{b_e,0} & \text{in case (ii) with } \vartheta(3)=v_{\min}\text{, (iii) with } e=e_2; \\ \mathfrak{R}_{\hbar=0}\left\{\frac{\dot{\mathcal{Z}}_{n;\varnothing}(\alpha_i,\hbar,q)}{\hbar^{b_e+1}}\right\} - \delta_{b_e,0} & \text{in case (ii) with } \vartheta(2)=v_{\min}\text{, (iii) with } e=e_3; \\ \mathfrak{R}_{\hbar=0}\left\{\frac{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\alpha_i,\hbar,q)}{\hbar^{b_e+1}}\right\} & \text{in case (iv).} \end{cases}$$

Thus, the contribution to $\sum_{d=0}^{\infty}\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};3}\}_i^1(d)q^d$ from all graphs Γ of types (ii) with $\vartheta(3)=v_{\min}$ and $\vartheta(2)=v_{\min}$ is

$$\mathcal{F}^{(0,0,0)}_{n;\mathbf{a};0}(\alpha_i,q)\sum_{b=0}^{\infty}\frac{(-\xi_{n;\mathbf{a}}(\alpha_i,q))^b}{b!}\left(\mathfrak{R}_{\hbar=0}\left\{\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)}{\hbar^{b+1}}\right\} - \delta_{b,0}\right) = \left(\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)}{\dot{I}_0(q)} - 1\right)\mathcal{F}^{(0,0,0)}_{n;\mathbf{a};0}(\alpha_i,q)$$

and

$$\sum_{b=0}^{\infty}\sum_{r=1}^{\infty}\mathcal{F}^{(0,0,b)}_{n;\mathbf{a};r}(\alpha_i,q)\,\mathfrak{R}_{\hbar=0}\left\{\frac{(-1)^b}{\hbar^{b+1}}[\![\dot{\mathcal{Z}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;r}q^r\right\},$$

respectively. Similarly, the contribution from all graphs Γ of type (iii) is

$$\begin{aligned}\sum_{b_2,b_3\geq 0}^{\infty}\sum_{r=1}^{\infty}\mathcal{F}^{(0,0,b_3)}_{n;\mathbf{a};r}(\alpha_i,q)&\Bigg(\frac{(-\xi_{n;\mathbf{a}}(\alpha_i,q))^{b_2}}{b_2!}\left(\mathfrak{R}_{\hbar=0}\left\{\frac{\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\hbar,q)}{\hbar^{b_2+1}}\right\} - \delta_{b_2,0}\right) \\ &\qquad\times\mathfrak{R}_{\hbar=0}\left\{\frac{(-1)^{b_3}}{\hbar^{b_3+1}}[\![\dot{\mathcal{Z}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;r}q^r\right\}\Bigg) \\ = \left(\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)}{\dot{I}_0(q)} - 1\right)&\sum_{b=0}^{\infty}\sum_{r=1}^{\infty}\mathcal{F}^{(0,0,b)}_{n;\mathbf{a};r}(\alpha_i,q)\,\mathfrak{R}_{\hbar=0}\left\{\frac{(-1)^b}{\hbar^{b+1}}[\![\dot{\mathcal{Z}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;r}q^r\right\}.\end{aligned}$$

Finally, the contribution from all graphs Γ of type (iv) is given by

$$-\sum_{b=1}^{\infty}\frac{\xi_{n;\mathbf{a}}(\alpha_i,q)^b}{b!}\,\mathfrak{R}_{\hbar=1}\left\{\frac{(-1)^b}{\hbar^b}\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\alpha_i,\hbar,q)\right\}.$$

Putting this all together and using the first equation in (3-12), but now with $\mathbf{a}=\varnothing$ and thus $\dot{I}_0=1$, we conclude that

$$\begin{aligned}\sum_{d=0}^{\infty}&\{\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}\}^1_i(d)q^d\\ &=\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)}{\dot{I}_0(q)}\sum_{b=0}^{\infty}\sum_{r=0}^{\infty}\mathcal{F}^{(0,0,b)}_{n;\mathbf{a};r}(\alpha_i,q)\,\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{(-1)^b}{\hbar^{b+1}}[\![\dot{\mathcal{Y}}_{n;\varnothing}(\alpha_i,\hbar,q)]\!]_{q;r}q^r\right\}\\ &\qquad-\sum_{b=1}^{\infty}\frac{\xi_{n;\mathbf{a}}(\alpha_i,q)^b}{b!}\,\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{(-1)^b}{\hbar^b}\dot{\mathcal{Z}}^{(\mathbf{0},\mathbf{1})}_{n;\mathbf{a};2}(\alpha_i,\hbar,q)\right\}.\end{aligned}$$

This is the same degree-recursive relation as (9-15) if and only if (9-3) holds.

10. Proof of (3-14)

The equivariant cohomology of $\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}$ is given by

$$\begin{aligned}H^*_{\mathbb{T}}&(\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}\times\mathbb{P}^{n-1})\\ &=\mathbb{Q}[\alpha_1,\dots,\alpha_n,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3]\Big/\left\{\prod_{k=1}^n(\mathbf{x}_1-\alpha_k),\prod_{k=1}^n(\mathbf{x}_2-\alpha_k),\prod_{k=1}^n(\mathbf{x}_3-\alpha_k)\right\}.\end{aligned}$$

Thus, by the defining property of the cohomology pushforward [Zinger 2009, Equation (3.11)], the three-point power series $\dot{\mathcal{Z}}_{n;\mathbf{a}}$ in (3-3) is completely determined by the n^3 power series

(10-1) $$\begin{aligned}\dot{\mathcal{Z}}_{n;\mathbf{a}}&(\alpha_{i_1},\alpha_{i_2},\alpha_{i_3},\hbar_1,\hbar_2,\hbar_3,q)\\ &=\sum_{d=0}^{\infty}q^d\int_{\overline{Q}_{0,3}(\mathbb{P}^{n-1},d)}\frac{\mathbf{e}(\dot{\mathcal{V}}^d_{n;\mathbf{a}})\,\mathrm{ev}_1^*\phi_{i_1}\,\mathrm{ev}_2^*\phi_{i_2}\,\mathrm{ev}_3^*\phi_{i_3}}{(\hbar_1-\psi_1)(\hbar_2-\psi_2)(\hbar_3-\psi_3)}.\end{aligned}$$

The localization formula (7-1) reduces this expression to a sum over decorated trees as in Section 7. Each of these trees has a unique special vertex v_0: the vertex where the branches from the three marked points come together (one or more of the marked points may be attached to this vertex). We compute this sum by breaking each such tree Γ at v_0 into up to 4 "subgraphs":

(i) Γ_0 consisting of the vertex v_0 only, with 3 marked points and with the same μ and $\mathfrak{d}$-values as in Γ;

(ii) for each marked point $t=1,2,3$ of Γ with $\vartheta(t)\neq v_0$, Γ_t consisting of the branch of Γ running between the vertices $\vartheta(t)$ and v_0, with the $\mathfrak{d}$-value of v_0 replaced by 0 and with one new marked point attached to v_0;

see Figure 9. The contribution of the vertex graphs (i) is accounted for by the Hurwitz numbers of Proposition 4.1, while the contribution of each of the strands is accounted for by the SQ analogue of the double Givental's J-function computed by (3-11), (3-12), and (3-15). Putting these contributions together, we will obtain (3-14).

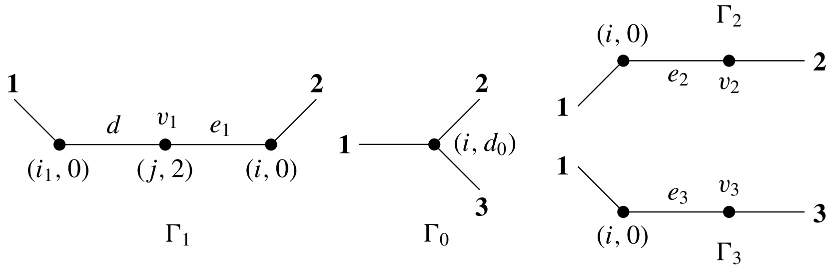


Figure 9. The four subgraphs of the second graph in Figure 1, with label i replaced by i_1.

Let $i=\mu(v_0)$ and $d_0=\mathfrak{d}(v_0)$. For each $t=1,2,3$ with $\vartheta(t)\neq v_0$, let $e_t=\{v_0,v_t\}$ be the edge leaving v_0 in the direction of $\vartheta(t)$. By (7-5),

$$Q_\Gamma \approx Q_{\Gamma_0}\times\prod_{t=1}^{3}Q_{\Gamma_t}=(\overline{\mathcal{M}}_{0,3|d_0}/\mathbb{S}_{d_0})\times\prod_{t=1}^{3}Q_{\Gamma_t}, \tag{10-2}$$

where the t-th factor is defined to be a point if $\vartheta(t)=v_0$. Let $\pi_0,\dots,\pi_3$ be the component projection maps in (10-2). By (7-7) and (7-8),

(10-3)

$$\frac{\mathbf{e}(\mathcal{N}Q_\Gamma)}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}=\prod_{k\neq i}\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}_1^{(d_0)}(\alpha_i-\alpha_k)\big)\cdot\prod_{t=1}^{3}\bigg(\pi_t^*\frac{\mathbf{e}(\mathcal{N}Q_{\Gamma_t})}{\mathbf{e}(T_{P_i}\mathbb{P}^{n-1})}(\omega_{e_t;v_0}-\pi_0^*\psi_t)\bigg),$$

$$\mathbf{e}\big(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma|)}\big)\big|_{Q_\Gamma}=\pi_0^*\mathbf{e}\big(\dot{\mathcal{V}}_{\mathbf{a}}^{(d_0)}(\alpha_i)\big)\cdot\prod_{t=1}^{3}\pi_t^*\mathbf{e}\big(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma_t|)}\big),$$

with the t-factor defined to be 1 if $\vartheta(t)=v_0$. Thus, the contribution of Γ to (10-1) is

(10-4)

$$\frac{1}{\prod_{k\neq i}(\alpha_i-\alpha_k)}\sum_{b_1,b_2,b_3\geq 0}\bigg(\frac{q^{d_0}}{d_0!}\int_{\overline{\mathcal{M}}_{0,3|d_0}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{\mathbf{a}}^{(d_0)}(\alpha_i))\prod_{t=1}^{3}\psi_t^{b_t}}{\prod_{k\neq i}\mathbf{e}(\dot{\mathcal{V}}_1^{(d_0)}(\alpha_i-\alpha_k))}$$
$$\times q^{|\Gamma_1|}\omega_{e_1;v_0}^{-(b_1+1)}\int_{Q_{\Gamma_1}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma_1|)})\mathrm{ev}_1^*\phi_{i_1}\mathrm{ev}_2^*\phi_i}{\mathbf{e}(\mathcal{N}Q_{\Gamma_1})(\hbar_1-\psi_1)}\prod_{t=2}^{3}q^{|\Gamma_t|}\omega_{e_t;v_0}^{-(b_t+1)}\int_{Q_{\Gamma_t}}\frac{\mathbf{e}(\dot{\mathcal{V}}_{n;\mathbf{a}}^{(|\Gamma_t|)})\mathrm{ev}_1^*\phi_i\mathrm{ev}_t^*\phi_{i_t}}{\mathbf{e}(\mathcal{N}Q_{\Gamma_t})(\hbar_t-\psi_t)}\bigg),$$

where the t-th factor on the second line is defined to be $\hbar_t^{-(b_t+1)}$ if $\vartheta(t)=v_0$.

We next sum up (10-4) over all possibilities for Γ. Let

$$\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)=\begin{cases}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_{i_1},\alpha_i,\hbar_1,\hbar,q) & \text{if } t=1;\\ \dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_i,\alpha_{i_t},\hbar,\hbar_t,q) & \text{if } t=2,3.\end{cases}$$

By (7-6) and Section 7B, the sum of the factor in (10-4) corresponding to each $t=1,2,3$ over all possibilities for Γ_t with $\mathfrak{d}(e_t)=d_t$ and $\mu(v_t)=j_t$ fixed is

$$(-1)^{b_t+1}\underset{\hbar=(\alpha_{j_t}-\alpha_i)/d_t}{\mathfrak{R}}\big\{\hbar^{-(b_t+1)}\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)\big\}.$$

Thus, by the residue theorem on $\mathbb{P}^1$ and Lemma 6.5, the sum of the factor in (10-4) corresponding to each $t=1,2,3$ over all possibilities for Γ_t nontrivial is

$$(-1)^{b_t}\underset{\hbar=0,\infty,-\hbar_t}{\mathfrak{R}}\left\{\frac{\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)}{\hbar^{b_t+1}}\right\} = (-1)^{b_t}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{\mathcal{Z}_i(\hbar,\alpha_{i_t},\hbar_t,q)}{\hbar^{b_t+1}}\right\}-\hbar_t^{-(b_t+1)}\prod_{k\neq i}(\alpha_{i_t}-\alpha_k).$$

Since the last term above is the contribution from the trivial subgraph Γ_t, the sum of the factor in (10-4) corresponding to each $t=1,2,3$ over all possibilities for Γ_t with $\mu(v_0)=i$ fixed is

$$\sum_{\Gamma_t}\big[t\text{-factor in (10-4)}\big]=(-1)^{b_t}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)}{\hbar^{b_t+1}}\right\}; \tag{10-5}$$

this takes into account the graphs Γ with $\vartheta(t)=i$.

By (10-4), (10-5), and Proposition 4.1,

$$\begin{aligned}\dot{\mathcal{Z}}_{n;\mathbf{a}}(\alpha_{i_1},\alpha_{i_2},\alpha_{i_3},\hbar_1,\hbar_2,\hbar_3,q) &\\ =\sum_{i=1}^{n}\frac{1}{\mathbf{s}_{n-1}(\alpha_i)\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)}&\prod_{t=1}^{3}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{1}{\hbar}\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)\right\}.\end{aligned}\tag{10-6}$$

By (3-11), (3-15), (3-12), and (4-9),

$$\begin{aligned}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{1}{\hbar}\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)\right\}=\sum_{\substack{s_t',s_t,r_t'\ge 0\\ s_t'+s_t+r_t'=n-1}}\Bigg((-1)^{r_t'}\mathbf{s}_{r_t'}&\\ \times\sum_{r_t''=0}^{s_t'}\widetilde{\mathcal{C}}^{(r_t'')}_{s_t'-\ell^-(\mathbf{a}),s_t'-r_t''-\ell^-(\mathbf{a})}(q)\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)L_{n;\mathbf{a}}(\alpha_i,q)^{s_t'-r_t''}}{I_0(q)\cdots I_{s_t'-r_t''}(q)}\ddot{\mathcal{Z}}^{(s_t)}_{n;\mathbf{a}}(\alpha_{i_t},\hbar_t,q)\Bigg)&\end{aligned}$$

for $t=2,3$. Combining this with (3-26), [Popa 2013, Proposition 4.4], and (2-16), we find that

$$\begin{aligned}\underset{\hbar=0}{\mathfrak{R}}\left\{\frac{1}{\hbar}\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_t},\hbar_t,q)\right\}&\\ =\sum_{s_t=0}^{n-1}\sum_{r_t=0}^{\hat{s}_t}\dot{\mathcal{C}}^{(r_t)}_{\hat{s}_t}(q)\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)L_{n;\mathbf{a}}(\alpha_i,q)^{\hat{s}_t-r_t}}{\ddot{\mathbb{I}}^{c}_{s_t+r_t}(q)}\ddot{\mathcal{Z}}^{(s_t)}_{n;\mathbf{a}}(\alpha_{i_t},\hbar_t,q)&\end{aligned}\tag{10-7}$$

for $t=2,3$. By the same reasoning,

$$\mathfrak{R}_{\hbar=0}\left\{\frac{1}{\hbar}\mathrm{e}^{-\xi_{n;\mathbf{a}}(\alpha_i,q)/\hbar}\,\dot{\mathcal{Z}}_i(\hbar,\alpha_{i_1},\hbar_1,q)\right\} = \sum_{s_1=0}^{n-1}\sum_{r_1=0}^{\hat{s}_1}\ddot{\mathcal{C}}^{(r_1)}_{\hat{s}_1}(q)\frac{\ddot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)L_{n;\mathbf{a}}(\alpha_i,q)^{\hat{s}_1-r_1}}{\dot{\mathbb{I}}^{c}_{s_1+r_1}(q)}\dot{\mathcal{Z}}^{(s_1)}_{n;\mathbf{a}}(\alpha_{i_1},\hbar_1,q), \tag{10-8}$$

where

$$\ddot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)=\left(\frac{L_{n;\mathbf{a}}(\alpha_i,q)}{\alpha_i}\right)^{-\ell(\mathbf{a})}\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q). \tag{10-9}$$

On the other hand, by (4-10) and (4-8),

$$\begin{aligned}\sum_{i=1}^{n}\frac{\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)^3L_{n;\mathbf{a}}(\alpha_i,q)^s}{\mathbf{s}_{n-1}(\alpha_i)\dot{\Phi}^{(0)}_{n;\mathbf{a}}(\alpha_i,q)}&\left(\frac{L_{n;\mathbf{a}}(\alpha_i,q)}{\alpha_i}\right)^{-\ell(\mathbf{a})}\\ &=\frac{1}{\mathbf{a}^{\mathbf{a}}}\sum_{i=1}^{n}L_{n;\mathbf{a}}(\alpha_i,q)^{s-|\mathbf{a}|}\frac{\mathrm{d}L}{\mathrm{d}q}\\ &=\frac{1}{\mathbf{a}^{\mathbf{a}}}\frac{\mathrm{d}}{\mathrm{d}q}\begin{cases}\ln\prod_{i=1}^{n}L_{n;\mathbf{a}}(\alpha_i,q) & \text{if } s=|\mathbf{a}|-1;\\ \frac{1}{s+1-|\mathbf{a}|}\sum_{i=1}^{n}L_{n;\mathbf{a}}(\alpha_i,q)^{s+1-|\mathbf{a}|} & \text{otherwise.}\end{cases}\end{aligned}$$

The collection $\{L_{n;\mathbf{a}}(\alpha_i,q)^{-1}\}$ is the set of n roots $\mathbf{y}$ of the equation

$$1-\mathbf{s}_1\mathbf{y}+\cdots+(-1)^n\mathbf{s}_n\mathbf{y}^n-\mathbf{a}^{\mathbf{a}}q\mathbf{y}^{\nu_n(\mathbf{a})}=0.$$

Thus, if $s\ge 0$ and $s+1<|\mathbf{a}|$,

$$\frac{\mathrm{d}}{\mathrm{d}q}\sum_{i=1}^{n}L_{n;\mathbf{a}}(\alpha_i,q)^{s+1-|\mathbf{a}|} =\frac{\mathrm{d}}{\mathrm{d}q}\mathcal{H}^{(|\mathbf{a}|-s-1)}\left(-\frac{\mathbf{s}_{n-1}}{\mathbf{s}_n},\frac{\mathbf{s}_{n-2}}{\mathbf{s}_n},\ldots,(-1)^{|\mathbf{a}|-s-1}\frac{\mathbf{s}_{\nu_n(\mathbf{a})+s+1}}{\mathbf{s}_n}\right)=0,$$

where $\mathcal{H}^{(r)}$ is as in (3-22). If $|\mathbf{a}|=n$, $\{L_{n;\mathbf{a}}(\alpha_i,q)\}$ is the set of n roots $\mathbf{y}$ of the equation

$$\mathbf{y}^n-(1-\mathbf{a}^{\mathbf{a}}q)^{-1}\mathbf{s}_1\mathbf{y}^{n-1}+(1-\mathbf{a}^{\mathbf{a}}q)^{-1}\mathbf{s}_2\mathbf{y}^{n-2}-\cdots+(-1)^n(1-\mathbf{a}^{\mathbf{a}}q)^{-1}\mathbf{s}_n=0.$$

Thus, if $s+1\le|\mathbf{a}|=n$,

$$\sum_{i=1}^{n}L_{n;\mathbf{a}}(\alpha_i,q)^{s-|\mathbf{a}|}\frac{\mathrm{d}L}{\mathrm{d}q}=\mathbf{a}^{\mathbf{a}}\mathcal{H}^{(s+1-n)}_{\nu_n(\mathbf{a})}(\mathbf{a}^{\mathbf{a}}q), \tag{10-10}$$

where $\mathcal{H}_{\nu}^{(r)}$ is as in (3-23). If $|\mathbf{a}|<n$, $\{L_{n;\mathbf{a}}(\alpha_i,q)\}$ is the set of n roots $\mathbf{y}$ of the equation

$$\mathbf{y}^n-\mathbf{s}_1\mathbf{y}^{n-1}+\cdots+(-1)^{\nu_n(\mathbf{a})-1}\mathbf{s}_{\nu_n(\mathbf{a})-1}\mathbf{y}^{|\mathbf{a}|+1}+(-1)^{\nu_n(\mathbf{a})}\big(\mathbf{s}_{\nu_n(\mathbf{a})}-(-1)^{\nu_n(\mathbf{a})}\mathbf{a}^{\mathbf{a}}q\big)\mathbf{y}^{|\mathbf{a}|} \\ +(-1)^{\nu_n(\mathbf{a})+1}\mathbf{s}_{\nu_n(\mathbf{a})+1}\mathbf{y}^{|\mathbf{a}|-1}+\cdots+(-1)^n\mathbf{s}_n=0.$$

Thus, if $s+1\le|\mathbf{a}|<n$, (10-10) still holds. Combining the equations in this paragraph, we find that

$$\sum_{i=1}^{n}\frac{\dot{\Phi}_{n;\mathbf{a}}^{(0)}(\alpha_i,q)^3L_{n;\mathbf{a}}(\alpha_i,q)^s}{\mathbf{s}_{n-1}(\alpha_i)\dot{\Phi}_{n;\mathbf{a}}^{(0)}(\alpha_i,q)}\left(\frac{L_{n;\mathbf{a}}(\alpha_i,q)}{\alpha_i}\right)^{-\ell(\mathbf{a})} \\ =\begin{cases}\mathcal{H}_{\nu_n(\mathbf{a})}^{(s+1-n)}(\mathbf{a}^{\mathbf{a}}q) & \text{if } s\ge n-1;\\ 0 & \text{if } 0\le s<n-1.\end{cases}$$

Combining this with (10-6)–(10-9) and (3-25), we obtain (3-14).

Acknowledgments

The author would like to thank Y. Cooper for many enlightening discussions concerning stable quotients invariants, R. Pandharipande for bringing moduli spaces of stable quotients to his attention, A. Popa for useful comments on a previous version of this paper, I. Ciocan-Fontanine for explaining [Ciocan-Fontanine and Kim 2013], and the referees for the detailed comments. He would also like to thank the School of Mathematics at IAS for its hospitality during the period when the results in this paper were obtained and the paper itself was completed. The author's research was partially supported by NSF grants DMS-0635607 and DMS-0846978 and the IAS Fund for Math.

References

[Artin 1991] M. Artin, *Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1991. MR 92g:00001 Zbl 0788.00001

[Atiyah and Bott 1984] M. F. Atiyah and R. Bott, "The moment map and equivariant cohomology", *Topology* **23**:1 (1984), 1–28. MR 85e:58041 Zbl 0521.58025

[Ciocan-Fontanine and Kim 2013] I. Ciocan-Fontanine and B. Kim, "Wall-crossing in genus zero quasimap theory and mirror maps", preprint, 2013. arXiv 1304.7056

[Ciocan-Fontanine et al. 2014] I. Ciocan-Fontanine, B. Kim, and D. Maulik, "Stable quasimaps to GIT quotients", *J. Geom. Phys.* **75** (2014), 17–47. MR 3126932 Zbl 06244048 arXiv 1106.3724

[Cooper and Zinger 2014] Y. Cooper and A. Zinger, "Mirror symmetry for stable quotients invariants", *Michigan Math. J.* **63**:3 (2014), 571–621. MR 3255692

[Elezi 2003] A. Elezi, "Mirror symmetry for concavex vector bundles on projective spaces", *Int. J. Math. Math. Sci.* 3 (2003), 159–197. MR 2004c:14106 Zbl 1011.14016

[Givental 1996] A. B. Givental, "Equivariant Gromov–Witten invariants", *Int. Math. Res. Notices* 13 (1996), 613–663. MR 97e:14015 Zbl 0881.55006

[Hori et al. 2003] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa, R. Vakil, and E. Zaslow, *Mirror symmetry*, Clay Mathematics Monographs **1**, American Mathematical Society, Providence, RI, 2003. MR 2004g:14042 Zbl 1044.14018

[Krattenthaler and Rivoal 2010] C. Krattenthaler and T. Rivoal, "On the integrality of the Taylor coefficients of mirror maps", *Duke Math. J.* **151**:2 (2010), 175–218. MR 2011d:11155 Zbl 1267.11077

[Lian et al. 1999] B. H. Lian, K. Liu, and S.-T. Yau, "Mirror principle, III", *Asian J. Math.* **3**:4 (1999), 771–800. MR 2002g:14080 Zbl 1009.14006

[Marian et al. 2011] A. Marian, D. Oprea, and R. Pandharipande, "The moduli space of stable quotients", *Geom. Topol.* **15**:3 (2011), 1651–1706. MR 2012h:14138 Zbl 1256.14057

[McDuff and Salamon 2004] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, American Mathematical Society Colloquium Publications **52**, American Mathematical Society, Providence, RI, 2004. MR 2004m:53154 Zbl 1064.53051

[Popa 2012] A. M. Popa, *Two-point Gromov–Witten formulas for symplectic toric manifolds*, thesis, University of New York, Stony Brook, 2012, Available at http://search.proquest.com/docview/1095122515. MR 3093933 arXiv 1206.2703

[Popa 2013] A. M. Popa, "The genus one Gromov–Witten invariants of Calabi–Yau complete intersections", *Trans. Amer. Math. Soc.* **365**:3 (2013), 1149–1181. MR 3003261 Zbl 1273.14116

[Popa and Zinger 2014] A. Popa and A. Zinger, "Mirror symmetry for closed, open, and unoriented Gromov–Witten invariants", *Adv. Math.* **259** (2014), 448–510. MR 3197664

[Voisin 1996] C. Voisin, "A mathematical proof of a formula of Aspinwall and Morrison", *Compositio Math.* **104**:2 (1996), 135–151. MR 98h:14069 Zbl 0951.14025

[Zinger 2009] A. Zinger, "Genus-zero two-point hyperplane integrals in the Gromov–Witten theory", *Comm. Anal. Geom.* **17**:5 (2009), 955–999. MR 2011c:14148 Zbl 1202.53085

[Zinger 2014] A. Zinger, "The genus 0 Gromov–Witten invariants of projective complete intersections", *Geom. Topol.* **18**:2 (2014), 1035–1114. MR 3190608

Received July 1, 2013. Revised February 11, 2014.

ALEKSEY ZINGER
DEPARTMENT OF MATHEMATICS
STONY BROOK UNIVERSITY
STONY BROOK, NY 11794-3651
UNITED STATES

azinger@math.sunysb.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.509

ACKNOWLEDGEMENT

The editors gratefully acknowledge the valuable advice of the referees who helped them select and better the papers appearing in 2014 in the *Pacific Journal of Mathematics* (reports dated November 12, 2013 through November 27, 2014).

Christoph Aisleitner, Alma L. Albujer, Shun-ichi Amari, Pablo Amster, Julio Andrade, Daniele Angella, Annamalai Anguraj, Anne-Marie Aubert, Irfan Bagci, Chengming Bai, Samuel A. Ballas, Dan Barbasch, Pilar Bayer, Jonathan Beck, Matthias Beck, Steven Bell, Misha Bialy, Joan S. Birman, Marek Biskup, Indranil Biswas, Don Blasius, Harold Philip Boas, Philippe Bolle, Mario Bonk, Brian H. Bowditch, Charles P. Boyer, Tara Elise Brendle, Ronald M. Brown, Aslak Bakke Buan, Michelle Bucher, Ashay A. Burungale, Esther Cabezas-Rivas, James W. Cannon, Huai-Dong Cao, Pierre-Emmanuel Caprace, Ana Caraiani, Claudio Carmeli, Valentina Casarino, Sagun Chanillo, Xiao-Li Chao, Vyjayanthi Chari, Moira Chas, Bo-yong Chen, Jan Cholewa, Martin Chuaqui, Tobias H. Colding, Charles H. Conley, Yves Cornulier, Alessio Corti, Balázs Csikós, Guizhen Cui, Marius Dadarlat, H. Garth Dales, Panagiota Daskalopoulos, Christopher William Davis, Matthew B. Day, Fusheng Deng, Shaoqiang Deng, B. N Dharmendra, Hongjie Dong, Sorin Dragomir, Todd Drumm, Semyon Dyatlov, Heather A. Dye, Ken Dykema, Richard Escobales, José M. Espinar, John Etnyre, Elisha Falbel, Hao Fang, Gavril Farkas, Maksym Fedorchuk, Alex Feingold, Yuval Zvi Flicker, William J. Floyd, Johanna Franklin, Stefan Friedl, Yu Fu, Nicola Fusco, V. Futorny, Anton Galaev, Giovanni P. Galdi, Wee Liang Gan, Yun Gao, Eduardo García-Río, Jianquan Ge, Yuxin Ge, Alfred Geroldinger, Paolo Ghiggini, Sudhir R. Ghorpade, Stefan Gille, Hiroshi Goda, David Goldberg, Cameron Gordon, Joshua Evan Greene, Sophie Grivaux, Elmar Große-Klönne, John Groves, Caroline Gruson, Robert Guralnick, P. Habegger, Farshid Hajir, Jeffrey Hakim, Ian Hambleton, Fengbo Hang, Kathryn E. Hare, Roger Heath-Brown, Matthew Hedden, David Hemmer, Sebastian Hensel, Jürgen Herzog, Kaoru Hiraga, Ilan Hirshberg, Gerald Hoehn, Ko Honda, Matias del Hoyo, Zejun Hu, Wen Huang, Xiaojun Huang, Alan Huckleberry, Patrick J. Iglesias-Zemmour, Jun-ichi Inoguchi, Oliver Jenkinson, David Jerison, Mattias Jonsson, Dominic Joyce, Boldizsar Kalmar, Vitali Kapovitch, Hendrik Kasten, Ely Kerman, Ju-Lee Kim, Kang-Tae Kim, Young Ho Kim, Alexander Kleshchev, Krzysztof Klosin, Andrew Knightly, Kevin P. Knudson, Dessislava H. Kochloukova, Leonid V. Kovalev, Ermanno Lanconelli, Urs Lang, Loredana Lanzani, Michael Larsen, Eike Lau, Christine Laurent-Thiébaut, Remi Leclercq, David B. Leep, Dimitry Leites, Romuald Lenczewski, Man Chun Leung, Gilbert Levitt, Chi Li, Fang Li, Jiayu Li, Ke-Pao Lin, Tao Li, Wei-Ping Li, Wen-Wei Li, Xiangdong Li, Yong Lin, Yuxiang Li, Yevgeniy Liokumovich, Kefeng Liu, José Ferrán Valdez Lorenzo, Jeremy Lovejoy, Richard Loy, Jean Ludwig, Yong Luo, Gennady Lyubeznik, Li Ma, Xiaonan Ma, Fares Maalouf, Yoshiaki Maeda, Romanos Malikiosis, Paolo Maremonti, Geoffrey Mason, Daniel V. Mathews, Carl Mautner, James McCoy, John McCuan, Eckhard Meinrenken, Alexander Merkurjev, Ralf Meyer, Enzo Luigi Mitidieri, Chikashi Miyazaki, Colette Mœglin, Susan Montgomery, Scott Morrison, Raymond Mortini, Paul Muhly, Ovidiu Munteanu, Jun Murakami, Alexey Muranov, Frédéric Naud, Andrés Navas, E. Neher, Sergey Neshveyev, Lenhard Ng, Bogdan Nica, Shin-ichi Ohta, Cezar Oniciuc, Dragos Oprea, Ye-Lin Ou, Valentin Ovsienko, Igor Pak, Dick Palais, Shengliang Pan, Pekka Pankka, Vassilis J. Papantoniou, Jae Keol Park, John R. Parker, Frédéric Paulin, Oscar M. Perdomo, Konstantinas Pileckas, Detlev Poguntke, Sorin Popa, T. Porter, Barbara Przebieracz, Patrizia Pucci, Zhenbo Qin, Jie Qing, Claudiu Raicu, Martin Raum, David A. Redett, Mark Reeder, Simeon Reich, Yudong Ren, Frédéric Robert, Richard Rochberg, Vered Rom-Kedar, Alfonso Romero, André Ronveaux, Min Ru, Daniel Ruberman, Dmitry V. Rutsky, Steen Ryom-Hansen, Mikael Rørdam, N. Saradha, Nikolai Saveliev, Renate Scheidler, Thomas Schick, Felix Schulze, Rasul Shafikov, Shuanglin Shao, A. R. Shastri, Thomas Richard Shemanske, Yi-bing Shen, Kenichi Shimizu, Tom Sideris, Joseph Silverman, Larry Smith, Frank Sottile, Viktor Starkov, Greg Stevenson, Peter Stiller, Stephan Stolz, Jeff Stopple, Jeffrey Streets, Michael Struwe, Pawel Strzelecki, Yucai Su, László Székelyhidi, Zizhou Tang, Hossein Teimoori, Chuu-Lian Terng, Abigail A. Thompson, Ulrike Tillmann, Adriano Tomassini, Vladimir Touraev, Martin Traizet, Tom Trotter, Ngô Việt Trung, Thomas Unger, Alexander Vasiliev, Joeri van der Veken, Luis Verde-Star, Thomas Vogel, Nolan Wallach, Gerard Walschap, Bixiang Wang, Jiaping Wang, Xu-Jia Wang, Yi Wang, Yong Wang, Youde Wang, Stephan Wehrli, Guofang Wei, Jun-cheng Wei, Ben Weinkove, Chris Wendl, Scott O. Wilson, Wilhelm Winter, Robert A. Wolak, Bun Wong, Jie Wu, William Wylie, Jie Xiao, Zhizhang Xie, Yuanlong Xin, Hao Xu, Xingwang Xu, Ergün Yalçin, Paul Yang, Xiaokui Yang, Xiangdong Ye, Wei Yuan, Tudor Zamfirescu, Yunus Zeytuncu, Lei Zhang, Qi S. Zhang, Bin Zhao, Entao Zhao, Kaiming Zhao, Lifeng Zhao, Shengmao Zhu, Roger Zierau, Wolfgang Ziller, Kang Zuo.

CONTENTS

Volume 272, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 272 No. 2 December 2014
