Pacific Journal of Mathematics

Volume 265 No. 2 October 2013

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 2013 is US $400/year for the electronic version, and $485/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 the 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. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.257

SINGULARITY REMOVABILITY AT BRANCH POINTS FOR WILLMORE SURFACES

YANN BERNARD AND TRISTAN RIVIÈRE

We consider a branched Willmore surface immersed in $\mathbb{R}^{m\geq 3}$ with square-integrable second fundamental form. We develop around each branch point local asymptotic expansions for the Willmore immersion, its first, and its second derivatives. Our expansions are given in terms of new integer-valued residues which are computed as circulation integrals around the branch point. We deduce explicit "point removability" conditions guaranteeing that the immersion is smooth through the branch point. These conditions are new, even in codimension one.

1. Introduction

1A. *Preliminaries.* Let $\vec{\Phi}$ be an immersion from a closed abstract two-dimensional manifold Σ into $\mathbb{R}^{m\geq 3}$. We denote by $g := \vec{\Phi}^* g_{\mathbb{R}^m}$ the pullback by $\vec{\Phi}$ of the flat canonical metric $g_{\mathbb{R}^m}$ of $\mathbb{R}^m$, also called the *first fundamental form of* $\vec{\Phi}$, and we let $d\mathrm{vol}_g$ be its associated *volume form*. The *Gauss map* of the immersion $\vec{\Phi}$ is the map taking values in the Grassmannian of oriented $(m-2)$-planes in $\mathbb{R}^m$ given by

$$\vec{n} := \star \frac{\partial_{x_1}\vec{\Phi} \wedge \partial_{x_2}\vec{\Phi}}{|\partial_{x_1}\vec{\Phi} \wedge \partial_{x_2}\vec{\Phi}|},$$

where $\star$ is the usual Hodge star operator in the Euclidean metric, and $\{x_1, x_2\}$ are local coordinates on the surface Σ.

Denoting by $\pi_{\vec{n}}$ the orthonormal projection of vectors in $\mathbb{R}^m$ onto the $(m-2)$-plane given by $\vec{n}$, the *second fundamental form* may be expressed as

$$\vec{\mathbb{I}}_p(X, Y) := \pi_{\vec{n}}\, d^2\vec{\Phi}(X, Y) \quad \text{for all } X, Y \in T_p\Sigma.$$

(In order to define $d^2\vec{\Phi}(X, Y)$ one has to extend locally around $T_p\Sigma$ the vector

Bernard is supported by the DFG Collaborative Research Center SFB/Transregio 71 (Project B3). Parts of this work were completed during his visits to the welcoming facilities of the Forschungsinstitut für Mathematik at the ETH in Zurich.

MSC2010: primary 30C70, 35J35, 35J50, 35J48, 35R01; secondary 49Q10, 53A30, 32S25, 58E15, 58E30.

fields X and Y. It is not difficult to check that $\pi_{\vec{n}}\, d^2\vec{\Phi}(X,Y)$ is independent of this extension.)

The *mean curvature vector* of the immersion at the point $p \in \Sigma$ is

$$\vec{H} := \tfrac{1}{2}\operatorname{Tr}_g(\vec{\mathbb{I}}_p) = \tfrac{1}{2}\big[\vec{\mathbb{I}}_p(\vec{e}_1,\vec{e}_1) + \vec{\mathbb{I}}_p(\vec{e}_2,\vec{e}_2)\big],$$

where $\{\vec{e}_1,\vec{e}_2\}$ is an orthonormal basis of $T_p\Sigma$ for the metric g.

In the present paper, we study the functional

$$W(\vec{\Phi}) := \int_\Sigma |\vec{H}|^2\, d\mathrm{vol}_g,$$

called *Willmore energy*. It has been extensively studied in the literature, due to its relevance to various areas of science. We refer the reader to [Rivière 2010] and the references therein for more extensive information on the properties and applications of the Willmore energy.

The Gauss–Bonnet theorem and Gauss equation imply that

$$W(\vec{\Phi}) = \frac{1}{4}\int_\Sigma |\vec{\mathbb{I}}|_g^2\, d\mathrm{vol}_g + \pi\chi(\Sigma) = \frac{1}{4}\int_\Sigma |d\vec{n}|_g^2\, d\mathrm{vol}_g + \pi\chi(\Sigma),$$

where $\chi(\Sigma)$ is the Euler characteristic of Σ, which is a topological invariant for a closed surface. From the variational point of view, the critical points of the Willmore functional, called *Willmore surfaces*, are thus also critical points of the Dirichlet energy of the Gauss map with respect to the induced metric g.

Minimal surfaces[1] are examples of Willmore surfaces. Not only is the Willmore energy invariant under reparametrization of the domain, but, more remarkably, it is invariant under Möbius transformations of $\mathbb{R}^m \cup \{\infty\}$; namely,

$$W(\Xi\circ\vec{\Phi}) = W(\vec{\Phi}) \quad \text{for any conformal diffeomorphism } \Xi \text{ of } \mathbb{R}^m\cup\{\infty\}.$$

Hence, the image of a Willmore immersion by a conformal transformation is again a Willmore immersion. It is thus no surprise that the class of Willmore immersions is considerably larger than that of minimal immersions (whose minimality is *not* preserved through conformal diffeomorphism).

An important task in the analysis of Willmore surfaces is to understand the closure of the space of Willmore immersions. Because the conformal group of transformations of $\mathbb{R}^m$ is not compact, one cannot expect the space of Willmore immersions to be closed in the strong C^l-topology. However, locally, in isothermic coordinates,[2] under some universal energy threshold, and as long as the conformal

[1] Minimal surfaces satisfy $\vec{H} = \vec{0}$ and are hence absolute minimizers of W.

[2] Analogously to other gauge-invariant problems, such as in Yang–Mills theory, isothermic coordinates (i.e., conformal parametrizations) provide the optimal *symmetry-breaking method*. A detailed discussion of this topic is available in [Rivière 2010].

parameter λ of the induced metric g is controlled in L^∞, the immersion is uniformly bounded in any C^l-norm. More precisely, the following *ε-regularity* result holds.

Theorem 1.1 [Rivière 2008]. *There exists* $\varepsilon_0 \equiv \varepsilon_0(m) > 0$ *such that, for any Willmore conformal immersion* $\vec{\Phi} : B_1 \to \mathbb{R}^m$ *satisfying*

$$\int_{B_1} |\nabla \vec{n}|^2 \, dx < \varepsilon_0,$$

and for any $l \in \mathbb{N}^*$, *we have*

$$\|e^{-\lambda} \nabla^l \vec{\Phi}\|^2_{L^\infty(B_{1/2})} \le C_l \left(\int_{B_1} |\nabla \vec{n}|^2 \, dx + 1 \right),$$

where C_l *only depends on* l, *while* λ *denotes the conformal parameter of* $\vec{\Phi}$. *Namely,* $\lambda = \|\log |\partial_{x_1} \vec{\Phi}|\|_{L^\infty(B_1)} = \|\log |\partial_{x_2} \vec{\Phi}|\|_{L^\infty(B_1)}$.

This theorem leads to the *concentration of compactness* "dialectic" developed by Sacks and Uhlenbeck. In a conformal parametrization, assuming that the conformal factor is L^∞-controlled in some subdomain of Σ, a sequence of Willmore immersions might fail to converge strongly in C^l only at finitely many isolated points, namely, at those points where the $W^{1,2}$-norm of the Gauss map concentrates. Assuming their induced metric generates a sequence of conformal classes which remains within a compact subdomain of the moduli space of Σ, the control of the conformal factor of a sequence of conformal immersions with uniformly bounded Willmore energy is also guaranteed, except again at those isolated points. This fact is established in [Rivière 2013], and it ultimately follows from the works [Toro 1995; Müller and Šverák 1995; Hélein 1996] on immersions with totally bounded curvature.

In this context, it appears natural to consider a branched Willmore immersion and study its local behavior near the point singularities.[3] In particular, we shall seek conditions that ensure the *removability* of the branch points.

In this paper, a *branch point* is a point where the immersion $\vec{\Phi}$ degenerates in the sense that $d\vec{\Phi}$ vanishes at that point. We focus on (conformal) locally Lipschitz and $W^{2,2}$ immersions $\vec{\Phi} : D^2 \setminus \{0\} \to \mathbb{R}^m$ with a branch point at the origin 0, and regular away from the origin. *A priori*, at a branch point, the mean curvature is singular. We will show that $\vec{\Phi}$, and thus the mean curvature, is actually smooth through a branch point, provided a certain set of sharp conditions are satisfied. The density $\theta_0 \in \mathbb{N}^*$ of the current $\vec{\Phi}_*[D^2]$ is called the *order* of the branch point. We shall use the words *branch point* and *singularity* interchangeably. This is of course an abuse of language, as the immersion $\vec{\Phi}$ is not singular at a branch point. In

[3] Such point singularities also naturally occur as blow-ups of the Willmore flow.

fact, both $\vec{\Phi}$ and $d\vec{\Phi}$ are well-defined there. It is the immersive nature of $\vec{\Phi}$ which degenerates at a branch point.

More generally, let Σ be an open surface, and suppose that $f:\Sigma\to B^m_R(0)\setminus\{0\}$, for some $R>0$, is a smooth proper Willmore immersion. We define the associated two-varifold

$$\mu := \big[x\mapsto \mathscr{H}^0(f^{-1}(x))\big]\mathscr{H}^2 \llcorner f(\Sigma),$$

and suppose that

$$0\in \operatorname{spt}(\mu),\quad \theta^2_*(\mu,0)<\infty,\quad \int_\Sigma |\mathbb{I}|^2_g\,d\mathrm{vol}_g<\infty.$$

It is shown in [Kuwert and Schätzle 2007] that $\theta^2(\mu,0)\in\mathbb{N}$ exists, and that $\operatorname{spt}(\mu)$ is a smooth, possibly multivalued graph, over some planes in $B^m_\rho(0)\setminus B^m_{\rho/2}(0)$ for some $\rho>0$. As we seek to understand the local behavior of our surface near the origin, we assume that there is exactly one graph of integer-multiplicity $\theta^2(\mu,0)\geq\theta_0\geq 1$. We can then switch to the parametric formulation used above and throughout this paper. A celebrated inequality of [Li and Yau 1982] for varifolds with compact support gives

$$\theta^2(\mu,0):=\lim_{r\searrow 0}\frac{\mu(B^m_r(0))}{\pi r^2}\leq\frac{1}{4\pi}W(\Sigma). \tag{1-1}$$

Accordingly, studying surfaces with a high-order branch point amounts to doing away with hypotheses demanding low upper bounds on the Willmore energy (such as the assumption $W(\Sigma)<8\pi$ in [Kuwert and Schätzle 2004]).

In the context of this paper, the word *removability* is to be understood with care. To say that a branch point is removable does not mean that it is the result of some "parametric illusion". Rather, it means that the map $\vec{\Phi}$ is smooth through the branch point, although it continues to fail to be an immersion at that point. In particular, the mean curvature, which is naturally singular at a branch point, turns out to be regular at a *removable* branch point. For instance, in this sense, the branched immersion $\vec{\Phi}:x\mapsto(x^2,x^3)$ has a removable branch point at the origin. In particular, the corresponding Gauss map which identifies to the $\mathbb{CP}^1$-projection of $\partial_x\vec{\Phi}$: $\vec{n}(x):=[2x,3x^2]\equiv[2,3x]$ is clearly smooth through the origin.

In [Bernard and Rivière 2011a], we delve deeper into the analysis of sequences of Willmore surfaces with uniformly bounded energy and nondegenerating conformal type. The results of the present paper play an important role there.

1B. ***Main results.*** Kuwert and Schätzle [2004] initiated the analytical study of point singularities of Willmore immersions by first considering unit-density singularities in codimension 1. They were able to find some removability criterion (extended in [Rivière 2008] to arbitrary codimension). Unfortunately, the energy restrictions necessary to ensure that the singularities occurring have unit-density are quite stringent (namely, one must assume the immersion has Willmore energy strictly below 8π). Still in codimension 1, Kuwert and Schätzle [2007] studied singularities of higher order, thereby allowing less stringent bounds on the energy. This time, however, no removability condition was found.

In the present work, we bridge the gaps left by previous studies. We work in arbitrary codimension and impose no restriction on the Willmore energy bound (i.e., we allow the order of the branch point to be arbitrarily large — although finite). Even in this general setting, we are able to find point-removability conditions.

Working in arbitrary codimension goes beyond the mere technical prowess. Willmore surfaces immersed into $\mathbb{R}^m$ for $m > 3$ are far from being devoid of interest, and the case $m = 4$ is particularly useful in geometry, as seen in [Burstall et al. 2002; Ejiri 1988; Montiel 2000].

While in codimension 1 techniques have been developed and used, analytical results in higher codimension require a suitable reformulation of the problem. Briefly speaking, the codimension-1 case involves a fourth-order nonlinear scalar equation, whereas, in higher codimension, one faces a *strongly coupled* fourth-order nonlinear *system* of equations. For this reason, we adopt a radically different approach to the problem, by using the original framework devised in [Rivière 2008]. In particular, working in a *conformal parametrization*, the aforementioned system is recast into a single equation in divergence form for the mean curvature *vector*. The analytical efficiency and benefits of this method have come to fruition in [Rivière 2008; 2013; Bernard and Rivière 2011a; 2011b].

Our main goal is twofold. Firstly, we study the regularity of the Gauss map, and we develop precise asymptotics for the immersion and the mean curvature near that point. Secondly, we bring into light explicit conditions ensuring the removability of the point singularity.

We assume that the point singularity lies at the origin, and we localize the problem by considering a map $\vec{\Phi} : D^2 \to \mathbb{R}^{m\geq 3}$, which is an immersion of $D^2 \setminus \{0\}$, and satisfying

(i) $\vec{\Phi} \in C^0(D^2) \cap C^\infty(D^2 \setminus \{0\})$;

(ii) $\mathcal{H}^2(\vec{\Phi}(D^2)) < \infty$;

(iii) $\int_{D^2} |\vec{\mathbb{I}}|_g^2 \, d\mathrm{vol}_g < \infty$.

As explained in [Rivière 2010], under the above assumptions, using the moving frame method of Chern and Hélein, one can construct a Lipschitz diffeomorphism f of the disk such that $\vec{\Phi} \circ f$ is conformal (an analogous procedure based on the work Müller and Sverak is presented in [Kuwert and Schätzle 2007]). We shall abusively continue to denote this reparametrization by $\vec{\Phi}$. It has properties (i)–(iii), and, moreover,

$$\vec{\Phi}(0)=\vec{0} \quad \text{and} \quad \vec{\Phi}(D^2)\subset B^m_R(0) \quad \text{for some } 0<R<\infty.$$

Hence, $\vec{\Phi}\in W^{1,\infty}_{\mathrm{loc}}\cap W^{2,2}(D^2\setminus\{0\})$. Away from the origin, we define the *Gauss map* $\vec{n}$ via

$$\vec{n}=\star\frac{\partial_{x_1}\vec{\Phi}\wedge\partial_{x_2}\vec{\Phi}}{|\partial_{x_1}\vec{\Phi}\wedge\partial_{x_2}\vec{\Phi}|},$$

where $\{x_1,x_2\}$ are standard Cartesian coordinates on the unit disk D^2, and $\star$ is the Euclidean Hodge-star operator. The immersion $\vec{\Phi}$ is conformal; i.e.,

$$|\partial_{x_1}\vec{\Phi}|=\mathrm{e}^{\lambda}=|\partial_{x_2}\vec{\Phi}| \quad \text{and} \quad \partial_{x_1}\vec{\Phi}\cdot\partial_{x_2}\vec{\Phi}=0, \tag{1-2}$$

where λ is the conformal parameter. An elementary computation shows that

$$d\mathrm{vol}_g=\mathrm{e}^{2\lambda}\,dx \quad \text{and} \quad |\nabla\vec{n}|^2\,dx=|d\vec{n}|^2_g\,d\mathrm{vol}_g=|\vec{\mathbb{I}}|^2_g\,d\mathrm{vol}_g. \tag{1-3}$$

Hence, by hypothesis, we see that $\vec{n}\in W^{1,2}(D^2\setminus\{0\})$. In dimension two, the 2-capacity of isolated points is null, so we actually have $\vec{n}\in W^{1,2}(D^2)$. Rescaling if necessary, we shall henceforth always assume that

$$\int_{D^2}|\nabla\vec{n}|^2\,dx<\varepsilon_0, \tag{1-4}$$

where the adjustable parameter $\varepsilon_0\equiv\varepsilon_0(m)$ is chosen to fit our various needs (in particular, we will need it to be "small enough" in Proposition C.1).

For the sake of the following paragraph, we consider a conformal immersion $\vec{\Phi}: D^2\to\mathbb{R}^m$, which is smooth across the unit disk. We introduce the local coordinates (x_1,x_2) for the flat metric on the unit disk $D^2=\{x=(x_1,x_2)\in\mathbb{R}^2 : x_1^2+x_2^2<1\}$. The operators $\nabla=(\partial_{x_1},\partial_{x_2})$, $\nabla^\perp=(-\partial_{x_2},\partial_{x_1})$, $\operatorname{div}=\nabla\cdot$, and $\Delta=\nabla\cdot\nabla$ will be understood in these coordinates. The conformal parameter λ is defined as in (1-2). We set

$$\vec{e}_j:=\mathrm{e}^{-\lambda}\partial_{x_j}\vec{\Phi} \quad \text{for } j\in\{1,2\}. \tag{1-5}$$

As $\vec{\Phi}$ is conformal, $\{\vec{e}_1(x),\vec{e}_2(x)\}$ forms an orthonormal basis of the tangent space $T_{\vec{\Phi}(x)}\vec{\Phi}(D^2)$. Owing to the topology of D^2, there exists for almost every $x\in D^2$ a positively oriented orthonormal basis $\{\vec{n}_1,\dots,\vec{n}_{m-2}\}$ of the normal space $N_{\vec{\Phi}(x)}\vec{\Phi}(D^2)$, such that $\{\vec{e}_1,\vec{e}_2,\vec{n}_1,\dots,\vec{n}_{m-2}\}$ forms a basis of $T_{\vec{\Phi}(x)}\mathbb{R}^m$. From the Plücker embedding, realizing the Grassmannian $\mathrm{Gr}_{m-2}(\mathbb{R}^m)$ as a submanifold of the projective space of the $(m-2)$-th exterior power $\mathbb{P}\big(\bigwedge^{m-2}\mathbb{R}^m\big)$, we can represent the Gauss map as the $(m-2)$-vector $\vec{n}=\bigwedge_{\alpha=1}^{m-2}\vec{n}_\alpha$. Via the Hodge operator $\star$, we identify vectors and $(m-1)$-vectors in $\mathbb{R}^m$; namely,

$$\star(\vec{n}\wedge\vec{e}_1)=\vec{e}_2,\quad \star(\vec{n}\wedge\vec{e}_2)=-\vec{e}_1,\quad \star(\vec{e}_1\wedge\vec{e}_2)=\vec{n}.$$

In this notation, the second fundamental form $\vec{\mathbb{I}}$, which is a symmetric 2-form on $T_{\vec{\Phi}(x)}\vec{\Phi}(D^2)$ into $N_{\vec{\Phi}(x)}\vec{\Phi}(D^2)$, is expressed as

$$\vec{\mathbb{I}}=\sum_{\alpha,i,j}\mathrm{e}^{-2\lambda}h^{\alpha}_{ij}\,\vec{n}_\alpha\,dx_i\otimes dx_j\equiv\sum_{\alpha,i,j}h^{\alpha}_{ij}\,\vec{n}_\alpha(\vec{e}_i)^*\otimes(\vec{e}_j)^*,$$

where

$$h^{\alpha}_{ij}=-\mathrm{e}^{-\lambda}\vec{e}_i\cdot\partial_{x_j}\vec{n}_\alpha.$$

The mean curvature vector is

$$\vec{H}=\sum_{\alpha=1}^{m-2}H^{\alpha}\vec{n}_\alpha=\frac{1}{2}\sum_{\alpha=1}^{m-2}(h^{\alpha}_{11}+h^{\alpha}_{22})\vec{n}_\alpha.$$

The Willmore equation [Weiner 1978] is cast in the form

$$\Delta_\perp\vec{H}+\sum_{\alpha,\beta,i,j}h^{\alpha}_{ij}h^{\beta}_{ij}H^{\beta}\vec{n}_\alpha-2|\vec{H}|^2\vec{H}=0, \tag{1-6}$$

with

$$\Delta_\perp\vec{H}:=\mathrm{e}^{-2\lambda}\pi_{\vec{n}}\operatorname{div}(\pi_{\vec{n}}(\nabla\vec{H})),$$

and $\pi_{\vec{n}}$ is the projection onto the normal space spanned by $\{\vec{n}_\alpha\}_{\alpha=1}^{m-2}$.

The Willmore equation (1-6) is a fourth-order nonlinear equation (in the coefficients of the induced metric, which depends on $\vec{\Phi}$). With respect to the coefficients H^α of the mean curvature vector, it is actually a strongly coupled nonlinear system whose study is particularly challenging. In codimension 1, there is one equation for the scalar curvature; in higher codimension, however, the situation becomes significantly more complicated, and one must seek different techniques to approach the problem. Fortunately, in a conformal parametrization, it is possible[4] to recast the system (1-6) in an equivalent, yet analytically more suitable, form [Rivière 2008]. Namely, we have[5]

$$\operatorname{div}\big(\nabla\vec{H}-3\pi_{\vec{n}}(\nabla\vec{H})+\star(\nabla^\perp\vec{n}\wedge\vec{H})\big)=0. \tag{1-7}$$

This remarkable reformulation in divergence form of the Willmore equation is the starting point of our analysis. In our singular situation, (1-7) holds only away from

[4]This procedure requires choosing the normal frame $\{\vec{n}_\alpha\}$ astutely. See [Rivière 2008] for details.

[5]The operators $\nabla:=(\partial_{x_1},\partial_{x_2})$, $\nabla^\perp:=(-\partial_{x_2},\partial_{x_1})$, and $\operatorname{div}:=\nabla\cdot$ are understood with respect to flat coordinates $\{x_1,x_2\}$ on the unit disk.

the origin, on $D^2 \setminus \{0\}$. In particular, we can define the constant $\vec{\gamma}_0 \in \mathbb{R}^m$, called *first residue*, by

$$\vec{\gamma}_0 := \frac{1}{4\pi} \int_{\partial D^2} \vec{\nu} \cdot \big(\nabla \vec{H} - 3\pi_{\vec{n}}(\nabla \vec{H}) + \star(\nabla^{\perp}\vec{n} \wedge \vec{H})\big), \tag{1-8}$$

where $\vec{\nu}$ denotes the unit outward normal vector to ∂D^2. We will see in Corollary 1.5 that the residue appears in the local asymptotic expansion of the mean curvature vector around the singularity. The residue $\vec{\gamma}_0$ as expressed in (1-8) already appears in [Rivière 2008], where the second author studies Willmore immersions with a unit-density point singularity, thereby generalizing in arbitrary codimension the results of [Kuwert and Schätzle 2004]. Although in the end identical to $\vec{\gamma}_0$, the residue used in the latter is defined differently.

We next state a result describing the regularity of the Gauss map around the point singularity.[6]

Proposition 1.2. *Let $\vec{\Phi} \in C^\infty(D^2 \setminus \{0\}) \cap (W^{2,2} \cap W^{1,\infty})(D^2)$ be a conformal Willmore immersion of the punctured disk into $\mathbb{R}^m$ whose Gauss map $\vec{n}$ lies in $W^{1,2}(D^2)$. Then $\nabla^2\vec{n} \in L^{2,\infty}(D^2)$, and thus in particular $\nabla\vec{n}$ is an element of* BMO. *Furthermore, $\vec{n}$ satisfies the pointwise estimate*

$$|\nabla\vec{n}(x)| \lesssim |x|^{-\epsilon} \quad \textit{for all } \epsilon > 0.$$

If the order of degeneracy of the immersion $\vec{\Phi}$ at the origin is at least two,[7] then $\nabla\vec{n}$ belongs to $L^\infty(B_1(0))$.

A conformal immersion of $D^2 \setminus \{0\}$ into $\mathbb{R}^m$ such that $\nabla\vec{\Phi}$ and the Gauss map $\vec{n}$ both extend to maps in $W^{1,2}(D^2)$ has a distinctive behavior near the point singularity located at the origin. One shows (see [Müller and Šverák 1995] and Lemma A.5 in [Rivière 2013]) that there exists a positive integer θ_0 with

$$|\vec{\Phi}(x)| \simeq |x|^{\theta_0} \quad \text{and} \quad |\nabla\vec{\Phi}(x)| \simeq |x|^{\theta_0 - 1} \quad \text{near the origin.} \tag{1-9}$$

In addition, we have

$$\lambda(x) := \tfrac{1}{2}\log\big(\tfrac{1}{2}|\nabla\vec{\Phi}(x)|^2\big) = (\theta_0 - 1)\log|x| + u(x),$$

where $u \in W^{2,1}(D^2)$, and one has

$$\begin{cases} \nabla\lambda \in L^2(D^2) & \text{when } \theta_0 = 1, \\ |\nabla\lambda(x)| \lesssim |x|^{-1} \in L^{2,\infty}(D^2) & \text{when } \theta_0 \geq 2. \end{cases} \tag{1-10}$$

The function $\mathrm{e}^{-u(x)} \equiv |x|^{\theta_0 - 1}\mathrm{e}^{-\lambda(x)}$ is continuous and strictly positive in a small neighborhood of the origin.

[6] In codimension 1, the statement of Proposition 1.2 was the object of [Kuwert and Schätzle 2007].

[7] Roughly speaking, if $\nabla\vec{\Phi}(0) = \vec{0}$. The notion of "order of degeneracy" is made precise below.

The integer θ_0 is the density of the current $\vec{\Phi}_*[D^2]$ at the image point $0 \in \mathbb{R}^m$.

When such a conformal immersion is Willmore on $D^2 \setminus \{0\}$, it is possible to refine the asymptotics (1-9). The following result describes the behavior of the immersion $\vec{\Phi}$ locally around the singularity at the origin.

Proposition 1.3. *Let $\vec{\Phi}$ be as in Proposition 1.2 with conformal parameter λ, and let θ_0 be as in (1-9). There exists a constant vector $\vec{A} = \vec{A}^1 + i\vec{A}^2 \in \mathbb{R}^2 \otimes \mathbb{R}^m$ satisfying the following conditions*:

(i) $\vec{A}^1 \cdot \vec{A}^2 = 0, \quad |\vec{A}^1| = |\vec{A}^2| = \theta_0^{-1} \lim_{x\to 0} \dfrac{e^{\lambda(x)}}{|x|^{\theta_0 - 1}}, \quad \pi_{\vec{n}(0)}\vec{A} = \vec{0}.$

(ii) *When $\theta_0 = 1$,*

$$\vec{\Phi}(x) = \Re(\vec{A}x) + \vec{\zeta}(x), \tag{1-11}$$

with $\vec{\zeta} \in \bigcap_{p<\infty} W^{2,p}(D^2)$ and

$$\vec{\zeta}(x) = \mathrm{O}(|x|^{2-\epsilon}), \quad \nabla\vec{\zeta}(x) = \mathrm{O}(|x|^{1-\epsilon}) \quad \textit{for all } \epsilon > 0.$$

(iii) *When $\theta_0 \geq 2$,*

$$\vec{\Phi}(x) = \Re\big(\vec{A}x^{\theta_0} + \vec{B}_1 x^{\theta_0+1} + \vec{C}_{\theta_0-1}|x|^{2\theta_0}x^{1-\theta_0}\big) + |x|^{\theta_0-1}\vec{\xi}(x), \tag{1-12}$$

where $\vec{B}_1$ and $\vec{C}_{\theta_0-1}$ are constant vectors in $\mathbb{C}^m$, and moreover for all $\epsilon > 0$,

$$\vec{\xi}(x) = \mathrm{O}(|x|^{3-\epsilon}), \quad \nabla\vec{\xi}(x) = \mathrm{O}(|x|^{2-\epsilon}), \quad \nabla^2\vec{\xi}(x) = \mathrm{O}(|x|^{1-\epsilon}).$$

The plane $\operatorname{span}\{\vec{A}^1, \vec{A}^2\}$ is tangent to the surface at the origin. If $\theta_0 = 1$, this plane is actually $T_0\Sigma$. One can indeed show that the tangent unit vectors $\vec{e}_j(0)$ spanning $T_0\Sigma$ (defined in (1-5)) satisfy $\vec{e}_j(0) = \vec{A}^j/|\vec{A}^j|$. In contrast, when $\theta_0 \geq 2$, the tangent plane $T_0\Sigma$ does not exist in the classical sense, and the vectors $\vec{e}_j(x)$ "spin" as x approaches the origin (see (2-21)). More precisely, $T_0\Sigma$ is the plane $\operatorname{span}\{\vec{A}^1, \vec{A}^2\}$ covered θ_0 times.

Remark 1.4. When $\theta_0 = 1$, the immersion $\vec{\Phi}$ belongs to $C^{1,\alpha}(D^2)$ for all $\alpha \in [0, 1)$. In general however, $\vec{\Phi}$ need *not* be $C^{1,1}(D^2)$. To see this, it suffices to invert the standard catenoid[8] about the origin, thereby yielding a Willmore surface[9] which comprises near the origin two identical graphs (mirror-symmetric), each degenerating with order $\theta_0 = 1$ at the origin. One then directly verifies that

$$|\nabla\vec{n}(x)| \simeq -\log|x| \in \mathrm{BMO} \setminus L^\infty(D^2).$$

Hence, we cannot expect in general $\epsilon = 0$ in (1-11). Moreover, $\vec{\Phi} \notin C^{1,1}(D^2)$.

[8]Conformally parametrized by $(r,\varphi) \mapsto \big((r + r^{-1})\cos(\varphi), (r + r^{-1})\sin(\varphi), -2\log(r)\big)$.

[9]For it is the image of a minimal (thus Willmore) surface under a Möbius transformation.

One also verifies that the inverted catenoid in $\mathbb{R}^3$ and centered around the $(0,0,1)$-axis has first residue $\vec{\gamma}_0 = -4(0,0,1)$.

Having obtained the asymptotic behavior of the immersion $\vec{\Phi}$ and its first two derivatives near the origin, it is possible to obtain analogous information for the mean curvature vector. This is the object of the next proposition.

Corollary 1.5. *Let $\vec{\Phi}$ be as in Proposition 1.2, λ be its conformal parameter, and θ_0 be as in (1-9). Locally around the singularity, the mean curvature vector satisfies*

(i) *when $\theta_0 = 1$,*

$$\vec{H}(x) + \vec{\gamma}_0 \log|x| \in \bigcap_{p<\infty} W^{1,p}(D^2),$$

where $\vec{\gamma}_0$ is the residue defined in (1-8);

(ii) *when $\theta_0 \geq 2$,*

$$e^{\lambda(x)}\vec{H}(x) = 2\theta_0 \mathrm{e}^{-u(x)} \Re\left[\vec{C}_{\theta_0-1}\left(\frac{|x|}{x}\right)^{\theta_0-1}\right] + \mathrm{O}(|x|^{1-\epsilon}) \quad \textit{for all } \epsilon > 0,$$

where $\vec{C}_{\theta_0-1} \in \mathbb{C}^m$ is the same constant vector as in Proposition 1.3(iii), and

$$\mathrm{e}^{u(x)} := |x|^{1-\theta_0}\mathrm{e}^{\lambda(x)} \in C^0(D^2, (0,\infty)).$$

In particular, since $\vec{H}$ is a normal vector, we note that $\pi_{\vec{n}(0)}\vec{C}_{\theta_0-1} = \vec{C}_{\theta_0-1}$.

When $\theta_0 \geq 2$, the weighted mean curvature vector $\mathrm{e}^{\lambda}\vec{H}$ is thus bounded across the singularity (unlike in the case $\theta_0 = 1$, where it behaves logarithmically). But its limit may not exist: $\mathrm{e}^{\lambda(x)}\vec{H}(x)$ is a "spinning vector" as x approaches the origin.[10] We may recast the expansion given in Corollary 1.5(ii) in the form

$$\vec{H}(x) = 2\theta_0 \mathrm{e}^{-2u(0)} \Re(\vec{C}_{\theta_0-1}x^{1-\theta_0}) + \mathrm{O}(|x|^{2-\theta_0-\epsilon}).$$

In this formulation, $\vec{H}$ appears as the sum of an harmonic function with a pole at the origin of order $(\theta_0 - 1)$ and of a rest of lower order. This feature persists even when $\vec{C}_{\theta_0-1} = \vec{0}$ and it can be precisely quantified, namely:

Proposition 1.6. *Let $\vec{\Phi}$ be as in Proposition 1.2, λ be its conformal parameter, and θ_0 be as in (1-9). There exists a complex-valued function $\vec{T}$ satisfying*

$$\partial_x \vec{T} = \mathrm{O}(|\vec{H}||\nabla\vec{n}|) \quad \textit{and} \quad \vec{T} = \mathrm{O}(|x|^{2-\theta_0-\epsilon}) \quad \textit{for all } \epsilon > 0,$$

and such that, locally around the singularity, the mean curvature vector satisfies

$$\vec{H}(x) + \vec{\gamma}_0 \log|x| = \Re(\vec{E}(x) - \vec{T}(x)), \tag{1-13}$$

[10] But the function $\mathrm{e}^{-u(x)}$ does have a finite, positive limit at $x = 0$, as shown in [Müller and Šverák 1995].

where $\vec{\gamma}_0$ *is the residue defined in* (1-8). *The function* $\vec{E}$ *is antiholomorphic with possibly a pole at the origin of order at most* $(\theta_0 - 1)$.

If the singularity of $\vec{E}$ *at the origin has order* $\alpha \in \{0, \dots, \theta_0 - 1\}$, *then the functions* $\vec{E}$ *and* $\vec{T}$ *can be adjusted to satisfy*

$$\vec{E} - \vec{T} = \vec{E}_\alpha \bar{x}^{-\alpha} - \vec{Q}$$

for some nonzero constant vector $\vec{E}_\alpha \in \mathbb{C}^m$, *and with*

$$\partial_x \vec{Q} = \mathrm{O}(|\vec{H}||\nabla \vec{n}|) \quad \textit{and} \quad \vec{Q} = \mathrm{O}(|x|^{1-\alpha-\epsilon}) \quad \textit{for all } \epsilon > 0.$$

Here $\bar{x}$ *denotes the complex conjugate of* $x \in \mathbb{R}^2 \simeq \mathbb{C}$.

We may view the function $\vec{E}$ from the previous proposition as a string of m complex-valued functions $\{E_j\}_{j=1,\dots,m}$, each of which is antiholomorphic and possibly has a pole at the origin of order at most $(\theta_0 - 1)$. This prompts us to introduce the following decisive quantity.

Definition 1.7. The *second residue* associated with the immersion $\vec{\Phi}$ at the origin is the $\mathbb{N}^m$-valued vector

$$\vec{\gamma} = (\gamma_1, \dots, \gamma_m) \quad \text{with } \gamma_j := \frac{1}{2i\pi} \int_{\partial D^2} d \log E_j \in \mathbb{N}. \tag{1-14}$$

The importance of $\vec{\gamma}$ cannot be overstated: it controls the leading-order singular behavior of the mean curvature at the origin, as the following statement shows.

Theorem 1.8. *Let* $\vec{\Phi}$ *be as in* Proposition 1.2 *and let* λ *be its conformal parameter*, θ_0 *as in* (1-9), *and the residues* $\vec{\gamma}_0$ *and* $\vec{\gamma}$ *as in* (1-8) *and* (1-14), *respectively. Define*

$$\alpha := \max_{1 \le j \le m} \gamma_j \in \{0, \dots, \theta_0 - 1\}.$$

Then $\nabla^{\theta_0+1-\alpha} \vec{n} \in L^{2,\infty}(D^2)$, *and thus* $\nabla^{\theta_0-\alpha} \vec{n} \in \mathrm{BMO}(D^2)$.

Locally around the origin, the immersion has the asymptotic expansion

$$\vec{\Phi}(x) = \Re\left(\vec{A} x^{\theta_0} + \sum_{j=1}^{\theta_0-\alpha} \vec{B}_j x^{\theta_0+j} + \vec{C}_\alpha |x|^{2\theta_0} x^{-\alpha}\right) - \vec{C} |x|^{2\theta_0} (\log |x|^{2\theta_0} - 4) + \vec{\xi}(x),$$

where $\vec{B}_j$ *and* $\vec{C}_\alpha \in \mathbb{C}^m$ *are constant vectors,* $\vec{A}$ *is as in* Proposition 1.3, *and*[11] $\vec{C} := \mathrm{e}^{2u(0)}/(2\theta_0^3)\vec{\gamma}_0$. *Furthermore, the function* $\vec{\xi}$ *satisfies the estimates*

$$\nabla^j \vec{\xi}(x) = \mathrm{O}(|x|^{2\theta_0-\alpha-j+1-\epsilon}) \quad \textit{for all } \epsilon > 0 \textit{ and } j \in \{0, \dots, \theta_0 - \alpha + 1\},$$
$$|x|^{1-\theta_0} \nabla^{\theta_0-\alpha+2} \vec{\xi} \in \bigcap_{p<\infty} L^p.$$

[11] The function u is as in Corollary 1.5.

In particular, we have

$$\vec{H}(x) = \Re(\vec{E}_\alpha \bar{x}^{-\alpha}) - \vec{\gamma}_0 \log|x| + \vec{\eta}(x),$$

where $\vec{E}_\alpha := 2\theta_0(\theta_0 - \alpha)\mathrm{e}^{-2u(0)}\vec{C}_\alpha^*$. *The function* $\vec{\eta}$ *satisfies*

$$\nabla^j \vec{\eta}(x) = \mathrm{O}(|x|^{1-j-\alpha-\epsilon}) \quad \text{for all } \epsilon > 0 \text{ and } j \in \{0, \dots, \theta_0 - \alpha - 1\},$$
$$|x|^{\theta_0 - 1}\nabla^{\theta_0-\alpha}\vec{\eta} \in \bigcap_{p<\infty} L^p.$$

From this result, it comes as no surprise to discover that the simultaneous vanishing of both residues $\vec{\gamma}_0$ and $\vec{\gamma}$ improves the regularity of the immersion $\vec{\Phi}$. This is the content of the next statement.

Theorem 1.9. *Under the hypotheses of Corollary 1.5, suppose that the first residue vanishes*: $\vec{\gamma}_0 = \vec{0}$.

(i) *When* $\theta_0 = 1$, *the immersion* $\vec{\Phi}$ *is smooth across the branch point.*

(ii) *When* $\theta_0 > 1$, *if in addition the second residue vanishes* $\vec{\gamma} = \vec{0}$, *the immersion* $\vec{\Phi}$ *is smooth across the branch point.*

Remark 1.10. Let $\vec{\Phi} : D^2 \to \mathbb{R}^{m\geq 3}$ be a minimal immersion with a branch point at the origin. Since minimal immersions have vanishing first and second residues (see Remark 2.6), Theorem 1.9 applies and singularities are removable.

We close this section with an important observation. When the Willmore immersion $\vec{\Phi}$ is smooth, the Willmore equation written in divergence form (1-7) holds on the whole unit disk D^2. Hence, the first residue $\vec{\gamma}_0$ defined in (1-8) vanishes about every point. In turn, the expansion (1-13) given in Proposition 1.6 shows that the antiholomorphic function $\vec{E}$ cannot be singular anywhere in D^2, and thus in particular that the second residue $\vec{\gamma}$ also vanishes about every point.

Given a branched Willmore immersion, we have found some explicit conditions ensuring the removability of the branch points. These conditions require that certain residues vanish. Naturally, if the immersion is explicitly given, one may *directly* verify smoothness at the bad points, without resorting to the residues. This is of course not the situation which we have in mind. Suppose now that a sequence of smooth Willmore immersions is given, with uniformly L^2-bounded second fundamental form (the uniform bound need *not* be smaller than 8π). If the sequence does not degenerate in moduli space,[12] it is known that one can extract a subsequence which converges strongly away from finitely many points to a limit-immersion which is Willmore and smooth away from finitely many isolated branch points. As residues are computed as circulation integrals along circles enlacing the singularities, one expects that they pass to the weak limit. This is indeed the case

[12] See [Bernard and Rivière 2011a] for further information on this condition.

for the first residue $\vec{\gamma}_0$ (as one easily verifies), but it remains an open problem for the second residue $\vec{\gamma}$. Because each immersion is smooth, its residues vanish about every point. If one knew that the second residue passed through weak limits, one could then conclude that the residues associated with the limit-immersion vanished as well, thereby making the limit into a branched *smooth* Willmore immersion.

1C. ***Some examples.*** Covering three times over the inverted catenoid of Remark 1.4 gives rise to a conformal Willmore immersion which degenerates at the origin with order $\theta_0 = 3$. Yet, the geometry of the image is identical to that of the inverted catenoid, i.e., with a degeneracy of order 1. This is an instance of a "false" third-order branch point simply resulting from having chosen a "bad" parametrization: the singularity truly has order one. Without surprise, in this case, we find that the first residue is $\vec{\gamma}_0 = -12(0,0,1)$ (i.e., three times that of the singly covered inverted catenoid of Remark 1.4). The second residue $\vec{\gamma}$ vanishes.

A conformal parametrization of the 3-Enneper (minimal) surface in $\mathbb{R}^3$ is given by

$$(r,\varphi) \mapsto \left(\frac{1}{3r^3}\cos 3\varphi - \frac{1}{r}\cos\varphi,\ \frac{1}{3r^3}\sin 3\varphi + \frac{1}{r}\sin\varphi,\ \frac{1}{r^2}\cos 2\varphi - 1\right).$$

Inverting this surface about the point $(0,0,0)$ gives rise to a compact Willmore surface whose conformal parametrization near $r=0$ satisfies

$$\vec{\Phi}(r,\varphi) = 3\big(r^3\cos 3\varphi, r^3\sin 3\varphi, 3r^4\cos 2\varphi\big) + \mathrm{O}(r^5, r^5, r^6).$$

This surface has a branch point of order $\theta_0 = 3$ at the origin, where the mean curvature is

$$\vec{H}(r,\varphi) = \left(0, 0, \frac{2}{3r^2}\cos 2\varphi\right) + \mathrm{O}(r^{-1}).$$

The first residue is computed to be $\vec{\gamma}_0 = \vec{0}$, and the second residue is $\vec{\gamma} = (0,0,2)$.

In codimension two, we consider now the following conformal parametrization of a minimal surface:

$$(r,\varphi) \mapsto \left(\frac{1}{r^3}\cos 3\varphi, \frac{1}{r^3}\sin 3\varphi, \frac{1}{r}\cos\varphi, \frac{1}{r}\sin\varphi - 1\right).$$

Inverting this surface about the origin in $\mathbb{R}^4$ gives rise to a compact Willmore surface whose conformal parametrization near $r=0$ satisfies

$$\vec{\Phi}(r,\varphi) = \big(r^3\cos 3\varphi, r^3\sin 3\varphi, r^5\cos\varphi, r^5\sin\varphi\big) + \mathrm{O}(r^7, r^7, r^9, r^6),$$

having too a branch point of order $\theta_0 = 3$ at the origin. The mean curvature is

$$\vec{H}(r,\varphi) = \left(0, 0, \frac{4}{r}\cos\varphi, \frac{4}{r}\sin\varphi\right) + \mathrm{O}(1).$$

The second residue is $\vec{\gamma} = (0, 0, 1, 1)$, while the first residue vanishes.

2. Proofs of the theorems

2A. ***Fundamental results and reformulation.*** We place ourselves in the situation described in Section 1B. Namely, we have a Willmore immersion $\vec{\Phi}$ on the punctured disk which degenerates at the origin in such a way that

$$|\vec{\Phi}(x)| \simeq |x|^{\theta_0} \quad \text{and} \quad |\nabla\vec{\Phi}(x)| = \sqrt{2}\mathrm{e}^{\lambda(x)} \simeq |x|^{\theta_0 - 1}$$

for some $\theta_0 \in \mathbb{N} \setminus \{0\}$.

Amongst the analytical tools available to the study of weak Willmore immersions with square-integrable second fundamental form, an important one is certainly the *ε-regularity*. The version appearing in Theorem 2.10 and Remark 2.11 of [Kuwert and Schätzle 2001] (see also Theorem I.5 in [Rivière 2008]) states that there exists $\varepsilon_0 > 0$ such that, if

$$\int_{B_1(0)} |\nabla\vec{n}|^2\, dx < \varepsilon_0, \tag{2-1}$$

then we have

$$\|\mathrm{e}^{-\lambda}\nabla\vec{n}\|_{L^\infty(B^g_\sigma)} \leq \frac{C}{\sigma}\|\nabla\vec{n}\|_{L^2(B^g_{2\sigma})} \quad \text{for all } B^g_{2\sigma} \subseteq \Omega := D^2 \setminus \{0\}, \tag{2-2}$$

where B^g_σ is a geodesic disk of radius σ for the induced metric $g = \vec{\Phi}^* g_{\mathbb{R}^m}$, and C is a universal constant. As always, λ denotes the conformal parameter.

The ε-regularity enables us to obtain the following result, already observed in [Kuwert and Schätzle 2007], and decisive to the remainder of the argument.

Lemma 2.1. *The function $\delta(r) := r \sup_{|x|=r} |\nabla\vec{n}(x)|$ satisfies*

$$\lim_{r\searrow 0} \delta(r) = 0 \quad \textit{and} \quad \int_0^1 \delta^2(r)\,\frac{dr}{r} < \infty.$$

Proof. From (1-3) and (1-9), the metric g satisfies

$$g_{ij}(x) \simeq |x|^{2(\theta_0-1)}\delta_{ij} \quad \text{on } B_{2r}(0) \setminus B_{r/2}(0) \text{ for all } r \in (0, 1/2).$$

A simple computation then shows that

$$B^g_{2cr^{\theta_0}}(x) \subset B_{2r}(0) \setminus B_{r/2}(0) \quad \text{for all } x \in \partial B_r(0), \tag{2-3}$$

where $0 < 2\theta_0 c < 1 - 2^{-\theta_0}$.

Since the metric g does not degenerate away from the origin, given $0 < r < 1/2$, we can always cover the flat circle $\partial B_r(0)$ with finitely many metric disks:

$$\partial B_r(0) \subset \bigcup_{j=1}^{N} B^g_{cr^{\theta_0}}(x_j) \quad \text{with } x_j \in \partial B_r(0).$$

Hence, per the latter, (2-2), and (2-3), we obtain that for some $x_0 \in \partial B_r(0)$ we have

$$\begin{aligned} r \sup_{|x|=r} |\nabla\vec{n}(x)| \simeq r^{\theta_0} \sup_{|x|=r} |\mathrm{e}^{-\lambda(x)}\nabla\vec{n}(x)| &\le r^{\theta_0} \|\mathrm{e}^{-\lambda}\nabla\vec{n}\|_{L^\infty(B^g_{cr^{\theta_0}}(x_0))} \\ &\lesssim \|\nabla\vec{n}\|_{L^2(B^g_{2cr^{\theta_0}}(x_0))} \le \|\nabla\vec{n}\|_{L^2(B_{2r}(0)\setminus B_{r/2}(0))}. \end{aligned} \tag{2-4}$$

As $\nabla\vec{n}$ is square-integrable by hypothesis, letting r tend to zero in the latter yields the first assertion.

The second assertion follows from (2-4), namely,

$$\int_0^{1/2} \delta^2(r)\,\frac{dr}{r} \lesssim \int_0^{1/2} \|\nabla\vec{n}\|^2_{L^2(B_{2r}(0)\setminus B_{r/2}(0))}\,\frac{dr}{r} = \log(4)\|\nabla\vec{n}\|^2_{L^2(B_1(0))},$$

which is by hypothesis finite. □

Recalling (1-3) linking the Gauss map to the mean curvature vector and the fact that $\mathrm{e}^{\lambda(x)} \simeq |x|^{\theta_0-1}$, we obtain from Lemma 2.1 that

$$r^{\theta_0} \sup_{|x|=r} |\vec{H}(x)| \le r^{\theta_0} \sup_{|x|=r} \mathrm{e}^{-\lambda(x)}|\nabla\vec{n}(x)| \lesssim \delta(r). \tag{2-5}$$

The Willmore equation (1-7) may be alternatively written

$$\operatorname{div}\big(\nabla\vec{H} - 3\pi_{\vec{n}}(\nabla\vec{H}) - \star(\vec{n}\wedge\nabla^\perp\vec{H})\big) = 0 \quad \text{on } \Omega := B_1(0)\setminus\{0\}.$$

It is elliptic [Rivière 2008]. Using the information on the gradient of $\vec{n}$ given by (2-2), and some standard analytical techniques for second-order elliptic equations in divergence form (see [Grüter and Widman 1982]), one deduces from (2-5) that

$$r^{\theta_0+1} \sup_{|x|=r} |\nabla\vec{H}(x)| \lesssim \delta(r). \tag{2-6}$$

These observations shall be helpful in the sequel.

Equation (1-7) implies that, for any ball $B_\rho(0)$ of radius ρ centered on the origin and contained in Ω, we have

$$\int_{\partial B_\rho(0)} \vec{\nu}\cdot\big(\nabla\vec{H} - 3\pi_{\vec{n}}(\nabla\vec{H}) + \star(\nabla^\perp\vec{n}\wedge\vec{H})\big) = 4\pi\vec{\gamma}_0 \quad \text{for all } \rho \in (0,1), \tag{2-7}$$

where $\vec{\gamma}_0$ is the residue defined in (1-8). Here $\vec{\nu}$ denotes the unit outward normal vector to $\partial B_\rho(0)$. An elementary computation shows that

$$\int_{\partial B_\rho(0)} \vec{\nu}\cdot\nabla\log|x| = 2\pi \quad \text{for all } \rho>0.$$

Thus, upon setting

$$\vec{X} := \nabla\vec{H} - 3\pi_{\vec{n}}(\nabla\vec{H}) + \star(\nabla^\perp\vec{n}\wedge\vec{H}) - 2\vec{\gamma}_0\nabla\log|x|, \tag{2-8}$$

we find

$$\operatorname{div}\vec{X} = 0 \quad \text{on } \Omega$$

and

$$\int_{\partial B_\rho(0)} \vec{\nu}\cdot\vec{X} = 0 \quad \text{for all } \rho\in(0,1).$$

As $\vec{X}$ is smooth away from the origin, the Poincaré lemma implies now the existence of an element $\vec{L}\in C^\infty(\Omega)$, defined up to an additive constant, such that

$$\vec{X} = \nabla^\perp\vec{L} \quad \text{on } \Omega. \tag{2-9}$$

We deduce from Lemma 2.1 and (2-5)–(2-9) that

$$\int_{B_1(0)} |x|^{2\theta_0}|\nabla\vec{L}|^2\,dx \lesssim \int_0^1 \delta^2(s)\,\frac{ds}{s} < \infty. \tag{2-10}$$

A classical Hardy–Sobolev inequality gives the estimate

$$\theta_0^2\int_{B_1(0)} |x|^{2(\theta_0-1)}|\vec{L}|^2\,dx \le \int_{B_1(0)} |x|^{2\theta_0}|\nabla\vec{L}|^2\,dx + \theta_0\int_{\partial B_1(0)} |\vec{L}|^2, \tag{2-11}$$

which is a finite quantity, owing to (2-10) and to the smoothness of $\vec{L}$ away from the origin. The immersion $\vec{\Phi}$ has near the origin the asymptotic behavior

$$|\nabla\vec{\Phi}(x)| \simeq |x|^{\theta_0-1}.$$

Hence (2-11) yields that

$$\vec{L}\cdot\nabla\vec{\Phi},\ \vec{L}\wedge\nabla\vec{\Phi}\in L^2(B_1(0)). \tag{2-12}$$

We next set $\vec{F}(x) := 2\vec{\gamma}_0\log|x|$, and define the functions g and $\vec{G}$ via

$$\begin{cases} \Delta g = \nabla\vec{F}\cdot\nabla\vec{\Phi}, & \Delta\vec{G} = \nabla\vec{F}\wedge\nabla\vec{\Phi} \quad \text{in } B_1(0),\\ g = 0, & \vec{G} = \vec{0} \quad \text{on } \partial B_1(0). \end{cases} \tag{2-13}$$

Since $|\nabla\vec{\Phi}(x)|\simeq|x|^{\theta_0-1}$ near the origin and $\vec{F}$ is the fundamental solution of the

Laplacian, by applying Calderón–Zygmund estimates to (2-13), we find[13]

$$\nabla^2 g,\ \nabla^2\vec{G} \in \begin{cases} L^{2,\infty}(B_1(0)), & \theta_0 = 1,\\ \mathrm{BMO}(B_1(0)), & \theta_0 \ge 2.\end{cases} \tag{2-14}$$

In [Bernard and Rivière 2011b] (see Lemma A.2), the authors derive the identities[14]

$$\begin{cases} \nabla\vec{\Phi}\cdot(\nabla^\perp\vec{L}+\nabla\vec{F}) = 0,\\ \nabla\vec{\Phi}\wedge(\nabla^\perp\vec{L}+\nabla\vec{F}) = -2\nabla\vec{\Phi}\wedge\nabla\vec{H}.\end{cases} \tag{2-15}$$

Accounted into (2-13), the latter yield that we have, in Ω,

$$\begin{cases} \mathrm{div}(\vec{L}\cdot\nabla^\perp\vec{\Phi}-\nabla g) = 0,\\ \mathrm{div}(\vec{L}\wedge\nabla^\perp\vec{\Phi}-2\vec{H}\wedge\nabla\vec{\Phi}-\nabla\vec{G}) = \vec{0},\end{cases} \tag{2-16}$$

where we have used the fact that

$$\Delta\vec{\Phi}\wedge\vec{H} = 2\mathrm{e}^{2\lambda}\vec{H}\wedge\vec{H} = \vec{0}.$$

The terms under the divergence symbols in (2-16) both belong to $L^2(B_1(0))$, owing to (2-12) and (2-14). The distributional equations (2-16), which are *a priori* to be understood on Ω, may thus be extended to all of $B_1(0)$. Indeed, a classical result of Laurent Schwartz states that the only distributions supported on $\{0\}$ are linear combinations of derivatives of the Dirac delta mass. Yet, none of these (including delta itself) belongs to $W^{-1,2}$. We shall thus understand (2-16) on $B_1(0)$. It is not difficult to verify (see Corollary IX.5 in [Dautray and Lions 1984]) that a divergence-free vector field in $L^2(B_1(0))$ is the curl of an element in $W^{1,2}(B_1(0))$. We apply this observation to (2-16) so as to infer the existence of two functions[15] S and $\vec{R}$ in the space $W^{1,2}(B_1(0))\cap C^\infty(\Omega)$, with

$$\begin{cases} \nabla^\perp S = \vec{L}\cdot\nabla^\perp\vec{\Phi}-\nabla g,\\ \nabla^\perp\vec{R} = \vec{L}\wedge\nabla^\perp\vec{\Phi}-2\vec{H}\wedge\nabla\vec{\Phi}-\nabla\vec{G}.\end{cases} \tag{2-17}$$

According to the identities (B-14) in the appendix, the functions S and $\vec{R}$ satisfy on $B_1(0)$ the following system of equations, called the *conservative conformal*

[13] The weak-L^2 Marcinkiewicz space $L^{2,\infty}(B_1(0))$ is defined as those functions f which satisfy $\sup_{\alpha>0}\alpha^2|\{x\in B_1(0): |f(x)|\ge\alpha\}|<\infty$. In dimension two, the prototype element of $L^{2,\infty}$ is $|x|^{-1}$. The space $L^{2,\infty}$ is also a Lorentz space, and in particular is a space of interpolation between Lebesgue spaces, which justifies the first inclusion in (2-14). See [Hélein 1996] or [Almeida 1995] for details.

[14] Observe that $\nabla^\perp\vec{L}+\nabla\vec{F}$ is exactly the divergence-free quantity appearing in (1-7).

[15] S is a scalar while $\vec{R}$ is $\bigwedge^2(\mathbb{R}^m)$-valued.

Willmore system:[16]

$$
\begin{cases} -\Delta S = \nabla(\star\vec{n})\cdot\nabla^{\perp}\vec{R} + \operatorname{div}\big((\star\vec{n})\cdot\nabla\vec{G}\big), \\ -\Delta\vec{R} = \nabla(\star\vec{n})\bullet\nabla^{\perp}\vec{R} - \nabla(\star\vec{n})\cdot\nabla^{\perp}S + \operatorname{div}\big((\star\vec{n})\bullet\nabla\vec{G} + \star\vec{n}\nabla g\big). \end{cases} \tag{2-18}
$$

Not only is this system independent of the codimension (which enters the equations in the guise of the operators $\star$ and $\bullet$), but it further displays two fundamental advantages. Analytically, (2-18) is uniformly elliptic. This is in sharp contrast with the Willmore equation (1-6) whose leading order operator $\Delta_{\perp}$ degenerates at the origin, owing to the presence of the conformal factor $\mathrm{e}^{\lambda(x)} \simeq |x|^{\theta_0-1}$. Structurally, the system (2-18) is in divergence form. We shall in the sequel capitalize on this remarkable feature to develop arguments of "integration by compensation". *A priori* however, since $\vec{n}$, S, and $\vec{R}$ are elements of $W^{1,2}$, the leading terms on the right-hand side of the conservative conformal Willmore system (2-18) are critical. This difficulty is nevertheless bypassed using the fact that the $W^{1,2}$-norm of the Gauss map $\vec{n}$ is chosen to be small enough (see (1-4)).

2B. *The general case when $\theta_0 \geq 1$.* We have gathered enough information about the functions involved to apply to the system (2-18) (a slightly extended version of) Proposition C.1 and thereby obtain that

$$
\nabla S,\ \nabla\vec{R} \in L^p(B_1(0)) \quad \text{for some } p > 2. \tag{2-19}
$$

It is shown at the end of Section B in the appendix that

$$
2\Delta\vec{\Phi} = (\nabla S - \nabla^{\perp}g)\cdot\nabla^{\perp}\vec{\Phi} - (\nabla\vec{R} - \nabla^{\perp}\vec{G})\bullet\nabla^{\perp}\vec{\Phi}. \tag{2-20}
$$

Hence, as $|\nabla\vec{\Phi}(x)| \simeq \mathrm{e}^{\lambda(x)} \simeq |x|^{\theta_0-1}$ around the origin, using (2-14) and (2-19), we may call upon Proposition C.2 with the weight $|\mu| = \mathrm{e}^{\lambda}$ and $a = \theta_0 - 1$ to conclude that

$$
(\partial_{x_1} + i\,\partial_{x_2})\vec{\Phi}(x) = \vec{P}(\bar{x}) + \mathrm{e}^{\lambda(x)}\vec{T}(x),
$$

where $\vec{P}$ is a $\mathbb{C}^m$-valued polynomial of degree at most $(\theta_0 - 1)$, and $\vec{T}(x) = \mathrm{O}(|x|^{1-2/p-\epsilon})$ for every $\epsilon > 0$. Because $\mathrm{e}^{-\lambda}\nabla\vec{\Phi}$ is a bounded function, we deduce more precisely that $\vec{P}(\bar{x}) = \theta_0\vec{A}^{*}\bar{x}^{\theta_0-1}$, for some constant vector $\vec{A} \in \mathbb{C}^m$ (we denote its complex conjugate by $\vec{A}^{*}$), so that

$$
\nabla\vec{\Phi}(x) = \begin{pmatrix} \Re \\ -\Im \end{pmatrix} (\theta_0\vec{A}x^{\theta_0-1}) + \mathrm{e}^{\lambda(x)}\vec{T}(x). \tag{2-21}
$$

Equivalently, upon writing $\vec{A} = \vec{A}^1 + i\,\vec{A}^2$, where $\vec{A}^1$ and $\vec{A}^2$ are two vectors in $\mathbb{R}^m$,

[16] Refer to Appendix A for the notation and the operators used.

the latter may be recast as

$$\begin{cases} \partial_{x_1}\vec{\Phi}(x) = \theta_0|x|^{\theta_0-1}\big[\vec{A}^1\cos((\theta_0-1)\varphi) - \vec{A}^2\sin((\theta_0-1)\varphi)\big] + e^{\lambda}\Re(\vec{T}(x)),\\ -\partial_{x_2}\vec{\Phi}(x) = \theta_0|x|^{\theta_0-1}\big[\vec{A}^2\cos((\theta_0-1)\varphi) + \vec{A}^1\sin((\theta_0-1)\varphi)\big] - e^{\lambda}\Im(\vec{T}(x)),\end{cases}$$

where $\varphi \in [0, 2\pi)$ denotes the argument of $x \in B_1(0)$. The conformality condition on $\vec{\Phi}$ shows easily that

$$|\vec{A}^1| = |\vec{A}^2| \quad \text{and} \quad \vec{A}^1\cdot\vec{A}^2 = 0. \tag{2-22}$$

Yet more precisely, as $|\nabla\vec{\Phi}|^2 = 2e^{2\lambda}$, we see that

$$|\vec{A}^1| = |\vec{A}^2| = \frac{1}{\theta_0}\lim_{x\to 0}\frac{e^{\lambda(x)}}{|x|^{\theta_0-1}} \in \,]0,\infty[\,. \tag{2-23}$$

Because $\vec{\Phi}(0) = \vec{0}$, we obtain from (2-21) the local expansion

$$\vec{\Phi}(x) = \Re(\vec{A}x^{\theta_0}) + O\big(|x|^{\theta_0+1-\frac{2}{p}-\epsilon}\big).$$

Since $\pi_{\vec{n}}\nabla\vec{\Phi} \equiv \vec{0}$, we deduce from (2-21) that

$$\pi_{\vec{n}(x)}\vec{A} = -\theta_0^{-1}x^{1-\theta_0}e^{\lambda}\pi_{\vec{n}}\vec{T}^*(x) = O\big(|x|^{1-\frac{2}{p}-\epsilon}\big) \quad \text{for all } \epsilon > 0. \tag{2-24}$$

Now let $\delta := 1 - 2/p \in (0, 1)$, and let $0 < \eta < p$ be arbitrary. We choose some ϵ satisfying

$$0 < \epsilon < \frac{2\eta}{p(p-\eta)} \equiv \delta - 1 + \frac{2}{p-\eta}.$$

We have observed that $\pi_{\vec{n}}\vec{A} = O(|x|^{\delta-\epsilon})$; hence $\pi_{\vec{n}}\vec{A} = o(|x|^{1-2/(p-\eta)})$, and, in particular, we find

$$\frac{1}{|x|}\pi_{\vec{n}(x)}\vec{A} \in L^{p-\eta}(B_1(0)) \quad \text{for all } \eta > 0. \tag{2-25}$$

This fact shall be helpful in the sequel.

When $\theta_0 = 1$, one directly deduces from the standard Calderón–Zygmund theorem applied to (2-20) that $\nabla^2\vec{\Phi} \in L^p$. In that case, e^{λ} is bounded from above and below, and thus the identity

$$|\nabla\vec{n}| = e^{-\lambda}|\pi_{\vec{n}}\nabla^2\vec{\Phi}| \tag{2-26}$$

(derived as (B-4) in the appendix) yields that $\nabla\vec{n} \in L^p$. When now $\theta_0 \geq 2$, we must proceed slightly differently to obtain analogous results. From (1-10), we know that $|x|\nabla\lambda(x)$ is bounded across the unit disk. We may thus apply Proposition C.2(ii) to (2-20) with the weight $|\mu| = \mathrm{e}^{\lambda}$ and $a = \theta_0$. The required hypothesis (C-13) is fulfilled, and so we obtain

$$\nabla^2 \vec{\Phi}(x) = \theta_0(1-\theta_0) \begin{pmatrix} -\Re & \Im \\ \Im & \Re \end{pmatrix} (\vec{A} x^{\theta_0 - 2}) + \mathrm{e}^{\lambda(x)} \vec{Q}(x), \tag{2-27}$$

where $\vec{A}$ is as in (2-21), and $\vec{Q}$ lies in $\mathbb{R}^4 \otimes L^{p-\epsilon}(B_1(0), \mathbb{R}^m)$ for every $\epsilon > 0$. The exponent $p > 2$ is the same as in (2-19).

Since $\mathrm{e}^{\lambda(x)} \simeq |x|^{\theta_0 - 1}$, we obtain from (2-27) that

$$\mathrm{e}^{-\lambda} |\pi_{\vec{n}} \nabla^2 \vec{\Phi}| \lesssim |x|^{-1} |\pi_{\vec{n}} \vec{A}| + |\pi_{\vec{n}} \vec{Q}|.$$

According to (2-25), the first summand on the right-hand side of the latter belongs to $L^{p-\eta}$ for all $\eta > 0$. Moreover, we have seen that $\pi_{\vec{n}} \vec{Q}$ lies in $L^{p-\epsilon}$ for all $\epsilon > 0$, whence it follows that $\mathrm{e}^{-\lambda} \pi_{\vec{n}} \nabla^2 \vec{\Phi}$ is itself an element of $L^{p-\epsilon}$ for all $\epsilon > 0$. Brought into (2-26), this information implies that

$$\nabla \vec{n} \in L^{p-\epsilon}(B_1(0)) \quad \text{for all } \epsilon > 0. \tag{2-28}$$

In light of this new fact, we may now return to (2-18). In particular, recalling (2-14), we find

$$\Delta S \equiv -\nabla(\star \vec{n}) \cdot (\nabla^{\perp} \vec{R} + \nabla \vec{G}) - (\star \vec{n}) \cdot \Delta \vec{G} \in L^q(B_1(0)),$$

with

$$\frac{1}{q} = \frac{1}{p} + \frac{1}{p - \epsilon}.$$

We attract the reader's attention on an important phenomenon occurring when $\theta_0 = 1$. In this case, if the aforementioned value of q exceeds 2 (i.e., if $p > 4$), then $\Delta S \notin L^q$, but rather only $\Delta S \in L^{2,\infty}$. This integrability "barrier" stems from that of $\Delta \vec{G}$, as given in (2-14). The same considerations apply of course with $\vec{R}$ and g in place of S and $\vec{G}$, respectively.

Our findings so far may be summarized as follows:

$$\nabla S,\ \nabla \vec{R} \in \begin{cases} W^{1,(2,\infty)} & \text{if } \theta_0 = 1 \text{ and } p > 4, \\ W^{1,q} & \text{otherwise.} \end{cases} \tag{2-29}$$

With the help of the Sobolev embedding theorem (and a result of [Tartar 2007] stating that $W^{1,(2,\infty)} \subset \mathrm{BMO}$), we infer that

$$\nabla S,\ \nabla \vec{R} \in \begin{cases} \mathrm{BMO} & \text{if } \theta_0 = 1 \text{ and } p > 4, \\ L^{\infty} & \text{if } \theta_0 \geq 2 \text{ and } p > 4, \\ L^s & \text{if } \theta_0 \geq 1 \text{ and } p \leq 4, \end{cases} \tag{2-30}$$

with

$$\frac{1}{s} = \frac{1}{q} - \frac{1}{2} = \frac{1}{p} + \frac{1}{p - \epsilon} - \frac{1}{2} < \frac{1}{p}.$$

Comparing (2-30) to (2-19), we see that the integrability has been improved. The process may thus be repeated until reaching that

$$\nabla S,\ \nabla\vec{R} \in L^b(B_1(0)) \quad \text{for all } b < \infty$$

holds in all configurations. With the help of this newly found fact, we reapply Proposition C.2 so as to improve (2-29) and (2-28) to

$$\nabla S,\ \nabla\vec{R} \in \begin{cases} W^{1,(2,\infty)}(B_1(0)) & \text{if } \theta_0 = 1, \\ W^{1,b}(B_1(0)) & \text{if } \theta_0 \geq 2 \text{ for all } b < \infty, \end{cases} \tag{2-31}$$

and

$$\nabla\vec{n} \in L^b(B_1(0)) \quad \text{for all } b < \infty. \tag{2-32}$$

The ε-regularity in the form (2-4) then yields a pointwise estimate for the Gauss map. Namely, in a neighborhood of the origin,

$$|\nabla\vec{n}(x)| \lesssim |x|^{-\epsilon} \quad \text{for all } \epsilon > 0. \tag{2-33}$$

2C. *The case $\theta_0 = 1$.* We shall now investigate further the case $\theta_0 = 1$, when $|\nabla\vec{\Phi}| \simeq \mathrm{e}^{\lambda}$ is bounded from both above and below around the origin. Setting

$$\vec{F}_1 := \nabla^{\perp}\vec{R} + \nabla\vec{G} \quad \text{and} \quad F_2 := \nabla^{\perp}S + \nabla g \tag{2-34}$$

in (2-20) gives

$$-2\Delta\vec{\Phi} = F_2 \cdot \nabla\vec{\Phi} - \vec{F}_1 \bullet \nabla\vec{\Phi}. \tag{2-35}$$

According to (2-14) and (2-29), the right-hand side of the latter has bounded mean oscillations. Hence $\nabla^2\vec{\Phi} \in \bigcap_{p<\infty} L^p$. Using that $2\mathrm{e}^{2\lambda}\vec{H} = \Delta\vec{\Phi}$, we differentiate (2-35) to obtain

$$-4\nabla(\mathrm{e}^{2\lambda}\vec{H}) = \nabla F_2 \cdot \nabla\vec{\Phi} - \nabla\vec{F}_1 \bullet \nabla\vec{\Phi} + F_2 \cdot \nabla^2\vec{\Phi} - \vec{F}_1 \bullet \nabla^2\vec{\Phi},$$

which, still owing to (2-14) and (2-29), and to the boundedness of $\nabla\vec{\Phi}$, shows that $\mathrm{e}^{2\lambda}\vec{H} \in W^{1,(2,\infty)}$. As $\mathrm{e}^{\pm\lambda}$ are bounded from below, we see that $\vec{H} \in \mathrm{BMO}$. Using that $\nabla\lambda \in L^2$ (see (1-10)), it follows that

$$\nabla\vec{H} = \mathrm{e}^{-2\lambda}\nabla(\mathrm{e}^{2\lambda}\vec{H}) - 2(\nabla\lambda)\vec{H} \in \bigcap_{p<2} L^p. \tag{2-36}$$

We shall now derive an asymptotic expansion for $\vec{H}(x)$ near the origin. To this end, we use a "generic" procedure, whose assumptions are fulfilled owing to our work from the previous section (in particular (2-32)) and to (2-36).

Proposition 2.2. *Let the immersion $\vec{\Phi}$ satisfy an expansion of the type (2-21) for all $p<\infty$. Suppose that $\vec{n}\in\bigcap_{p<\infty}W^{1,p}(B_1(0))$ and $\vec{H}\in\bigcap_{p<2}W^{1,p}(B_1(0))$. Then, locally around the origin,*

$$\vec{H}(x)+\vec{\gamma}_0\log|x|\in\bigcap_{p<\infty}W^{1,p}(B_1(0)),$$

where $\vec{\gamma}_0$ is the residue defined in (2-7).

Proof. In order to derive this result, one must return to (1-7):

$$\mathscr{L}(\vec{H}):=\operatorname{div}\big(\nabla\vec{H}-3\pi_{\vec{n}}\nabla\vec{H}+\star(\nabla^{\perp}\vec{n}\wedge\vec{H})\big)=0\quad\text{on } B_1(0)\setminus\{0\}.$$

Owing to the hypotheses on $\vec{n}$ and $\vec{H}$, this equation has a distributional sense. Since $\mathscr{L}(\vec{H})$ is supported on the origin and belongs to $W^{-1,p}$ for $p<2$, it can only be proportional to the Dirac mass δ_0. From (2-7), we deduce more precisely that

$$\mathscr{L}(\vec{H})=4\pi\vec{\gamma}_0\delta_0.$$

Let $\vec{A}\in\mathbb{C}^m$ be the constant vector appearing in the expansion (2-21). Since $\pi_{\vec{n}(0)}\vec{A}=\vec{0}$ (see (2-24)), an elementary computation gives

$$\begin{aligned}4\pi\vec{A}\cdot\vec{\gamma}_0\delta_0&=4\pi\,\pi_T\vec{A}\cdot\vec{\gamma}_0\delta_0=\pi_T\vec{A}\cdot\mathscr{L}(\vec{H})\\&=\operatorname{div}\big(\vec{H}\cdot\nabla\pi_T\vec{A}-\pi_T\vec{A}\cdot\star(\nabla^{\perp}\vec{n}\wedge\vec{H})\big)\\&\qquad+\nabla\pi_T\vec{A}\cdot\big(\nabla\vec{H}-3\pi_{\vec{n}}\nabla\vec{H}+\star(\nabla^{\perp}\vec{n}\wedge\vec{H})\big),\end{aligned}\tag{2-37}$$

where we have used the fact that $\pi_T\vec{H}\equiv\vec{0}$.

Because $\vec{A}$ is constant and $\nabla\vec{n}\in\bigcap_{p<\infty}L^p$, it follows from the fact that $\pi_{\vec{n}}=\vec{n}\llcorner\vec{n}\llcorner$ that $\nabla\pi_{\vec{n}}\vec{A}$ and thus $\nabla\pi_T\vec{A}$ lie in $\bigcap_{p<\infty}L^p$. Moreover, $\nabla\vec{H}\in\bigcap_{1\le p<2}L^p$ by hypothesis. Introducing this information into (2-37), we note that its right-hand side belongs to $W^{-1,p}$ for all $p<\infty$. Yet, its left-hand side is proportional to the Dirac mass, which does *not* belong to any $W^{-1,p}$ for $p\ge 2$. We accordingly conclude that $\vec{A}\cdot\vec{\gamma}_0=0$. Returning to the expansion (2-21) reveals now that

$$\vec{\gamma}_0\cdot\begin{pmatrix}\vec{e}_1(x)\\\vec{e}_2(x)\end{pmatrix}\simeq\vec{\gamma}_0\cdot\vec{T}(x)=\mathrm{O}(|x|^{1-\epsilon})\quad\text{for all }\epsilon>0,$$

whence

$$|x|^{-1}\pi_T(\vec{\gamma}_0)\in\bigcap_{p<\infty}L^p(B_1(0)).\tag{2-38}$$

A direct computation gives

$$\begin{aligned}\mathscr{L}(\vec{\gamma}_0\log|x|)&=-4\pi\vec{\gamma}_0\delta_0+\operatorname{div}\big(3\pi_T(\vec{\gamma}_0)\nabla\log|x|+\star(\nabla^{\perp}\vec{n}\wedge\vec{\gamma}_0)\log|x|\big)\\&=-\mathscr{L}(\vec{H})+\operatorname{div}\big(3\pi_T(\vec{\gamma}_0)\nabla\log|x|+\star(\nabla^{\perp}\vec{n}\wedge\vec{\gamma}_0)\log|x|\big).\end{aligned}$$

Using the fact that $\nabla\vec{n} \in \bigcap_{p<\infty} L^p$ and (2-38) shows that

$$\mathscr{L}(\vec{H} + \vec{\gamma}_0 \log|x|) \in \bigcap_{p<\infty} W^{-1,p}.$$

It is established in [Rivière 2008] that the operator $\mathscr{L}$ is second-order elliptic and in particular that it satisfies $\mathscr{L}^{-1} W^{-1,p} \subset W^{1,p}$. The desired claim ensues:

$$\vec{H}(x) + \vec{\gamma}_0 \log|x| \in \bigcap_{p<\infty} W^{1,p}. \qquad \square$$

We continue our study of the case $\theta_0 = 1$ by a slight improvement on the regularity of the Gauss map $\vec{n}$. It is shown as (B-7) in the appendix that the $\bigwedge^{m-2}(\mathbb{S}^{m-1})$-valued Gauss map $\vec{n}$ satisfies a perturbed harmonic map equation, namely

$$\Delta\vec{n} = 2\star(\nabla^\perp\vec{\Phi}\wedge\nabla\vec{H}) + 2\mathrm{e}^{2\lambda}K\vec{n} - 2\star\mathrm{e}^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22}), \tag{2-39}$$

where K denotes the Gauss curvature. Recall that

$$|\nabla\vec{n}| = \mathrm{e}^{-\lambda}|\pi_{\vec{n}}\nabla^2\vec{\Phi}| = \mathrm{e}^{\lambda}\begin{vmatrix}\vec{h}_{11} & \vec{h}_{12}\\ \vec{h}_{21} & \vec{h}_{22}\end{vmatrix},$$

so that $\mathrm{e}^{\lambda}\vec{h}_{ij}$ inherits the regularity of $\nabla\vec{n} \in \bigcap_{p<\infty} L^p$. Bringing this information and the expansion given in Proposition 2.2 into (2-39) shows that

$$|\Delta\vec{n}| \lesssim |x|^{-1} + \text{terms in} \bigcap_{p<\infty} L^p \in L^{2,\infty}.$$

Hence $\nabla^2\vec{n} \in L^{2,\infty}$, and thus $\nabla\vec{n} \in \mathrm{BMO}$.

2D. ***The case $\theta_0 \geq 2$.*** We now return to (2-20) in the case when $\theta_0 \geq 2$. Setting

$$\vec{F}_1 := \nabla^\perp\vec{R} + \nabla\vec{G} \quad \text{and} \quad F_2 := \nabla^\perp S + \nabla g, \tag{2-40}$$

it reads

$$-2\Delta\vec{\Phi} = F_2\cdot\nabla\vec{\Phi} - \vec{F}_1 \bullet \nabla\vec{\Phi}. \tag{2-41}$$

Owing to (2-14) and (2-29), the functions $\vec{F}_1$ and F_2 are Hölder continuous of any order $\alpha \in (0,1)$. It thus makes sense to define the constants

$$\vec{f}_1 := \vec{F}_1(0) \quad \text{and} \quad f_2 := F_2(0).$$

They are elements of $\mathbb{R}^2 \otimes \bigwedge^2(\mathbb{R}^m)$ and of $\mathbb{R}^2$, respectively. We will in the sequel view $\vec{f}_1$ as an element of $\mathbb{C}\otimes\bigwedge^2(\mathbb{R}^m)$ and f_2 as an element of $\mathbb{C}$.

For future purposes, let us define $\vec{\Gamma}$ via

$$\Delta\vec{\Gamma} = 4\theta_0\Re(\vec{C}^*_{\theta_0-1}x^{\theta_0-1}) \quad \text{with } -8\vec{C}^*_{\theta_0-1} := f_2\vec{A} - \vec{f}_1 \bullet \vec{A}, \tag{2-42}$$

where $\vec{A}\in\mathbb{C}^m$ is the constant vector appearing in (2-21). More precisely, writing $\vec{f}_1=\vec{f}_1^{\,1}+i\vec{f}_1^{\,2}$ and $f_2=f_2^1+if_2^2$, then

$$\begin{cases}-8\Re(\vec{C}^*_{\theta_0-1})=f_2^1\vec{A}^1-f_2^2\vec{A}^2-\vec{f}_1^{\,1}\bullet\vec{A}^1+\vec{f}_1^{\,2}\bullet\vec{A}^2,\\ -8\Im(\vec{C}^*_{\theta_0-1})=f_2^1\vec{A}^2+f_2^2\vec{A}^1-\vec{f}_1^{\,1}\bullet\vec{A}^2-\vec{f}_1^{\,2}\bullet\vec{A}^1.\end{cases}$$

Equation (2-42) is solved explicitly (up to an unimportant harmonic function):

$$\vec{\Gamma}(x)=\Re(\vec{C}_{\theta_0-1}|x|^{2\theta_0}x^{1-\theta_0}),\tag{2-43}$$

where $\vec{C}_{\theta_0-1}$ is the complex conjugate of $\vec{C}^*_{\theta_0-1}$.

Note next that (2-41) and (2-42) yield

$$2\Delta(\vec{\Phi}-\vec{\Gamma})=(F_2-f_2)\cdot\nabla\vec{\Phi}-(\vec{F}_1-\vec{f}_1)\bullet\nabla\vec{\Phi}+\mathrm{e}^{\lambda}[f_2\cdot\vec{T}-\vec{f}_1\bullet\vec{T}],\tag{2-44}$$

where we have used the representation (2-21).

Since $|\vec{T}(x)|\lesssim|x|^{1-\epsilon}$ for all $\epsilon>0$, and $|\vec{F}_1(x)-\vec{f}_1|+|F_2(x)-f_2|\lesssim|x|^{\alpha}$ for all $\alpha\in(0,1)$, while, as previously explained, the weight $|\nabla\vec{\Phi}(x)|\simeq\mathrm{e}^{\lambda}$ satisfies the condition (C-13), we may apply Corollary C.3 to (2-44), thereby obtaining

$$\nabla(\vec{\Phi}-\vec{\Gamma})(x)=\vec{P}(\bar{x})+\mathrm{e}^{\lambda(x)}\vec{U}(x),\tag{2-45}$$

$$\nabla^2(\vec{\Phi}-\vec{\Gamma})(x)=\nabla\vec{P}(\bar{x})+\mathrm{e}^{\lambda(x)}\vec{V}(x),\tag{2-46}$$

where $\vec{P}$ is a polynomial of degree at most θ_0. Moreover, $\vec{U}(x)=\mathrm{O}(|x|^{2-\epsilon})$, and

$$|x|^{-(1-\epsilon)}\vec{V}\in\bigcap_{p<\infty}L^p\quad\text{with }\operatorname{Tr}\vec{V}(x)=\mathrm{O}(|x|^{1-\epsilon})\quad\text{for all }\epsilon>0.\tag{2-47}$$

One sees in (2-43) that $\nabla\vec{\Gamma}(x)=\mathrm{O}(|x|^{\theta_0})$. Hence, from (2-45) and the fact that $|\nabla\vec{\Phi}|(x)\simeq|x|^{\theta_0-1}$, it follows that the polynomial $\vec{P}$ contains exactly one monomial of degree (θ_0-1) and one monomial of degree θ_0. More precisely, identifying the representations (2-27) and (2-46) yields

$$\nabla^2\vec{\Phi}(x)=\begin{pmatrix}-\Re&\Im\\ \Im&\Re\end{pmatrix}\big(\theta_0(1-\theta_0)\vec{A}x^{\theta_0-2}-\theta_0(1+\theta_0)\vec{B}_1x^{\theta_0-1}\big)+\nabla^2\vec{\Gamma}(x)+\mathrm{e}^{\lambda(x)}\vec{V}(x),\tag{2-48}$$

where $\vec{B}_1\in\mathbb{C}^m$ is constant. The constant vector $\vec{A}$ is as in (2-21).

Remark 2.3. The estimates (2-47) may be slightly improved. To do so, one differentiates (2-44) throughout with respect to x_j, and applies Proposition C.2(i) and Corollary C.3 to the resulting equation (however, Proposition C.2(ii) will not be available, as some of the weights appearing — of the order $|\nabla^2\vec{\Phi}|$ — need not *a priori* fulfill the condition (C-13)). One eventually obtains that

$$\vec{V}(x) = \mathrm{O}(|x|^{1-\epsilon}) \quad \text{for all } \epsilon > 0. \tag{2-49}$$

Now recall that $|\nabla\vec{\Phi}(0)| = 0 = |\vec{\Phi}(0)|$. We deduce from (2-48) and (2-43) that

$$\vec{\Phi}(x) = \Re\big(\vec{A}x^{\theta_0} + \vec{B}_1 x^{\theta_0+1} + \vec{C}_{\theta_0-1}|x|^{2\theta_0}x^{1-\theta_0}\big) + |x|^{\theta_0-1}\vec{\xi}(x), \tag{2-50}$$

where

$$\vec{\xi}(x) = \mathrm{O}(|x|^{3-\epsilon}), \quad \nabla\vec{\xi}(x) = \mathrm{O}(|x|^{2-\epsilon}), \quad \nabla^2\vec{\xi}(x) = \mathrm{O}(|x|^{1-\epsilon}) \quad \text{for all } \epsilon > 0.$$

Moreover, as $2\mathrm{e}^{2\lambda}\vec{H} = \Delta\vec{\Phi}$, the representation (2-48) along with (2-42) gives the local asymptotic expansion

$$\vec{H}(x) = 2\theta_0\mathrm{e}^{2u(x)}\Re(\vec{C}_{\theta_0-1}x^{1-\theta_0}) + \mathrm{O}(|x|^{2-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0, \tag{2-51}$$

where $\vec{C}_{\theta_0-1}$ is as above, and $\mathrm{e}^{u(x)} := |x|^{1-\theta_0}\mathrm{e}^{\lambda(x)}$, which is known to have a positive limit at the origin. This shows in particular that $\mathrm{e}^{\lambda(x)}\vec{H}(x)$ is a bounded function (unlike the case $\theta_0 = 1$, whereby $\mathrm{e}^{\lambda}\vec{H} \simeq \vec{H}$ behaves logarithmically). However, it "spins" as x approaches the origin: its limit need not exist, and, if it does exist, then it must be zero (i.e., $\vec{C}_{\theta_0-1} = \vec{0}$). Note in addition that, because $\vec{H}$ is a normal vector, we have always $\pi_{\vec{n}(0)}\vec{C}_{\theta_0-1} = \vec{C}_{\theta_0-1}$.

We close this section by proving that $\nabla^2\vec{n} \in L^{2,\infty}$ and that $\nabla\vec{n} \in L^{\infty}$. We have seen that $\mathrm{e}^{\lambda}\vec{H}$ is bounded. Applying standard elliptic techniques to (1-7) then yields that $|x|\mathrm{e}^{\lambda}\nabla\vec{H}$ is bounded as well, and hence that $\mathrm{e}^{\lambda}\nabla\vec{H} \in L^{2,\infty}$. Going back to the perturbed harmonic map equation (2-39) satisfied by the Gauss map $\vec{n}$, and using the fact that $\mathrm{e}^{\lambda}\vec{h}_{ij}$ inherits the regularity of $\nabla\vec{n} \in \bigcap_{p<\infty} L^p$, we deduce that $\Delta\vec{n}$ lies in $L^{2,\infty}$, and therefore indeed that $\nabla^2\vec{n} \in L^{2,\infty}$. In particular, this implies that $\nabla\vec{n} \in \mathrm{BMO}$. However, it is possible to show that $\nabla\vec{n} \in L^{\infty}(B_1(0))$. To see this, we first note that (2-50) yields

$$\nabla\vec{\Phi}(x) = \begin{pmatrix} \Re \\ -\Im \end{pmatrix}(\theta_0\vec{A}x^{\theta_0-1}) + \nabla\big(|x|^{\theta_0-1}\vec{\xi}(x)\big) + \mathrm{O}(|x|^{\theta_0}).$$

Since $\pi_{\vec{n}}\nabla\vec{\Phi} \equiv 0$, the latter and the estimates on $\vec{\xi}$ give

$$|\pi_{\vec{n}(x)}\vec{A}| = \mathrm{O}(|x|). \tag{2-52}$$

A quick inspection of the identity (2-48) then reveals that

$$|\pi_{\vec{n}(x)}\nabla^2\vec{\Phi}(x)| \lesssim |\pi_{\vec{n}(x)}\vec{A}|\,|x|^{\theta_0-2} + \mathrm{O}(|x|^{\theta_0-1}) = \mathrm{O}(|x|^{\theta_0-1}).$$

Combining this with (2-26) gives that $\nabla\vec{n}$ is bounded across the singularity.

2D1. *An observation.* In this section, we adopt the previously encountered complex notation $x := x_1 + i x_2$ and $\bar{x} := x_1 - i x_2$. We set $\partial_x := \frac{1}{2}(\partial_{x_1} - i\partial_{x_2})$ and $\partial_{\bar{x}} := \frac{1}{2}(\partial_{x_1} + i\partial_{x_2})$. We may then deduce from (2-50) that

$$2\partial_x \vec{\Phi} = \theta_0 \vec{A} x^{\theta_0 - 1} + \mathrm{O}(|x|^{\theta_0}),$$

and thus

$$\vec{A} = \frac{2}{\theta_0} \lim_{x \to 0} x^{1-\theta_0} \partial_x \vec{\Phi}. \tag{2-53}$$

On the other hand, when $\theta_0 \geq 2$, recalling (2-40) and (2-17), we have

$$\vec{F}_1 := \nabla^{\perp} \vec{R} + \nabla \vec{G} = \vec{L} \wedge \nabla^{\perp} \vec{\Phi} - 2\vec{H} \wedge \nabla \vec{\Phi} \equiv 2i(\vec{L} + 2i\vec{H}) \wedge \partial_{\bar{x}} \vec{\Phi}, \tag{2-54}$$

$$F_2 := \nabla^{\perp} S + \nabla g = \vec{L} \cdot \nabla^{\perp} \vec{\Phi} \equiv 2i\vec{L} \cdot \partial_{\bar{x}} \vec{\Phi}. \tag{2-55}$$

From (2-42), we now find

$$\begin{aligned} -8\vec{C}^{*}_{\theta_0-1} &:= F_2(0)\vec{A} - \vec{F}_1(0) \bullet \vec{A} \\ &= \frac{4i}{\theta_0} \lim_{x\to 0} x^{1-\theta_0} [\vec{L} \cdot \partial_{\bar{x}} \vec{\Phi}] \partial_x \vec{\Phi} - \frac{4i}{\theta_0} \lim_{x\to 0} x^{1-\theta_0} \big[(\vec{L} + 2i\vec{H}) \wedge \partial_{\bar{x}} \vec{\Phi}\big] \bullet \partial_x \vec{\Phi}. \end{aligned}$$

Rearranging the computations leading to the identities (B-11) yields without much effort that

$$(\vec{V} \wedge \partial_{\bar{x}} \vec{\Phi}) \bullet \partial_x \vec{\Phi} = (\vec{V} \cdot \partial_{\bar{x}} \vec{\Phi}) \partial_x \vec{\Phi} + \tfrac{1}{2} \mathrm{e}^{2\lambda} \vec{V}$$

holds for all 1-vectors $\vec{V}$. As $\vec{H}$ is a normal vector, we thus find

$$-8\vec{C}^{*}_{\theta_0-1} = -\frac{2i}{\theta_0} \lim_{x\to 0} x^{1-\theta_0} \mathrm{e}^{2\lambda} (\vec{L} + 2i\vec{H}).$$

Introducing, as in (2-51), the function $\mathrm{e}^{u(x)} := |x|^{1-\theta_0} \mathrm{e}^{\lambda(x)}$, which is known to be continuous, bounded from above and below across the origin, we reach the expression

$$-\vec{C}_{\theta_0-1} = \frac{\mathrm{e}^{2u(0)}}{4\theta_0} \lim_{x\to 0} x^{\theta_0-1} (2\vec{H} + i\vec{L}). \tag{2-56}$$

The importance of the function $2\vec{H} + i\vec{L}$ further arises in Section 2E.

2E. *Removability results.*

2E1. *Preparation.* We now return to the defining equation for $\vec{L}$, namely

$$\nabla^{\perp} \vec{L} := \nabla \vec{H} - 3\pi_{\vec{n}} \nabla \vec{H} + \star(\nabla^{\perp} \vec{n} \wedge \vec{H}) - 2\vec{\gamma}_0 \nabla \log |x|.$$

We will first recast this equation in the form[17]

$$\nabla^{\perp}\vec{L} = -2\nabla\vec{H} + 3\pi_T\nabla\vec{H} - \star(\vec{n}\wedge\pi_T\nabla^{\perp}\vec{H}) - 2\vec{\gamma}_0\nabla\log|x|, \tag{2-57}$$

where we have used the fact that $\vec{H}$ is a normal vector, so $\vec{n}\wedge\vec{H}=\vec{0}$. In turn, the latter is equivalently expressed as[18]

$$\partial_x(\vec{L} + 2i\vec{H} + 2i\vec{\gamma}_0\log|x|) = 3i\pi_T\partial_x\vec{H} - \star(\vec{n}\wedge\pi_T\partial_x\vec{H}). \tag{2-58}$$

Using the fact that $\vec{H}$ is normal and (B-2), a simple computation reveals that

$$\pi_T\partial_x\vec{H} = -\sum_{j=1,2}(\vec{H}\cdot\pi_{\vec{n}}\partial_x\vec{e}_j)\vec{e}_j = -(\vec{H}\cdot\vec{H})\partial_x\vec{\Phi} - \big(\vec{H}\cdot\overline{\vec{H}_0}\big)\partial_{\bar{x}}\vec{\Phi}, \tag{2-59}$$

where $\vec{H}_0$ denotes the Weingarten operator:

$$\vec{H}_0 := \tfrac{1}{2}(\vec{h}_{11} - \vec{h}_{22} - 2i\vec{h}_{12}).$$

From this and the elementary identities

$$\star(\vec{n}\wedge\partial_x\vec{\Phi}) = i\,\partial_x\vec{\Phi} \quad\text{and}\quad \star(\vec{n}\wedge\partial_{\bar{x}}\vec{\Phi}) = -i\,\partial_{\bar{x}}\vec{\Phi},$$

we obtain

$$\star(\vec{n}\wedge\pi_T\partial_x\vec{H}) = -i(\vec{H}\cdot\vec{H})\partial_x\vec{\Phi} + i(\vec{H}\cdot\vec{H}_0)\partial_{\bar{x}}\vec{\Phi}. \tag{2-60}$$

Altogether (2-59)–(2-60) brought into (2-58) give

$$\partial_x\big(i\vec{L} - 2\vec{H} - 2\vec{\gamma}_0\log|x|\big) = 2(\vec{H}\cdot\vec{H})\partial_x\vec{\Phi} + 4(\vec{H}\cdot\vec{H}_0)\partial_{\bar{x}}\vec{\Phi}.$$

This equation, like the original one introducing $\vec{L}$, is valid only on the punctured disk $D^2\setminus\{0\}$. For notational convenience, we will henceforth write it

$$\partial_x\big(i\vec{L} - 2\vec{H} - 2\vec{\gamma}_0\log|x|\big) = 2\vec{q}. \tag{2-61}$$

Owing to the fact that $|\vec{H}||\nabla\vec{\Phi}|$ and $|\vec{H}_0||\nabla\vec{\Phi}|$ are controlled by $|\nabla\vec{n}|$, we note that

$$|\vec{q}| \lesssim |\nabla\vec{n}||\vec{H}|. \tag{2-62}$$

Lemma 2.4. *If, locally around the origin, for some integer $k\in\{1,\dots,\theta_0\}$, we have*

$$\vec{H} = \mathrm{O}(|x|^{k-\theta_0-\epsilon}) \quad \textit{for all } \epsilon > 0, \tag{2-63}$$

[17] Recall that $\pi_T := \mathrm{id} - \pi_{\vec{n}}$ denotes projection onto the tangent space.

[18] With the same notation as in Section 2D1.

then we have

$$\vec{H}+\vec{\gamma}_0\log|x|-\frac{i}{2}\vec{L}=\vec{E}-\vec{T}. \tag{2-64}$$

The function $\vec{E}$ is antimeromorphic with a pole at the origin of order at most (θ_0-k). Moreover,

$$\partial_x\vec{T}=\vec{q}\quad \textit{on } D^2\setminus\{0\},\qquad \vec{T}=\mathrm{O}(|x|^{1+k-\theta_0-\epsilon})\quad \textit{for all } \epsilon>0.$$

The function $\vec{T}$ is unique up to addition of antimeromorphic summands.

Proof. Suppose that for some integer $k\in\{1,\dots,\theta_0\}$ we have

$$\vec{H}=\mathrm{O}(|x|^{k-\theta_0-\epsilon})\quad \text{for all } \epsilon>0.$$

Owing to (2-63), we have as well

$$|x|^{\theta_0-k}\vec{q}\in\bigcap_{p<\infty}L^p. \tag{2-65}$$

We consider any $\vec{w}$ satisfying

$$\partial_x\vec{w}=2\bar{x}^{\theta_0-k}\vec{q}\quad \text{on } D^2. \tag{2-66}$$

Per (2-65), $\vec{w}$ is $C^{0,1-\epsilon}$-Hölder continuous for any $\epsilon>0$. From (2-61), we have

$$\partial_x\big[\bar{x}^{\theta_0-k}(i\vec{L}-2\vec{H}-2\vec{\gamma}_0\log|x|)-\vec{w}\big]=0\quad \text{on } D^2\setminus\{0\}. \tag{2-67}$$

We will extend this equation to all of the unit disk D^2. To do so, it suffices to show that the function to which the operator ∂_x is applied on the left-hand side of (2-67) lies in L^2. Since $\vec{w}$ is Hölder continuous, while $\vec{H}$ satisfies (2-63), it only remains to verify that $|x|^{\theta_0-k}\vec{L}$ lies in the space L^2. Exactly as we derived (2-12) from (2-5), we infer here that $|x|^{\theta_0+1-k}\nabla\vec{H}\in\bigcap_{p<\infty}L^p$, and then per (2-57) that

$$|x|^{\theta_0+1-k}\nabla\vec{L}\in\bigcap_{p<\infty}L^p, \tag{2-68}$$

from which we obtain that $|x|^{\theta_0-k}\vec{L}\in L^2$. Accordingly, (2-67) holds on the unit disk, whence

$$\vec{H}+\vec{\gamma}_0\log|x|-\frac{i}{2}\vec{L}=\vec{P}-\bar{x}^{k-\theta_0}\vec{w},$$

where $\vec{P}$ is antimeromorphic with a pole at the origin of order at most (θ_0-k). Putting in the latter

$$\vec{E}:=\vec{P}+\bar{x}^{k-\theta_0}\vec{w}(0)\quad \text{and}\quad \vec{T}:=(\vec{w}-\vec{w}(0))\bar{x}^{k-\theta_0}$$

gives the desired representation (2-64). Moreover, we have

$$\partial_x \vec{T} = \vec{q} \quad \text{on } D^2 \setminus \{0\}, \qquad \vec{T} = \mathrm{O}(|x|^{1+k-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0.$$

The function $\vec{w}$ is clearly unique up to addition of antimeromorphic terms. The same is also true for $\vec{T}$. Should the "first" found $\vec{T}$ happen to contain an antimeromorphic summand, it will necessarily be of order at most $\mathrm{O}(|x|^{1+k-\theta_0})$ and could thus safely be fed into $\vec{E}$ without affecting the desired statement. □

We now come to a central result in our study.

Proposition 2.5. *There exists a unique function $\vec{T}$ containing no monomial of $\bar{x}$, satisfying*

$$\partial_x \vec{T} = \vec{q} \quad \text{on } D^2 \setminus \{0\} \quad \text{and} \quad \vec{T} = \mathrm{O}(|x|^{2-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0, \tag{2-69}$$

and such that locally around the singularity, we have

$$\vec{H}(x) + \vec{\gamma}_0 \log |x| - \frac{i}{2}\vec{L}(x) = \vec{E}(x) - \vec{T}(x), \tag{2-70}$$

where $\vec{\gamma}_0$ is the residue defined in (1-8), while the function $\vec{E}$ is antiholomorphic with possibly a pole at the origin of order at most $(\theta_0 - 1)$.

If the singularity of $\vec{E}$ at the origin has order $\alpha \in \{0, \dots, \theta_0 - 1\}$, then $\vec{E}$ and $\vec{T}$ may be adjusted to satisfy

$$\vec{E} - \vec{T} = \vec{E}_\alpha \bar{x}^{-\alpha} - \vec{Q}$$

for some nonzero constant $\vec{E}_\alpha \in \mathbb{C}^m$, and with

$$\partial_x \vec{Q} = \vec{q} \quad \text{on } D^2 \setminus \{0\} \quad \text{and} \quad \vec{Q} = \mathrm{O}(|x|^{1-\alpha-\epsilon}) \quad \text{for all } \epsilon > 0.$$

Proof. We have seen in Proposition 2.2 and in (2-51) that $\vec{H} = \mathrm{O}(|x|^{1-\theta_0-\epsilon})$ for all $\epsilon > 0$. The desired representation (2-70) was obtained in Lemma 2.4.

For simplicity, we will only prove the second part of the lemma for the first three cases $\alpha \in \{\theta_0 - 3, \dots, \theta_0 - 1\}$. All other cases are obtained *mutatis mutandis*.

Case $\alpha = \theta_0 - 1$. We can write locally

$$\vec{E} = \vec{E}_{\theta_0 - 1}\bar{x}^{1-\theta_0} + \vec{E}^0,$$

where $\vec{E}_{\theta_0-1} \in \mathbb{C}^m$ is constant, and $\vec{E}^0$ is an antimeromorphic function with a pole at the origin of order at most $(\theta_0 - 2)$; i.e., $|\vec{E}^0| \lesssim |x|^{2-\theta_0}$. We may then let $\vec{Q} := \vec{T} - \vec{E}^0$ with $\partial_x \vec{Q} = \partial_x \vec{T}$ on $D^2 \setminus \{0\}$, and $\vec{Q}$ and $\vec{T}$ have the same asymptotic behavior at the origin.

Case $\alpha = \theta_0 - 2$. In this case, without loss of generality, $\theta_0 \geq 2$, so that $\nabla\vec{n} \in L^\infty$. As $\vec{E}$ is antimeromorphic with a pole of order $(\theta_0 - 2)$ at the origin, we have $|\vec{E}| \simeq |x|^{2-\theta_0}$ near the origin. The second condition in (2-69) put into (2-70) shows that $\vec{H}$ is controlled by $|x|^{2-\theta_0-\epsilon}$ for all $\epsilon > 0$. Calling upon the representation (2-64) with $k = 2$ gives that

$$-\frac{i}{2}\vec{L} + \vec{H} + \vec{\gamma}_0 \log|x| = \vec{E}^1 - \vec{T}^1,$$

where $\vec{E}^1$ is an antimeromorphic function with a pole at the origin of order at most $(\theta_0 - 2)$, and a function $\vec{T}^1$ satisfies

$$\partial_x \vec{T}^1 = \vec{q} \quad \text{and} \quad \vec{T}^1 = \mathrm{O}(|x|^{3-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0. \tag{2-71}$$

As we did in the case $\alpha = \theta_0 - 1$, we can write

$$\vec{E}^1(\bar{x}) = \vec{E}_{\theta_0-2}\bar{x}^{2-\theta_0} + \vec{E}^0(\bar{x}),$$

where $\vec{E}_{\theta_0-2} \in \mathbb{C}^m$ is constant, and $\vec{E}^0$ is an antimeromorphic function with a pole at the origin of order at most $(\theta_0 - 3)$. Clearly, the function $\vec{Q} := \vec{T}^1 - \vec{E}^0$ satisfies the two conditions (2-71). Furthermore, we have

$$-\frac{i}{2}\vec{L} + \vec{H} + \vec{\gamma}_0 \log|x| = \vec{E}^1 - \vec{T}^1 = \vec{E}_{\theta_0-2}\bar{x}^{2-\theta_0} - \vec{Q},$$

as desired.

Case $\alpha = \theta_0 - 3$. We start with the representation (2-64) with $k = 1$, which as we have seen is equivalent to (2-70):

$$\vec{H} + \vec{\gamma}_0 \log|x| - \frac{i}{2}\vec{L} = \vec{E} - \vec{T},$$

and assume that the antimeromorphic function $\vec{E}$ has a pole of order $(\theta_0 - 3)$ at the origin, while $\vec{T} = \mathrm{O}(|x|^{2-\theta_0-\epsilon})$ for all $\epsilon > 0$. Exactly as we did in the case $\alpha = \theta_0 - 2$, we obtain

$$\vec{H} = \mathrm{O}(|x|^{2-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0.$$

Calling upon Lemma 2.4 with $k = 2$ gives us the alternative representation

$$-\frac{i}{2}\vec{L} + \vec{H} + \vec{\gamma}_0 \log|x| = \vec{E}^1 - \vec{T}^1, \tag{2-72}$$

where $\vec{E}^1$ is an antimeromorphic function with a pole at the origin of order at most $(\theta_0 - 2)$, while $\vec{T}^1 = \mathrm{O}(|x|^{3-\theta_0-\epsilon})$. Hence,

$$\vec{T} \equiv \vec{E} - \vec{E}^1 + \vec{T}^1 = -\vec{E}^1 + \mathrm{O}(|x|^{3-\theta_0-\epsilon}).$$

If $\vec{E}^1$ had a pole of order (θ_0-2), then $\vec{T}$ would contain a monomial term of $\bar{x}$, which we have ruled out by hypothesis. Thus, the pole of $\vec{E}^1$ has order at most (θ_0-3). The representation (2-72) then yields

$$\vec{H} = \mathrm{O}(|x|^{3-\theta_0-\epsilon}) \quad \text{for all } \epsilon > 0.$$

Finally, calling once more upon Lemma 2.4 with this time $k=3$ gives us the desired representation. □

The regularity of the function $\vec{Q}$ is closely tied to that of $\vec{H}$ (and ultimately to that of the Gauss map). A quick inspection of the proof of Proposition 2.5 reveals that, if the local behavior of the mean curvature improves to $\vec{H} = \mathrm{O}(|x|^{-\alpha})$ for some $\alpha \in \{0,\dots,\theta_0-2\}$, then we get the corresponding improvement $\vec{Q} = \mathrm{O}(|x|^{-\alpha+1-\epsilon})$ for all $\epsilon > 0$. In this case, the order of the pole of the antimeromorphic function $\vec{E}$ is at most α. On the other hand, if $\vec{E}$ happens to be regular at the origin, the identity (2-70) shows that the regularity of $\vec{H}$ improves with that of $\vec{Q}$, a condition which, per our previous observation, makes it possible to implement a bootstrapping procedure. The obstruction induced by the singular behavior of the function $\vec{E}$ at the origin is studied in detail in the next section. We view $\vec{E}$ as a string of m complex-valued functions $\{E_j\}_{j=1,\dots,m}$, all of which are antimeromorphic and may have a pole at the origin of order at most (θ_0-1). In particular, we define the $\mathbb{N}^m$-valued *second residue*

$$\vec{\gamma} = (\gamma_1,\dots,\gamma_m) \quad \text{with } \gamma_j := \frac{1}{2i\pi}\int_{\partial D^2} d\log E_j. \tag{2-73}$$

Remark 2.6. Branched minimal surfaces have vanishing first and second residues. Indeed, if $\vec{H} \equiv \vec{0}$, from the very definition (1-8) of the first residue, we see that $\vec{\gamma}_0 = \vec{0}$. Furthermore, the function $\vec{q}$ introduced in (2-61) is identically vanishing, thereby yielding that $\vec{Q} \equiv \vec{0}$. According to (2-70), we have $\vec{E} = -\frac{1}{2}i\vec{L}$. But, as seen in (2-9), the function $\vec{L}$ must be constant when $\vec{H} \equiv \vec{0}$. The function $\vec{E}$ is thus regular, and hence the second residue $\vec{\gamma}$ vanishes.

2E2. *How the second residue $\vec{\gamma}$ controls the regularity.* We start by defining

$$\alpha := \max_{1\le j\le m} \gamma_j \in \{0,\dots,\theta_0-1\}.$$

Per Proposition 2.5, we may choose $\vec{E} = \vec{E}_\alpha \bar{x}^{-\alpha}$ for some constant vector $\vec{E}_\alpha \in \mathbb{C}^m$. According to Proposition 2.5, we have

$$\vec{Q} = \mathrm{O}(|x|^{1-\alpha-\epsilon}) \quad \text{for all } \epsilon > 0. \tag{2-74}$$

Because $\vec{L}$ is real-valued, (2-70) yields

$$\vec{H} + \vec{\gamma}_0 \log|x| = \Re(\vec{E} - \vec{Q}). \tag{2-75}$$

We define next the two-component vector field $\vec{U} := (\Re, \Im)(e^{\lambda}\vec{Q})$. As $\partial_x \vec{Q} = \vec{q}$, we have[19]

$$\text{(2-76)} \qquad \operatorname{div}\vec{U} = \nabla\lambda\cdot\vec{U} + 2e^{\lambda}\Re(\vec{q}), \quad \operatorname{curl}\vec{U} = \nabla^{\perp}\lambda\cdot\vec{U} + 2e^{\lambda}\Im(\vec{q}).$$

As $|\nabla\lambda| \lesssim |x|^{-1}$, we use (2-75) along with the estimates (2-62) and (2-74) to find

$$\text{(2-77)} \quad |\operatorname{div}\vec{U}| + |\operatorname{curl}\vec{U}| \lesssim |x|^{-1}|\vec{U}| + e^{\lambda}|x|^{-\epsilon}|\vec{H}| \lesssim |x|^{\theta_0-1-\alpha-\epsilon} \quad \text{for all } \epsilon > 0.$$

With the help of a simple Hodge decomposition, (2-77) along with the fact that $|\vec{U}| \simeq e^{\lambda}|\vec{Q}| = \mathrm{O}(|x|^{\theta_0-\epsilon})$ yields

$$|\nabla(e^{\lambda}\vec{Q})| \simeq |\nabla\vec{U}| \lesssim |x|^{\theta_0-1-\epsilon} \quad \text{for all } \epsilon > 0.$$

Again since $|\nabla\lambda| \lesssim |x|^{-1}$, the latter shows that

$$\text{(2-78)} \qquad |\nabla\vec{Q}| \lesssim |x|^{-\epsilon} \quad \text{for all } \epsilon > 0.$$

Putting (2-78) into (2-75), and recalling that $\vec{E}$ is a power function, then yields

$$\text{(2-79)} \qquad |\nabla\vec{H}| \lesssim |x|^{-1-\alpha}.$$

As $\alpha \le \theta_0 - 1$, we thus find $e^{\lambda}\nabla\vec{H} \in L^{2,\infty}$. It is proved as (B-8) in the appendix that the $\Lambda^{m-2}(\mathbb{S}^{m-1})$-valued Gauss map $\vec{n}$ satisfies the perturbed harmonic map equation

$$\text{(2-80)} \qquad \Delta\vec{n} - 2e^{2\lambda}K\vec{n} = 2\star(\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H}) - 2\star e^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22}),$$

where K is the Gauss curvature, whence

$$\text{(2-81)} \qquad |\Delta\vec{n}| \lesssim e^{\lambda}|\nabla\vec{H}| + |\nabla\vec{n}|^2 \lesssim |x|^{\theta_0-2} \in L^{2,\infty}.$$

Accordingly, $\nabla^2\vec{n} \in L^{2,\infty}$, and in particular $\nabla\vec{n} \in \mathrm{BMO}$.

We have seen in the Introduction that the conformal parameter satisfies

$$\text{(2-82)} \qquad \lambda = (\theta_0 - 1)\log|x| + u,$$

where the function u belongs to $W^{2,1}$. More precisely, from the Liouville equation, we know that

$$-\Delta u = e^{2\lambda}K,$$

with K denoting the Gauss curvature. As explained, $e^{2\lambda}K$ inherits the regularity of $|\nabla\vec{n}|^2$ (as it is made of products of terms of the type $e^{\lambda}\vec{h}_{ij}$, each of which inherits the regularity of $|\nabla\vec{n}|$). Owing to (2-81), we thus have $\nabla^2 u \in \bigcap_{p<\infty} L^p$, and in particular

[19] Although the equation for $\vec{Q}$ holds only on $D^2\setminus\{0\}$, the system for $\vec{U}$ may easily be extended to the whole unit disk D^2 owing to the fact that $\vec{U} = \mathrm{O}(|x|^{1-\epsilon}) \in L^{\infty}$.

that ∇u is Hölder continuous. Hence, (2-82) shows that

$$|\nabla\lambda| \lesssim |x|^{-1}. \tag{2-83}$$

Furthermore, we may write

$$2e^{2\lambda} = (T_1 + R_1)|x|^{2(\theta_0-1)}, \tag{2-84}$$

where T_1 is the first-order Taylor polynomial expansion of $2e^{2u} \in C^{1,1-\epsilon}$ (for all $\epsilon > 0$) near the origin, and R_1 is the corresponding remainder. Hence

$$\nabla^j R_1 = O(|x|^{2-j-\epsilon}), \quad j \in \{0,1\}, \text{ for all } \epsilon > 0. \tag{2-85}$$

With the help of (2-75), we write

$$\Delta\vec{\Phi} \equiv 2e^{2\lambda}\vec{H} = \Delta\vec{\Phi}_0 + \Delta\vec{\Phi}_1,$$

where

$$\begin{cases} \Delta\vec{\Phi}_0 = T_1|x|^{2(\theta_0-1)}\Re(\vec{E} - \vec{\gamma}_0\log|x|), \\ \Delta\vec{\Phi}_1 = -2e^{2\lambda}\Re(\vec{Q}) + |x|^{2(\theta_0-1)}R_1\Re(\vec{E} - \vec{\gamma}_0\log|x|). \end{cases}$$

Since T_1 and $\vec{E}$ are power functions, we easily obtain via solving explicitly and handling the remainder with Corollary C.3 that

$$\vec{\Phi}_0 = \Re(\vec{P}_0) + C_\alpha|x|^{2\theta_0}\Re(\vec{E}) - \vec{C}|x|^{2\theta_0}(\log|x|^{2\theta_0} - 4) + \vec{\xi}_0,$$

where $\vec{P}_0$ is a $\mathbb{C}^m$-valued holomorphic polynomial of degree at most $(2\theta_0 - \alpha)$, and

$$C_\alpha := \frac{e^{2u(0)}}{2\theta_0(\theta_0-\alpha)} \quad \text{and} \quad \vec{C} := \frac{e^{2u(0)}}{2\theta_0^3}\vec{\gamma}_0. \tag{2-86}$$

The remainder $\vec{\xi}_0$ satisfies

$$\nabla^j\vec{\xi}_0 = O(|x|^{2\theta_0-\alpha+1-j-\epsilon}) \quad \text{for all } j \in \{0,\dots,2\}, \text{ for all } \epsilon > 0,$$
$$|x|^{2+\alpha-2\theta_0}\nabla^3\vec{\xi}_0 \in \bigcap_{p<\infty} L^p.$$

To obtain information on $\vec{\Phi}_1$, we differentiate once its partial differential equation in each coordinate x_1 and x_2, and apply Corollary C.3 to the respective results using (2-85), the fact that $e^\lambda\vec{Q} \in \bigcup_{p<\infty} W^{1,p}$, that $\nabla\lambda = O(|x|^{-1})$, and the fact that $\vec{F}$ is a power function. Without much effort, it ensues that we can write

$$\vec{\Phi}_1 = \Re(\vec{P}_1) + \vec{\xi}_1,$$

where $\vec{P}_1$ is a $\mathbb{C}^m$-valued holomorphic polynomial of degree at most $(2\theta_0 - \alpha)$, and the $\mathbb{R}^m$-valued function $\vec{\xi}_1$ satisfies

$$\nabla^j \vec{\xi}_1 = \mathrm{O}(|x|^{2\theta_0-\alpha+1-j-\epsilon}) \quad \text{for all } j \in \{0,\dots,2\}, \text{ for all } \epsilon > 0,$$
$$|x|^{2+\alpha-2\theta_0} \nabla^3 \vec{\xi}_1 \in \bigcap_{p<\infty} L^p.$$

Comparing $\vec{\Phi}_0 + \vec{\Phi}_1$ to the previously found expression (2-21), we deduce

$$\vec{\Phi} = \Re\big(\vec{A} x^{\theta_0} + \vec{B}_1 x^{\theta_0+1} + C_\alpha |x|^{2\theta_0} \vec{E}\big) - \vec{C} |x|^{2\theta_0} (\log|x|^{2\theta_0} - 4) + (\vec{\xi}_0 + \vec{\xi}_1), \tag{2-87}$$

where $\vec{B}_1 \in \mathbb{C}^m$ is constant, while $\vec{A}$ is as in Proposition 1.3.

Note that

$$|\nabla^j \vec{\Phi}| = \mathrm{O}(|x|^{\theta_0 - j}) \quad \text{for all } j \in \{0,\dots,2\}. \tag{2-88}$$

Suppose next that $\alpha \le \theta_0 - 2$. Then (2-79) gives

$$\mathrm{e}^\lambda \nabla \vec{H} \in L^\infty. \tag{2-89}$$

In turn brought into (2-81), the latter shows that

$$\nabla^2 \vec{n} \in \bigcap_{p<\infty} L^p. \tag{2-90}$$

Accordingly, the function u appearing in (2-82) lies in $C^{2,1-\epsilon}(D^2)$ for all $\epsilon > 0$, whence

$$\nabla^2 \lambda = \mathrm{O}(|x|^{-2}).$$

When $\alpha \le \theta_0 - 2$, we have that $|x|^{-1} \mathrm{e}^\lambda |\vec{E}| \simeq |x|^{\theta_0 - 2 - \alpha} \in L^\infty$. Hence (2-74) and (2-75) yield

$$|x|^{-1} \mathrm{e}^\lambda \vec{H} \in L^\infty. \tag{2-91}$$

We now need to improve the regularity of $\vec{q}$. Recall that

$$\vec{q} := |\vec{H}|^2 \partial_x \vec{\Phi} + 2(\vec{H} \cdot \vec{H}_0) \partial_{\bar{x}} \vec{\Phi}.$$

As $\mathrm{e}^\lambda \vec{H}$ and $\mathrm{e}^\lambda \vec{H}_0$ inherit the regularity of $\nabla \vec{n}$, we find

$$\begin{aligned} \big|\nabla\big(\mathrm{e}^\lambda (\vec{H} \cdot \vec{H}_0) \partial_{\bar{x}} \vec{\Phi}\big)\big| &\lesssim |\mathrm{e}^\lambda \vec{H}| \, |\nabla^2 \vec{n}| + |\nabla \vec{n}| \big(\mathrm{e}^\lambda |\nabla \vec{H}| + |\nabla^2 \vec{\Phi}| \, |\vec{H}|\big) \\ &\lesssim |x| \, |\nabla^2 \vec{n}| + |\nabla \vec{n}| + |x|^{-1} \mathrm{e}^\lambda |\vec{H}| \in \bigcap_{p<\infty} L^p, \end{aligned} \tag{2-92}$$

where we have used (2-91), (2-89), (2-88), and (2-90). Exactly in the same fashion, one verifies that

$$\mathrm{e}^\lambda |\vec{H}|^2 \partial_x \vec{\Phi} \in \bigcap_{p<\infty} W^{1,p}.$$

Together, the latter and (2-92) brought into the definition of $\vec{q}$ show that

$$\mathrm{e}^\lambda \vec{q} \in \bigcap_{p<\infty} W^{1,p}. \tag{2-93}$$

We next return to the system (2-76). Proceeding as in (2-77) with the information that $\alpha \le \theta_0 - 2$, we infer that

$$|\mathrm{div}(|x|^{-1}\vec{U})| + |\mathrm{curl}(|x|^{-1}\vec{U})| \lesssim |x|^{\theta_0-2-\alpha-\epsilon} \lesssim |x|^{-\epsilon} \quad \text{for all } \epsilon > 0,$$

so that $|x|^{-1}\mathrm{e}^{\lambda}\vec{Q} \equiv |x|^{-1}\vec{U}$ is an element of $W^{1,p}$ for all finite p. By a similar token, using (2-93), it is not difficult to see that

$$|\nabla^2\vec{U}| \lesssim |\nabla(|x|^{-1}\vec{U})| + |\nabla(\mathrm{e}^{\lambda}\vec{q})| \in \bigcap_{p<\infty} L^p.$$

Hence, $\mathrm{e}^{\lambda}\vec{Q} \equiv \vec{U} \in \bigcap_{p<\infty} W^{2,p}$. Using that $\nabla\lambda = \mathrm{O}(|x|^{-1})$ now gives

$$|x|^{-1}\mathrm{e}^{\lambda}|\nabla\vec{Q}| \lesssim |\nabla(|x|^{-1}\mathrm{e}^{\lambda}\vec{Q})| + |x|^{-2}\mathrm{e}^{\lambda}|\vec{Q}| \in \bigcap_{p<\infty} L^p,$$

where have used that $\alpha \le \theta_0 - 2$ and $\vec{Q} = \mathrm{O}(|x|^{1-\alpha-\epsilon})$ for all $\epsilon < 0$. In particular, owing to (2-75), we have

$$|x|^{-1}\mathrm{e}^{\lambda}\big|\nabla\big(\vec{H} + \vec{\gamma}_0 \log|x| - \Re(\vec{E})\big)\big| \lesssim |x|^{-1}\mathrm{e}^{\lambda}|\nabla\vec{Q}| \in \bigcap_{p<\infty} L^p.$$

Analogously, using now additionally that $\nabla^2\lambda = \mathrm{O}(|x|^{-2})$ yields

$$\mathrm{e}^{\lambda}|\nabla^2\vec{Q}| \lesssim |x|^{-2}|\vec{U}| + |\nabla(|x|^{-1}\vec{U})| + |\nabla^2\vec{U}|.$$

As we have shown above, each of these terms lies in L^p for all finite p. Accordingly, differentiating twice (2-75) yields

$$\mathrm{e}^{\lambda}\big|\nabla^2\big(\vec{H} + \vec{\gamma}_0 \log|x| - \Re(\vec{E})\big)\big| \lesssim \mathrm{e}^{\lambda}|\nabla^2\vec{Q}| \in \bigcap_{p<\infty} L^p. \tag{2-94}$$

We have pointed out that the function u in (2-82) lies in $C^{2,1-\epsilon}$ for all $\epsilon > 0$, owing to the fact that $\vec{n} \in W^{2,p}$ for all $p < \infty$. We may now replace (2-84) by

$$2\mathrm{e}^{2\lambda} = (T_2 + R_2)|x|^{2(\theta_0-1)},$$

where T_2 is the second-order Taylor polynomial expansion of $2\mathrm{e}^{2u}$, and R_2 is the corresponding remainder. Hence

$$\nabla^j R_2 = \mathrm{O}(|x|^{3-j-\epsilon}), \quad j \in \{0, \dots, 2\}, \text{ for all } \epsilon > 0. \tag{2-95}$$

As before, we write the decomposition

$$\Delta\vec{\Phi} \equiv 2\mathrm{e}^{2\lambda}\vec{H} = \Delta\vec{\Phi}_0 + \Delta\vec{\Phi}_1,$$

with now

$$\begin{cases} \Delta\vec{\Phi}_0 = T_2|x|^{2(\theta_0-1)}\Re(\vec{E} - \vec{\gamma}_0 \log|x|), \\ \Delta\vec{\Phi}_1 = -2\mathrm{e}^{2\lambda}\Re(\vec{Q}) + |x|^{2(\theta_0-1)}R_2\Re(\vec{E} - \vec{\gamma}_0 \log|x|). \end{cases}$$

Since T_2 and $\vec{E}$ are power functions, we easily obtain via solving explicitly and handling the remainder with Corollary C.3 that

$$\vec{\Phi}_0 = \Re(\vec{P}_0) + C_\alpha |x|^{2\theta_0} \Re(\vec{E}) - \vec{C}|x|^{2\theta_0}(\log|x|^{2\theta_0} - 4) + \vec{\xi}_0,$$

where $\vec{P}_0$ is a $\mathbb{C}^m$-valued holomorphic polynomial of degree at most $(2\theta_0 - \alpha)$, and the constants C_α and $\vec{C}$ are as in (2-86). The remainder $\vec{\xi}_0$ satisfies

$$\nabla^j \vec{\xi}_0 = \mathrm{O}(|x|^{2\theta_0 - \alpha + 1 - j - \epsilon}) \quad \text{for all } j \in \{0, \dots, 3\}, \text{ for all } \epsilon > 0,$$
$$|x|^{3+\alpha-2\theta_0} \nabla^4 \vec{\xi}_0 \in \bigcap_{p<\infty} L^p.$$

To obtain information on $\vec{\Phi}_1$, we differentiate twice its partial differential equation in each coordinate x_1 and x_2, and apply Corollary C.3 to the results using (2-95), the fact that $\mathrm{e}^\lambda \vec{Q} \in \bigcap_{p<\infty} W^{2,p}$, that $\nabla^2 \lambda = \mathrm{O}(|x|^{-2})$, and the fact that $\vec{E}$ is a power function. Without much effort, it ensues that we can write

$$\vec{\Phi}_1 = \Re(\vec{P}_1) + \vec{\xi}_1,$$

where $\vec{P}_1$ is a $\mathbb{C}^m$-valued holomorphic polynomial of degree at most $(2\theta_0 - \alpha)$, and the $\mathbb{R}^m$-valued function $\vec{\xi}_1$ satisfies

$$\nabla^j \vec{\xi}_1 = \mathrm{O}(|x|^{2\theta_0 - \alpha + 1 - j - \epsilon}) \quad \text{for all } j \in \{0, \dots, 3\}, \text{ for all } \epsilon > 0,$$
$$|x|^{3+\alpha-2\theta_0} \nabla^4 \vec{\xi}_1 \in \bigcap_{p<\infty} L^p.$$

Comparing $\vec{\Phi}_0 + \vec{\Phi}_1$ to the previously found expression (2-87), we deduce

$$\begin{aligned}\vec{\Phi} = \Re\big(\vec{A} z^{\theta_0} + \vec{B}_1 z^{\theta_0+1} + \vec{B}_2 z^{\theta_0+2} + C_\alpha |x|^{2\theta_0} \vec{E}\big)& \\ - \vec{C}|x|^{2\theta_0}(\log|x|^{2\theta_0} - 4) + (\vec{\xi}_0 + \vec{\xi}_1)&,\end{aligned} \tag{2-96}$$

where $\vec{A}$ and $\vec{B}_1$ are as in (2-87), while $\vec{B}_2 \in \mathbb{C}^m$ is a constant.

Note that

$$|\nabla^j \vec{\Phi}| = \mathrm{O}(|x|^{\theta_0 - j}) \quad \text{for all } j \in \{0, \dots, 3\}. \tag{2-97}$$

Finally, we return to (2-80). Using the previously noted fact that $\mathrm{e}^\lambda \vec{h}_{ij}$ inherit the regularity of $\nabla \vec{n}$, along with (2-90), (2-94), (2-97), we now obtain

$$\begin{aligned}|\Delta \nabla \vec{n}| &\lesssim |\nabla^2 \vec{n}| + |\nabla \vec{n}|^2 |\nabla \vec{n}| + |\nabla \vec{\Phi}||\nabla^2 \vec{H}| + |\nabla^2 \vec{\Phi}||\nabla \vec{H}| \\ &\lesssim |\nabla^2 \vec{n}| + |\nabla \vec{n}|^2 |\nabla \vec{n}| + \mathrm{e}^\lambda |\nabla^2 \vec{H}| + |x|^{-1} \mathrm{e}^\lambda |\nabla \vec{H}| \\ &\simeq |x|^{\theta_0 - 3 - \alpha} + \text{terms in } \bigcap_{p<\infty} L^p.\end{aligned}$$

This shows that $\nabla^3 \vec{n} \in L^{2,\infty}$ if $\alpha = \theta_0 - 2$. On the other hand, if $\alpha \leq \theta_0 - 3$, we obtain that $\vec{n} \in \bigcap_{p<\infty} W^{3,p}$. We may then start over again the above procedure gaining one order of decay at every step. A clear pattern emerges. Repeating finitely many times the steps performed above, one eventually reaches that

$$\nabla^{\theta_0-\alpha+1}\vec{n} \in L^{2,\infty} \quad \text{and thus} \quad \nabla^{\theta_0-\alpha}\vec{n} \in \mathrm{BMO}\,. \tag{2-98}$$

Furthermore, for all $j \in \{0, \dots, \theta_0 - \alpha\}$, we have

$$|x|^{\alpha+j-1}\nabla^j\big(\vec{H} + \vec{\gamma}_0 \log|x| - \Re(\vec{E})\big) \in \bigcap_{p<\infty} L^p. \tag{2-99}$$

We also obtain a local expansion for the immersion, namely

$$\vec{\Phi} = \Re\left(\vec{A}x^{\theta_0} + \sum_{j=1}^{\theta_0-\alpha} \vec{B}_j x^{\theta_0+j} + C_\alpha |x|^{2\theta_0}\vec{E}\right) - \vec{C}|x|^{2\theta_0}(\log|x|^{2\theta_0} - 4) + \vec{\xi}, \tag{2-100}$$

where $\vec{B}_j \in \mathbb{C}^m$ are constant vectors, while $\vec{A}$ is as in (2-70). The constants C_α and $\vec{C}$ are

$$C_\alpha := \frac{e^{2u(0)}}{2\theta_0(\theta_0-\alpha)} \quad \text{and} \quad \vec{C} := \frac{e^{2u(0)}}{2\theta_0^3}\vec{\gamma}_0.$$

The remainder $\vec{\xi}$ satisfies

$$\nabla^j\vec{\xi} = O(|x|^{2\theta_0-\alpha+1-j-\epsilon}) \quad \text{for all } j \in \{0, \dots, \theta_0-\alpha+1\}, \text{ for all } \epsilon > 0,$$
$$|x|^{1-\theta_0}\nabla^{\theta_0-\alpha+2}\vec{\xi} \in \bigcap_{p<\infty} L^p.$$

Of course, when $\alpha > 0$, the “remainder” term $\vec{\xi}$ in (2-100) dominates the logarithmic term, written here to indicate the presence and the influence of the (modified) first residue $\vec{\gamma}_0$ of which it is a multiple.

2E3. *When both residues vanish: smoothness of the immersion.* This last section is devoted to proving Theorem 1.9. We shall assume that the first and second residues defined respectively in (1-8) and in (2-73) vanish.

When $\theta_0 = 1$, we have seen at the end of Section 2C that $\nabla\vec{n} \in \mathrm{BMO}$. In the same section, Proposition 2.2 states that $\vec{H} \in W^{1,p}$ for all $p < \infty$. Hence, $\nabla\vec{H} \in L^p$ for all finite p. On the other hand, when $\theta \geq 2$, we proved in (2-100) that $\nabla^{\theta_0}\vec{n} \in \mathrm{BMO}$ and in (2-99) that $|x|^{j-1}\nabla^j\vec{H} \in \bigcap_{p<\infty} L^p$ for all $j \in \{1, \dots, \theta_0\}$. Altogether, in all cases, we thus have

$$\nabla^{\theta_0}\vec{n} \in \mathrm{BMO} \quad \text{and} \quad |x|^{j-1}\nabla^j\vec{H} \in \bigcap_{p<\infty} L^p \quad \text{for all } j \in \{1, \dots, \theta_0\}.$$

Observe that (2-100) implies

$$|\nabla^j\vec{\Phi}(x)| \lesssim |x|^{\theta_0-j} \quad \text{for all } j \in \{0, \dots, \theta_0\}.$$

Owing to

$$\left|\nabla^{\theta_0-1}(\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H})\right| \lesssim \sum_{j=1}^{\theta_0}|\nabla^j\vec{H}|\,|\nabla^{\theta_0-j+1}\vec{\Phi}| \lesssim \sum_{j=1}^{\theta_0}|x|^{j-1}|\nabla^j\vec{H}|,$$

whence we find

$$\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H}\in\bigcap_{p<\infty}W^{\theta_0-1,p}. \tag{2-101}$$

Recall next (B-8) satisfied by the Gauss map, namely

$$\Delta\vec{n} = 2\star(\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H}) + 2\mathrm{e}^{2\lambda}K\vec{n} - 2\star\mathrm{e}^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22}). \tag{2-102}$$

As previously noticed, $\mathrm{e}^{\lambda}\vec{h}_{ij}$ inherits the regularity of $|\nabla\vec{n}|$, so that

$$\mathrm{e}^{2\lambda}K\vec{n} - 2\star\mathrm{e}^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22})\in\bigcap_{p<\infty}W^{\theta_0-1,p}. \tag{2-103}$$

Introducing (2-101) and (2-103) into (2-102) now shows that $\vec{n}\in W^{\theta_0+1,p}$ for all $p<\infty$, thereby improving the regularity of $\vec{n}$. It suffices now to repeat the procedure outlined in Section 2E2 and in the above paragraph until reaching that $\vec{n}$ is smooth, from which it immediately follows that the immersion $\vec{\Phi}$ is smooth as well. This concludes the proof of Theorem 1.9.

Appendix A. Notational conventions

We place an arrow on all letters referring to elements of $\mathbb{R}^m$. To simplify the notation, by $\vec{\Phi}\in X(D^2)$ is meant $\vec{\Phi}\in X(D^2,\mathbb{R}^m)$ whenever X is a function space. Similarly, we write $\nabla\vec{\Phi}\in X(D^2)$ for $\nabla\vec{\Phi}\in\mathbb{R}^2\otimes X(D^2,\mathbb{R}^m)$.

Although this custom may seem at first odd, we allow the differential operators classically acting on scalars to act on elements of $\mathbb{R}^m$. Thus, for example, $\nabla\vec{\Phi}$ is the element of $\mathbb{R}^2\otimes\mathbb{R}^m$ that can be written $(\partial_{x_1}\vec{\Phi},\partial_{x_2}\vec{\Phi})$. If S is a scalar and $\vec{R}$ an element of $\mathbb{R}^m$, we let

$$\begin{aligned}
\vec{R}\cdot\nabla\vec{\Phi} &:= (\vec{R}\cdot\partial_{x_1}\vec{\Phi},\ \vec{R}\cdot\partial_{x_2}\vec{\Phi}),\\
\nabla^{\perp}S\cdot\nabla\vec{\Phi} &:= \partial_{x_1}S\,\partial_{x_2}\vec{\Phi}-\partial_{x_2}S\,\partial_{x_1}\vec{\Phi},\\
\nabla^{\perp}\vec{R}\cdot\nabla\vec{\Phi} &:= \partial_{x_1}\vec{R}\cdot\partial_{x_2}\vec{\Phi}-\partial_{x_2}\vec{R}\cdot\partial_{x_1}\vec{\Phi},\\
\nabla^{\perp}\vec{R}\wedge\nabla\vec{\Phi} &:= \partial_{x_1}\vec{R}\wedge\partial_{x_2}\vec{\Phi}-\partial_{x_2}\vec{R}\wedge\partial_{x_1}\vec{\Phi}.
\end{aligned}$$

Analogous quantities are defined according to the same logic.

Two operations between multivectors are useful. The *interior multiplication* $\llcorner$ maps a pair comprising a q-vector γ and a p-vector β to a $(q-p)$-vector. It is defined via

$$\langle\gamma\llcorner\beta,\alpha\rangle = \langle\gamma,\beta\wedge\alpha\rangle \quad \text{for each } (q-p)\text{-vector } \alpha.$$

Let α be a k-vector. The *first-order contraction* operation $\bullet$ is defined inductively through

$$\alpha \bullet \beta = \alpha \llcorner \beta \quad \text{when } \beta \text{ is a 1-vector,}$$

and

$$\alpha \bullet (\beta \wedge \gamma) = (\alpha \bullet \beta) \wedge \gamma + (-1)^{pq} (\alpha \bullet \gamma) \wedge \beta$$

when β and γ are respectively a p-vector and a q-vector.

Appendix B. Miscellaneous facts

On the Gauss map. Let $\vec{\Phi}$ be a conformal immersion of the unit disk into $\mathbb{R}^m$. For $j \in \{1, 2\}$, we let

$$\vec{e}_j := \mathrm{e}^{-\lambda} \partial_{x_j} \vec{\Phi} \quad \text{with } 2\mathrm{e}^{2\lambda} = |\nabla \vec{\Phi}|^2.$$

One easily verifies (see details in [Bernard and Rivière 2011b, Section III.2.2]) that

$$\pi_T \nabla \vec{e}_j = (\nabla^\perp \lambda) \vec{e}_{j'} \quad \text{where } (\vec{e}_{1'}, \vec{e}_{2'}) := (\vec{e}_2, -\vec{e}_1), \tag{B-1}$$

where π_T denotes projection onto the tangent space spanned by $\{\vec{e}_1, \vec{e}_2\}$. Moreover,

$$\pi_{\vec{n}} \nabla \vec{e}_j \equiv \mathrm{e}^{-\lambda} \pi_{\vec{n}} \nabla \partial_j \vec{\Phi} =: \mathrm{e}^{\lambda} \begin{pmatrix} \vec{h}_{1j} \\ \vec{h}_{2j} \end{pmatrix}. \tag{B-2}$$

where $\pi_{\vec{n}}$ denotes projection onto the normal space: $\pi_{\vec{n}} = \mathrm{id} - \pi_T$. With this notation, the mean curvature vector takes the form

$$\vec{H} = \tfrac{1}{2} (\vec{h}_{11} + \vec{h}_{22}). \tag{B-3}$$

The $(m-2)$-vector $\vec{n}$ satisfies $\vec{n} := \star(\vec{e}_1 \wedge \vec{e}_2)$. Accordingly, using (B-1), we have

$$\nabla \vec{n} = \star\big[(\pi_{\vec{n}} \nabla \vec{e}_1) \wedge \vec{e}_2 + \vec{e}_1 \wedge (\pi_{\vec{n}} \nabla \vec{e}_2)\big], \tag{B-4}$$

so that

$$\begin{aligned} \Delta \vec{n} = \star\big[\mathrm{div}(\pi_{\vec{n}} \nabla \vec{e}_1) \wedge \vec{e}_2 + \vec{e}_1 \wedge \mathrm{div}(\pi_{\vec{n}} \nabla \vec{e}_2)\big] &+ 2 \star [\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2] \\ &+ \star[\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_T \nabla \vec{e}_2 + \pi_T \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2]. \end{aligned}$$

The identities (B-1) yield

$$\pi_T \nabla \vec{e}_k \wedge \pi_{\vec{n}} \nabla \vec{e}_l = (\nabla^\perp \lambda) \cdot (\vec{e}_{k'} \wedge \pi_{\vec{n}} \nabla \vec{e}_l),$$

and thus

$$\begin{aligned} \Delta \vec{n} = \star\big[\mathrm{div}(\pi_{\vec{n}} \nabla \vec{e}_1) \wedge \vec{e}_2 + \vec{e}_1 \wedge \mathrm{div}(\pi_{\vec{n}} \nabla \vec{e}_2)\big] &+ 2 \star [\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2] \\ &+ \star(\nabla^\perp \lambda) \cdot [\vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_1 + \vec{e}_2 \wedge \pi_{\vec{n}} \nabla \vec{e}_2]. \end{aligned} \tag{B-5}$$

Next, using the definition of $\vec{e}_k$ and again (B-1), we obtain[20]

$$\begin{aligned}\operatorname{div} \pi_{\vec{n}} \nabla \vec{e}_k &\equiv \pi_{\vec{n}} \operatorname{div} \pi_{\vec{n}} \nabla \vec{e}_k + \pi_T \operatorname{div} \pi_{\vec{n}} \nabla \vec{e}_k \\ &= \pi_{\vec{n}} \operatorname{div} \pi_{\vec{n}} \nabla (\mathrm{e}^{-\lambda} \partial_{x_k} \vec{\Phi}) + (\vec{e}_l \cdot \operatorname{div} \pi_{\vec{n}} \nabla \vec{e}_k) \vec{e}_l \\ &= \mathrm{e}^{-\lambda} \pi_{\vec{n}} \operatorname{div} \pi_{\vec{n}} \nabla \partial_{x_k} \vec{\Phi} - \mathrm{e}^{-\lambda} \pi_{\vec{n}} (\nabla \lambda \cdot \nabla \partial_{x_k} \vec{\Phi}) - (\pi_{\vec{n}} \nabla \vec{e}_l \cdot \pi_{\vec{n}} \nabla \vec{e}_k) \vec{e}_l \\ &= \mathrm{e}^{-\lambda} \pi_{\vec{n}} \operatorname{div} \pi_{\vec{n}} \nabla \partial_{x_k} \vec{\Phi} - (\pi_{\vec{n}} \nabla \vec{e}_l \cdot \pi_{\vec{n}} \nabla \vec{e}_k) \vec{e}_l - \nabla \lambda \cdot \pi_{\vec{n}} \nabla \vec{e}_k .\end{aligned}$$

Introducing the latter into (B-5) gives, after a few elementary manipulations,

$$\begin{aligned}\Delta \vec{n} = \star \mathrm{e}^{-\lambda} \big[\pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_1} \vec{\Phi}) \wedge \vec{e}_2 + \vec{e}_1 \wedge \pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_2} \vec{\Phi}) \big] \\ - \big[|\pi_{\vec{n}} \nabla \vec{e}_1|^2 + |\pi_{\vec{n}} \nabla \vec{e}_2|^2 \big] \star (\vec{e}_1 \wedge \vec{e}_2) + 2 \star [\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2] \\ + \star (\nabla^{\perp} \lambda) \cdot \big[\vec{e}_1 \wedge \pi_{\vec{n}} (\nabla \vec{e}_1 - \nabla^{\perp} \vec{e}_2) - \pi_{\vec{n}} (\nabla^{\perp} \vec{e}_1 + \nabla \vec{e}_2) \wedge \vec{e}_2 \big].\end{aligned}$$

Owing to (B-2) and (B-4), we find

$$\begin{aligned}\Delta \vec{n} + |\nabla \vec{n}|^2 \vec{n} = \star \mathrm{e}^{-\lambda} \big[\pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_1} \vec{\Phi}) \wedge \vec{e}_2 + \vec{e}_1 \wedge \pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_2} \vec{\Phi}) \big] \\ + 2 \star \mathrm{e}^{-2\lambda} [\pi_{\vec{n}} \nabla \partial_{x_1} \vec{\Phi} \wedge \pi_{\vec{n}} \nabla \partial_{x_2} \vec{\Phi}] + 2 \star \mathrm{e}^{\lambda} \vec{H} \wedge [\partial_{x_2} \lambda \vec{e}_1 - \partial_{x_1} \lambda \vec{e}_2].\end{aligned}$$

Equivalently,

$$\begin{aligned}\Delta \vec{n} + |\nabla \vec{n}|^2 \vec{n} = \star \vec{e}_1 \wedge \mathrm{e}^{-\lambda} \big[\pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_2} \vec{\Phi}) - 2 \mathrm{e}^{2\lambda} \vec{H} \partial_{x_2} \lambda \big] \\ - \star \vec{e}_2 \wedge \mathrm{e}^{-\lambda} \big[\pi_{\vec{n}} \operatorname{div} (\pi_{\vec{n}} \nabla \partial_{x_1} \vec{\Phi}) - 2 \mathrm{e}^{2\lambda} \vec{H} \partial_{x_1} \lambda \big] \\ + 2 \star [\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2].\end{aligned} \tag{B-6}$$

Moreover, (B-1) gives $\pi_T \nabla \partial_{x_j} \vec{\Phi} = \nabla (\mathrm{e}^{\lambda}) \vec{e}_j + \nabla^{\perp} (\mathrm{e}^{\lambda}) \vec{e}_{j'}$. Hence, calling upon (B-2) implies

$$\pi_{\vec{n}} \operatorname{div} \pi_T \nabla \partial_{x_j} \vec{\Phi} = \nabla (\mathrm{e}^{\lambda}) \cdot \pi_{\vec{n}} \nabla \vec{e}_j + \nabla^{\perp} (\mathrm{e}^{\lambda}) \cdot \pi_{\vec{n}} \nabla \vec{e}_{j'} = \vec{H} \partial_{x_j} \mathrm{e}^{2\lambda},$$

and thus, as $\Delta \vec{\Phi} = 2 \mathrm{e}^{2\lambda} \vec{H}$,

$$\pi_{\vec{n}} \operatorname{div} \pi_{\vec{n}} \nabla \partial_{x_j} \vec{\Phi} \equiv \pi_{\vec{n}} \partial_{x_j} \Delta \vec{\Phi} - \pi_{\vec{n}} \operatorname{div} \pi_T \nabla \partial_{x_j} \vec{\Phi} = 2 \pi_{\vec{n}} \partial_{x_j} (\mathrm{e}^{2\lambda} \vec{H}) - \vec{H} \partial_{x_j} \mathrm{e}^{2\lambda}.$$

The interested reader will note that this equation is equivalent to the Codazzi–Mainardi identities. Substituted into (B-6), the latter gives

$$\begin{aligned}\Delta \vec{n} + |\nabla \vec{n}|^2 \vec{n} &= 2 \star \big(\partial_{x_1} \vec{\Phi} \wedge \pi_{\vec{n}} \partial_{x_2} \vec{H} - \partial_{x_2} \vec{\Phi} \wedge \pi_{\vec{n}} \partial_{x_1} \vec{H} \big) + 2 \star \big[\pi_{\vec{n}} \nabla \vec{e}_1 \wedge \pi_{\vec{n}} \nabla \vec{e}_2 \big] \\ &= 2 \star \big(\nabla^{\perp} \vec{\Phi} \wedge \pi_{\vec{n}} \nabla \vec{H} \big) - 2 \star \mathrm{e}^{2\lambda} \vec{h}_{12} \wedge (\vec{h}_{11} - \vec{h}_{22}).\end{aligned} \tag{B-7}$$

One also notes from (B-2) and the fact that $\vec{H}$ is normal that

$$\pi_T \partial_{x_j} \vec{H} \equiv \langle \vec{e}_k , \partial_{x_j} \vec{H} \rangle \vec{e}_k = - \mathrm{e}^{\lambda} (\vec{H} \cdot \vec{h}_{jk}) \vec{e}_k ,$$

[20] Implicit summations over repeated indices are understood.

whence

$$\nabla^{\perp}\vec{\Phi}\wedge\pi_T\nabla\vec{H}\equiv\partial_{x_1}\vec{\Phi}\wedge\pi_T\partial_{x_2}\vec{H}-\partial_{x_2}\vec{\Phi}\wedge\pi_T\partial_{x_1}\vec{H}=-2\mathrm{e}^{2\lambda}|\vec{H}|^2(\star\vec{n}).$$

Equation (B-7) may thus be recast as

$$\Delta\vec{n}+|\nabla\vec{n}|^2\vec{n}=2\star(\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H})+4\mathrm{e}^{2\lambda}|\vec{H}|^2\vec{n}-2\star\mathrm{e}^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22}).$$

Finally, since

$$|\nabla\vec{n}|^2-4\mathrm{e}^{2\lambda}|\vec{H}|^2=-2\mathrm{e}^{2\lambda}K,$$

where K is the Gauss curvature, we obtain

$$\Delta\vec{n}-2\mathrm{e}^{2\lambda}K\vec{n}=2\star(\nabla^{\perp}\vec{\Phi}\wedge\nabla\vec{H})-2\star\mathrm{e}^{2\lambda}\vec{h}_{12}\wedge(\vec{h}_{11}-\vec{h}_{22}).\tag{B-8}$$

Conservative conformal Willmore system. We establish in this section a few general identities. As before, we let $\vec{\Phi}$ be a (smooth) conformal immersion of the unit disk into $\mathbb{R}^m$, and set $\vec{e}_j:=\mathrm{e}^{-\lambda}\partial_{x_j}\vec{\Phi}$, where λ is the conformal parameter. Since $\vec{\Phi}$ is conformal, $\{\vec{e}_1,\vec{e}_2\}$ forms an orthonormal basis of the tangent space. As $\vec{n}=\star(\vec{e}_1\wedge\vec{e}_2)$, if $\vec{V}$ is a 1-vector, we find

$$(\star\vec{n})\cdot(\vec{V}\wedge\partial_{x_j}\vec{\Phi})=\mathrm{e}^{-\lambda}(\vec{e}_1\wedge\vec{e}_2)\cdot(\vec{V}\wedge\vec{e}_j)=-\mathrm{e}^{-\lambda}\vec{e}_{j'}\cdot\vec{V}=-\partial_{x_{j'}}\vec{\Phi}\cdot\vec{V},$$

where

$$(\vec{e}_{1'},\vec{e}_{2'}):=(\vec{e}_2,-\vec{e}_1).$$

Hence,

$$(\star\vec{n})\cdot(\vec{V}\wedge\nabla\vec{\Phi})=\vec{V}\cdot\nabla^{\perp}\vec{\Phi},\quad(\star\vec{n})\cdot(\vec{V}\wedge\nabla^{\perp}\vec{\Phi})=-\vec{V}\cdot\nabla\vec{\Phi}.\tag{B-9}$$

We choose next an orthonormal basis $\{\vec{n}_\alpha\}_{\alpha=1}^{m-2}$ of the normal space such that $\{\vec{e}_1,\vec{e}_2,\vec{n}_1,\dots,\vec{n}_{m-2}\}$ is a positive oriented orthonormal basis of $\mathbb{R}^m$.

Recalling the definition of the interior multiplication operator $\llcorner$ given in Appendix A, it is not hard to obtain

$$(\star\vec{n})\llcorner\vec{e}_j=(\vec{e}_1\wedge\vec{e}_2)\llcorner\vec{e}_j=\delta_{j2}\vec{e}_1-\delta_{j1}\vec{e}_2\quad\text{and}\quad(\star\vec{n})\llcorner\vec{n}_\alpha=0.$$

Hence,

$$(\star\vec{n})\bullet(\vec{e}_j\wedge\vec{n}_\alpha)\equiv((\star\vec{n})\llcorner\vec{e}_j)\wedge\vec{n}_\alpha+((\star\vec{n})\llcorner\vec{n}_\alpha)\wedge\vec{e}_j=\delta_{j2}\vec{e}_1\wedge\vec{n}_\alpha-\delta_{j1}\vec{e}_2\wedge\vec{n}_\alpha.$$

Moreover, we have trivially

$$(\star\vec{n})\bullet(\vec{e}_j\wedge\vec{e}_k)=\pm(\star\vec{n})\bullet(\star\vec{n})=0.$$

From this one easily deduces that, for every 1-vector $\vec{V}$,

$$(\star\vec{n})\bullet(\vec{V}\wedge\nabla\vec{\Phi})=\pi_{\vec{n}}\vec{V}\wedge\nabla^{\perp}\vec{\Phi},\quad(\star\vec{n})\bullet(\vec{V}\wedge\nabla^{\perp}\vec{\Phi})=-\pi_{\vec{n}}\vec{V}\wedge\nabla\vec{\Phi}.\tag{B-10}$$

We have furthermore

$$(\vec{V}\wedge\vec{e}_j)\bullet\vec{e}_i=(\vec{e}_i\cdot\vec{V})\vec{e}_j-\delta_{ij}\vec{V}.$$

From this, and $\vec{e}_i:=\mathrm{e}^{-\lambda}\partial_{x^i}\vec{\Phi}$, it follows that whenever $\vec{V}=V^i\vec{e}_i+V^\alpha\vec{n}_\alpha$ then

$$\text{(B-11)}\qquad\begin{cases}(\vec{V}\wedge\nabla^\perp\vec{\Phi})\bullet\nabla^\perp\vec{\Phi}=\mathrm{e}^{2\lambda}(\pi_T\vec{V}-2\vec{V}),\\ \ (\vec{V}\wedge\nabla\vec{\Phi})\bullet\nabla^\perp\vec{\Phi}=\mathrm{e}^{2\lambda}(V^2\vec{e}_1-V^1\vec{e}_2)\equiv(\vec{V}\cdot\nabla\vec{\Phi})\cdot\nabla^\perp\vec{\Phi}.\end{cases}$$

We are now sufficiently geared to prove:

Lemma B.1. *Let $\vec{\Phi}$ be a smooth conformal immersion of the unit disk into $\mathbb{R}^m$ with corresponding mean curvature vector $\vec{H}$, and let $\vec{L}$ be a* 1*-vector. We define $A\in\mathbb{R}^2\otimes\bigwedge^0(\mathbb{R}^m)$ and $\vec{B}\in\mathbb{R}^2\otimes\bigwedge^2(\mathbb{R}^m)$ via*

$$A=\vec{L}\cdot\nabla\vec{\Phi},\qquad\vec{B}=\vec{L}\wedge\nabla\vec{\Phi}+2\vec{H}\wedge\nabla^\perp\vec{\Phi}.$$

Then the following identities hold:

$$\text{(B-12)}\qquad A=-(\star\vec{n})\cdot\vec{B}^\perp,\qquad\vec{B}=-(\star\vec{n})\bullet\vec{B}^\perp+(\star\vec{n})A^\perp,$$

where $\star\vec{n}:=(\partial_{x_1}\vec{\Phi}\wedge\partial_{x_2}\vec{\Phi})/|\partial_{x_1}\vec{\Phi}\wedge\partial_{x_2}\vec{\Phi}|$.

Moreover, we have

$$\text{(B-13)}\qquad 2\Delta\vec{\Phi}=A\cdot\nabla^\perp\vec{\Phi}-\vec{B}\bullet\nabla^\perp\vec{\Phi}.$$

Proof. The identities (B-9) give immediately (recall that $\vec{H}$ is a normal vector, so that $\vec{H}\cdot\nabla^\perp\vec{\Phi}=0$) the required

$$(\star\vec{n})\cdot\vec{B}^\perp=-\vec{L}\cdot\nabla\vec{\Phi}+2\vec{H}\cdot\nabla^\perp\vec{\Phi}=-\vec{L}\cdot\nabla\vec{\Phi}=-A.$$

Analogously, the identities (B-10) give (again, $\vec{H}$ is normal, so $\pi_{\vec{n}}\vec{H}=\vec{H}$)

$$\begin{aligned}(\star\vec{n})\bullet\vec{B}^\perp&=-\pi_{\vec{n}}\vec{L}\wedge\nabla\vec{\Phi}-2\vec{H}\wedge\nabla^\perp\vec{\Phi}=-\vec{B}+\pi_T\vec{L}\wedge\nabla\vec{\Phi}\\&=-\vec{B}+\mathrm{e}^\lambda\big((\vec{L}\cdot\vec{e}_1)\vec{e}_1+(\vec{L}\cdot\vec{e}_2)\vec{e}_2\big)\wedge\begin{pmatrix}\vec{e}_1\\\vec{e}_2\end{pmatrix}=-\vec{B}+\mathrm{e}^\lambda\begin{pmatrix}-\vec{L}\cdot\vec{e}_2\\+\vec{L}\cdot\vec{e}_1\end{pmatrix}\vec{e}_1\wedge\vec{e}_2\\&=-\vec{B}+(\vec{L}\cdot\nabla^\perp\vec{\Phi})(\star\vec{n})=-\vec{B}+(\star\vec{n})A^\perp,\end{aligned}$$

which is the second equality in (B-12).

In order to prove (B-13), we will use (B-11). Namely, since $\vec{H}=H^\alpha\vec{n}_\alpha$, we find

$$\vec{B}\bullet\nabla^\perp\vec{\Phi}=(\vec{L}\cdot\nabla\vec{\Phi})\cdot\nabla^\perp\vec{\Phi}-4\mathrm{e}^{2\lambda}\vec{H}=A\cdot\nabla^\perp\vec{\Phi}-4\mathrm{e}^{2\lambda}\vec{H}.$$

Hence,

$$\vec{B}\bullet\nabla^\perp\vec{\Phi}-A\cdot\nabla^\perp\vec{\Phi}=-4\mathrm{e}^{2\lambda}\vec{H}.$$

Finally, there remains to recall that $\Delta\vec{\Phi}=2\mathrm{e}^{2\lambda}\vec{H}$ to reach the desired identity. □

We choose now

$$A = \nabla S - \nabla^{\perp} g \quad \text{and} \quad \vec{B} = \nabla \vec{R} - \nabla^{\perp} \vec{G},$$

where S and g are scalars, while $\vec{R}$ and $\vec{G}$ are 2-vectors. Then Lemma B.1 yields

$$\begin{cases} \nabla S = -(\star\vec{n}) \cdot (\nabla^{\perp} \vec{R} + \nabla \vec{G}) + \nabla^{\perp} g, \\ \nabla \vec{R} = -(\star\vec{n}) \bullet (\nabla^{\perp} \vec{R} + \nabla \vec{G}) + (\star\vec{n})(\nabla^{\perp} S + \nabla g) + \nabla^{\perp} \vec{G}, \end{cases}$$

thereby giving

$$\text{(B-14)} \qquad \begin{cases} -\Delta S = \nabla(\star\vec{n}) \cdot \nabla^{\perp} \vec{R} + \operatorname{div}\big((\star\vec{n}) \cdot \nabla \vec{G}\big), \\ -\Delta \vec{R} = \nabla(\star\vec{n}) \bullet \nabla^{\perp} \vec{R} - \nabla(\star\vec{n}) \cdot \nabla^{\perp} S + \operatorname{div}\big((\star\vec{n}) \bullet \nabla \vec{G} - \star\vec{n} \nabla g\big). \end{cases}$$

Furthermore, we have

$$\text{(B-15)} \qquad 2\Delta\vec{\Phi} = (\nabla S - \nabla^{\perp} g) \cdot \nabla^{\perp} \vec{\Phi} - (\nabla \vec{R} - \nabla^{\perp} \vec{G}) \bullet \nabla^{\perp} \vec{\Phi}.$$

Appendix C. Nonlinear and weighted elliptic results

Proposition C.1. *Let $u \in W^{1,2}(B_1(0)) \cap C^2(B_1(0) \setminus \{0\})$ satisfy the equation*

$$\text{(C-1)} \qquad -\Delta u = \nabla b \cdot \nabla^{\perp} u + \operatorname{div}(b \nabla f) \quad \textit{on } B_1(0),$$

where $f \in W_0^{2,(2,\infty)}(B_1(0))$, and moreover

$$\text{(C-2)} \qquad b \in W^{1,2} \cap L^{\infty}(B_1(0)) \quad \textit{with } \|\nabla b\|_{L^2(B_1(0))} < \varepsilon_0$$

for some ε_0 chosen to be "small enough". Then

$$\nabla u \in L^p(B_{1/4}(0)) \quad \textit{for some } p > 2.$$

Proof. Before delving into the proof of the statement, one important remark is in order. Let D be any disk included (properly or not) in $B_1(0)$. From the very definition of the space $L^{2,\infty}$ (see [Tartar 2007]), we have

$$\text{(C-3)} \qquad \|\Delta f\|_{L^1(D)} \le |D|^{\frac{1}{2}} \|\Delta f\|_{L^{2,\infty}(D)} \lesssim |D|^{\frac{1}{2}} \|\nabla^2 f\|_{L^{2,\infty}(D)}.$$

Moreover, an embedding result of [Tartar 2007] states that ∇f has bounded mean oscillations, whence in particular

$$\text{(C-4)} \qquad \|\nabla f\|_{L^2(D)} \lesssim |D|^{\frac{1}{2}-\epsilon} \quad \text{for all } \epsilon > 0.$$

These inequalities shall be helpful in the sequel.

We now return to the proof of the proposition. Let us fix some point $x_0 \in B_{1/2}(0)$ and some radius $\sigma \in \big(0, \frac{1}{2}\big)$, and we let $k \in (0, 1)$. Note that $B_{k\sigma}(x_0)$ is properly contained in $B_1(0)$. To reach the desired result, we decompose the solution to (C-1) as the sum $u = u_0 + u_1$, where

$$\begin{cases} -\Delta u_0 = \operatorname{div}(b\nabla f), & -\Delta u_1 = \nabla b \cdot \nabla^{\perp} u \quad \text{in } B_\sigma(x_0), \\ \quad u_0 = u, & \qquad u_1 = 0 \qquad\qquad \text{on } \partial B_\sigma(x_0). \end{cases}$$

Accounting for the hypotheses (C-2) and (C-4) in standard elliptic estimates (see [Almeida 1995, Proposition 4]) yields

$$\begin{aligned} \|\nabla u_0\|_{L^2(B_{k\sigma}(x_0))} &\lesssim \|b\nabla f\|_{L^2(B_{k\sigma}(x_0))} + k\|\nabla u\|_{L^2(B_\sigma(x_0))} \\ &\lesssim (k\sigma)^{1-\epsilon} + k\|\nabla u\|_{L^2(B_\sigma(x_0))}, \end{aligned} \tag{C-5}$$

up to some unimportant multiplicative constants. On the other hand, applying Wente's inequality (see [Hélein 1996, Theorem 3.4.1]) gives

$$\begin{aligned} \|\nabla u_1\|_{L^2(B_{k\sigma}(x_0))} \le \|\nabla u_1\|_{L^2(B_\sigma(x_0))} &\lesssim \|\nabla b\|_{L^2(B_\sigma(x_0))}\|\nabla u\|_{L^2(B_\sigma(x_0))} \\ &\le \varepsilon_0\|\nabla u\|_{L^2(B_\sigma(x_0))}, \end{aligned} \tag{C-6}$$

again up to some multiplicative constant without bearing on the sequel. Hence, combining (C-5) and (C-6), we obtain the estimate

$$\begin{aligned} \|\nabla u\|_{L^2(B_{k\sigma}(x_0))} &\le \|\nabla u_0\|_{L^2(B_{k\sigma}(x_0))} + \|\nabla u_1\|_{L^2(B_{k\sigma}(x_0))} \\ &\lesssim (k+\varepsilon_0)\|\nabla u\|_{L^2(B_\sigma(x_0))} + (k\sigma)^{1-\epsilon}. \end{aligned}$$

Because ε_0 and ϵ are small adjustable parameters, we may always choose k so as to arrange for $(k+\varepsilon_0)$ to be less than 1. A standard "controlled-growth" argument (see, e.g., [Hélein 1996, Lemma 3.5.11]) enables us to conclude that there exists some $\beta \in (0,1)$ for which

$$\|\nabla u\|_{L^2(B_\sigma(x))} \le C_0\sigma^\beta \quad \text{for all } \sigma \in \left(0, \tfrac{1}{2}\right),\ x \in B_{1/2}(0), \tag{C-7}$$

and for some constant C_0.

With the help of the Poincaré inequality, this estimate may be used to show that u is locally Hölder continuous. We are however interested in another implication of (C-7). Consider the maximal function

$$M_{2-\beta}\, g(x) := \sup_{\sigma>0} \sigma^{-\beta} \int_{B_\sigma(x)} |g(y)|\, dy. \tag{C-8}$$

We recast (C-1) in the form

$$-\Delta u = b\Delta f + \nabla b \cdot (\nabla^{\perp} u + \nabla f).$$

Calling upon (C-2)–(C-4) and upon the estimate (C-7), we derive that, for $x \in B_{1/2}(0)$, we have

(C-9)
$$\begin{aligned} M_{2-\beta}(\chi_{B_{1/2}(0)}\Delta u)(x) &\leq \|b\|_{L^\infty(B_1(0))} \sup_{0<\sigma<1/2} \sigma^{-\beta}\|\Delta f\|_{L^1(B_\sigma(x))} \\ &\qquad + \|\nabla b\|_{L^2(B_1(0))} \sup_{0<\sigma<1/2} \sigma^{-\beta}\big(\|\nabla u\|_{L^2(B_\sigma(x))} + \|\nabla f\|_{L^2(B_\sigma(x))}\big) \\ &\lesssim \sup_{0<\sigma<1/2} \sigma^{-\beta+1} + \varepsilon_0 \sup_{0<\sigma<1/2} (\sigma^{-\beta+\beta} + \sigma^{-\beta+1-\epsilon}) < \infty \end{aligned}$$

for all $0 < \epsilon \leq 1-\beta$. Moreover, it is clear that Δu is integrable on $B_{1/2}(0)$. We may thus use Proposition 3.2 from [Adams 1975][21] to deduce that

$$\frac{1}{|x|} * \chi_{B_{1/2}(0)}\Delta u \in L^{r,\infty}(B_{1/2}(0)) \quad \text{with } r := \frac{2-\beta}{1-\beta} > 2.$$

A classical estimate about Riesz kernels states we have in general

$$|\nabla u|(y) \lesssim \frac{1}{|x|} * \chi_{B_{1/2}(0)}\Delta u + C \quad \text{for all } y \in B_{1/4}(0)),$$

where C is a constant depending on the C^1-norm of u on $\partial B_{1/2}(0)$, hence finite by hypothesis. It follows in particular that, as announced,

$$\nabla u \in L^p(B_{1/4}(0)) \quad \text{for all } p < r. \qquad \square$$

Proposition C.2. *Let $u \in C^2(B_1(0) \setminus \{0\})$ solve*

(C-10)
$$\Delta u(x) = \mu(x) f(x) \quad \textit{in } B_1(0),$$

where $f \in L^p(B_1(0))$ for some $p > 2$. The weight μ satisfies

(C-11)
$$|\mu(x)| \simeq |x|^a \quad \textit{for some } a \in \mathbb{N}.$$

Then:

(i) *We have*[22]

(C-12)
$$\nabla u(x) = P(\bar{x}) + |\mu(x)|T(x),$$

where $P(\bar{x})$ is a complex-valued polynomial of degree at most a, and near the origin $T(x) = \mathrm{O}(|x|^{1-2/p-\epsilon})$ for every $\epsilon > 0$.

---

[21] Namely, $\||x|^{-1} * g\|^r_{L^{r,\infty}} \lesssim \|M_{2-\beta}g\|^{1-1/r}_{L^\infty}\|g\|^{1/r}_{L^1}$ for $r = (2-\beta)/(1-\beta)$ and $\beta \in (0,1)$.

[22] $\bar{x}$ is the complex conjugate of x. We parametrize $B_1(0)$ by $x = x_1 + ix_2$, and then $\bar{x} := x_1 - ix_2$. In this notation, ∇u in (C-12) is understood as $\partial_{x_1}u + i\partial_{x_2}u$.

(ii) *Furthermore, if* $\mu \in C^1(B_1(0) \setminus \{0\})$, *if* $a \neq 0$, *and if*

$$|x|^{1-a} \nabla \mu(x) \in L^\infty(B_1(0)), \tag{C-13}$$

we have

$$\nabla^2 u(x) = \nabla P(\bar{x}) + |\mu(x)| Q(x), \tag{C-14}$$

where P *is as in* (i), *and*

$$Q \in L^{p-\epsilon}(B_1(0), \mathbb{C}^2) \quad \textit{for all } \epsilon > 0.$$

As a (2×2) *real-valued matrix*, Q *satisfies in addition*

$$\operatorname{Tr} Q \in L^p(B_1(0)).$$

Naturally, if $a = 0$, *the standard Calderón–Zygmund theorem yields that* $u \in W^{2,p}(B_1(0))$. *The hypothesis* (C-13) *becomes unnecessary, and* (C-14) *holds with* P *being constant and* $\epsilon = 0$.

Proof. Using Green's formula for the Laplacian, an exact expression for the solution u may be found and used to obtain, for all $x \in B_1(0)$ and with $\vec{\nu}$ the outer normal unit vector to the boundary of $B_1(0)$,

$$\begin{aligned} \nabla u(x) = \frac{1}{2\pi} \int_{\partial B_1(0)} \biggl(\frac{x-y}{|x-y|^2} \partial_{\vec{\nu}} u(y) - u(y) \partial_{\vec{\nu}} \frac{x-y}{|x-y|^2} \biggr) \, d\sigma(y) \\ - \frac{1}{2\pi} \int_{B_1(0)} \frac{x-y}{|x-y|^2} \mu(y) f(y) \, dy \\ =: J_0(x) + J_1(x). \end{aligned} \tag{C-15}$$

Without loss of generality, and to avoid notational clutter, because u is twice differentiable away from the origin, we shall henceforth assume that $|x| < 1/2$.

We will estimate separately J_0 and J_1, and open the discussion by noting that, when $|y| > |x|$, we have the expansion

$$\frac{x-y}{|x-y|^2} = -\sum_{m \geq 0} P^m(x, y) \quad \text{with } P^m(x, y) := \bar{x}^m \bar{y}^{-(m+1)}.$$

Hence, we deduce the identity

$$\begin{aligned} J_0(x) &= -\frac{1}{2\pi} \sum_{m \geq 0} \int_{\partial B_1(0)} \bigl[P^m(x, y) \partial_{\vec{\nu}} u(y) - u(y) \partial_{\vec{\nu}} P^m(x, y) \bigr] \, dS(y) \\ &= -\frac{1}{2\pi} \sum_{m \geq 0} \bar{x}^m \int_0^{2\pi} \bigl[(m+1) u(\mathrm{e}^{i\varphi}) - (\partial_{\vec{\nu}} u)(\mathrm{e}^{i\varphi}) \bigr] \mathrm{e}^{i(m+1)\varphi} \, d\varphi \\ &= \sum_{m \geq 0} C_m \bar{x}^m, \end{aligned} \tag{C-16}$$

where the C_m are (complex-valued) constants depending only on the C^1-norm of u along $\partial B_1(0)$. As u is continuously differentiable on the boundary of the unit disk by hypothesis, and $|x| < 1$, it is clear that $|J_0(x)|$ is bounded above by some constant C for all $x \in B_1(0)$. Since $|C_m|$ grows sublinearly in m, we can surely find two constants γ and δ such that

$$|C_m| < \gamma\delta^m \quad \text{for all } m \geq 0.$$

Hence, when $|x| \leq R < \delta^{-1}$, we have

$$\Big|\sum_{m\geq a+1} C_m \bar{x}^m\Big| \leq \gamma\delta^{a+1}|x|^{a+1}\sum_{m\geq 0}(\delta R)^m \lesssim |x|^{a+1}.$$

And because J_0 is bounded, when $R < |x| < 1$, we find some large enough constant $K = K(C, a, \gamma, \delta)$ such that

$$\begin{aligned}\Big|\sum_{m\geq a+1} C_m \bar{x}^m\Big| &\leq |J_0(x)| + \sum_{0\leq m\leq a} C_m |x|^m \leq C + (a+1)\gamma\delta^a \\ &\leq K\delta^{a+1} \leq K(R^{-1}\delta)^{a+1}|x|^{a+1} \lesssim |x|^{a+1}.\end{aligned}$$

As by hypothesis $|\mu(x)| \simeq |x|^a$, we may now return to (C-16) and write

$$J_0(x) = P_0(\bar{x}) + |\mu(x)|T_0(x), \tag{C-17}$$

where P_0 is a polynomial of degree at most a, and the remainder T_0 is controlled by some constant depending on the C^1-norm of u on $\partial B_1(0)$. Moreover, $T_0(x) = \mathrm{O}(|x|)$ near the origin.

We next estimate the integral J_1. To do so, we proceed as above and write

$$J_1(x) = I_1(x) + \sum_{m=a+1}^{\infty} I_2^m(x) - \sum_{m=0}^{a} I_1^m(x) + \sum_{m=0}^{a} I_1^m(x) + I_2^m(x), \tag{C-18}$$

where we have put

$$\begin{aligned}I_1(x) &:= \frac{1}{2\pi}\int_{B_1(0)\cap B_{2|x|}(0)} \frac{x-y}{|x-y|^2}\mu(y)f(y)\,dy,\\ I_1^m(x) &:= \frac{1}{2\pi}\int_{B_1(0)\cap B_{2|x|}(0)} P^m(x,y)\mu(y)f(y)\,dy,\\ I_2^m(x) &:= \frac{1}{2\pi}\int_{B_1(0)\setminus B_{2|x|}(0)} P^m(x,y)\mu(y)f(y)\,dy.\end{aligned}$$

We first observe that the last sum in (C-18) may be written

$$P_1(x) := \sum_{0\leq m\leq a} I_1^m(x) + I_2^m(x) = \sum_{0\leq m\leq a}\int_{B_1(0)} P^m(x,y)\mu(y)f(y)\,dy = \sum_{0\leq m\leq a} A_m\bar{x}^m,$$

where

$$A_m := -\int_{B_1(0)} \overline{y}^{-(m+1)} \mu(y) f(y)\, dy.$$

From the fact that $f \in L^p(B_1(0))$ for $p > 2$, and the hypothesis $|\mu(y)| \simeq |y|^a$, it follows easily that $|A_m| < \infty$ for $m \le a$, and thus that P_1 is a polynomial of degree at most a.

We have next to handle the other summands appearing in (C-18), beginning with I_1. We find

$$\begin{aligned}|I_1(x)| &\lesssim |\mu(x)| \int_{B_{2|x|}(0)} \frac{|f(y)|}{|x-y|}\, dy \lesssim |\mu(x)| \int_{B_{3|x|}(x)} \frac{|f(y)|}{|x-y|}\, dy \\ &\lesssim |\mu(x)|\,|x| M_0 f(x) \lesssim |x|^{1-\frac{2}{p}} |\mu(x)|,\end{aligned} \tag{C-19}$$

where we have used the fact that $B_{2|x|}(0) \subset B_{3|x|}(x)$, and a classical estimate bounding convolution with the Riesz kernel by the maximal function[23] (see [Ziemer 1989, Proposition 2.8.2]). We have also used the simple estimate $M_0 f(x) \lesssim |x|^{-2/p} \|f\|_{L^p}$.

Next, let $q \in [1,2)$ be the conjugate exponent of p. We immediately deduce for $0 \le m \le a$ that

$$\begin{aligned}|I_1^m(x)| &\lesssim |x|^m \int_{B_{2|x|}(0)} |y|^{-1-m+a} |f(y)|\, dy \\ &\lesssim |x|^a \big\| |y|^{-1} \big\|_{L^q(B_{2|x|}(0))} \|f\|_{L^p(B_1(0))} \lesssim |x|^{1-\frac{2}{p}} |\mu(x)|.\end{aligned} \tag{C-20}$$

We next estimate I_2^m. As $m \ge a+1$, we note that, for any $\epsilon > 0$, we have

$$a + 1 - m - \epsilon - \frac{2}{p} < 0.$$

With again q being the conjugate exponent of p, we find thus

$$\begin{aligned}|I_2^m(x)| &\lesssim |x|^m \int_{B_1(0)\setminus B_{2|x|}(0)} |y|^{a-1-m} |f(y)|\, dy \\ &= |x|^m \int_{B_1(0)\setminus B_{2|x|}(0)} |y|^{a+1-m-\epsilon-\frac{2}{p}} |y|^{\epsilon - \frac{2}{q}} |f(y)|\, dy \\ &\le 2^{a+1-m-\epsilon-\frac{2}{p}} |x|^{a+1-\frac{2}{p}-\epsilon} \big\| |y|^{\epsilon-\frac{2}{q}} \big\|_{L^q(B_1(0))} \|f\|_{L^p(B_1(0))} \\ &\lesssim 2^{a+1-m-\epsilon-\frac{2}{p}} |x|^{1-\frac{2}{p}-\epsilon} |\mu(x)|.\end{aligned} \tag{C-21}$$

Combining altogether in (C-18) our findings (C-19)–(C-21), we obtain that

$$J_1(x) = P_1(\bar{x}) + |\mu(x)| T_1(x), \tag{C-22}$$

[23] See (C-8) for the definition of $M_0 f$.

where P_1 is a polynomial of degree at most a, and the remainder T_1 satisfies the estimate

$$|T_1(x)| \lesssim |x|^{1-\frac{2}{p}-\epsilon} \quad \text{for all } \epsilon > 0. \tag{C-23}$$

Altogether, (C-17) and (C-22) put into (C-15) show that we have

$$\nabla u(x) = P(\bar{x}) + |\mu(x)| T(x), \tag{C-24}$$

where $P := P_0 + P_1$ is a polynomial of degree at most a, and the remainder $T := T_0 + T_1$ satisfies the same estimate (C-23) as T_1. The announced statement (i) ensues immediately.

We prove next statement (ii). Comparing (C-14) to (C-24), we see that

$$\begin{aligned} |\mu(x)| Q(x) &= \nabla\big(|\mu(x)| T(x)\big) \\ &= \nabla\big(|\mu(x)| T_0(x)\big) + \nabla I_1(x) + \sum_{m\geq a+1} \nabla I_2^m(x) - \sum_{0\leq m\leq a} \nabla I_1^m(x). \end{aligned} \tag{C-25}$$

By definition,

$$|\mu(x)| T_0(x) = \sum_{m\geq a+1} C_m \bar{x}^m,$$

with the constants C_m depending only on the C^1-norm of u along $\partial B_1(0)$ and growing sublinearly in m. Using similar arguments to those leading to (C-17), it is clear from (C-11) that

$$|\mu(x)|^{-1} \nabla\big(|\mu(x)| T_0(x)\big) \in L^\infty(B_1(0)). \tag{C-26}$$

Controlling the gradients of I_1^m and I_2^m is done *mutatis mutandis* the estimates (C-20) and (C-21). For the sake of brevity, we only present in detail the case of I_1^m. Namely,

$$\begin{aligned} \nabla I_1^m(x) &= \frac{1}{2\pi} \int_{B_1(0)\cap B_{2|x|}(0)} \nabla_x P^m(x,y)\mu(y) f(y)\,dy \\ &\quad + \frac{1}{2\pi} \frac{x}{|x|} \otimes \int_{\partial B_{2|x|}(0)} P^m(x,y)\mu(y) f(y)\,dy. \end{aligned} \tag{C-27}$$

After some elementary computations, and using the hypothesis $|\mu(y)| \simeq |y|^a$, we reach

$$\begin{aligned} |\nabla I_1^m(x)| &\lesssim m|x|^{a-2} \int_{B_1(0)\cap B_{2|x|}(0)} |f(y)|\,dy + |x|^{a-1} \int_{\partial B_{2|x|}(0)} |f(y)|\,dy \\ &\lesssim m|x|^{a-\frac{2}{p}} \|f\|_{L^p(B_1(0))} + |x|^{a-1} \int_{\partial B_{2|x|}(0)} |f(y)|\,dy, \end{aligned}$$

so that immediately

$$\big\| |x|^{-a}\nabla I_1^m(x)\big\|_{L^{p-\epsilon}(B_1(0))} < \infty \quad \text{for all } \epsilon > 0.$$

Proceeding analogously for ∇I_2^m, we reach that for any $\epsilon > 0$ we have

$$\text{(C-28)} \quad \sum_{m\geq a+1} \big\| |x|^{-a}\nabla I_2^m(x)\big\|_{L^{p-\epsilon}(B_1(0))} + \sum_{0\leq m\leq a} \big\| |x|^{-a}\nabla I_1^m(x)\big\|_{L^{p-\epsilon}(B_1(0))} < \infty.$$

Hence, there remains only to estimate ∇I_1. This is slightly more delicate. For notational convenience, we write

$$\text{(C-29)} \quad \nabla I_1(x) = \frac{1}{2\pi}\nabla\int_{B_1(0)\cap B_{2|x|}(0)} \frac{x-y}{|x-y|^2}\mu(y)f(y)\,dy =: \frac{1}{2\pi}(L(x)+K(x)),$$

with

$$K(x) = \chi_{B_{1/2}(0)}(x)\frac{x}{|x|}\otimes\int_{\partial B_{2|x|}(0)} \frac{x-y}{|x-y|^2}\mu(y)f(y)\,dy,$$

and the convolution

$$L(x) = \big(\Omega * f(y)\mu(y)\chi_{B_1(0)\cap B_{2|x|}(0)}(y)\big)(x),$$

where Ω is the (2×2)-matrix made of the Calderón–Zygmund kernels:

$$\Omega(z) := \frac{|z|^2\mathbb{I}_2 - 2z\otimes z}{|z|^4}.$$

The boundary integral K is easily estimated:

$$|x|^{-a}|K(x)| \lesssim \frac{1}{|x|}\int_{\partial B_{2|x|}(0)} |f(y)|\,dy,$$

thereby yielding

$$\text{(C-30)} \quad \big\| |x|^{-a}K(x)\big\|_{L^p(B_1(0))} \lesssim \|f\|_{L^p(B_1(0))}.$$

To estimate L, we proceed as follows:

$$\text{(C-31)} \quad L(x) - \mu(x)\big(\Omega * f\chi_{B_1(0)\cap B_{2|x|}(0)}\big)(x) = \int_{B_1(0)\cap B_{2|x|}(0)} \Omega(x-y)f(y)(\mu(y)-\mu(x))\,dy.$$

Let S_x be the cone with apex the point $x/2$ and such that the disk $B_{|x|/4}(0)$ is inscribed in it. Note that, for $y\in S_x$, we have $2|x-y| > |x|$. Hence, we find

$$\text{(C-32)} \quad \int_{S_x\cap B_1(0)\cap B_{2|x|}(0)} \Omega(x-y)f(y)(\mu(y)-\mu(x))\,dy \lesssim |\mu(x)|\,|x|^{-2}\int_{B_{2|x|}(0)} |f(y)|\,dy.$$

By hypothesis, the function μ is continuously differentiable away from the origin. Thus, to each point y in the complement of the cone S_x, there corresponds some $\alpha \equiv \alpha(x, y) \in [0, 1]$ with

$$\mu(y) - \mu(x) = (x - y) \cdot \nabla\mu(\alpha x + (1 - \alpha)y).$$

Using (C-13), we deduce easily

$$|\mu(y) - \mu(x)| \lesssim |x|^{a-1}|x - y| \quad \text{for all } y \in S_x^c \cap B_1(0) \cap B_{2|x|}(0).$$

Accordingly, we have

$$\text{(C-33)} \quad \int_{S_x^c \cap B_1(0) \cap B_{2|x|}(0)} \Omega(x - y) f(y)(\mu(y) - \mu(x))\, dy \\ \lesssim |x|^{a-1} \int_{B_{2|x|}(0)} \frac{|f(y)|}{|x - y|}\, dy \lesssim |\mu(x)| M_0 f(x),$$

where we have used the same estimate as in (C-19). Bringing (C-32) and (C-33) into (C-31) and using the fact that $|\mu(x)| \simeq |x|^a$ yields

$$|\mu(x)|^{-1}|L(x)| \lesssim \big(\Omega * f(y)\chi_{B_1(0)\cap B_{2|x|}(0)}(y)\big)(x) \\ + \frac{1}{|x|^2} \int_{B_{2|x|}(0)} |f(y)|\, dy + M_0 f(x).$$

Because f is L^p, standard estimates on Calderón–Zygmund operators and on the maximal function, together with a classical Hardy inequality then give us

$$\big\| |\mu|^{-1} L \big\|_{L^p(B_1(0))} \lesssim \|f\|_{L^p(B_1(0))} < \infty.$$

Owing to the latter and to (C-30), we obtain from (C-29) that $|\mu|^{-1}\nabla I_1 \in L^p(B_1(0))$. With (C-26) and (C-28), the identity (C-25) thus implies that Q belongs to $L^{p-\epsilon}$ for all $\epsilon > 0$. This completes the first part of statement (ii).

We shall now prove the second part of (ii), and show that the trace of Q is in L^p. To this end, let us note that

$$\text{(C-34)} \qquad \operatorname{Tr} \nabla\bar{x} = \operatorname{Tr} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = 0.$$

We have seen in (C-25) that

$$\text{(C-35)} \qquad |\mu| Q = \nabla(|\mu| T_0) + \nabla I_1 + \sum_{m \geq a+1} \nabla I_2^m - \sum_{0 \leq m \leq a} \nabla I_1^m.$$

By definition, $|\mu(x)| T_0(x) = \sum_{m \geq a+1} C_m \bar{x}^m$, so that (C-34) gives

$$\text{(C-36)} \qquad \operatorname{Tr} \nabla\big(|\mu(x)| T_0(x)\big) = 0.$$

Owing to the fact that $P^m(x,y) = \bar{x}^m \bar{y}^{-(m+1)}$, it then easily follows from (C-34) and (C-27) that

$$\operatorname{Tr} \nabla I_1^m(x) = \frac{1}{2\pi} \operatorname{Tr} \frac{x}{|x|} \otimes \int_{\partial B_{2|x|}(0)} P^m(x,y)\mu(y) f(y)\, dy;$$

whence the estimate

$$|\mu(x)|^{-1} \big|\operatorname{Tr} \nabla I_1^m(x)\big| \lesssim 2^{a-m-1} \frac{1}{|x|} \int_{\partial B_{2|x|}(0)} |f(y)|\, dy,$$

and thus

$$\big\| |\mu|^{-1} \operatorname{Tr} \nabla I_1^m \big\|_{L^p} \lesssim 2^{a-m-1} \|f\|_{L^p}. \tag{C-37}$$

In exactly the same fashion, one finds

$$\big\| |\mu|^{-1} \operatorname{Tr} \nabla I_2^m \big\|_{L^p} \lesssim 2^{a-m-1} \|f\|_{L^p}. \tag{C-38}$$

Finally, there remains to handle the term $|\mu|^{-1} \operatorname{Tr} \nabla I_1$. But this term belongs to L^p, as we have shown that $|\mu|^{-1} \nabla I_1$ does. Using this in (C-35), together with (C-36)–(C-38), yields the announced result. □

Corollary C.3. *Let $u \in C^2(B_1(0) \setminus \{0\})$ solve*

$$\Delta u(x) = \mu(x) f(x) \quad \text{in } B_1(0),$$

where

$$|f(x)| \lesssim |x|^{n+r} \quad \text{and} \quad |\mu(x)| \simeq |x|^a$$

for two nonnegative integers n and a, and $r \in (0,1)$.

Then

$$\nabla u(x) = P(\bar{x}) + |\mu(x)| T(x), \tag{C-39}$$

where P is a complex-valued polynomial of degree at most $(a+n+1)$, and near the origin $T(x) = \mathrm{O}(|x|^{n+1+r-\epsilon})$ for every $\epsilon > 0$.

If in addition μ satisfies (C-13), then $|x|^{-(n+r)} |\mu|^{-1} \nabla(|\mu| T)$ belongs to L^p for all finite p. Furthermore, we have the estimate

$$\big|\operatorname{Tr} \nabla\big(|\mu(x)| T(x)\big)\big| \lesssim |x|^{n+r} |\mu(x)|. \tag{C-40}$$

Proof. The argument goes along the same lines as that of Proposition C.2. We set

$$\omega(x) := |x|^{n+r} \mu(x) \quad \text{and} \quad h(x) := |x|^{-(n+r)} f(x).$$

From the given hypotheses, we see that $h \in L^\infty$, and ω satisfies (C-11) with $(a+n+r)$ in place of a. If μ satisfies (C-13), then so does ω, again with $(a+n+r)$ in place of a.

Using the representation (C-15) gives

$$\begin{aligned}\nabla u(x) &= \frac{1}{2\pi}\int_{\partial B_1(0)}\left[\frac{x-y}{|x-y|^2}\partial_{\vec{\nu}}u(y) - u(y)\partial_{\vec{\nu}}\frac{x-y}{|x-y|^2}\right]d\sigma(y)\\ &\qquad - \frac{1}{2\pi}\int_{B_1(0)}\frac{x-y}{|x-y|^2}\omega(y)h(y)\,dy\\ &=: J_0(x)+J_1(x) \qquad\qquad \text{for all } x\in B_1(0),\end{aligned}$$

where $\vec{\nu}$ is the outer normal unit vector to the boundary of $B_1(0)$.

The integral J_0 is estimated as in (C-17) so as to yield

$$J_0(x) = P_0(\bar{x}) + |\mu(x)|T_0(x),$$

where P_0 is a polynomial of degree at most $(a+n+1)$, and $T_0(x) = \mathrm{O}(|x|^{n+2})$ with $|\mu|^{-1}\nabla(|\mu|T_0) = \mathrm{O}(|x|^{n+1})$.

We next estimate the integral J_1. We proceed again as we did in the proof of Proposition C.2. Namely,

$$J_1(x) = I_1(x) + \sum_{m=a+n+2}^{\infty} I_2^m(x) - \sum_{m=0}^{a+n+1} I_1^m(x) + \sum_{m=0}^{a+n+1} I_1^m(x) + I_2^m(x),$$

where we have put

$$\begin{aligned}I_1(x) &:= \frac{1}{2\pi}\int_{B_1(0)\cap B_{2|x|}(0)}\frac{x-y}{|x-y|^2}\omega(y)h(y)\,dy,\\ I_1^m(x) &:= \frac{1}{2\pi}\int_{B_1(0)\cap B_{2|x|}(0)} P^m(x,y)\omega(y)h(y)\,dy,\\ I_2^m(x) &:= \frac{1}{2\pi}\int_{B_1(0)\setminus B_{2|x|}(0)} P^m(x,y)\omega(y)h(y)\,dy.\end{aligned}$$

As before, $P^m(x,y) := \bar{x}^m\bar{y}^{-(m+1)}$. We first observe that the last sum in the expression for J_1 may be written

$$\begin{aligned}P_1(x) &:= \sum_{0\le m\le a+n+1} I_1^m(x) + I_2^m(x)\\ &= \sum_{0\le m\le a+n+1}\int_{B_1(0)} P^m(x,y)\omega(y)h(y)\,dy = \sum_{0\le m\le a+n+1} A_m\bar{x}^m,\end{aligned}$$

where

$$A_m := \int_{B_1(0)} \bar{y}^{-(m+1)}\omega(y)h(y)\,dy.$$

From the boundedness of h and the hypothesis $|\omega(y)|\simeq|y|^{a+n+r}$, it follows easily that $|A_m|<\infty$ for $m<a+n+1+r$, and thus, since $r>0$, that P_1 is a polynomial of degree at most $(a+n+1)$. Once this has been observed, the remainder of the proof follows *mutatis mutandis* that of Proposition C.2. Namely, we write

$$J_1(x) = P_1(\bar{x}) + |\omega(x)| T_1(x),$$

with $T_1(x) = \mathrm{O}(|x|^{1-\epsilon})$ for all $\epsilon > 0$. Moreover, $|\omega|^{-1}\nabla(|\omega| T_1) \in L^p$ for all $p < \infty$, and $|\omega|^{-1} \operatorname{Tr} \nabla(|\omega| T_1) \in L^\infty$.

Finally, setting $P = P_0 + P_1$ and $T = T_0 + |x|^{n+r} T_1 = \mathrm{O}(|x|^{n+r+1-\epsilon})$ gives the desired representation (C-39). Clearly, from (C-13) and the above, we have

$$\big| |\mu|^{-1} \nabla(|\mu| T) \big| \lesssim \big| |\mu|^{-1} \nabla(|\mu| T_0) \big| + |x|^{n+r} \big| |\omega|^{-1} \nabla(|\omega| T_1) \big|,$$

so that indeed $|x|^{-(n+r)} |\mu|^{-1} \nabla(|\mu| T)$ belongs to L^p for all finite p. Furthermore, we have, as announced,

$$\begin{aligned} \big| \operatorname{Tr} \nabla(|\mu| T) \big| &\leq \big| \operatorname{Tr} \nabla(|\mu| T_0) \big| + \big| \operatorname{Tr} \nabla(|\omega| T_1) \big| \\ &\lesssim |x|^{n+1} |\mu| + |\omega| \lesssim |x|^{n+r} |\mu|. \end{aligned}$$

□

References

[Adams 1975] D. R. Adams, "A note on Riesz potentials", *Duke Math. J.* **42**:4 (1975), 765–778. MR 56 #16361 Zbl 0336.46038

[Almeida 1995] L. Almeida, "The regularity problem for generalized harmonic maps into homogeneous spaces", *Calc. Var. Partial Differential Equations* **3**:2 (1995), 193–242. MR 97a:58040 Zbl 0820.58013

[Bernard and Rivière 2011a] Y. Bernard and T. Rivière, "Energy quantization for Willmore surfaces and applications", preprint, 2011. arXiv 1106.3780v1

[Bernard and Rivière 2011b] Y. Bernard and T. Rivière, "Local Palais–Smale sequences for the Willmore functional", *Comm. Anal. Geom.* **19**:3 (2011), 563–599. MR 2012j:58013 Zbl 1245.53051

[Burstall et al. 2002] F. E. Burstall, D. Ferus, K. Leschke, F. Pedit, and U. Pinkall, *Conformal geometry of surfaces in S^4 and quaternions*, Lecture Notes in Mathematics **1772**, Springer, Berlin, 2002. MR 2004a:53058 Zbl 1033.53001

[Dautray and Lions 1984] R. Dautray and J.-L. Lions, *Analyse mathématique et calcul numérique pour les sciences et les techniques*, Masson, Paris, 1984. Translated as *Mathematical analysis and numerical methods for science and technology, 3: Spectral theory and applications*, Springer, Berlin, 1990. MR 91h:00004a Zbl 0642.35001

[Ejiri 1988] N. Ejiri, "Willmore surfaces with a duality in $S^N(1)$", *Proc. London Math. Soc.* (3) **57**:2 (1988), 383–416. MR 89h:53117 Zbl 0671.53043

[Grüter and Widman 1982] M. Grüter and K.-O. Widman, "The Green function for uniformly elliptic equations", *Manuscripta Math.* **37**:3 (1982), 303–342. MR 83h:35033 Zbl 0485.35031

[Hélein 1996] F. Hélein, *Applications harmoniques, lois de conservation et repères mobiles*, Diderot, Paris, 1996. Translated as *Harmonic maps, conservation laws and moving frames*, 2nd ed., Cambridge Tracts in Mathematics **150**, Cambridge University Press, Cambridge, 2002. MR 2003g:58024 Zbl 0920.58022

[Kuwert and Schätzle 2001] E. Kuwert and R. Schätzle, "The Willmore flow with small initial energy", *J. Differential Geom.* **57**:3 (2001), 409–441. MR 2003a:53097 Zbl 1035.53092

[Kuwert and Schätzle 2004] E. Kuwert and R. Schätzle, "Removability of point singularities of Willmore surfaces", *Ann. of Math.* (2) **160**:1 (2004), 315–357. MR 2006c:58015 Zbl 1078.53007

[Kuwert and Schätzle 2007] E. Kuwert and R. Schätzle, "Branch points of Willmore surfaces", *Duke Math. J.* **138**:2 (2007), 179–201. MR 2008j:53108 Zbl 1130.53007

[Li and Yau 1982] P. Li and S. T. Yau, "A new conformal invariant and its applications to the Willmore conjecture and the first eigenvalue of compact surfaces", *Invent. Math.* **69**:2 (1982), 269–291. MR 84f:53049 Zbl 0503.53042

[Montiel 2000] S. Montiel, "Willmore two-spheres in the four-sphere", *Trans. Amer. Math. Soc.* **352**:10 (2000), 4469–4486. MR 2001b:53080 Zbl 0961.53035

[Müller and Šverák 1995] S. Müller and V. Šverák, "On surfaces of finite total curvature", *J. Differential Geom.* **42**:2 (1995), 229–258. MR 97b:53007 Zbl 0853.53003

[Rivière 2008] T. Rivière, "Analysis aspects of Willmore surfaces", *Invent. Math.* **174**:1 (2008), 1–45. MR 2009k:53154 Zbl 1155.53031

[Rivière 2010] T. Rivière, "Conformally invariant two-dimensional variational problems", preprint, Cours joint de l'Institut Henri Poincaré et Paris XII-Créteil, November 2010.

[Rivière 2013] T. Rivière, "Variational principles for immersed surfaces with L^2-bounded second fundamental form", *J. Reine Angew. Math.* (2013). arXiv 1007.2997v1

[Tartar 2007] L. Tartar, *An introduction to Sobolev spaces and interpolation spaces*, Lecture Notes of the Unione Matematica Italiana **3**, Springer, Berlin, 2007. MR 2008g:46055 Zbl 1126.46001

[Toro 1995] T. Toro, "Geometric conditions and existence of bi-Lipschitz parameterizations", *Duke Math. J.* **77**:1 (1995), 193–227. MR 96b:28006 Zbl 0847.42011

[Weiner 1978] J. L. Weiner, "On a problem of Chen, Willmore, et al.", *Indiana Univ. Math. J.* **27**:1 (1978), 19–35. MR 57 #7466 Zbl 0343.53038

[Ziemer 1989] W. P. Ziemer, *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, Graduate Texts in Mathematics **120**, Springer, New York, 1989. MR 91e:46046 Zbl 0692.46022

Received April 27, 2012. Revised January 10, 2013.

YANN BERNARD
MATHEMATISCHES INSTITUT
ALBERT-LUDWIGS-UNIVERSITÄT
79004 FREIBURG
GERMANY
yann.bernard@math.uni-freiburg.de

TRISTAN RIVIÈRE
DEPARTMENT OF MATHEMATICS
ETH ZENTRUM
8093 ZURICH
SWITZERLAND
riviere@math.ethz.ch

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.313

ON BACH FLAT WARPED PRODUCT EINSTEIN MANIFOLDS

QIANG CHEN AND CHENXU HE

We show that a compact warped product Einstein manifold with vanishing Bach tensor of dimension $n \geq 4$ is either Einstein or a finite quotient of a warped product with an $(n-1)$-dimensional Einstein fiber. The fiber has constant curvature if $n = 4$.

1. Introduction

Let $\lambda, m \in \mathbb{R}$ be constants. A $(\lambda, n+m)$-Einstein manifold is a complete Riemannian manifold with a smooth function f that satisfies the $(\lambda, n+m)$-Einstein equation

$$\mathrm{Ric}_f^m = \mathrm{Ric} + \mathrm{Hess}\, f - \frac{1}{m}\, df \otimes df = \lambda g. \tag{1-1}$$

When m is a positive integer, $(\lambda, n+m)$-Einstein metrics are exactly those n-dimensional manifolds which are the base of an $n+m$ dimensional Einstein warped product, i.e., $(M \times F^m, g + e^{-2f/m} g_F)$ is an Einstein manifold with Einstein constant λ, where (F^m, g_F) is another Einstein manifold [Besse 1987].

We call Ric_f^m in (1-1) the *m-Bakry–Emery tensor*. Lower bounds on this tensor are given by various comparison theorems for the measure $e^{-f}\, d\mathrm{vol}_g$; see, for example, [Villani 2009, Part II; Wei and Wylie 2009]. In view of these comparison theorems, the $(\lambda, n+m)$-Einstein equation is the natural Einstein condition of having a constant m-Bakry–Emery Ricci tensor. The study of $(\lambda, n+1)$-Einstein metrics often needs special techniques, so we do not consider these metrics here: *we assume $m \neq 1$ throughout this paper*. Taking $m \to \infty$, one also obtains the gradient Ricci soliton equation

$$\mathrm{Ric} + \mathrm{Hess}\, f = \lambda g.$$

We could then also call a gradient Ricci soliton a (λ, ∞)-Einstein manifold. Ricci solitons have been studied because of their connection to Ricci flow and because they are a natural generalization of Einstein manifolds. We refer to the survey paper [Cao 2010] for recent progress on this subject.

MSC2000: 53B20, 53C21, 53C25.
Keywords: warped product metric, Einstein manifold, Bach flat, Bakry–Emery tensor.

In a series of papers with P. Petersen and W. Wylie, the second author studied warped product Einstein manifolds under various curvature and symmetry conditions [He et al. 2010; 2011; 2012]. Many interesting results on gradient Ricci solitons have also been obtained on warped product Einstein manifolds. In [He et al. 2010] we found nontrivial examples of manifolds on homogeneous spaces that are neither Einstein nor products of Einstein manifolds. This contrasts with the gradient Ricci soliton case, where all homogeneous gradient Ricci solitons are Einstein or products of Einstein manifolds.

In this paper, we consider an interesting class of complete warped product Einstein manifolds: those with a vanishing Bach tensor. This well-known tensor was first introduced by R. Bach [1921] to study conformal relativity. On any Riemannian manifold $(M^n, g)(n \geq 4)$, the Bach tensor is defined by

$$B(X, Y) = \frac{1}{n-3}(\nabla^2_{E_i, E_j} W)(X, E_i, E_j, Y) + \frac{1}{n-2} \operatorname{Ric}(E_i, E_j) W(X, E_i, E_j, Y).$$

Here, $\{E_i\}_{i=1}^n$ is an orthonormal frame, $\nabla^2_{E_i, E_j}$ is the covariant derivative of tensors, and W is the Weyl curvature tensor. If the manifold is Einstein or locally conformally flat, the Bach tensor vanishes. The dimension 4 is most interesting since on any compact 4-manifold (M^4, g), Bach flat metrics are precisely the critical points of the conformally invariant functional on the space of metrics

$$\mathcal{W}(g) = \int_M |W_g|^2 \, d\mathrm{vol}_g,$$

where W_g is the Weyl tensor of the metric g. Other than Einstein and locally conformally flat metrics, there are two more classes of compact 4-manifolds with a vanishing Bach tensor: metrics that are locally conformal to an Einstein one, and half conformally flat metrics (self-dual or anti-self-dual) if M^4 is orientable. The aim of this paper is to show that a stronger converse of warped product Einstein metrics holds. The proof is motivated by a recent corresponding result on gradient Ricci solitons [Cao and Chen 2011].

Theorem 1.1. *Suppose (M^4, g, f) is a compact $(\lambda, 4+m)$-Einstein manifold with $m \neq 0, 1,$ or -2. If the Bach tensor B vanishes everywhere, then M is locally conformally flat.*

Theorem 1.1 is a direct consequence of the following more general result combined with some early results in [He et al. 2012].

Theorem 1.2. *Suppose $(M^n, g, f)(n \geq 4)$ is a compact $(\lambda, n+m)$-Einstein manifold with $m \neq 0, 1,$ or $2-n$. If the Bach tensor B vanishes everywhere, then M has a harmonic Weyl tensor and $W(X, Y, Z, \nabla f) = 0$ for any vector fields $X, Y,$ and Z.*

Remark 1.3. Theorem 1.2 is analogous to [Cao and Chen 2011, Theorem 5.1].

Remark 1.4. In the case where $m = 2 - n$, our argument breaks down, since some coefficient vanishes in the key identity (3-1). On the other hand, it is observed in [Catino et al. 2011] that in this case, a $(\lambda, n + (2 - n))$-Einstein metric is globally conformally Einstein. In particular, it has a vanishing Bach tensor when $n = 4$.

Remark 1.5. Böhm [1998] constructed compact, rotationally symmetric $(\lambda, n+m)$-Einstein metrics on $\mathbb{S}^n$ for $n = 3, 4, 5, 6, 7$ that are not Einstein. This is in sharp contrast to the gradient Ricci solitons. These examples also show that our conclusion in dimension 4 cannot be strengthened.

Remark 1.6. Theorem 1.1 was first obtained by G. Catino [2012] under a stronger assumption that (M^4, g) is half conformally flat.

If m is positive, then from [Kim and Kim 2003] and the comparison theorem of m-Bakry–Emery tensors in [Qian 1997, Theorem 5], a $(\lambda, n+m)$-Einstein manifold is compact if and only if $\lambda > 0$. Using [He et al. 2012, Theorem 1.5], the global classification of warped product Einstein manifolds with harmonic Weyl tensor, and $W(\nabla f, \cdot, \cdot, \nabla f) = 0$, Theorem 1.2 has the following corollary:

Corollary 1.7. *Let $m \neq 1$ be a positive number. Suppose that $(M^n, g, f)(n \geq 4)$ is a simply connected $(\lambda, n+m)$-Einstein manifold with $\lambda > 0$ and has a vanishing Bach tensor. Then (M^n, g, f) is either*

(1) *Einstein with constant function f, or*

(2) *$g = dt^2 + \psi^2(t) g_L$ and $f = f(t)$, where g_L is Einstein with nonnegative Ricci curvature, and has constant curvature if $n = 4$.*

Remark 1.8. In the proof of [He et al. 2012, Theorem 1.5], the authors made the assumption that $m > 1$. In fact, the whole argument carries over for the case when $0 < m < 1$.

Remark 1.9. Since m can be an arbitrary constant in our definition of a $(\lambda, n+m)$-Einstein manifold, we would like to discuss the case when $m < 0$. In this case, a $(\lambda, n+m)$-Einstein manifold with positive λ is not necessarily compact, as the proof of the comparison theorem [Qian 1997, Theorem 5] is no longer valid. On the other hand, if we assume further that $m \neq 2 - n$, the argument of the local classification result in [He et al. 2012, Theorem 7.9] carries over (see also [Catino et al. 2011] and Remark 3.4). The global result in [He et al. 2012, Theorem 1.5] also holds, except for the statement that (L, g_L) has nonnegative Ricci curvature when $\lambda \geq 0$. The argument of that statement relies on the positivity of m (see [He et al. 2012, Theorem 7.10]).

The paper is organized as follows. In Section 2, we recall definitions and basic properties of Bach, Cotton, and Weyl tensors, and the D-tensor defined in [Cao and Chen 2011]. We also list some relevant properties of warped product Einstein metrics. In Section 3, we show how the D-tensor characterizes the geometry of the level set of f — see Proposition 3.3. In Section 4, we prove Theorem 1.1 and Theorem 1.2.

2. Preliminaries

In this section, we set up our notation and recall some well-known facts on warped product Einstein manifolds. For more detail, see for example [He et al. 2012] and references therein.

We use the convention that the Riemann curvature tensor $R(X, Y, Y, X)$ has the same sign as the sectional curvature of the 2-plane spanned by X and Y. For $n \geq 4$, the Weyl curvature tensor is defined as

$$R = W + \frac{2}{n-2}\operatorname{Ric} \odot g - \frac{\text{scal}}{(n-1)(n-2)} g \odot g,$$

where, for two symmetric $(0, 2)$-tensors s and r, we define the Kulkarni–Nomizu product $s \odot r$ to be the $(0, 4)$-tensor

$$\begin{aligned}&(s \odot r)(X, Y, Z, W)\\ &\quad = \tfrac{1}{2}\big(r(X, W)s(Y, Z) + r(Y, Z)s(X, W) - r(X, Z)s(Y, W) - r(Y, W)s(X, Z)\big).\end{aligned}$$

Recall that for any $X, Y \in TM$, the Bach tensor B is the symmetric $(0, 2)$-tensor defined by

$$B(X, Y) = \frac{1}{n-3}\sum_{i,j}(\nabla^2_{E_i, E_j} W)(X, E_i, E_j, Y) + \frac{1}{n-2}\sum_{i,j}\operatorname{Ric}(E_i, E_j) W(X, E_i, E_j, Y), \tag{2-1}$$

where $\{E_i\}_{i=1}^n$ is an orthonormal frame and $\nabla^2_{E_i, E_j} W$ is the covariant derivative of the Weyl tensor.

The Schouten tensor is the $(0, 2)$-tensor

$$S = \operatorname{Ric} - \frac{\text{scal}}{2(n-1)} g$$

and the Cotton tensor C is defined as

$$C(X, Y, Z) = (\nabla_X S)(Y, Z) - (\nabla_Y S)(X, Z) \quad \text{for any } X, Y, Z \in TM.$$

Using the fact that $(\operatorname{div} R)(X, Y, Z) = (\nabla_X \operatorname{Ric})(Y, Z) - (\nabla_Y \operatorname{Ric})(X, Z)$, we have

$$\begin{aligned}&C(X, Y, Z)\\ &\quad = (\operatorname{div} R)(X, Y, Z) - \frac{1}{2(n-1)}\big((\nabla_X \text{scal}) g(Y, Z) - (\nabla_Y \text{scal}) g(X, Z)\big).\end{aligned} \tag{2-2}$$

Definition 2.1. A Riemannian manifold (M^n, g) has a *harmonic Weyl tensor* if the Cotton tensor vanishes.

Remark 2.2. For $n \geq 4$, the Cotton tensor is, up to a constant factor, the divergence of the Weyl tensor

$$C(X, Y, Z) = \frac{n-2}{n-3}(\operatorname{div} W)(X, Y, Z) \quad \text{for any } X, Y, Z \in TM. \tag{2-3}$$

So we can rewrite the Bach tensor as

$$B(X,Y) = \frac{1}{n-2}\Big(\sum_i (\nabla_{E_i} C)(E_i, X, Y) + \sum_{i,j} \operatorname{Ric}(E_i, E_j) W(X, E_i, E_j, Y)\Big), \tag{2-4}$$

where $\{E_i\}_{i=1}^n$ is an orthonormal frame.

Remark 2.3. If $n = 3$, then $W = 0$, and thus a harmonic Weyl tensor is equivalent to (M^3, g) being locally conformally flat. If $n \geq 4$, then M has a harmonic Weyl tensor if and only if $\operatorname{div} W = 0$, and M is locally conformally flat if and only if $W = 0$.

On a $(\lambda, n+m)$-Einstein manifold (M, g, f) for any $X, Y, Z \in TM$ we define the D-tensor, which is identical to the one in [Cao and Chen 2011] for Ricci solitons, as follows:

$$\begin{aligned} D(X, Y, Z) = {} & \frac{1}{(n-1)(n-2)}\big(\operatorname{Ric}(X, \nabla f) g(Y, Z) - \operatorname{Ric}(Y, \nabla f) g(X, Z)\big) \\ & + \frac{1}{n-2}\big(\operatorname{Ric}(Y, Z) g(X, \nabla f) - \operatorname{Ric}(X, Z) g(Y, \nabla f)\big) \\ & - \frac{\operatorname{scal}}{(n-1)(n-2)}\big(g(X, \nabla f) g(Y, Z) - g(Y, \nabla f) g(X, Z)\big). \end{aligned} \tag{2-5}$$

Both C and D are skew-symmetric in their first two indices and trace-free in any two indices:

$$C(X, Y, Z) = -C(Y, X, Z), \quad \sum_i C(E_i, E_i, X) = \sum_i C(E_i, X, E_i) = 0,$$
$$D(X, Y, Z) = -D(Y, X, Z), \quad \sum_i D(E_i, E_i, X) = \sum_i D(E_i, X, E_i) = 0.$$

Next we recall some properties of warped product Einstein manifolds; the proofs can be found in [He et al. 2012]. The function ρ is defined by

$$\operatorname{scal} = (n-1)\lambda - (m-1)\rho.$$

Note that a $(\lambda, n+1)$-Einstein manifold has constant scalar curvature $(n-1)\lambda$. The modified Ricci and Riemann curvature tensors are defined by

$$P = \operatorname{Ric} + \rho g$$

and

$$Q = R + \frac{2}{m}\,\mathrm{Ric} \odot g - \frac{\lambda+\rho}{m} g \odot g = R + \frac{2}{m} P \odot g + \frac{\rho-\lambda}{m} g \odot g.$$

Proposition 2.4. *Let (M, g, f) be a $(\lambda, n+m)$-Einstein manifold with $m \neq 1$. Then*

$$P(\nabla f) = -\frac{m}{2}\nabla\rho, \text{ or equivalently } \mathrm{Ric}(\nabla f) = -\frac{m}{2}\nabla\rho + \rho\nabla f; \tag{2-6}$$

$$(\mathrm{div}\, R) = Q(X, Y, Z, \nabla f) - \frac{1}{m}(g \odot g)(X, Y, Z, P(\nabla f)). \tag{2-7}$$

Equation (2-6) is (3.12) in [Case et al. 2011], and (2-7) was shown in [He et al. 2012, Proposition 6.3].

3. The covariant 3-tensor D

In this section, we extend some known results regarding the 3-tensor D from gradient Ricci solitons to warped product Einstein manifolds. Since the $(\lambda, n+m)$-Einstein equation (1-1) contains the extra term in $df \otimes df$, we provide the calculations in detail, though we essentially follow proofs in [Cao and Chen 2011; 2012].

For gradient Ricci solitons, the D tensor relates the Cotton tensor and Weyl tensor in the following way (see [Cao and Chen 2011, Lemma 3.1]):

$$C(X, Y, Z) = D(X, Y, Z) + W(X, Y, Z, \nabla f) \quad \text{for any } X, Y, Z \in TM.$$

For warped product manifolds, we have a similar relation for these three tensors. (The case $n = 4$ was given in [Catino 2012].)

Lemma 3.1. *Suppose (M^n, g, f) is a $(\lambda, n+m)$-Einstein manifold. The Cotton tenor C, D-tensor, and Weyl tensor W satisfy the identity*

$$C(X, Y, Z) = W(X, Y, Z, \nabla f) + \frac{m+n-2}{m} D(X, Y, Z) \quad \textit{for } X, Y, Z \in TM. \tag{3-1}$$

Proof. From the formula (2-7) for div R, the definition of Q-tensor, and the decomposition curvature tensor R, we have (with a dash standing for the three first arguments X, Y, Z)

$$\begin{aligned}
&(\mathrm{div}\, R)(\text{-}) \\
&= Q(\text{-}, \nabla f) - \frac{1}{m}(g \odot g)(\text{-}, P(\nabla f)) \\
&= R(\text{-}, \nabla f) + \frac{2}{m}(\mathrm{Ric} \odot g)(\text{-}, \nabla f) - \frac{\lambda+\rho}{m}(g \odot g)(\text{-}, \nabla f) - \frac{1}{m}(g \odot g)(\text{-}, P(\nabla f)) \\
&= W(\text{-}, \nabla f) + \frac{2}{n-2}(\mathrm{Ric} \odot g)(\text{-}, \nabla f) - \frac{(n-1)\lambda - (m-1)\rho}{(n-1)(n-2)}(g \odot g)(\text{-}, \nabla f) \\
&\qquad + \frac{2}{m}(\mathrm{Ric} \odot g)(\text{-}, \nabla f) - \frac{\lambda+\rho}{m}(g \odot g)(\text{-}, \nabla f) - \frac{1}{m}(g \odot g)(\text{-}, P(\nabla f))
\end{aligned}$$

$$= W(\text{-}, \nabla f) - \frac{1}{m}(g \odot g)(\text{-}, P(\nabla f)) + \frac{2(m+n-2)}{m(n-2)}(\mathrm{Ric} \odot g)(\text{-}, \nabla f)$$
$$- \frac{(n-1)(m+n-2)\lambda + \big((n-1)(n-2) - m(m-1)\big)\rho}{m(n-1)(n-2)}(g \odot g)(\text{-}, \nabla f).$$

Using the fact that $P = \mathrm{Ric} - \rho g$, we have

$$\begin{aligned}
&(\operatorname{div} R)(X, Y, Z) \\
&\quad = W(X, Y, Z, \nabla f) + \frac{1}{n-2}\big(\mathrm{Ric}(X, \nabla f) g(Y, Z) - \mathrm{Ric}(Y, \nabla f) g(X, Z)\big) \\
&\qquad + \frac{m+n-2}{m(n-2)}\big(\mathrm{Ric}(Y, Z) g(X, \nabla f) - \mathrm{Ric}(X, Z) g(Y, \nabla f)\big) \\
&\qquad - \frac{(n-1)(m+n-2)\lambda - m(m-1)\rho}{m(n-1)(n-2)}\big(g(X, \nabla f) g(Y, Z) - g(Y, \nabla f) g(X, Z)\big).
\end{aligned}$$

From the formula (2-6) for $\mathrm{Ric}(\nabla f)$, we have

$$\begin{aligned}
&(\operatorname{div} R)(X, Y, Z, \nabla f) - W(X, Y, Z, \nabla f) \\
&\quad = -\frac{m}{2(n-2)}\big((\nabla_X \rho) g(Y, Z) - (\nabla_Y \rho) g(X, Z)\big) \\
&\qquad + \frac{m+n-2}{m(n-2)}\big(\mathrm{Ric}(Y, Z) g(X, \nabla f) - \mathrm{Ric}(X, Z) g(Y, \nabla f)\big) \\
&\qquad - \frac{m+n-2}{m(n-1)(n-2)}\big((n-1)\lambda - m\rho\big)\big(g(X, \nabla f) g(Y, Z) - g(Y, \nabla f) g(X, Z)\big).
\end{aligned}$$

From the defining equation (2-2) of the Cotton tensor C and using $\mathrm{scal} = (n-1)\lambda - (m-1)\rho$, we have

$$C(X, Y, Z) = (\operatorname{div} R)(X, Y, Z) + \frac{m-1}{2(n-1)}\big((\nabla_X \rho) g(Y, Z) - (\nabla_Y \rho) g(X, Z)\big),$$

and then

$$\begin{aligned}
&\frac{m}{m+n-2}\big(C(X, Y, Z) - W(X, Y, Z, \nabla f)\big) \\
&\qquad = -\frac{m}{2(n-1)(n-2)}\big((\nabla_X \rho) g(Y, Z) - (\nabla_Y \rho) g(X, Z)\big) \\
&\qquad\quad + \frac{1}{n-2}\big(\mathrm{Ric}(Y, Z) g(X, \nabla f) - \mathrm{Ric}(X, Z) g(Y, \nabla f)\big) \\
&\qquad\quad - \frac{(n-1)\lambda - m\rho}{(n-1)(n-2)}\big(g(X, \nabla f) g(Y, Z) - g(Y, \nabla f) g(X, Z)\big),
\end{aligned}$$

which is exactly equal to $D(X, Y, Z)$ by the formula of $\mathrm{Ric}(\nabla f)$. □

For gradient Ricci solitons, one amazing fact is that the norm of the D-tensor is linked to the geometry of the level set of the potential function f (see [Cao and Chen 2012, (4.5); Cao and Chen 2011, Lemma 3.2]). We have the following extension to warped product Einstein manifolds.

Lemma 3.2. *Suppose (M^n, g, f) is a $(\lambda, n+m)$-Einstein manifold. Let Σ^{n-1} be a level set of f with $\nabla f(p) \neq 0$, and let $h_{ab}(a, b = 2, \ldots, n)$ and $H = (n-1)\sigma$ be its second fundamental form and mean curvature, respectively. Then we have*

$$|D|^2 = \frac{2|\nabla f|^4}{(n-2)^2} \sum_{a,b=2}^{n} |h_{ab} - \sigma g_{ab}|^2 + \frac{m^2}{2(n-1)(n-2)(m-1)^2} |\nabla^\Sigma \mathrm{scal}|^2, \tag{3-2}$$

where $|\nabla^\Sigma \mathrm{scal}|^2 = |\nabla \mathrm{scal}|^2 - \big(\nabla \mathrm{scal} \cdot (\nabla f / |\nabla f|)\big)^2$.

Proof. Let $\{e_i\}_{i=1}^n$ be an orthonormal frame with $e_1 = \nabla f / |\nabla f|$ at the point where ∇f is nonzero. The second fundamental form h_{ab} and the mean curvature H of the level hypersurface Σ are given by

$$\begin{aligned} h_{ab} &= g\Big(\nabla_{e_a} \frac{\nabla f}{|\nabla f|}, e_b\Big) = \frac{1}{|\nabla f|} \nabla_{e_a} \nabla_{e_b} f = \frac{1}{|\nabla f|} \big(\lambda g_{ab} - \mathrm{Ric}(e_a, e_b)\big), \\ H &= \frac{1}{|\nabla f|} \big((n-1)\lambda - \mathrm{scal} + \mathrm{Ric}(e_1, e_1)\big). \end{aligned}$$

So we have

$$\begin{aligned} \sum_{a,b=2}^{n} |h_{ab}|^2 &= \frac{1}{|\nabla f|^2} \sum_{a,b=2}^{n} |\lambda g_{ab} - \mathrm{Ric}(e_a, e_b)|^2 \\ &= \frac{1}{|\nabla f|^2} \Big((n-1)\lambda^2 - 2\lambda\big(\mathrm{scal} - \mathrm{Ric}(e_1, e_1)\big) + \sum_{a,b=2}^{n} |\,\mathrm{Ric}(e_a, e_b)|^2 \Big), \end{aligned}$$

$$H^2 = \frac{1}{|\nabla f|^2} \Big((n-1)^2\lambda^2 - 2(n-1)\lambda\big(\mathrm{scal} - \mathrm{Ric}(e_1, e_1) + (\mathrm{scal} - \mathrm{Ric}(e_1, e_1))^2\big)\Big).$$

From $\mathrm{Ric}(\nabla f) = \rho \nabla f - \frac{m}{2} \nabla \rho$, it follows that

$$\begin{aligned} R_{11} &= \mathrm{Ric}(e_1, e_1) = \rho - \frac{m}{2|\nabla f|^2} \nabla \rho \cdot \nabla f, \\ R_{1a} &= \mathrm{Ric}(e_1, e_a) = -\frac{m}{2|\nabla f|} \nabla_a \rho. \end{aligned}$$

So we have

$$\begin{aligned} &\sum_{a,b=2}^{n} |h_{ab} - \sigma g_{ab}|^2 \\ &\qquad = \frac{1}{|\nabla f|^2} |\mathrm{Ric}|^2 - \frac{2}{|\nabla f|^2} \sum_{a=2}^{n} R_{1a}^2 - \frac{1}{|\nabla f|^2} R_{11}^2 - \frac{1}{(n-1)|\nabla f|^2} (\mathrm{scal} - R_{11})^2, \end{aligned}$$

where

$$\begin{aligned}
-\frac{2}{|\nabla f|^2}\sum_{a=2}^{n} R_{1a}^2 &= -\frac{m^2}{2|\nabla f|^4}|\nabla\rho|^2 + \frac{m^2}{2|\nabla f|^6}(\nabla\rho\cdot\nabla f)^2,\\
-\frac{1}{|\nabla f|^2}R_{11}^2 &= -\frac{1}{|\nabla f|^2}\rho^2 + \frac{m\rho}{|\nabla f|^4}\nabla\rho\cdot\nabla f - \frac{m^2}{4|\nabla f|^6}(\nabla\rho\cdot\nabla f)^2,\\
\mathrm{scal} - R_{11} &= (n-1)\lambda - m\rho + \frac{m}{2|\nabla f|^2}\nabla\rho\cdot\nabla f,
\end{aligned}$$

and

$$\begin{aligned}
-\frac{1}{(n-1)|\nabla f|^2}(\mathrm{scal}-R_{11})^2 = -\frac{(n-1)\lambda^2}{|\nabla f|^2} - \frac{m^2\rho^2}{(n-1)|\nabla f|^2} + \frac{2m\lambda\rho}{|\nabla f|^2}\\
-\frac{m^2}{4(n-1)|\nabla f|^6}(\nabla\rho\cdot\nabla f)^2 + \frac{m\big(m\rho-(n-1)\lambda\big)}{(n-1)|\nabla f|^4}\nabla\rho\cdot\nabla f.
\end{aligned}$$

Adding them together yields

$$\begin{aligned}
&\sum_{a,b=2}^{n} |h_{ab} - \sigma g_{ab}|^2\\
&= \frac{1}{|\nabla f|^2}|\mathrm{Ric}|^2 - \frac{m^2}{2|\nabla f|^4}|\nabla\rho|^2 + \frac{m^2(n-2)}{4(n-1)|\nabla f|^6}(\nabla\rho\cdot\nabla f)^2\\
&\quad + \frac{m\big((m+n-1)\rho-(n-1)\lambda\big)}{(n-1)|\nabla f|^4}\nabla\rho\cdot\nabla f - \frac{m^2+n-1}{(n-1)|\nabla f|^2}\rho^2 + \frac{2m}{|\nabla f|^2}\lambda\rho - \frac{n-1}{|\nabla f|^2}\lambda^2.
\end{aligned}$$

Let $D_{ijk} = D(e_i, e_j, e_k)$. Then we have

$$D_{ijk} = b_1(\nabla_i\rho\delta_{jk} - \nabla_j\rho\delta_{ik}) + b_2(\nabla_i f R_{jk} - \nabla_j f R_{ik}) + b_3(\nabla_i f\delta_{jk} - \nabla_j f\delta_{ik}),$$

where $\nabla_i = \nabla_{e_i}$ and

$$b_1 = -\frac{m}{2(n-1)(n-2)},\quad b_2 = \frac{1}{n-2},\quad b_3 = -\frac{(n-1)\lambda - m\rho}{(n-1)(n-2)}.$$

So we have

$$\begin{aligned}
|D|^2 &= \sum_{i,j,k=1}^{n} D_{ijk}^2\\
&= b_1^2\big(2(n-1)|\nabla\rho|^2\big) + b_2^2\big(2|\nabla f|^2|\mathrm{Ric}|^2 - 2\,\mathrm{Ric}^2(\nabla f,\nabla f)\big)\\
&\qquad + b_3^2\big(2(n-1)|\nabla f|^2\big) + 2b_1b_2\big(2\mathrm{scal}\nabla\rho\cdot\nabla f - 2\,\mathrm{Ric}(\nabla f,\nabla\rho)\big)\\
&\qquad + 2b_1b_3\big(2(n-1)\nabla\rho\cdot\nabla f\big) + 2b_2b_3\big(2|\nabla f|^2\mathrm{scal} - 2\,\mathrm{Ric}(\nabla f,\nabla f)\big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2(n-1)(n-2)^2}\big(4(n-1)|\nabla f|^2|\mathrm{Ric}|^2 - m^2 n|\nabla\rho|^2\big)\\
&\quad + \frac{4m\big((m+n-1)\rho-(n-1)\lambda\big)}{2(n-1)(n-2)^2}(\nabla\rho\cdot\nabla f)\\
&\quad - \frac{4\big(((n-1)\lambda-m\rho)^2+(n-1)\rho^2\big)}{2(n-1)(n-2)^2}|\nabla f|^2.
\end{aligned}
$$

A straightforward computation shows that

$$
|D|^2 = \frac{2|\nabla f|^4}{(n-2)^2}\sum_{a,b=2}^{n}|h_{ab}-\sigma g_{ab}|^2 + \frac{m^2}{2(n-1)(n-2)}|\nabla^\Sigma\rho|^2.
$$

Substituting the function ρ by scal gives us the desired identity. □

Similarly, the vanishing of the D-tensor implies many nice properties about the geometry of the warped product Einstein manifold (M^n, g, f) and the level sets of f.

Proposition 3.3. *Suppose $(M^n, g, f)(n \geq 3)$ is a $(\lambda, n+m)$-Einstein manifold with $m \neq 1$ and $D = 0$. Let c be a regular value of f and $\Sigma = \{x \in M \mid f(x) = c\}$ be the level hypersurface of f. Then*:

(1) *Both the scalar curvature and $|\nabla f|^2$ are constant on Σ.*

(2) *On Σ, the Ricci tensor has either a unique eigenvalue or two distinct eigenvalues with multiplicity 1 and $n-1$; moreover, the eigenvalue with multiplicity 1 is in the direction of ∇f.*

(3) *The second fundamental form h_{ab} of Σ is of the form $h_{ab} = \dfrac{H}{n-1}g_{ab}$.*

(4) *The mean curvature H is constant on Σ.*

(5) *$R(\nabla f, X, Y, Z) = 0$ for any vectors X, Y, and Z tangent to Σ.*

Proof. It follows the argument in the proof of [Cao and Chen 2011, Proposition 3.1] by using Lemma 3.2. □

Remark 3.4. If a $(\lambda, n+m)$-Einstein manifold with $m \neq 2-n$ has a harmonic Weyl tensor and $W(\nabla f, \cdot, \cdot, \cdot)$, then the D-tensor vanishes by Lemma 3.1. So Proposition 3.3 offers an alternative proof of [He et al. 2012, Theorem 7.9], which is the main step for the global classification in [ibid., Theorem 7.10].

4. The proof of Theorems 1.1 and 1.2

In this section, we first prove Theorem 1.2, which says that a compact Bach flat $(\lambda, n+m)$-Einstein manifold with $m \neq 0$, 1, or $2-n$ has a harmonic Weyl tensor and $W(X, Y, Z, \nabla f) = 0$ for any $X, Y, Z \in TM$. Then Theorem 1.1 follows by using [He et al. 2012, Theorem 7.9].

Proof of Theorem 1.2. We follow the argument in [Cao and Chen 2011]. Fix a point $p \in M$ and assume that $\{E_i\}_{i=1}^n$ is an orthonormal frame with $\nabla E_i(p) = 0$. Using (2-3), the Bach tensor equation (2-4), and Lemma 3.1, a direct computation shows that for any $X, Y \in TM$, we have

$$
\begin{aligned}
(n-2)B(X,Y) &= \sum_i (\nabla_{E_i} C)(E_i, X, Y) + \sum_{i,j} \mathrm{Ric}(E_i, E_j) W(X, E_i, E_j, Y) \\
&= (\nabla_{E_i} W)(E_i, X, Y, \nabla f) + W(E_i, X, Y, \nabla_{E_i} \nabla f) \\
&\qquad + \frac{m+n-2}{m} (\nabla_{E_i} D)(E_i, X, Y) + \mathrm{Ric}(E_i, E_j) W(X, E_i, E_j, Y) \\
&= (\operatorname{div} W)(\nabla f, Y, X) + \frac{m+n-2}{m} (\nabla_{E_i} D)(E_i, X, Y) \\
&\qquad + W(X, E_i, E_j, Y) \big(\mathrm{Ric}(E_i, E_j) + \mathrm{Hess}\, f(E_i, E_j)\big) \\
&= \frac{n-3}{n-2} C(\nabla f, Y, X) + \frac{m+n-2}{m} (\nabla_{E_i} D)(E_i, X, Y) \\
&\qquad + W(X, E_i, E_j, Y) \Big(\frac{1}{m} g(\nabla f, E_i) g(\nabla f, E_j) + \lambda g(E_i, E_j)\Big) \\
&= \frac{n-3}{n-2} C(\nabla f, Y, X) + \frac{m+n-2}{m} (\nabla_{E_i} D)(E_i, X, Y) \\
&\qquad + \frac{1}{m} W(\nabla f, X, Y, \nabla f).
\end{aligned}
$$

Letting $X = Y = \nabla f$ and integrating over M yields

$$
\begin{aligned}
\text{(4-1)} \qquad \frac{m(n-2)}{m+n-2} \int_M B(\nabla f, \nabla f)\, d\mathrm{vol} &= \int_M \sum_i (\nabla_{E_i} D)(E_i, \nabla f, \nabla f)\, d\mathrm{vol} \\
&= -\int_M \sum_i D(E_i, \nabla f, \nabla_{E_i} \nabla f)\, d\mathrm{vol}.
\end{aligned}
$$

For the integrand, using the fact that the D-tensor is trace-free for any two indices, we have

$$
\begin{aligned}
&-\sum_i D(E_i, \nabla f, \nabla_{E_i} \nabla f) \\
&\quad = \sum_{i,j} D(E_i, \nabla f, E_j) \big(\mathrm{Ric}(E_i, E_j) - \frac{1}{m} g(E_i, \nabla f) g(E_j, \nabla f) - \lambda g(E_i, E_j)\big) \\
&\quad = \sum_{i,j,k} D(E_i, E_k, E_j)\, \mathrm{Ric}(E_i, E_j) g(E_k, \nabla f) \\
&\quad = \frac{1}{2} \sum_{i,j,k} D(E_i, E_k, E_j)\, \big(\mathrm{Ric}(E_i, E_j) g(E_k, \nabla f) - \mathrm{Ric}(E_k, E_j) g(E_i, \nabla f)\big) \\
&\quad = -\frac{1}{2} \sum_{i,j,k} |D(E_i, E_j, E_k)|^2.
\end{aligned}
$$

It follows that

$$\frac{m(n-2)}{m+n-2}\int_M B(\nabla f,\nabla f)\,d\mathrm{vol} = -\frac{1}{2}\int_M |D|^2\,d\mathrm{vol}. \tag{4-2}$$

So a vanishing Bach tensor implies that the D-tensor vanishes on M.

From (3-1), we have $C(X, Y, Z) = W(X, Y, Z, \nabla f)$. We show that both are zero on the regular points of f, and thus on M, since f is an analytic function (see [He et al. 2012, Proposition 2.8]). At a regular point of f, we choose $E_1 = \nabla f/|\nabla f|$ and let $C_{ijk} = C(E_i, E_j, E_k)$. By the symmetry of the Weyl tensor, we have $C_{ij1} = 0$. Let $a, b, c \geq 2$ be integers. From Proposition 3.3, we have $\mathrm{Ric}(E_1, E_a) = 0$ and $R(E_1, E_a, E_b, E_c) = 0$, and thus $W(E_a, E_b, E_c, E_1) = R(E_a, E_b, E_c, E_1) = 0$. So we have $C_{abc} = W(E_a, E_b, E_c, \nabla f) = 0$. It remains to show that $C_{1ij} = 0$ for any $i, j = 1, \dots, n$. Since $D = 0$, Bach flatness implies that

$$\begin{aligned} 0 = (n-2)B(E_i, E_j) &= \frac{n-3}{n-2}C_{1ij}|\nabla f| + \frac{1}{m}W(E_1, E_i, E_j, \nabla f)|\nabla f| \\ &= \frac{n-3}{n-2}C_{1ij}|\nabla f| + \frac{1}{m}C_{1ij}|\nabla f|. \end{aligned}$$

It follows that $C_{1ij} = 0$ if $m \neq -(n-2)/(n-3)$. We have $-(n-2)/(n-3) = -2$ when $n = 4$, which is excluded in the theorem. When $n \geq 5$, an extension of [Cao and Chen 2011, Proposition 5.1] shows that $C_{1ij} = 0$ for all $m \neq 0$, 1, or $2-n$. □

Proof of Theorem 1.1. From Theorem 1.2, we know that (M^4, g, f) has a harmonic Weyl tensor and $W(\nabla f, X, Y, Z) = 0$ for any $X, Y, Z \in TM$. We assume that M is not Einstein. At a regular point p of f, we assume that the Ricci tensor has distinct eigenvalues. The complement of such points can not contain an open set, as g and f are analytic in the harmonic coordinate (see [He et al. 2012, Proposition 2.8]). So it is enough to show that the metric g is locally conformally flat around p. [He et al. 2012, Theorem 7.9] says that the metric is locally a warped product over an interval, meaning that $g = dt^2 + \psi(t)^2 g_L$, where (L^3, g_L) is an Einstein metric and thus has constant curvature. A computation shows that such a metric has vanishing Weyl tensors, i.e., that it is locally conformally flat.

An alternative approach is to use the symmetries of Weyl tensors to show that they are zeros, as in the proof of [Cao and Chen 2011, Theorem 1.1]. □

Remark 4.1. In [He et al. 2012], the authors considered a warped product Einstein manifold with a nonempty boundary. Let $w = \exp(-f/m)$ in the interior of M and $w = 0$ on the boundary ∂M. Both Theorem 1.1 and Theorem 1.2 can also be extended to the case when M has a nonempty boundary. For any small $\epsilon > 0$, we define $M_\epsilon = \{x \in M : w(x) \geq \epsilon\}$, and we only have to show that $D = 0$ on M_ϵ. Then taking the limit $\epsilon \to 0$ implies that $D = 0$ on M. In fact, the boundary term of the integral (4-1) vanishes:

$$\int_{\partial M_\epsilon} D(\nu, \nabla f, \nabla f)\, d\mathrm{vol} = 0,$$

since the unit normal vector ν of ∂M_ϵ is parallel to ∇f. So the integral equation (4-2) holds on M_ϵ, and then $D = 0$ on M_ϵ.

5. Acknowledgements

The authors would like to thank Huai-Dong Cao for enlightening conversations and helpful suggestions. We also thank the anonymous referees for their valuable comments on our previous version.

References

[Bach 1921] R. Bach, "Zur Weylschen Relativitätstheorie und der Weylschen Erweiterung des Krümmungstensorbegriffs", *Math. Z.* **9**:1-2 (1921), 110–135. MR 1544454 JFM 48.1035.01

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Böhm 1998] C. Böhm, "Inhomogeneous Einstein metrics on low-dimensional spheres and other low-dimensional spaces", *Invent. Math.* **134**:1 (1998), 145–176. MR 99i:53046 Zbl 0965.53033

[Cao 2010] H.-D. Cao, "Recent progress on Ricci solitons", pp. 1–38 in *Recent advances in geometric analysis*, edited by Y.-I. Lee et al., Adv. Lect. Math. (ALM) **11**, International Press, Somerville, MA, 2010. MR 2011d:53061 Zbl 1201.53046

[Cao and Chen 2011] H.-D. Cao and Q. Chen, "On Bach-flat gradient shrinking Ricci solitons", preprint, 2011. arXiv 1105.3163v3

[Cao and Chen 2012] H.-D. Cao and Q. Chen, "On locally conformally flat gradient steady Ricci solitons", *Trans. Amer. Math. Soc.* **364**:5 (2012), 2377–2391. MR 2888210 Zbl 1245.53038

[Case et al. 2011] J. Case, Y.-J. Shu, and G. Wei, "Rigidity of quasi–Einstein metrics", *Differential Geom. Appl.* **29**:1 (2011), 93–100. MR 2012d:53124 Zbl 1215.53033

[Catino 2012] G. Catino, "A note on four-dimensional (anti-)self-dual quasi–Einstein manifolds", *Differential Geom. Appl.* **30**:6 (2012), 660–664. MR 2996860 Zbl 06126062

[Catino et al. 2011] G. Catino, C. Mantegazza, L. Mazzieri, and M. Rimoldi, "Locally conformally flat quasi–Einstein manifolds", preprint, 2011. arXiv 1010.1418v3

[He et al. 2010] C. He, P. Petersen, and W. Wylie, "Warped product Einstein metrics over spaces with constant scalar curvature", preprint, 2010. arXiv 1012.3446v1

[He et al. 2011] C. He, P. Petersen, and W. Wylie, "Uniqueness of warped product Einstein metrics and applications", preprint, 2011. arXiv 1110.2456

[He et al. 2012] C. He, P. Petersen, and W. Wylie, "On the classification of warped product Einstein metrics", *Comm. Anal. Geom.* **20**:2 (2012), 271–311. MR 2928714

[Kim and Kim 2003] D.-S. Kim and Y. H. Kim, "Compact Einstein warped product spaces with nonpositive scalar curvature", *Proc. Amer. Math. Soc.* **131**:8 (2003), 2573–2576. MR 2004b:53063 Zbl 1029.53027

[Qian 1997] Z. Qian, "Estimates for weighted volumes and applications", *Quart. J. Math. Oxford Ser.* (2) **48**:190 (1997), 235–242. MR 98e:53058 Zbl 0902.53032

[Villani 2009] C. Villani, *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften **338**, Springer, Berlin, 2009. MR 2010f:49001 Zbl 1156.53003

[Wei and Wylie 2009] G. Wei and W. Wylie, "Comparison geometry for the Bakry–Emery Ricci tensor", *J. Differential Geom.* **83**:2 (2009), 377–405. MR 2011a:53064 Zbl 1189.53036

Received July 5, 2012. Revised December 3, 2012.

QIANG CHEN
DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PA 18015
UNITED STATES
qic208@lehigh.edu

CHENXU HE
DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PA 18015
UNITED STATES
Current address:
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73019
UNITED STATES
he.chenxu@ou.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.327

ON PLANE SEXTICS WITH DOUBLE SINGULAR POINTS

ALEX DEGTYAREV

We compute the fundamental groups of five maximizing sextics with double singular points only; in four cases, the groups are as expected. The approach used would apply to other sextics as well, given their equations.

1. Introduction

The fundamental group $\pi_1 := \pi_1(\mathbb{P}^2 \smallsetminus D)$ of a plane curve $D \subset \mathbb{P}^2$, introduced by O. Zariski [1929], is an important topological invariant of the curve. Apart from distinguishing the connected components of the equisingular moduli spaces, this group can be used as a seemingly inexpensive way of studying algebraic surfaces, the curve serving as the branch locus of a projection of the surface onto $\mathbb{P}^2$.

At present, the fundamental groups of all curves of degree up to five are known, and the computation of the groups of irreducible curves of degree six (sextics) is close to its completion; see [Degtyarev 2012] for the principal statements and further references. In higher degrees, little is known: there are a few general theorems, usually bounding the complexity of the group of a curve with sufficiently "moderate" singularities, and a number of sporadic examples scattered in the literature. For further details on this fascinating subject, we refer the reader to the recent surveys [Artal-Bartolo et al. 2008; Libgober 2007a; 2007b].

1A. ***Principal results.*** If a sextic $D \subset \mathbb{P}^2$ has a singular point P of multiplicity three or higher, then, projecting from this point, we obtain a trigonal (or, even better, bi- or monogonal) curve in a Hirzebruch surface; see Section 3A. By means of the so-called *dessins d'enfants*, such curves and their topology can be studied in purely combinatorial terms, as certain graphs in the plane. The classification of such curves and the computation of their fundamental groups were completed in [Degtyarev 2012]. If all singular points are double, the best that one can obtain is a tetragonal curve, which is a much more complicated object. (A reduction of tetragonal curves to trigonal curves in the presence of a section is discussed in Section 3B; see Remark 3.6. It is the extra section that makes the problem difficult.) At present, I do not know how the group of a tetragonal curve can be computed

MSC2000: primary 14H45; secondary 14H30, 14H50.
Keywords: plane sextic, torus type, fundamental group, tetragonal curve.

unless the curve is real and its defining equation is known (and, even then, the approach suggested in the paper may still fail; cf. Remark 2.1).

There is a special class of irreducible sextics, the so-called $\mathbb{D}_{2n}$-*sextics* and, in particular, sextics of *torus type* (see Section 2A for the precise definitions), for which the fundamental group is nonabelian for some simple homological reasons; see [Degtyarev 2008]. (The fact that a sextic is of torus type is usually indicated by the presence of a pair of parentheses in the notation; their precise meaning is explained in Section 2A.) On the other hand, thanks to the special structures and symmetries of these curves, their explicit equations are known; see [Degtyarev 2009b; Degtyarev and Oka 2009; Oka and Pho 2002]. In this paper, we almost complete the computation of the fundamental groups of $\mathbb{D}_{2n}$-sextics (with one pair of complex conjugate sextics of torus type left). Our principal results can be stated as follows.

Theorem 1.1. *The fundamental group of the $\mathbb{D}_{14}$-special sextic with the set of singularities $3\boldsymbol{A}_6 \oplus \boldsymbol{A}_1$, line 37 in Table 1, is $\mathbb{Z}_3 \times \mathbb{D}_{14}$.*

Theorem 1.2. *The fundamental groups of the irreducible sextics of torus type with the sets of singularities $(\boldsymbol{A}_{14} \oplus \boldsymbol{A}_2) \oplus \boldsymbol{A}_3$, line 8, $(\boldsymbol{A}_{14} \oplus \boldsymbol{A}_2) \oplus \boldsymbol{A}_2 \oplus \boldsymbol{A}_1$, line 9, and $(\boldsymbol{A}_{11} \oplus 2\boldsymbol{A}_2) \oplus \boldsymbol{A}_4$, line 17 in Table 1, are isomorphic to $\Gamma := \mathbb{Z}_2 * \mathbb{Z}_3$. The group of the curve with the set of singularities $(\boldsymbol{A}_8 \oplus 3\boldsymbol{A}_2) \oplus \boldsymbol{A}_4 \oplus \boldsymbol{A}_1$, line 33, is*

$$\begin{aligned}(1.3)\quad \pi_1 = \bigl\langle \alpha_2, \alpha_3, \alpha_4 \bigm| [\alpha_3, \alpha_4] = \{\alpha_2, \alpha_3\}_3 = \{\alpha_2, \alpha_4\}_9 = 1, \\ \alpha_4\alpha_2\alpha_3^{-1}\alpha_4\alpha_2\alpha_4(\alpha_4\alpha_2)^{-2}\alpha_3 = (\alpha_2\alpha_4)^2\alpha_3^{-1}\alpha_2\alpha_4\alpha_3\alpha_2\bigr\rangle,\end{aligned}$$

where $\{\alpha, \beta\}_{2k+1} := (\alpha\beta)^k\alpha(\alpha\beta)^{-k}\beta^{-1}$.

Theorem 1.1 is proved in Section 4C, and Theorem 1.2 is proved in Sections 4E–4H, one curve at a time. I do not know whether the last group (1.3) is isomorphic to Γ: all "computable" invariants seem to coincide (see Remark 4.7), but the presentations obtained resist all simplification attempt. The quotient of (1.3) by the extra relation $\{\alpha_2, \alpha_4\}_3 = 1$ is Γ.

The next proposition is proved in Section 4I. (The perturbation $3\boldsymbol{A}_6 \oplus \boldsymbol{A}_1 \to 3\boldsymbol{A}_6$ excluded in the statement results in a $\mathbb{D}_{14}$-special sextic and the fundamental group equals $\mathbb{Z}_3 \times \mathbb{D}_{14}$; see [Degtyarev and Oka 2009].)

Proposition 1.4. *Let D' be a nontrivial perturbation of a sextic as in Theorem 1.1 or 1.2. Unless the set of singularities of D' is $3\boldsymbol{A}_6$, the group $\pi_1(\mathbb{P}^2 \smallsetminus D')$ is Γ or $\mathbb{Z}_6$, depending on whether D' is, or, respectively, is not, of torus type.*

With Theorem 1.1 in mind, the fundamental groups of all $\mathbb{D}_{2n}$-special sextics, $n \geq 5$, are known; see [Degtyarev 2012]. Modulo the feasible conjecture that any sextic of torus type degenerates to a maximizing one, the only such sextic whose group remains unknown is $(\boldsymbol{A}_8 \oplus \boldsymbol{A}_5 \oplus \boldsymbol{A}_2) \oplus \boldsymbol{A}_4$, line 32 in Table 1. (This conjecture has been proved, and *all* groups except the one just mentioned are indeed known; details will appear elsewhere.) Most of these groups are isomorphic to Γ; see [Degtyarev 2012] for details and further references.

I would like to mention an alternative approach (see [Artal Bartolo et al. 2002]) reducing a plane sextic with large Milnor number to a trigonal curve equipped with a number of sections, all but one splitting in the covering elliptic surface. It was used in [Artal Bartolo et al. 2002] to handle the curves in lines 1–6 in Table 1. This approach is also used in a forthcoming paper to produce the defining equations of most sextics listed in Table 1; then, the fundamental groups of most *real* ones can be computed using Theorem 3.16. All groups that could be found are abelian. Together with the classification of sextics, which is also almost completed, this fact implies that, with very few exceptions, the fundamental group of a nonspecial irreducible simple sextic is abelian.

1B. ***Idea of the proof*** (***see Section 4A for more details***)**.** We use the classical Zariski–van Kampen method (see Theorem 3.16), expressing the fundamental group of a curve in terms of its braid monodromy with respect to an appropriate pencil of lines. The curves and pencils considered are *real*, and the braid monodromy in a neighborhood of the real part of the pencil is computed in terms of the real part of the curve. (This approach originates in topology of real algebraic curves; historically, it goes back to Viro, Fiedler, Kharlamov, Rokhlin, and Klein.) Our main contribution is the description of the monodromy along a real segment where all four branches of the curve are nonreal; see Proposition 3.12. Besides, the curves are not required to be *strongly real*; i.e., nonreal singular fibers are allowed. Hence, we follow [Orevkov 1999] and attempt to extract information about such nonreal fibers from the real part of the curve. The outcome is Theorem 3.16, which gives us an "upper bound" on the fundamental group in question. The applicability issues and a few other common tricks are discussed in Section 4A.

1C. ***Contents of the paper.*** In Section 2, we introduce the terminology related to plane sextics, list the sextics that are still to be investigated, and discuss briefly the few known results. In Section 3, we outline an approach to the (partial) computation of the braid monodromy of a real tetragonal curve and state an appropriate version of the Zariski–van Kampen theorem. Finally, in Section 4 the results of Section 3 and known equations are used to prove Theorems 1.1 and 1.2 and Proposition 1.4.

1D. ***Conventions.*** All group actions are *right*. Given a right action $X \times G \to X$ and a pair of elements $x \in X$, $g \in G$, the image of (x, g) is denoted by $x \uparrow g \in X$. The same postfix notation and multiplication convention is often used for maps: it is under this convention that the monodromy $\pi_1(\text{base}) \to \operatorname{Aut}(\text{fiber})$ of a locally trivial fibration is a homomorphism rather than an antihomomorphism.

The assignment symbol := is used as a shortcut for "is defined as".

We use the conventional symbol $\square$ to mark the ends of the proofs. Some statements are marked with $\triangleleft$, meaning that the proof has already been explained (for example, most corollaries).

2. Preliminaries

2A. ***Special classes of sextics.*** A plane sextic $D \in \mathbb{P}^2$ is called *simple* if all its singularities are simple, i.e., those of type $\boldsymbol{A}$–$\boldsymbol{D}$–$\boldsymbol{E}$. The total Milnor number μ of a simple sextic D does not exceed 19; see [Persson 1985]; if $\mu = 19$, then D is called *maximizing*. Maximizing sextics are always defined over algebraic number fields and their moduli spaces are discrete: two such sextics are equisingular deformation equivalent if and only if they are related by a projective transformation of $\mathbb{P}^2$.

A sextic D is said to be of *torus type* if its equation can be represented in the form $f_2^3 + f_3^2 = 0$, where f_2 and f_3 are some polynomials of degree 2 and 3, respectively. The points of intersection of the conic $\{f_2 = 0\}$ and cubic $\{f_3 = 0\}$ are always singular for D. These singular points play a very special rôle; they are called the *inner* singularities (with respect to the given torus structure). For the vast majority of curves, a torus structure is unique, and in this case it is common to parenthesize the inner singularities in the notation.

An irreducible sextic D is called $\mathbb{D}_{2n}$-*special* if its fundamental group $\pi_1(\mathbb{P}^2 \smallsetminus D)$ admits a dihedral quotient $\mathbb{D}_{2n} := \mathbb{Z}_n \rtimes \mathbb{Z}_2$. According to [Degtyarev 2008], only $\mathbb{D}_6$-, $\mathbb{D}_{10}$-, and $\mathbb{D}_{14}$-special sextics exist, and an irreducible sextic is of torus type if and only if it is $\mathbb{D}_6$-special. (In particular, torus type is a topological property.)

Any sextic D of torus type is a degeneration of Zariski's six-cuspidal sextic, which is obtained from a generic pair (f_2, f_3). It follows that the fundamental group of D factors to the modular group $\Gamma := SL(2, \mathbb{Z}) = \mathbb{Z}_2 * \mathbb{Z}_3 = \mathbb{B}_3/(\sigma_1\sigma_2\sigma_1)^2$; see [Zariski 1929]; in particular, this group is infinite. Conjecturally, the fundamental groups of all other irreducible simple sextics are finite.

2B. ***Sextics to be considered.*** It is expected that, with few explicit exceptions (e.g., $9\boldsymbol{A}_2$), any simple sextic degenerates to a maximizing one. (The proof of this conjecture, which relies upon the theory of $K3$-surfaces, is currently a work in progress. In fact, most curves degenerate to one of those whose groups are already known.) Hence, it is essential to compute the fundamental groups of the maximizing sextics; the others would follow. The groups of all irreducible sextics with a singular point of multiplicity three or higher are known (see [Degtyarev 2012] for a summary of the results), and those with $\boldsymbol{A}$ type singularities only are still to be investigated.

A list of irreducible maximizing sextics with $\boldsymbol{A}$ type singular points only can be compiled using the results of [Yang 1996] (a list of the sets of singularities realized

#	Singularities	(r,c)	Equation, π_1, remarks
1.	A_{19}	$(2,0)$	$\pi_1=\mathbb{Z}_6$, see [Artal Bartolo et al. 2002]
2.	$A_{18}\oplus A_1$	$(1,1)$	$\pi_1=\mathbb{Z}_6$, see [ibid.]
3.	$(A_{17}\oplus A_2)$	$(1,0)^*$	$\pi_1=\Gamma$, see [ibid.] and [Degtyarev 2009a] (torus type)
4.	$A_{16}\oplus A_3$	$(2,0)$	$\pi_1=\mathbb{Z}_6$, see [Artal Bartolo et al. 2002]
5.	$A_{16}\oplus A_2\oplus A_1$	$(1,1)$	$\pi_1=\mathbb{Z}_6$, see [ibid.]
6.	$A_{15}\oplus A_4$	$(0,1)^*$	$\pi_1=\mathbb{Z}_6$, see [ibid.]
7.	$A_{14}\oplus A_4\oplus A_1$	$(0,3)$	
8.	$(A_{14}\oplus A_2)\oplus A_3$	$(1,0)$	$\pi_1=\Gamma$, see Section 4E (torus type)
9.	$(A_{14}\oplus A_2)\oplus A_2\oplus A_1$	$(1,0)$	$\pi_1=\Gamma$, see Section 4F (torus type)
10.	$A_{13}\oplus A_6$	$(0,2)$	
11.	$A_{13}\oplus A_4\oplus A_2$	$(2,0)$	
12.	$A_{12}\oplus A_7$	$(0,1)$	
13.	$A_{12}\oplus A_6\oplus A_1$	$(1,1)$	
14.	$A_{12}\oplus A_4\oplus A_3$	$(1,0)$	
15.	$A_{12}\oplus A_4\oplus A_2\oplus A_1$	$(1,1)$	
16.	$A_{11}\oplus 2A_4$	$(2,0)$	
17.	$(A_{11}\oplus 2A_2)\oplus A_4$	$(1,0)$	$\pi_1=\Gamma$, see Section 4G (torus type)
18.	$A_{10}\oplus A_9$	$(2,0)^*$	
19.	$A_{10}\oplus A_8\oplus A_1$	$(1,1)$	
20.	$A_{10}\oplus A_7\oplus A_2$	$(2,0)$	
21.	$A_{10}\oplus A_6\oplus A_3$	$(0,1)$	
22.	$A_{10}\oplus A_6\oplus A_2\oplus A_1$	$(1,1)$	
23.	$A_{10}\oplus A_5\oplus A_4$	$(2,0)$	
24.	$A_{10}\oplus 2A_4\oplus A_1$	$(1,1)$	
25.	$A_{10}\oplus A_4\oplus A_3\oplus A_2$	$(1,0)$	
26.	$A_{10}\oplus A_4\oplus 2A_2\oplus A_1$	$(2,0)$	
27.	$A_9\oplus A_6\oplus A_4$	$(1,1)^*$	
28.	$A_9\oplus 2A_4\oplus A_2$	$(1,0)^*$	$\pi_1=$(2.2), see [Degtyarev 2009b] ($\mathbb{D}_{10}$-sextic)
29.	$(2A_8)\oplus A_3$	$(1,0)$	$\pi_1=\Gamma$, see [Degtyarev 2009a] (torus type)
30.	$A_8\oplus A_7\oplus A_4$	$(0,1)$	
31.	$A_8\oplus A_6\oplus A_4\oplus A_1$	$(1,1)$	
32.	$(A_8\oplus A_5\oplus A_2)\oplus A_4$	$(0,1)$	`nt104` in [Oka and Pho 2002] (torus type)
33.	$(A_8\oplus 3A_2)\oplus A_4\oplus A_1$	$(1,0)$	$\pi_1=$(1.3), see Section 4H (torus type)
34.	$A_7\oplus 2A_6$	$(0,1)$	
35.	$A_7\oplus A_6\oplus A_4\oplus A_2$	$(2,0)$	
36.	$A_7\oplus 2A_4\oplus 2A_2$	$(1,0)$	
37.	$3A_6\oplus A_1$	$(1,0)$	$\pi_1=\mathbb{Z}_3\times\mathbb{D}_{14}$, see Section 4C ($\mathbb{D}_{14}$-sextic)
38.	$2A_6\oplus A_4\oplus A_2\oplus A_1$	$(2,0)$	
39.	$A_6\oplus A_5\oplus 2A_4$	$(2,0)$	

An * marks sets of singularities that are realized by reducible sextics as well.
There are 42 real and 20 pairs of complex conjugate curves.

Table 1. Irreducible maximizing sextics with $\boldsymbol{A}$ type singularities.

by such sextics) and [Shimada 2007] (a description of the moduli spaces). We represent the result in Table 1, where the column (r, c) shows the number of classes: r is the number of real sextics, and c is the number of pairs of complex conjugate ones. The approach developed further in the paper lets one compute (or at least estimate) the fundamental group of a sextic with $\boldsymbol{A}$ type singularities, provided that its equation is known. In the literature, I could find explicit equations for lines 1–6, 8, 9, 17, 28, 29, 32, 33, and 37. With the results of this paper (Theorems 1.1 and 1.2) taken into account, the groups of all these sextics except $(\boldsymbol{A}_8 \oplus \boldsymbol{A}_5 \oplus \boldsymbol{A}_2) \oplus \boldsymbol{A}_4$, line 32 (which is not real), are known.

Remark 2.1. Unfortunately, our approach does not always work even if the curve is real. Thus, each of the two sextics with the set of singularities $\boldsymbol{A}_{19}$, line 1, has a single real point (the isolated singular point of type $\boldsymbol{A}_{19}$; see [Artal Bartolo et al. 2002] for the equations) and Theorem 3.16 does not provide enough relations to compute the group.

2C. ***Known results.*** The fundamental group of the $\mathbb{D}_{10}$-special sextic with the set of singularities $\boldsymbol{A}_9 \oplus 2\boldsymbol{A}_4 \oplus \boldsymbol{A}_2$, line 28 in Table 1, can be described as follows; see [Degtyarev 2009b] (where $'$ temporarily stands for the commutant of a group):

$$\pi_1/\pi_1'' = \mathbb{Z}_3 \times \mathbb{D}_{10}, \quad \pi_1'' = SL(2, \Bbbk_9), \tag{2.2}$$

where $\Bbbk_9$ is the field of nine elements. The fundamental groups of the first twelve sextics, lines 1–6, have been found in [Artal Bartolo et al. 2002]: with the exception of $(\boldsymbol{A}_{17} \oplus \boldsymbol{A}_2)$, line 3 (sextic of torus type, $\pi_1 = \Gamma$), they are all abelian. To my knowledge, the groups not mentioned in Table 1 have not been computed yet.

3. The braid monodromy

3A. ***Hirzebruch surfaces.*** A *Hirzebruch surface* Σ_d, $d > 0$, is a geometrically ruled rational surface with a (unique) *exceptional section* E of self-intersection $-d$. Typically, we use affine coordinates (x, y) in Σ_d such that E is given by $y = \infty$; then, x can be regarded as an affine coordinate in the base of the ruling. (The line $\{x = \infty\}$ plays no special rôle; usually, it is assumed sufficiently generic.) The fiber of the ruling over a point x in the base is denoted by F_x, and the *affine fiber* over x is $F_x^\circ := F_x \smallsetminus E$. This is an affine space over $\mathbb{C}$; in particular, one can speak about convex hulls in F_x°.

An *n-gonal curve* is a reduced curve $C \subset \Sigma_d$ intersecting each fiber at n points, i.e., such that the restriction to C of the ruling $\Sigma_d \to \mathbb{P}^1$ is a map of degree n. A *singular fiber* of an n-gonal curve C is a fiber F of the ruling intersecting $C + E$ geometrically at fewer than $(n + 1)$ points. A singular fiber F is *proper* if C does not pass through $F \cap E$. The curve C is *proper* if so are all its singular fibers. In other words, C is proper if it is disjoint from E.

In affine coordinates (x, y) as above an n-gonal curve $C \subset \Sigma_d$ is given by a polynomial of the form $\sum_{i=0}^{n} a_i(x)y^i$, where $\deg a_i \leqslant m + d(n-i)$ for some $m \geq 0$ (in fact, $m = C \cdot E$) and at least one polynomial a_i does have the prescribed degree (so that C does not contain the fiber $\{x = \infty\}$). The curve is proper if and only if $m = 0$; in this case $a_n(x) = \text{const}$.

A proper n-gonal curve $C \subset \Sigma_d$ defines a distinguished *zero section* $Z \subset \Sigma_d$, sending each point $x \in \mathbb{P}^1$ to the barycenter of the n points of $F_x^\circ \cap C$. Certainly, this section does not need to coincide with $\{y = 0\}$, which depends on the choice of the coordinates.

3B. ***The cubic resolvent.*** Consider a reduced *real* quartic polynomial

$$f(x, y) := y^4 + p(x)y^2 + q(x)y + r(x), \tag{3.1}$$

so that its roots y_1, y_2, y_3, y_4 (at each point x) satisfy $y_1 + y_2 + y_3 + y_4 = 0$, and consider the (*modified*) *cubic resolvent* of f

$$y^3 - 2p(x)y^2 + b_1(x)y + q(x)^2, \quad b_1 := p^2 - 4r, \tag{3.2}$$

and its reduced form

$$\bar{y}^3 + g_2(x)\bar{y} + g_3(x) \tag{3.3}$$

obtained by the substitution $y = \bar{y} + \frac{2}{3}p$. The discriminants of (3.1)–(3.3) are equal:

$$D = 16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r - 27q^4 + 256r^3. \tag{3.4}$$

Recall that $D = 0$ if and only if (3.1), or, equivalently, (3.2) or (3.3), has a multiple root. Otherwise, $D < 0$ if and only if exactly two roots of (3.1) are real. The roots of (3.2) are

$$\begin{aligned} \alpha &:= (y_1 + y_2)(y_3 + y_4) = -(y_1 + y_2)^2, \\ \beta &:= (y_1 + y_3)(y_2 + y_4) = -(y_1 + y_3)^2, \\ \gamma &:= (y_1 + y_4)(y_2 + y_3) = -(y_1 + y_4)^2, \end{aligned} \tag{3.5}$$

and those of (3.3) are obtained from (3.5) by shifting the barycenter $\frac{1}{3}(\alpha + \beta + \gamma)$ to zero.

Remark 3.6. If $\{f(x, y) = 0\}$ is a proper tetragonal curve in a Hirzebruch surface Σ_d, then (3.2) defines a proper trigonal curve $C' \subset \Sigma_{2d}$ and a distinguished section $S := \{y = 0\}$ (in general, *other* than the zero section) which is tangent (more precisely, has even intersection index at each intersection point) to C'. Conversely, (3.1) can be recovered from (3.2) (together with the section $S = \{y = 0\}$) uniquely up to the automorphism $y \mapsto -y$, which takes q to $-q$.

Remark 3.7. One has

$$q = -(y_1 + y_2)(y_1 + y_3)(y_1 + y_4);$$

hence, q vanishes if and only if two of the roots of (3.1) are opposite. If all roots are nonreal, $y_{1,2} = \alpha \pm \beta' i$, $y_{3,4} = -\alpha \pm \beta'' i$, α, β', $\beta'' \in \mathbb{R}$, then

$$q = 2\alpha\bigl(\beta'^2 - \beta''^2\bigr), \quad b_1 = -8\alpha^2\bigl(\beta'^2 + \beta''^2\bigr) + \bigl(\beta'^2 - \beta''^2\bigr)^2.$$

Hence, $q(x) = 0$ if and only if either $\alpha = 0$ (and then $b_1(x) > 0$, assuming that $D(x) \ne 0$) or $\beta' = \pm\beta''$ (and then $b_1(x) < 0$). If $y_1 = y_2$, i.e., $\beta' = 0$, then $q(x) > 0$ if and only if one has the inequality $y_1 < \operatorname{Re} y_3 = \operatorname{Re} y_4$ equivalent to $y_1 < 0$.

Remark 3.8. Observe also that, if $y_1 = y_2$, then g_3 takes the form

$$g_3 = \tfrac{2}{27}(y_1 - y_4)^3(y_1 - y_3)^3.$$

Hence, $g_3(x) < 0$ if and only if the two other roots are real and separated by the double root $y_1 = y_2$. Otherwise, either $y_1 < \operatorname{Re} y_3, \operatorname{Re} y_4$ or $y_1 > \operatorname{Re} y_3, \operatorname{Re} y_4$, and, in view of Remark 3.7, the former holds if and only if $q(x) > 0$.

3C. *The real monodromy.* Choose affine coordinates (x, y) in the Hirzebruch surface Σ_d so that the exceptional section E is $\{y = \infty\}$. Consider a *real* proper tetragonal curve $C \subset \Sigma_d$; it is given by a real polynomial $f(x, y)$ as in (3.1). Over a generic real point $x \in \mathbb{R}$, the four points $y_1, \dots, y_4$ of the intersection $C \cap F_x^\circ$ can be ordered lexicographically, according to the *decreasing* of $\operatorname{Re} y$ first and $\operatorname{Im} y$ second. We always assume this ordering. Then, choosing a real reference point $y \gg 0$, we have a *canonical* geometric basis $\{\alpha_1, \dots, \alpha_4\}$ for the fundamental group $\pi(x) := \pi_1(F_x^\circ \smallsetminus C, y)$; see Figure 1.

Let $x_1, \dots, x_r$ be all *real* singular fibers of C, ordered by increasing. For each i, consider a pair of nonsingular fibers $x_i^- := x_i - \epsilon$ and $x_i^+ := x_i + \epsilon$, where ϵ is a sufficiently small positive real number; see Figure 2. Define $x_0 = x_{r+1} = \infty$ and,

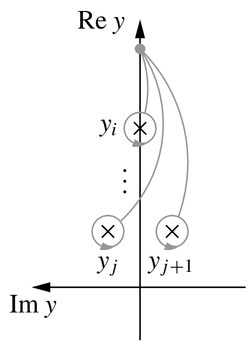


Figure 1. The canonical basis.

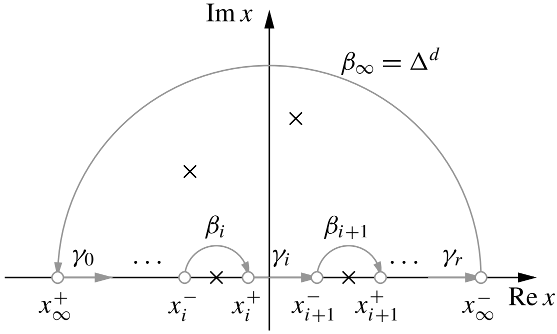


Figure 2. The monodromies β_i and γ_j.

assuming the fiber $x = \infty$ is nonsingular, pick also a pair of real nonsingular fibers $x_{r+1}^- = x_\infty^- := R \gg 0$ and $x_0^+ = x_\infty^+ := -R$. Identify all groups $\pi(x_i^\pm)$ with the free group $\mathbb{F}_4$ by means of their respective canonical bases. (All reference points are chosen in a real section $y = \text{const} \gg 0$, which is assumed disjoint from the fiberwise convex hull of C over the disk $|x| \leqslant R$.) Consider the semicircles $t \mapsto x_i + \epsilon e^{i\pi(1-t)}$, $t \in [0, 1]$, and the line segments $t \mapsto t$, $t \in [x_j^+, x_{j+1}^-]$; see Figure 2. These paths give rise to the *monodromy isomorphisms*

$$\beta_i : \pi(x_i^-) \to \pi(x_i^+), \quad \gamma_j : \pi(x_j^+) \to \pi(x_{j+1}^-),$$

$i = 1, \ldots, r$, $j = 0, \ldots, r$. In addition, we also have the monodromy $\beta_0 = \beta_\infty = \beta_{r+1} : \pi(x_\infty^-) \to \pi(x_\infty^+)$ along the semicircle $t \mapsto Re^{i\pi t}$, $t \in [0, 1]$, and the *local monodromies*

$$\mu_i : \pi(x_i^+) \to \pi(x_i^+), \quad i = 1, \ldots, r$$

along the circles $t \mapsto x_i + \epsilon e^{2\pi i t}$, $t \in [0, 1]$. Using the identifications $\pi(x_i^\pm) = \mathbb{F}_4$ fixed above, all β_i, μ_i, γ_j can be regarded as elements of the automorphism group $\operatorname{Aut}\mathbb{F}_4$, and as such they belong to the braid group $\mathbb{B}_4$. Recall (see [Artin 1947]) that Artin's *braid group* $\mathbb{B}_4 \subset \operatorname{Aut}\langle\alpha_1, \ldots, \alpha_4\rangle$ is the subgroup consisting of the automorphisms taking each generator α_i to a conjugate of a generator and preserving the product $\alpha_1\alpha_2\alpha_3\alpha_4$. It is generated by the three braids

$$\sigma_i : \alpha_i \mapsto \alpha_i\alpha_{i+1}\alpha_i^{-1}, \quad \alpha_{i+1} \mapsto \alpha_i, \quad i = 1, 2, 3,$$

the defining relations being $\{\sigma_1, \sigma_2\}_3 = \{\sigma_2, \sigma_3\}_3 = [\sigma_1, \sigma_3] = 1$.

3D. ***The computation.*** The braids β_i, μ_i, and γ_j introduced in the previous section are easily computed from the real part $C_\mathbb{R} \subset \mathbb{R}^2$ of the curve. In the figures, we use the following notation:

- real branches of C are represented by solid bold lines;

- pairs y_i, y_{i+1} of complex conjugate branches are represented by dotted lines (showing the common real part $\operatorname{Re} y_i = \operatorname{Re} y_{i+1}$);
- relevant fibers of Σ_d are represented by vertical dotted grey lines.

Certainly, the dotted lines are not readily seen in the figures; however, in most cases, it is only the intersection indices that matter, and the latter are determined by the indexing of the branches at the starting and ending positions.

We summarize the results in the next three statements. The first one is obvious: essentially, one speaks about the link of the singularity $y^4 - x^{4d}$.

Lemma 3.9. *Assume that $R \gg 0$ is so large that the disk $\{|x| < R\}$ contains all singular fibers of C. Then one has $\beta_\infty = \Delta^d$, where $\Delta := \sigma_1\sigma_2\sigma_3\sigma_1\sigma_2\sigma_1 \in \mathbb{B}_4$ is the* Garside element. ◁

The following lemma is easily proved by considering the local normal forms of the singularities. (In the simplest case of a vertical tangent, the circumventing braids β are computed, e.g., in [Orevkov 1999]; the general case is completely similar.) For the statement, we extend the standard notation $\boldsymbol{A}_m$, $m \geq 1$, to $\boldsymbol{A}_0$ to designate a simple tangency of C and the fiber.

Lemma 3.10. *The braids β_j and μ_j about a singular fiber x_j of type $\boldsymbol{A}_m$, $m \geq 0$, depend only on m and the pair $(i, i+1)$ of indices of the branches that merge at the singular point. They are as shown in Figure 3.* ◁

Remark 3.11. At a point of type $\boldsymbol{A}_{2k-1}$, it is *not* important whether the two branches of C at this point are real or complex conjugate. On the other hand, at a point of type $\boldsymbol{A}_{2k}$ it *does* matter whether the number of real branches increases or decreases. If a fiber contains two double points, with indices $(1, 2)$ and $(3, 4)$, then the powers of σ_1 and σ_3 contributed to β or μ by each of the points are multiplied; since σ_1 and σ_3 commute, the order is not important.

The following statement is our principal technical tool, most important being Figure 4, right, describing the behavior of the "invisible" branches. (Note that the

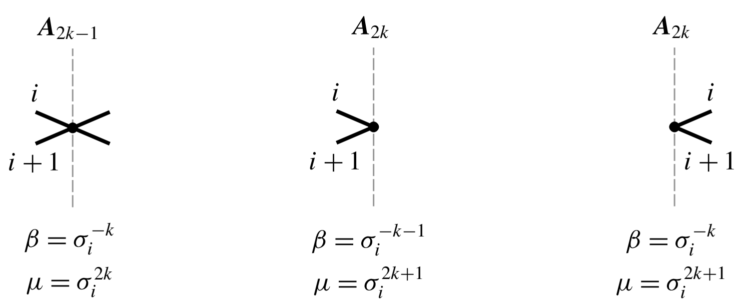


Figure 3. The braids β and μ.

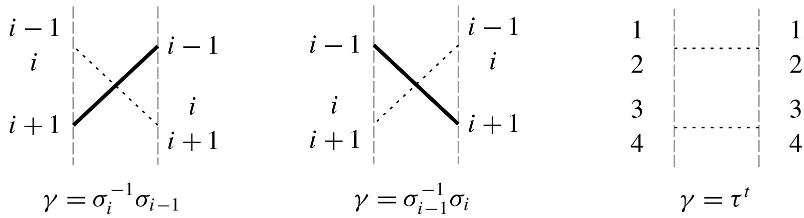


Figure 4. The braids γ.

two dotted lines in the figure may cross; the permutation of the branches depends on the parity of the twist parameter t introduced in the statement.)

Proposition 3.12. *Let I be a real segment in the x-axis free of singular fibers of C. Then the monodromy γ over I is*

- *identity, if all four branches of C over I are real, and*
- *as shown in Figure 4 otherwise.*

Here, $\tau := \sigma_2^{-1}\sigma_3\sigma_1^{-1}\sigma_2$ and the twist parameter *t in Figure 4, right, is the number of roots $x' \in I$ of the coefficient $q(x)$ (see (3.1)) such that $b_1(x') > 0$ (see (3.2)) and q changes sign at x'; each root x' contributes $+1$ or -1 depending on whether q is increasing or decreasing at x', respectively.*

Proof. The only case that needs consideration, viz. that of four nonreal branches (see Figure 4, right), is given by Remark 3.7. Indeed, the canonical basis in the fiber F_x° over $x \in I$ changes when the real parts of all four branches vanish, and this happens when $q(x) = 0$ and $b_1(x) > 0$. This change contributes $\tau^{\pm 1}$ to γ, and the sign ± 1 (the direction of rotation) depends on whether q increases or decreases. $\square$

Remark 3.13. A longer segment I with exactly two real branches of C over it can be divided into smaller pieces $I_1, I_2, \ldots$, each containing a single crossing point as in Figure 4; then, the monodromy γ over I is the product of the contributions of each piece. In fact, as explained above, the precise position and number of crossings is irrelevant; what only matters is the final permutation between the endpoints of I. For example, to minimize the number of elementary pieces, one can always assume that the branches, both bold and dotted, are monotonous.

3E. ***The Zariski–van Kampen theorem.*** We are interested in the fundamental group $\pi_1 := \pi_1(\Sigma_{\tilde{d}} \smallsetminus (\tilde{C} \cup E))$, where $\tilde{C} \subset \Sigma_{\tilde{d}}$ is a real tetragonal curve, possibly improper, and $E \subset \Sigma_{\tilde{d}}$ is the exceptional section. To compute π_1, we consider the *proper model* $C \subset \Sigma_d$, obtained from $\tilde{C}$ by blowing up all points of intersection $\tilde{C} \cap E$ and blowing down the corresponding fibers. In addition to the braids β_i, μ_i, and γ_j introduced in Section 3C, to each real singular fiber x_i of C we assign its *local slope* $\varkappa_i \in \pi(x_i^+)$, which depends on the type of the corresponding singular fiber of the original curve $\tilde{C}$. Roughly, consider a small analytic disk $\Phi \subset \Sigma_d$ transversal to the fiber F_{x_i} and disjoint from C and E, and a similar disk $\tilde{\Phi} \subset \Sigma_{\tilde{d}}$ with respect to $\tilde{C}$. Let $\tilde{\Phi}' \subset \Sigma_d$ be the image of $\tilde{\Phi}$, and assume that the boundaries $\partial\Phi$ and $\partial\tilde{\Phi}'$ have a common point in the fiber over x_i^+. Then the loop $[\partial\tilde{\Phi}'] \cdot [\partial\Phi]^{-1}$ is homotopic to a certain class $\varkappa_i \in \pi(x_i^+)$, well defined up to a few moves irrelevant in the sequel. This class is the slope.

Roughly, the slope measures (in the form of the twisted monodromy; see the definitions prior to Theorem 3.16) the deviation of the braid monodromy of an improper curve $\tilde{C}$ from that of its proper model C. Slopes appear in the relation at infinity as well, compensating for the fact that, near improper singular fibers, the curve intersects *any* section of $\Sigma_{\tilde{d}}$. Details and further properties are found in [Degtyarev 2012, Section 5.1.3]; in this paper, slopes are used in Theorem 3.16.

Remark 3.14. In all examples considered below, $\tilde{C} \subset \Sigma_{d-1}$ has a single improper fiber F, where $\tilde{C}$ has a singular point of type $\tilde{\boldsymbol{A}}_m$, $m \geq 1$, maximally transversal to both E and F. If $F = \{x = 0\}$, such a curve $\tilde{C}$ is given by a polynomial $\tilde{f}$ of the form $\sum_{i=0}^4 y^i a_i(x)$ with $a_4(x) = x^2$ and $x \mid a_3(x)$, and the defining polynomial of its transform $C \subset \Sigma_d$ is $f_{\mathrm{nr}}(x, y) := x^2 \tilde{f}(x, y/x)$. The corresponding singular fiber of C has a node $\boldsymbol{A}_1$ at $(0, 0)$ and another double point $\boldsymbol{A}_{m-2}$ (assuming $m \geq 2$).

Thus, the only nontrivial example relevant in the sequel is the one described below. (By the very definition, at each singular fiber x_i proper for $\tilde{C}$ the slope is $\varkappa_i = 1$.) A great deal of other examples of both computing the slopes and using them in the study of the fundamental group are found in [Degtyarev 2012].

Example 3.15. At the only improper fiber $x_i = 0$ described in Remark 3.14 the slope is the class of $\alpha_j\alpha_{j+1}$, where $(j, j+1)$ are the two branches merging at the node; see Figure 3. This fact can easily be seen using a local model. In a small neighborhood of $x = 0$, one can assume that $\tilde{C}$ is given by $(y-a)(y-b) = 0$. Let $\tilde{\Phi} \subset \Sigma_{\tilde{d}}$ and $\Phi \subset \Sigma_d$ be the disk $\{y = c, |x| \leqslant 1\}$, $c \in \mathbb{R}$ and $c \gg |a|, |b|$. Then, the relevant part of C is the node $(y - ax)(y - bx) = 0$, and $\tilde{\Phi}$ projects onto the disk $\tilde{\Phi}' = \{y = cx, |x| \leqslant 1\}$, which meets Φ at $(1, c)$. Now, consider one full turn $x = \exp(2\pi i t)$, $t \in [0, 1]$, and follow the point (x, cx) in $\partial\tilde{\Phi}'$: it describes the circle $y = c\exp(2\pi i t)$ encompassing once the two points of the intersection $C \cap F_1^\circ$. The class $\alpha_j\alpha_{j+1}$ of this circle is the slope. Even more precisely, one should start with the constant path $[0, 1] \to (1, c)$ and homotope this path in $F_x^\circ \smallsetminus C$, keeping one end in Φ and the other in $\tilde{\Phi}'$. In the terminal position, the path is a loop again, and its class $\alpha_j\alpha_{j+1}$ is the slope.

Define the *twisted local monodromy* $\tilde{\mu}_i := \mu_i \cdot \operatorname{inn} \varkappa_i$, where $\operatorname{inn} : G \to \operatorname{Aut} G$ is the homomorphism sending an element g of a group G to the inner automorphism $\operatorname{inn} g : h \mapsto g^{-1}hg$. Thus, $\tilde{\mu}_i : \pi(x_i^+) \to \pi(x_i^+)$ is the map $\alpha \mapsto \varkappa_i^{-1}(\alpha \uparrow \mu_i)\varkappa_i$. In general, $\tilde{\mu}_i$ is not a braid. Take $x_0^+ = x_\infty^+$ for the reference fiber and consider the braids

$$\rho_i := \prod_{j=1}^{i} \gamma_{j-1}\beta_j : \pi(x_0^+) \to \pi(x_i^+), \quad i = 1, \dots, r+1 = \infty$$

(left to right product), the (global) slopes

$$\bar{\varkappa}_i := \varkappa_i \uparrow \rho_1^{-1} \in \pi(x_0^+), \quad i = 1, \dots, r,$$

and the twisted monodromy homomorphisms

$$\tilde{\mathfrak{m}}_i := \rho_i \tilde{\mu}_i \rho_i^{-1} : \pi(x_0^+) \to \pi(x_0^+), \quad i = 1, \dots, r.$$

The following theorem is essentially due to Zariski and van Kampen [van Kampen 1933], and the particular case of improper curves in Hirzebruch surfaces, treated by means of the slopes, is considered in detail in [Degtyarev 2012, Section 5.1.3]. Here, we state and outline the proof of a very special case of this approach, incorporating the (partial) computation of the braid monodromy of a real tetragonal curve in terms of its real part.

We use the following common convention: given an automorphism β of the free group $\langle \alpha_1, \dots, \alpha_4 \rangle$, the *braid relation* $\beta = \mathrm{id}$ stands for the quadruple of relations $\alpha_j \uparrow \beta = \alpha_j$, $j = 1, \dots, 4$. Note that, since β is an automorphism, this is equivalent to the infinitely many relations $\alpha = \alpha \uparrow \beta$, $\alpha \in \langle \alpha_1, \dots, \alpha_4 \rangle$.

Theorem 3.16. *In the notation above, the inclusion of the reference fiber induces an epimorphism*

$$\pi(x_0^+) = \langle \alpha_1, \dots, \alpha_4 \rangle \twoheadrightarrow \pi_1,$$

and the relations $\tilde{\mathfrak{m}}_i = \mathrm{id}$, $i = 1, \dots, r$, *hold in* π_1. *If the fiber* $x = \infty$ *is nonsingular and all nonreal singular fibers are proper for* $\tilde{C}$, *then one also has the* relations at infinity $\rho_\infty = \mathrm{id}$ *and* $(\alpha_1 \cdots \alpha_4)^d = \bar{\varkappa}_r \cdots \bar{\varkappa}_1$. *If, in addition,* C *has at most one pair of conjugate nonreal singular fibers, then the relations listed define* π_1.

Proof. The assertion is a restatement of the classical Zariski–van Kampen theorem modified for the case of improper curves; see [Degtyarev 2012, Theorem 5.50]. The relation at infinity $(\alpha_1 \cdots \alpha_4)^d = \bar{\varkappa}_r \cdots \bar{\varkappa}_1$ holds in π_1 whenever all slopes not accounted for, namely those at the nonreal fibers, are known to be trivial. The automorphism $\rho_{r+1} : \pi(x_0^+) \to \pi(x_{r+1}^+) = \pi(x_0^+)$ is the monodromy along the "boundary" of the upper half-plane $\operatorname{Im} x > 0$ (see Figure 2), i.e., the product of the monodromies about all singular fibers in this half-plane; if the slopes at these fibers are all trivial, then $\rho_{r+1} = \mathrm{id}$ in π_1. Finally, if $\tilde{C}$ has at most one pair of conjugate nonreal singular fibers, then all but possibly one braid relations are present and hence they define the group; see [Degtyarev 2012, Lemma 5.59]. □

4. The computation

4A. ***The strategy.*** We start with a plane sextic $D \subset \mathbb{P}^2$ and choose homogeneous coordinates $(z_0 : z_1 : z_2)$ so that D has a singular point of type $\boldsymbol{A}_m$, $m \geq 3$, at $(0:0:1)$ tangent to the axis $\{z_1 = 0\}$. Then, in the affine coordinates $x := z_1/z_0$, $y := z_2/z_0$, the curve D is given by a polynomial $\tilde{f}$ as in Remark 3.14, and the same polynomial $\tilde{f}$ defines a certain tetragonal curve $\tilde{C} \subset \Sigma_1$, viz. the proper transform of D under the blow-up of $(0:0:1)$. The common fundamental group

$$\pi_1 := \pi_1(\mathbb{P}^2 \smallsetminus D) = \pi_1(\Sigma_1 \smallsetminus (\tilde{C} \cup E))$$

is computed using Theorem 3.16 applied to $\tilde{C}$ and its transform $C \subset \Sigma_2$, with the only nontrivial slope $\varkappa = \alpha_1\alpha_2$ or $\alpha_3\alpha_4$ over $x = 0$ given by Example 3.15. (Here, $E \subset \Sigma_1$ is the exceptional section, i.q. the exceptional divisor over the point $(0:0:1)$ blown up.) *A priori*, Theorem 3.16 may only produce a certain group $\mathfrak{g}$ that surjects onto π_1 rather than π_1 itself; however, in most cases this group $\mathfrak{g}$ is "minimal expected" (see Section 4D below) and we do obtain π_1.

The assumption that the fiber $x = \infty$ is nonsingular is not essential as long as the singularity over ∞ is taken into consideration: one can always move ∞ to a generic point by a real projective change of coordinates. To keep the defining equations as simple as possible, we assume such a change of coordinates implicitly. Furthermore, it is only the cyclic order of the singular fibers in the circle $\mathbb{P}^1_{\mathbb{R}}$ that matters, and sometimes we reorder the fibers by applying a cyclic permutation to their "natural" indices. In other words, the braid $\beta_\infty = \Delta^2$ is in the center of $\mathbb{B}_4$ and, hence, it can be inserted at any place in the relation $\gamma_0\beta_1\gamma_1 \cdots \gamma_r\beta_\infty = \mathrm{id}$.

To compute the braids, we outline the real (bold lines) and imaginary (dotted lines) branches of C in the figures. Recall that it is only the mutual position of the real branches and their intersection indices with the imaginary ones that matters; see Remark 3.13. The "special" node that contributes the only nontrivial slope (the blow-up center in the passage from C to $\tilde{C}$; see Remark 3.14) is marked with a white dot; the other singular points of C (including those of type $\boldsymbol{A}_0$) are marked with black dots. The shape of the curve can mostly be recovered using Remarks 3.7 and 3.8; however, it is usually easier to determine the mutual position of the roots directly via `Maple`. The braids β_i, μ_i, and γ_j are computed from the figures as explained in Section 3D.

Warning 4.1. The polynomial f_{nr} given by Remark 3.14 is used to determine the slope and mutual position of the two singular points over $x = 0$: the "special" node is always at $(0, 0)$. For all other applications, e.g., for Proposition 3.12, this polynomial should be converted to the reduced form (3.1).

4B. ***Relations.*** Recall that a braid relation $\tilde{\mathfrak{m}}_i = \mathrm{id}$ stands for a quadruple of relations $\alpha_j \uparrow \tilde{\mathfrak{m}}_i = \alpha_j$, $j = 1, \ldots, 4$. Alternatively, this can be regarded as an infinite sequence of relations $\alpha \uparrow \tilde{\mathfrak{m}}_i = \alpha$, $\alpha \in \mathbb{F}_4$, or, equivalently, as a quadruple of relations $\alpha'_j \uparrow \tilde{\mathfrak{m}}_i = \alpha'_j$, $j = 1, \ldots, 4$, where $\alpha'_1, \ldots, \alpha'_4$ is *any* basis for $\mathbb{F}_4$. For this reason, in the computation below we start with the braid relations $\alpha'_j \uparrow \tilde{\mu}_i = \alpha'_j$ in the canonical basis over x_i^+ and translate them to x_0^+ via ρ_i^{-1}. In the most common case $\tilde{\mu}_i = \sigma_r^p$, $r = 1, 2, 3$, $p \in \mathbb{Z}$, the whole quadruple is equivalent to the single relation $\{\alpha'_r, \alpha'_{r+1}\}_p = 1$, where

$$\{\alpha, \beta\}_{2k} := (\alpha\beta)^k(\beta\alpha)^{-k}, \quad \{\alpha, \beta\}_{2k+1} := (\alpha\beta)^k\alpha(\alpha\beta)^{-k}\beta^{-1}.$$

Remark 4.2. The braid relations about the fiber $x_k = 0$ with the only nontrivial slope (see Example 3.15) can also be presimplified. Let $\alpha'_1, \ldots, \alpha'_4$ be the canonical basis in x_k^+. If $\varkappa_k = \alpha'_1\alpha'_2$ and $\mu_k = \sigma_1^2\sigma_3^p$, the braid relations $\tilde{\mu}_k = \mathrm{id}$ and relation at infinity $(\alpha'_1 \cdots \alpha'_4)^2 = \varkappa_k$ together are equivalent to

$$\alpha'_1\alpha'_2(\alpha'_3\alpha'_4)^2 = \{\alpha'_3, \alpha'_4\}_{p+4} = 1.$$

Similarly, if $\varkappa_k = \alpha'_3\alpha'_4$ and $\mu_k = \sigma_1^p\sigma_3^2$, we obtain

$$(\alpha'_1\alpha'_2)^2\alpha'_3\alpha'_4 = \{\alpha'_1, \alpha'_2\}_{p+4} = 1.$$

Certainly, these relations should be translated back to x_0^+ via ρ_k^{-1}. Note, though, that we do not use this simplification in the sequel.

Remark 4.3. In some cases, simpler relations are obtained if another point x_i^+, $i > 0$, is taken for the reference fiber. To do so, one merely replaces the braids ρ_j, $j = 1, \ldots, r + 1 = \infty$, with $\rho'_j := \rho_i^{-1}\rho_j$.

All computations below were performed using GAP [GAP Group 2008], with the help of the simple braid manipulation routines contained in [Degtyarev 2012]. The GAP code can be found at http://www.fen.bilkent.edu.tr/~degt/papers/papers.htm. The processing is almost fully automated, the input being the braids β_i, μ_i, γ_j and the only nontrivial slope $\varkappa_k = \alpha_1\alpha_2$ or $\alpha_3\alpha_4$, which are read off from the diagrams depicting the curves.

4C. ***The set of singularities $3A_6 \oplus A_1$, line 37.*** Any sextic with this set of singularities is $\mathbb{D}_{14}$-special (see [Degtyarev 2008]), and, according to [Degtyarev and Oka 2009], any $\mathbb{D}_{14}$-special sextic can be given by an equation of the form

$$\begin{aligned}
&2t(t^3-1)(z_0^4z_1z_2 + z_1^4z_2z_0 + z_2^4z_0z_1)\\
&\quad + (t^3-1)(z_0^4z_1^2 + z_1^4z_2^2 + z_2^4z_0^2) + t^2(t^3-1)(z_0^4z_2^2 + z_1^4z_0^2 + z_2^4z_1^2)\\
&\quad + 2t(t^3+1)(z_0^3z_1^3 + z_1^3z_2^3 + z_2^3z_0^3) + 4t^2(t^3+2)(z_0^3z_1^2z_2 + z_1^3z_2^2z_0 + z_2^3z_0^2z_1)\\
&\quad + 2(t^6+4t^3+1)(z_0^3z_1z_2^2 + z_1^3z_2z_0^2 + z_2^3z_0z_1^2) + t(t^6+13t^3+10)z_0^2z_1^2z_2^2,
\end{aligned}$$

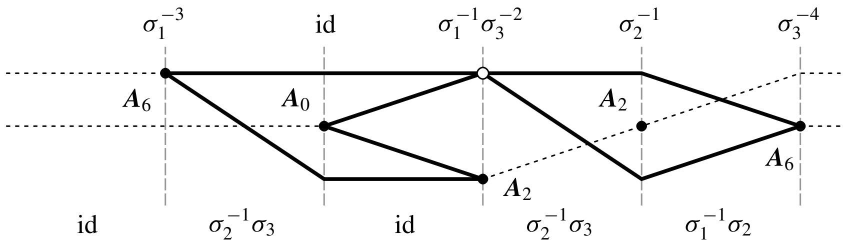


Figure 5. The set of singularities $3A_6 \oplus A_1$, line 37.

$t^3 \neq 1$. The set of singularities of this curve is $3A_6 \oplus A_1$ if and only if $t^3 = -27$; we use the real value $t = -3$. After the substitution $z_0 = 1$, $z_1 = x + \frac{1}{3}$, and $z_2 = y/x$ the equation is brought to the form considered in Remark 3.14. Up to a positive factor, the discriminant (3.4) with respect to y is

$$-x^5(27x^3 - 648x^2 + 6363x + 7)(3x-2)^2(3x+1)^7,$$

which has real roots

$$x_1 = -\tfrac{1}{3}, \quad x_2 \approx -0.001, \quad x_3 = 0, \quad x_4 = \tfrac{2}{3}, \quad x_5 = \infty$$

and two simple imaginary roots. Hence, Theorem 3.16 does compute the group.

The only root of q on the real segment $[-\infty, x_1]$ is $x' \approx -3.48$, and $b_1(x') < 0$; hence, one has $\gamma_0 = \mathrm{id}$; see Proposition 3.12. The other braids β_i, γ_j are easily found from Figure 5, and, using Theorem 3.16 and GAP, we obtain a group of order 42. This concludes the proof of Theorem 1.1. □

4D. ***Sextics of torus type.*** All maximal, in the sense of degeneration, sextics of torus type are described in [Oka and Pho 2002], where a sextic D is represented by a pair of polynomials $f_2(x, y)$, $f_3(x, y)$ of degree 2 and 3, respectively, so that the defining polynomial of D is $f_{\mathrm{tor}} := f_2^3 + f_3^2$. (Below, these equations are cited in a slightly simplified form: I tried to clear the denominators by linear changes of variables and appropriate coefficients.) Each curve (at least, each of those considered below) has a type A_m, $m \geq 3$, singularity at $(0, 0)$ tangent to the y-axis. Hence, we start with the substitution $\tilde{f}(x, y) := y^6 f_{\mathrm{tor}}(x/y, 1/y)$ to obtain a polynomial $\tilde{f}$ as in Remark 3.14; then we proceed as in Section 4A.

To identify the group g given by Theorem 3.16 as Γ, we use the following GAP code, which was suggested to me by E. Artal:

(4.4)
```
P := PresentationNormalClosure(g, Subgroup(g, a));
SimplifyPresentation(P);
```

here, a is an appropriate ratio $\alpha_i \alpha_j^{-1}$ which normally generates the commutant of g. If the resulting presentation has two generators and no relations, we conclude that $\mathrm{g} = \pi_1 = \Gamma$, *even when the statement of Theorem 3.16 does not guarantee a complete set of relations*. Indeed, *a priori* we have epimorphisms $\mathrm{g} \twoheadrightarrow \pi_1 \twoheadrightarrow \Gamma$ (the latter follows from the fact that the curve is assumed to be of torus type), which induce epimorphisms $[\mathrm{g}, \mathrm{g}] \twoheadrightarrow [\pi_1, \pi_1] \twoheadrightarrow [\Gamma, \Gamma] = \mathbb{F}_2$ of the commutants. If $[\mathrm{g}, \mathrm{g}] = \mathbb{F}_2$, both these epimorphisms are isomorphisms (since $\mathbb{F}_2$ is Hopfian) and the 5-lemma implies that $\mathrm{g} \twoheadrightarrow \pi_1 \twoheadrightarrow \Gamma$ are also isomorphisms.

In some cases (e.g., in Sections 4E and 4F), the call `SimplifiedFpGroup(g)` returns a recognizable presentation of Γ.

4E. ***The set of singularities $(A_{14} \oplus A_2) \oplus A_3$, line 8.*** The curve in question is `nt139` in [Oka and Pho 2002]:

$$f_2 = 80(-36y^2 + 120xy - 82x^2 + 2x),$$
$$f_3 = 100(-1512y^3 + 7794y^2x - 18y^2 - 11664yx^2 + 144xy + 5313x^3 - 194x^2 + x).$$

Up to a positive coefficient, the discriminant of f_{nr} is

$$x^{13}(5120x^4 + 36864x^3 + 3456x^2 - 2160x - 405)(x-1)^3.$$

It has five real roots, which we reorder cyclically as follows:

$$x_1 = 0, \quad x_2 \approx 0.27, \quad x_3 = 1, \quad x_4 = \infty, \quad x_5 \approx -7.1.$$

Besides, there are two conjugate imaginary singular fibers, which are of type A_0.

The curve is depicted in Figure 6, from which all braids β_i, γ_j are easily found. Taking x_0^+ for the reference fiber and using $\mathtt{a} = \alpha_1\alpha_2^{-1}$ in (4.4), we obtain $\pi_1 = \Gamma$.

4F. ***The set of singularities $(A_{14} \oplus A_2) \oplus A_2 \oplus A_1$, line 9.*** The curve is `nt142` in [Oka and Pho 2002]:

$$f_2 = -45y^2 - 240yx - 106x^2 + 90x,$$
$$f_3 = 1025y^3 + 6045y^2x - 375y^2 + 5490yx^2 - 4050yx + 1354x^3 - 2040x^2 + 750x.$$

Up to a positive coefficient, the discriminant of f_{nr} is

$$x^{13}(8x^3 - 10720x^2 + 14250x - 5625)(x+1)^2(14x+15)^3,$$

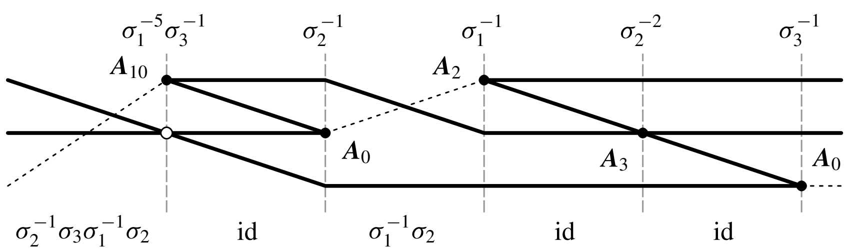


Figure 6. The set of singularities $(A_{14} \oplus A_2) \oplus A_3$, line 8.

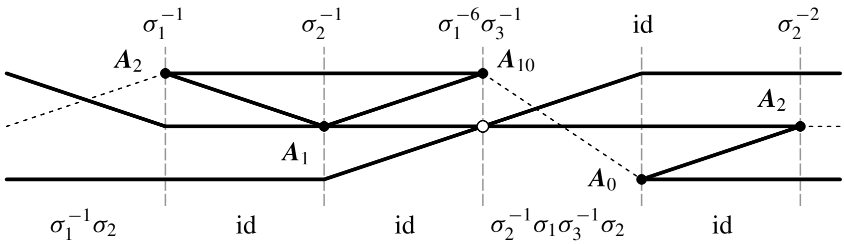


Figure 7. The set of singularities $(A_{14} \oplus A_2) \oplus A_2 \oplus A_1$, line 9.

and all its roots are real:

$$x_1 = -\tfrac{15}{14}, \quad x_2 = -1, \quad x_3 = 0, \quad x_4 \approx 1338, \quad x_5 = \infty.$$

The braids β_i, γ_j are found from Figure 7 and, using x_0^+ as the reference fiber and $\mathtt{a} = \alpha_1\alpha_2^{-1}$ in (4.4), we conclude that $\pi_1 = \Gamma$.

4G. ***The set of singularities $(A_{11} \oplus 2A_2) \oplus A_4$, line 17.*** This is `nt118` in [Oka and Pho 2002]:

$$\begin{aligned} f_2 &= \tfrac{1}{5}(-3456y^2 + 1200yx - 3005x^2 + 240x), \\ f_3 &= \tfrac{1}{5}(-89856y^3 + 130464y^2x - 6912y^2 - 112680yx^2 + 8640yx \\ &\qquad + 91345x^3 - 13320x^2 + 480x). \end{aligned}$$

Up to a positive coefficient, the discriminant of f_{nr} is

$$-x^{10}(25x^3 + 290x^2 + 360x + 162)(35x^2 - 384x + 1152)^3.$$

It has three real roots, which we reorder cyclically as follows:

$$x_1 = 0, \quad x_2 = \infty, \quad x_3 \approx -10.26.$$

In addition, there are two pairs of complex conjugate singular fibers, of types A_2 and A_0. Thus, *a priori* Theorem 3.16 only gives us a certain epimorphism $\mathtt{g} \twoheadrightarrow \pi_1$. However, using $\mathtt{a} = \alpha_1\alpha_2^{-1}$ in (4.4), we conclude that $\mathtt{g} = \pi_1 = \Gamma$. (All braids are found from Figure 8 and the reference fiber is x_1^+; see Remark 4.3.)

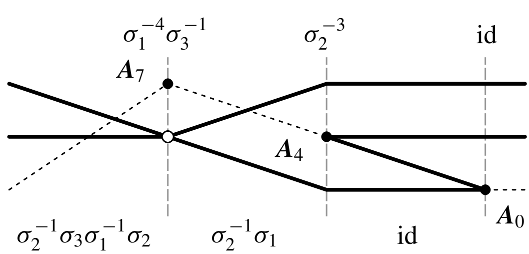


Figure 8. The set of singularities $(A_{11} \oplus 2A_2) \oplus A_4$, line 17.

4H. ***The set of singularities*** $(\boldsymbol{A_8 \oplus 3A_2) \oplus A_4 \oplus A_1}$, ***line 33.*** This curve is `nt83` in [Oka and Pho 2002]:

$$
\begin{aligned}
&f_2 = -565y^2 - 14yx + 176y - 5x^2 + 104x - 16,\\
(4.5)\qquad &f_3 = 13321y^3 + 3135y^2x - 6294y^2 + 207yx^2 - 3516yx + 1056y\\
&\qquad + 25x^3 - 558x^2 + 624x - 64.
\end{aligned}
$$

Up to a positive coefficient, the discriminant of f_{nr} is

$$x^3(x+3)(x+9)^2(11915x^3 + 96579x^2 - 14823x + 729)^3(x-9)^9.$$

It has five real roots, which we reorder cyclically as follows:

$$x_1 = 0, \quad x_2 = 9, \quad x_3 = -9, \quad x_4 \approx -8.26, \quad x_5 = -3.$$

We conclude that the curve has only two nonreal singular fibers, which are cusps. Hence, Theorem 3.16 gives us a complete presentation of π_1.

In the interval (x_5, x_1), where f has four imaginary branches, q has four roots

$$x_1' \approx -2.93, \quad x_2' = -1.92, \quad x_3' \approx -0.79, \quad x_4' \approx -0.14,$$

with b_1 negative at x_1', x_3' and positive at x_2', x_4'; at the latter two points one also has $q' < 0$. Hence, $\gamma_0 = \tau^{-2}$; see Proposition 3.12. All other braids are easily found from Figure 9.

Remark 4.6. For a further simplification, observe that the braid ρ_∞ appearing in Theorem 3.16 equals

$$\sigma_2^{-1}\sigma_1\sigma_3^{-1}\sigma_1\sigma_3^{-1}\sigma_2 \cdot \sigma_1^{-1} \cdot \sigma_2^{-1}\sigma_1 \cdot \sigma_2^{-4} \cdot \sigma_3^{-1} \cdot \sigma_2^{-2} \cdot \sigma_3^{-1}\sigma_2 \cdot \sigma_1^{-1} \cdot (\sigma_3\sigma_1\sigma_2)^4,$$

and one can check that $\rho_\infty = \rho_{\mathrm{im}}^{-1}\sigma_1^3\rho_{\mathrm{im}}$, where $\rho_{\mathrm{im}} := \sigma_2\sigma_1^{-1}\sigma_3^2\sigma_2$. (Note that ρ_∞ represents the monodromy about a single imaginary cusp of the curve; hence, it is expected to be conjugate to σ_1^3.) Thus, we can replace the quadruple of relations $\rho_\infty = \mathrm{id}$ with a single relation $\{\alpha_1, \alpha_2\}_3 \uparrow \rho_{\mathrm{im}} = 1$; see Section 4B.

Now, taking x_3^+ for the reference fiber (see Remark 4.3), using Remark 4.6, and applying `SimplifiedFpGroup(g)`, we arrive at (1.3). This presentation has three

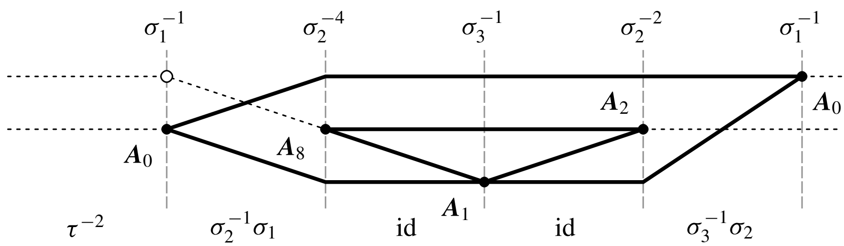


Figure 9. The set of singularities $(\boldsymbol{A}_8 \oplus 3\boldsymbol{A}_2) \oplus \boldsymbol{A}_4 \oplus \boldsymbol{A}_1$, line 33, projected from $\boldsymbol{A}_4$.

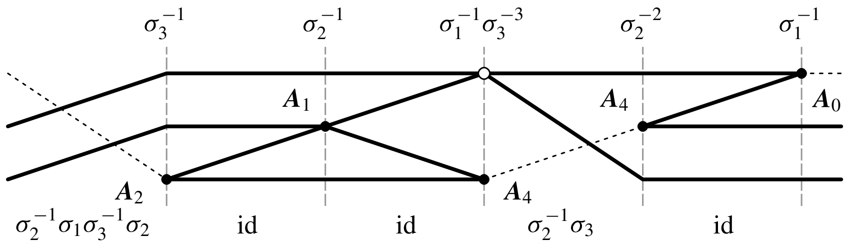


Figure 10. The set of singularities $(A_8 \oplus 3A_2) \oplus A_4 \oplus A_1$, line 33, projected from A_8.

generators and four relations of total length 48. Together with the previous sections, this concludes the proof of Theorem 1.2. □

Remark 4.7. The Alexander module of the group π_1 considered in this section is $\mathbb{Z}[t, t^{-1}]/(t^2 - t + 1)$, and the finite quotients π_1/α_2^p, $p = 2, 3, 4$, are isomorphic to the similar quotients of Γ. My laptop failed to compute the order of π_1/α_2^5.

Remark 4.8. In (4.5), the singular point at the origin is of type A_4. One can start with a change of variables $x \mapsto y + 9$, $y \mapsto x + 1$ and resolve the type A_8 point instead. The tetragonal model is depicted in Figure 10, and the computation becomes slightly simpler, but the resulting presentation is of the same complexity, even with the additional observation that $\rho_\infty = \rho_{\mathrm{im}}^{-1}\sigma_1^3\rho_{\mathrm{im}}$, where $\rho_{\mathrm{im}} := \sigma_2\sigma_1^{-1}\sigma_3\sigma_2$; see Remark 4.6.

4I. ***Proof of Proposition 1.4.*** For the sets of singularities $(A_{14} \oplus A_2) \oplus A_3$, line 8, $(A_{14} \oplus A_2) \oplus A_2 \oplus A_1$, line 9, and $(A_{11} \oplus 2A_2) \oplus A_4$, line 17, the statement is an immediate consequence of [Degtyarev 2012, Theorem 7.48]. For $3A_6 \oplus A_1$, line 37, the only proper quotient of the commutant $[\pi_1, \pi_1] = \mathbb{Z}_7$ is trivial; hence, the group π_1' of any perturbation D' is either abelian, $\pi_1' = \mathbb{Z}_6$, or isomorphic to π_1, the latter being the case if and only if D' is $\mathbb{D}_{14}$-special; see [Degtyarev 2008].

For the remaining set of singularities $(A_8 \oplus 3A_2) \oplus A_4 \oplus A_1$, line 33, we proceed as follows. Any proper perturbation factors through a maximal one, where a single singular point P of type A_m splits into two points $A_{m'}$, $A_{m''}$, so that $m' + m'' = m - 1$. Assume that $P \neq (0 : 0 : 1)$; see Section 4A. Then this point corresponds to a certain singular fiber x_i of the tetragonal model C and gives rise to a braid relation $\{\alpha_k, \alpha_{k+1}\}_{m+1} \uparrow \rho_i^{-1} = 1$; see Section 4B. For the new curve D', this relation changes to $\{\alpha_k, \alpha_{k+1}\}_s \uparrow \rho_i^{-1} = 1$, where $s := \text{g.c.d.}(m' + 1, m'' + 1)$.

For any perturbation of any point P, we have $s = 3$ if P is of type A_8 or A_2 and the result is still of torus type, and $s = 1$ otherwise. Now, the statement is easily proved by repeating the computation with the braid $\mu_i = \sigma_k^{m+1}$ replaced with σ_k^s. (If it is the type A_4 point that is perturbed, one can use the alternative tetragonal model given by Remark 4.8.) □

Acknowledgements

This paper was written during my sabbatical stay at *l'Instutut des Hautes Études Scientifiques* and *Max-Planck-Institut für Mathematik*; I would like to thank these institutions for their support and hospitality. I am also grateful to M. Oka, who kindly clarified for me the results of [Oka and Pho 2002], to V. Kharlamov, who patiently introduced me to the history of the subject, and to the anonymous referee of the paper, who made a number of valuable suggestions and checked and confirmed the somewhat unexpected result of Section 4H.

References

[Artal Bartolo et al. 2002] E. Artal Bartolo, J. Carmona Ruber, and J. I. Cogolludo Agustín, "On sextic curves with big Milnor number", pp. 1–29 in *Trends in singularities*, edited by A. Libgober and M. Tibăr, Birkhäuser, Basel, 2002. MR 2003d:14034 Zbl 1018.14009

[Artal-Bartolo et al. 2008] E. Artal-Bartolo, J. I. Cogolludo, and H.-o. Tokunaga, "A survey on Zariski pairs", pp. 1–100 in *Algebraic geometry in East Asia* (Hanoi, 2005), edited by K. Konno and V. Nguyen-Khac, Adv. Stud. Pure Math. **50**, Math. Soc. Japan, Tokyo, 2008. MR 2009g:14030 Zbl 1141.14015

[Artin 1947] E. Artin, "Theory of braids", *Ann. of Math.* (2) **48** (1947), 101–126. MR 8,367a Zbl 0030.17703

[Degtyarev 2008] A. Degtyarev, "Oka's conjecture on irreducible plane sextics", *J. Lond. Math. Soc.* (2) **78**:2 (2008), 329–351. MR 2009f:14054 Zbl 1158.14026

[Degtyarev 2009a] A. Degtyarev, "Fundamental groups of symmetric sextics, II", *Proc. Lond. Math. Soc.* (3) **99**:2 (2009), 353–385. MR 2011c:14091 Zbl 1173.14019

[Degtyarev 2009b] A. Degtyarev, "On irreducible sextics with non-abelian fundamental group", pp. 65–91 in *Singularities* (Toyama, 2007), edited by J.-P. Brasselet et al., Adv. Stud. Pure Math. **56**, Math. Soc. Japan, Tokyo, 2009. MR 2011e:14051 Zbl 1193.14037

[Degtyarev 2012] A. Degtyarev, *Topology of algebraic curves: An approach via dessins d'enfants*, de Gruyter Studies in Mathematics **44**, de Gruyter, Berlin, 2012. MR 2952675 Zbl 06055189

[Degtyarev and Oka 2009] A. Degtyarev and M. Oka, "A plane sextic with finite fundamental group", pp. 93–108 in *Singularities* (Toyama, 2007), edited by J.-P. Brasselet et al., Adv. Stud. Pure Math. **56**, Math. Soc. Japan, Tokyo, 2009. MR 2011g:14075 Zbl 1193.14038

[GAP Group 2008] GAP Group, *GAP – groups, algorithms, and programming, version 4.4.12*, 2008, http://www.gap-system.org.

[van Kampen 1933] E. R. van Kampen, "On the fundamental group of an algebraic curve", *Amer. J. Math.* **55**:1-4 (1933), 255–267. MR 1506962 Zbl 0006.41502

[Libgober 2007a] A. Libgober, "Lectures on topology of complements and fundamental groups", pp. 71–137 in *Singularity theory*, edited by D. Chéniot et al., World Scientific, Hackensack, NJ, 2007. MR 2008i:14043 Zbl 1125.14014

[Libgober 2007b] A. Libgober, "Problems in topology of the complements to plane singular curves", pp. 370–387 in *Singularities in geometry and topology* (Trieste, 2005), edited by J.-P. Brasselet et al., World Scientific, Hackensack, NJ, 2007. MR 2008c:14044 Zbl 1123.14018

[Oka and Pho 2002] M. Oka and D. T. Pho, "Classification of sextics of torus type", *Tokyo J. Math.* **25**:2 (2002), 399–433. MR 2003k:14030 Zbl 1062.14036

[Orevkov 1999] S. Y. Orevkov, "Link theory and oval arrangements of real algebraic curves", *Topology* **38**:4 (1999), 779–810. MR 2000b:14066 Zbl 0923.14032

[Persson 1985] U. Persson, "Double sextics and singular K-3 surfaces", pp. 262–328 in *Algebraic geometry* (Barcelona, 1983), edited by E. Casas-Alvero et al., Lecture Notes in Math. **1124**, Springer, Berlin, 1985. MR 87i:14036 Zbl 0603.14023

[Shimada 2007] I. Shimada, "On the connected components of the moduli of polarized $K3$ surfaces", preprint, 2007, http://tinyurl.com/shimada-connK3.

[Yang 1996] J.-G. Yang, "Sextic curves with simple singularities", *Tohoku Math. J.* (2) **48**:2 (1996), 203–227. MR 98e:14026 Zbl 0866.14014

[Zariski 1929] O. Zariski, "On the problem of existence of algebraic functions of two variables possessing a given branch curve", *Amer. J. Math.* **51**:2 (1929), 305–328. MR 1506719 JFM 55.0806.01

Received July 10, 2012. Revised December 29, 2012.

ALEX DEGTYAREV
DEPARTMENT OF MATHEMATICS
BILKENT UNIVERSITY
06800 ANKARA
TURKEY
degt@fen.bilkent.edu.tr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.349

A COMPUTATIONAL APPROACH TO THE KOSTANT–SEKIGUCHI CORRESPONDENCE

HEIKO DIETRICH AND WILLEM A. DE GRAAF

Let $\mathfrak{g}$ be a real form of a simple complex Lie algebra. Based on ideas of Đoković and Vinberg, we describe an algorithm to compute representatives of the nilpotent orbits of $\mathfrak{g}$ using the Kostant–Sekiguchi correspondence. Our algorithms are implemented for the computer algebra system GAP and, as an application, we have built a database of nilpotent orbits of all real forms of simple complex Lie algebras of rank at most 8. In addition, we consider two real forms $\mathfrak{g}$ and $\mathfrak{g}'$ of a complex simple Lie algebra $\mathfrak{g}^c$ with Cartan decompositions $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and $\mathfrak{g}' = \mathfrak{k}' \oplus \mathfrak{p}'$. We describe an explicit construction of an isomorphism $\mathfrak{g} \to \mathfrak{g}'$, respecting the given Cartan decompositions, which fails if and only if $\mathfrak{g}$ and $\mathfrak{g}'$ are not isomorphic. This isomorphism can be used to map the representatives of the nilpotent orbits of $\mathfrak{g}$ to other realizations of the same algebra.

1. Introduction

When considering the action of a Lie group on its Lie algebra, the question arises as to what its orbits are. This question has mainly been studied for complex simple Lie algebras $\mathfrak{g}^c$, with their adjoint groups G^c. Particularly the theory concerning nilpotent orbits (that is, G^c-orbits consisting of nilpotent elements) has seen many interesting developments over the past decades; we refer to [Collingwood and McGovern 1993] for a detailed account. These orbits have been classified in terms of combinatorial objects called weighted Dynkin diagrams, using a beautiful connection between nilpotent orbits and orbits of $\mathfrak{sl}_2$-triples. If $\mathfrak{g}^c$ is of classical type, then the nilpotent orbits also have been classified in terms of certain sets of partitions (of the dimension of the natural representation).

For real Lie algebras $\mathfrak{g}$, with the action of the adjoint group G, it is much harder to classify the nilpotent (G-)orbits. The main problem compared to the complex case is that a weighted Dynkin diagram can correspond to several nilpotent orbits.

The research leading to these results has received funding from the European Union's Seventh Framework Program FP7/2007-2013 under grant agreement No. 271712.
MSC2010: 17B45, 20G20.
Keywords: real Lie algebra, real nilpotent orbit, computational methods, Kostant–Sekiguchi correspondence.

To illustrate this phenomenon consider $G^c = \mathrm{PSL}_n(\mathbb{C})$ and $G = \mathrm{PSL}_n(\mathbb{R})$ with their Lie algebras $\mathfrak{g}^c = \mathfrak{sl}_n(\mathbb{C})$ and $\mathfrak{g} = \mathfrak{sl}_n(\mathbb{R})$. The nilpotent orbits in $\mathfrak{g}^c$ are parametrized by partitions of n, whose parts correspond to the sizes of the Jordan blocks of a representative of the orbit. The nilpotent orbits in $\mathfrak{g}$ are associated with the same partitions, with the difference that the partitions with only even terms correspond to two nilpotent orbits.

More generally, the nilpotent orbits of the simple real Lie algebras of classical type have been classified in terms of combinatorial objects such as partitions or certain types of Young diagrams; see [Collingwood and McGovern 1993, Section 9.3]. For the classification in Lie algebras of exceptional types the main ingredient is the Kostant–Sekiguchi correspondence: Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be a Cartan decomposition of the simple real Lie algebra $\mathfrak{g}$ with complexification $\mathfrak{g}^c = \mathfrak{k}^c \oplus \mathfrak{p}^c$. Let G^c be the adjoint group of $\mathfrak{g}^c$ and denote by G, K, and K^c the connected Lie subgroups of G^c with corresponding Lie algebras $\mathfrak{g}$, $\mathfrak{k}$, and $\mathfrak{k}^c$, respectively. The Kostant–Sekiguchi correspondence states a one-to-one correspondence between the nilpotent orbits in $\mathfrak{g}$ and the nilpotent K^c-orbits in $\mathfrak{p}^c$. Although this correspondence can be described explicitly (as we will do in Section 3), it is difficult to obtain concrete representatives of nilpotent orbits in $\mathfrak{g}$. Most classification results therefore are on the *complex side*, that is, consider nilpotent K^c-orbits in $\mathfrak{p}^c$; see for example [Đoković 1988; Galina 2009; Noël 1998; 2001a; 2001b]. However, in tedious work, Đoković [1998; 1999; 2000] has used this correspondence to obtain representatives of the nilpotent orbits for each of the simple real Lie algebras of exceptional type.

The aim of this paper is to describe methods for constructing representatives of the nilpotent orbits of a real simple Lie algebra on a computer. One approach to obtain representatives is to take the existing classifications in the literature, to set up isomorphisms to the algebras given, and to map the given representatives. However, it is not straightforward to verify the correctness of the representatives given in the literature, so this approach is rather error-prone. (In fact, in each of his papers cited above, Đoković corrected some errors, due to typos, in his previous papers.) For this reason we devise algorithms that effectively carry out the Kostant–Sekiguchi correspondence. Since the correctness of each step can be checked algorithmically, we get a certified list of representatives.

1.1. ***Main results.*** We describe computational methods to achieve three aims:

(A) Construct isomorphism type representatives for all real forms of a simple complex Lie algebra.

(B) Construct representatives of all nilpotent orbits of a real form constructed in (A).

(C) Construct an isomorphism between two given real forms of a simple complex Lie algebra.

For computational purposes it is often needed that the Lie algebras are given by means of a multiplication table (with respect to some basis). We describe in Section 2 how to construct multiplication tables for all real forms of simple complex Lie algebras (up to isomorphism).

In Sections 3–6 we describe our algorithms to construct representatives of the nilpotent orbits of a Lie algebra constructed in (A). We combine the Kostant–Sekiguchi correspondence (see Section 3) with the theory of carrier algebras developed in [Vinberg 1979] (see Section 4). This is inspired by Đoković's [1987] proof of the Kostant–Sekiguchi correspondence. In Section 5 we discuss the construction of so-called Chevalley systems; results obtained there will also be important for (C). In Section 6 we discuss the main computational problem for applying the Kostant–Sekiguchi correspondence, namely, to construct a complex Cayley triple in a K^c-orbit of homogeneous $\mathfrak{sl}_2$-triples; we give more details in Section 3.

In order to use our lists of representatives of nilpotent orbits also in other realizations of the Lie algebras (for instance in the split real forms, in their natural representation), we devise algorithms to construct isomorphisms between real simple Lie algebras. More precisely, in Section 7 we discuss the isomorphism problem for two real forms $\mathfrak{g}$ and $\mathfrak{g}'$ of a complex simple Lie algebra $\mathfrak{g}^c$. If $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and $\mathfrak{g}' = \mathfrak{k}' \oplus \mathfrak{p}'$ are given Cartan decompositions, then we describe an explicit construction of an isomorphism $\mathfrak{g} \to \mathfrak{g}'$, respecting the given Cartan decompositions, which fails if and only if such an isomorphism does not exist.

1.2. ***Related work.*** Đoković [1998; 1999; 2000] first used the Kostant–Sekiguchi correspondence to obtain representatives of nilpotent orbits for the real forms of Lie algebras of exceptional type. His methods vary somewhat from paper to paper. However, in all these publications the main idea is to start with a complex nilpotent orbit $\mathbb{O}^c \subset \mathfrak{g}^c$ meeting $\mathfrak{g}$ nontrivially. Then some real representatives of $\mathbb{O}^c$ in $\mathfrak{g}$ are computed. The Kostant–Sekiguchi correspondence is used to decide whether these real representatives lie in the same G-orbit or not. The process stops when enough elements lying in different G-orbits are found. This ad hoc approach has worked for the Lie algebras of exceptional type, but there is no guarantee that it will always yield representatives of all nilpotent orbits. Furthermore, it is rather tedious to apply and difficult to translate into a systematic approach suitable for a computer.

In our approach the problem is reduced to finding a complex Cayley triple in a carrier algebra. Most carrier algebras that occur are *principal* and for those we have an automatic procedure for finding the triple (see Section 6.2). However, some carrier algebras are not principal, and for those we translate the problem into a set of polynomial equations that has to be solved. For dealing with the latter problem we use a simple-minded systematic technique (see Section 6.3) which turned out to work well in all our examples, which include all Lie algebras of rank at most 8.

1.3. ***Computational remark.*** Our algorithms are implemented for the computer algebra system GAP [GAP 2012], as part of a package for doing computations with real Lie algebras, called CoReLG [CoReLG 2012]. The functions for obtaining the multiplication tables of the real simple Lie algebras in this package have been implemented by Paolo Faccin; see [Dietrich et al. 2013]. As an application, we created a database containing representatives of nilpotent orbits for all simple real forms of rank at most 8; this database will also be contained in the package CoReLG. As mentioned in the previous paragraph, we construct certain complex Cayley triples in carrier algebras. It is possible that isomorphic carrier algebras will turn up when dealing with different simple Lie algebras. To avoid dealing with the same problem twice, we have also built a database of nonprincipal carrier algebras, together with the Cayley triples that we found (see Section 6.3.1).

Our approach works uniformly for all simple real Lie algebras. However, our database is currently limited to the Lie algebras of ranks up to 8 for two reasons. Firstly, it includes all exceptional types. Secondly, in the SLA package, the current implementations of the algorithms for listing the nilpotent orbits of a θ-group are not very efficient when θ is an outer automorphism. This makes it currently difficult to go beyond rank 8 when the real form is defined relative to an outer involution.

There is the question of which base field to use for the computations. The Lie algebras with which we work are defined over $\mathbb{R}$ or $\mathbb{C}$. However, we want to perform exact computations, and the field $\mathbb{Q}$ is not suitable as we often need square roots of rational numbers. For this reason we work over the field $\mathbb{Q}^{\sqrt{}} = \mathbb{Q}(\{\sqrt{p} \mid p \text{ a prime}\})$. In the Appendix we indicate how the arithmetic of that field is implemented. Since we often work in the complex Lie algebra $\mathfrak{g}^c$ in order to obtain results in the real Lie algebra $\mathfrak{g}$, we also use the field $\mathbb{Q}^{\sqrt{}}(\imath)$ where $\imath = \sqrt{-1} \in \mathbb{C}$.

1.4. ***Notation.*** Throughout this paper we retain the previous notation and denote by θ the Cartan involution associated with the Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. By $\mathfrak{g}^c = \mathfrak{k}^c \oplus \mathfrak{p}^c$ we denote the complexification of $\mathfrak{g}$, and σ is the complex conjugation of $\mathfrak{g}^c$ with respect to $\mathfrak{g}$. By abuse of notation, we also denote by θ its extension to $\mathfrak{g}^c$. Let G^c be the adjoint group of $\mathfrak{g}^c$ and denote by G, K, and K^c the connected Lie subgroups of G^c with corresponding Lie algebras $\mathfrak{g}$, $\mathfrak{k}$, and $\mathfrak{k}^c$, respectively.

2. Constructing the Lie algebras

The aim of this section is to describe the construction of the real forms we consider. Our computational setup is as in [de Graaf 2000]; that is, in our algorithms we suppose the Lie algebras are given by multiplication tables, usually with respect to Chevalley bases. For the sake of completeness, we first recall the relevant definitions, and then construct certain bases of all real forms (up to isomorphism) of simple complex Lie algebras.

2.1. ***Canonical generators.*** Let $\mathfrak{g}^c$ be a complex semisimple Lie algebra with Cartan subalgebra $\mathfrak{h}^c$. Let Φ be the corresponding root system with basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$. Then $\mathfrak{g}^c$ has a Chevalley basis with respect to Φ; see [Humphreys 1978, Section 25.2]:

Definition 1. A basis $\{h_1, \dots, h_l, x_\alpha \mid \alpha \in \Phi\}$ of $\mathfrak{g}^c$ is a *Chevalley basis* if $\{h_1, \dots, h_l\}$ spans the Cartan subalgebra $\mathfrak{h}^c$ of $\mathfrak{g}^c$, and for all $\alpha, \beta \in \Phi$ the following hold:

- x_α spans the root space $\mathfrak{g}_\alpha = \{x \in \mathfrak{g}^c \mid [h_i, x] = \alpha(h_i)x \text{ for all } i\}$ corresponding to α,
- $[x_\alpha, x_{-\alpha}] = h_\alpha$, where h_α is the unique element in $[\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}]$ with $\alpha(h_\alpha) = 2$; in particular, $h_i = h_{\alpha_i}$ for all $i = 1, \dots, l$,
- $[x_\alpha, x_\beta] = N_{\alpha,\beta} x_{\alpha+\beta}$ if $\alpha + \beta \in \Phi$, where $N_{\alpha,\beta} \neq 0$ is an integer with $N_{\alpha,\beta} = -N_{-\alpha,-\beta}$,
- $[x_\alpha, x_\beta] = 0$ if $\alpha + \beta \notin \Phi$ and $\alpha \neq -\beta$.

Note that we see the roots in Φ as elements of the dual space $(\mathfrak{h}^c)^*$ via $[h, x_\alpha] = \alpha(h)x_\alpha$. For two roots $\alpha, \beta \in \Phi$, the corresponding Cartan integer now is $\langle \alpha, \beta \rangle = \alpha(h_\beta)$; the Cartan matrix of Φ defined by Δ is $(\langle \alpha_i, \alpha_j \rangle)_{ij}$; see [Humphreys 1978, pp. 39, 55]. In the sequel, we usually denote by $\{h_1, \dots, h_l, x_\alpha \mid \alpha \in \Phi\}$ a fixed Chevalley basis of $\mathfrak{g}^c$, and by $\{h_i, x_i, y_i \mid i = 1, \dots, l\}$ with $x_i = x_{\alpha_i}$ and $y_i = x_{-\alpha_i}$ the canonical generating set it contains:

Definition 2. A generating set $\{c_i, a_i, b_i \mid i = 1, \dots, l\}$ of $\mathfrak{g}^c$ is a *canonical generating set* if for all $i, j \in \{1, \dots, l\}$ the following hold:

- $c_i \in \mathfrak{h}^c$, $a_i \in \mathfrak{g}_{\alpha_i}$, and $b_i \in \mathfrak{g}_{-\alpha_i}$,
- $[c_i, c_j] = 0$ and $[a_i, b_j] = \delta_{ij} c_i$, where δ_{ij} is the Kronecker delta,
- $[c_i, a_j] = \langle \alpha_j, \alpha_i \rangle a_j$ and $[c_i, b_j] = -\langle \alpha_j, \alpha_i \rangle b_j$.

Let $\{c'_i, a'_i, b'_i \mid i = 1, \dots, l\}$ be a second canonical generating set of $\mathfrak{g}^c$, possibly relative to a different basis of simple roots Δ'. If Δ and Δ' define the same Cartan matrix, then there exists a unique automorphism of $\mathfrak{g}^c$ which maps (c_i, a_i, b_i) to (c'_i, a'_i, b'_i) for every $i = 1, \dots, l$; see [Jacobson 1962, Chapter IV, Theorem 3]. We freely use this property throughout the paper. Also, if Φ and l follow from the context, then we write $\{h_i, x_\alpha \mid \alpha, i\}$ and $\{h_i, x_i, y_i \mid i\}$ for the Chevalley basis and canonical generating set. We end this section with a proposition, which yields a straightforward algorithm to obtain a canonical generating set. For its proof, as well as the algorithm, we refer to [de Graaf 2000, Section 5.11].

Proposition 3. *For $i = 1, \dots, l$ let $a_i \in \mathfrak{g}_{\alpha_i}$ and $b_i \in \mathfrak{g}_{-\alpha_i}$, and write $c_i = [a_i, b_i]$. If $[c_i, a_i] = 2a_i$ for all i, then $\{c_i, a_i, b_i \mid i\}$ is a canonical generating set of $\mathfrak{g}^c$.*

2.2. *Real forms.* We now turn to the construction of the real forms of a complex semisimple Lie algebra $\mathfrak{g}^c$; without loss of generality, we may assume that $\mathfrak{g}^c$ is simple. We continue to use the notation of Section 2.1; that is, $\mathfrak{h}^c$ is a Cartan subalgebra of $\mathfrak{g}^c$ with root system Φ, having basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$. Let $\{h_i, x_\alpha \mid i, \alpha\}$ and $\{h_i, x_i, y_i \mid i\}$ be a corresponding Chevalley basis and canonical generating set. Recall that a real Lie algebra $\mathfrak{g}'$ is a real form of $\mathfrak{g}^c$ if $\mathfrak{g}^c = \mathfrak{g}' \oplus \imath\mathfrak{g}'$ as real vectorspaces.

2.2.1. *Real forms defined by involutions.* It is proved in [Onishchik 2004, Theorem 3.1] that the real subalgebra $\mathfrak{u}$ of $\mathfrak{g}^c$ defined as

$$\mathfrak{u} = \mathrm{Span}_{\mathbb{R}}(\{\imath h_1, \dots, \imath h_l, x_\alpha - x_{-\alpha}, \imath(x_\alpha + x_{-\alpha}) \mid \alpha \in \Phi^+\})$$

is a (compact) real form of $\mathfrak{g}^c$. Let τ be the corresponding real structure; that is, $\tau\colon \mathfrak{g}^c \to \mathfrak{g}^c$ is the complex conjugation of $\mathfrak{g}^c = \mathfrak{u} \oplus \imath\mathfrak{u}$ with respect to $\mathfrak{u}$. This implies that $\tau(x_\alpha) = -x_{-\alpha}$ for all $\alpha \in \Phi$; in particular, for all $i = 1, \dots, l$ we have

$$\tau(h_i) = -h_i, \quad \tau(x_i) = -y_i, \quad \text{and} \quad \tau(y_i) = -x_i.$$

It follows from [Onishchik 2004, Theorem 3.2] that, up to isomorphism, every real form of $\mathfrak{g}^c$ is constructed as follows: Let θ be an involutive automorphism of $\mathfrak{g}^c$ commuting with τ. Then $\mathfrak{u} = \mathfrak{u}_0 \oplus \mathfrak{u}_1$, where $\mathfrak{u}_i$ is the eigenspace of θ in $\mathfrak{u}$ with eigenvalue $(-1)^i$, and the real form defined by $\mathfrak{u}$ and θ is

$$\mathfrak{g} = \mathfrak{g}(\theta, \mathfrak{u}) = \mathfrak{k} \oplus \mathfrak{p} \quad \text{with } \mathfrak{k} = \mathfrak{u}_0 \text{ and } \mathfrak{p} = \imath\mathfrak{u}_1.$$

This decomposition of $\mathfrak{g}$ is a Cartan decomposition whose Cartan involution is the restriction of θ to $\mathfrak{g}$; see [Onishchik 2004, Section 5]. We denote by $\sigma\colon \mathfrak{g}^c \to \mathfrak{g}^c$ the complex conjugation of $\mathfrak{g}^c = \mathfrak{g} \oplus \imath\mathfrak{g}$ relative to $\mathfrak{g}$.

Two such real forms $\mathfrak{g}(\theta, \mathfrak{u})$ and $\mathfrak{g}(\theta', \mathfrak{u})$ are isomorphic if and only if θ and θ' are conjugate in $\mathrm{Aut}(\mathfrak{g}^c)$. The finite order automorphisms of $\mathfrak{g}^c$ are, up to conjugacy, classified by so-called Kac diagrams; see [Vinberg et al. 1990, Section 3.3.7] or [Helgason 1978, Section X.5]. By running through these diagrams we can efficiently construct all involutions of $\mathfrak{g}^c$ up to conjugacy, and hence all real forms of $\mathfrak{g}^c$ up to isomorphism.

2.2.2. *Real forms of inner type.* Let θ be an inner involutive automorphism of $\mathfrak{g}^c$. Up to conjugacy, θ maps (h_i, x_i, y_i) to $(h_i, \lambda_i x_i, \lambda_i^{-1} y_i)$ with $\lambda_i \in \{\pm 1\}$ for all i. Clearly, such an automorphism commutes with τ, and bases of $\mathfrak{k}$ and $\mathfrak{p}$ in $\mathfrak{g} = \mathfrak{g}(\theta, \mathfrak{u}) = \mathfrak{k} \oplus \mathfrak{p}$ are

$$\mathcal{K} = \{x_\alpha - x_{-\alpha}, \imath(x_\alpha + x_{-\alpha}) \mid \alpha \in \Phi^+ \text{ with } \theta(x_\alpha) = x_\alpha\} \cup \{\imath h_1, \dots, \imath h_l\},$$
$$\mathcal{P} = \{\imath(x_\alpha - x_{-\alpha}), x_\alpha + x_{-\alpha} \mid \alpha \in \Phi^+ \text{ with } \theta(x_\alpha) = -x_\alpha\}.$$

We define $\mathfrak{g}$ by the multiplication table constructed via the basis $\mathcal{B} = \mathcal{K} \cup \mathcal{P}$. We note that $\{\imath h_1, \dots, \imath h_l\}$ spans a Cartan subalgebra $\mathfrak{h}_0$ of $\mathfrak{k}$, which is also a Cartan subalgebra of $\mathfrak{g}$. It is straightforward to see that $\sigma(x_\alpha) = -x_{-\alpha}$ if $\theta(x_\alpha) = x_\alpha$, and $\sigma(x_\alpha) = x_{-\alpha}$ if $\theta(x_\alpha) = -x_\alpha$.

2.2.3. *Real forms of outer type.* Let θ be an outer involutive automorphism of $\mathfrak{g}^c$. Up to conjugacy, $\theta = \varphi \circ \chi$, where φ is an involutive diagram automorphism and χ is an inner involutive automorphism; clearly, χ and φ commute. As above, we can assume that χ maps (h_i, x_i, y_i) to $(h_i, \lambda_i x_i, \lambda_i^{-1} y_i)$ with $\lambda_i \in \{\pm 1\}$ for all i. Further, φ maps (h_i, x_i, y_i) to $(h_{\pi(i)}, x_{\pi(i)}, y_{\pi(i)})$ for all i, where π is an involutive permutation of $\{1, \dots, l\}$ with $(\langle \alpha_j, \alpha_i \rangle)_{ij} = (\langle \alpha_{\pi(j)}, \alpha_{\pi(i)} \rangle)_{ij}$; note that $\{h_{\pi(i)}, x_{\pi(i)}, y_{\pi(i)} \mid i\}$ also is a canonical generating set, and $\lambda_{\pi(i)} = \lambda_i$ since χ and φ commute. The permutation π induces an automorphism of Φ, which we also denote by φ; that is, $\varphi(\alpha_i) = \alpha_{\pi(i)}$.

Let $\mathfrak{g} = \mathfrak{g}(\theta, \mathfrak{u}) = \mathfrak{k} \oplus \mathfrak{p}$. We now determine bases $\mathcal{K}$ and $\mathcal{P}$ for $\mathfrak{k}$ and $\mathfrak{p}$, and, as before, define $\mathfrak{g}$ by the multiplication table constructed via $\mathcal{B} = \mathcal{K} \cup \mathcal{P}$. Since $\mathfrak{g}^c$ admits outer automorphisms, it is of type A, D, or E_6, in particular, simply laced; see [Onishchik 2004, Table 1]. We first consider the case where $\mathfrak{g}^c$ is not of type A_l with l even. In this case there exists a Chevalley basis $\{h_i, \hat{x}_\alpha \mid i, \alpha\}$ such that, when defining $\widehat{N}_{\alpha,\beta}$ by $[\hat{x}_\alpha, \hat{x}_\beta] = \widehat{N}_{\alpha,\beta} \hat{x}_{\alpha+\beta}$, we have $\widehat{N}_{\varphi(\alpha),\varphi(\beta)} = \widehat{N}_{\alpha,\beta}$ for all $\alpha, \beta \in \Phi$; see [Kac 1990, Section 7.9] or [de Graaf 2000, Section 5.15]. (This result does not hold if $\mathfrak{g}^c$ is of type A_l with l even; we consider this case in the following section.) Induction on the height of α now proves that $\varphi(\hat{x}_\alpha) = \hat{x}_{\varphi(\alpha)}$ for all $\alpha \in \Phi$. Thus, if $\varphi(\alpha) = \alpha$, then φ acts as the identity on $\mathfrak{g}_\alpha$, which implies that $\varphi(x_\alpha) = x_\alpha$.

For $\alpha \in \Phi$ define

$$v_\alpha = x_\alpha - \varphi(x_\alpha) \quad \text{and} \quad u_\alpha = \begin{cases} x_\alpha & \text{if } \varphi(\alpha) = \alpha, \\ x_\alpha + \varphi(x_\alpha) & \text{if } \varphi(\alpha) \neq \alpha. \end{cases}$$

Let Ψ^+ be the set consisting of all $\alpha \in \Phi^+$ such that $\varphi(\alpha) = \alpha$, along with one element of each pair $(\alpha, \varphi(\alpha))$ where $\varphi(\alpha) \neq \alpha$. Let $\mathcal{J} \subseteq \{1, \dots, l\}$ be a set of representatives of the π-orbits on $\{1, \dots, l\}$ of length 2. Now we define $\mathcal{K}$ as the union of the three sets

$$\begin{gathered} \mathcal{H}_0 = \{\imath h_i \mid i = 1, \dots, l \text{ with } \pi(i) = i\} \cup \{\imath(h_i + h_{\pi(i)}) \mid i \in \mathcal{J}\}, \\ \{u_\alpha - u_{-\alpha}, \imath(u_\alpha + u_{-\alpha}) \mid \alpha \in \Psi^+ \text{ with } \chi(x_\alpha) = x_\alpha\}, \quad \text{and} \\ \{v_\alpha - v_{-\alpha}, \imath(v_\alpha + v_{-\alpha}) \mid \alpha \in \Psi^+ \text{ with } \chi(x_\alpha) = -x_\alpha \text{ and } \varphi(\alpha) \neq \alpha\}; \end{gathered}$$

note that, if $\varphi(\alpha) = \alpha$ and $\chi(x_\alpha) = x_\alpha$, then $\theta(x_\alpha) = x_\alpha$, whence $u_\alpha - u_{-\alpha}$ and $\imath(u_\alpha + u_{-\alpha})$ lie in $\mathfrak{k}$. We define $\mathcal{P}$ to be the union of

$$\{h_i - h_{\pi(i)} \mid i \in \mathcal{I}\}, \quad \{\imath(u_\alpha - u_{-\alpha}), u_\alpha + u_{-\alpha} \mid \alpha \in \Psi^+ \text{ with } \chi(x_\alpha) = -x_\alpha\},$$
$$\text{and} \quad \{\imath(v_\alpha - v_{-\alpha}), v_\alpha + v_{-\alpha} \mid \alpha \in \Psi^+ \text{ with } \chi(x_\alpha) = x_\alpha \text{ and } \varphi(\alpha) \neq \alpha\}.$$

It is straightforward to verify that $\mathcal{K}$ and $\mathcal{P}$ are bases of $\mathfrak{k}$ and $\mathfrak{p}$. Further, $\mathcal{H}_0$ spans a Cartan subalgebra $\mathfrak{h}_0$ of $\mathfrak{k}$, but this time the complexification $\mathfrak{h}_0^c$ is not a Cartan subalgebra of $\mathfrak{g}^c$. We have $\sigma(u_\alpha) = -u_{-\alpha}$ and $\sigma(v_\alpha) = v_{-\alpha}$ if $\chi(x_\alpha) = x_\alpha$, and $\sigma(u_\alpha) = u_{-\alpha}$ and $\sigma(v_\alpha) = -v_{-\alpha}$ otherwise.

Remark 4. We consider the weight space decomposition of $\mathfrak{g}^c$ with respect to $\mathfrak{h}_0^c$ and show that each weight space in $\mathfrak{k}^c$ and $\mathfrak{p}^c$ (corresponding to a nonzero weight) is 1-dimensional. Note that φ fixes $\mathfrak{h}_0^c$ pointwise and, if $h \in \mathfrak{h}_0^c$, then $\alpha_i(h)\varphi(x_i) = \varphi([h, x_i]) = [h, \varphi(x_i)] = \varphi(\alpha_i)(h)\varphi(x_i)$ for all i, implying that $\alpha(h) = \varphi(\alpha)(h)$ for all $\alpha \in \Phi$. Now write $\Psi = \Psi^+ \cup (-\Psi^+)$ and define $\Psi_0 = \{\alpha|_{\mathfrak{h}_0^c} \mid \alpha \in \Psi\}$ as a subset of $(\mathfrak{h}_0^c)^\star$. Consider the simple Lie algebra $\mathfrak{l} = \{x \in \mathfrak{g}^c \mid \varphi(x) = x\}$; see [Kac 1990, Section 7.9]. It is easy to verify that for all $\alpha \in \Psi$ we have $u_\alpha \in \mathfrak{l}$, and, further, if $h \in \mathfrak{h}_0^c$, then $[h, u_\alpha] = \alpha(h)u_\alpha$. Since $\mathfrak{l}$ is simple, this proves that the root space decomposition of $\mathfrak{l}$ with respect to $\mathfrak{h}_0^c$ is $\mathfrak{l} = \mathfrak{h}_0^c \oplus \bigoplus_{\alpha\in\Psi} \mathfrak{l}_\alpha$, where $\mathfrak{l}_\alpha$ is spanned by u_α; in particular, $|\Psi| = |\Psi_0|$. So we have the $\mathfrak{h}_0^c$-weight space decompositions

$$\mathfrak{k}^c = \mathfrak{h}_0^c \oplus \bigoplus_{\alpha\in\Psi_0} \mathfrak{k}_\alpha^c, \quad \mathfrak{p}^c = \mathrm{Span}_{\mathbb{C}}(\{h_i - h_{\pi(i)} \mid i = 1, \ldots, l \text{ with } \pi(i) \neq i\}) \oplus \bigoplus_{\alpha\in\Psi_0} \mathfrak{p}_\alpha^c,$$

where each $\mathfrak{k}_\alpha^c = \{x \in \mathfrak{k}^c \mid [h, x] = \alpha(h)x \text{ for all } h \in \mathfrak{h}_0^c\}$ (and similarly $\mathfrak{p}_\alpha^c$) has dimension at most one. More precisely, if $\alpha \in \Phi$ and $\bar\alpha = \alpha|_{\mathfrak{h}_0^c}$, then the following hold:

- if $\varphi(\alpha) \neq \alpha$ and $\chi(x_\alpha) = x_\alpha$, then $\mathfrak{k}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(u_\alpha)$ and $\mathfrak{p}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(v_\alpha)$,
- if $\varphi(\alpha) \neq \alpha$ and $\chi(x_\alpha) \neq x_\alpha$, then $\mathfrak{k}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(v_\alpha)$ and $\mathfrak{p}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(u_\alpha)$,
- if $\varphi(\alpha) = \alpha$ and $\chi(x_\alpha) = x_\alpha$, then $\mathfrak{k}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(u_\alpha)$ and $\mathfrak{p}_{\bar\alpha}^c = 0$,
- if $\varphi(\alpha) = \alpha$ and $\chi(x_\alpha) \neq x_\alpha$, then $\mathfrak{k}_{\bar\alpha}^c = 0$ and $\mathfrak{p}_{\bar\alpha}^c = \mathrm{Span}_{\mathbb{C}}(u_\alpha)$.

2.2.4. *Real forms of A_l, l even, of outer type.* It remains to consider the case where $\mathfrak{g}^c$ is of type A_l with $l = 2m$ even; we use the notation of the previous section. Up to conjugacy, we can assume that χ is the identity; thus $\theta = \varphi$ is the unique diagram automorphism. (This follows directly from looking at the possible Kac diagrams of an outer involution in this case.) Since $\mathfrak{g}^c$ is simply laced, $N_{\alpha,\beta} = \pm 1$ for all $\alpha, \beta \in \Phi$ with $\alpha + \beta \in \Phi$, and induction on the height of α proves that $\varphi(x_\alpha) = \pm x_{\varphi(\alpha)}$ for all $\alpha \in \Phi$. By [Kac 1990, Section 7.10], there is a Chevalley basis of $\mathfrak{g}^c$ such that $\varphi(x_\alpha) = -x_\alpha$ for all $\alpha \in \Phi$ with $\varphi(\alpha) = \alpha$. Let $\mathfrak{g} = \mathfrak{g}(\theta, \mathfrak{u}) = \mathfrak{k} \oplus \mathfrak{p}$. A basis of $\mathfrak{k}$ is the set $\mathcal{K}$ defined as the union of

$$\mathcal{H}_0 = \{\imath(h_i + h_{\pi(i)}) \mid i \in \mathcal{I}\} \quad \text{and} \quad \{u_\alpha - u_{-\alpha}, \imath(u_\alpha + u_{-\alpha}) \mid \alpha \in \Psi^+ \text{ with } \varphi(\alpha) \neq \alpha\};$$

note that $|\mathcal{I}| = m$ since π acts fixed-point freely on $\{1, \dots, 2m\}$. A basis $\mathcal{P}$ of $\mathfrak{p}$ is the union of

$$\{h_i - h_{\pi(i)} \mid i \in \mathcal{I}\}, \quad \{\imath(u_\alpha - u_{-\alpha}), u_\alpha + u_{-\alpha} \mid \alpha \in \Psi^+ \text{ with } \varphi(\alpha) = \alpha\},$$
$$\text{and} \quad \{\imath(v_\alpha - v_{-\alpha}), v_\alpha + v_{-\alpha} \mid \alpha \in \Psi^+ \text{ with } \varphi(\alpha) \neq \alpha\}.$$

Again, $\mathcal{H}_0$ spans a Cartan subalgebra $\mathfrak{h}_0$ of $\mathfrak{k}$, and $\mathfrak{h}_0^c$ is not a Cartan subalgebra of $\mathfrak{g}^c$. We obtain weight space decompositions of $\mathfrak{k}^c$ and $\mathfrak{p}^c$ as in Section 2.2.3. All nonzero weight spaces with respect to $\mathfrak{h}_0^c$ are 1-dimensional and spanned by a u_α or v_α. Again, $\sigma(u_\alpha) = u_{-\alpha}$ and $\sigma(v_\alpha) = v_{-\alpha}$.

3. Kostant–Sekiguchi correspondence

Let $\mathfrak{g}^c$ be a complex semisimple Lie algebra with real form $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, associated complex conjugation σ, and Cartan involution θ. Recall that we denote by G and K^c the connected Lie subgroups of the adjoint group G^c of $\mathfrak{g}^c$ with Lie algebras $\mathfrak{g}$ and $\mathfrak{k}^c$, respectively. The Kostant–Sekiguchi correspondence is a one-to-one correspondence between the nilpotent G-orbits in $\mathfrak{g}$ and the nilpotent K^c-orbits in $\mathfrak{p}^c$. The latter orbits can be constructed using the algorithms in [de Graaf 2011; 2012]; note that K^c, together with its action on $\mathfrak{p}^c$, is a so-called θ-group. An implementation of the Kostant–Sekiguchi correspondence would therefore allow us to construct the nilpotent G-orbits in $\mathfrak{g}$.

We now describe this correspondence in more detail. Its proof has been completed independently in [Đoković 1987] and [Sekiguchi 1987]; here we follow the description in the first of these papers, and refer the reader to it for an historical account and (references to) proofs. First, we need some notation. The following definitions are as in [Đoković 1987] with the exception that our "f" has been replaced by "$-f$". An $\mathfrak{sl}_2$-*triple* in $\mathfrak{g}$ (or $\mathfrak{g}^c$) is a triple (f, h, e) of elements in $\mathfrak{g}$ (or $\mathfrak{g}^c$) such that $[h, e] = 2e$, $[h, f] = -2f$, and $[e, f] = h$. The *characteristic* (element) of this triple is h.

Definition 5. (a) An $\mathfrak{sl}_2$-triple (f, h, e) in $\mathfrak{g}^c$ is *homogeneous* if $e, f \in \mathfrak{p}^c$ and $h \in \mathfrak{k}^c$.

(b) An $\mathfrak{sl}_2$-triple (f, h, e) in $\mathfrak{g}^c$ is a *complex Cayley triple* if it is homogeneous and $\sigma(e) = f$.

(c) An $\mathfrak{sl}_2$-triple (f, h, e) in $\mathfrak{g}$ is a *real Cayley triple* if $\theta(e) = -f$.

The Kostant–Sekiguchi correspondence can now be stated as in Figure 1, where all maps are bijections. We provide some details. Every nonzero nilpotent $e \in \mathfrak{p}^c$ lies in some homogeneous $\mathfrak{sl}_2$-triple (f, h, e) of $\mathfrak{g}^c$, and the projection $(f, h, e) \mapsto e$ induces a bijection between the K^c-orbits of homogeneous $\mathfrak{sl}_2$-triples in $\mathfrak{g}^c$ and the K^c-orbits of nonzero nilpotent elements in $\mathfrak{p}^c$; let φ_1 denote the inverse of this bijection. Every K^c-orbit of homogeneous $\mathfrak{sl}_2$-triples in $\mathfrak{g}^c$ contains a complex Cayley triple and, conversely, every K-orbit of complex Cayley triples in $\mathfrak{g}^c$ is contained in a unique K^c-orbit of homogeneous $\mathfrak{sl}_2$-triples in $\mathfrak{g}^c$. Thus, inclusion gives rise to a bijection between the K-orbits of complex Cayley triples and the K^c-orbits of homogeneous $\mathfrak{sl}_2$-triples in $\mathfrak{g}^c$; again, let φ_2 denote the inverse of this bijection. Let (f, h, e) be a real Cayley triple. Then its *Cayley transform* is the triple

$$\left(\tfrac{1}{2}(\imath e + \imath f + h), \imath(e - f), \tfrac{1}{2}(-\imath e - \imath f + h)\right),$$

which is a complex Cayley triple. The inverse Cayley transform maps a complex Cayley triple (f, h, e) to the real Cayley triple

$$\left(\tfrac{1}{2}\imath(e - f + h), e + f, \tfrac{1}{2}\imath(e - f - h)\right).$$

Taking inverse Cayley transforms induces a bijection φ_3 between the K-orbits of complex Cayley triples in $\mathfrak{g}^c$ and the K-orbits of real Cayley triples in $\mathfrak{g}$. The projection $(f, h, e) \mapsto e$ yields a bijection φ_4 between these K-orbits of real Cayley triples and the G-orbits of nonzero nilpotent elements in $\mathfrak{g}$. In conclusion, the Kostant–Sekiguchi correspondence states that $\varphi_4 \circ \varphi_3 \circ \varphi_2 \circ \varphi_1$ is a bijection.

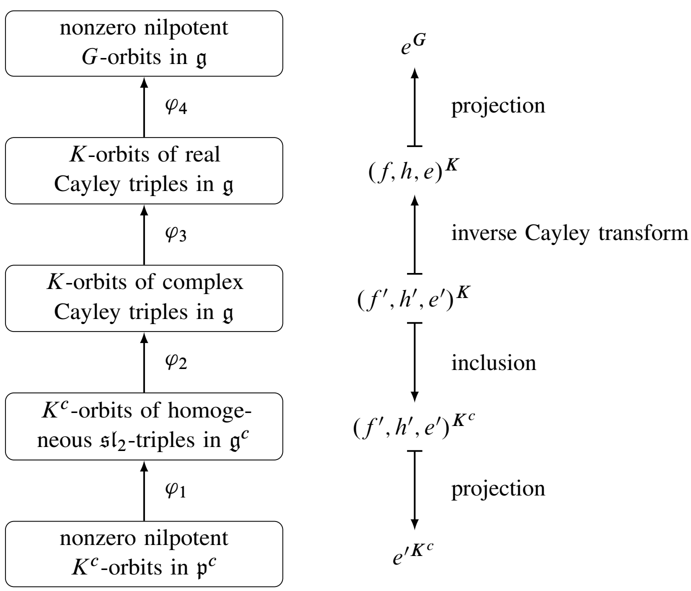


Figure 1. Kostant–Sekiguchi correspondence.

Using the algorithms of [de Graaf 2011; 2012], we can compute all K^c-orbits of homogeneous $\mathfrak{sl}_2$-triples in $\mathfrak{g}^c$, which also gives us the bijection φ_1. A realization of the map $\varphi_4 \circ \varphi_3$ is straightforward. Thus, computationally, it remains to realize φ_2, that is:

Main Problem. *Find a complex Cayley triple in a K^c-orbit of homogeneous $\mathfrak{sl}_2$-triples.*

We discuss our approach to this problem in Section 6. For this purpose, we require some preliminary results; the subsequent sections therefore introduce carrier algebras and Chevalley systems.

4. Carrier algebras

We briefly review the theory of carrier algebras as developed in [Vinberg 1979]. In general, carrier algebras are connected to $\mathbb{Z}_m$-graded Lie algebras. Since we exclusively deal with $\mathbb{Z}_2$-gradings (coming from Cartan decompositions), we only consider this case here.

Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be as in Section 2.2, and consider the $\mathbb{Z}_2$-grading $\mathfrak{g}^c = \mathfrak{g}_0^c \oplus \mathfrak{g}_1^c$, where $\mathfrak{g}_0^c = \mathfrak{k}^c$ and $\mathfrak{g}_1^c = \mathfrak{p}^c$. Recall that $G_0 = K^c$ is the connected Lie subgroup of G^c with Lie algebra $\mathfrak{g}_0^c$. Let $e \in \mathfrak{g}_1^c$ be nilpotent, and consider the normalizer $N_0(e) = \{x \in \mathfrak{g}_0^c \mid [x, e] = \lambda e \text{ for some } \lambda \in \mathbb{C}\}$. Let $\mathfrak{t}$ be a maximal torus of $N_0(e)$, that is, a maximal abelian subalgebra consisting of semisimple elements, and let $\mu \in \mathfrak{t}^*$ be defined by $[t, e] = \mu(t)e$ for $t \in \mathfrak{t}$. Let $\mathfrak{a}^c = \bigoplus_{k\in\mathbb{Z}} \mathfrak{a}_k$ be the $\mathbb{Z}$-graded Lie algebra defined by

$$\mathfrak{a}_k = \{x \in \mathfrak{g}^c_{k \bmod 2} \mid [t, x] = k\mu(t)x \text{ for all } t \in \mathfrak{t}\}.$$

The *carrier algebra* of e is the commutator algebra of $\mathfrak{a}^c$ with the inherited $\mathbb{Z}$-grading; that is,

$$\mathfrak{s}^c = \mathfrak{s}(e, \mathfrak{t}) = \bigoplus_{k\in\mathbb{Z}} \mathfrak{s}_k = [\mathfrak{a}^c, \mathfrak{a}^c].$$

As shown in [Vinberg 1979], it has the following properties:

- $\mathfrak{s}^c$ is semisimple with $\dim \mathfrak{s}_0 = \dim \mathfrak{s}_1$,
- $\mathfrak{s}^c$ is not a proper subalgebra of a $\mathbb{Z}$-graded semisimple subalgebra of $\mathfrak{g}^c$ of the same rank,
- $\mathfrak{s}_k \subseteq \mathfrak{k}^c$ if k is even, and $\mathfrak{s}_k \subseteq \mathfrak{p}^c$ otherwise,
- $\mathfrak{s}^c$ is normalized by a Cartan subalgebra of $\mathfrak{g}_0^c$.

Moreover, $e \in \mathfrak{s}_1$ is in *general position*; that is, $[\mathfrak{s}_0, e] = \mathfrak{s}_1$; every element in $\mathfrak{s}_1$ in general position is G_0-conjugate to e. If (f, h, e) is a homogeneous $\mathfrak{sl}_2$-triple in $\mathfrak{s}^c$, that is, $h \in \mathfrak{s}_0$ and $f \in \mathfrak{s}_{-1}$, then $h/2$ is the unique *defining element* of $\mathfrak{s}^c$; that is, for all k,

$$\mathfrak{s}_k = \left\{x \in \mathfrak{s}^c \;\middle|\; \left[\tfrac{h}{2}, x\right] = kx\right\}.$$

Since all maximal tori of $N_0(e)$ are conjugate, this yields a bijection between the nilpotent G_0-orbits in $\mathfrak{g}_1^c$ and the G_0-conjugacy classes of $\mathbb{Z}$-graded subalgebras with the above properties. This bijection can be used for an algorithm to list the nilpotent G_0-orbits in $\mathfrak{g}_1^c$; see [de Graaf 2011; Littelmann 1996].

Remark 6. Suppose (f,h,e) is a homogeneous $\mathfrak{sl}_2$-triple in $\mathfrak{g}^c$ and let $\mathfrak{s}^c = \mathfrak{s}^c(e,\mathfrak{t})$ be a carrier algebra. Since $h \in N_0(e)$, we can choose a torus containing h; thus $h \in \mathfrak{s}^c$. By the Jacobson–Morozov theorem (see [Knapp 2002, Theorem X.10.3]) there is $f' \in \mathfrak{s}^c$ such that (f',h,e) is an $\mathfrak{sl}_2$-triple in $\mathfrak{s}^c$, hence also in $\mathfrak{g}^c$. The same theorem shows $f = f'$; thus we can assume that $\mathfrak{s}^c$ contains f,h,e. We also call such an $\mathfrak{s}^c$ a carrier algebra of the triple (f,h,e); note that $h/2$ is its defining element.

Let $\mathfrak{h}_0^c$ be a fixed Cartan subalgebra of $\mathfrak{g}_0^c$. A carrier algebra $\mathfrak{s}^c$ is *standard* if it is normalized by $\mathfrak{h}_0^c$, and $[\mathfrak{h}_0^c, \mathfrak{s}_k] \subseteq \mathfrak{s}_k$ for all k. Since the Cartan subalgebras of $\mathfrak{g}_0^c$ are G_0-conjugate, every nilpotent G_0-orbit in $\mathfrak{g}_1^c$ corresponds to at least one standard carrier algebra $\mathfrak{s}^c$. Now, as shown in [Vinberg 1979, p. 23], the defining element of $\mathfrak{s}^c$ lies in $\mathfrak{h}_0^c$, and $\mathfrak{h}_0^c \cap \mathfrak{s}_0$ is a Cartan subalgebra of $\mathfrak{s}^c$; let $\Phi_{\mathfrak{s}^c}$ be the corresponding root system of $\mathfrak{s}^c$. Clearly, the homogeneous components $\mathfrak{s}_k$ are sums of root spaces, which allows us to define the degree of $\alpha \in \Phi_{\mathfrak{s}^c}$ as $\deg(\alpha) = k$ if $\mathfrak{s}_\alpha \subseteq \mathfrak{s}_k$. If $\Delta_{\mathfrak{s}^c}$ is a basis of simple roots such that $\deg(\alpha) \geq 0$ for all $\alpha \in \Delta_{\mathfrak{s}^c}$, then in fact $\deg(\alpha) \in \{0,1\}$; see [Vinberg 1979, p. 29]. If $\deg(\alpha) = 1$ for all $\alpha \in \Delta_{\mathfrak{s}^c}$, then $\mathfrak{s}^c$ is *principal*. In that case $\mathfrak{s}_0 = \mathfrak{s}_0 \cap \mathfrak{h}_0^c$ is a torus (in particular, abelian) and $\mathfrak{s}_1$ is spanned by $\mathfrak{s}_\alpha$ with $\alpha \in \Delta_{\mathfrak{s}^c}$.

5. Chevalley systems

Again, we consider $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with Cartan involution θ and complexification $\mathfrak{g}^c$ with complex conjugation σ. We suppose that $\mathfrak{h}^c = \mathfrak{h} \oplus \imath\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}^c$, where $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$ with $\mathfrak{h} = (\mathfrak{h} \cap \mathfrak{k}) \oplus (\mathfrak{h} \cap \mathfrak{p})$; write $\mathfrak{h}_0 = \mathfrak{h} \cap \mathfrak{k}$ and $\mathfrak{a} = \mathfrak{h} \cap \mathfrak{p}$. In this situation, $\mathfrak{h}$ is called *standard* and an adjoint $\mathrm{ad}(h)$ has only purely imaginary eigenvalues if $h \in \mathfrak{h}_0$, and only real ones if $h \in \mathfrak{a}$ (see [Rothschild 1972, p. 405] or [Onishchik 2004, Proposition 5.1(ii)]). This condition on $\mathfrak{h}$ is not a serious restriction since every Cartan subalgebra of $\mathfrak{g}$ is conjugate to a standard Cartan subalgebra; see [Rothschild 1972, Proposition 1.3]. We let Φ be the root system of $\mathfrak{g}^c$ with respect to $\mathfrak{h}^c$, with basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$. Further we assume that we have a canonical generating set $\{h_i, x_i, y_i \mid i\}$ such that for every i either $\theta(x_i) = \lambda_i x_i$, with $\lambda_i = \pm 1$, or $\theta(x_i) = x_j$ with $i \neq j$. We extend these canonical generators to a Chevalley basis $\{h_i, x_\alpha \mid i, \alpha\}$. If $\alpha \in \Phi$ is a root, then $\beta = \alpha \circ \theta$ is a root with $\theta(x_\alpha) \in \mathfrak{g}_\beta$ and $\theta(h_\alpha) = h_\beta$; hence, by our assumptions, Δ is stable under $\alpha \mapsto \alpha \circ \theta$. Let π be the permutation of $\{1, \dots, l\}$ defined by $\alpha_i \circ \theta = \alpha_{\pi(i)}$. We retain this notation throughout this section.

Lemma 7. *For every $\alpha \in \Phi$ the following hold.*

(a) $\theta(x_\alpha) = \lambda_\alpha x_{\alpha\circ\theta}$ *for some* $\lambda_\alpha \in \{\pm 1\}$, *and* $\lambda_\alpha = \lambda_\alpha^{-1} = \lambda_{-\alpha} = \lambda_{\alpha\circ\theta}$.

(b) $\sigma(x_\alpha) = r_\alpha x_{-\alpha\circ\theta}$ *for some* $r_\alpha \in \mathbb{R}$, *and* $r_\alpha^{-1} = r_{-\alpha} = r_{-\alpha\circ\theta}$.

(c) $\theta(h_\alpha) = h_{\alpha\circ\theta}$ *and* $\sigma(h_\alpha) = h_{-\alpha\circ\theta} = -h_{\alpha\circ\theta}$.

Proof. (a) We already know that $\lambda_{\alpha_i} = \lambda_i \in \{\pm 1\}$ and now use induction: If $\lambda_\alpha, \lambda_\beta \in \{\pm 1\}$, then $N_{\alpha,\beta}\theta(x_{\alpha+\beta}) = \lambda_\alpha\lambda_\beta N_{\alpha\circ\theta,\beta\circ\theta} x_{(\alpha+\beta)\circ\theta}$, and hence $\lambda_{\alpha+\beta} = N_{\alpha,\beta}^{-1} N_{\alpha\circ\theta,\beta\circ\theta}\lambda_\alpha\lambda_\beta \in \{\pm 1\}$ since $|N_{\alpha,\beta}| = |N_{\alpha\circ\theta,\beta\circ\theta}|$; the latter holds since $|N_{\alpha,\beta}| = r+1$ where r is the largest integer with $\alpha - r\beta \in \Phi$; see [Humphreys 1978, Theorem 25.2]. Since θ is an involution, $\lambda_{\alpha\circ\theta} = \lambda_\alpha^{-1}$, and $\lambda_{-\alpha} = \lambda_\alpha^{-1}$ follows from $h_{\alpha\circ\theta} = \theta(h_\alpha) = \theta([x_\alpha, x_{-\alpha}]) = \lambda_\alpha\lambda_{-\alpha} h_{\alpha\circ\theta}$.

(b+c) Let $\{k_1, \dots, k_l\}$ be a basis of $\mathfrak{h}^c = \mathfrak{h}_0^c \oplus \mathfrak{a}^c$ such that $\{k_1, \dots, k_m\}$ and $\{k_{m+1}, \dots, k_l\}$ form bases of $\mathfrak{a}$ and $\mathfrak{h}_0$, respectively. If $i \in \{1, \dots, m\}$, then $[k_i, \sigma(x_\alpha)] = \sigma([k_i, x_\alpha]) = \sigma(\alpha(k_i)x_\alpha) = \alpha(k_i)\sigma(x_\alpha)$ as $\alpha(k_i)$ is real. Analogously, if $i \in \{m+1, \dots, l\}$, then $\alpha(k_i)$ is purely imaginary and $[k_i, \sigma(x_\alpha)] = -\alpha(k_i)\sigma(x_\alpha)$. Hence $\sigma(x_\alpha) = r_\alpha x_{-\alpha\circ\theta}$ with $r_\alpha \in \mathbb{C}$. Note that $h_{-\beta} = -h_\beta$ for all $\beta \in \Phi$. Now it follows from $[\sigma(h_\alpha), \sigma(x_\alpha)] = 2\sigma(x_\alpha)$ that $-\alpha \circ \theta(\sigma(h_\alpha)) = 2$; hence $\sigma(h_\alpha) \in [\mathfrak{g}_{-\alpha\circ\theta}, \mathfrak{g}_{\alpha\circ\theta}]$ implies that

$$\sigma(h_\alpha) = h_{-\alpha\circ\theta} = -h_{\alpha\circ\theta}.$$

Since $\sigma(h_\alpha) = r_\alpha r_{-\alpha}[x_{-\alpha\circ\theta}, x_{\alpha\circ\theta}] = -r_\alpha r_{-\alpha} h_{\alpha\circ\theta}$, this already proves that $r_\alpha r_{-\alpha} = 1$ for all $\alpha \in \Phi$. On the other hand, $r_\alpha \overline{r_{-\alpha}} = 1$ (with $\overline{\cdot}$ denoting the complex conjugate in $\mathbb{C}$) follows from

$$r_\alpha \lambda_{-\alpha\circ\theta} x_{-\alpha} = \theta(r_\alpha x_{-\alpha\circ\theta}) = \theta\circ\sigma(x_\alpha) = \sigma\circ\theta(x_\alpha) = \lambda_\alpha \sigma(x_{\alpha\circ\theta}) = \lambda_\alpha \overline{r_{-\alpha}}^{-1} x_{-\alpha};$$

recall that $\sigma \circ \theta = \theta \circ \sigma$ and $\lambda_{-\alpha\circ\theta} = \lambda_\alpha$ by (a). Together, we have $r_\alpha \in \mathbb{R}$ for all $\alpha \in \Phi$. Since σ has order two, $r_{-\alpha\circ\theta} = r_\alpha^{-1} = r_{-\alpha}$ for all $\alpha \in \Phi$. □

As for $\lambda_i = \lambda_{\alpha_i}$, we sometimes write $r_i = r_{\alpha_i}$. We now consider Chevalley systems as defined in [Bourbaki 1975, Chapter VIII, Section 3, Definition 3]; see also [Đoković 1987, Lemma 2].

Definition 8. We use the previous notation. A *Chevalley system* of $\mathfrak{g}^c$ with respect to $\mathfrak{h}^c$ is a family $(w_\alpha)_{\alpha\in\Phi}$ where $w_\alpha \in \mathfrak{g}_\alpha$ with $[w_\alpha, w_{-\alpha}] = -h_\alpha$ for all $\alpha \in \Phi$ and such that the linear map defined by $h \mapsto -h$ for $h \in \mathfrak{h}^c$ and $w_\alpha \mapsto w_{-\alpha}$ for $\alpha \in \Phi$ is a Lie automorphism, called *Chevalley automorphism*. If $\theta(w_\alpha) = \lambda_\alpha w_{\alpha\circ\theta}$ and $\sigma(w_\alpha) = \lambda_\alpha w_{-\alpha\circ\theta}$ for all $\alpha \in \Phi$ (with λ_α as in Lemma 7), then $(w_\alpha)_{\alpha\in\Phi}$ is called *adapted* with respect to $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ (and the Chevalley basis $\{h_i, x_\alpha \mid i, \alpha\}$).

We first show that adapted Chevalley systems exist. Then, for real forms of inner type, we construct an adapted Chevalley system from our given Chevalley basis; see [Đoković 1987, Lemma 2].

Lemma 9. *There is an adapted Chevalley system $(v_\alpha)_{\alpha\in\Phi}$ of $\mathfrak{g}^c$ with respect to $\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{p}$ and $\mathfrak{h}^c$.*

Proof. For $\alpha\in\Phi$ let $z_\alpha=\varepsilon_\alpha x_\alpha$ where $\varepsilon_\alpha=-1$ if $\alpha\in\Phi^-$ is negative, and $\varepsilon_\alpha=1$ otherwise. We first prove that $(z_\alpha)_{\alpha\in\Phi}$ is a Chevalley system of $\mathfrak{g}^c$. Clearly, $[z_\alpha,z_{-\alpha}]=\varepsilon_\alpha\varepsilon_{-\alpha}h_\alpha=-h_\alpha$. Let ψ be the linear map defined by $\psi(h)=-h$ for $h\in\mathfrak{h}^c$ and $\psi(z_\alpha)=z_{-\alpha}$ for $\alpha\in\Phi$. If $\alpha\in\Phi$, then $\varepsilon_{-\alpha}=-\varepsilon_\alpha$ and $\psi(x_\alpha)=\psi(\varepsilon_\alpha z_\alpha)=\varepsilon_\alpha z_{-\alpha}=-x_{-\alpha}$. If $\alpha,\beta\in\Phi$ with $\alpha+\beta\in\Phi$, then $\psi([z_\alpha,z_\beta])=\psi(\varepsilon_\alpha\varepsilon_\beta N_{\alpha,\beta}x_{\alpha+\beta})=\varepsilon_\alpha\varepsilon_\beta\varepsilon_{\alpha+\beta}N_{\alpha,\beta}z_{-\alpha-\beta}$, and $N_{\alpha,\beta}=-N_{-\alpha,-\beta}$ yields $\psi([z_\alpha,z_\beta])=[\psi(z_\alpha),\psi(z_\beta)]$. Also, $\psi([z_\alpha,z_{-\alpha}])=[\psi(z_\alpha),\psi(z_{-\alpha})]$ and $\psi([h,z_\alpha])=[\psi(h),\psi(z_\alpha)]$; thus ψ is an automorphism and $(z_\alpha)_{\alpha\in\Phi}$ is a Chevalley system with respect to $\mathfrak{h}^c$.

We have seen in Section 2.2 that

$$\tilde{\mathfrak{u}}=\operatorname{Span}_{\mathbb{R}}(\{\imath h_1,\dots,\imath h_l,x_\alpha-x_{-\alpha},\imath(x_\alpha+x_{-\alpha})\mid\alpha\in\Phi^+\})$$

is a compact real form of $\mathfrak{g}^c$. If $\tilde\tau$ is the corresponding complex conjugation, then $\tilde\tau(x_\alpha)=-x_{-\alpha}$ for all $\alpha\in\Phi$ and $\tilde\tau(h_i)=-h_i$ for all i. In particular, $\tilde\tau$ and θ commute, and $\tilde\sigma=\theta\circ\tilde\tau$ is a real structure defining a real form $\tilde{\mathfrak{g}}=\mathfrak{g}(\theta,\tilde{\mathfrak{u}})=\tilde{\mathfrak{k}}\oplus\tilde{\mathfrak{p}}$ with Cartan involution θ (or, more precisely, the restriction of θ to $\tilde{\mathfrak{g}}$). If $\pi(i)=i$, then $\imath h_i\in\tilde{\mathfrak{k}}$; otherwise $\imath(h_i+h_{\pi(i)})\in\tilde{\mathfrak{k}}$ and $h_i-h_{\pi(i)}\in\tilde{\mathfrak{p}}$ (see Section 2.2); thus $\tilde{\mathfrak{g}}$ has a standard Cartan subalgebra $\tilde{\mathfrak{h}}$ with $(\tilde{\mathfrak{h}})^c=\mathfrak{h}^c$. It follows readily from the definition of $\tilde\sigma$ that $\tilde\sigma(z_\alpha)=\lambda_\alpha z_{-\alpha\circ\theta}$ for all $\alpha\in\Phi$. Clearly, $\theta(z_\alpha)=\lambda_\alpha z_{\alpha\circ\theta}$, which shows that $(z_\alpha)_{\alpha\in\Phi}$ is an adapted Chevalley system with respect to $\tilde{\mathfrak{g}}=\tilde{\mathfrak{k}}\oplus\tilde{\mathfrak{p}}$ and $\mathfrak{h}^c$.

Set $\mathfrak{u}=\mathfrak{k}\oplus\imath\mathfrak{p}$. Then $\mathfrak{u}$ is the compact form of $\mathfrak{g}^c$ associated with the real form $\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{p}$ (cf. Section 2.2). Let $\tau\colon\mathfrak{g}^c\to\mathfrak{g}^c$ be the complex conjugation with respect to $\mathfrak{u}$; then $\sigma=\theta\circ\tau$, and θ and τ commute. Thus, $\mathfrak{g}$ and $\tilde{\mathfrak{g}}$ both are real forms defined by the automorphism θ and the compact real structures τ and $\tilde\tau$, respectively. Using Lemma 7, we get $\tau(x_i)=r_i\lambda_i y_i$ and $\tau(y_i)=r_i^{-1}\lambda_i x_i$. Let $\eta\colon\mathfrak{g}^c\to\mathfrak{g}^c$ be the automorphism which maps (h_i,x_i,y_i) to $(h_i,|r_i|^{-1/2}x_i,|r_i|^{1/2}y_i)$ for all i. A short calculation shows that the compact structures $\eta^{-1}\circ\tau\circ\eta$ and $\tilde\tau$ commute. As shown in [Onishchik 2004, Proposition 3.5], commuting compact structures are equal; hence $\tau\circ\eta=\eta\circ\tilde\tau$. Again, using Lemma 7, we see that $\theta\circ\eta=\eta\circ\theta$, whence also $\sigma\circ\eta=\eta\circ\tilde\sigma$. Now consider $(v_\alpha)_{\alpha\in\Phi}$ with $v_\alpha=\eta(z_\alpha)$. Clearly, this is a Chevalley system: First, $v_\alpha\in\mathfrak{g}_\alpha$ and $[v_\alpha,v_{-\alpha}]=\eta(-h_\alpha)=-h_\alpha$ for all $\alpha\in\Phi$. Second, if ψ is the Chevalley automorphism corresponding to $(z_\alpha)_{\alpha\in\Phi}$, then $\eta\circ\psi\circ\eta^{-1}$ is the Chevalley automorphism corresponding to $(v_\alpha)_{\alpha\in\Phi}$. Also, for $\alpha\in\Phi$ we have $\sigma(v_\alpha)=\sigma\circ\eta(z_\alpha)=\eta\circ\tilde\sigma(z_\alpha)=\lambda_\alpha v_{-\alpha\circ\theta}$ and $\theta(v_\alpha)=\theta\circ\eta(z_\alpha)=\eta\circ\theta(z_\alpha)=\lambda_\alpha v_{\alpha\circ\theta}$, so $(v_\alpha)_{\alpha\in\Phi}$ is adapted with respect to $\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{p}$. ☐

Proposition 10. *We use the previous notation. For $\alpha\in\Phi$ let $z_\alpha=\varepsilon_\alpha x_\alpha$ where $\varepsilon_\alpha=-1$ if $\alpha\in\Phi^-$ is negative, and $\varepsilon_\alpha=1$ otherwise. Since $\sigma(z_{\alpha_i})=-r_i z_{-\alpha_{\pi(i)}}$*

with $r_i = r_{\pi(i)}$ by Lemma 7, there are $\tau_i = \tau_{\pi(i)} \in \mathbb{R}$ such that $\tau_i\tau_{\pi(i)}r_i \in \{\pm 1\}$. Let ψ be the automorphism of $\mathfrak{g}^c$ mapping (h_i, x_i, y_i) to $(h_i, \tau_i x_i, \tau_i^{-1} y_i)$ for all i; then ψ commutes with θ. Define $w_\alpha = \psi(z_\alpha)$ for $\alpha \in \Phi$.

(a) *$(z_\alpha)_{\alpha\in\Phi}$ and $(w_\alpha)_{\alpha\in\Phi}$ are Chevalley systems with respect to $\mathfrak{h}^c$.*

(b) *$\theta(w_{\alpha_i}) = \lambda_i w_{\alpha_{\pi(i)}}$ and $\sigma(w_{\alpha_i}) = \lambda_i w_{-\alpha_{\pi(i)}}$ for all i.*

(c) *If $\mathfrak{g}$ is of inner type, then $(w_\alpha)_{\alpha\in\Phi}$ is adapted with respect to $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$.*

Proof. (a) This follows as in the proof of Lemma 9.

(b) Recall $z_{\alpha_i} = x_i$ and $z_{-\alpha_{\pi(i)}} = -y_{\pi(i)}$ for all i. Now $w_{-\alpha_{\pi(i)}} = \psi(z_{-\alpha_{\pi(i)}}) = \tau_{\pi(i)}^{-1} z_{-\alpha_{\pi(i)}}$ yields

$$\sigma(w_{\alpha_i}) = \sigma(\psi(z_{\alpha_i})) = \sigma(\tau_i z_{\alpha_i}) = -\tau_i r_i z_{-\alpha_{\pi(i)}} = -\tau_i r_i \tau_{\pi(i)} w_{-\alpha_{\pi(i)}} = r_i' w_{-\alpha_{\pi(i)}},$$

where $r_i' = -\tau_i\tau_{\pi(i)}r_i \in \{\pm 1\}$. We have $\theta(x_i) = \lambda_i x_i$ if $\pi(i) = i$, and $\theta(x_i) = x_{\pi(i)}$ otherwise, and, therefore, $\tau_i = \tau_{\pi(i)}$ implies that $\theta(w_{\alpha_i}) = \lambda_i w_{\alpha_i}$ if $\pi(i) = i$, and $\theta(w_{\alpha_i}) = w_{\alpha_{\pi(i)}}$ otherwise. By Lemma 9, there exists an adapted Chevalley system $(v_\alpha)_{\alpha\in\Phi}$ with respect to $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and $\mathfrak{h}^c$; each v_α can be written as $v_\alpha = c_\alpha w_\alpha$ for some $c_\alpha \in \mathbb{C}$. It follows from

$$-h_i = [v_{\alpha_i}, v_{-\alpha_i}] = c_{\alpha_i} c_{-\alpha_i} [w_{\alpha_i}, w_{-\alpha_i}] = -c_{\alpha_i} c_{-\alpha_i} h_i$$

that $c_{-\alpha_i} = c_{\alpha_i}^{-1}$ for all i. If $\pi(i) \neq i$, then

$$c_{\alpha_i} w_{\alpha_{\pi(i)}} = \theta(c_{\alpha_i} w_{\alpha_i}) = \theta(v_{\alpha_i}) = v_{\alpha_{\pi(i)}} = c_{\alpha_{\pi(i)}} w_{\alpha_{\pi(i)}};$$

hence $c_{\alpha_i} = c_{\alpha_{\pi(i)}}$ for all i. Thus $\overline{c_{\alpha_i}} c_{\alpha_{\pi(i)}} > 0$ is real for every i, and $r_i' = \lambda_i$ follows from $r_i', \lambda_i \in \{\pm 1\}$ and

$$\lambda_i v_{-\alpha_{\pi(i)}} = \sigma(v_{\alpha_i}) = \overline{c_{\alpha_i}} \sigma(w_{\alpha_i}) = \overline{c_{\alpha_i}} r_i' c_{-\alpha_{\pi(i)}}^{-1} v_{-\alpha_{\pi(i)}} = r_i' \overline{c_{\alpha_i}} c_{\alpha_{\pi(i)}} v_{-\alpha_{\pi(i)}}.$$

(c) By (b) we know that $\sigma(w_{\alpha_i}) = \lambda_{\alpha_i} w_{-\alpha_{\pi(i)}}$, and Lemma 7 yields $\sigma(w_{-\alpha_i}) = \lambda_{-\alpha_i} w_{\alpha_{\pi(i)}}$ for $i = 1, \dots, l$. For $\alpha, \beta \in \Phi$ with $\alpha + \beta \in \Phi$ write $[w_\alpha, w_\beta] = M_{\alpha,\beta} w_{\alpha+\beta}$ where $M_{\alpha,\beta} = M_{-\alpha,-\beta}$ is real (in fact, integral). Suppose now that for $\alpha, \beta \in \Phi$ we have $\sigma(w_\alpha) = \lambda_\alpha w_{-\alpha\circ\theta}$ and $\sigma(w_\beta) = \lambda_\beta w_{-\beta\circ\theta}$. Then

$$M_{\alpha,\beta}\sigma(w_{\alpha+\beta}) = \sigma([w_\alpha, w_\beta]) = [\sigma(w_\alpha), \sigma(w_\beta)] = \lambda_\alpha\lambda_\beta M_{-\alpha\circ\theta,-\beta\circ\theta} w_{-(\alpha+\beta)\circ\theta}.$$

If $\mathfrak{g}$ is of inner type, then $\alpha \circ \theta = \alpha$ and $\lambda_\alpha\lambda_\beta = \lambda_{\alpha+\beta}$ for all $\alpha, \beta \in \Phi$. Thus, in this case, $M_{\alpha,\beta} = M_{-\alpha\circ\theta,-\beta\circ\theta} = M_{-\alpha,-\beta}$, and induction on the height of α proves that $\sigma(w_\alpha) = \lambda_\alpha w_{-\alpha\circ\theta}$. Similarly, $\theta(w_\alpha) = \lambda_\alpha w_{\alpha\circ\theta}$ for all α; thus $(w_\alpha)_{\alpha\in\Phi}$ is adapted with respect to $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. □

The proof of Proposition 10(b) has the following important corollary, which we use in Section 7. Recall that $\theta(x_i) = \lambda_i x_{\pi(i)}$ and $\sigma(x_i) = r_i y_{\pi(i)}$ for all i.

Corollary 11. *The coefficients r_i and $-\lambda_i$ have the same sign for all i.*

Proof. In the proof of Proposition 10(b) we have shown that $\lambda_i = r_i' = -\tau_i \tau_{\pi(i)} r_i = -\tau_i^2 r_i$. □

6. Constructing complex Cayley triples

Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be as in Section 2.2, with complexification $\mathfrak{g}^c$, Cartan involution θ, and complex conjugation σ. As usual, we denote by Φ a root system of $\mathfrak{g}^c$ with basis of simple roots Δ; let $\{h_i, x_\alpha \mid i, \alpha\}$ be a corresponding Chevalley basis. We now discuss our Main Problem (see Section 3); that is, given a homogeneous $\mathfrak{sl}_2$-triple (f, h, e) in $\mathfrak{g}^c$, we want to construct a complex Cayley triple (f', h', e') which is K^c-conjugate to (f, h, e). As constructed in Section 4, we also assume we have a standard carrier algebra $\mathfrak{s}^c = \mathfrak{s}^c(e, \mathfrak{t})$ containing f, h, e (see Remark 6) and normalized by the Cartan subalgebra $\mathfrak{h}_0^c = \mathfrak{h}_0 + \imath\mathfrak{h}_0$ of $\mathfrak{k}^c$ with $\mathfrak{h}_0 \subseteq \mathfrak{k}$ as in Section 2.2.

We will see in Section 6.1 that $\mathfrak{s}^c$ is σ-stable; hence $\mathfrak{s} = \mathfrak{s}^c \cap \mathfrak{g}$ is a real form of $\mathfrak{s}^c$. Also, we will see that $\mathfrak{s}^c$ is θ-stable; thus

$$(\star) \qquad \mathfrak{s} = (\mathfrak{s}^c \cap \mathfrak{k}) \oplus (\mathfrak{s}^c \cap \mathfrak{p})$$

is a Cartan decomposition whose Cartan involution is the restriction of θ to $\mathfrak{s}$. Note that $\mathfrak{s}_0 \cap \mathfrak{k}^c$ and $\mathfrak{s}_0 \cap \mathfrak{k}$ contain Cartan subalgebras of $\mathfrak{s}^c$ and $\mathfrak{s}$, respectively, namely, $\mathfrak{h}_0^c \cap \mathfrak{s}_0$ and $\mathfrak{h}_0 \cap \mathfrak{s}_0$. In particular, the real form $\mathfrak{s}$ is always of inner type and $\mathfrak{h}_0 \cap \mathfrak{s}_0$ is a standard Cartan subalgebra. Thus the results of Section 5 can be applied: we show in Section 6.1 how to construct an adapted Chevalley system for $\mathfrak{s}^c$; here adapted always means with respect to $\mathfrak{h}_0^c \cap \mathfrak{s}_0$, the Cartan decomposition ($\star$), and a chosen Chevalley basis of $\mathfrak{s}^c$.

By construction, the triple (f, h, e) is also a homogeneous $\mathfrak{sl}_2$-triple in $\mathfrak{s}^c$. The approach of [Đoković 1987] is to find $x \in \mathfrak{s}_1$ with $[x, \sigma(x)] = h$ so that $(\sigma(x), h, x)$ is a complex Cayley triple in $\mathfrak{s}^c$, thus also in $\mathfrak{g}^c$. By the Kostant–Sekiguchi correspondence and [Kostant and Rallis 1971, Lemma 4], such an x exists and $(\sigma(x), h, x)$ is K^c-conjugate to (f, h, e). If $\mathfrak{s}^c$ is principal, then Chevalley systems can be used to find x; see Section 6.2. If $\mathfrak{s}^c$ is not principal, then we make a case distinction and use induction; see Section 6.3.

6.1. ***Constructing an adapted Chevalley system.*** In the following, let $\Phi_{\mathfrak{s}^c}$ be the root system of $\mathfrak{s}^c$ with respect to $\mathfrak{h}_0^c \cap \mathfrak{s}_0$; let $\Delta_{\mathfrak{s}^c} = \{\beta_1, \ldots, \beta_s\}$ be a basis of simple roots. As mentioned in Section 4, we can assume that each root space $\mathfrak{s}_{\beta_i}$ either lies in $\mathfrak{s}_0$ or $\mathfrak{s}_1$.

6.1.1. *Inner type.* If $\mathfrak{g}$ is of inner type, then $\mathfrak{h}_0^c \le \mathfrak{k}^c$ is also a Cartan subalgebra of $\mathfrak{g}^c$; hence $\Phi_{\mathfrak{s}^c}$ can be considered as a root subsystem of Φ. This implies that $\{x_\alpha \mid \alpha \in \Phi_{\mathfrak{s}^c}\}$, along with certain elements of $\mathfrak{h}_0^c \cap \mathfrak{s}_0$, forms a Chevalley basis of $\mathfrak{s}^c$. We denote it by $\{k_i, w_\alpha \mid \alpha \in \Phi_{\mathfrak{s}^c}, i = 1, \dots, s\}$ and let $\{k_i, a_i, b_i \mid i\}$ be the canonical generating set it contains. As usual, write $k_\alpha = [w_\alpha, w_{-\alpha}]$ for $\alpha \in \Phi_{\mathfrak{s}^c}$. We have seen in Section 2.2.2 that $\sigma(w_\alpha) \in \{\pm w_{-\alpha}\}$ and $\theta(w_\alpha) \in \{\pm w_\alpha\}$; hence $\mathfrak{s}^c$ is σ- and θ-stable. For $\Phi_{\mathfrak{s}^c}$ we use an ordering compatible with that of Φ. Let $z_\alpha = w_\alpha$ and $z_{-\alpha} = -w_{-\alpha}$ for $\alpha \in \Phi^+_{\mathfrak{s}^c}$; then $(z_\alpha)_{\alpha\in\Phi_{\mathfrak{s}^c}}$ is an adapted Chevalley system of $\mathfrak{s}^c$.

6.1.2. *Outer type.* Now let $\mathfrak{g}$ be of outer type with defining outer automorphism $\theta = \varphi \circ \chi$. By construction, each homogeneous component $\mathfrak{s}_k$ lies either in $\mathfrak{k}^c$ or in $\mathfrak{p}^c$, which shows that $\mathfrak{s}^c$ is θ-stable. By definition, each $\mathfrak{s}_k$ is normalized by $\mathfrak{h}_0^c$; thus it is a sum of weight spaces (with respect to $\mathfrak{h}_0^c$) as considered in Sections 2.2.3 and 2.2.4; in the following we use the notation introduced in these sections. Let $\alpha \in \Phi_{\mathfrak{s}^c}$; then $\mathfrak{s}_\alpha$ is an $\mathfrak{h}_0^c$-weight space, and it is either contained in $\mathfrak{k}^c$ or in $\mathfrak{p}^c$ (since it lies in a homogeneous component $\mathfrak{s}_k$). These observations show that there is an $\alpha' \in \Phi$ such that either $\mathfrak{s}_\alpha = \mathrm{Span}_{\mathbb{C}}(u_{\alpha'})$ or $\mathfrak{s}_\alpha = \mathrm{Span}_{\mathbb{C}}(v_{\alpha'})$, and, accordingly, $\mathfrak{s}_{-\alpha} = \mathrm{Span}_{\mathbb{C}}(u_{-\alpha'})$ or $\mathfrak{s}_{-\alpha} = \mathrm{Span}_{\mathbb{C}}(v_{-\alpha'})$. Since $\sigma(u_{\alpha'}) = \pm u_{-\alpha'}$ and $\sigma(v_\alpha) = \pm v_{-\alpha'}$, this shows that $\mathfrak{s}^c$ is stable under σ. We can now define a new set of canonical generators $\{k_i, a_i, b_i \mid i = 1, \dots, s\}$ for $\mathfrak{s}$; we make a case distinction:

- If $\mathfrak{s}_{\beta_i}$ is spanned by $u_\alpha = x_\alpha$ with $\varphi(\alpha) = \alpha$, then define $a_i = x_\alpha$, $b_i = x_{-\alpha}$ and $k_i = [a_i, b_i]$.
- Now let $\mathfrak{s}_{\beta_i}$ be spanned by $u_\alpha = x_\alpha + x_{\varphi(\alpha)}$ with $\varphi(\alpha) \neq \alpha$. Note that $\beta = \alpha - \varphi(\alpha)$ is not a root because φ maps positive roots on positive roots but $\varphi(\beta) = -\beta$. This proves $[u_\alpha, u_{-\alpha}] = h_\alpha + h_{\varphi(\alpha)}$. Also, it follows that $\langle \alpha, \varphi(\alpha)\rangle \le 0$ (see [Humphreys 1978, Lemma 9.4]) and finally $\langle \alpha, \varphi(\alpha)\rangle \in \{0, -1\}$, as Φ is simply laced, which means that there is only one root length; in particular, $\langle \alpha, \varphi(\alpha)\rangle = \langle \varphi(\alpha), \alpha\rangle$. The latter now implies that $[h_\alpha + h_{\varphi(\alpha)}, u_\alpha] = (2 + \langle \varphi(\alpha), \alpha\rangle)u_\alpha$ since $\varphi(\alpha)(h_\alpha) = \langle \varphi(\alpha), \alpha\rangle = \langle \alpha, \varphi(\alpha)\rangle = \alpha(h_{\varphi(\alpha)})$. If $\langle \varphi(\alpha), \alpha\rangle = 0$, then we define $a_i = u_\alpha$, $b_i = u_{-\alpha}$, and $k_i = [a_i, b_i]$. Otherwise, we set $a_i = \sqrt{2}u_\alpha$, $b_i = \sqrt{2}u_{-\alpha}$, and $k_i = [a_i, b_i]$.
- If $\mathfrak{s}_{\beta_i}$ is spanned by $v_\alpha = x_\alpha - x_{\varphi(\alpha)}$, then we do exactly the same as in the previous case with u replaced by v.

In all cases we find $a_i \in \mathfrak{s}_{\beta_i}$, $b_i \in \mathfrak{s}_{-\beta_i}$, and $k_i = [a_i, b_i]$ such that $[k_i, a_i] = 2b_i$ for all i. By Proposition 3, $\{k_i, a_i, b_i \mid i\}$ is a canonical generating set for $\mathfrak{s}^c$, and, by construction, $\sigma(a_i) = \pm b_i$ for all i. We extend this canonical generating set to a Chevalley basis $\{k_i, w_\alpha \mid i, \alpha\}$ of $\mathfrak{s}^c$ such that $w_{\alpha_i} = a_i$ and $w_{-\alpha_i} = b_i$; as usual, write $k_\alpha = [w_\alpha, w_{-\alpha}]$ for all α. We now define $z_\alpha = w_\alpha$ for $\alpha > 0$ and $z_\alpha = -w_\alpha$ for $\alpha < 0$; it is straightforward to verify that $(z_\alpha)_{\alpha\in\Phi_{\mathfrak{s}^c}}$ is an adapted Chevalley system of $\mathfrak{s}^c$; see Proposition 10.

The conclusion is that for all $\mathfrak{g}$ we can find an adapted Chevalley system of $\mathfrak{s}^c$, and the coefficients of its elements with respect to the given basis of $\mathfrak{g}$ lie in $\mathbb{Q}(\imath, \sqrt{2})$; in particular, in $\mathbb{Q}^{\surd}(\imath)$.

6.2. ***The principal case.*** This construction follows [Đoković 1987, Lemma 3]. We use the previous notation and suppose that the carrier algebra $\mathfrak{s}^c$ of (f, h, e) is principal; that is, there is a basis $\Delta_{\mathfrak{s}^c}$ of $\Phi_{\mathfrak{s}^c}$ such that for every $\alpha \in \Delta_{\mathfrak{s}^c}$ we have $\mathfrak{s}_\alpha \subseteq \mathfrak{s}_1$. Let $(z_\alpha)_{\alpha\in\Phi_{\mathfrak{s}^c}}$ be the adapted Chevalley system for $\mathfrak{s}^c$ as constructed in the previous section. We want to find $x \in \mathfrak{s}_1$ with $[x, \sigma(x)] = h$ of the form $x = \sum_{\alpha\in\Delta_{\mathfrak{s}^c}} c_\alpha z_\alpha$ with all c_α real. Note that $\sigma(x) = -\sum_{\alpha\in\Delta_{\mathfrak{s}^c}} c_\alpha z_{-\alpha}$ and $\alpha - \beta \notin \Phi_{\mathfrak{s}^c}$ for all $\alpha, \beta \in \Delta_{\mathfrak{s}^c}$. Thus, the equation we have to solve is $h = [x, \sigma(x)] = \sum_{\alpha\in\Delta_{\mathfrak{s}^c}} c_\alpha^2 k_\alpha$; recall that $[z_\alpha, z_{-\alpha}] = -k_\alpha$. Note that $\beta(h) = 2$ for all $\beta \in \Delta_{\mathfrak{s}^c}$ since $z_\beta \in \mathfrak{s}_1$ and $h/2$ is the defining element of $\mathfrak{s}^c$. This shows that our equation is equivalent to the system of equations $2 = \sum_{\alpha\in\Delta_{\mathfrak{s}^c}} d_\alpha \beta(k_\alpha)$ with $d_\alpha = c_\alpha^2$ where β ranges over $\Delta_{\mathfrak{s}^c}$. The coefficients $\beta(k_\alpha)$ of this system are the entries of the Cartan matrix of $\Phi_{\mathfrak{s}^c}$, whose inverse has nonnegative entries; see [Humphreys 1978, Section 13.1]. Thus, the system has a solution with all $d_\alpha \geq 0$ real. In conclusion, to construct x, we first compute $h = \sum_{\alpha\in\Delta_{\mathfrak{s}^c}} d_\alpha k_\alpha$, and then set $x = \sum_{\alpha\in\Delta_{\mathfrak{s}^c}} c_\alpha z_\alpha$ where $c_\alpha = \sqrt{d_\alpha}$ is real for every $\alpha \in \Delta_{\mathfrak{s}^c}$.

We now show that each $c_\alpha \in \mathbb{Q}^{\surd}$. For every $\alpha \in \Delta_{\mathfrak{s}^c} \subseteq \Phi_{\mathfrak{s}^c}$, the element $k_\alpha = [w_\alpha, w_{-\alpha}]$ is a $\mathbb{Z}$-linear combination of $k_1, \dots, k_s$, the elements of the Chevalley basis of $\mathfrak{s}^c$ that span its Cartan subalgebra $\mathfrak{h}_0^c \cap \mathfrak{s}_0$; see [Humphreys 1978, Theorem 25.2]. As shown in the previous paragraph, these elements are $\mathbb{Z}$-linear combinations of $h_1, \dots, h_l$, the elements of the Chevalley basis of $\mathfrak{g}^c$ that span $\mathfrak{h}^c$. Similarly, the element h, which is the characteristic of an $\mathfrak{sl}_2$-triple, is a $\mathbb{Z}$-linear combination of $h_1, \dots, h_l$. Together, all this implies that the d_α are in fact rational; thus $c_\alpha \in \mathbb{Q}^{\surd}$.

6.3. ***Nonprincipal case.*** Now suppose that the carrier algebra $\mathfrak{s}^c$ of (f, h, e) is nonprincipal. As mentioned above, there exists $x \in \mathfrak{s}_1$ such that $(\sigma(x), h, x)$ is a complex Cayley triple in the same K^c-orbit as (f, h, e). However, constructing x is not straightforward. We first set up the system of rational polynomial equations in the coefficients of x with respect to a basis of $\mathfrak{s}_1$, equivalent to $[x, \sigma(x)] = h$. Note that this is a system of $\dim \mathfrak{s}_0$ polynomial equations in $\dim \mathfrak{s}_1$ variables. Then in order to solve them we use a brute-force approach; that is, for $i = 1, 2, 3, \dots,$ we set all but i indeterminates in these equations to zero. For each equation system that arises we check, using Gröbner bases (see for example [Cox et al. 1992]), whether a solution over $\mathbb{C}$ exists. We stop when we find an equation system consisting of equations of the form $T^2 = a$, where $a \in \mathbb{Q}$ and T is an indeterminate, or $T_c = a_{c_1} T_{c_1}^2 + \cdots + a_{c_m} T_{c_m}^2$, where each T_{c_i} satisfies an equation of the first type. It is then straightforward to obtain a solution over $\mathbb{Q}^{\surd}(\imath)$. This systematic approach for constructing a complex Cayley triple $(\sigma(x), h, x)$ can easily be carried out automatically by a computer. It turned out to work well in all our computations for the carrier algebras in the real forms constructed in Section 2.2; our experiments include all simple real Lie algebras of rank at most 8. Unfortunately, we have no proof that a solution of the equation system always exists over the field $\mathbb{Q}^{\surd}(\imath)$; hence we cannot prove that our approach will always work.

6.3.1. *A database.* To reduce work, we have constructed a database of the simple nonprincipal carrier algebras that appeared during our calculations. Let $\mathfrak{s}^c$ be such a carrier algebra. As shown in Section 6.1, there is a canonical generating set $\{k_i, a_i, b_i \mid i = 1, \ldots, s\}$ of $\mathfrak{s}^c$ such that $a_i \in \mathfrak{s}_{\varepsilon_i}$ with $\varepsilon_i \in \{0, 1\}$ and $\sigma(a_i) = \lambda_i b_i$ with $\lambda_i \in \{\pm 1\}$ for all i. Since $\sigma(k_i) = -k_i$ for all i, the map σ is determined by the signs $\lambda_1, \ldots, \lambda_s$. Moreover, $k_1, \ldots, k_s \in \mathfrak{s}_0$, and, if $a_i \in \mathfrak{s}_k$, then $b_i \in \mathfrak{s}_{-k}$. Thus, the following data describes $\mathfrak{s}^c$, its grading, and σ completely; we store this data in our database:

- a multiplication table, canonical generators $\{k_i, a_i, b_i \mid i\}$, and Cartan matrix C,
- the signs $\lambda_1, \ldots, \lambda_s$ and $\varepsilon_1, \ldots, \varepsilon_s$,
- a complex Cayley triple (f, h, e) in $\mathfrak{s}$ such that $e \in \mathfrak{s}_1$ is in general position.

Suppose in our computations we consider a real semisimple Lie algebra $\mathfrak{g}' = \mathfrak{k}' \oplus \mathfrak{p}'$ with complexification $(\mathfrak{g}')^c = (\mathfrak{k}')^c \oplus (\mathfrak{p}')^c$ and complex conjugation σ'. Let (f', h', e') be a homogeneous $\mathfrak{sl}_2$-triple in $(\mathfrak{g}')^c$, and we want to find a conjugate complex Cayley triple in $(\mathfrak{g}')^c$. As before, we first construct the carrier algebra $(\mathfrak{s}')^c$ of the triple. If it is principal, then we proceed as in Section 6.2, so let it be nonprincipal. Recall that $(\mathfrak{s}')^c$ is semisimple and, by considering its simple components separately, we can assume that $(\mathfrak{s}')^c$ itself is simple. Suppose in our database there exists a simple carrier algebra $\mathfrak{s}^c$ whose parameters as described above satisfy the following:

(1) $(\mathfrak{s}')^c$ has canonical generators $\{k_i', a_i', b_i' \mid i\}$ with Cartan matrix C,

(2) if $\sigma'(a_i') = \lambda_i' b_i'$, then $\operatorname{sgn}(\lambda_i') = \operatorname{sgn}(\lambda_i)$ for all i,

(3) if $a_i' \in \mathfrak{s}'_{\varepsilon_i'}$, then $\varepsilon_i' = \varepsilon_i$ for all i.

If all this holds, then we can get a complex Cayley triple in $(\mathfrak{s}')^c$ as follows. Let φ be the isomorphism from $\mathfrak{s}^c$ to $(\mathfrak{s}')^c$ which maps (k_i, a_i, b_i) to $(k_i', \mu_i a_i', \mu_i^{-1} b_i')$, where $\mu_i = \sqrt{\lambda_i / \lambda_i'}$ for all i. Obviously φ is an isomorphism of $\mathbb{Z}$-graded Lie algebras. A short calculation shows that the antilinear homomorphisms $\varphi \circ \sigma$ and $\sigma' \circ \varphi$ agree on the canonical generators of $\mathfrak{s}^c$; thus $\varphi \circ \sigma = \sigma' \circ \varphi$. Since φ maps the unique defining element $h/2$ of $\mathfrak{s}^c$ onto the unique defining element $h'/2$ of $(\mathfrak{s}')^c$, we have $h' = \varphi(h)$. Let $x = \varphi(e)$ and $y = \varphi(f)$; then (y, h', x) is a complex Cayley triple in $(\mathfrak{s}')^c$. Since $x \in \mathfrak{s}_1'$ is in general position, (y, h', x) is $(K')^c$-conjugate to (f', h', e'). The conclusion is that by storing the simple carrier algebras in a database we can find a complex Cayley triple in a carrier algebra by a look-up in the database.

7. Isomorphisms

Let $\mathfrak{g}^c$ be a simple complex Lie algebra with real form $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and Cartan involution θ and complex conjugation σ. As usual, we extend θ to an automorphism of $\mathfrak{g}^c$. Let $(\mathfrak{g}')^c$ be a second simple complex Lie algebra with real form $\mathfrak{g}' = \mathfrak{k}' \oplus \mathfrak{p}'$, Cartan involution θ', and complex conjugation σ'. We consider the problem to decide whether $\mathfrak{g}$ and $\mathfrak{g}'$ are isomorphic, and, if they are, to find an isomorphism. For this we may obviously assume that $\mathfrak{g}^c$ and $(\mathfrak{g}')^c$ are isomorphic.

Recall that a Cartan decomposition is unique up to conjugacy; see [Onishchik 2004, Theorem 5.1]. Thus, if $\mathfrak{g}$ and $\mathfrak{g}'$ are isomorphic, then there also exists an isomorphism $\psi\colon \mathfrak{g} \to \mathfrak{g}'$ with $\psi(\mathfrak{k}) = \mathfrak{k}'$ and $\psi(\mathfrak{p}) = \mathfrak{p}'$. Clearly, such an isomorphism extends to an isomorphism $\psi\colon \mathfrak{g}^c \to (\mathfrak{g}')^c$ with $\psi \circ \theta = \theta' \circ \psi$ and $\psi \circ \sigma = \sigma' \circ \psi$. Conversely, if we find an isomorphism

$$(*) \qquad \psi\colon \mathfrak{g}^c \to (\mathfrak{g}')^c \quad \text{with } \psi \circ \theta = \theta' \circ \psi \text{ and } \psi \circ \sigma = \sigma' \circ \psi,$$

then ψ restricts to an isomorphism $\psi\colon \mathfrak{g} \to \mathfrak{g}'$ with $\psi(\mathfrak{k}) = \mathfrak{k}'$ and $\psi(\mathfrak{p}) = \mathfrak{p}'$.

We now describe a construction of the isomorphism $(*)$, which fails if and only if $\mathfrak{g}$ and $\mathfrak{g}'$ are not isomorphic. Our main tool is the technique described in the following preliminary section.

7.1. ***Weyl group action.*** We consider the following setup. Let $\mathfrak{h}^c \le \mathfrak{g}^c$ be a Cartan subalgebra of $\mathfrak{g}^c$ with corresponding root system Φ and basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$. Let W be the Weyl group associated to Φ. As usual, let $\{h_i, x_i, y_i \mid i\}$ be a canonical generating set contained in a Chevalley basis $\{h_i, x_\alpha \mid i, \alpha\}$ of $\mathfrak{g}^c$. Note that $\theta(x_\alpha) \in \mathfrak{g}_{\alpha\circ\theta}$, and we suppose that $\alpha \mapsto \alpha \circ \theta$ preserves Δ. Then $\theta = \varphi \circ \chi = \chi \circ \varphi$, where φ is a diagram automorphism permuting Δ, and χ is an inner automorphism with $\chi(h) = h$ for all $h \in \mathfrak{h}^c$. Let the permutation π be defined by $\varphi(\alpha_i) = \alpha_{\pi(i)}$. We further suppose that $\theta(x_\alpha) = \lambda_\alpha x_{\alpha\circ\theta}$ with $\lambda_\alpha = 1$ if $\alpha \circ \theta \ne \alpha$. Thus, $\lambda_\alpha \in \{\pm 1\}$ for all $\alpha \in \Phi$; we write $\lambda_i = \lambda_{\alpha_i}$ and call $\lambda_1, \dots, \lambda_l$ the parameters of θ.

By abuse of notation, to $w \in W$ we associate the automorphism $w \in \mathrm{Aut}(\mathfrak{g}^c)$ which maps (h_i, x_i, y_i) to $(h_{w(\alpha_i)}, x_{w(\alpha_i)}, x_{-w(\alpha_i)})$ for all i. Let $\Delta_\theta = \{\alpha \in \Delta \mid \alpha \circ \theta = \alpha\}$, let Φ_θ be the root subsystem of Φ with basis Δ_θ, and let W_θ be its Weyl group.

Lemma 12. *If $w = s_{\alpha_k} \in W_\theta$, then $\alpha \to \alpha \circ \theta$ preserves the basis of simple roots $w(\Delta)$.*

Proof. This follows readily if θ is inner since then φ is the identity and $\alpha \circ \theta = \alpha$ for all $\alpha \in \Phi$. So suppose φ is not the identity; hence Φ is simply laced; see [Onishchik 2004, Table 1]. Note that $\pi(k) = k$; thus $\theta(w(x_k)) = \theta(y_k) = \lambda_k y_k = \lambda_k w(x_k)$, which shows $w(\alpha_k) \circ \theta = w(\alpha_k) \in w(\Delta)$. If j is such that $\langle \alpha_j, \alpha_k \rangle = -1$, then $w(\alpha_j) = \alpha_j + \alpha_k$ and $w(x_j) = x_{\alpha_k + \alpha_j}$. Since Φ is simply laced, $N_{\alpha,\beta} = \pm 1$ for all $\alpha, \beta \in \Phi$, and $[x_k, x_j] = N_{\alpha_k,\alpha_j} x_{\alpha_k+\alpha_j}$ implies that $\theta(w(x_j)) = \pm x_{\alpha_k+\alpha_{\pi(j)}}$. Since also $\langle \alpha_{\pi(j)}, \alpha_k \rangle = -1$, we have $x_{\alpha_k+\alpha_{\pi(j)}} = w(x_{\pi(j)})$; hence $w(\alpha_j) \circ \theta = w(\alpha_{\pi(j)}) \in w(\Delta)$. Analogously, if $\langle \alpha_j, \alpha_k \rangle = 0$, then $w(x_j) = x_j$; hence $w(\alpha_j) \circ \theta = w(\alpha_{\pi(j)}) \in w(\Delta)$. □

Suppose θ has parameters $\lambda_1, \dots, \lambda_l$, that is, $\theta(x_i) = \lambda_i x_{\pi(i)}$ for all i, and let $w = s_{\alpha_k} \in W_\theta$. Clearly, $\{w(h_i), w(x_i), w(y_i) \mid i\}$ is a canonical generating set, and we modify it as follows: whenever $\pi(i) > i$, we replace $w(x_{\pi(i)})$ and $w(y_{\pi(i)})$ by $\theta(w(x_i))$ and $\theta(w(y_i))$; let $\{\bar{h}_i, \bar{x}_i, \bar{y}_i \mid i\}$ be the resulting canonical generating set with corresponding basis of simple roots $w(\Delta)$, which still is θ-stable by Lemma 12. By construction, if $\pi(i) \neq i$, then $\theta(\bar{x}_i) = \bar{x}_{\pi(i)}$. Now let $\pi(j) = j$ and recall that $w(\alpha_j) = \alpha_j - \langle \alpha_j, \alpha_k \rangle \alpha_k$ and $\pi(k) = k$. A case distinction on the value of $\langle \alpha_j, \alpha_k \rangle$ shows that

$$\theta(w(x_j)) = \lambda_j \lambda_k^{\langle \alpha_j, \alpha_k \rangle} w(x_j).$$

In conclusion, if we replace our original canonical generators and basis of simple roots by their (modified) images under $w \in \mathrm{Aut}(\mathfrak{g}^c)$, then for the parameters $\tilde{\lambda}_j$ of θ we have $\tilde{\lambda}_j = 1$ if $\pi(j) \neq j$ and $\tilde{\lambda}_j = \lambda_j \lambda_k^{\langle \alpha_j, \alpha_k \rangle}$ if $\pi(j) = j$.

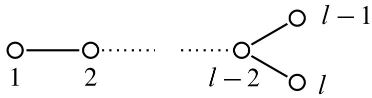
7.2. ***Inner type.*** First we suppose that $\mathfrak{g}$ is of inner type; that is, $\mathfrak{k}$ contains a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$. Let Φ be the root system of $\mathfrak{g}^c$ with respect to $\mathfrak{h}^c$, with basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$. Let $\{h_i, x_i, y_i \mid i\}$ be a canonical generating set corresponding to Δ. If $\mathfrak{g}'$ is not of inner type, that is, if a Cartan subalgebra of $\mathfrak{k}'$ is not a Cartan subalgebra of $\mathfrak{g}'$, then $\mathfrak{g}$ and $\mathfrak{g}'$ are not isomorphic. Hence, we assume that $\mathfrak{g}'$ is of inner type and define $\mathfrak{h}'$, Φ', and Δ' in the same way. Since $\mathfrak{g}^c$ and $(\mathfrak{g}')^c$ are isomorphic we may assume that Δ and Δ' are ordered so that the corresponding Cartan matrices are the same. Recall that each root space $\mathfrak{g}_\alpha$ with $\alpha \in \Phi$ lies either in $\mathfrak{k}^c$ or $\mathfrak{p}^c$; thus we have $\theta(x_i) = \lambda_i x_i$ with $\lambda_i \in \{\pm 1\}$ for all i. Let $\lambda'_1, \dots, \lambda'_l$ be defined similarly.

Suppose that we are in the situation $\lambda_i = \lambda'_i$ for all i, and write $\sigma(x_i) = r_i y_i$ and $\sigma(x'_i) = r'_i y'_i$. By Corollary 11, we have $\mathrm{sgn}(r_i) = \mathrm{sgn}(r'_i)$ for all i, which allows us to define the reals $\mu_i = \sqrt{r_i/r'_i}$. Now the isomorphism $\psi\colon \mathfrak{g}^c \to (\mathfrak{g}')^c$ which maps (h_i, x_i, y_i) to $(h'_i, \mu_i x'_i, \mu_i^{-1} y'_i)$ for all i satisfies $\psi \circ \theta = \theta' \circ \psi$ and $\psi \circ \sigma = \sigma' \circ \psi$, and we are done.

In the remainder of this section we show how to achieve $\lambda_i = \lambda'_i$ for all i in the case that $\mathfrak{g}$ and $\mathfrak{g}'$ are isomorphic. The idea is to use the results of Section 7.1 to find a new basis of simple roots such that θ and its parameters $\lambda_1, \dots, \lambda_l$ are in a *standard form*. As explained below, this means that at most one parameter λ_k is negative, with certain restrictions on k. The possible standard forms are obtained by listing the Kac diagrams of the inner involutions of $\mathfrak{g}^c$; to each Kac diagram corresponds exactly one standard form, and $\mathfrak{g}$ and $\mathfrak{g}'$ are isomorphic if and only if the standard forms of θ and θ' coincide.

In the following we explain this in detail for the simple Lie algebra of type D_l.

Example 13. Let the notation be as above and suppose $\mathfrak{g}^c$ is of type D_l with $l > 4$. We suppose that our basis of simple roots $\Delta = \{\alpha_1, \dots, \alpha_l\}$ corresponds to the labels of the following Dynkin diagram of D_l:



Up to conjugacy, the involutive inner automorphisms of $\mathfrak{g}^c$ are $\bar{\chi}_j$ with $j = 1, \dots, \lfloor l/2 \rfloor$ or $j = l-1$, where $\bar{\chi}_j(x_i) = (-1)^{\delta_{ij}} x_i$ for all i. If we do not have that $\theta = \bar{\chi}_j$ for some j, then we proceed as follows. Recall that the parameters of θ are $\lambda_1, \dots, \lambda_l$ where $\theta(x_i) = \lambda_i x_i$. For $k = 1, \dots, l$ write $w_k = s_{\alpha_k} \in W$ for the reflection defined by k-th simple root α_k. Let $\tilde{x}_i = w_k(x_i)$, $\tilde{y}_i = w_k(y_i)$, and $\tilde{h}_i = w_k(h_i)$ be the images of the canonical generators under w_k. As seen in Section 7.1, with respect to this new canonical generating set, θ has the same parameters as before, except that λ_j is replaced by $\lambda_j \lambda_k$ if $\langle \alpha_j, \alpha_k \rangle = -1$ (or, equivalently, if α_j and α_k are connected in the Dynkin diagram). We will now iterate this modification of parameters. We stress that in each iteration step the reflections $w_1, \dots, w_l$ are defined with respect to the new basis of simple roots constructed in the previous step; thus, acting with w_i and then with w_j means we first apply the reflection s_{α_i} and then the reflection $s_{w_i(\alpha_j)}$. Similarly, in each iteration step we have new parameters λ_i and a new canonical generating set. By abuse of notation, in each iteration step we always denote these by the same symbols.

We now show that we can apply a sequence of simple reflections to find a new set of canonical generators such that for the parameters of θ there is a unique $k \in \{1, \dots, \lfloor l/2 \rfloor, l-1, l\}$ with $\lambda_k = -1$; that is, $\theta = \bar{\chi}_k$. The details are as follows:

- The first step is to achieve that at most one of $\lambda_1, \dots, \lambda_{l-2}$ has value -1. If this is not already the case, then there exist $i < k \leq l-2$ with $\lambda_i, \lambda_k = -1$ and $\lambda_j = 1$ whenever $i < j < k$ or $k < j \leq l-2$. If we act with $w_i, w_{i+1}, \dots, w_{k-1}$, then we obtain new parameters of θ with $\lambda_{k-1} = -1$ and $\lambda_k = \cdots = \lambda_{l-2} = 1$. Now either $k-1$ is the only index in $\{1, \dots, l-2\}$ with $\lambda_{k-1} = -1$, or we iterate this process. Eventually, at most one value of $\lambda_1, \dots, \lambda_{l-2}$ is -1.

- Next, a case distinction with four cases $\lambda_{l-1}, \lambda_l \in \{\pm 1\}$ shows that we can in fact assume that at most one value of $\lambda_1, \ldots, \lambda_l$ is -1: For example, suppose $\lambda_i = -1$ with $i < l-1$ and $\lambda_{l-1} = -1$ are the only negative parameters. If we act with $w_{l-1}, w_{l-2}, \ldots, w_{i+1}$, then among the new parameters the only negative ones are $\lambda_{i+1} = -1$ and $\lambda_l = -1$. By an iteration, the only negative parameters are $\lambda_{l-2} = \lambda_l = -1$ (or $\lambda_{l-2} = \lambda_{l-1} = -1$), and acting with w_l (or w_{l-1}) yields the assertion.
- If the only negative parameter $\lambda_k = -1$ satisfies $k \in \{1, \ldots, \lfloor l/2 \rfloor, l-1, l\}$, then we are done. Thus, suppose we have $\lambda_k = -1$ with $\lfloor l/2 \rfloor < k \le l-2$. Let $t = l-k-1$, and act with $w_k, w_{k+1}, \ldots, w_{k+t}; w_{k-1}, w_k, \ldots, w_{k-1+t}; \ldots; w_1, w_2, \ldots, w_{1+t}$. This gives new parameters with only negative parameter $\lambda_{t+1} = -1$.

If $\lambda_l = -1$ is the only negative parameter, then we apply the diagram automorphism which fixes $\alpha_1, \ldots, \alpha_{l-1}$ and interchanges α_{l-1} and α_l; the resulting new basis of simple roots still defines the same Cartan matrix, and now we have $\theta = \bar{\chi}_{l-1}$. Thus, every inner automorphisms θ of order two can be brought into *standard form*; that is, there is exactly one negative parameter $\lambda_k = -1$, and $k \in \{1, \ldots, \lfloor l/2 \rfloor, l-1\}$.

Our approach for the other simple Lie algebras is the same: We act with the Weyl group (as described in Section 7.1) and certain diagram automorphisms to find a new basis of simple roots such that θ has standard form; that is, at most one parameter $\lambda_k = -1$ is negative, with the following restrictions: $k \le \lceil l/2 \rceil$ for A_l, $k = l$ or $k \le \lfloor l/2 \rfloor$ for C_l, $k = 1$ for G_2, $k \in \{2, 3\}$ for F_4, $k \in \{1, 2\}$ for E_6, $k \in \{1, 2, 7\}$ for E_7, and $k \in \{1, 8\}$ for E_8.

Remark 14. A more uniform approach to the problem of finding the standard form of θ is by using the classification of finite order inner automorphisms as, for example, given in [Reeder 2010]. In this approach one acts with the affine Weyl group, and finding the Kac diagram of an automorphism is equivalent to finding a point in the fundamental alcove conjugate to a given point. It can be worked out how acting by an element of the affine Weyl group amounts to choosing a different basis of simple roots. For the purposes of this paper, as we are dealing with involutions only, we have chosen the more elementary method outlined above.

7.3. ***Outer type.*** Suppose θ is an outer involutive automorphism of $\mathfrak{g}^c$. We apply the following four steps to $\mathfrak{g}$ (and then $\mathfrak{g}'$).

(1) The first step is to construct a θ-stable Cartan subalgebra of $\mathfrak{g}^c$: For this purpose let $\mathfrak{h}_0^c$ be a Cartan subalgebra of $\mathfrak{k}^c$ and define $\mathfrak{h}^c = C_{\mathfrak{g}^c}(\mathfrak{h}_0^c)$ as its centralizer in $\mathfrak{g}^c$. It is shown in [Knapp 2002, Proposition 6.60] that $\mathfrak{h}^c$ is a Cartan subalgebra of $\mathfrak{g}^c$; clearly, it is fixed by θ. Now $\mathfrak{h} = \mathfrak{h}^c \cap \mathfrak{g}$ is a *maximally compact* Cartan subalgebra of $\mathfrak{g}$ (see [Knapp 2002, Proposition 6.61]), and all

Cartan subalgebras of $\mathfrak{g}$ constructed this way are conjugate in $\mathfrak{g}$ (see [Knapp 2002, Proposition 6.61]). Thus, if $\mathfrak{g}$ and $\mathfrak{g}'$ are isomorphic and $\mathfrak{h}$ and $\mathfrak{h}'$ are Cartan subalgebras constructed as above, then there is an isomorphism $\mathfrak{g} \to \mathfrak{g}'$ which maps $\mathfrak{h}$ to $\mathfrak{h}'$.

(2) The second step is to construct a basis of simple roots which is stable under $\alpha \mapsto \alpha \circ \theta$: Let Φ be the root system with respect to $\mathfrak{h}^c$, and recall that, if $\alpha \in \Phi$, then $\beta = \alpha \circ \theta$ is a root with $\theta(x_\alpha) \in \mathfrak{g}_\beta$ and $\theta(h_\alpha) = h_\beta$. This shows that the $\mathbb{R}$-span $\mathfrak{h}_\mathbb{R}$ of all h_α with $\alpha \in \Phi$ is invariant under θ. Moreover, $h_{0,\mathbb{R}} = \mathfrak{h}_\mathbb{R} \cap \mathfrak{h}_0^c$ is the 1-eigenspace of θ in $\mathfrak{h}_\mathbb{R}$. Since $\mathfrak{h}_{0,\mathbb{R}}$ spans $\mathfrak{h}_0^c$ as a $\mathbb{C}$-vector space, the restriction of each $\alpha \in \Phi$ to $\mathfrak{h}_{0,\mathbb{R}}$ is nonzero: if $\alpha(\mathfrak{h}_{0,\mathbb{R}}) = \{0\}$, then $\mathfrak{g}_\alpha \subseteq C_{\mathfrak{g}^c}(\mathfrak{h}_0^c) = \mathfrak{h}^c$ yields a contradiction. This shows that there is $h_0 \in \mathfrak{h}_{0,\mathbb{R}}$ with $\alpha(h_0) \neq 0$ for all $\alpha \in \Phi$: such an h_0 can be chosen as any element outside a finite number of hyperplanes in $\mathfrak{h}_{0,\mathbb{R}}$, namely, the kernels of α in $h_{0,\mathbb{R}}$. We use h_0 to define $\alpha > 0$ if and only if $\alpha(h_0) > 0$; note that elements in $\mathfrak{h}_\mathbb{R}$ only have real eigenvalues. It is easy to check that this defines a root ordering, and, if $\alpha > 0$, then also $\alpha \circ \theta > 0$. Therefore the corresponding set of simple roots $\Delta = \{\alpha_1, \ldots, \alpha_l\}$ is θ-stable. Let π be the permutation of $\{1, \ldots, l\}$ defined by $\alpha_i \circ \theta = \alpha_{\pi(i)}$, and denote by $\{h_i, x_i, y_i \mid i\}$ a canonical generating set corresponding to Δ.

(3) The third step is to adjust the coefficients of θ: If $\pi(i) = i$, then set $\tilde{h}_i = h_i$, $\tilde{x}_i = x_i$ and $\tilde{y}_i = y_i$. Otherwise, for all $(i, \pi(i))$ with $\pi(i) > i$ set $\tilde{h}_i = h_i$, $\tilde{x}_i = x_i$, $\tilde{y}_i = y_i$, and $\tilde{h}_{\pi(i)} = \theta(h_i)$, $\tilde{x}_{\pi(i)} = \theta(x_i)$, $\tilde{y}_{\pi(i)} = \theta(y_i)$. By replacing $\{h_i, x_i, y_i \mid i\}$ with the canonical generating set $\{\tilde{h}_i, \tilde{x}_i, \tilde{y}_i \mid i\}$ (see Proposition 3), we may assume that $\theta(x_i) = \lambda_i x_{\pi(i)}$ with $\lambda_i = 1$ if $\pi(i) \neq i$.

(4) Finally, we decompose θ: Let φ be the diagram automorphism defined by π with respect to the new canonical generating set defined in (3) (see Section 2.2); that is, $\varphi(x_i) = x_{\pi(i)}$, $\varphi(y_i) = y_{\pi(i)}$, and $\varphi(h_i) = h_{\pi(i)}$ for all i. By construction,

$$\chi = \varphi \circ \theta = \theta \circ \varphi$$

is an involutive inner automorphism of $\mathfrak{g}^c$ with $\chi(x_i) = x_i$ if $\pi(i) \neq i$, and $\chi(x_i) = \lambda_i x_i$ if $\pi(i) = i$ and $\theta(x_i) = \lambda_i x_i$; clearly, $\lambda_i = \pm 1$. The analogous statement holds for y_i.

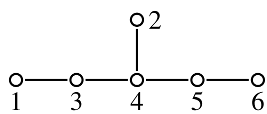
We use the same procedure to construct a θ'-stable set of positive roots Δ', and automorphisms φ' and χ' of $(\mathfrak{g}')^c$. We also assume that the bases Δ and Δ' are ordered such that the corresponding Cartan matrices are the same, and $\pi = \pi'$ as permutations of $\{1, \ldots, l\}$. If the latter is not possible, then $\mathfrak{g}^c$ and $(\mathfrak{g}')^c$ are not isomorphic. Let $\{h_i, x_i, y_i \mid i\}$ and $\{h_i', x_i', y_i' \mid i\}$ be the associated sets of canonical generators as constructed in Step (3) above, and let $\psi\colon \mathfrak{g}^c \to (\mathfrak{g}')^c$ be the associated isomorphism. We now try to modify ψ so that it is compatible with θ, θ' and σ, σ'.

7.3.1. *Make ψ compatible with θ and θ'.* Recall that $\theta = \chi \circ \varphi$, $\theta' = \chi' \circ \varphi'$, and $\psi \circ \varphi = \varphi' \circ \psi$. If $\pi(i) \neq i$, then

$$\theta' \circ \psi(x_i) = \theta'(x_i') = x'_{\pi'(i)} = x'_{\pi(i)} = \psi(x_{\pi(i)}) = \psi \circ \theta(x_i);$$

similarly, $\theta' \circ \psi$ and $\psi \circ \theta$ coincide on the whole subspace of $\mathfrak{g}^c$ spanned by all x_i, y_i, h_i with $\pi(i) \neq i$. If $\pi(i) = i$ with $\theta(x_i) = \lambda_i x_i$ and $\theta'(x_i') = \lambda_i' x_i'$, then $\theta' \circ \psi(x_i) = \psi \circ \theta(x_i)$ if and only if $\lambda_i = \lambda_i'$.

- Type A_l: If $l = 2m$ is even, then π acts fixed-point freely; thus ψ as constructed above already satisfies $\theta' \circ \psi = \psi \circ \theta$. If $l = 2m + 1$, then π has exactly one fixed point, say $i = 1$, and either $\lambda_1 = 1$ or $\lambda_1 = -1$. On the other hand, up to conjugacy, A_l has two outer automorphisms, so each choice for λ_1 corresponds to a different conjugacy class of automorphisms. Thus, if $\mathfrak{g}^c$ and $(\mathfrak{g}')^c$ are isomorphic, then $\lambda_1 = \lambda_1'$, and ψ is an isomorphism with $\theta' \circ \psi = \psi \circ \theta$.
- Type E_6: Here π has two fixed points, say $i = 2, 4$; thus there are four possible combinations of signs for λ_2 and λ_4. However, up to conjugacy, E_6 has two outer automorphisms. Suppose our root basis $\Delta = \{\alpha_1, \ldots, \alpha_6\}$ corresponds to the labels of the following Dynkin diagram of E_6:

  Up to conjugacy, E_6 has two outer automorphisms $\varphi \circ \bar{\chi}$, where φ is the diagram automorphism acting via the permutation $\pi = (1, 6)(3, 5)$, and $\bar{\chi}$ is an inner automorphisms which satisfies $\bar{\chi}(x_4) = \pm x_4$ and $\bar{\chi}(x_i) = x_i$ if $i \neq 4$. As outlined in Section 7.1 we now act with $w_2 = s_{\alpha_2}$ and $w_4 = s_{\alpha_4}$ in order to find a new canonical generating set (with respect to a new basis of simple roots), relative to which we have $\lambda_2 = \lambda_4 = 1$, or $\lambda_2 = 1$ and $\lambda_4 = -1$. It is straightforward to see that this can always be done. For example, if $\lambda_2 = -1$ and $\lambda_4 = 1$, then we first act with w_2 to get $\lambda_2 = \lambda_4 = -1$ and subsequently with w_4 to get $\lambda_2 = 1$ and $\lambda_4 = -1$. Finally we use the same trick as in the beginning of Section 7.3 to obtain $\lambda_i = 1$ for all $i \neq 2, 4$ (that is, we set $x_5 = \theta(x_3)$, etc.). The conclusion is that we can arrange that $\lambda_i = \lambda_i'$ for every i; hence ψ is an isomorphism with $\theta' \circ \psi = \psi \circ \theta$.
- Type D_l: We proceed as for E_6 and suppose that our basis of simple roots Δ corresponds to the Dynkin diagram of D_l as shown on page 370. Up to conjugacy, the involutive outer automorphisms of D_l are $\varphi \circ \bar{\chi}$, where φ is the diagram automorphism defined by $\pi = (l - 1, l)$, and $\bar{\chi}$ is an inner

automorphism with $\bar{\chi}(x_i) = \bar{\lambda}_i x_i$, where either $\bar{\lambda}_i = 1$ for all i, or there exists a unique negative $\bar{\lambda}_k$ and $k \in \{1, \dots, \lceil l/2 \rceil - 1\}$. As in Example 13, we act with reflections $s_{\alpha_j} \in W$, $j \in \{1, \dots, l-2\}$, to find a canonical generating set relative to which there is a unique negative parameter λ_k, and $k \in \{1, \dots, l-2\}$. If $k \le \lceil l/2 \rceil - 1$, then we are done; otherwise we proceed as follows. Set $\beta_i = \alpha_{l-i-1}$ for $i = 1, \dots, l-2$, and $\beta_{l-1} = -\alpha_1 - \dots - \alpha_{l-2} - \alpha_{l-1}$, and $\beta_l = -\alpha_1 - \dots - \alpha_{l-2} - \alpha_l$. Then $\bar{\Delta} = \{\beta_1, \dots, \beta_l\}$ is also a basis of simple roots with the same Dynkin diagram. Now we take a canonical generating set with respect to $\bar{\Delta}$. With respect to this new canonical generating set, χ has a unique negative parameter $\lambda_k = -1$, and $k \in \{1, \dots, \lceil l/2 \rceil - 1\}$.

Using these constructions, we can arrange that $\lambda_i = \lambda'_i$ for all i; thus the corresponding isomorphism ψ (defined on the newly constructed canonical generating sets) is compatible with θ and θ'.

7.3.2. *Make ψ compatible with σ and σ'.* Using the construction in the previous paragraphs, we have established that either $\mathfrak{g}$ and $\mathfrak{g}'$ are not isomorphic, or we have an isomorphism $\psi \colon \mathfrak{g}^c \to (\mathfrak{g}')^c$ compatible with θ and θ'. We assume the latter holds, and we now adjust ψ so it is also compatible with the complex conjugations σ and σ'; this yields the desired isomorphism between $\mathfrak{g}$ and $\mathfrak{g}'$.

By our previous construction, if $i \neq \pi(i)$, then $\theta(x_i) = x_{\pi(i)}$, and $\theta(x_i) = \lambda_i x_i$ otherwise. Lemma 7 shows that $\sigma(x_i) = r_i y_{\pi(i)}$ for some $r_i \in \mathbb{R}$ with $r_i = r_{\pi(i)}$. If $i \neq \pi(i)$, then $r_i < 0$, and, if $i = \pi(i)$, then r_i and $-\lambda_i$ have the same sign; see Corollary 11. Now define $\mu_i = \sqrt{1/|r_i|}$ for $i = 1, \dots, l$. If we replace $x_i, y_i, x_{\pi(i)}, y_{\pi(i)}$ by $\tilde{x}_i = \mu_i x_i$, $\tilde{y}_i = \mu_i^{-1} y_i$, $\tilde{x}_{\pi(i)} = \mu_i x_{\pi(i)}$, $\tilde{y}_{\pi(i)} = \mu_i^{-1} y_{\pi(i)}$, then we get a new set of canonical generators where θ acts in the same way and $\sigma(\tilde{x}_i) = \pm \tilde{y}_{\pi(i)}$ for all i. In a similar way, we obtain a new set of canonical generators $\{\tilde{x}'_i, \tilde{y}'_i, h'_i \mid i\}$ of $(\mathfrak{g}')^c$; recall that $\lambda_i = \lambda'_i$ for all i. The associated isomorphism $\mathfrak{g}^c \to (\mathfrak{g}')^c$ now is compatible with θ, σ, and θ', σ', and restricts to an isomorphism $\mathfrak{g} \to \mathfrak{g}'$ preserving the Cartan decompositions.

Remark 15. In the algorithms described in this section we compute root systems of $\mathfrak{g}^c$ and $(\mathfrak{g}')^c$ with respect to Cartan subalgebras $\mathfrak{h}^c$ and $(\mathfrak{h}')^c$. In order for that to work well we need Cartan subalgebras that split over $\mathbb{Q}\sqrt{}(\imath)$ (or an extension thereof of small degree). However, the problem of finding such Cartan subalgebras is very difficult; see [Ivanyos et al. 2012]. Therefore, in our algorithms we assume that we have a Cartan subalgebra with a small splitting field.

7.4. ***Nilpotent orbits under isomorphisms.*** Suppose $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and $\mathfrak{g}' = \mathfrak{k}' \oplus \mathfrak{p}'$ are semisimple real Lie algebras and $\psi \colon \mathfrak{g} \to \mathfrak{g}'$ is an isomorphism such that $\psi(\mathfrak{k}) = \mathfrak{k}'$ and $\psi(\mathfrak{p}) = \mathfrak{p}'$. As described in the previous sections, we can extend this to an isomorphism $\psi \colon \mathfrak{g}^c \to (\mathfrak{g})^c$ compatible with the corresponding Cartan involutions θ, θ'

and complex conjugations σ, σ'. Let G be the connected Lie subgroup of the adjoint group G^c of $\mathfrak{g}^c$, having Lie algebra $\mathfrak{g}$. Similarly, let G' be defined for $\mathfrak{g}'$.

Lemma 16. *The isomorphism $\psi\colon \mathfrak{g} \to \mathfrak{g}'$ maps nilpotent orbits to nilpotent orbits.*

Proof. Clearly, $e \in \mathfrak{g}$ is nilpotent if and only if $\psi(e) \in \mathfrak{g}'$ is nilpotent. We show that, if two nilpotent $e, f \in \mathfrak{g}$ are conjugate under G, then $e' = \psi(e)$ and $f' = \psi(f)$ are conjugate under G'; then the same argument with ψ replaced by ψ^{-1} proves the assertion. As shown in [Helgason 1978, pp. 126–127], the adjoint group G is generated by all $\exp \operatorname{ad} x$ with $x \in \mathfrak{g}$, and the isomorphism ψ lifts to an isomorphism $\tilde{\psi}\colon G \to G'$, $\beta \mapsto \psi \circ \beta \circ \psi^{-1}$. Thus, if $\beta(e) = f$ for some $\beta \in G$, then $\tilde{\psi}(\beta)(\psi(e)) = \psi(f)$, and $\psi(e)$ and $\psi(f)$ are G'-conjugate in $\mathfrak{g}'$. □

Appendix: Comment on the implementation

For computing with semisimple Lie algebras we use the package `SLA` [de Graaf 2012] for the computer algebra system `GAP` [GAP 2012]. This package provides the functionality, for example, to compute Chevalley bases, canonical generators, and involutive automorphisms. In Section 6.3 we use the Gröbner bases functionality of the computer algebra system SINGULAR [Decker et al. 2011] via the linkage package `Singular` [Costantini and de Graaf 2006].

A.1. ***The field*** $\mathbb{Q}^{\surd}$**.** We now comment on the field $\mathbb{Q}^{\surd} = \mathbb{Q}(\{\sqrt{p} \mid p \text{ a prime}\})$. `GAP` already allows us to work with subfields of cyclotomic fields $\mathbb{Q}(\zeta_n)$, where ζ_n is a complex primitive n-th root of unity. However, if $x = \sum_{i=1}^{m} \sqrt{p_i}$ for primes $p_1, \dots, p_m$, then the smallest n with $x \in \mathbb{Q}(\zeta_n)$ is $n = \operatorname{lcm}(e_1, \dots, e_m)$, where $e_i = p_i$ if $p_i \equiv 1 \bmod 4$, and $e_i = 4p_i$ otherwise; see Lemma 17. Thus, already for small m this requires to work in large cyclotomic fields. Alternatively, one could work in an algebraic extension defined by an irreducible polynomial over $\mathbb{Q}$. The disadvantage here is that we do not know in the beginning which irrationals turn up, so we would have to extend and therefore change the underlying field several times. To avoid all this, we have implemented our own realization of $\mathbb{Q}^{\surd}(\imath)$. Every element of $\mathbb{Q}^{\surd}(\imath)$ can be written uniquely as $u = \sum_{i=1}^{n} r_i \sqrt{z_i}$ where $z_i > 0$ are pairwise distinct squarefree integers and $r_i \in \mathbb{Q}(\imath)$. Internally, we represent u as a list with entries (r_i, z_i), which allows efficient addition and multiplication in $\mathbb{Q}^{\surd}(\imath)$. A computational bottleneck is the construction of the multiplicative inverse of such a $u \neq 0$: We compute powers $\{1, u, u^2, \dots, u^m\}$ until u^m can be expressed as a $\mathbb{Q}$-linear combination of $\{1, u, \dots, u^{m-1}\}$, say $u^m = \sum_{i=0}^{m-1} q_i u^i$. The minimal polynomial of u over $\mathbb{Q}$ is $f(x) = x^m - \sum_{i=0}^{m-1} q_i x^i = xg(x) + q_0$; therefore $u^{-1} = -g(u)/q_0$. Although all this can done with linear algebra, m can become rather large; see Lemma 19.

Often we had to deal with the following problem: Suppose $v \in \mathbb{Q}^{\surd}(\imath)$ is given as an element of $\mathbb{Q}(\zeta_n)$ for some n; write it as an element of $\mathbb{Q}^{\surd}(\imath)$, that is, $v = \sum_{i=1}^{m} r_i \sqrt{k_i}$ for pairwise distinct positive squarefree integers k_i and Gaussian rationals r_i. Clearly, it is sufficient to consider v real. The first step is to determine the set $\mathcal{S}$ of all positive squarefree k with $\sqrt{k} \in \mathbb{Q}(\zeta_n)$; we do this in Corollary 18. The second step is to make the ansatz $v = \sum_{k \in \mathcal{S}} r_k \sqrt{k}$ in $\mathbb{Q}(\zeta_n)$ with indeterminates $r_k \in \mathbb{Q}$. Linear algebra can be used to find a solution of this equation; we prove in Lemma 19 that such a solution always exists. We now provide the theoretical background of this approach; our starting point is the following lemma; see [Shirali and Yogananda 2004, p. 56, Proposition 3], and its corollary.

Lemma 17. *If p is an odd prime, then* $\sqrt{(-1)^{(p-1)/2} p} \in \mathbb{Q}(\zeta_p)$.

Corollary 18. *Let k and n be positive integers. Suppose k is squarefree and let e be the number of primes $p \equiv 3 \bmod 4$ dividing k.*

(a) *If $\sqrt{2} \in \mathbb{Q}(\zeta_n)$, then $8 \mid n$. If $\sqrt{k} \in \mathbb{Q}(\zeta_n)$, then $k \mid n$.*

(b) *If n is odd, then $\mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{2n})$, and $\sqrt{k} \in \mathbb{Q}(\zeta_n)$ if and only if e is even and $k \mid n$.*

(c) *If $4 \mid n$ and $8 \nmid n$, then $\sqrt{k} \in \mathbb{Q}(\zeta_n)$ if and only if k is odd and $k \mid n$.*

(d) *If $8 \mid n$, then $\sqrt{k} \in Q(\zeta_n)$ if and only if $k \mid n$.*

(e) *Let n be minimal with $\sqrt{k} \in \mathbb{Q}(\zeta_n)$. If k is odd and e is even, then $n = k$, and $n = 4k$ otherwise.*

Lemma 19. (a) *Let $n, k_1, \dots, k_m$ be pairwise distinct positive squarefree integers and suppose there exists a prime $p \mid n$ with $p \nmid k_i$ for all i. Then $\sqrt{n} \notin \mathbb{Q}(\sqrt{k_1}, \dots, \sqrt{k_m})$.*

(b) *Let $v = \sum_{i=1}^{m} r_i \sqrt{k_i} \in \mathbb{Q}^{\surd}$ for rational $r_i \neq 0$ and pairwise distinct positive squarefree integers k_i. Then v is a primitive element of $\mathbb{Q}(\sqrt{k_1}, \dots, \sqrt{k_m})$.*

Proof. (a) We use induction on m. The assertion clearly holds if $m = 1$; thus let $m > 1$ and write $K' = \mathbb{Q}(\sqrt{k_1}, \dots, \sqrt{k_{m-1}})$ and $K = K'(b)$ with $b = \sqrt{k_m}$. Suppose that $\sqrt{n} \in K$. Since $\sqrt{n} \notin K'$ by the induction hypothesis, $b \notin K'$ and, therefore, $\sqrt{n} = r + bs$ for unique $r, s \in K'$. Note that $s, r \neq 0$ since otherwise $\sqrt{n}$ or $\sqrt{nk_m}/k_m$ would lie in K', a contradiction. Now squaring yields $b = (n - r^2 - s^2 k_m)/(2rs) \in K'$, the final contradiction.

(b) Suppose $K = \mathbb{Q}(\sqrt{k_1}, \dots, \sqrt{k_m}) = \mathbb{Q}(\sqrt{k_1}, \dots, \sqrt{k_s})$ has degree $d = 2^s$ over $\mathbb{Q}$ with $s \leq m$. Since K is the splitting field of the separable polynomial $(x^2 - k_1) \cdots (x^2 - k_s)$, the extension is Galois and therefore $\mathcal{G} = \mathrm{Gal}(K/\mathbb{Q})$ has order d. Clearly, every map defined by $\sqrt{k_i} \mapsto \pm\sqrt{k_i}$ for $i = 1, \dots, s$ gives rise to a Galois automorphism, and an order argument shows that $\mathcal{G}$ consists exactly of these automorphisms. We now show that $1, \sqrt{k_1}, \dots, \sqrt{k_s}$ are linearly independent over $\mathbb{Q}$. Clearly, this is true for $s = 1$, so let $s \geq 2$. For a contradiction, assume (†) $\sum_{i=1}^{s} r_i \sqrt{k_i} + r_{m+1} = 0$ for rationals r_i. Let p be a prime dividing $k_1 \cdots k_s$. Now (†) implies that $\sqrt{p}$ lies in the field generated by $\sqrt{k_1'}, \dots, \sqrt{k_s'}$ with $k_i' = k_i / \gcd(k_i, p)$, contradicting part (a). Let f be the minimal polynomial of v over $\mathbb{Q}$. Clearly, $\gamma(v)$ is a root of f for every $\gamma \in \mathcal{G}$. Since $\sqrt{k_1}, \dots, \sqrt{k_s}$ are $\mathbb{Q}$-linearly independent, it follows that $\gamma(v) \neq \gamma'(v)$ for all $\gamma \neq \gamma'$ in $\mathcal{G}$. This shows that f has at least d different zeros, which implies that f has in fact degree d and v is primitive. □

References

[Bourbaki 1975] N. Bourbaki, *Éléments de mathématique: Fasc. XXXVIII: Groupes et algèbres de Lie. Chapitre VII: Sous-algèbres de Cartan, éléments réguliers. Chapitre VIII: Algèbres de Lie semi-simples déployées*, Actualités Scientifiques et Industrielles **1364**, Hermann, Paris, 1975. MR 56 #12077 Zbl 0329.17002

[Collingwood and McGovern 1993] D. H. Collingwood and W. M. McGovern, *Nilpotent orbits in semisimple Lie algebras*, Van Nostrand Reinhold, New York, 1993. MR 94j:17001 Zbl 0972.17008

[CoReLG 2012] H. Dietrich, P. Faccin, and W. A. de Graaf, "CoReLG - computing with real Lie groups", a GAP 4 package, 2012, Available at http://www.science.unitn.it/∼corelg.

[Costantini and de Graaf 2006] M. Costantini and W. A. de Graaf, "Singular: a GAP 4 package", The GAP Group, 2006, Available at http://www.gap-system.org/Packages/singular.html.

[Cox et al. 1992] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Springer, New York, 1992. MR 93j:13031 Zbl 0756.13017

[Decker et al. 2011] W. Decker, G.-M. Greuel, G. Pfister, and M. Schönemann, "Singular: a computer algebra system for polynomial computations", 2011, Available at http://www.singular.uni-kl.de. Version 3-1-3.

[Dietrich et al. 2013] H. Dietrich, P. Faccin, and W. A. de Graaf, "Computing with real Lie algebras: Real forms, Cartan decompositions, and Cartan subalgebras", *J. Symbolic Comput.* **56** (2013), 27–45. MR 3061707

[Đoković 1987] D. Ž. Đoković, "Proof of a conjecture of Kostant", *Trans. Amer. Math. Soc.* **302**:2 (1987), 577–585. MR 88j:17007 Zbl 0631.17004

[Đoković 1988] D. Ž. Đoković, "Classification of nilpotent elements in simple exceptional real Lie algebras of inner type and description of their centralizers", *J. Algebra* **112**:2 (1988), 503–524. MR 89b:17010 Zbl 0639.17005

[Đoković 1998] D. Ž. Đoković, "Explicit Cayley triples in real forms of G_2, F_4 and E_6", *Pacific J. Math.* **184**:2 (1998), 231–255. MR 99e:17005 Zbl 1040.17004

[Đoković 1999] D. Ž. Đoković, "Explicit Cayley triples in real forms of E_7", *Pacific J. Math.* **191**:1 (1999), 1–23. MR 2001c:22012 Zbl 1040.17006

[Đoković 2000] D. Ž. Đoković, "Explicit Cayley triples in real forms of E_8", *Pacific J. Math.* **194**:1 (2000), 57–82. MR 2001c:22013 Zbl 1013.22003

[Galina 2009] E. Galina, "Weighted Vogan diagrams associated to real nilpotent orbits", pp. 239–253 in *New developments in Lie theory and geometry* (Córdoba, 2007), edited by C. S. Gordon et al., Contemp. Math. **491**, Amer. Math. Soc., Providence, RI, 2009. MR 2010m:17007 Zbl 1180.22017

[GAP 2012] The GAP Group, "GAP: groups, algorithms, and programming", 2012, Available at http://www.gap-system.org. Version 4.5.5.

[de Graaf 2000] W. A. de Graaf, *Lie algebras: theory and algorithms*, North-Holland Mathematical Library **56**, North-Holland, Amsterdam, 2000. MR 2001j:17011 Zbl 1122.17300

[de Graaf 2011] W. A. de Graaf, "Computing representatives of nilpotent orbits of θ-groups", *J. Symbolic Comput.* **46**:4 (2011), 438–458. MR 2012f:17008 Zbl 1222.17002

[de Graaf 2012] W. A. de Graaf, "SLA: computing with simple Lie algebras, a GAP 4 package", 2012, Available at http://www.science.unitn.it/~degraaf/sla.html.

[Helgason 1978] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics **80**, Academic Press, New York, 1978. MR 80k:53081 Zbl 0451.53038

[Humphreys 1978] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, 2nd ed., Graduate Texts in Mathematics **9**, Springer, New York, 1978. MR 81b:17007 Zbl 0447.17001

[Ivanyos et al. 2012] G. Ivanyos, L. Rónyai, and J. Schicho, "Splitting full matrix algebras over algebraic number fields", *J. Algebra* **354** (2012), 211–223. MR 2879232 Zbl 06078371

[Jacobson 1962] N. Jacobson, *Lie algebras*, Interscience Tracts in Pure and Applied Mathematics **10**, Wiley Interscience, New York, 1962. MR 26 #1345 Zbl 0121.27504

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, Cambridge, 1990. MR 92k:17038 Zbl 0716.17022

[Knapp 2002] A. W. Knapp, *Lie groups beyond an introduction*, 2nd ed., Progress in Mathematics **140**, Birkhäuser, Boston, 2002. MR 2003c:22001 Zbl 1075.22501

[Kostant and Rallis 1971] B. Kostant and S. Rallis, "Orbits and representations associated with symmetric spaces", *Amer. J. Math.* **93** (1971), 753–809. MR 47 #399 Zbl 0224.22013

[Littelmann 1996] P. Littelmann, "An effective method to classify nilpotent orbits", pp. 255–269 in *Algorithms in algebraic geometry and applications* (Santander, 1994), edited by L. González-Vega and T. Recio, Progress in Mathematics **143**, Birkhäuser, Basel, 1996. MR 98a:20041 Zbl 0866.20037

[Noël 1998] A. G. Noël, "Nilpotent orbits and theta-stable parabolic subalgebras", *Represent. Theory* **2** (1998), 1–32. MR 99g:17023 Zbl 0891.17006

[Noël 2001a] A. G. Noël, "Classification of admissible nilpotent orbits in simple exceptional real Lie algebras of inner type", *Represent. Theory* **5** (2001), 455–493. MR 2002i:17015a Zbl 1005.17006

[Noël 2001b] A. G. Noël, "Classification of admissible nilpotent orbits in simple real Lie algebras $E_{6(6)}$ and $E_{6(-26)}$", *Represent. Theory* **5** (2001), 494–502. MR 2002i:17015b Zbl 1005.17005

[Onishchik 2004] A. L. Onishchik, *Lectures on real semisimple Lie algebras and their representations*, European Mathematical Society, Zürich, 2004. MR 2005b:17014 Zbl 1080.17001

[Reeder 2010] M. Reeder, "Torsion automorphisms of simple Lie algebras", *Enseign. Math.* (2) **56**:1-2 (2010), 3–47. MR 2012b:17040 Zbl 1223.17020

[Rothschild 1972] L. P. Rothschild, "Orbits in a real reductive Lie algebra", *Trans. Amer. Math. Soc.* **168** (1972), 403–421. MR 50 #2271 Zbl 0222.17009

[Sekiguchi 1987] J. Sekiguchi, "Remarks on real nilpotent orbits of a symmetric pair", *J. Math. Soc. Japan* **39**:1 (1987), 127–138. MR 88g:53053 Zbl 0627.22008

[Shirali and Yogananda 2004] S. Shirali and C. S. Yogananda, *Number theory*, Universities Press, Hyderabad, 2004.

[Vinberg 1979] È. B. Vinberg, "Classification of homogeneous nilpotent elements of a semisimple graded Lie algebra", *Trudy Sem. Vektor. Tenzor. Anal.* **19** (1979), 155–177. In Russian; translated in *Selecta Mathematica Sovietica* **6**:1 (1987), 15–35. MR 80k:17006 Zbl 0431.17006

[Vinberg et al. 1990] È. B. Vinberg, V. V. Gorbatsevich, and A. L. Onishchik, Строение групп и алгебр ли, Encyclopaedia of Mathematical Sciences **41**, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1990. In Russian; translated as *Lie groups and Lie algebras, III: Structure of Lie groups and Lie algebras*, by V. V. Minakhin, Springer, Berlin, 1994. MR 91b:22001 Zbl 0733.22003

Received July 27, 2012. Revised September 17, 2012.

HEIKO DIETRICH
SCHOOL OF MATHEMATICAL SCIENCES
MONASH UNIVERSITY
CLAYTON, VIC 3800
AUSTRALIA
heiko.dietrich@monash.edu

WILLEM A. DE GRAAF
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TRENTO
I-38050 POVO
ITALY
degraaf@science.unitn.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.381

LANDAU–TOEPLITZ THEOREMS FOR SLICE REGULAR FUNCTIONS OVER QUATERNIONS

GRAZIANO GENTILI AND GIULIA SARFATTI

The theory of slice regular functions of a quaternionic variable extends the notion of holomorphic function to the quaternionic setting. This theory, already rich in results, is sometimes surprisingly different from the theory of holomorphic functions of a complex variable; however, several fundamental results in the two environments are similar even if their proofs for the case of quaternions need new technical tools.

In this paper we prove the Landau–Toeplitz theorem for slice regular functions in a formulation that involves an appropriate notion of *regular* 2-*diameter*. We show that the Landau–Toeplitz inequalities hold in the case of the regular n-diameter for all $n \geq 2$. Finally, a 3-diameter version of the Landau–Toeplitz theorem is proved using the notion of *slice* 3-*diameter*.

1. Introduction

The Schwarz lemma, in its different flavors, is the basis of a chapter of fundamental importance in the geometric theory of holomorphic functions of one and several complex variables. Its classic formulation in one variable is the following:

Theorem 1.1 (Schwarz lemma). *Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disc of $\mathbb{C}$ centered at the origin, and let $f : \mathbb{D} \to \mathbb{D}$ be a holomorphic function such that $f(0) = 0$. Then*

$$|f(z)| \leq |z| \tag{1}$$

for all $z \in \mathbb{D}$, and

$$|f'(0)| \leq 1. \tag{2}$$

Equality holds in (1) for some $z \in \mathbb{D} \setminus \{0\}$, or in (2), if and only if there exists $u \in \mathbb{C}$ with $|u| = 1$ such that $f(z) = uz$ for all $z \in \mathbb{D}$.

Gentili and Sarfatti were partially supported by GNSAGA of the INdAM and by PRIN "Proprietà geometriche delle varietà reali e complesse" of the MIUR. Sarfatti was partially supported by PRIN "Geometria differenziale e analisi globale" of the MIUR.

MSC2010: primary 30G35; secondary 30C80.

Keywords: functions of hypercomplex variables, geometric theory of regular functions of a quaternionic variable, Schwarz lemma and generalizations.

The Schwarz lemma and its extension due to Pick lead in a natural way to the construction of the Poincaré metric, which plays a key role in the study of the hyperbolic geometry of complex domains and manifolds. In the same year of the first formulation of the Schwarz lemma, the Landau–Toeplitz theorem [1907] was proven. This less known but quite interesting result concerns the study of the possible shapes of the image of the unit disc under a holomorphic function and it is formulated in terms of the diameter of the image set.

Theorem 1.2 (Landau–Toeplitz [1907]; see also Burckel et al. 2006). *Let f be holomorphic in $\mathbb{D}$ and such that the diameter* $\operatorname{diam} f(\mathbb{D})$ *of* $f(\mathbb{D})$ *equals* 2. *Then*

$$\operatorname{diam} f(r\mathbb{D}) \leq 2r \tag{3}$$

for all $r \in (0, 1)$ *and*

$$|f'(0)| \leq 1. \tag{4}$$

Equality holds in (3) *for some* $r \in (0, 1)$, *or in* (4), *if and only if f is of the form* $f(z) = a + bz$, *with* $a, b \in \mathbb{C}$ *and* $|b| = 1$.

This result can be interpreted as a generalization of the classical Schwarz lemma in which the diameter of the image set takes over the role of the maximum modulus of the function; indeed, there exist infinite subsets of the plane that have constant diameter and are different from a disc; the Reuleaux polygons are a well-known example of such sets [Gardner 2006; Lachand-Robert and Oudet 2007].

The recent definition of slice regularity for quaternionic functions of one quaternionic variable, inspired by Cullen [1965] and developed in [Gentili and Struppa 2006; 2007], identifies a large class of functions, which includes natural quaternionic power series and polynomials. The study of a geometric theory for this class of functions has by now produced several interesting results, sometimes analogous to those valid for holomorphic functions; the Schwarz lemma is among these results [Gentili and Struppa 2007], together with the Bohr theorem and the Bloch–Laudau theorem [Della Rocchetta et al. 2012; 2013; Sarfatti 2013].

Fairly new developments in the theory of holomorphic functions of one complex variable include the analogue of the Schwarz lemma for meromorphic functions, and open new fascinating perspectives for future research. In this setting, Solynin [2008] recalls into the scenery the approach of Landau and Toeplitz and its modern reinterpretation and generalization due to Burckel, Marshall, Minda, Poggi-Corradini and Ransford [Burckel et al. 2008].

In our paper, we first prove an analogue of the Landau–Toeplitz theorem for slice regular functions. To this purpose we need to introduce a new tool to “measure” the image of the open unit ball $\mathbb{B}$ of the space of quaternions $\mathbb{H}$ through a slice regular function, the *regular diameter*.

Definition 1.3. Let f be a slice regular function on $\mathbb{B} = \{q \in \mathbb{H} : |q| < 1\}$ and let

$$f(q) = \sum_{n\geq 0} q^n a_n$$

be its power series expansion. For $r \in (0, 1)$, we define the *regular diameter* of the image of $r\mathbb{B}$ under f as

$$\tilde{d}_2(f(r\mathbb{B})) = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |f_u(q) - f_v(q)|,$$

where

$$f_u(q) = \sum_{n\geq 0} q^n u^n a_n \quad \text{and} \quad f_v(q) = \sum_{n\geq 0} q^n v^n a_n.$$

We define the *regular diameter* of the image of $\mathbb{B}$ under f as

$$\tilde{d}_2(f(\mathbb{B})) = \lim_{r\to 1^-} \tilde{d}_2(f(r\mathbb{B})).$$

The introduction of this new geometric quantity is necessary because of the peculiarities of the quaternionic environment, and in particular since a composition of slice regular functions is not slice regular in general. The regular diameter can play the role of the diameter; in fact, the former is finite if and only if the latter is finite. The regular diameter hence appears in the statement of the announced result.

Theorem 3.9 (Landau–Toeplitz for regular functions). *Let f be a slice regular function on $\mathbb{B}$ such that $\tilde{d}_2(f(\mathbb{B})) = 2$ and let $\partial_c f(0)$ be its* slice derivative *in* 0. *Then*

$$\tilde{d}_2(f(r\mathbb{B})) \leq 2r \tag{5}$$

for all $r \in (0, 1)$, and

$$|\partial_c f(0)| \leq 1. \tag{6}$$

Equality holds in (5) for some $r \in (0, 1)$, or in (6), if and only if f is an affine function; *that is, $f(q) = a + qb$ with $a, b \in \mathbb{H}$ and $|b| = 1$.*

As in the complex setting, this theorem can be interpreted as a generalization of the Schwarz lemma.

The new version of the Landau–Toeplitz theorem proposed in [Burckel et al. 2008] concerns holomorphic functions whose image is measured with a notion of diameter more general than the classic one, the n-diameter. In the quaternionic setting, the analogue of this geometric quantity is defined:

Definition 1.4. Let $E \subset \mathbb{H}$. For every $n \in \mathbb{N}$, $n \geq 2$, the *n-diameter* of E is defined as

$$d_n(E) = \sup_{w_1,\ldots,w_n\in E} \left(\prod_{1\leq j<k\leq n} |w_k - w_j| \right)^{\frac{2}{n(n-1)}}.$$

Retracing the approach used in the complex setting, we are able to obtain only the generalization of the first part of the statement of the Landau–Toeplitz theorem for the n-diameter. As in the case $n = 2$, we need a notion of regular n-diameter $\tilde{d}_n(f(\mathbb{B}))$ for the image of $\mathbb{B}$ through a slice regular function f. This notion is a generalization of Definition 1.3 modeled on Definition 1.4 and given in terms of the $*$-product between slice regular functions (see Section 2). For all $n \geq 2$, the regular n-diameter turns out to be finite when the n-diameter is finite. For this reason, even if it may appear awkward, it makes sense to use the regular n-diameter in the following statement:

Theorem 1.5. *Let f be a slice regular function on $\mathbb{B}$ such that $\tilde{d}_n(f(\mathbb{B})) = d_n(\mathbb{B})$. Then*

$$\tilde{d}_n(f(r\mathbb{B})) \leq d_n(r\mathbb{B}) \quad \textit{for all } r \in (0, 1),$$

and

$$|\partial_c f(0)| \leq 1.$$

Since the 3-diameter of a 4-dimensional subset of $\mathbb{H}$ is attained on a (specific) bidimensional section, we are encouraged to introduce an appropriate notion $\hat{d}_3 f(\mathbb{B})$ of *slice* 3-*diameter* for $f(\mathbb{B})$ inspired by the power series expansion of the regular 3-diameter. This leads to the following complete result:

Theorem 5.7 (Landau–Toeplitz theorem for the slice 3-diameter). *Let f be a slice regular function on $\mathbb{B}$ such that $\hat{d}_3(f(\mathbb{B})) = d_3(\mathbb{B})$. Then*

$$\hat{d}_3(f(r\mathbb{B})) \leq d_3(r\mathbb{B}) \tag{7}$$

for every $r \in (0, 1)$, and

$$|\partial_c f(0)| \leq 1. \tag{8}$$

Equality holds in (7) for some $r \in (0, 1)$, or in (8), if and only if f is an affine function $f(q) = a + qb$ with $a, b \in \mathbb{H}$ and $|b| = 1$.

We point out that all the extensions of the Landau–Toeplitz results presented in this paper generalize the Schwarz lemma to a much larger class of image sets; in fact, for all $n \geq 2$ there exist infinitely many subsets of the space $\mathbb{H}$ which have fixed n-diameter, do not coincide with a 4-ball, and neither contain nor are contained in the 4-ball. The 4-bodies of constant width are examples of such subsets, presented for instance in [Gardner 2006; Lachand-Robert and Oudet 2007].

2. Preliminaries

Let $\mathbb{H}$ be the skew field of quaternions obtained by endowing $\mathbb{R}^4$ with the multiplication operation defined on the standard basis $\{1, i, j, k\}$ by $i^2 = j^2 = k^2 = -1$ and $ij = k$, and then extended by distributivity to all quaternions

$q = x_0 + x_1 i + x_2 j + x_3 k$. For every $q \in \mathbb{H}$, we define the *real* and *imaginary* part of q as $\operatorname{Re} q = x_0$ and $\operatorname{Im} q = x_1 i + x_2 j + x_3 k$, its *conjugate* as $\bar{q} = \operatorname{Re} q - \operatorname{Im} q$, and its *modulus* as $|q|^2 = q\bar{q}$. The multiplicative inverse of $q \neq 0$ is $q^{-1} = \bar{q}/|q|^2$. Let $\mathbb{S}$ be the unit 2-sphere of purely imaginary quaternions, $\mathbb{S} = \{q \in \mathbb{H} \mid q^2 = -1\}$. Then for any $I \in \mathbb{S}$, we will denote by L_I the complex plane $\mathbb{R} + \mathbb{R}I$, and if $\Omega \subset \mathbb{H}$, we further set $\Omega_I = \Omega \cap L_I$. Notice that to every $q \in \mathbb{H} \setminus \mathbb{R}$, we can associate a unique element in $\mathbb{S}$ by the map $q \mapsto \operatorname{Im}(q)/|\operatorname{Im}(q)| = I_q$; therefore, for any $q \in \mathbb{H} \setminus \mathbb{R}$ there exist and are unique $x, y \in \mathbb{R}$ with $y > 0$ and $I_q \in \mathbb{S}$ such that $q = x + yI_q$. If q is real then I_q can be any element of $\mathbb{S}$.

The preliminary results stated in this section will be given for slice regular functions defined on open balls of type $B = B(0, R) = \{q \in \mathbb{H} \mid |q| < R\}$. We point out that in most cases these results hold, with appropriate changes, for a more general class of domains introduced in [Colombo et al. 2009]. Let us now recall the definition of slice regularity.

Definition 2.1. A function $f : B = B(0, R) \to \mathbb{H}$ is said to be *slice regular* (often abbreviated to *regular* later on) if for every $I \in \mathbb{S}$, its restriction f_I to B_I has continuous partial derivatives and satisfies

$$\bar{\partial}_I f(x + yI) = \frac{1}{2}\left(\frac{\partial}{\partial x} + I\frac{\partial}{\partial y}\right) f_I(x + yI) = 0 \quad \text{for every } x + yI \in B_I.$$

In the sequel we may refer to the vanishing of $\bar{\partial}_I f$ by saying that the restriction f_I is holomorphic on B_I. In what follows, for the sake of shortness we will omit the prefix slice when referring to slice regular functions. A notion of derivative, called *slice* (*or Cullen*) *derivative*, can be given for regular functions by

$$\partial_c f(x + yI) = \frac{\partial}{\partial x} f(x + yI) \quad \text{for } x + yI \in B.$$

This definition is well-posed because it is applied only to regular functions; moreover, slice regularity is preserved by slice differentiation. A basic result connects slice regularity and classical holomorphy:

Lemma 2.2 (splitting lemma; see [Gentili and Struppa 2007]). *If f is a regular function on $B = B(0, R)$ then for every $I \in \mathbb{S}$ and for every $J \in \mathbb{S}$ with J orthogonal to I, there exist two holomorphic functions $F, G : B_I \to L_I$ such that*

$$f_I(z) = F(z) + G(z)J \quad \textit{for every } z = x + yI \in B_I.$$

Theorem 2.3 [Gentili and Struppa 2007]. *A function f is regular on $B = B(0, R)$ if and only if f has a power series expansion*

$$f(q) = \sum_{n \geq 0} q^n a_n \quad \textit{with} \quad a_n = \frac{1}{n!}\frac{\partial^n f}{\partial x^n}(0)$$

converging absolutely and uniformly on compact sets in $B(0, R)$.

The next two results will be needed later.

Theorem 2.4 (identity principle, weak version; see [Gentili and Struppa 2007]). *Let $f : B = B(0, R) \to \mathbb{H}$ be a regular function. Denote by Z_f the zero set of f, $Z_f = \{q \in B \mid f(q) = 0\}$. If there exists $I \in \mathbb{S}$ such that $B_I \cap Z_f$ has an accumulation point in B_I then f vanishes identically on B.*

Theorem 2.5 (representation formula; see [Colombo et al. 2009]). *Let f be a regular function on $B = B(0, R)$ and let $J \in \mathbb{S}$. Then for all $x + yI \in B$, the following equality holds*:

$$f(x+yI) = \tfrac{1}{2}\big[f(x+yJ) + f(x-yJ)\big] + I\tfrac{1}{2}\big[J\big(f(x-yJ) - f(x+yJ)\big)\big].$$

The product of two regular functions is not, in general, regular. To guarantee regularity we need to introduce the following multiplication operation:

Definition 2.6. Let $f(q) = \sum_{n\geq 0} q^n a_n$ and $g(q) = \sum_{n\geq 0} q^n b_n$ be regular functions on $B = B(0, R)$. The $*$-*product* of f and g is the regular function

$$f * g : B \to \mathbb{H}$$

defined by

$$f * g(q) = \sum_{n\geq 0} q^n \sum_{k=0}^{n} a_k b_{n-k}.$$

The $*$-product is associative but not, in general, commutative. The following result clarifies the relation between the $*$-product and the pointwise product of regular functions.

Proposition 2.7 [Gentili et al. 2013]. *Let $f(q) = \sum\limits_{n\geq 0} q^n a_n$ and $g(q) = \sum\limits_{n\geq 0} q^n b_n$ be regular functions on $B = B(0, R)$. Then*

$$f * g(q) = \begin{cases} f(q)g(f(q)^{-1}qf(q)) & \textit{if } f(q) \neq 0, \\ 0 & \textit{if } f(q) = 0. \end{cases}$$

Notice that if $q = x + yI$ (and if $f(q) \neq 0$) then $f(q)^{-1}qf(q)$ has the same modulus and same real part as q; hence, $T_f(q) = f(q)^{-1}qf(q)$ lies in $x + y\mathbb{S}$, the same 2-sphere as q. A zero $x_0 + y_0 I$ of the function g is not necessarily a zero of $f * g$, but one element on the same sphere $x_0 + y_0\mathbb{S}$ is.

To conclude this section we recall a result that is basic for our purposes.

Theorem 2.8 (maximum modulus principle [Gentili and Struppa 2007]). *Let $f : B \to \mathbb{H}$ be a regular function. If there exists $I \in \mathbb{S}$ such that the restriction $|f_I|$ has a local maximum in B_I then f is constant in B. In particular, if $|f|$ has a local maximum in B then f is constant in B.*

3. The Landau–Toeplitz theorem for regular functions

In this section, we will prove our analogue of the Landau–Toeplitz theorem for (quaternionic) regular functions. To reach the aim, we will need a few steps.

Denote by $\langle\ ,\ \rangle$ the scalar product of $\mathbb{R}^4$, and by $\times$ the vector product of $\mathbb{R}^3$. Recall the equality $uv = -\langle u, v\rangle + u \times v$, valid for purely imaginary quaternions u, v. Also, if $w = x + yL \in \mathbb{H}$, then $\langle w, I\rangle = \langle yL, I\rangle = -\operatorname{Re}(yLI) = -\operatorname{Re}(wI)$ for all $I \in \mathbb{S}$.

Definition 3.1. Let $I \in \mathbb{S}$. For any $w \in \mathbb{H}$, we define the imaginary component of w along I as $\operatorname{Im}_I(w) = \langle w, I\rangle = -\operatorname{Re}(wI)$.

Proposition 3.2. *Let $w \in B = B(0, R)$, $0 < |w| = r < R$, and let g be a holomorphic function on $B \cap L_{I_w}$. If*

$$g(w) = w \quad \textit{and} \quad r = \max_{z \in r\overline{\mathbb{B}}_{I_w}} |g(z)| \tag{9}$$

then $\operatorname{Im}_{I_w}(\partial_c g(w)) = 0$.

Proof. To simplify the notation, set $I = I_w$. Define $\varphi : \mathbb{R} \to \mathbb{R}$ by $\varphi(\theta) = |g(we^{I\theta})|^2$. The splitting lemma (2.2) implies that for every $J \in \mathbb{S}$ orthogonal to I there exist holomorphic functions $F, G : B_I \to L_I$ such that $g(z) = F(z) + G(z)J$ for every $z \in B_I$. A direct computation shows that $\varphi(\theta) = F(we^{I\theta})\overline{F(we^{I\theta})} + G(we^{I\theta})\overline{G(we^{I\theta})}$; hence,

$$\varphi'(\theta) = -2\operatorname{Im}_I \big(we^{I\theta}\big(F'(we^{I\theta})\overline{F(we^{I\theta})} + G'(we^{I\theta})\overline{G(we^{I\theta})}\big)\big),$$

where F' and G' are the complex derivatives of F and G in B_I. Since, by hypothesis, $\theta = 0$ is a maximum for φ, we have

$$0 = \varphi'(0) = -2\operatorname{Im}_I \big(w\big(F'(w)\overline{F(w)} + G'(w)\overline{G(w)}\big)\big). \tag{10}$$

Moreover, $w = g(w) = F(w) + G(w)J$, which implies $F(w) = w$ and $G(w) = 0$. Putting these values in (10), we have $0 = -2\operatorname{Im}_I(wF'(w)\bar{w}) = -2|w|^2 \operatorname{Im}_I(F'(w))$, which yields $\operatorname{Im}_I(F'(w)) = 0$. Finally, recalling the definition of the slice derivative and Definition 3.1, we get

$$\operatorname{Im}_I(\partial_c g(w)) = \operatorname{Im}_I\big(F'(w) + G'(w)J\big) = \operatorname{Im}_I(F'(w)) = 0. \qquad \square$$

Remark 3.3. The proposition can be interpreted as a consequence of the Julia–Wolff–Carathéodory theorem (see for instance [Abate 1989; Burckel 1979]); in fact, the hypotheses in (9) yield that $g : r\mathbb{B}_I \to r\mathbb{B}$ and that w is a boundary fixed point for the restriction of g to $r\mathbb{B}_I$; hence, if we split the function g as $g(z) = F(z) + G(z)J$, for $z \in r\mathbb{B}_I$, we have that w is a Wolff point for $F : r\mathbb{B}_I \to r\mathbb{B}_I$.

The proof of the classical Landau–Toeplitz theorem in the setting of holomorphic maps [Burckel et al. 2008] relies upon the analogue of Proposition 3.2, which is not sufficient for our purposes in the quaternionic environment; in fact, we need the following:

Proposition 3.4. *Let $g : \mathbb{B} \to \mathbb{H}$ be a regular function such that $\mathrm{Im}_{I_q}(g(q)) = 0$ for every $q \in \mathbb{B}$. Then g is a real constant function.*

Proof. Let $g(q) = \sum_{n \geq 0} q^n a_n$ on $\mathbb{B}$. For any $I \in \mathbb{S}$, we split the coefficient a_n as $b_n + c_n J$ with $b_n, c_n \in L_I$ and $J \in \mathbb{S}$ orthogonal to I. By hypothesis, we have

$$0 = \mathrm{Im}_I(g(z)) = \mathrm{Im}_I\Bigl(\sum_{n \geq 0} z^n (b_n + c_n J)\Bigr) = \mathrm{Im}_I\Bigl(\sum_{n \geq 0} z^n b_n\Bigr) \quad \text{for all } z \in \mathbb{B}_I.$$

As a consequence of the open mapping theorem, the holomorphic map $\sum_{n \geq 0} z^n b_n$ is constant; that is, $b_n = 0$ for all $n > 0$. Therefore, the component of each a_n along L_I vanishes for all $n > 0$. Since $I \in \mathbb{S}$ is arbitrary, this implies $a_n = 0$ for all $n > 0$. The hypothesis yields that $a_0 \in \mathbb{R}$. □

A basic notion used to state the classical Landau–Toeplitz theorem is the diameter of the images of holomorphic functions. In the new quaternionic setting, due to the fact that composition of regular functions is not regular in general, the definition of a "regular" diameter for the images of regular functions requires a peculiar approach.

Definition 3.5. Let $f : \mathbb{B} \to \mathbb{H}$ be a regular function $f(q) = \sum_{n \geq 0} q^n a_n$, and let $u \in \mathbb{H}$. We define the *regular composition* of f with the function $q \mapsto qu$ as

$$f_u(q) = \sum_{n \geq 0} (qu)^{*n} a_n = \sum_{n \geq 0} q^n u^n a_n.$$

If $|u| = 1$, the radius of convergence of the series expansion for f_u is the same as that for f. Moreover, if u and q_0 lie in the same plane L_I then u and q_0 commute, hence $f_u(q_0) = f(q_0 u)$. In particular, if $u \in \mathbb{R}$ then $f_u(q) = f(qu)$ for every q.

Definition 3.6. Let $f : \mathbb{B} \to \mathbb{H}$ be a regular function. For $r \in (0, 1)$, we define the *regular diameter* of the image of $r\mathbb{B}$ under f as

$$\tilde{d}_2(f(r\mathbb{B})) = \max_{u,v \in \overline{\mathbb{B}}} \max_{|q| \leq r} |f_u(q) - f_v(q)|.$$

We define the *regular diameter* of the image of $\mathbb{B}$ under f as

$$\tilde{d}_2(f(\mathbb{B})) = \lim_{r \to 1^-} \tilde{d}_2(f(r\mathbb{B})). \tag{11}$$

Remark 3.7. By the maximum modulus principle for regular functions, $\tilde{d}_2(f(r\mathbb{B}))$ is an increasing function of r; hence the limit (11) always exists. So $\tilde{d}_2(f(\mathbb{B}))$ is well defined.

Let E be a subset of $\mathbb{H}$. We will denote by $\operatorname{diam} E = \sup_{q,w\in E}|q-w|$ the classical diameter of E.

Proposition 3.8. *Let f be a regular function on $\mathbb{B}$. Then*

$$\operatorname{diam} f(\mathbb{B}) \le \tilde{d}_2(f(\mathbb{B})) \le 2\operatorname{diam} f(\mathbb{B}).$$

Proof. To prove the first inequality, let $r\in(0,1)$ and consider $q,w\in r\overline{\mathbb{B}}$. We want to bound $|f(q)-f(w)|$. Suppose without loss of generality that $|w|\ge|q|$ and $w\neq 0$. Then

$$|f(q)-f(w)| = \left|f\left(q\frac{|w|}{|w|}\right)-f\left(w\frac{|w|}{|w|}\right)\right| = \left|f_{\frac{q}{|w|}}(|w|)-f_{\frac{w}{|w|}}(|w|)\right|, \tag{12}$$

where the last equality is due to the fact that $|w|$, being real, commutes with both $q/|w|$ and $w/|w|$. Since $q/|w|\in\overline{\mathbb{B}}$ and $w/|w|\in\partial\mathbb{B}$, (12) yields

$$\begin{aligned}|f(q)-f(w)| &\le \max_{u,v\in\overline{\mathbb{B}}}|f_u(|w|)-f_v(|w|)|\\ &\le \max_{u,v\in\overline{\mathbb{B}}}\max_{|q|\le r}|f_u(q)-f_v(q)| = \tilde{d}_2(f(r\mathbb{B})).\end{aligned}$$

This implies that $\operatorname{diam} f(r\overline{\mathbb{B}})\le\tilde{d}_2(f(r\mathbb{B}))$. Since this inequality holds for any $r\in(0,1)$, we obtain

$$\operatorname{diam} f(\mathbb{B}) = \lim_{r\to1^-}\operatorname{diam} f(r\overline{\mathbb{B}}) \le \lim_{r\to1^-}\tilde{d}_2(f(r\mathbb{B})) = \tilde{d}_2(f(\mathbb{B})).$$

To show the missing inequality, let $u,v\in\overline{\mathbb{B}}$, $r\in(0,1)$, and let J,K be elements of $\mathbb{S}$ such that $u\in L_J$ and $v\in L_K$. Using the representation formula (see Theorem 2.5) and taking into account that u and $x+yJ$ commute as well as v and $x+yK$, we get, for all $q=x+yI\in r\overline{\mathbb{B}}$,

$$\begin{aligned}&|f_u(q)-f_v(q)|\\ &=\tfrac12|(f((x+yJ)u)-(f(x+yK)v))+(f((x-yJ)u)-f((x-yK)v))\\ &\quad+IJ(f((x-yJ)u)-f((x+yJ)u))-IK(f((x-yK)v)-f((x+yK)v))|\\ &\le\tfrac12|f((x+yJ)u)-(f(x+yK)v)|+\tfrac12|f((x-yJ)u)-f((x-yK)v)|\\ &\quad+\tfrac12|f((x-yJ)u)-f((x+yJ)u)|+\tfrac12|f((x-yK)v)-f((x+yK)v)|\\ &\le 2\operatorname{diam} f(r\mathbb{B}).\end{aligned} \tag{13}$$

Since inequality (13) holds for every $u,v\in\overline{\mathbb{B}}$ and for every $q\in r\overline{\mathbb{B}}$, we get

$$\tilde{d}_2(f(r\mathbb{B})) = \max_{u,v\in\overline{\mathbb{B}}}\max_{|q|\le r}|f_u(q)-f_v(q)| \le 2\operatorname{diam} f(r\mathbb{B}); \tag{14}$$

and since this holds for every $r\in(0,1)$, we get $\tilde{d}_2(f(\mathbb{B}))\le 2\operatorname{diam} f(\mathbb{B})$. ☐

Notice that if f is an affine function, say $f(q) = a + qb$, then for every $r \in (0, 1)$ we have $\tilde{d}_2(f(r\mathbb{B})) = |b| \operatorname{diam} r\mathbb{B} = |b| r \operatorname{diam} \mathbb{B}$. In particular, if f is constant then $\tilde{d}_2(f(r\mathbb{B})) = 0$. Moreover, the regular diameter $\tilde{d}_2(f(r\mathbb{B}))$ is invariant under translations; in fact, if $g(q) = f(q) - f(0)$ then $\tilde{d}_2(g(r\mathbb{B})) = \tilde{d}_2(f(r\mathbb{B}))$ for every $r \in (0, 1)$.

Theorem 3.9 (Landau–Toeplitz for regular functions). *Let $f : \mathbb{B} \to \mathbb{H}$ be a regular function such that $\tilde{d}_2(f(\mathbb{B})) = \operatorname{diam} \mathbb{B} = 2$. Then*

$$\tilde{d}_2(f(r\mathbb{B})) \leq 2r \tag{15}$$

for every $r \in (0, 1)$, and

$$|\partial_c f(0)| \leq 1. \tag{16}$$

Equality holds in (15) for some $r \in (0, 1)$, or in (16), if and only if f is an affine function $f(q) = a + qb$ with $a, b \in \mathbb{H}$ and $|b| = 1$.

Proof. To prove the first inequality, take $u, v \in \overline{\mathbb{B}}$ and consider the auxiliary function

$$g_{u,v}(q) = \tfrac{1}{2} q^{-1}(f_u(q) - f_v(q)).$$

This function is regular on $\mathbb{B}$; indeed, if the power series expansion of f in $\mathbb{B}$ is $\sum_{n\geq 0} q^n a_n$, then

$$g_{u,v}(q) = \tfrac{1}{2} q^{-1}\Bigl(\sum_{n\geq 0} q^n u^n a_n - \sum_{n\geq 0} q^n v^n a_n\Bigr) = \tfrac{1}{2} \sum_{n\geq 0} q^n (u^{n+1} - v^{n+1}) a_{n+1}.$$

From this expression of $g_{u,v}$ we can recover its value at $q = 0$:

$$g_{u,v}(0) = \tfrac{1}{2}(u - v) a_1 = \tfrac{1}{2}(u - v) \partial_c f(0). \tag{17}$$

Since $g_{u,v}$ is a regular function, using the maximum modulus principle we get that

$$r \mapsto \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |g_{u,v}(q)|$$

is increasing on $(0, 1)$. Moreover, the regularity of the function $q \mapsto f_u(q) - f_v(q)$ yields that for any fixed $r \in (0, 1)$ we can write

$$\max_{|q|\leq r} |g_{u,v}(q)| = \max_{|q|\leq r} \frac{|f_u(q) - f_v(q)|}{2|q|} = \frac{1}{2r} \max_{|q|\leq r} |f_u(q) - f_v(q)|,$$

which leads to

$$\frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = \frac{1}{2r} \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |f_u(q) - f_v(q)| = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |g_{u,v}(q)|; \tag{18}$$

therefore, $\tilde{d}_2(f(r\mathbb{B}))/2r$ is an increasing function of r and so always less than or equal to the limit

$$\lim_{r\to 1^-} \frac{1}{2r}\tilde{d}_2(f(r\mathbb{B})) = \frac{1}{2}\tilde{d}_2(f(\mathbb{B})) = 1.$$

This means that

$$\tilde{d}_2(f(r\mathbb{B})) \le 2r \quad \text{for every } r \in (0,1), \tag{19}$$

proving inequality (15). To prove (16), consider the odd part of f,

$$f_{\mathrm{odd}}(q) = \frac{f(q) - f(-q)}{2}.$$

It satisfies the hypotheses of the Schwarz lemma for regular functions (see [Gentili and Struppa 2007]); indeed, f_{odd} is a regular function on $\mathbb{B}$, $f_{\mathrm{odd}}(0) = 0$, and

$$|f_{\mathrm{odd}}(q)| = \tfrac{1}{2}|f(q) - f(-q)| \le \tfrac{1}{2}\tilde{d}_2(f(\mathbb{B})) = 1 \quad \text{for every } q \in \mathbb{B};$$

hence,

$$\begin{aligned} 1 \ge |\partial_c f_{\mathrm{odd}}(0)| &= \tfrac{1}{2}|\partial_c f(q) - \partial_c(f(-q))|\big|_{q=0} \\ &= \tfrac{1}{2}|\partial_c f(q) + \partial_c f(-q)|\big|_{q=0} = |\partial_c f(0)|. \end{aligned} \tag{20}$$

We will now prove the last part of the statement, covering the case of equality. To begin with, notice that if $f(q) = a + qb$ with $a, b \in \mathbb{H}$ and $|b| = 1$, then equality holds in both (15) and (16).

Conversely, suppose that equality holds in (16), so $|\partial_c f(0)| = 1$. In this case we have $|\partial_c f_{\mathrm{odd}}(0)| = 1$; therefore, by the Schwarz lemma (see [Gentili and Struppa 2007]),

$$f_{\mathrm{odd}}(q) = q\partial_c f(0). \tag{21}$$

We want to show that in this case $\tilde{d}_2(f(r\mathbb{B})) = 2r$ for every $r \in (0,1)$; in fact, from (17) and (18) it follows that

$$\frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} \ge \max_{u,v\in\overline{\mathbb{B}}} |g_{u,v}(0)| = \max_{u,v\in\overline{\mathbb{B}}} \tfrac{1}{2}|(u-v)\partial_c f(0)| = 1 \quad \text{for every } r \in (0,1).$$

Comparing the last inequality with (19) we get

$$\tilde{d}_2(f(r\mathbb{B})) = 2r \quad \text{for every } r \in (0,1). \tag{22}$$

We now introducd a new auxiliary function. Take $w \in \mathbb{B}$ with $0 < |w| = r < 1$ and set

$$h_w(q) = \tfrac{1}{2}(f(q) - f(-w))\partial_c f(0)^{-1}.$$

The function h_w is regular on $\mathbb{B}$ and fixes w; indeed,

$$h_w(w) = \tfrac{1}{2}(f(w) - f(-w))\partial_c f(0)^{-1} = f_{\mathrm{odd}}(w)\partial_c f(0)^{-1} = w,$$

where the last equality is due to (21). We need now to restrict our attention to what happens in L_{I_w}. By the maximum modulus principle (Theorem 2.8), we are able to find $z_0 \in L_{I_w}$, $|z_0| = r$ such that for $z \in L_{I_w}$,

$$\max_{|z|\le r} |h_w(z)| = \tfrac{1}{2} \max_{|z|\le r} |f(z) - f(-w)| = \tfrac{1}{2}|f(z_0) - f(-w)|.$$

Let $\hat{u} \in L_{I_w}$ with $|\hat{u}| = 1$ be such that $-w = z_0\hat{u}$. Then again for $z \in L_{I_w}$, due to the fact that z_0 and $\hat{u}$ commute,

$$\begin{aligned}\max_{|z|\le r} |h_w(z)| &= \tfrac{1}{2}|f(z_0) - f(z_0\hat{u})| \\ &= \tfrac{1}{2}|f(z_0) - f_{\hat{u}}(z_0)| \le \tfrac{1}{2} \max_{u,v\in\overline{\mathbb{B}}} \max_{|z|\le r} |f_u(z) - f_v(z)|.\end{aligned}$$

Recalling (22) for $z \in L_{I_w}$ and $q \in \mathbb{H}$ we obtain

$$\max_{|z|\le r} |h_w(z)| \le \tfrac{1}{2} \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\le r} |f_u(q) - f_v(q)| = \tfrac{1}{2}\tilde{d}_2(f(r\mathbb{B})) = r = |h_w(w)|.$$

The function h_w then satisfies the hypotheses of Proposition 3.2; hence,

$$0 = \mathrm{Im}_{I_w}(\partial_c h_w(q)|_{q=w}) = \mathrm{Im}_{I_w}\big(\tfrac{1}{2}\partial_c f(w)\partial_c f(0)^{-1}\big).$$

Now recall that w is an arbitrary element of $\mathbb{B} \setminus \{0\}$. By continuity, we get that the function $w \mapsto \frac{1}{2}\partial_c f(w)\partial_c f(0)^{-1}$, regular on $\mathbb{B}$, satisfies the hypotheses of Proposition 3.4. Consequently, $\frac{1}{2}\partial_c f(w)\partial_c f(0)^{-1}$ is a real constant function hence $\partial_c f(w)$ is constant as well; therefore, f has the required form $f(q) = f(0) + q\partial_c f(0)$.

We will show now how equality in (15) for some $s \in (0, 1)$ implies equality in (16); this and the preceding step will conclude the proof. Suppose that there exists $s \in (0, 1)$ such that $\tilde{d}_2(f(s\mathbb{B}))/2s = 1$. By (19) and since $\tilde{d}_2(f(r\mathbb{B}))/2r$ is increasing in r, we have

$$\frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = 1 \quad \text{for every } r \in [s, 1).$$

Let us prove that this equality holds for all $r \in (0, 1)$. Let $\hat{u}, \hat{v} \in \overline{\mathbb{B}}$ be such that

$$\frac{\tilde{d}_2(f(s\mathbb{B}))}{2s} = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\le s} |g_{u,v}(q)| = \max_{|q|\le s} |g_{\hat{u},\hat{v}}(q)|,$$

where the first equality follows from (18). Let $r > s$. By the choice of $\hat{u}, \hat{v} \in \overline{\mathbb{B}}$, we get

$$1 = \frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\le r} |g_{u,v}(q)| \ge \max_{|q|\le r} |g_{\hat{u},\hat{v}}(q)| \ge \max_{|q|\le s} |g_{\hat{u},\hat{v}}(q)| = 1.$$

By the maximum modulus principle, the function $g_{\hat{u},\hat{v}}$ must be constant in $q \in \mathbb{B}$ and equal to 1 in modulus. Now consider $r \in (0, s)$. Then

$$1 \geq \frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |g_{u,v}(q)| \geq \max_{|q|\leq r} |g_{\hat{u},\hat{v}}(q)| = 1,$$

which implies $\tilde{d}_2(f(r\mathbb{B}))/2r = 1$ for every $r \in (0,1)$. The claim is now that $|\partial_c f(0)| = 1$. By (20), we first of all obtain

$$\lim_{r\to 0^+} \frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = 1 \geq |\partial_c f(0)|. \tag{23}$$

Recalling that

$$\frac{\tilde{d}_2(f(r\mathbb{B}))}{2r} = \max_{u,v\in\overline{\mathbb{B}}} \max_{|q|\leq r} |g_{u,v}(q)|,$$

we can get, for every $n \in \mathbb{N}$, the existence of $u_n, v_n \in \overline{\mathbb{B}}$ and q_n with $|q_n| = \frac{1}{n}$ (converging up to subsequences), such that

$$1 = \lim_{n\to\infty} \frac{\tilde{d}_2(f(\frac{1}{n}\mathbb{B}))}{2\frac{1}{n}} = \lim_{n\to\infty} |g_{u_n,v_n}(q_n)| = |g_{\tilde{u},\tilde{v}}(0)| \leq \max_{u,v\in\overline{\mathbb{B}}} |g_{u,v}(0)| = |\partial_c f(0)|.$$

(The last equality is due to (17).) A comparison with (23) concludes the proof. □

4. The *n*-diameter case

We next prove the n-diameter version of the Landau–Toeplitz theorem for regular functions. Recall from Definition 1.4 the definition of the n-diameter of a subset of $\mathbb{H}$. As in the complex case (see [Burckel et al. 2008]), we have:

Proposition 4.1. *For all $n \geq 2$, we have $d_n(E) \leq d_2(E) = \operatorname{diam} E$. Moreover, $d_n(E)$ is finite if and only if $d_2(E)$ is finite.*

As we did in Section 3, in the case of the classical diameter d_2, we will adopt a specific definition for the n-diameter of the image of a subset of $\mathbb{H}$ under a regular function. We will always consider images of open balls of the form $r\mathbb{B}$.

Definition 4.2. Let $n \geq 2$ and let f be a regular function on $\mathbb{B}$. For $r \in (0,1)$ we define, in terms of the $*$-product, the *regular n-diameter* of the image of $r\mathbb{B}$ under f as

$$\tilde{d}_n(f(r\mathbb{B})) = \max_{w_1,\ldots,w_n\in\overline{\mathbb{B}}} \max_{|q|\leq r} \left| \prod^{*}_{1\leq j<k\leq n} (f_{w_k}(q) - f_{w_j}(q)) \right|^{\frac{2}{n(n-1)}}.$$

We define the *regular n-diameter* of the image of $\mathbb{B}$ under f as

$$\tilde{d}_n(f(\mathbb{B})) = \lim_{r\to 1^-} \tilde{d}_n(f(r\mathbb{B})).$$

The same argument used for the regular diameter in Remark 3.7 guarantees that $\tilde{d}_n(f(\mathbb{B}))$ is well-defined. Notice that because of the noncommutativity of quaternions, the order of the factors of a $*$-product has its importance. We can choose any order we like, but it has to be fixed once chosen. In what follows, when we write $1 \le j < k \le n$ we always mean to order the couples (j,k) with the lexicographic order. To simplify the notation, we will sometimes write $j < k$, meaning $1 \le j < k \le n$.

The first step toward understanding the relation between the n-diameter and the regular n-diameter is the following result:

Proposition 4.3. *Let $f : \mathbb{B} \to \mathbb{H}$ be a regular function, and let $n \ge 2$. Then $\tilde{d}_n(f(\mathbb{B})) \le \tilde{d}_2(f(\mathbb{B}))$.*

Proof. We omit the (technical) proof. The idea is to turn the $*$-product into a usual product with an iterated application of Proposition 2.7. □

Notice that Proposition 4.1 and Proposition 4.3 imply that if $d_n(f(\mathbb{B}))$ is finite then $\tilde{d}_n(f(\mathbb{B}))$ is finite as well (for any regular function f and $n \ge 2$).

Let us make some simple remarks about the definition of regular n-diameter. As for the case $n = 2$, the regular n-diameter is invariant under translation; in fact, if f is a regular function on $\mathbb{B}$ and g is defined as $g(q) = f(q) - f(0)$ then $\tilde{d}_n(g(r\mathbb{B})) = \tilde{d}_n(f(r\mathbb{B}))$. Moreover, if $f(q) = qb$ with $b \in \mathbb{H}$ then $\tilde{d}_n(f(r\mathbb{B})) = |b| d_n(r\mathbb{B})$. In particular, if f is constant then $\tilde{d}_n(f(r\mathbb{B})) = 0$; hence, if f is of the form $f(q) = a + qb$ for some quaternions a and b then the regular n-diameter of $f(r\mathbb{B})$ coincides with its n-diameter.

In order to obtain analogues of inequalities (15) and (16) in the n-diameter case, we study the ratio between the regular n-diameter of the image of $r\mathbb{B}$ under a regular function f and the n-diameter of the domain $r\mathbb{B}$ of f.

Lemma 4.4. *Let f be a regular function on $\mathbb{B}$ and let $n \in \mathbb{N}$, $n \ge 2$. Then*

$$\varphi_n(r) = \frac{\tilde{d}_n(f(r\mathbb{B}))}{d_n(r\mathbb{B})} = \frac{\tilde{d}_n(f(r\mathbb{B}))}{d_n(\mathbb{B})r}$$

is an increasing function of r on the open interval $(0,1)$, *and*

$$\lim_{r \to 0^+} \varphi_n(r) = |\partial_c f(0)|.$$

Proof. If f is a constant or affine function, $\varphi_n(r)$ is a constant function. So let f be neither constant nor affine. Fix $w_1, \dots, w_n \in \overline{\mathbb{B}}$ and consider the auxiliary function

$$g_{w_1,\dots,w_n}(q) = d_n(\mathbb{B})^{-\frac{n(n-1)}{2}} q^{-\frac{n(n-1)}{2}} \prod_{1 \le j < k \le n}^{*} (f_{w_k}(q) - f_{w_j}(q)).$$

Since $f_{w_j}(0) = f(0)$ for every $j = 1, \dots, n$, we get that $g_{w_1,\dots,w_n}$ is regular on $\mathbb{B}$.

Moreover, using the maximum modulus principle as in (18), we can write

$$\varphi_n(r)^{\frac{n(n-1)}{2}} = \max_{w_1,\dots,w_n\in\overline{\mathbb{B}}} \max_{|q|\le r} |g_{w_1,\dots,w_n}(q)|;$$

hence, we can conclude that $\varphi_n(r)$ is increasing in r.

In turn, to prove the second part of the statement, we proceed as follows:

$$\begin{aligned}\lim_{r\to 0^+} \varphi_n(r) &= \lim_{r\to 0^+} \frac{\tilde{d}_n(f(r\mathbb{B}))}{d_n(\mathbb{B})r} \\ &= \lim_{r\to 0^+} d_n(\mathbb{B})^{-1} r^{-1} \max_{w_1,\dots,w_n\in\overline{\mathbb{B}}} \max_{|q|\le r} \left| \prod^{*}_{j<k} (f_{w_k}(q) - f_{w_j}(q)) \right|^{\frac{2}{n(n-1)}}.\end{aligned}$$

We can turn the $*$-product into a usual product with an iterated application of Proposition 2.7 (omitting the points where some factor $f_{w_k}(q) - f_{w_j}(q)$ vanishes), thus obtaining

$$\begin{aligned}&\lim_{r\to 0^+} \varphi_n(r) \\ &\quad= \lim_{r\to 0^+} d_n(\mathbb{B})^{-1} r^{-1} \max_{w_1,\dots,w_n\in\overline{\mathbb{B}}} \max_{|q|\le r} \left| \prod_{j<k} \big(f_{w_k}(T_{j,k}(q)) - f_{w_j}(T_{j,k}(q)) \big) \right|^{\frac{2}{n(n-1)}},\end{aligned}$$

where for all $j<k$, $T_{j,k}(q)$ is a suitable quaternion belonging to the same sphere $\operatorname{Re} q + |\operatorname{Im} q|\mathbb{S}$ of q. Since for every $j<k$ we have $|T_{k,j}(q)| = |q|$ if $|q| = r$, using the power series expansion of f we can write

$$\begin{aligned}&\lim_{r\to 0^+} \varphi_n(r) \\ &\quad= \lim_{r\to 0^+} d_n(\mathbb{B})^{-1} \max_{w_1,\dots,w_n\in\overline{\mathbb{B}}} \max_{|q|\le r} \prod_{j<k} \left| \sum_{n\ge 1} (T_{k,j}(q))^{n-1} (w_k^n - w_j^n) a_n \right|^{\frac{2}{n(n-1)}}.\end{aligned}$$

Since $\varphi_n(r)$ is lowerbounded by 0 and it is increasing in r then the limit of $\varphi_n(r)$ as r goes to 0 always exists. Proceeding as in the proof of Theorem 3.9, we can find a sequence of points $\{q_m\}_{m\in\mathbb{N}}$ such that $|q_m| = \frac{1}{m}$ for any $m \in \mathbb{N}$, and a sequence of n-tuples $\{(w_{1,m},\dots,w_{n,m})\}_{m\in\mathbb{N}} \subset \overline{\mathbb{B}}^n$ converging to some $(\hat{w}_1,\dots,\hat{w}_n) \in \overline{\mathbb{B}}^n$ such that

$$\lim_{m\to\infty} \varphi_n\Big(\frac{1}{m}\Big) = d_n(\mathbb{B})^{-1} \prod_{j<k} \left| \sum_{n\ge 1} (T_{k,j}(0))^{n-1} (\hat{w}_k^n - \hat{w}_j^n) a_n \right|^{\frac{2}{n(n-1)}};$$

therefore, by Definition 1.4 we obtain

$$\lim_{m\to\infty} \varphi_n\left(\frac{1}{m}\right) = d_n(\mathbb{B})^{-1} |a_1| \prod_{j<k} |(\hat{w}_k - \hat{w}_j)|^{\frac{2}{n(n-1)}} \le |a_1| = |\partial_c f(0)|.$$

To prove the opposite inequality, notice that for every choice of $\{\tilde{w}_1,\dots,\tilde{w}_n\}\subset\overline{\mathbb{B}}$,

$$\lim_{r\to 0^+}\max_{w_1,\dots,w_n\in\overline{\mathbb{B}}}\max_{|q|=r}\prod_{j<k}\left|\sum_{n\geq 1}(T_{k,j}(q))^{n-1}(w_k^n-w_j^n)a_n\right|^{\frac{2}{n(n-1)}}$$
$$\geq\lim_{r\to 0^+}\max_{|q|=r}\prod_{j<k}\left|\sum_{n\geq 1}(T_{k,j}(q))^{n-1}(\tilde{w}_k^n-\tilde{w}_j^n)a_n\right|^{\frac{2}{n(n-1)}},$$

whence

$$\lim_{r\to 0^+}\max_{w_1,\dots,w_n\in\overline{\mathbb{B}}}\max_{|q|=r}\prod_{j<k}\left|\sum_{n\geq 1}(T_{k,j}(q))^{n-1}(w_k^n-w_j^n)a_n\right|^{\frac{2}{n(n-1)}}$$
$$\geq\max_{\tilde{w}_1,\dots,\tilde{w}_n\in\overline{\mathbb{B}}}\lim_{r\to 0^+}\max_{|q|=r}\prod_{j<k}\left|\sum_{n\geq 1}(T_{k,j}(q))^{n-1}(\tilde{w}_k^n-\tilde{w}_j^n)a_n\right|^{\frac{2}{n(n-1)}};$$

therefore, we conclude that

$$\lim_{r\to 0^+}\varphi_n(r)\geq d_n(\mathbb{B})^{-1}\max_{w_1,\dots,w_n\in\overline{\mathbb{B}}}\prod_{j<k}|(w_j-w_k)a_1|^{\frac{2}{n(n-1)}}=|a_1|=|\partial_c f(0)|.\quad\square$$

Using Lemma 4.4, one easily proves the following result:

Theorem 4.5. *let f be a regular function on $\mathbb{B}$ such that $\tilde{d}_n(f(\mathbb{B}))=d_n(\mathbb{B})$. Then*

$$\tilde{d}_n(f(r\mathbb{B}))\leq d_n(r\mathbb{B})\tag{24}$$

for every $r\in(0,1)$, and

$$|\partial_c f(0)|\leq 1.\tag{25}$$

We believe that if equality holds in (24) for some $r\in(0,1)$ or in (25) then f is affine, but we were not able to prove this statement. On the one hand, it is easy to see that if f is affine, $f(q)=a+qb$ with $a,b\in\mathbb{H}$, $|b|=1$ then equality holds both in (24) and in (25); on the other hand, we do not yet know, in general, if the converse holds using the notion of regular n-diameter (for $n>2$).

5. A 3-diameter version of the Landau–Toeplitz theorem

In this section we prove a complete 3-diameter version of the Landau–Toeplitz theorem. The proof relies upon the elementary fact that three points lie always in the same plane. For this reason, the 3-diameter of a subset of $\mathbb{H}$, which has dimension 4, is always attained on a bidimensional section of the set. To compute the 3-diameter of the unit ball of $\mathbb{H}$ we need to recall a preliminary result about what happens in the complex case (for a proof, see [Burckel et al. 2008], for instance). Let $\mathbb{D}$ be the open unit disc of $\mathbb{C}$.

Lemma 5.1. *Given n points $\{w_1, \dots, w_n\} \subset \overline{\mathbb{D}}$, we have*

$$\prod_{1 \le j < k \le n} |w_j - w_k| \le n^{\frac{n}{2}}.$$

Equality holds if and only if (after relabeling) $w_j = u\alpha^j$, where $u \in \mathbb{S}^1$ and $\alpha = e^{i2\pi/n}$ is an n-th root of unity.

Lemma 5.2. *Fix any $I \in \mathbb{S}$ and $u \in \partial\mathbb{B}$. The 3-diameter of the unit ball of $\mathbb{H}$ is*

$$d_3(\mathbb{B}) = (|\alpha_2 - \alpha_1||\alpha_3 - \alpha_1||\alpha_3 - \alpha_2|)^{1/3},$$

where $\alpha_j = ue^{I2\pi j/3}$ for $j = 1, 2, 3$.

Proof. The result can be easily proved showing that the 3-diameter is attained on a maximal disc that, without loss of generality, can be chosen to be some $\mathbb{B}_I$. □

In particular, $d_3(\mathbb{B}) = d_3(\mathbb{D}) = \sqrt{3}$. To prove our 3-diameter version of the Landau–Toeplitz theorem, we introduce an appropriate notion of "slicewise" 3-diameter, inspired by the power series expansion of the regular 3-diameter.

Definition 5.3. Let $f : \mathbb{B} \to \mathbb{H}$ be a regular function, and let $\sum_{n\ge0} q^n a_n$ be its power series expansion. If a_N is the first nonvanishing coefficient, let $\hat{f}$ be the function obtained by multiplying f (on the right) by $a_N^{-1}|a_N|$:

$$\hat{f}(q) = \sum_{n\ge0} q^n a_n a_N^{-1} |a_N| = \sum_{n\ge0} q^n b_n.$$

This is regular on $\mathbb{B}$ as well. For any $I \in \mathbb{S}$, let w_1, w_2, w_3 be points in the closed disc $\overline{\mathbb{B}}_I$, and consider the function

$$\hat{g}_{w_1,w_2,w_3}(z) = \sum_{n\ge0} z^n \sum_{k=0}^{n} \sum_{j=0}^{k} \big(w_2^j - w_1^j\big)\big(w_3^{k-j} - w_1^{k-j}\big)\big(w_3^{n-k} - w_2^{n-k}\big) b_j b_{k-j} b_{n-k},$$

which is holomorphic in all variables z, w_1, w_2, w_3 on $\mathbb{B}_I$. We define the *slice 3-diameter* of $f(r\mathbb{B})$ by

$$\hat{d}_3(f(r\mathbb{B})) = \sup_{I\in\mathbb{S}} \max_{w_1,w_2,w_3\in\overline{\mathbb{B}}_I} \max_{z\in r\overline{\mathbb{B}}} \left|\hat{g}_{w_1,w_2,w_3}(z)\right|^{1/3}, \tag{26}$$

and the *slice 3-diameter* of $f(\mathbb{B})$ as the limit

$$\hat{d}_3(f(\mathbb{B})) = \lim_{r\to1^-} \hat{d}_3(f(r\mathbb{B})).$$

By the maximum modulus principle (Theorem 2.8), the function $r \mapsto \hat{d}_3(f(r\mathbb{B}))$ is increasing; hence, the previous definition is well posed. It is not difficult to prove that $\hat{g}_{w_1,w_2,w_3}(z)$ is continuous as a function of I and of the real and imaginary parts of z, w_1, w_2, w_3; hence, the supremum in (26) is actually a maximum.

Remark 5.4. For any regular function $f:\mathbb{B}\to\mathbb{H}$, the slice 3-diameter $\hat{d}_3(f(\mathbb{B}))$ is the same as the slice 3-diameter $\hat{d}_3((f-f(0))(\mathbb{B}))=\hat{d}_3(f(\mathbb{B})-f(0))$. Moreover, it is easy to prove that if the slice 3-diameter $\hat{d}_3(f(\mathbb{B}))$ vanishes then f is constant.

Lemma 5.5. *In analogy with what happens in the regular n-diameter case, let f be a regular function on $\mathbb{B}$, and for $r\in(0,1)$ let $\hat{\varphi}_3(r)$ be the ratio defined as*

$$\hat{\varphi}_3(r)=\frac{\hat{d}_3(f(r\mathbb{B}))}{d_3(r\mathbb{B})}=\frac{\hat{d}_3(f(r\mathbb{B}))}{d_3(\mathbb{B})r}.$$

Then $\hat{\varphi}_3(r)$ is increasing in r and

$$\lim_{r\to 0^+}\hat{\varphi}_3(r)=|\partial_c f(0)|.$$

Proof. One proves that

$$\hat{\varphi}_3(r)^3=d_3(\mathbb{B})^{-3}\max_{I\in\mathbb{S}}\max_{w_1,w_2,w_3\in\overline{\mathbb{B}}_I}\max_{z\in r\overline{\mathbb{B}}_I}\left|z^{-3}\hat{g}_{w_1,w_2,w_3}(z)\right|,$$

(see Definition 5.3) and uses the technique of the proof of Lemma 4.4 on each slice. □

The fundamental tool to prove the "equality case" is the following:

Theorem 5.6. *Let f be a regular function on $\mathbb{B}$ and for $r\in(0,1)$, let*

$$\hat{\varphi}_3(r)=\frac{\hat{d}_3(f(r\mathbb{B}))}{d_3(\mathbb{B})r}.$$

Then $\hat{\varphi}_3(r)$ is strictly increasing in r except if f is a constant or affine function; that is, if $f(q)=a+qb$ with $a,b\in\mathbb{H}$.

Proof. Thanks to Remark 5.4 we can suppose $f(0)=0$. Since $\hat{\varphi}_3(r)$ is increasing for $r\in(0,1)$, if it is not strictly increasing then there exist s,t, $0<s<t<1$ such that $\hat{\varphi}_3$ is constant on $[s,t]$. We will show that this yields that $\hat{\varphi}_3$ is constant on $(0,t]$. Let $I\in\mathbb{S}$ and $w_1,w_2,w_3\in\overline{\mathbb{B}}_I$ be such that

$$\hat{\varphi}_3(s)^3=d_3(\mathbb{B})^{-3}\max_{z\in s\overline{\mathbb{B}}_I}\left|z^{-3}\hat{g}_{w_1,w_2,w_3}(z)\right|.$$

For $r\in[s,t]$, we have $\hat{\varphi}_3(r)=\hat{\varphi}_3(s)$ and by the choice of w_1,w_2,w_3,

$$\begin{aligned}\hat{\varphi}_3(r)^3&\geq d_3(\mathbb{B})^{-3}\max_{z\in r\overline{\mathbb{B}}_I}\left|z^{-3}\hat{g}_{w_1,w_2,w_3}(z)\right|\\&\geq d_3(\mathbb{B})^{-3}\max_{z\in s\overline{\mathbb{B}}_I}\left|z^{-3}\hat{g}_{w_1,w_2,w_3}(z)\right|=\hat{\varphi}_3(s)^3;\end{aligned}$$

hence, by the maximum modulus principle (see Theorem 2.8), we get that the function $z\mapsto z^{-3}\hat{g}_{w_1,w_2,w_3}(z)$ is constant on $\mathbb{B}_I$. If we now consider $r\in(0,s)$

then $\hat{\varphi}_3(r) \leq \hat{\varphi}_3(s)$, and

$$\begin{aligned}\hat{\varphi}_3(r)^3 &\geq d_3(\mathbb{B})^{-3} \max_{z \in r\overline{\mathbb{B}}_I} \left| z^{-3} \hat{g}_{w_1,w_2,w_3}(z) \right| \\ &= d_3(\mathbb{B})^{-3} \max_{z \in s\overline{\mathbb{B}}_I} \left| z^{-3} \hat{g}_{w_1,w_2,w_3}(z) \right| = \hat{\varphi}_3(s)^3;\end{aligned}$$

hence, $\hat{\varphi}_3(r) = \hat{\varphi}_3(s)$ for all $r \in (0,t]$. Thanks to Lemma 5.5, we obtain

$$\hat{\varphi}_3(r) \equiv \lim_{r \to 0^+} \hat{\varphi}_3(r) = |\partial_c f(0)| = |a_1| \quad \text{for } r \in [0,t].$$

Recalling Remark 5.4, we get that either f is constant or $a_1 = \partial_c f(0) \neq 0$. Let us suppose that f is not constant (so that $b_n = a_n a_1^{-1} |a_1|$ for any $n \in \mathbb{N}$). Recalling the definition of $\hat{g}_{w_1,w_2,w_3}(z)$, and since the (constant) function $z \mapsto z^{-3} \hat{g}_{w_1,w_2,w_3}(z)$ is equal to its limit at 0, we have

$$\begin{aligned}&|a_1|^3 \\ &= \hat{\varphi}_3^3(r) \\ &= \frac{1}{d_3(\mathbb{B})^3} \left| \sum_{n \geq 3} z^{n-3} \sum_{k=0}^{n} \sum_{j=0}^{k} \left(w_2^j - w_1^j\right)\left(w_3^{k-j} - w_1^{k-j}\right)\left(w_3^{n-k} - w_2^{n-k}\right) b_j b_{k-j} b_{n-k} \right| \\ &= \frac{1}{d_3(\mathbb{B})^3} \left| (w_2 - w_1)(w_3 - w_1)(w_3 - w_2) b_1^3 \right|\end{aligned}$$

for any $z \in \mathbb{B}_I$; therefore, thanks to Lemma 5.2, without loss of generality we can suppose that $w_1 = 1, w_2, w_3$ are cube roots of unity in L_I. Now let J be an imaginary unit, $J \neq I$, and consider v_1, v_2, v_3 cube roots of unity in L_J. Then, for any $r \in [0,t]$,

$$\begin{aligned}|a_1| = \hat{\varphi}_3(r) &= d_3(\mathbb{B})^{-1} \max_{I \in \mathbb{S}} \max_{w_1,w_2,w_3 \in \overline{\mathbb{B}}_I} \max_{z \in r\overline{\mathbb{B}}_I} |z^{-3} \hat{g}_{w_1,w_2,w_3}(z)|^{1/3} \\ &\geq d_3(\mathbb{B})^{-1} \max_{z \in r\overline{\mathbb{B}}_J} |z^{-3} \hat{g}_{v_1,v_2,v_3}(z)|^{1/3} \geq d_3(\mathbb{B})^{-1} |z^{-3} \hat{g}_{v_1,v_2,v_3}(z)|_{z=0}^{1/3} = |a_1|;\end{aligned}$$

therefore, for any $J \in \mathbb{S}$, if v_1, v_2, v_3 are cube roots of unity in L_J, the function $z \mapsto z^{-3} \hat{g}_{v_1,v_2,v_3}(z) \equiv c_J$ is constant on $\mathbb{B}_J$. Notice that $|c_J|$ does not depend on $J \in \mathbb{S}$. Now let I be an imaginary unit in $\mathbb{S}$, fix $z \in t\mathbb{B}_I$ with $|z| = r$, and let $w_1 = 1, w_2, w_3$ be cube roots of unity in L_I. Consider the function defined for $\zeta \in \mathbb{B}_I$ by

$$\begin{aligned}h_z^I(\zeta) &= z^{-3} \hat{g}_{\zeta,w_2,w_3}(z) \\ &= \sum_{n \geq 3} z^{n-3} \sum_{k=0}^{n} \sum_{j=0}^{k} \left(w_2^j - \zeta^j\right)\left(w_3^{k-j} - \zeta^{k-j}\right)\left(w_3^{n-k} - w_2^{n-k}\right) b_j b_{k-j} b_{n-k}.\end{aligned}$$

By construction, $\zeta \mapsto h_z^I(\zeta)$ is holomorphic on a neighborhood of $\overline{\mathbb{B}}_I$, and

$$|h_z^I(\zeta)| \leq \hat{\varphi}_3(r)^3 d_3(\mathbb{B})^3 = |a_1|^3 d_3(\mathbb{B})^3.$$

Its value at $\zeta = 1$ is

$$h_z^I(1) = z^{-3} g_{1,w_2,w_3}(z) = (w_2 - 1)(w_3 - 1)(w_3 - w_2) b_1^3 = -3\sqrt{3} I |a_1|^3.$$

Then the function

$$\zeta \mapsto h_z^I(\zeta)(h_z^I(1))^{-1} = h_z^I(\zeta) I (3\sqrt{3})^{-1} |a_1|^{-3}$$

fixes the point $\zeta = 1$ and maps the closed unit disc $\overline{\mathbb{B}}_I$ to itself; in fact,

$$\begin{aligned} \left| h_z^I(\zeta) I (3\sqrt{3})^{-1} |a_1|^{-3} \right| &= |h_z^I(\zeta)| (3\sqrt{3})^{-1} |a_1|^{-3} \\ &\leq |a_1|^3 d_3(\mathbb{B})^3 (3\sqrt{3})^{-1} |a_1|^{-3} = 1. \end{aligned}$$

We can therefore apply Proposition 3.2 and we get

$$\mathrm{Im}_I \left(\frac{\partial}{\partial \zeta}\Big|_{\zeta=1} h_z^I(\zeta) I (3\sqrt{3})^{-1} |a_1|^{-3} \right) = 0; \quad \text{that is,} \quad \mathrm{Re}\left(\frac{\partial}{\partial \zeta} h_z^I(1) \right) = 0.$$

Doing the same construction for any $J \in \mathbb{S}$, we get that

$$\mathrm{Re}\left(\frac{\partial}{\partial \zeta} h_z^J(1) \right) = 0 \tag{27}$$

for any fixed $z \in t\mathbb{B}_J$. An easy computation shows that

$$\begin{aligned} &\frac{\partial}{\partial \zeta} h_z^I(1) \\ &= -\sum_{n \geq 3} z^{n-3} \sum_{k=2}^{n-1} \sum_{j=1}^{k-1} \big(j(w_3^{k-j} - 1) + (k-j)(w_2^j - 1) \big)(w_3^{n-k} - w_2^{n-k}) b_j b_{k-j} b_{n-k}. \end{aligned}$$

Thanks to the uniform convergence of the series expansion and since (27) holds for any $z \in t\mathbb{B}_I$, we get that the real part of each coefficient must vanish. Namely, for any $n \in \mathbb{N}$, $n \geq 3$,

$$\mathrm{Re}\left(\sum_{k=2}^{n-1} \sum_{j=1}^{k-1} \big(j(w_3^{k-j} - 1) + (k-j)(w_2^j - 1) \big)(w_3^{n-k} - w_2^{n-k}) b_j b_{k-j} b_{n-k} \right) = 0.$$

That this is true for any $I \in \mathbb{S}$ will allow us to show that $b_n = a_n(a_1^{-1}|a_1|)$ is real for any $n \in \mathbb{N}$. We do this by induction. The first step is trivial; $b_0 = 0$ and

$$b_1 = a_1(a_1^{-1}|a_1|) = |a_1|.$$

Suppose then that $b_1, \dots, b_{s-1}$ are real numbers. The first coefficient of the series expansion of $(\partial/\partial\zeta)h_z^I(1)$ that contains b_s is the one for which $n = s+2$, and which has to satisfy

(28)
$$\operatorname{Re}\sum_{k=2}^{s+1}\sum_{j=1}^{k-1}\bigl(j(w_3^{k-j}-1)+(k-j)(w_2^j-1)\bigr)(w_3^{s+2-k}-w_2^{s+2-k})b_j b_{k-j} b_{s+2-k}=0.$$

Three terms in this sum involve b_s: those with $(k,j)=(s+1,s)$, $(s+1,1)$, $(2,1)$. After some manipulations, their sum is seen to equal

$$\bigl(\sqrt{3}I(3s+2-(w_2^s+w_3^s))-3(w_2^s-w_3^s))\bigr)|a_1|^2 b_s.$$

hence, we can split the sum in (28) into

$$\bigl(\sqrt{3}I(3s+2-(w_2^s+w_3^s))-3(w_2^s-w_3^s))\bigr)|a_1|^2 b_s+\Sigma_1+\Sigma_2, \tag{29}$$

where

$$\Sigma_1=\sum_{j=2}^{s-1}\bigl(j(w_3^{s+1-j}-1)+(s+1-j)(w_2^j-1)\bigr)(w_3-w_2)b_j b_{s+1-j} b_1,$$

$$\Sigma_2=\sum_{k=3}^{s}\sum_{j=1}^{k-1}\bigl(j(w_3^{k-j}-1)+(k-j)(w_2^j-1)\bigr)(w_3^{s+2-k}-w_2^{s+2-k})b_j b_{k-j} b_{s+2-k}.$$

We claim that $\operatorname{Re}\Sigma_1$ and $\operatorname{Re}\Sigma_2$ vanish. Indeed, if s is even, we can express Σ_1 as

$$\sum_{j=2}^{s/2}\bigl(j(w_3^{s+1-j}+w_2^{s+1-j}-2)+(s+1-j)(w_2^j+w_3^j-2)\bigr)(w_3-w_2)b_j b_{s+1-j} b_1.$$

The real part of each summand vanishes because $w_2^n+w_3^n\in\mathbb{R}$ and $w_2^n-w_3^n\in I\mathbb{R}$ for any $n\in\mathbb{N}$, while $b_n\in\mathbb{R}$ for any $n=1,\dots,s-1$. This shows that $\operatorname{Re}\Sigma_1=0$ when s is even. The proofs for Σ_1 with s odd and for Σ_2 are similar.

We have reduced (28) to

$$\operatorname{Re}\Bigl(\bigl(\sqrt{3}I(3s+2-(w_2^s+w_3^s))-3(w_2^s-w_3^s))\bigr)|a_1|^2 b_s\Bigr)=0.$$

Therefore, for any $s\in\mathbb{N}$, there exists $\alpha_s\in\mathbb{R}$ such that $\operatorname{Re}(\alpha_s I b_s)=\alpha_s\operatorname{Re}(I b_s)=\operatorname{Im}_I(b_s)=0$ for all $I\in\mathbb{S}$; hence, we get $b_s\in\mathbb{R}$ for all s. Recalling that the b_n are the coefficients of the power series of $\hat{f}$, we get that $\hat{f}(\mathbb{B}_I)\subseteq L_I$ for all $I\in\mathbb{S}$; hence, $\hat{f}$ is complex holomorphic on each slice.

We now claim that for any $r \in (0,t]$ the slice 3-diameter of $f(r\mathbb{B})$ coincides with the usual 3-diameter of $\hat{f}(r\mathbb{B})$. Indeed, for any $I \in \mathbb{S}$, we have

$$\begin{aligned}
&\hat{g}_{w_1,w_2,w_3}(z)\\
&\qquad = \sum_{n\geq 0} z^n \sum_{k=0}^{n}\sum_{j=0}^{k}(w_2^j - w_1^j)(w_3^{k-j} - w_1^{k-j})(w_3^{n-k} - w_2^{n-k})b_j b_{k-j} b_{n-k}\\
&\qquad = \Big(\sum_{n\geq 0}\big((zw_2)^n - (zw_1)^n\big)\, b_n\Big)\Big(\sum_{n\geq 0}\big((zw_3)^n - (zw_1)^n\big)\, b_n\Big)\\
&\qquad\qquad \times \Big(\sum_{n\geq 0}\big((zw_3)^n - (zw_2)^n\big)\, b_n\Big),
\end{aligned}$$

which is to say

$$\hat{g}_{w_1,w_2,w_3}(z) = \big(\hat{f}(zw_2) - \hat{f}(zw_1)\big)\big(\hat{f}(zw_3) - \hat{f}(zw_1)\big)\big(\hat{f}(zw_3) - \hat{f}(zw_2)\big).$$

For each side we take the absolute value, the third root and the maximum over $z \in r\overline{\mathbb{B}}_I$ and $w_1, w_2, w_3 \in \overline{\mathbb{B}}_I$, to obtain $\hat{d}_3(f(r\mathbb{B})) = d_3(\hat{f}(r\mathbb{B}_I))$, as desired.

Thanks to the complex n-diameter version of the Landau–Toeplitz theorem [Burckel et al. 2008] we conclude that $\hat{f}$ is an affine function,

$$\hat{f}(q) = b_0 + qb_1 = a_0 a_1^{-1}|a_1| + q|a_1|.$$

Hence f is affine as well: $f(q) = a_0 + qa_1$. □

Theorem 5.7 (Landau–Toeplitz theorem for the slice 3-diameter). *Let f be a regular function on $\mathbb{B}$ such that $\hat{d}_3 f(\mathbb{B}) = d_3(\mathbb{B})$. Then*

$$\hat{d}_3(f(r\mathbb{B})) \leq d_3(r\mathbb{B}) \tag{30}$$

for every $r \in (0,1)$, and

$$|\partial_c f(0)| \leq 1. \tag{31}$$

Equality holds in (30) for some $r \in (0,1)$, or in (31), if and only if f is an affine function $f(q) = a + qb$ with $a, b \in \mathbb{H}$ and $|b| = 1$.

Proof. By Lemma 5.5, both inequalities hold true. For the equality case, if $f(q) = a + qb$ with $a, b \in \mathbb{H}$, $|b| = 1$, it is easy to see that equality holds in both statements; otherwise, if equality holds in (30) or in (31) then $\hat{\varphi}_3(r)$ defined in Lemma 5.5 is not strictly increasing. Theorem 5.6 then implies that f is an affine function. Since $\hat{d}_3((f(\mathbb{B})) = d_3(\mathbb{B})$, the coefficient of the first degree term of f has unitary modulus. □

Notice that the notion of the slice 3-diameter does not make sense for $n \geq 4$. Moreover, the n-diameter of $\mathbb{B}$, when $n \geq 4$, is not anymore attained at points that lie on the same plane L_I; in fact, the following result holds true:

Proposition 5.8. *For all $I \in \mathbb{S}$ the inequality $d_4(\mathbb{B}) > d_4(\mathbb{B}_I)$ holds.*

Proof. The proof follows from the direct computation of the 4-diameter of a maximal tetrahedron contained in $\mathbb{B}$. □

The proof of Theorem 5.6 heavily relies upon the fact that both the 3-diameter of $\mathbb{B}$ and the slice 3-diameter of $f(\mathbb{B})$ are attained at a complete set of cube roots of unity lying on a same plane L_I. We have no alternative proof to use when $n \geq 4$.

References

[Abate 1989] M. Abate, *Iteration theory of holomorphic maps on taut manifolds*, Mediterranean Press, Rende, 1989. MR 92i:32032 Zbl 0747.32002

[Burckel 1979] R. B. Burckel, *An introduction to classical complex analysis, 1*, Pure and Applied Mathematics **82**, Academic Press, New York, 1979. MR 81d:30001 Zbl 0434.30002

[Burckel et al. 2006] R. B. Burckel, D. E. Marshall, and P. Poggi-Corradini, "On a theorem of Landau and Toeplitz", preprint, 2006. arXiv math/0603579v1

[Burckel et al. 2008] R. B. Burckel, D. E. Marshall, D. Minda, P. Poggi-Corradini, and T. J. Ransford, "Area, capacity and diameter versions of Schwarz's lemma", *Conform. Geom. Dyn.* **12** (2008), 133–152. MR 2010j:30050 Zbl 1233.30016

[Colombo et al. 2009] F. Colombo, G. Gentili, I. Sabadini, and D. Struppa, "Extension results for slice regular functions of a quaternionic variable", *Adv. Math.* **222**:5 (2009), 1793–1808. MR 2010j:30102 Zbl 1179.30052

[Cullen 1965] C. G. Cullen, "An integral theorem for analytic intrinsic functions on quaternions", *Duke Math. J.* **32** (1965), 139–148. MR 30 #3227 Zbl 0173.09001

[Della Rocchetta et al. 2012] C. Della Rocchetta, G. Gentili, and G. Sarfatti, "The Bohr theorem for slice regular functions", *Math. Nachr.* **285**:17-18 (2012), 2093–2105. MR 3002603 Zbl 1257.30053

[Della Rocchetta et al. 2013] C. Della Rocchetta, G. Gentili, and G. Sarfatti, "A Bloch–Landau theorem for slice regular functions", pp. 55–74 in *Advances in hypercomplex analysis*, edited by G. Gentili et al., Springer–INdAM series **1**, Springer, Milan, 2013. Zbl 06148976

[Gardner 2006] R. J. Gardner, *Geometric tomography*, 2nd ed., Encyclopedia of Mathematics and its Applications **58**, Cambridge University Press, 2006. MR 2007i:52010 Zbl 1102.52002

[Gentili and Struppa 2006] G. Gentili and D. C. Struppa, "A new approach to Cullen-regular functions of a quaternionic variable", *C. R. Math. Acad. Sci. Paris* **342**:10 (2006), 741–744. MR 2006m:30095 Zbl 1105.30037

[Gentili and Struppa 2007] G. Gentili and D. C. Struppa, "A new theory of regular functions of a quaternionic variable", *Adv. Math.* **216**:1 (2007), 279–301. MR 2008h:30052 Zbl 1124.30015

[Gentili et al. 2013] G. Gentili, C. Stoppato, and D. C. Struppa, *Regular functions of a quaternionic variable*, Springer, Heidelberg, 2013. MR 3013643 Zbl 06090418

[Lachand-Robert and Oudet 2007] T. Lachand-Robert and É. Oudet, "Bodies of constant width in arbitrary dimension", *Math. Nachr.* **280**:7 (2007), 740–750. MR 2008d:52002 Zbl 1121.52009

[Landau and Toeplitz 1907] E. Landau and O. Toeplitz, "Über die größte Schwankung einer analytischen Funktion in einem Kreise", *Arch. Math. Phys.* (3) **11** (1907), 302–307. JFM 38.0433.02

[Sarfatti 2013] G. Sarfatti, *Elements of function theory in the unit ball of quaternions*, Ph.D. thesis, Università di Firenze, 2013, www.students.math.unifi.it/users/sarfatti/PhDGiulia.pdf.

[Solynin 2008] A. Y. Solynin, "A Schwarz lemma for meromorphic functions and estimates for the hyperbolic metric", *Proc. Amer. Math. Soc.* **136**:9 (2008), 3133–3143. MR 2009c:30064 Zbl 1151.30017

Received June 28, 2011. Revised April 14, 2013.

GRAZIANO GENTILI
DIPARTIMENTO DI MATEMATICA E INFORMATICA "U. DINI"
UNIVERSITÀ DI FIRENZE
VIALE MORGAGNI 67/A
50134 FIRENZE
ITALY
gentili@math.unifi.it

GIULIA SARFATTI
DIPARTIMENTO DI MATEMATICA E INFORMATICA "U. DINI"
UNIVERSITÀ DI FIRENZE
VIALE MORGAGNI 67/A
50134 FIRENZE
ITALY
sarfatti@math.unifi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.405

ON SURGERY CURVES FOR GENUS-ONE SLICE KNOTS

PATRICK M. GILMER AND CHARLES LIVINGSTON

If a knot K bounds a genus-one Seifert surface $F \subset S^3$ and F contains an essential simple closed curve α that has induced framing 0 and is smoothly slice, then K is smoothly slice. Conjecturally, the converse holds. It is known that if K is slice and the determinant of K is not 1, then there are strong constraints on the algebraic concordance class of such α, and it was thought that these constraints might imply that α is at least algebraically slice. We present a counterexample; in the process we answer negatively a question of Cooper and relate the result to a problem of Kauffman. Results of this paper depend on the interplay between the Casson–Gordon invariants of K and algebraic invariants of α.

1. Introduction

For $n > 1$, if a smooth knotted S^{2n-1} in S^{2n+1} bounds an embedded disk in B^{2n+2}, such a smooth slicing disk can be constructed from a $2n$-manifold bounded by K in S^{2n+1} by ambient surgery. Whether the same is true for knots in S^3 has remained an open question for 40 years, though counterexamples exist in the topological category [Freedman and Quinn 1990].

One well-known and simply stated conjecture [Kirby 1978, Problem 1.38] is a special case: the untwisted Whitehead double of a knot $J \subset S^3$ is smoothly slice if and only if J is smoothly slice. More generally, if K is a knot in S^3 that bounds a genus-one Seifert surface F and is algebraically slice, then up to isotopy and orientation change, there are exactly two essential simple closed curves on F, J_1 and J_2, with self-linking 0 with respect to the Seifert form of F. In this situation, we will call J_1 and J_2 surgery curves for F. Conjecturally, if K is smoothly slice, then one of J_1 or J_2 is necessarily smoothly slice (see [Kauffman 1987, Strong conjecture, page 226] for instance).

Shortly after Casson and Gordon [1986] developed obstructions to slicing algebraically slice knots, it was noticed that Casson–Gordon invariants could be

Gilmer was partially supported by NSF-DMS-0905736. Livingston was partially supported by NSF-DMS-1007196.
MSC2010: 57M25.
Keywords: knot, slice knot, genus 1.

expressed in terms of signature invariants of curves on Seifert surfaces [Gilmer 1983; Litherland 1984]. Moreover, Casson–Gordon invariants could be interpreted in this way as obstructions to slicing K by slicing a surgery curve on a genus-one Seifert surface for K. Casson–Gordon invariants actually obstruct topological locally flat slice disks.

A genus-one knot K is algebraically slice if and only if it has an Alexander polynomial of the form

$$\begin{aligned}\Delta_K(t) &= (mt-(m+1))((m+1)t-m)\\ &= m(m+1)t^2-(m^2+(m+1)^2)t+m(m+1)\end{aligned}$$

for some $m \geq 0$. Observe that if Δ_K has the form above, then the nonnegative integer m is determined. For a genus-one algebraically slice knot K, let $m(K)$ denote this number; note that the determinant of K is $(2m(K)+1)^2$.

We let $\sigma_K(t)$ denote the Levine–Tristram signature function of K [Levine 1969; Tristram 1969], as defined on the unit interval [0, 1] and redefined to be the average of the one-sided limits at the jumps. Casson–Gordon theory implies that if a genus-one knot K is slice and $m(K) \neq 0$, then the signature function of one of the surgery curves satisfies strong constraints. To state these, we make the following definition.

Definition 1. A knot J satisfies the (m, p)-signature conditions for integers $m > 0$ and p relatively prime to m and $m+1$ if

$$\sum_{i=0}^{r-1} \sigma_J(ca^i/p) = 0$$

for all $c \in \mathbb{Z}_p{}^*$, and $a = (m+1)/m \bmod p$,where r is the order of a modulo p.

To get a feeling for this summation, consider the case of $m(K)=1$ and $p=73$. In $\mathbb{Z}_{73}$, the number 2 generates the multiplicative subgroup $\{1,2,4,8,16,32,64,55,37\}$. This subgroup has 8 cosets in the group of units $(\mathbb{Z}_{73})^*$. For instance, the coset containing $c=5$ is {5, 7, 10, 14, 20, 28, 39, 40, 56}. Thus the following arises as one of the sums in the $(1, 73)$-signature condition:

$$\sigma_J\left(\tfrac{5}{73}\right)+\sigma_J\left(\tfrac{7}{73}\right)+\sigma_J\left(\tfrac{10}{73}\right)+\sigma_J\left(\tfrac{14}{73}\right)+\sigma_J\left(\tfrac{20}{73}\right)+\sigma_J\left(\tfrac{28}{73}\right)+\sigma_J\left(\tfrac{39}{73}\right)+\sigma_J\left(\tfrac{40}{73}\right)+\sigma_J\left(\tfrac{56}{73}\right).$$

Notice that the cosets appear to be fairly randomly distributed in the unit interval. Nonetheless, as we show, the vanishing of all such sums is not sufficient to imply the vanishing of the signature function itself. Consider the following simple consequence of Theorem 8 below.

Theorem 2. *If K is a genus-one smoothly slice knot, then one of the surgery curves J satisfies $(m(K), p)$-signature conditions for an infinite set of primes p.*

In his thesis, Cooper states a stronger result:

Theorem 3 [Cooper 1982]. *Let K be a genus-one smoothly slice knot, then one of the surgery curves J satisfies the $(m(K),\, p)$-signature conditions for all p relatively prime to m and $m+1$.*

One quick corollary, first observed by Cooper, of either of these theorems is that for a genus one slice knot K with $m(K) > 0$, the integral of the signature function of one of the slice curves J is 0. This follows by summing the signature sums in the theorem over all values of c to get a sum of the form $\sum_{i=1}^{p-1} \sigma_J(i/p) = 0$ and then noting that for large p, this sum approximates the integral. (This integral condition was later seen to follow from the L^2-signature approach of [Cochran et al. 2003, Theorem(1.4)].)

Clearly, the constraints given by these theorems are quite extensive. One explicit question asked by Cooper is whether the vanishing of the combined sum $\sum_{i=1}^{p-1} \sigma_J(i/p)$ for the appropriate infinite sets of p implies the vanishing of the signature function [Cooper 1982, Question (3.16)]. We will show that the answer is no. In fact, the much stronger constraints given in Theorems 2 and 3 are not sufficient to imply the vanishing of the signature function of one of the surgery curves. Here is the algebraic formulation of the question.

Question 4. Let σ be an integer-valued step function defined on [0, 1] with the property that $\sigma(x) = \sigma(1-x)$ for all x. Assume also $\sigma(0) = \sigma(1) = 0$, that there are no jumps at points with denominator a prime power, and that σ is equal to the average of the one-sided limits at the jumps. Suppose that for all $p > 1$ coprime to m and $m+1$, for G the multiplicative subgroup of $(\mathbb{Z}_p)^*$ generated by $(m+1)/m$, and for all $n \in \mathbb{Z}_p$, we have

$$\sum_{r \in nG} \sigma(r/p) = 0.$$

Then does $\sigma(t) = 0$ for all t?

For each $m > 0$, the answer to the above question is emphatically no. Let $K_{(r,s)}$ denote the (r, s)-cable of K (that is, r longitudes, and s meridians). Let $-K$ denote the mirror image of K.

Theorem 5. *Let K be a knot with a nonzero signature function, and $m > 0$. The signature function of $K_{(m,1)} \# -K_{(m+1,1)}$ is nonzero and satisfies the $(m,\, p)$-signature conditions for all p relatively prime to m and $m+1$.*

We have a perhaps nicer family to work with in the case $m = 1$. Let $T_{r,s}$ denote the (r, s)-torus knot, which is the (r, s)-cable of the unknot.

Theorem 6. *If r is an odd number and $r \geq 3$, the signature function of $(T_{2,r})_{(2,-r)}$ is nonzero and satisfies the $(1,\, p)$-signature conditions for odd p.*

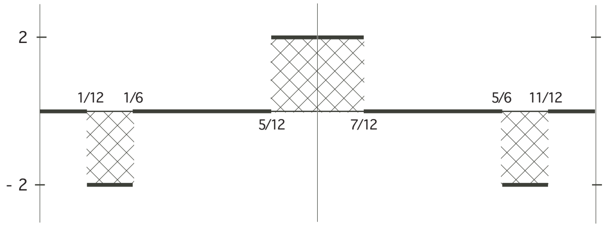


Figure 1. Signature function of $(T_{2,3})_{(2,-3)}$ satisfying the $(1, p)$-signature conditions for odd p.

Although Casson–Gordon theory gives a somewhat weaker version of Cooper's theorem, it provides access to the more powerful Witt class analogs of Theorem 2, which carry more information than is given by signatures. Also, Casson–Gordon theory obstructs topological sliceness, whereas Cooper worked in the smooth category. We now describe these Witt class invariants.

If K is a knot, let $V_t = (1-t)V + (1-t^{-1})V^t$, where V is a Seifert matrix of K and t is an indeterminate. For p a prime power and $j/p \in \mathbb{Z}[1/p]/\mathbb{Z}$, let $w_K(j/p)$ denote the element represented by $V_{e^{2\pi ij/p}}$ in $W(\mathbb{Q}(\zeta_p)) \otimes \mathbb{Z}_{(2)}$. Here, $W(\mathbb{Q}(\zeta_p))$ denotes the Witt group of hermitian forms over the field $\mathbb{Q}(\zeta_p)$ and $\mathbb{Z}_{(2)}$ denotes $\mathbb{Z}$ localized at 2. An elementary proof shows that this defines a homomorphism on the concordance group.

Definition 7. We say a knot J satisfies the (m, p)-Witt conditions for integers $m > 0$ and p relatively prime to m and $m + 1$ if

$$\sum_{i=0}^{r-1} w_J[(c+ai)/p] = 0 \in W(\mathbb{Q}(\zeta_p)) \otimes \mathbb{Z}_{(2)}$$

for all $c \in \mathbb{Z}_p{}^*$, $a = (m+1)/m \mod p$, and r the order of a modulo p.

If a knot J satisfies the (m, p)-Witt conditions, it satisfies the (m, p)-signature conditions as well. But the Witt conditions are stronger. For instance, one can define a discriminant invariant on $W(\mathbb{Q}(\zeta_p)) \otimes \mathbb{Z}_{(2)}$, which is discussed in [Gilmer and Livingston 1992b].

Theorem 8. *Let K be a genus one topologically slice knot. There is some finite set of bad primes P such that one of the surgery curves J satisfies the $(m(K), p)$-Witt conditions for all p in the set*

$$\{r^n \mid n \in \mathbb{Z}_+, r \text{ is prime}, r \notin P, r^n \text{ divides } (m+1)^q - (m)^q \text{ for some prime power } q\}.$$

Consider $\mathrm{Wh}(J, n)$, the n-twisted Whitehead double of J. It is well-known that this knot is algebraically slice if and only if $n = m(m + 1)$. Moreover $m(\mathrm{Wh}(J, m(m + 1))) = m$. It is also known that the two surgery curves for

$\mathrm{Wh}(J, m(m+1))$ both have the isotopy type of $J \# T_{(m,m+1)}$. One can see this using the techniques discussed in [Kauffman 1987, pages 214–223]. Using this fact, for these knots one can sometimes remove the exceptions created by the unknown set of bad primes.

Theorem 9. *Let $m > 0$. If $\mathrm{Wh}(J, m(m+1))$ is topologically slice, then $J \# T_{(m,m+1)}$ satisfies the (m, p)-Witt conditions for all p in the set*

$$\{p \mid p \textit{ is prime}, \ \gcd(p^2, (m+1)^q - (m)^q) = p \textit{ for some odd prime power } q\}.$$

Our examples of knots satisfying (m, p)-signature conditions also satisfy Witt conditions.

Theorem 10. *For any knot K and $m > 0$, $K_{(m,1)} \# -K_{(m+1,1)}$ satisfies the (m, p)-Witt conditions for all p relatively prime to m and $m+1$. For any odd integer n, $(T_{2,n})_{(2,-n)}$ satisfies the $(1, p)$-Witt conditions for all odd p.*

In the next theorems, we focus on some particularly nice examples.

Theorem 11. *Let $J = (T_{2,3})_{(2,-3)}$, the $(2, -3)$-cable of trefoil knot $T_{2,3}$. Let $K = \mathrm{Wh}(J, 2)$.*

(1) *K is a genus one algebraically slice knot with both surgery curves having the same knot type, J.*

(2) *J satisfies the $(1, p)$-Witt conditions for all odd p. In particular J satisfies the $(1, p)$-signature conditions for all odd p. Another consequence is that the constraints of Theorems 3, 8, and 9 on K are satisfied.*

(3) *The signature function of J is nonzero.*

(4) *$\Delta_J(t) = (t^{-1} - 1 + t)(t^{-2} - 1 + t^2)$ does not satisfy the Fox–Milnor condition; that is, $\Delta_J(t)$ cannot be written as $f(t) f(t^{-1})$ for $f(t) \in \mathbb{Z}[t, t^{-1}]$.*

(5) $\mathrm{Arf}\, J \neq 0$.

We do not know whether $\mathrm{Wh}((T_{2,3})_{(2,-3)}, 2)$ is topologically locally flat slice or smoothly slice. A conjecture made by Kauffman [1987, Weak Conjecture, page 226; Kirby 1978, Problem 1.52] implies that $\mathrm{Wh}((T_{2,3})_{(2,-3)}, 2)$ is not smoothly slice since $\mathrm{Arf}((T_{2,3})_{(2,-3)}) \neq 0$. Thus examples such as this one offer a route to possible counterexamples to this conjecture.

By modifying the example slightly (without changing the relevant signature function, Alexander polynomial, Arf invariant, or even Witt class invariant), using results of Hedden [2007; 2009] on the Ozsváth–Szabó invariant of cables and Whitehead doubles, obstructing sliceness becomes possible. This is described in the first part of the following theorem. We also give a second example of a knot with similar properties.

Theorem 12. *Let $J' = (T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0))_{(2,-3)}$. Then $K' = \mathrm{Wh}(J', 2)$ is not smoothly slice. Moreover the conclusions of Theorem 11 hold when K is replaced by K' and J is replaced by J'.*

Let $J'' = (T_{2,3})_{(2,-3)} \# (T_{2,3})_{(2,-3)}$. Then $K'' = \mathrm{Wh}(J'', 2)$ is not smoothly slice. Moreover conclusions (1), (2), *and* (3) *of Theorem 11 hold when K is replaced by K'' and J is replaced by J''.*

In Section 2, we outline the proofs of Theorems 5, 6 and 10. Section 3 presents the proof of Theorem 12 using tools from Heegaard–Floer theory. In Section 4 and Appendices B and C, we review Casson–Gordon theory and prove Theorems 8 and 9. Similar arguments have appeared, but some depend on Theorem 1 of [Gilmer 1993], whose proof contains a gap (shared with [Gilmer 1983, Theorem (0.1)]). We show how to modify this proof to obtain the results stated above. In Section 5, we give some restrictions on signature functions which satisfy the m-signature averaging conditions.

2. Proofs of Theorems 5, 6 and 10

Let S be a finite set in $\mathbb{R}/\mathbb{Z}$. For any function $f(t)$ on $\mathbb{R}/\mathbb{Z}$ taking values in an abelian group, define $\mu_S(f(t)) = \sum_{s\in S} f(s)$. We let $\phi_k : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ denote multiplication by the integer k. Observe that if ϕ_k is injective on S, then $\mu_{\phi_k(S)}(f(t)) = \mu_S(f(kt))$. In particular, we have the following.

Lemma 13. *If $S \subset \mathbb{R}/\mathbb{Z}$ is a finite set on which ϕ_m and ϕ_n are both injective and $\phi_m(S) = \phi_n(S)$, then for all f, $\mu_S(f(mt) - f(nt)) = 0$.*

In the current case of interest, we have an integer m, an integer p relatively prime to $m(m+1)$, and an integer c representing an element in $\mathbb{Z}_p^*$. We let $a = (m+1)/m$ mod p and $S = \{ca^i/p\} \subset \mathbb{Q}/\mathbb{Z}$. Notice that $ma^i = (m+1)a^{i-1}$. Thus, in this setting $\phi_m(S) = \phi_{m+1}(S)$.

Corollary 14. *With the notation of the previous paragraph, for all f,*

$$\mu_S\big(f((m+1)t) - f(mt)\big) = 0.$$

An immediate application is the case that f is the signature function of a knot J, in which case $f(mt)$ is the signature function of the knot $J_{m,\pm 1}$.

In the proof of Lemma 13, it is not required that f be defined on all of $\mathbb{R}/\mathbb{Z}$, but only on the sets S, $\phi_m(S)$ and $\phi_n(S)$. For instance, for a knot J and prime power p, there is the function $w_J : \{j/p\} \to W(\mathbb{Q}(\zeta_p)) \otimes \mathbb{Z}_{(2)}$ defined by

$$w_J(j/p) = (1 - \zeta_p^j)V + (1 - \zeta_p^{-j})V^t,$$

where $\zeta_p = e^{2\pi i/p}$.

The only missing ingredient in the proofs of Theorems 5, 6, and 10 is the following theorem.

Theorem 15. *If $\mathbb{S}$ is a satellite of C with orbit P and winding number n, then*

$$w_{\mathbb{S}}(j/p) = w_P(j/p) + w_C(nj/p).$$

This result is very close to a result of Litherland [1984, Theorem 1], which states that if $V_t(K) = (1-t)V + (1-t^{-1})V^t$, where V is the Seifert form of K, then $V_t(\mathbb{S})$ is Witt equivalent to the form $V_{t^n}(C) \oplus V_t(P)$ in the Witt group $\mathrm{Wh}(\mathbb{Q}(t))$ of the function field. One would like to argue at this point that the substitution of ζ_p for t defines a map $W(\mathbb{Q}(t)) \to W(\mathbb{Q}(\zeta_p))$, and Theorem 15 results. Unfortunately, this procedure does not lead to a well-defined map $W(\mathbb{Q}(t)) \to W(\mathbb{Q}(\zeta_p))$, as a class in $W(\mathbb{Q}(t))$ may be represented by a matrix whose entries have poles at ζ_p. We leave it to Appendix A to show how this hurdle can be overcome.

3. Smooth obstructions to slicing

In [Ozsváth and Szabó 2003], an invariant τ is defined with the property that if K is smoothly slice, then $\tau(K) = 0$. In order to apply this, we need to modify our knot K slightly. Let $K' = \mathrm{Wh}((T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0))_{(2,-3)}, 2)$. We show $\tau(K') = 1$.

As a first step, it follows from [Ozsváth and Szabó 2003] that $\tau(T_{2,3}) = 1$. Next, Hedden [2007] proved that for any J, $\tau(\mathrm{Wh}(J, t)) = 1$ for all $t < 2\tau(J)$. Thus, $\tau(\mathrm{Wh}(T_{2,3}, 0)) = 1$. By additivity, $\tau(T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0)) = 2$.

According to another theorem of Hedden [2009], if $\tau(J) = \mathrm{genus}(J)$, then

$$\tau(J_{(s,sn+1)}) = s\tau(J) + \tfrac{1}{2}sn(s-1) + s - 1.$$

In the case of interest to us, we have $s = 2$ and $n = -2$, so $\tau(J_{(2,-3)}) = 2\tau(J) - 1$. We do have $\tau(T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0)) = \mathrm{genus}(T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0)) = 2$, so

$$\tau((T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0))_{(2,-3)}) = 2\tau(T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0)) - 1 = 2(2) - 1 = 3.$$

Finally, again by Hedden's computation of τ of doubled knots,

$$\tau(\mathrm{Wh}((T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0))_{(2,-3)}, t)) = 1$$

if $t < 6$. So in particular, $\tau(\mathrm{Wh}((T_{2,3} \# \mathrm{Wh}(T_{2,3}, 0))_{(2,-3)}, 2)) = 1$.

We can also consider $K'' = \mathrm{Wh}((T_{2,3})_{(2,-3)} \# (T_{2,3})_{(2,-3)}, 2)$. Using the same formula of Hedden's for cables, we have

$$\tau((T_{2,3})_{(2,-3)}) = 1,$$

which gives us

$$\tau((T_{2,3})_{(2,-3)} \# (T_{2,3})_{(2,-3)}) = 2.$$

Then using Hedden's formula for doubles, $\tau(K'') = 1$.

4. Casson–Gordon theory

By a character χ on X, we mean a homomorphism $\chi : H_1(X) \to \mathbb{Q}/\mathbb{Z}$. This is a d-character if $\chi : H_1(X) \to (1/d)\mathbb{Z}/\mathbb{Z} \subset \mathbb{Q}/\mathbb{Z}$. Given a knot K and a prime power q, let S_q denote the q-fold branched cyclic cover of S^3 along K. Given a d-character on S_q, Casson and Gordon [1986] defined an invariant $\tau(K, \chi)$ taking values in $W(\mathbb{Q}[\zeta_d](t)) \otimes \mathbb{Q}$. Here, $W(\mathbb{Q}[\zeta_d](t))$ is the Witt group of Hermitian forms over $\mathbb{Q}[\zeta_d](t)$. If d is odd (as will be the case when K is a genus one algebraically slice knot), then $\tau(K, \chi)$ may be refined to take values in $W(\mathbb{Q}[\zeta_d](t)) \otimes \mathbb{Z}_{(2)}$ [Gilmer and Livingston 1992a; 1992b]. This refinement is useful as these Witt groups have 2-torsion. Here is the theorem of Casson and Gordon [1986; 1978] which asserts that certain $\tau(K, \chi)$ vanish when K is slice. (Casson and Gordon proved this theorem for smooth slice disks, and later, based on [Freedman and Quinn 1990], it was seen to hold in the topological locally flat category.)

Theorem 16 [Casson and Gordon 1986]. *Let K be a slice knot bounding a slice disk $\Delta \subset B^4$. Let W_q be the q-fold cyclic branched cover of B^4 over Δ.*

- *If χ is a character on S_q of prime power order that extends to W_q, then $\tau(K, \chi) = 0$.*
- *A character χ on S_q extends to W_q if and only if it vanishes on $\kappa(\Delta, q)$, the kernel of $H_1(S_q) \to H_1(W_q)$.*
- *The kernel $\kappa(\Delta, q)$ is a metabolizer for the linking form on $H_1(S_q)$ and is invariant under the group of covering transformations.*
- *The set of characters χ on S_q that extend to W_q form a metabolizer $\mathfrak{m}(q, \Delta)$ for the linking form on $H^1(S_q, \mathbb{Q}/\mathbb{Z})$.*

If p is a prime and G is an abelian group, let $G_{(p)}$ denote the p-primary summand of G. Note that the obstruction to sliceness given by Theorem 16 can be reduced to a sequence of obstructions associated to each prime p: $\tau(K, \chi) = 0$ for $\chi \in \mathfrak{m}(q, \Delta)_{(p)}$.

Let F be a Seifert surface K. Then $F \cup \Delta$ bounds a 3-manifold $R \subset B^4$. In [Gilmer 1993, Theorem 1], the first author related $\mathfrak{m}(q, \Delta)$ to the metabolizer H for Seifert form on $H_1(F)$ that arises as the kernel of the map induced by inclusion $H_1(F) \to H_1(R)/\text{Torsion}(H_1(R))$. However, Stefan Friedl [2004] found a gap in the proof, appearing in the second to last sentence on page 6 of [Gilmer 1993]. We now want to state a corrected version of the theorem.

Theorem 17. *Assume the notations and suppositions of Theorem 16, and let R and H be as above. Let p be a prime relatively prime to $|\text{Torsion}(H_1(R))|$. Let $\{x_i'\}$ be a basis for H. Let $\{y_i'\}$ be a complementary dual basis in $H_1(F)$ to $\{x_i'\}$ with respect to the intersection pairing. View F as built from a disk by adding $2g$ bands, with* *cores representing the x'_i and y'_i. Let the linking circles to those bands be denoted x_i and y_i. Let $\mathbb{Y}$ be the subgroup of $H_1(S_q)$ generated by the lifts of the y_i to a single component of the inverse image of $S^3 \setminus F$ in S_q. Then $\kappa(\Delta, q)_{(p)} = \mathbb{Y}_{(p)}$.*

Two independent proofs of Theorem 17 are presented in Appendices B and C. In [Friedl 2004, Theorem 8.6] and [Cochran et al. 2003, page 511], an equivalent result is asserted for almost all primes p (rather than for all primes not dividing $|\mathrm{Torsion}(H_1(R))|$).

To each element $z \in H_1(S_q)_{(p)}$, there is an associated character

$$\chi_z : H_1(S_q)_{(p)} \to \mathbb{Z}_{p^k} \subset \mathbb{Q}/\mathbb{Z}$$

(for some value of k), defined by $\chi_z(w) = \ell k(w, z) \in \mathbb{Q}/\mathbb{Z}$.

Corollary 18. *Assuming the notations and suppositions of Theorems 16 and 17, then $\mathfrak{m}(q, \Delta)_{(p)} = \{\chi_z \mid z \in \mathbb{Y}_{(p)}\}$.*

We can now summarize the proof of Theorem 8. Details follow as in [Gilmer 1993].

Proof of Theorem 8. By Theorem 17, one needs to show that the vanishing of the Casson–Gordon invariants for characters χ_z with $z \in \mathbb{Y}_p$ implies the surgery curve J satisfies the specified $(m(K), p)$-Witt conditions. There are two steps. First, one considers a new knot K', formed from K by tying a knot $-J$ in the band of the Seifert surface representing J. This new knot is slice, since it has surgery curve $J \# -J$, which is slice. The manifold R for K' is built by adding a two-handle to $F \times [0, 1]$, and can be seen to be a solid handlebody, in fact, a solid torus. Thus, Theorem 17 implies that for all the relevant characters, the Casson–Gordon invariants vanish. The proof is completed by proving that the effect of changing K to K' on the Casson–Gordon invariants is to add the sum of invariants appearing in the $(m(K), p)$-Witt conditions. □

(We take this opportunity to remark that Theorem (3.5) of [Gilmer 1983] remains valid. Although its proof uses Theorem 1.1 of the same paper,[1] it only does so in the case that R is a handlebody. For similar reasons, the proof of [Naik 1996, Theorem 7] is valid.)

Proof of Theorem 9. If K is an algebraically slice knot of genus one, $m = m(K)$, and q is odd, then $H_1(S_q)$ is the direct sum of two cyclic groups of order $(m+1)^q - m^q$. For each odd prime p such that $\gcd(p^2, (m+1)^q - (m)^q) = p$, the p-primary part of $H_1(S_q)$ (denoted $H_1(S_q)_{(p)}$) is a two-dimensional vector space over $\mathbb{Z}_p$. An analysis of $H_1(S_q)$ (as in the proof of [Gilmer 1993, top of page 16]) shows that the two metabolizers for the Seifert form spanned by the two surgery curves, say J_1 and J_2, lead to two distinct metabolizers for the linking form restricted to $H_1(S_q)_{(p)}$. In

[1] This is the same as [Gilmer 1993, Theorem 1] in the case $q = 2$.

fact, these metabolizers are eigenspaces for a generator of the group of covering transformations with the distinct eigenvalues $(m+1)/m$ and $m/(m+1)$. Thus this linking form on $H_1(S_q)_{(p)}$ is hyperbolic. It follows that an element in $H_1(S_q)_{(p)}$ in the complement of the union of these two metabolizers cannot have self-linking zero. Therefore, the linking form on $H_1(S_q)_{(p)}$ has only these two metabolizers.

If K is slice, then $\kappa(\Delta, q)_{(p)}$ must be one of these two metabolizers. Thus by Theorem 16, if $\chi : H_1(S_q) \to (1/p)\mathbb{Z}/\mathbb{Z}$ vanishes on $\kappa(\Delta, q)_{(p)}$, then $\tau(K, \chi) = 0$. By [Gilmer 1993, proof of Theorem 3], for each of these p, either J_1 or J_2 must satisfy the (m, p)-Witt conditions. But for $K = W(J, m(m+1))$, both J_1 and J_2 have the isotopy type of $J \# T_{(m,m+1)}$. □

5. The averaging conditions restrict where the jumps can occur

We consider the family $\mathcal{J}$ of step functions f on [0, 1] that vanish at 0 and 1 and have a finite number of jumps, with value at the jumps the average of the one-sided limits. For $f \in \mathcal{J}$, define

$$\Sigma_p(f) = \sum_{i=1}^{p-1} f(i/p).$$

Consider, also, the family of symmetric jump functions

$$\mathcal{S} = \{f \in \mathcal{J} \mid f(x) = f(1-x)\}.$$

These include the knot signature functions.

We say that $\sigma \in \mathcal{S}$ satisfies the m-signature averaging condition if $\Sigma_p(\sigma) = 0$ for each p relatively prime to m and $m+1$. The m-signature averaging condition is a consequence of the (m, p)-signature conditions for all p relatively prime to m and $m+1$.

The Alexander polynomial of the knot 5_2 is $2 - 3t + 2t^2$ [Cha and Livingston 2011] which has simple roots at $\frac{1}{4}(3 \pm i\sqrt{7})$. These roots lie on the unit circle and have argument $\pm 2\pi a$, where $a = \frac{1}{2\pi i} \log\bigl(\frac{1}{4}(3 + i\sqrt{7})\bigr) \approx 0.115$.

Proposition 19. *The number a is irrational. The signature function of $5_2 \# -(5_2)_{2,1}$ has jumps in the interval $\bigl[0, \frac{1}{2}\bigr]$ at $a/2$, a, and $(1-a)/2$, and this signature function satisfies the $(1, p)$-signature conditions for all odd p.*

Proof. If a were rational, $2 - 3t + 2t^2$ would have to be a factor of some cyclotomic polynomial; but these are monic. The signature function of 5_2 viewed as a function on [0, 1] has jumps at a and $1 - a$. Using [Litherland 1984] or [Livingston and Melvin 1985], the signature function of the knot $(5_2) \# -(5_2)_{2,1}$ jumps at exactly $a/2$, a, $(1-a)/2$, $(1+a)/2$, $1-a$, and $1-a/2$. By Theorem 5, $(5_2) \# -(5_2)_{2,1}$ satisfies the $(1, p)$-signature conditions for all odd p. □

This example contradicts a claim that we once (see the sentence beginning on the first line of [Gilmer and Livingston 1992a, page 486]) deferred to a future publication, but now retract. Note that the locations of the irrational jumps $a/2$, a, and $(1-a)/2$ in the first half interval together with 1 are linearly dependent over $\mathbb{Q}$. Our next theorem says that this is necessary for the jumps of a signature function which satisfies the m-signature averaging condition.

For $0 < a < 1$, let χ_a denote the characteristic function taking the value 1 on $[0, a)$, value $\frac{1}{2}$ at a and value 0 on $(a, 1]$. We have

$$\Sigma_p(\chi_a) = \begin{cases} \lfloor pa \rfloor & \text{if } pa \notin \mathbb{Z}, \\ \lfloor pa \rfloor - \frac{1}{2} & \text{if } pa \in \mathbb{Z}, \end{cases}$$

where $\lfloor x \rfloor$ denotes the greatest integer in x.

For $0 < a < \frac{1}{2}$, consider the symmetric jump function $S_a = \chi_{1-a} - \chi_a$ on $[0, 1]$. Then $S_a \in \mathcal{S}$ and

$$\Sigma_p(S_a) = \lfloor p(1-a) \rfloor - \lfloor pa \rfloor.$$

We define $F_p(a)$ by

$$\Sigma_p(S_a) - p \int_0^1 S_a(x)\,dx = F_p(a) = \begin{cases} 2\langle pa \rangle - 1 & \text{if } pa \notin \mathbb{Z}, \\ 0 & \text{if } pa \in \mathbb{Z}, \end{cases} \tag{5-1}$$

where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x.

Theorem 20. *Let $\sigma \in \mathcal{S}$ and let $\{j_1, \dots, j_s\}$ be the irrational points of discontinuity of σ that lie in the interval $[0, \frac{1}{2}]$. Suppose $s \geq 1$. If σ satisfies the m-signature averaging condition, then $\{j_1, \dots, j_s, 1\}$ are linearly dependent over $\mathbb{Q}$.*

Proof. It is easily seen that the integral of σ must be zero. We assume that there is a jump at an irrational point. Thus $s \geq 1$.

We have that σ can be written uniquely as $\sum_{i=1}^r c_i S_{a_i}$ with the c_i nonzero and the a_i distinct. By reordering, we can assume that a_i is rational if and only if $i > s$, for some $s \leq r$. Thus $\{j_1, \dots, j_s\} = \{a_1, \dots, a_s\}$. For each $i > s$, write $a_i = b_i/d_i$ in lowest terms. Let D be the least common multiple of the elements of $\{d_i \mid i > s\} \cup \{m, m+1\}$. Let $N = \{p \mid p > 0,\ p \equiv -1 \pmod{D}\}$. For all $p \in N$, $\Sigma_p \sigma = 0$, and $pa_i \notin \mathbb{Z}$. Hence, using (5-1), we have that $\sum_{i=1}^r c_i \langle pa_i \rangle = r/2$ for all $p \in N$.

Since $p \in N$ is constant modulo D, $\sum_{i=s+1}^r c_i \langle pa_i \rangle$ is constant for $p \in N$. Hence the sum over the irrational terms, $\sum_{i=1}^s c_i \langle pa_i \rangle$ is constant for $p \in N$, as well. Thus

$$\mathcal{I} = \{(\langle pa_1 \rangle, \langle pa_2 \rangle, \dots, \langle pa_s \rangle) \mid p \in N\}$$

is not dense in I^s. Kronecker's Theorem [Hardy and Wright 1938, Theorem 442] states that if the fractional parts of the positive integral multiples of a vector $(a_1, a_2, \dots, a_s)$ are not dense in I^s, then $\{a_1, \dots, a_s, 1\}$ are linearly dependent over

$\mathbb{Q}$. It is not hard to see that the same holds for multiples by any arithmetic sequence, like N. $\square$

The above theorem still holds if one relaxes the hypothesis by removing the condition that the value of σ at the jump points be given by the average of the one sided limits, as one could redefine the values at these points without changing the values of $\Sigma_p(\sigma)$ for the specified p's.

Note that, if a is a rational whose denominator divides d, then

$$F_p(a) = F_{p+kd}(a) = -F_{-p+kd}(a). \tag{5-2}$$

Definition 21. Given an odd number $d > 1$, let $\mathbb{D}(d)$ be the determinant of the $(d-1)/2 \times (d-1)/2$ matrix indexed by $1 \le i, j \le (d-1)/2$ with entries

$$F_i(j/d) = \begin{cases} 2\langle ij/d\rangle - 1 & \text{if } d \nmid ij, \\ 0 & \text{if } d \mid ij. \end{cases}$$

Conjecture 22. *For all odd numbers $d > 1$, $\mathbb{D}(d) \neq 0$.*

This conjecture is true for d prime according to the next proposition. We have verified the conjecture for $d < 1500$ using Mathematica.

Proposition 23. *If s is an odd prime, then $\mathbb{D}(s) = \pm 2^{(s-3)/2} h_s/s$, where h_s is the first factor of the class number of the cyclotomic ring $\mathbb{Z}[\zeta_s]$. Thus $\mathbb{D}(s) \neq 0$.*

Proof.

The result follows from Equations (1.7), (2.3), (2.4), and (2.5) of [Carlitz and Olson 1955]. $\square$

Theorem 24. *Let $d > 1$ be a fixed odd integer for which $\mathbb{D}(d) \neq 0$. Suppose $\sigma \in \mathcal{S}$ has all jumps at rational points whose denominator divides d. If $\Sigma_p(\sigma) = 0$ for all odd p, then $\sigma = 0$.*

Proof. We have $\sigma = \sum_{j=1}^{(d-1)/2} a_j S_{j/d}$ for some a_j. Since $\Sigma_p(\sigma) = 0$ for all odd p, we have $\int_0^1 \sigma(x)\,dx = 0$. We pick an odd integer $p(i)$ congruent to i modulo p for every i in the range $0 \le i \le (d-1)/2$. For each i, we have $\Sigma_{p(i)}(\sigma) - p(i) \int_0^1 \sigma(x)\,dx = 0$. Using Equations (5-1) and (5-2), this gives us the linear equation $\sum_{j=1}^{(d-1)/2} a_j F_i(j/d) = 0$. The resulting system of $(d-1)/2$ equations in the $(d-1)/2$ unknowns a_j has only the trivial solution if $\mathbb{D}(d) \neq 0$. $\square$

Corollary 25. *Suppose $d > 1$ is an odd integer and $\mathbb{D}(d) \neq 0$. A nonzero knot signature function satisfying the* 1*-signature averaging condition cannot have jumps only at points whose denominator is a divisor of d.*

Since knot signature functions cannot jump at points with prime denominators [Tristram 1969], Proposition 23 does not say anything about knots, except to the extent that it makes Conjecture 22 plausible.

Appendix A: Witt invariants of cable knots

The proof of Theorem 15 follows fairly readily from work of Litherland, some basic knot theoretic results, and consideration of Witt groups.

We begin with an observation: if $\mathbb{S}$ is a satellite of K with orbit P and winding number n, then for an appropriate choice of Seifert surfaces for K, P, and $\mathbb{S}$, the Seifert matrix for $\mathbb{S}$ is the direct sum of a Seifert matrix for P and one for $C_{n,1}$. The construction of the Seifert surfaces for a satellite knot, which leads to the above result, goes back to [Seifert 1950].

Thus, to prove Theorem 15 we need only prove the following:

Theorem 26. *For $C_{(n,1)}$, the $(n, 1)$-cable of C,*

$$w_{C_{(n,1)}}(j/p) = w_C(nj/p).$$

Proof. The proof is largely contained in a diagram; note in the following description that the central square of the diagram is not apparently commutative, while one has commutativity around the other interior faces of the diagram.

$$\begin{tikzcd}
\mathscr{C} \arrow[r, "\alpha"] \arrow[d, "\lambda_n"] & \mathscr{G} \arrow[r, "\beta"] \arrow[d, "\lambda'_n"] & W(\mathbb{Q}[t,t^{-1}]_{(\phi_p)}) \arrow[r, "\gamma"] \arrow[d, "\eta'_n"] & W(\mathbb{Q}(t)) \arrow[d, "\eta_n"] \\
\mathscr{C} \arrow[r, "\alpha"] & \mathscr{G} \arrow[r, "\beta"] \arrow[dr, "\rho'"'] & W(\mathbb{Q}[t,t^{-1}]_{(\phi_p)}) \arrow[r, "\gamma"] \arrow[d, "\rho"] & W(\mathbb{Q}(t)) \\
 & & W(\mathbb{Q}(\zeta_p)) &
\end{tikzcd}$$

Here is the notation and necessary background:

- $\mathscr{C}$ is the concordance group; $\mathscr{G}$ is Levine's algebraic concordance group of Seifert matrices; α is the homomorphism induced by $K \to V_K$.
- $W(\mathbb{Q}[t,t^{-1}]_{(\phi_p)})$ is the Witt group of the localization of $\mathbb{Q}[t,t^{-1}]$ at the p-cyclotomic polynomial ϕ_p (that is, the domain formed by inverting all polynomials relatively prime to ϕ_p); β is the map induced by
$$V \to (1-t)V + (1-t^{-1})V^t.$$
- $W(\mathbb{Q}(t))$ is the Witt group of the field of fractions of $\mathbb{Q}[t,t^{-1}]$; γ is induced by inclusion. The inclusion map is injective (see [Milnor and Husemoller 1973, Corollary IV 3.3] in the symmetric case, and [Ranicki 1981, Proposition 4.2.1 iii] for the hermitian case that arises here).
- λ_n is the function induced by forming the $(n, 1)$-cable; λ'_n is the homomorphism induced by λ_n. This map can be given explicitly in terms of Seifert matrices. That this induces a map on $\mathscr{G}$ and that the map is a homomorphism is elementary (see [Cha et al. 2008; Kawauchi 1980] for further discussion).

- The map ρ is induced by the map $t \to \zeta_p$.
- The map η_n (respectively η'_n) is induced by the embedding of $\mathbb{Q}(t)$ (respectively $\mathbb{Q}[t, t^{-1}]_{(\phi_p)}$) into itself which sends t to t^n.

The proof of Theorem 15 is seen to be equivalent to showing that

$$\rho' \circ \alpha \circ \lambda_n = \rho \circ \eta'_n \circ \beta \circ \alpha.$$

By writing $\rho' = \rho \circ \beta$, we see this will follow from

$$\beta \circ \alpha \circ \lambda_n = \eta'_n \circ \beta \circ \alpha.$$

According to Litherland [1984], we have

$$\gamma \circ \beta \circ \alpha \circ \lambda_n = \eta_n \circ \gamma \circ \beta \circ \alpha.$$

Using commutativity of the rightmost square, we have $\eta_n \circ \gamma = \gamma \circ \eta'_n$, so Litherland's equality can be rewritten as

$$\gamma \circ \beta \circ \alpha \circ \lambda_n = \gamma \circ \eta'_n \circ \beta \circ \alpha.$$

Finally, because γ is injective, this implies $\beta \circ \alpha \circ \lambda_n = \eta'_n \circ \beta \circ \alpha$, as desired. □

Appendix B: One approach to Theorem 17

Let $\mathbb{Q}' = \{r/s \in \mathbb{Q} \mid \gcd(s, r) = \gcd(s, |\mathrm{Torsion}(H_1(R))|) = 1\}$.

Lemma 27. *If T is a finitely generated torsion group, and the prime divisors of $|T|$ are all divisors of $|\mathrm{Torsion}(H_1(R))|$, then $T \otimes (\mathbb{Q}'/\mathbb{Z}) = 0$, and $\mathrm{Tor}(T, \mathbb{Q}'/\mathbb{Z}) = 0$.*

Proof. It suffices to prove this for T a finite cyclic group of order k relatively prime to all the denominators of elements of $\mathbb{Q}'$. From the short exact sequence

$$0 \to \mathbb{Z} \xrightarrow{k\cdot} \mathbb{Z} \to T \to 0,$$

we obtain

$$0 \to \mathrm{Tor}(T, \mathbb{Q}'/\mathbb{Z}) \to \mathbb{Q}'/\mathbb{Z} \xrightarrow{k\cdot} \mathbb{Q}'/\mathbb{Z} \to T \otimes \mathbb{Q}'/\mathbb{Z} \to 0.$$

Suppose s is a denominator of an element in $\mathbb{Q}'$; then $\gcd(k, s) = 1$, so there exists $a, b \in \mathbb{Z}$ such that $ka + sb = 1$. It follows that $k \cdot a/s \equiv 1/s \pmod 1$. Thus $k\cdot : \mathbb{Q}'/\mathbb{Z} \to \mathbb{Q}'/\mathbb{Z}$ is surjective. It is easy to see that it is also injective. □

Lemma 28. *A short exact sequence of the form*

$$0 \to T_1 \xrightarrow{\psi} T_2 \oplus F_2 \xrightarrow{\phi} T_3 \oplus F_3 \to 0,$$

where the F_i are free abelian groups and the T_i are torsion groups, induces a short exact sequence

$$0 \to T_1 \xrightarrow{\pi_{T_2} \circ \psi} T_2 \xrightarrow{\phi_{|T_2}} T_3 \to 0.$$

Proof. Exactness on the left and at the middle of this sequence is immediate. We only need to show that $\phi_{|T_2}$ is surjective. Let $x \in T_3$; there exist $(y, z) \in T_2 \oplus F_2$ with $\phi((y, z)) = x$. We wish to show that $z = 0$. There exist nonzero integers n and m such that $nx = 0$ and $my = 0$. Then $\phi((0, mnz)) = \phi((mny, mnz)) = mnx = 0$. By exactness of the original sequence, $(0, mnz) \in \psi(T_1)$. Since $z \in F_2$, we have that $z = 0$. □

Lemma 29. *If* $\mathcal{T} = \text{Torsion}(H_1(R))$ *and* H *denotes the kernel of* $H_1(F) \to H_1(R)/\mathcal{T}$, *then* $H \otimes \mathbb{Q}'/\mathbb{Z}$ *is the kernel of the natural map* $H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to H_1(R) \otimes (\mathbb{Q}'/\mathbb{Z})$.

Proof. Let $\mathcal{I}$ and $\hat{\mathcal{I}}$ be the images of $H_1(F) \to H_1(R)$ and $H_1(F) \to H_1(R)/\mathcal{T}$, respectively. We have a short exact sequence

$$0 \to H \to H_1(F) \to \hat{\mathcal{I}} \to 0.$$

As $\hat{\mathcal{I}}$ is free abelian, $\text{Tor}(\hat{\mathcal{I}}, \mathbb{Q}'/\mathbb{Z}) = 0$, and we then have a short exact sequence

$$0 \to H \otimes (\mathbb{Q}'/\mathbb{Z}) \to H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to \hat{\mathcal{I}} \otimes (\mathbb{Q}'/\mathbb{Z}) \to 0.$$

Let $\mathcal{R}$ denote $H_1(R)$, and note that $\mathcal{I}/(\mathcal{I} \cap \mathcal{T}) = \hat{\mathcal{I}}$. Consider the lattice of subgroups consisting of $\mathcal{R}$, $\mathcal{I}$, $\mathcal{T}$, and $\mathcal{I} \cap \mathcal{T}$. Their inclusions fit into the following commutative diagram with exact rows and columns:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & \mathcal{I} \cap \mathcal{T} \arrow[r] \arrow[d] & \mathcal{T} \arrow[r] \arrow[d] & \mathcal{T}/(\mathcal{I} \cap \mathcal{T}) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathcal{I} \arrow[r] \arrow[d] & \mathcal{R} \arrow[r] \arrow[d] & \mathcal{R}/\mathcal{I} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \hat{\mathcal{I}} \arrow[r] \arrow[d] & \mathcal{R}/\mathcal{T} \arrow[r] \arrow[d] & (\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}} \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd}$$

To see exactness, view the first two columns as the inclusion of one chain complex into another. The third column is the quotient chain complex. Thus we have a short exact sequence of chain complexes. The first two chain complexes are clearly exact. It follows that the third column is exact, using the associated long exact sequence of homology groups.

Using the long exact sequence of the pair (R, F), we may identify $\mathcal{R}/\mathcal{I}$ with $H_1(R, F)$. Using Lefschetz duality and the universal coefficient theorem, we have $H_1(R, F) \approx H^2(R, \Delta) \approx H^2(R) \approx \mathcal{T} \oplus \mathbb{Z}^{\beta_2(R)}$. With these identifications, the last column of the diagram becomes a short exact sequence

$$0 \to \mathcal{T}/(\mathcal{I} \cap \mathcal{T}) \to \mathcal{T} \oplus \mathcal{F} \to \mathrm{Torsion}((\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}}) \oplus \mathcal{F}' \to 0,$$

where $\mathcal{F}$ and $\mathcal{F}'$ are free abelian groups. By Lemma 28, there is a short exact sequence

$$0 \to \mathcal{T}/(\mathcal{I} \cap \mathcal{T}) \to \mathcal{T} \to \mathrm{Torsion}((\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}}) \to 0.$$

We conclude that $|\mathrm{Torsion}(\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}})| = |\mathcal{I} \cap \mathcal{T}|$. By Lemma 27, we have

$$\mathrm{Tor}((\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}}, \mathbb{Q}'/\mathbb{Z}) = \mathrm{Tor}(\mathrm{Torsion}((\mathcal{R}/\mathcal{T})/\hat{\mathcal{I}}), \mathbb{Q}'/\mathbb{Z}) = 0,$$

so the sequence obtained from the last row of the diagram upon tensoring with $\mathbb{Q}'/\mathbb{Z}$ is exact. In particular, the map $\hat{\mathcal{I}} \otimes (\mathbb{Q}'/\mathbb{Z}) \to (H_1(R)/\mathcal{T}) \otimes (\mathbb{Q}'/\mathbb{Z})$ is injective. It follows that $H \otimes (\mathbb{Q}'/\mathbb{Z})$, the kernel of $H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to \hat{\mathcal{I}} \otimes (\mathbb{Q}'/\mathbb{Z})$, is the same as the kernel of $H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to (H_1(R)/\mathcal{T}) \otimes (\mathbb{Q}'/\mathbb{Z})$.

Considering the middle column, we obtain the following exact sequence:

$$\mathcal{T} \otimes (\mathbb{Q}'/\mathbb{Z}) \to H_1(R) \otimes (\mathbb{Q}'/\mathbb{Z}) \to H_1(R)/\mathcal{T} \otimes (\mathbb{Q}'/\mathbb{Z}) \to 0.$$

Since $\mathcal{T} \otimes (\mathbb{Q}'/\mathbb{Z}) = 0$ by Lemma 27, we see that

$$H_1(R) \otimes (\mathbb{Q}'/\mathbb{Z}) \to (H_1(R)/\mathcal{T}) \otimes (\mathbb{Q}'/\mathbb{Z})$$

is injective. Thus the kernel of $H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to (H_1(R)/\mathcal{T}) \otimes (\mathbb{Q}'/\mathbb{Z})$ is also the kernel of $H_1(F) \otimes (\mathbb{Q}'/\mathbb{Z}) \to H_1(R) \otimes (\mathbb{Q}'/\mathbb{Z})$. □

The second to last sentence of [Gilmer 1993, page 6] asserts without justification, in the situation of Lemma 29, that $H \otimes \mathbb{Q}/\mathbb{Z}$ is the kernel of the natural map $H_1(F) \otimes (\mathbb{Q}/\mathbb{Z}) \to H_1(R) \otimes (\mathbb{Q}/\mathbb{Z})$. The original proof of the theorem may then be modified using Lemma 29 and replacing $\mathbb{Q}/\mathbb{Z}$ by $\mathbb{Q}'/\mathbb{Z}$ judiciously. This proof then yields the conclusion: ${A^q}_p \cap (H \otimes \mathbb{Q}/\mathbb{Z})$ (in the notation of [Gilmer 1993]) is equal to $\mathfrak{m}(q, \Delta)_{(p)}$ for primes p relatively prime to $|\mathrm{Torsion}(H_1(R)|$. This, in turn, can be rephrased as Theorem 17.

Appendix C: Another approach to Theorem 17

C.1. ***Notation.***

- K is a slice knot with a genus g Seifert surface F; K bounds a slice disk Δ; $R \subset B^4$ is a 3-manifold bounded by $F \cup \Delta$.
- S_q is the q-fold branched cover of S^3 branched over K; W_q is the q-fold branched cover of B^4 branched over Δ.
- H is the kernel of $H_1(F) \to H_1(R)/\,\mathrm{Torsion}(H_1(R))$; $\kappa(q, \Delta)$ is the kernel of $H_1(S_q) \to H_1(W_q)$.

We further choose generators for various homology groups:

- $\{x_i'\} \cup \{y_i'\}$ is a symplectic basis of $H_1(F)$ where the x_i generate H.
- F is built from a disk with 1-handles added corresponding to this basis. The dual linking circles to the bands represent homology classes in $H_1(S^3 \setminus F)$ denoted $\{x_i\} \cup \{y_i\}$.

Recall (see [Rolfsen 1976]) that S_q is built from q copies of $S^3 \setminus F$. These copies can be enumerated cyclically, corresponding to translates under the deck transformation. There is a corresponding enumeration of the lifts of F to S_q.

- The lifts of the x_i are denoted $\tilde{x}_{i,\alpha}$, and similarly for the $\tilde{y}_i$, $\tilde{x}_i'$ and $\tilde{y}_i'$. The α are indices denoting the appropriate lift of $S^3 \setminus F$ and F. Here, $\alpha \in \mathbb{Z}_q$.
- $\mathcal{Y}$ denotes the subgroup of $H_1(S_q)$ generated by the $\tilde{y}_{i,\alpha}$. Similarly for $\mathcal{X}$, $\mathcal{Y}'$, and $\mathcal{X}'$.
- $\mathbb{Y}$ denotes the subgroup of $H_1(S_q)$generated by a single set of lifts $\{\tilde{y}_{i,0}\}$.

C.2. ***Statement and proof summary.*** Theorem 17 can now be stated succinctly: if p is relatively prime to the order of Torsion$(H_1(R))$, then $\kappa(\Delta, q)_{(p)} = \mathbb{Y}_{(p)}$. The proof has several steps:

- Lemma 30: $H_1(S_q)_{(p)} = \mathcal{Y}_{(p)} \oplus \mathcal{X}_{(p)}$ and $|\mathcal{Y}_{(p)}| = |\mathcal{X}_{(p)}|$.
- Lemma 31: $\mathcal{X}'_{(p)} = \mathcal{Y}_{(p)}$.
- Lemma 32: $\mathcal{X}'_{(p)} \subset \kappa(\Delta, q)_{(p)}$ and $|\kappa(\Delta, q)_{(p)}|^2 = |H_1(S_q)_{(p)}|$.
- Lemma 33: $\mathbb{Y}_{(p)} = \mathcal{Y}_{(p)}$.

Proof of Theorem 17. We want to show that $\kappa(\Delta, q)_{(p)} = \mathbb{Y}_{(p)}$. By Lemma 33, this is equivalent to showing that $\kappa(\Delta, q)_{(p)} = \mathcal{Y}_{(p)}$. By Lemmas 30 and 32, the orders of these two groups are the same. By Lemmas 31 and 32, $\mathcal{Y}_{(p)} \subset \kappa(\Delta, q)$, and the proof is complete. $\square$

C.3. ***Proofs of lemmas.***

Lemma 30. *$H_1(S_q)_{(p)} = \mathcal{Y}_{(p)} \oplus \mathcal{X}_{(p)}$ and $|\mathcal{Y}_{(p)}| = |\mathcal{X}_{(p)}|$.*

Proof. We use the convention that the Seifert form V is the pairing $V(a, b) = \text{link}(i_+(a), b)$, where i_+ is the positive push-off. For transformations, we have matrices acting on the left; in presentation matrices, the rows give the relations.

The Seifert matrix of V for the surface F with respect to the basis $\{x_i'\} \cup \{y_i'\}$ for $H_1(F)$ is of the form

$$\begin{pmatrix} 0 & M \\ M^t + I & B \end{pmatrix}$$

for some g dimensional square matrices M and B, with B symmetric.

The first homology of S_q is generated by (all) the lifts of the x_i and y_i, which we have denoted $\tilde{x}_{i,\alpha}$ and $\tilde{y}_{i,\alpha}$. As described, for instance in [Rolfsen 1976, page 213], a presentation matrix of the first homology of S_q with respect to this basis is determined by V. In this case, the result is a matrix of the form

$$\begin{pmatrix} 0 & \mathcal{M} \\ \mathcal{M}' & \mathcal{B} \end{pmatrix},$$

where $\mathcal{M}$ and $\mathcal{B}$ are qg dimensional matrices that are built out of the blocks of V as follows (we illustrate in the case $q = 3$):

$$\mathcal{M} = \begin{pmatrix} M+I & -M & 0 \\ 0 & M+I & -M \\ -M & 0 & M+I \end{pmatrix}, \quad \mathcal{M}' = \begin{pmatrix} M^t & -M^t - I & 0 \\ 0 & M^t & -M^t - I \\ -M^t - I & 0 & M^t \end{pmatrix},$$

$$\mathcal{B} = \begin{pmatrix} B & -B & 0 \\ 0 & B & -B \\ -B & 0 & B \end{pmatrix}.$$

The first columns correspond to the $\tilde{x}_{i,\alpha}$ and the later columns to the $\tilde{y}_{i,\alpha}$.

Notice first that $|\mathcal{M}| = |\mathcal{M}'|$ and $|\mathcal{M}|^2 = |H_1(S_q)|$.

Forming the quotient, $H_1(S_q)/\mathcal{Y}$ yields a group $\overline{\mathcal{X}}$ generated by the image of $\mathcal{X}$. This quotient is presented by $\mathcal{M}'$, and thus has order $\sqrt{|H_1(S_q)|}$, so $\mathcal{X}$ has order at least this large. Thus $|\mathcal{X}_{(p)}|^2 \geq |H_1(S_q)_{(p)}|$. On the other hand, since $\mathcal{X}_{(p)}$ is a self-annihilating subgroup for a nonsingular form, $|\mathcal{X}_{(p)}|^2 \leq |H_1(S_q)_{(p)}|$.

We now have $|\mathcal{X}_{(p)}|^2 = |H_1(S_q)_{(p)}|$, and thus $|\mathcal{X}_{(p)}| = |\overline{\mathcal{X}}_{(p)}|$. From this we can conclude that $\mathcal{X}_p \cap \mathcal{Y}_{(p)} = 0$, so $H_1(S_q)_{(p)} = \mathcal{X}_{(p)} \oplus \mathcal{Y}_{(p)}$. □

Lemma 31. $\mathcal{X}'_{(p)} = \mathcal{Y}_{(p)}$.

Proof. The positive and negative push-off maps $i_\pm : H_1(F) \to S^3 \setminus F$ send the span of the x'_i to the span of the y_i. Denote the restriction of these maps by $j_\pm : \langle\{x'_i\}\rangle \to \langle\{y_i\}\rangle$. With respect to these bases, the maps $j_\pm$ are given by the matrices M^t and $M^t + I$. Now view these matrices as defining maps from $\mathbb{Z}^g$ to itself with M^t corresponding to an automorphism T. Then any element $y \in \mathbb{Z}^g$ can be written $y = \mathrm{Id}(y) = (T+\mathrm{Id})(y) - T(y)$. Thus, $\mathrm{Image}(j_+) + \mathrm{Image}(j_-) = \mathrm{Span}(\{y_i\})$. Lifting to the q-fold branched covers, we see that the $\tilde{y}_{i,\alpha}$ are all in the image of the $\tilde{x}'_{i,\alpha}$ (in more detail, each $\tilde{y}_{i,\alpha}$ is in the span of the images of the $\{\tilde{x}'_{i,\alpha}\}$ and $\{\tilde{x}'_{i,\alpha+1}\}$). Also, the images of the $\tilde{x}'_{i,\alpha}$ are all in $\mathrm{Span}(\{\tilde{y}_{i,\alpha}\})$. The same thus holds on the level of the p-torsion, completing the proof of the lemma. □

Lemma 32. $\mathcal{X}'_{(p)} \subset \kappa(\Delta, q)_{(p)}$ *and* $|\kappa(\Delta, q)_{(p)}|^2 = |H_1(S_q)_{(p)}|$.

Proof. Let $\gamma = |\mathrm{Torsion}(H_1(R))|$. Then $\gamma z = 0 \in H_1(R)$ for all $z \in H$. Lifting, we see that $z' \in \mathcal{X}'_{(p)}$ for all $\gamma z' = 0 \in H_1(W_q)$, so $\gamma \mathcal{X}'_{(p)} \subset \kappa(\Delta, q)_{(p)}$. But multiplication by γ is an isomorphism on $\mathcal{X}'_{(p)}$ since p is relatively prime to γ.

We have from Theorem 16 that $|\kappa(\Delta, q)|^2 = |H_1(S_q)|$, so the same holds for the p-torsion. □

Lemma 33. $\mathbb{Y}_{(p)} = \mathcal{Y}_{(p)}$.

Proof. Let $\Lambda = \mathbb{Z}[\mathbb{Z}_q]$, the group ring of the cyclic group. We write $\mathbb{Z}_q$ multiplicatively, generated by t. The standard derivation of a presentation of the homology $H_1(S_q)$, such as in [Rolfsen 1976], is a Mayer–Vietoris argument. The homology groups involved are all modules over Λ, where t acts by the deck transformation. From this viewpoint, the Mayer–Vietoris sequence now yields that as a Λ-module the homology is given as a quotient $H_1(S_q) \cong \Lambda^{2g}/(V - tV^t)\Lambda^{2g}$.

Since $V - V^t$ is invertible, we can multiply the quotienting submodule by $(V - V^t)^{-1}$ without changing the quotient space. Some elementary algebra then shows that

$$H_1(S_q) \cong \Lambda^{2g}/(\Gamma + t(I - \Gamma))\Lambda^{2g},$$

where $\Gamma = (V - V^t)^{-1}V$.

It is clear from this that for any $z \in \Lambda^{2g}$, we have $\Gamma z = t(\Gamma - I)z \in H_1(S_q)$. Thus, $\Gamma^q z = t^q(\Gamma - I)^q z \in H_1(S_q)$. However, $t^q = 1$, so $\Gamma^q - (\Gamma - I)^q$ annihilates $H_1(S_q)$.

Expanding, we have that for some polynomial f with constant term 0 and of degree $q-1$, the action of $f(\Gamma)$ on $H_1(S_q)$ coincides with I. The leading coefficient of f is q. If p does not divide the order $|H_1(S_q)|$, the lemma is immediately true, so assume p divides the order $|H_1(S_q)|$. We know that p is relatively prime to q. Thus, we can switch to $\mathbb{Z}_{(p)}$-coefficients, in which case the leading coefficient of f is a unit, and we see that with $\mathbb{Z}_{(p)}$-coefficients, Γ is invertible.

We now focus on the Seifert matrix V of the algebraically slice knot. In the coordinates we have been using, we see that

$$\Gamma = \begin{pmatrix} M^t + I & B \\ 0 & -M \end{pmatrix}.$$

From this we conclude that with $\mathbb{Z}_{(p)}$-coefficients, M and $M + I$ are both invertible. Recall that for each k, M and $M + I$ determine the maps from $\operatorname{Span}(\tilde{x}'_{i,k})$ and $\operatorname{Span}(\tilde{x}'_{i,k+1})$ to $\operatorname{Span}(\tilde{y}_{i,k})$. Thus, any element in $\operatorname{Span}(\tilde{y}_{i,k})$ is also in $\operatorname{Span}(\tilde{y}_{i,k+1})$. This completes the proof of the lemma. □

Acknowledgments

We thank John Ewing for his help with the number theoretic aspects of this project. In particular, John pointed out to us the relationship between the determinant $\mathbb{D}(s)$ and class numbers. We also appreciate Stefan Friedl's careful reading and thoughtful commentary on our initial draft of this article.

References

[Carlitz and Olson 1955] L. Carlitz and F. R. Olson, "Maillet's determinant", *Proc. Amer. Math. Soc.* **6**:2 (1955), 265–269. MR 16,999d Zbl 0065.02703

[Casson and Gordon 1978] A. Casson and C. M. Gordon, Lecture given by C. McA. Gordon at Cambridge University, 1978.

[Casson and Gordon 1986] A. J. Casson and C. M. Gordon, "Cobordism of classical knots", pp. 181–199 in *À la recherche de la topologie perdue*, edited by L. Guillou and A. Marin, Progr. Math. **62**, Birkhäuser, Boston, MA, 1986. MR 900252 Zbl 0597.57001

[Cha and Livingston 2011] J. C. Cha and C. Livingston, "KnotInfo: Table of knot invariants", 2011, Available at http://www.indiana.edu/∼knotinfo.

[Cha et al. 2008] J. C. Cha, C. Livingston, and D. Ruberman, "Algebraic and Heegaard–Floer invariants of knots with slice Bing doubles", *Math. Proc. Cambridge Philos. Soc.* **144**:2 (2008), 403–410. MR 2009h:57013 Zbl 1144.57002

[Cochran et al. 2003] T. D. Cochran, K. E. Orr, and P. Teichner, "Knot concordance, Whitney towers and L^2-signatures", *Ann. of Math.* (2) **157**:2 (2003), 433–519. MR 2004i:57003 Zbl 1044.57001

[Cooper 1982] D. Cooper, *Signatures of surfaces with applications to knot and link cobordism*, Ph.D. thesis, University of Warwick, 1982, Available at http://library.msri.org/books/cooper.

[Freedman and Quinn 1990] M. H. Freedman and F. Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series **39**, Princeton University Press, 1990. MR 94b:57021 Zbl 0705.57001

[Friedl 2004] S. Friedl, "Eta invariants as sliceness obstructions and their relation to Casson–Gordon invariants", *Algebr. Geom. Topol.* **4** (2004), 893–934. MR 2005j:57016 Zbl 1067.57003

[Gilmer 1983] P. M. Gilmer, "Slice knots in S^3", *Quart. J. Math. Oxford Ser.* (2) **34**:135 (1983), 305–322. MR 85d:57004 Zbl 0542.57007

[Gilmer 1993] P. Gilmer, "Classical knot and link concordance", *Comment. Math. Helv.* **68**:1 (1993), 1–19. MR 94c:57007 Zbl 0805.57005

[Gilmer and Livingston 1992a] P. Gilmer and C. Livingston, "The Casson–Gordon invariant and link concordance", *Topology* **31**:3 (1992), 475–492. MR 93h:57037 Zbl 0797.57001

[Gilmer and Livingston 1992b] P. Gilmer and C. Livingston, "Discriminants of Casson–Gordon invariants", *Math. Proc. Cambridge Philos. Soc.* **112**:1 (1992), 127–139. MR 94e:57007 Zbl 0791.57015

[Hardy and Wright 1938] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford University Press, 1938. Zbl 0020.29201

[Hedden 2007] M. Hedden, "Knot Floer homology of Whitehead doubles", *Geom. Topol.* **11** (2007), 2277–2338. MR 2008m:57030 Zbl 1187.57015

[Hedden 2009] M. Hedden, "On knot Floer homology and cabling, II", *Int. Math. Res. Not.* **2009**:12 (2009), 2248–2274. MR 2011f:57015 Zbl 1172.57008

[Kauffman 1987] L. H. Kauffman, *On knots*, Annals of Mathematics Studies **115**, Princeton University Press, 1987. MR 89c:57005 Zbl 0627.57002

[Kawauchi 1980] A. Kawauchi, "On links not cobordant to split links", *Topology* **19**:4 (1980), 321–334. MR 82d:57011 Zbl 0805.57005

[Kirby 1978] R. Kirby, "Problems in low dimensional manifold theory", pp. 273–312 in *Algebraic and geometric topology* (Stanford, CA, 1976), edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, R.I., 1978. MR 80g:57002 Zbl 0394.57002

[Levine 1969] J. Levine, "Invariants of knot cobordism", *Invent. Math.* **8** (1969), 98–110. MR 40 #6563 Zbl 0179.52401

[Litherland 1984] R. A. Litherland, "Cobordism of satellite knots", pp. 327–362 in *Four-manifold theory* (Durham, NH, 1982)), edited by C. Gordon and R. Kirby, Contemp. Math. **35**, Amer. Math. Soc., Providence, RI, 1984. MR 86k:57003 Zbl 0563.57001

[Livingston and Melvin 1985] C. Livingston and P. Melvin, "Abelian invariants of satellite knots", pp. 217–227 in *Geometry and topology* (College Park, MD, 1983/84)), edited by J. Alexander and J. Harer, Lecture Notes in Math. **1167**, Springer, Berlin, 1985. MR 87e:57009 Zbl 0605.57001

[Milnor and Husemoller 1973] J. Milnor and D. Husemoller, *Symmetric bilinear forms*, Ergebnisse der Mathematik und ihrer Grenzgebiete **73**, Springer, New York, 1973. MR 58 #22129 Zbl 0292.10016

[Naik 1996] S. Naik, "Casson–Gordon invariants of genus one knots and concordance to reverses", *J. Knot Theory Ramifications* **5**:5 (1996), 661–677. MR 97h:57020 Zbl 0890.57014

[Ozsváth and Szabó 2003] P. Ozsváth and Z. Szabó, "Knot Floer homology and the four-ball genus", *Geom. Topol.* **7** (2003), 615–639. MR 2004i:57036 Zbl 1037.57027

[Ranicki 1981] A. Ranicki, *Exact sequences in the algebraic theory of surgery*, Mathematical Notes **26**, Princeton University Press, 1981. MR 82h:57027 Zbl 0471.57012

[Rolfsen 1976] D. Rolfsen, *Knots and links*, Mathematics Lecture Series **7**, Publish or Perish, Berkeley, 1976. MR 58 #24236 Zbl 0339.55004

[Seifert 1950] H. Seifert, "On the homology invariants of knots", *Quart. J. Math., Oxford Ser.* (2) **1** (1950), 23–32. MR 11,735b Zbl 0035.11103

[Tristram 1969] A. G. Tristram, "Some cobordism invariants for links", *Proc. Cambridge Philos. Soc.* **66** (1969), 251–264. MR 40 #2104 Zbl 0191.54703

Received September 7, 2012. Revised December 27, 2012.

PATRICK M. GILMER
DEPARTMENT OF MATHEMATICS
LOUISIANA STATE UNIVERSITY
376 LOCKETT HALL
BATON ROUGE, LA 70803
UNITED STATES
gilmer@math.lsu.edu

CHARLES LIVINGSTON
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
RAWLES HALL
BLOOMINGTON, IN 47405-5701
UNITED STATES
livingst@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.427

CHARACTERIZING ABELIAN VARIETIES BY THE REDUCTION OF THE MORDELL–WEIL GROUP

CHRIS HALL AND ANTONELLA PERUCCA

Let A be an abelian variety defined over a number field K. Let $\mathfrak{p}$ be a prime of K of good reduction and $A_{\mathfrak{p}}$ the fiber of A over the residue field $k_{\mathfrak{p}}$. We call $A(K)_{\mathfrak{p}}$ the image of the Mordell–Weil group via reduction modulo $\mathfrak{p}$, which is a subgroup of $A_{\mathfrak{p}}(k_{\mathfrak{p}})$. We prove in particular that the size of $A(K)_{\mathfrak{p}}$, by varying $\mathfrak{p}$, encodes enough information to characterize the K-isogeny class of A, provided that the following necessary condition holds: the Mordell–Weil group $A(K)$ is Zariski dense in A. This is an analogue to a 1983 result of Faltings, considering instead the size of $A_{\mathfrak{p}}(k_{\mathfrak{p}})$.

1. Introduction

Statement of the theorems. Let K be a number field and A, A' be abelian varieties over K. Let $S(A, A')$ be the set of primes of K of good reduction for A and A', and let $A_{\mathfrak{p}}$, $A'_{\mathfrak{p}}$ be the respective fibers of A, A' over the residue field $k_{\mathfrak{p}}$ for $\mathfrak{p} \in S(A, A')$.

Faltings [1983] proved the following local-global principle for any $S \subseteq S(A, A')$ of Dirichlet density 1: A, A' are K-isogenous if and only if $A_{\mathfrak{p}}$, $A'_{\mathfrak{p}}$ are $k_{\mathfrak{p}}$-isogenous for every $\mathfrak{p} \in S$. The latter is equivalent, for a large class of abelian varieties, to the identities $\#A_{\mathfrak{p}}(k_{\mathfrak{p}}) = \#A'_{\mathfrak{p}}(k_{\mathfrak{p}})$ for $\mathfrak{p} \in S$. The motivation for this paper was to consider instead identities using the reductions of the Mordell–Weil groups $A(K)$, $A'(K)$, which we denote by $A(K)_{\mathfrak{p}}$, $A'(K)_{\mathfrak{p}}$ and which are subgroups of $A_{\mathfrak{p}}(k_{\mathfrak{p}})$, $A'_{\mathfrak{p}}(k_{\mathfrak{p}})$. We prove in particular the following result:

Theorem 1.1. *Suppose A, A' are abelian varieties over a number field K such that $A(K)$, $A'(K)$ are Zariski dense in A, A', respectively. Let $S \subseteq S(A, A')$ have Dirichlet density* 1. *If $\#A(K)_{\mathfrak{p}} = \#A'(K)_{\mathfrak{p}}$ holds for every $\mathfrak{p} \in S$, then A and A' are K-isogenous.*

In other words, if $A(K)$ is Zariski dense in A, then the function $\mathfrak{p} \in S \mapsto \#A(K)_{\mathfrak{p}}$ characterizes the K-isogeny class of A. Note, we define this function via a global object, namely the Mordell–Weil group $A(K)$, and it only "sees" the Zariski closure of $A(K)$, hence the reason we assume $A(K)$ is Zariski dense in A. This assumption

MSC2010: 11G05, 11G10.
Keywords: abelian varieties, Mordell–Weil group.

is equivalent to the following: for every nontrivial abelian subvariety $B \subseteq A$, the Mordell–Weil group $B(K)$ is infinite.

For each prime number l and finite group G, we denote by $\mathrm{ord}_l(G)$, $\exp_l(G)$, and $\mathrm{rad}_l(G)$ the l-adic valuation of the order, exponent, and radical of the order of G, respectively. (Recall that the radical of a positive integer is the product of the primes dividing it, the empty product being 1; so $\mathrm{rad}_l(G)$ is 1 or 0 depending on whether or not l divides the order of G.)

We say an abelian variety A is *square-free* if and only if $B = 0$ is the only abelian variety for which there exists a K-homomorphism $B^2 \to A$ with finite kernel. Then we can prove the following stronger result:

Theorem 1.2. *Suppose A, A' are abelian varieties over a number field K such that $A(K)$, $A'(K)$ are Zariski dense in A, A' respectively. Let $S \subseteq S(A, A')$ have Dirichlet density* 1.

(i) *If for some prime number l and for some $m \geq 0$ the inequalities*

$$|\mathrm{ord}_l(A(K)_{\mathfrak{p}}) - \mathrm{ord}_l(A'(K)_{\mathfrak{p}})| \leq m \quad \textit{for all } \mathfrak{p} \in S$$

hold, then A and A' are K-isogenous.

(ii) *Suppose that A, A' are square-free.*
If for some prime number l and for some $m \geq 0$ the inequalities

$$|\exp_l(A(K)_{\mathfrak{p}}) - \exp_l(A'(K)_{\mathfrak{p}})| \leq m \quad \textit{for all } \mathfrak{p} \in S$$

hold, then A and A' are K-isogenous.

(iii) *Suppose that A, A' are square-free.*
There exists l_0 depending only on A, A', K such that if for some prime number $l \geq l_0$ the equalities

$$\mathrm{rad}_l(A(K)_{\mathfrak{p}}) = \mathrm{rad}_l(A'(K)_{\mathfrak{p}}) \quad \textit{for all } \mathfrak{p} \in S$$

hold, then A and A' are K-isogenous.

The assumption on A and A' that we required in the previous statement is in general necessary. Indeed, if C is a nonzero abelian variety having trivial Mordell–Weil group then A and $A' = A \times C$ can not be distinguished with the data as in the statement of the theorem, and indeed $A'(K) = A(K) \times \{0\}$ is not dense in A'. Moreover, we cannot distinguish between A and $A' = A^2$ just by looking at the exponent (or the radical of the size) of $A(K)_{\mathfrak{p}}$.

We generalize the previous result by considering more general objects than the Mordell–Weil group. If A is an abelian variety, we call a subgroup $\Gamma \subseteq A(K)$ *dense* if and only if it is Zariski dense in A, and a *submodule* if and only if it is an $\mathrm{End}_K(A)$-submodule of $A(K)$. We denote $\Gamma_{\mathfrak{p}} \subseteq A(k_{\mathfrak{p}})$ the reduction of Γ modulo $\mathfrak{p}$.

Theorem 1.3. *Let A, A' be abelian varieties defined over a number field K, and $\Gamma \subseteq A(K)$, $\Gamma' \subseteq A'(K)$ be submodules. Suppose Γ is dense. Consider the following property*:

$$\exists \varphi \in \mathrm{Hom}_K(A, A') \text{ such that } \ker(\varphi) \text{ and } [\varphi(\Gamma) : \varphi(\Gamma) \cap \Gamma'] \text{ are finite.} \tag{1}$$

Let $S \subseteq S(A, A')$ have Dirichlet density 1.

(i) *If for some prime number l and for some $m \geq 0$*

$$\mathrm{ord}_l(\Gamma_{\mathfrak{p}}) \leq \mathrm{ord}_l(\Gamma'_{\mathfrak{p}}) + m \quad \text{for all } \mathfrak{p} \in S \tag{2}$$

holds, then (1) *holds.*

(ii) *Suppose that A is square-free. If for some prime number l and for some $m \geq 0$*

$$\exp_l(\Gamma_{\mathfrak{p}}) \leq \exp_l(\Gamma'_{\mathfrak{p}}) + m \quad \text{for all } \mathfrak{p} \in S \tag{3}$$

holds, then (1) *holds.*

(iii) *Suppose that A is square-free. There exists l_0 depending only on A, A', K, Γ, Γ' such that if for some prime number $l \geq l_0$*

$$\mathrm{rad}_l(\Gamma_{\mathfrak{p}}) \leq \mathrm{rad}_l(\Gamma'_{\mathfrak{p}}) \quad \text{for all } \mathfrak{p} \in S \tag{4}$$

holds, then (1) *holds.*

Conversely, property (1) implies that for all primes in $S(A, A')$ and for all but finitely many prime numbers l, we have $\mathrm{ord}_l(\Gamma_{\mathfrak{p}}) \leq \mathrm{ord}_l(\Gamma'_{\mathfrak{p}})$, and similarly for $\exp_l$ and rad_l (compare Lemma 3.4).

If A is a simple abelian variety and $\psi \in \mathrm{End}_K(A)\backslash\mathbb{Z}$, then for any nontorsion point $P \in A(K)$ and sufficiently large prime l, the subgroups $\Gamma = \mathbb{Z}P + \mathbb{Z}\psi(P)$ and $\Gamma' = \mathbb{Z}P$ satisfy $\exp_l(\Gamma_{\mathfrak{p}}) = \exp_l(\Gamma'_{\mathfrak{p}})$ for all $\mathfrak{p} \in S(A)$; however, Γ' does not contain a finite index subgroup of $\varphi(\Gamma)$ for $\varphi \neq 0 \in \mathrm{End}_K(A)$. This is the reason we suppose Γ, Γ' are submodules and not merely subgroups in Theorem 1.3.

Our results relate to the so-called support problem, and especially to the following result:

Theorem 1.4 [Demeyer and Perucca 2013, Theorem 1.2]. *Let A be an abelian variety defined over a number field K, and $P \in A(K)$ be a rational point. Suppose $\mathbb{Z}P$ is Zariski dense in A. If $S \subseteq S(A)$ has Dirichlet density* 1, *then the function $\mathfrak{p} \in S \mapsto \#(\mathbb{Z}P)_{\mathfrak{p}}$ characterizes the K-isomorphism class of A.*

It is not possible to characterize the K-isomorphism class of A by knowing the order and the exponent of $A(K)_{\mathfrak{p}}$ for every $\mathfrak{p} \in S(A)$. Indeed, there exist pairs of elliptic curves over a number field K which are not K-isomorphic, but such that for every prime number l there is a K-isogeny between them of degree coprime to l, as shown by Zarhin [2008, Section 12].

We will deduce Theorems 1.1 and 1.2 from Theorem 1.3. An overview of the proof of Theorem 1.3 is given at the end of this section (page 431). In Section 2 we develop the notion of *almost independent* points to compensate for the fact that the submodules we consider are in general not free. We also define what it means for points to dominate a submodule (page 434), and for an infinite submodule we show how to construct a finite dominating subset consisting of almost independent points. We bring these notions together in Section 3, with preparatory theorems (3.1 and 3.2) about the reduction of submodules and the proof of Theorem 1.3 (page 438).

Notation and conventions. We assume all abelian varieties, subvarieties, homomorphisms, etc. are defined over a fixed number field K. Given abelian varieties $A_1, \dots, A_r$ we write $S(A_1, \dots, A_r)$ for the primes of K of common good reduction.

If A is an abelian variety and $\mathfrak{p}$ is a prime in $S(A)$, we write $k_\mathfrak{p}$ for the residue field and $A_\mathfrak{p}$ for the fiber of A over $k_\mathfrak{p}$. Given a subgroup $\Gamma \subseteq A(K)$, we write $\Gamma_\mathfrak{p} \subseteq A_\mathfrak{p}(k_\mathfrak{p})$ for the reduction of Γ modulo $\mathfrak{p}$. The symbol l always denotes a prime number, and we define $\operatorname{ord}_l(\Gamma_\mathfrak{p})$, $\exp_l(\Gamma_\mathfrak{p})$, $\operatorname{rad}_l(\Gamma_\mathfrak{p})$ to be the l-adic valuation of the size, exponent, and radical of the size of $\Gamma_\mathfrak{p}$ respectively.

By the order of a point we mean the order of the subgroup that it generates.

Two main ingredients. The proof of Theorem 1.3 is based on two main ingredients: Theorem 1 of [Perucca 2011] and a basic structure theorem for abelian varieties, known as Poincaré's reducibility theorem. We recall these statements for the convenience of the reader and for future reference; aside from these two inputs, this paper will be self contained.

Proposition 1.5 [Perucca 2011]. *Let $A_1, \dots, A_r$ be abelian varieties over K, and $P_i \in A_i(K)$ be a rational point for $1 \le i \le r$. If l is a prime number and $e_1, \dots, e_r$ are nonnegative integers, then the set of primes*

$$\{\mathfrak{p} \in S(A_1, \dots, A_r) : \operatorname{ord}_l(P_i \bmod p) = e_i \textit{ for all } i\}$$

admits a Dirichlet density.

If the rational point $P = (P_1, \dots, P_r)$ on $A = A_1 \times \cdots \times A_r$ generates a Zariski dense subgroup, then this Dirichlet density is positive.

Proof. This is the special case of [Perucca 2011, Theorem 1] where all semiabelian varieties are abelian and we consider only one prime number l. The existence of the density (in fact, it is a natural density) is proven there with a method from [Jones and Rouse 2010]. The fact that under the additional assumption the density is nonzero was first proven in [Pink 2004] and can also be proven with a method from [Khare and Prasad 2004]. The proof uses Kummer theory, results on the l-adic representation, and the Chebotarev density theorem. □

Theorem 1.6 (Poincaré's reducibility theorem [Mumford 1970, Theorem 1, p. 160]). *If A is an abelian variety over K and $B \subseteq A$ is an abelian subvariety, then there exists an abelian subvariety $C \subseteq A$ such that $B \cap C$ is finite, $A = B + C$, and A is isogenous to $B \times C$.*

By applying this result finitely many times, we have:

Corollary 1.7. *Let A be an abelian variety over K. There exist pairwise nonisogenous simple abelian varieties $B_1, \ldots, B_r$ uniquely determined up to isogeny and ordering, and positive integers $e_1, \ldots, e_r$ such that A is isogenous to $B_1^{e_1} \times \cdots \times B_r^{e_r}$.*

Overview of the proof of Theorem 1.3. Under the additional hypothesis that A and A' are simple and that Γ and Γ' are each generated by a rational point of infinite order, the proof of Theorem 1.3 becomes technically much easier. We present the proof of (i) in this special case.

Proposition 1.8. *Let A and A' be simple abelian varieties over K and let $\Gamma \subseteq A(K)$ and $\Gamma' \subseteq A'(K)$ be the submodules generated by points of infinite order $P \in A(K)$ and $P' \in A'(K)$ respectively. Let l be a prime number. Suppose that there exists a set $S \subseteq S(A, A')$ of Dirichlet density* 1 *and an integer $m \geq 0$ such that*

$$\operatorname{ord}_l(\Gamma_{\mathfrak{p}}) \leq \operatorname{ord}_l(\Gamma'_{\mathfrak{p}}) + m \quad \textit{for every } \mathfrak{p} \in S \tag{5}$$

holds. Then there exists an isogeny $\varphi : A \to A'$ such that the index $[\varphi(\Gamma) : \varphi(\Gamma) \cap \Gamma']$ is finite.

Proof. Consider the subgroup of $A \times A'$ generated by (P, P'), and denote by $B \subseteq A \times A'$ the connected component of the unity of its Zariski closure. A closed algebraic subgroup of an abelian variety has only finitely many connected components, hence there exists an integer $n \geq 1$ such that (nP, nP') is a rational point of B. Since A is simple and P is of infinite order, the Zariski closure of $\mathbb{Z}P$ in A is equal to A. The projection $\pi : B \to A$ is therefore surjective. For the same reason, the projection $\pi' : B \to A'$ is surjective. Again, because A and A' are simple, there are now two possibilities: either π and π' are isogenies, or else B is equal to $A \times A'$. In the first case, there exists an isogeny $\psi : A \to B$ such that $\psi \circ \pi$ is the multiplication-by-n' endomorphism of B for some nonzero integer n'. The composite isogeny $\varphi := \pi' \circ \psi : A \to A'$ has the required properties, since indeed

$$\varphi(nP) = \pi'(\psi(nP)) = \pi'(\psi(\pi(nP, nP'))) = \pi'(nn'P, nn'P') = nn'P'$$

holds. We are now left to show that (5) excludes the second possibility, that $B = A \times A'$. For this we use Proposition 1.5. Indeed, if $B = A \times A'$, then there exists by Proposition 1.5 a set $S' \subseteq S(A, A')$ of positive Dirichlet density, such that for all $\mathfrak{p} \in S'$,

$$\operatorname{ord}_l(P \bmod \mathfrak{p}) = m + 1 \quad \text{and} \quad \operatorname{ord}_l(P' \bmod \mathfrak{p}) = 0$$

holds. Hence, we have $\mathrm{ord}_l(\Gamma_{\mathfrak{p}}) > m$ and $\mathrm{ord}_l(\Gamma'_{\mathfrak{p}}) = 0$ for all $\mathfrak{p} \in S'$, because on one hand $\Gamma_{\mathfrak{p}}$ contains $(P \bmod \mathfrak{p})$, and on the other hand $\Gamma'_{\mathfrak{p}}$ consists of images under endomorphisms of $(P' \bmod \mathfrak{p})$. The complement of S' has Dirichlet density < 1, contradicting (5). □

2. Preliminaries

Dense submodules. Let A be an abelian variety. We call a subgroup $\Gamma \subseteq A(K)$ a *submodule* if and only if Γ is an $\mathrm{End}(A)$-submodule of $A(K)$. We say that a subgroup of $A(K)$ is *dense* if and only if it satisfies the equivalent conditions of the following lemma:

Lemma 2.1. *If A is an abelian variety and Γ is a subgroup of $A(K)$, then the following are equivalent*:

(i) Γ *is Zariski dense in A.*

(ii) $\varphi(\Gamma) \neq \{0\}$ *for every abelian variety B and* $\varphi \neq 0 \in \mathrm{Hom}(A, B)$.

Proof. Let $C \subseteq A$ be the Zariski closure of Γ. If $\varphi(\Gamma) = 0$ for some nonzero $\varphi \in \mathrm{Hom}(A, B)$, then the kernel of φ is a proper subgroup containing Γ, so (i) implies (ii). Conversely, if $C \neq A$, then the projection $A \to A/C$ is a nonzero morphism between abelian varieties which kills Γ; therefore (ii) implies (i). □

If $\Gamma \subseteq A(K)$ is a finite subgroup, then either Γ is not dense or $A = 0$. If A is simple and if $\Gamma \subseteq A(K)$ is an infinite subgroup, then Γ is dense.

Almost independent points. In this subsection, we suppose $A_1, \dots, A_r$ are nonzero abelian varieties and $P_i \in A_i(K)$ is a rational point for $1 \leq i \leq r$. We let $A = A_1 \times \cdots \times A_r$ and $P = (P_1, \dots, P_r)$.

We say that $P_1, \dots, P_r$ are *independent* if and only if the Zariski closure of $\mathbb{Z}P$ satisfies $\overline{\mathbb{Z}P} = A$. Note that if $A_1 = \cdots = A_r$, the points $P_1, \dots, P_r$ are independent if and only if they form a basis for a free $\mathrm{End}(A_1)$-submodule of $A_1(K)$ (see Definition 3 and Remark 6 in [Perucca 2009]).

Lemma 2.2. *The following are equivalent*

(i) $P_1, \dots, P_r$ *are independent.*

(ii) *For every abelian variety B, the following implication holds*:

$$(6)\quad \sum_{i=1}^{r} \phi_i(P_i) = 0 \textit{ for } (\phi_1, \dots, \phi_r) \in \mathrm{Hom}(A_1 \times \cdots \times A_r, B) \implies \phi_i = 0 \textit{ for all } i.$$

Proof. If $A = A_1 \times \cdots \times A_r$ and $\Gamma = \mathbb{Z}P$, then conditions (i) and (ii) are equivalent to the respective conditions of Lemma 2.2. □

A weaker condition is the following:

Definition 2.3. We say $P_1, \dots, P_r$ are *almost independent* if and only if $\overline{\mathbb{Z}P}_1, \dots, \overline{\mathbb{Z}P}_r$ are nontrivial, connected, and satisfy

$$\overline{\mathbb{Z}P} = \overline{\mathbb{Z}P}_1 \times \overline{\mathbb{Z}P}_2 \times \cdots \times \overline{\mathbb{Z}P}_r.$$

The analogue of (6) for almost independent points is this:

$$\text{(7)} \quad \sum_{i=1}^{r} \phi_i(P_i) = 0 \text{ for } (\phi_1, \dots, \phi_r) \in \operatorname{Hom}(A_1 \times \cdots \times A_r, B) \implies \phi_i(P_i) = 0 \text{ for all } i.$$

Lemma 2.4. *Let $B_1, \dots, B_s$ be simple abelian varieties such that A is isogenous to $B_1 \times \cdots \times B_s$. If $\overline{\mathbb{Z}P}_1, \dots, \overline{\mathbb{Z}P}_r$ are connected and nontrivial, then the following are equivalent*:

(i) *$P_1, \dots, P_r$ are almost independent.*

(ii) *The implication (7) holds for every abelian variety B.*

(iii) *The implication (7) holds for $B = B_1, \dots, B_s$.*

Proof. If $P_1, \dots, P_r$ are almost independent and if B and $\phi \in \operatorname{Hom}(A, B)$ satisfy $\phi(P) = 0$, then $\phi_i(\overline{\mathbb{Z}P}_i) \subseteq \phi(\overline{\mathbb{Z}P}) = 0$, and so $\phi_i(P_i) = 0$ for each i; therefore (i) implies (ii). Conversely, if B is the quotient $A/\overline{\mathbb{Z}P}$ and $\phi : A \to B$ is the natural homomorphism, then (2) implies $\phi_i(\overline{\mathbb{Z}P}_i) = 0$, thus $\overline{\mathbb{Z}P}_1 \times \cdots \times \overline{\mathbb{Z}P}_r \subseteq \overline{\mathbb{Z}P}$. The reverse inclusion is trivial, thus (ii) implies (i).

It is clear (ii) implies (iii). We assume the latter holds and prove the converse. Suppose B is an abelian variety and $\phi \in \operatorname{Hom}(A, B)$ satisfies $\phi(P) = 0$. We must show $\phi_1(P_1) = \cdots = \phi_r(P_r) = 0$. In fact, the only finite quotients of $\overline{\mathbb{Z}P}_1, \dots, \overline{\mathbb{Z}P}_r$ are trivial since they are connected, hence it suffices to show $\phi_1(P_1), \dots, \phi_r(P_r)$ are torsion.

Up to rearranging $B_1, \dots, B_s$, there exists an isogeny $\psi : \phi(A) \to B_1 \times \cdots \times B_t$ for some $t \le s$. Let π_j be the projection onto the factor B_j. We have $0 = \pi_j \psi \phi(P) = \sum_i \pi_j \psi \phi_i(P_i)$, and by (iii) we deduce $\pi_j \psi \phi_i(P_i) = 0$ for every i, j. Then $\psi \phi_i(P_i) = 0$ for every i. The latter implies $\phi_1(P_1), \dots, \phi_r(P_r)$ are torsion as claimed since ψ is an isogeny. □

The following generalizes Proposition 1.5 to almost independent points:

Proposition 2.5. *Suppose $A_1, \dots, A_r$ are abelian varieties and $P_1 \in A_1(K), \dots, P_r \in A_r(K)$ are almost independent points. Let l be a prime number. If $e_1, \dots, e_r$ are nonnegative integers, then the following set has a positive Dirichlet density*:

$$\{\mathfrak{p} \in S(A_1, \dots, A_r) : \operatorname{ord}_l(P_i \bmod \mathfrak{p}) = e_i \textit{ for all } i\}.$$

Proof. Let $B_i \subseteq A_i$ be the abelian subvariety $\overline{\mathbb{Z}P}_i$. The point $P = (P_1, \dots, P_r)$ of $B := B_1 \times \cdots \times B_r$ satisfies $\overline{\mathbb{Z}P} = B$. The statement follows from Proposition 1.5. □

Domination of subgroups. Suppose $A, A_1, \dots, A_r$ are abelian varieties and let $\Gamma \subseteq A(K)$ be a subgroup.

Definition 2.6. Given subsets $M_i \subseteq A_i(K)$ for $1 \le i \le r$, we say $M_1, \dots, M_r$ *dominate* Γ if and only if the submodule $\Gamma' \subseteq A(K)$, which is generated by $\mathrm{Hom}(A_1, A)M_1, \dots, \mathrm{Hom}(A_r, A)M_r$, is such that $\Gamma \cap \Gamma'$ has finite index in Γ.

We understand that an empty set of points dominates any finite subgroup.

Lemma 2.7. *If A is an abelian variety and $\Gamma, \Gamma' \subseteq A(K)$ are submodules, then the following are equivalent:*

(i) $\Gamma \cap \Gamma'$ *has finite index in* Γ.

(ii) $\Gamma \cap \Gamma' \cap B(K)$ *has finite index in* $\Gamma \cap B(K)$ *for every simple abelian subvariety* $B \subseteq A$.

Proof. The implication (i) $\Rightarrow$ (ii) is an easy remark about abelian groups, so we only have to prove the converse. If A is simple, then (i) and (ii) are trivially equivalent, so suppose $A_1, A_2 \subseteq A$ are nontrivial complementary abelian subvarieties.

If C is an abelian variety and $\varphi : A \to C$ is an isogeny, then $\varphi(\Gamma), \varphi(\Gamma')$ have finite index in the respective submodules $\Gamma_0, \Gamma'_0 \subseteq C(K)$ they generate, thus $[\Gamma : \Gamma \cap \Gamma']$ is finite if and only if $[\Gamma_0 : \Gamma_0 \cap \Gamma'_0]$ is finite. Moreover, if $B \subseteq A$ is a simple abelian subvariety, then $[\Gamma \cap B(K) : \Gamma \cap \Gamma' \cap B(K)]$ is finite if and only if $[\Gamma_0 \cap \varphi(B)(K) : \Gamma_0 \cap \Gamma'_0 \cap \varphi(B)(K)]$ is finite. We may then suppose without loss of generality that $A = A_1 \times A_2$, so that $\Gamma = \Gamma_1 \times \Gamma_2$ and $\Gamma' = \Gamma'_1 \times \Gamma'_2$, where Γ_i, Γ'_i are submodules of $A_i(K)$ for $i = 1, 2$. By induction on $\dim(A)$, we suppose that (i) and (ii) are equivalent for $A = A_1, A_2$. If (ii) holds for Γ, Γ', A, it also holds for Γ_i, Γ'_i, A_i, where $i = 1, 2$. We deduce that $[\Gamma_i : \Gamma_i \cap \Gamma'_i]$ is finite for $i = 1, 2$, hence $[\Gamma : \Gamma \cap \Gamma']$ is finite. Thus, (ii) implies (i) as claimed. □

Proposition 2.8. *Suppose $A, A_1, \dots, A_r$ are abelian varieties and $\Gamma \subseteq A(K)$ is a submodule. If $P_1 \in A_1(K), \dots, P_r \in A_r(K)$ are almost independent, then either they dominate Γ, or there exists a simple abelian subvariety $A_{r+1} \subseteq A$ and a point $P_{r+1} \in \Gamma \cap A_{r+1}(K)$ such that $P_1, \dots, P_{r+1}$ are almost independent.*

Proof. Let $\Gamma' \subseteq A(K)$ be the submodule generated by $\mathrm{Hom}(A_i, A)P_i$ for $1 \le i \le r$. Suppose $P_1, \dots, P_r$ do not dominate Γ and thus $\Gamma \cap \Gamma'$ has infinite index in Γ. Then Lemma 2.7 implies $\Gamma \cap \Gamma' \cap A_{r+1}(K)$ has infinite index in $\Gamma \cap A_{r+1}(K)$ for some simple abelian subvariety $A_{r+1} \subseteq A$. Let $\Gamma_0 = \Gamma \cap A_{r+1}(K)$ and $\Gamma'_0 = \Gamma' \cap A_{r+1}(K)$. The index of $\Gamma_0 \cap \Gamma'_0$ in Γ_0 is infinite. Then since Γ_0 is a finitely generated abelian group, there exists a point $P_{r+1} \in \Gamma_0$ of infinite order such that $\mathbb{Z}P_{r+1} \cap \Gamma'_0 = \{0\}$. We will show $P_1, \dots, P_{r+1}$ are almost independent.

Let $\Gamma''_0 \subseteq \Gamma_0$ be the $\mathrm{End}(A_{r+1})$-submodule generated by P_{r+1}. If $\varphi \in \mathrm{End}(A_{r+1})$ is nonzero, then there exist $\psi \in \mathrm{End}(A_{r+1})$ and $m \ge 1$ such that $\psi\varphi$ is multiplication by m. In particular, the identity $\mathbb{Z}P_{r+1} \cap \Gamma_0' = \{0\}$ implies $\Gamma_0'' \cap \Gamma_0' = \{0\}$. Since P_{r+1} has infinite order and A_{r+1} is simple, we have that $\overline{\mathbb{Z}P}_{r+1} = A_{r+1}$ is nontrivial and connected.

Suppose $B_1, \ldots, B_s$ are simple abelian varieties such that $A_1 \times \cdots \times A_{r+1}$ is isogenous to $B_1 \times \cdots \times B_s$, and either $B_i = B_j$ or B_i, B_j are nonisogenous if $i \neq j$. We may suppose $B_s = A_{r+1}$. Let $B \in \{B_1, \ldots, B_s\}$ and $(\phi_1, \ldots, \phi_{r+1}) \in \operatorname{Hom}(A_1 \times \cdots \times A_{r+1}, B)$ satisfy $\sum_{i=1}^{r+1} \phi_i(P_i) = 0$. Let $Q = \sum_{i=1}^{r} \phi_i(P_i)$. If $B \neq A_{r+1}$, then $\operatorname{Hom}(A_{r+1}, B) = \{0\}$, hence $\phi_{r+1}(P_{r+1}) = 0$. If $B = A_{r+1} \subseteq A$, we have $\phi_{r+1}(P_{r+1}) = -Q \in \Gamma_0'$, hence $\phi_{r+1}(P_{r+1})$ lies in $\Gamma_0' \cap \Gamma_0'' = \{0\}$.

Either way, $\phi_{r+1}(P_{r+1}) = 0$, hence Lemma 2.4 implies $\phi_i(P_i) = 0$ for $1 \leq i \leq r$ since $P_1, \ldots, P_r$ are almost independent. In particular, Lemma 2.4 also implies $P_1, \ldots, P_{r+1}$ are almost independent as claimed. $\square$

Any infinite submodule $\Gamma \subseteq A(K)$ contains an almost independent point (it suffices to take any point P in Γ of infinite order and should $\overline{\mathbb{Z}P}$ not be connected, replacing P by a suitable multiple). One can then use the following corollary to find finitely many points of Γ which are almost independent and dominate Γ:

Corollary 2.9. *Suppose $A, A_1, \ldots, A_r$ are abelian varieties and let $\Gamma \subseteq A(K)$ be a submodule. If $P_1 \in A_1(K), \ldots, P_r \in A_r(K)$ are almost independent, then either they dominate Γ, or there exist $s > r$ and points $P_{r+1}, \ldots, P_s \in \Gamma$ such that $P_1, \ldots, P_s$ are almost independent and dominate Γ.*

Proof. Repeated application of Proposition 2.8 yields a sequence $P_1, \ldots, P_r, P_{r+1}, \ldots$ of almost independent points and a strictly increasing sequence of subgroups of Γ which are dominated by those points. This process must terminate after finitely many iterations because Γ is a finitely generated abelian group, and when it does, by Proposition 2.8, the given points dominate Γ. $\square$

3. Proof of the theorems

Order of reductions of submodules.

Theorem 3.1. *Let A, B be abelian varieties, and suppose that no element of* $\operatorname{Hom}(A, B)$ *has finite kernel. Let $\Gamma \subseteq A(K)$ be a dense submodule and l be a prime number.*

(i) *For every $e \geq 0$ the following set has positive Dirichlet density*:

$$O(A, B, \Gamma)_e := \{\mathfrak{p} \in S(A, B) : \operatorname{ord}_l(\Gamma_\mathfrak{p}) \geq \operatorname{ord}_l(B(K)_\mathfrak{p}) + e\}.$$

(ii) *If A is square-free, then for every $e \geq 0$ the following set has positive Dirichlet density*:

$$E(A, B, \Gamma)_e := \{\mathfrak{p} \in S(A, B) : \exp_l(\Gamma_\mathfrak{p}) \geq \exp_l(B(K)_\mathfrak{p}) + e\}.$$

(iii) *If A is square-free and if l is larger than a constant depending only on A, B, Γ, K, then the following set has positive Dirichlet density*:

$$R(A, B, \Gamma) := \{\mathfrak{p} \in S(A, B) : \operatorname{rad}_l(\Gamma_\mathfrak{p}) = 1, \operatorname{rad}_l(B(K)_\mathfrak{p}) = 0\}.$$

Proof. By our hypothesis on the elements of $\operatorname{Hom}(A, B)$, there exists a simple abelian variety C which occurs with multiplicity a as an isogeny factor of A, and with strictly smaller multiplicity b as an isogeny factor of B. Let $A' \subseteq A$ and $B' \subseteq B$ be abelian subvarieties for which there exist isogenies $\varphi : A \to A' \times C^a$ and $\psi : C^b \times B' \to B$. There is $d_1 > 0$ satisfying

$$\left|\operatorname{ord}_l(B(K)_\mathfrak{p}) - \operatorname{ord}_l((B' \times C^b)(K)_\mathfrak{p})\right| \leq v_l(d_1) \quad \text{for every } \mathfrak{p} \in S(A, B).$$

Moreover, if $\varphi_*(\Gamma) \subseteq (A' \times C^a)(K)$ is the submodule generated by $\varphi(\Gamma)$, then there exists $d_2 > 0$ satisfying

$$|\operatorname{ord}_l(\Gamma_\mathfrak{p}) - \operatorname{ord}_l(\varphi_*(\Gamma)_\mathfrak{p})| \leq v_l(d_2)$$

because $\varphi(\Gamma)$ has finite index in $\varphi_*(\Gamma)$; therefore,

$$O(A' \times C^a, B' \times C^b, \varphi_*(\Gamma))_{v_l(d_1 d_2)+e} \subseteq O(A, B, \Gamma)_e.$$

Similarly we have $d_3 > 0$ such that

$$E(A' \times C^a, B' \times C^b, \varphi_*(\Gamma))_{v_l(d_3)+e} \subseteq E(A, B, \Gamma)_e,$$

and such that for $l \nmid d_3$ we have:

$$R(A' \times C^a, B' \times C^b, \varphi_*(\Gamma)) \subseteq R(A, B, \Gamma).$$

Up to replacing A, B, Γ by $A' \times C^a, B' \times C^b, \varphi_*(\Gamma)$, we may suppose without loss of generality that φ, ψ are the respective identity maps.

Lemma 2.1 implies $\operatorname{Hom}(A, C)\Gamma$ is infinite since Γ is dense. It follows that $\Gamma \cap \operatorname{Hom}(C, A)(C(K))$ is infinite; thus it contains a point P which is almost independent. Let $\Gamma_0 \subseteq \Gamma$ and $\Gamma'_0 \subseteq B(K)$ be the respective submodules generated by P and $\operatorname{Hom}(A, B)P$. They are respectively isomorphic to a and b copies of the submodule of $C(K)$ generated by P; thus

$$\operatorname{ord}_l(\Gamma_\mathfrak{p}) \geq \operatorname{ord}_l(\Gamma_{0,\mathfrak{p}}) \geq \operatorname{ord}_l(\Gamma'_{0,\mathfrak{p}}) + (a-b)\operatorname{ord}_l(P \bmod \mathfrak{p}) \quad \text{for } \mathfrak{p} \in S(A, B).$$

Moreover, if A is square-free, then $a = 1$ and $b = 0$, so in particular, $\operatorname{Hom}(C, B) = \{0\}$ and $\Gamma'_0 = \{0\}$.

Corollary 2.9 implies there exist $t \geq 0$ points $Q_i \in B(K)$ such that together with P they are almost independent and dominate $B(K)$. Moreover, for every $m \geq 0$, the set

$$S_m := \{\mathfrak{p} \in S(A, B) : \operatorname{ord}_l(P \bmod \mathfrak{p}) = m, \operatorname{ord}_l(Q_i \bmod \mathfrak{p}) = 0 \text{ for all } i\}$$

has a positive density by Proposition 2.5. Thus it suffices to show that each of $O(A, B, \Gamma)_e$, $E(A, B, \Gamma)_e$, $R(A, B, \Gamma)$ contains S_m for some $m \geq 1$.

Let $\Gamma'_1 \subseteq B(K)$ be the submodule generated by the points Q_i. If $\Gamma' = \Gamma'_0 + \Gamma'_1$, then the index d of Γ' in $B(K)$ is finite; therefore, if $\mathfrak{p} \in S_m$, then

$$\operatorname{ord}_l(B(K)_\mathfrak{p}) \leq v_l(d) + \operatorname{ord}_l(\Gamma'_{0,\mathfrak{p}}) + \operatorname{ord}_l(\Gamma'_{1,\mathfrak{p}})$$

and

$$\operatorname{ord}_l(\Gamma_\mathfrak{p}) - \operatorname{ord}_l(B(K)_\mathfrak{p}) \geq (a-b)\cdot m - v_l(d) \geq m - v_l(d).$$

In particular, if $m \geq e + v_l(d)$, then $S_m \subseteq O(A, B, \Gamma)_e$, hence (i) holds. If $\mathfrak{p} \in S_m$ and if $\Gamma'_0 = \{0\}$ (for example, if A is square-free), then

$$\exp_l(B(K)_\mathfrak{p}) \leq v_l(d) + \exp_l(\Gamma'_{1,\mathfrak{p}}) = v_l(d),$$

while

$$\exp_l(\Gamma_\mathfrak{p}) \geq \operatorname{ord}_l(P \mod \mathfrak{p}) = m \geq \exp_l(B(K)_\mathfrak{p}) + (m - v_l(d)).$$

Thus, if A is square-free and $m \geq e + v_l(d)$, then $S_m \subseteq E(A, B, \Gamma)_e$, hence (ii) holds. If $l \nmid d$, then $S_1 \subseteq R(A, B, \Gamma)$, hence (iii) holds. $\square$

Exponents of reductions of submodules.

Theorem 3.2. *Let $A_1, \dots, A_r$ be abelian varieties, and $\Gamma_i \subseteq A_i(K)$ for $1 \leq i \leq r$ be submodules. Suppose that $e_r \geq \cdots \geq e_1 \geq 0$, and that for $1 \leq i < r$ we have $e_{i+1} = 0$ whenever $\Gamma_1, \dots, \Gamma_i$ dominate Γ_{i+1}. Then there exists $d \geq 1$ (depending only on $A_1, \dots, A_r, K, \Gamma_1, \dots, \Gamma_r$) such that the following set has positive Dirichlet density for every prime number l:*

$$E_{l,d} = \{\mathfrak{p} \in S(A_1, \dots, A_r) : e_i \leq \exp_l(\Gamma_i \mod \mathfrak{p}) \leq e_i + v_l(d) \text{ for all } i\}.$$

Proof. For $i = 1, \dots, r$ we apply Corollary 2.9 and choose $M_i \subseteq \Gamma_i$ such that the elements of $\mathcal{B}_i = M_1 \cup \cdots \cup M_i$ are almost independent and dominate Γ_i. Let $\Gamma'_i \subseteq A_i(K)$ be the submodule generated by $\operatorname{Hom}(A_1, A_i)M_1, \dots, \operatorname{Hom}(A_i, A_i)M_i$ so that $d_i = [\Gamma_i : \Gamma_i \cap \Gamma'_i]$ is finite.

If $\mathcal{B}_i = \mathcal{B}_{i-1}$, then $\Gamma_1, \dots, \Gamma_{i-1}$ dominate Γ_i, hence $e_i = 0$ by hypothesis. In particular, if we define $\exp_l(M \mod \mathfrak{p}) = \max_{P \in M \cup \{0\}} \exp_l(P \mod \mathfrak{p})$ for a finite set M, then Proposition 2.5 implies the following set has positive density for every l:

$$S_l = \{\mathfrak{p} \in S(A_1, \dots, A_r) : \exp_l(M_i \mod \mathfrak{p}) = e_i \text{ for all } i\}.$$

We claim S_l is contained in $E_{l,d}$ for $d = d_1 \dots d_r$, and thus the latter has positive density.

If $\mathfrak{p} \in S(A_1, \dots, A_r)$, then

$$\exp_l(\Gamma_i \cap \Gamma'_i \mod \mathfrak{p}) \leq \exp_l(\Gamma_i \mod \mathfrak{p}) \leq \exp_l(\Gamma_i \cap \Gamma'_i \mod \mathfrak{p}) + v_l(d).$$

If, moreover, $\mathfrak{p} \in S_l$, then

$$e_i = \exp_l(M_i \mod \mathfrak{p}) \le \exp_l(\Gamma_i \cap \Gamma_i' \mod \mathfrak{p}) \le \max\{e_1, \dots, e_i\} = e_i$$

because either $M_i = \varnothing$ and $e_i = 0$, or $\varnothing \ne M_i \subseteq \Gamma_i \cap \Gamma_i'$; therefore, $S_l \subseteq E_{l,d}$ as claimed. □

In particular, we may apply this theorem as soon as no Γ_i is dominated by the other submodules. We deduce that this is a necessary and sufficient condition for the set

$$\{\mathfrak{p} \in S(A_1, \dots, A_r) : \exp_l(\Gamma_i \mod \mathfrak{p}) = e_i \text{ for all } i\}$$

to have a positive Dirichlet density for all but finitely many prime numbers l, and for every $e_1, \dots, e_r \ge 0$.

Corollary 3.3. *Suppose $A_1, \dots, A_r$ are abelian varieties, and $\Gamma_i \subseteq A_i(K)$ for $1 \le i \le r$ are submodules. If $\Gamma_1, \dots, \Gamma_i$ do not dominate Γ_{i+1} for $1 \le i < r$, then for every $m \ge 0$ and for every prime number l, the following set has positive Dirichlet density*:

$$O_{l,m} = \{\mathfrak{p} \in S(A_1, \dots, A_r) : \mathrm{ord}_l(\Gamma_{i+1} \mod \mathfrak{p}) > \mathrm{ord}_l(\Gamma_i \mod \mathfrak{p}) + m \text{ for } 1 \le i < r\}.$$

Proof. Let $d \ge 1$ and $E_{l,d}$ be as in Theorem 3.2. Choose $e_1 \ge 0$ and $e_{i+1} > \dim(A_i)(e_i + v_l(d)) + m$ for $1 \le i < r$. Then $E_{l,d}$ has positive density by Theorem 3.2, and it lies in $O_{l,m}$ since

$$\exp_l(\Gamma_i \mod \mathfrak{p}) \le \mathrm{ord}_l(\Gamma_i \mod \mathfrak{p}) \le \dim(A_i(K)) \cdot \exp_l(\Gamma_i \mod \mathfrak{p})$$

holds for $\mathfrak{p} \in S(A_i)$. □

Proof of the main theorems.

Proof of Theorem 1.3. Suppose that property (1) fails. We show that (2), (3), and (4) fail accordingly.

If there is no homomorphism $A \to A'$ with finite kernel, then Theorem 3.1 (i) shows that (2) fails for every l and m, that (3) fails for every l and m if A is square-free, and that (4) fails if A is square-free and l is greater than a constant depending only on A, A', Γ, K.

Suppose now that there is $\varphi \in \mathrm{Hom}(A, A')$ with finite kernel. Since (1) fails, then $\varphi(\Gamma) \cap \Gamma'$ has infinite index in $\varphi(\Gamma)$, which means that Γ' does not dominate $\varphi(\Gamma)$. Consequently, Γ' does not dominate Γ. Let $A_1 = A'$, $A_2 = A$, $\Gamma_1 = \Gamma'$, and $\Gamma_2 = \Gamma$. Corollary 3.3 implies (2) fails for every l and m. Theorem 3.2 (applied with $e_1 = 0$ and $e_2 > v_l(d) + m$) implies (3) fails for every l and m, moreover (applied with $e_1 = 0$ and $e_2 = 1$), it implies (4) fails for l greater than a constant depending only on A, A', K, Γ, Γ'. □

Proof of Theorem 1.2. Applying Theorem 1.3 by taking $\Gamma = A(K)$, $\Gamma' = A'(K)$, we find in particular that A is isogenous to an abelian subvariety of A'. Moreover, by reversing the roles of Γ, Γ' we analogously find that A' is isogenous to an abelian subvariety of A, so we deduce that A, A' are isogenous. □

Proof of Theorem 1.1. This is an immediate consequence of Theorem 1.2. □

We conclude with a converse to Theorem 1.3:

Lemma 3.4. *With the notations of Theorem 1.3, property (1) implies that, for some integer $d > 0$, we have*

$$\mathrm{ord}_l(\Gamma_{\mathfrak{p}}) \le \mathrm{ord}_l(\Gamma'_{\mathfrak{p}}), \quad \exp_l(\Gamma_{\mathfrak{p}}) \le \exp_l(\Gamma'_{\mathfrak{p}}), \quad \mathrm{rad}_l(\Gamma_{\mathfrak{p}}) \le \mathrm{rad}_l(\Gamma'_{\mathfrak{p}})$$

for every $\mathfrak{p} \in S(A, A')$ and for every prime number $l \nmid d$.

Proof. Let $\varphi \in \mathrm{Hom}(A, A')$ be as in (1). Let k be the size of the kernel of φ, and c the index of $\varphi(\Gamma) \cap \Gamma'$ in $\varphi(\Gamma)$. If $\mathfrak{p} \in S(A, A')$ and letting $d = kc$, we have

$$\mathrm{ord}_l(\Gamma_{\mathfrak{p}}) \le v_l(k) + v_l(c) + \mathrm{ord}_l((\varphi(\Gamma) \cap \Gamma')_{\mathfrak{p}}) \le v_l(d) + \mathrm{ord}_l(\Gamma'_{\mathfrak{p}}).$$

Similarly, we have

$$\exp_l(\Gamma_{\mathfrak{p}}) \le v_l(d) + \exp_l(\Gamma'_{\mathfrak{p}}), \quad \mathrm{rad}_l(\Gamma_{\mathfrak{p}}) \le v_l(d) + \mathrm{rad}_l(\Gamma'_{\mathfrak{p}}).$$ □

Acknowledgements

Both authors would like to thank the anonymous referee, whose generous remarks helped to improve the exposition greatly.

References

[Demeyer and Perucca 2013] J. Demeyer and A. Perucca, "The constant of the support problem for abelian varieties", *J. Number Theory* **133**:9 (2013), 2843–2856. MR 3057049

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. Translated as "Finiteness theorems for abelian varieties over number fields", pp. 9–27 in *Arithmetic geometry* (Storrs, CT, 1984), edited by G. Cornell and J. H. Silverman, Springer, New York, 1986. MR 85g:11026a Zbl 0588.14026

[Jones and Rouse 2010] R. Jones and J. Rouse, "Galois theory of iterated endomorphisms", *Proc. Lond. Math. Soc.* (3) **100**:3 (2010), 763–794. MR 2011d:11120 Zbl 1244.11057

[Khare and Prasad 2004] C. Khare and D. Prasad, "Reduction of homomorphisms mod p and algebraicity", *J. Number Theory* **105**:2 (2004), 322–332. MR 2005f:11118 Zbl 1080.11046

[Mumford 1970] D. Mumford, *Abelian varieties*, Studies in Mathematics **5**, Tata Institute of Fundamental Research, Bombay, 1970. Second edition, 1974; reprinted 2008 by the Hindustan Book Agency, Delhi, and distributed by the AMS. The reprint has different pagination but the original page numbers are visible in the margins. MR 44 #219 Zbl 0223.14022

[Perucca 2009] A. Perucca, "Prescribing valuations of the order of a point in the reductions of abelian varieties and tori", *J. Number Theory* **129**:2 (2009), 469–476. MR 2009h:14073 Zbl 1166.14028

[Perucca 2011] A. Perucca, "On the reduction of points on abelian varieties and tori", *Int. Math. Res. Not.* **2011**:2 (2011), 293–308. MR 2012h:14115 Zbl 05856665

[Pink 2004] R. Pink, "On the order of the reduction of a point on an abelian variety", *Math. Ann.* **330**:2 (2004), 275–291. MR 2005g:11101 Zbl 1077.11046

[Zarhin 2008] Y. G. Zarhin, "Homomorphisms of abelian varieties over finite fields", pp. 315–343 in *Higher-dimensional geometry over finite fields* (Göttingen, 2007), edited by D. Kaledin and Y. Tschinkel, NATO Sci. Peace Secur. Ser. D Inf. Commun. Secur. **16**, IOS, Amsterdam, 2008. MR 2010a:11117 Zbl 1183.14035 arXiv 0711.1615

Received November 22, 2011. Revised May 6, 2013.

CHRIS HALL
MATHEMATICS
UNIVERSITY OF WYOMING
ROSS HALL
LARAMIE, WY 82071
UNITED STATES
chall14@uwyo.edu

ANTONELLA PERUCCA
FACULTAT MATHEMATIK
UNIVERSITY OF REGENSBURG
UNIVERSITSTRASSE 31
D-93040 REGENSBURG
GERMANY
antonella.perucca@mathematik.uni-regensburg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.441

VARIATION OF COMPLEX STRUCTURES AND THE STABILITY OF KÄHLER–RICCI SOLITONS

STUART J. HALL AND THOMAS MURPHY

We investigate the linear stability of Kähler–Ricci solitons for perturbations induced by varying the complex structure within a fixed Kähler class. We calculate stability for the known examples of Kähler–Ricci solitons.

0. Introduction

We consider a stability problem for shrinking Kähler–Ricci solitons. These are critical points of the ν-functional, defined by Perelman on the space of Riemannian metrics on a closed manifold M. The main result is a formula for the second variation of this functional when restricted to perturbations obtained by varying the complex structure within a fixed Kähler class. Such perturbations were first studied by Tian and Zhu [2008] for Kähler–Einstein manifolds, and our paper attempts to extend their results to Kähler–Ricci solitons. Definitions and notation from the main theorem are explained below.

Theorem 0.1 (Main Theorem). *Let (M, g, f) be a normalised Kähler–Ricci soliton and let h be an f-essential variation. The second variation of the ν-functional at g, $\langle Nh, h\rangle_f$, is given as*

$$\langle Nh, h\rangle_f = 2\int_M f\|h\|^2 e^{-f}\, dV_g.$$

The main utility of this result is that if one had explicit knowledge of the metric and the function f then it is possible to calculate the quantity $\langle Nh, h\rangle_f$ quite easily. In Section 4, we do this for all the known examples of Kähler–Ricci solitons. Notice also that for Kähler–Einstein metrics $f = 0$ and so $N(h) = 0$, recovering a result of Tian and Zhu.

The structure of this paper is as follows: In Section 1, we begin with background on Ricci solitons and the stability problem. In Section 2, the space $\mathcal{W}(g)$ and the space of f-essential variations in the above theorem are studied. We obtain several

Murphy is supported by an A.R.C. grant. We acknowledge the support of a Dennison research grant from the University of Buckingham, which funded a research visit by Murphy.
MSC2010: primary 53C44; secondary 53C25.
Keywords: Kähler–Ricci soliton, Perelman functional, linear stability.

useful characterizations of elements of these spaces. In Section 3, we give a proof of the main theorem. In Section 4, the stability of the known examples of Ricci solitons is investigated.

After a preliminary version of this work was posted on the arXiv, Yuanqi Wang kindly made us aware that he had independently obtained our Main Theorem as part of his Ph.D. thesis [Wang 2011]. His proof is similar to ours but proceeds by direct calculation rather than using the results of Dai, Wang and Wei. His thesis also contains interesting results about convergence of the Kähler–Ricci flow to a Kähler–Einstein metric when the complex structure is allowed to vary.

1. Ricci solitons and stability

Background on solitons. Throughout this paper, (M, g) is a smooth closed Riemannian manifold.

Definition 1.1 (Ricci soliton). Let $X \in \Gamma(TM)$ be a smooth vector field. The triple (M, g, X) is called a Ricci soliton if it satisfies the equation

$$\mathrm{Ric}(g) + L_X g = cg \tag{1-1}$$

for a constant $c \in \mathbb{R}$. If $c < 0, c = 0, c > 0$ then the soliton is referred to as expanding, steady and shrinking respectively. When $c \neq 0$, set $c = 1/2\tau$. If $X = \nabla f$ for a smooth function f then the soliton is called a gradient Ricci soliton and (1-1) becomes

$$\mathrm{Ric}(g) + \mathrm{Hess}(f) = \frac{1}{2\tau} g. \tag{1-2}$$

When the vector field X is Killing, an Einstein metric is recovered; Einstein metrics are therefore referred to as trivial Ricci solitons. We can set $c = 1$ to factor out homothety, and as one may change the soliton potential f by a constant, let us also require that

$$\int_M f e^{-f} \, dV_g = 0.$$

A soliton with these choices will be referred to as a *normalised gradient Ricci soliton.*

As well as being interesting as generalisations of Einstein metrics, Ricci solitons also occur as the fixed points of the Ricci flow

$$\frac{\partial g}{\partial t} = -2 \, \mathrm{Ric}(g) \tag{1-3}$$

up to diffeomorphism. In this paper we will be considering nontrivial Ricci solitons on compact manifolds. Foundational results due to Hamilton [1995] and Perelman [2002] imply that expanding and steady Ricci solitons on compact manifolds must be trivial. Hence our focus is on shrinking Ricci solitons. Perelman also showed that such solitons are necessarily gradient Ricci solitons. We will henceforth refer to these metrics as nontrivial shrinkers.

Due to the work of many people [Cao 1996; Dancer and Wang 2011; Koiso 1990; Podestà and Spiro 2010; Wang and Zhu 2004] there are now many (infinitely many) examples of nontrivial shrinkers. One striking feature all known nonproduct examples share is that they are Kähler. This means that $\mathrm{Hess}(f)$ is J-invariant and so the real vector field ∇f is holomorphic (see [Besse 1987, 2.124]). In this case the underlying manifold M is in fact a smooth Fano variety.

Perelman [2002] showed that gradient Ricci solitons are the critical points of a functional, which is usually denoted by $\nu(g)$. Let $f \in C^\infty(M)$ and $\tau \in \mathbb{R}$. We say that (f, τ) is compatible if

$$\int_M e^{-f}(4\pi\tau)^{-n/2} = 1.$$

Definition 1.2. The ν-functional is given by

$$\nu(g) = \inf_{\text{compatible } (f,\tau)} \int_M [(R + |\nabla f|^2)\tau + f - n]e^{-f}(4\pi\tau)^{-n/2}\, dV_g,$$

where R is the scalar curvature of g.

As well as giving a variational characterization of Ricci solitons, Perelman showed that the functional is monotonically increasing under the Ricci flow. Hence, if one could perturb a soliton in a direction that increases ν and then continue the flow, one would not flow back to the soliton and the soliton would be regarded as unstable.

Linear stability. In order to determine the behaviour of the flow around a soliton one can investigate the second variation of $\nu(g)$ for an admissible perturbation.

Definition 1.3. Let $h \in s^2(T^*M)$. Then $g + th$, $t \in \mathbb{R}^+$ is said to be an admissible perturbation. We have $\partial g/\partial t|_{t=0} = h$.

If the second variation is strictly negative then the fixed point is stable and attracting. If the second variation has positive directions then one may perturb the soliton and then flow away. Natasha Sesum [2006] has obtained fundamental results on this topic.

Proposition 1.4 [Cao et al. 2004; Cao and Zhu 2012]. *Let $h \in s^2(TM^*)$ be an admissible variation of a Ricci soliton g. The second variation of ν is given by*

$$D_g^2\nu(h,h) = \frac{\tau}{(4\pi\tau)^{n/2}} \int_M \langle Nh, h\rangle e^{-f}\, dV_g,$$

where

$$Nh = \tfrac{1}{2}\Delta_f h + \mathrm{Rm}(h,\cdot) + \mathrm{div}^*\mathrm{div}_f h + \tfrac{1}{2}\mathrm{Hess}(v_h) - C(h,g)\mathrm{Ric}. \tag{1-4}$$

Here $\Delta_f(\cdot) = \Delta(\cdot) - \nabla_{\nabla f}(\cdot)$, $\mathrm{div}_f(\cdot) = \mathrm{div}(\cdot) - \iota_{\nabla f}$, v_h *is the solution of the equation*

$$\Delta_f v_h + \frac{v_h}{2\tau} = \mathrm{div}_f\mathrm{div}_f(h),$$

and

$$C(h,g) = \frac{\int_M \langle \mathrm{Ric}, h\rangle e^{-f}\,dV_g}{\int_M Re^{-f}\,dV_g}.$$

This operator allows us to define the concept of linear stability.

Definition 1.5. Let (M, g, f) be a Ricci soliton. The soliton is *linearly stable* if the operator N is nonpositive definite, and *unstable* otherwise.

We now focus upon Kähler–Ricci solitons. The first result regarding stability is the following:

Theorem 1.6 [Cao et al. 2004; Hall and Murphy 2011; Tian and Zhu 2008]. *Let* (M, g, f) *be a Kähler–Ricci soliton. If* $dim H^{(1,1)}(M) > 1$ *then* (M, g, f) *is unstable.*

Kähler–Ricci solitons can be viewed as fixed points of a flow related to the Ricci flow (1-3) called the Kähler–Ricci flow, which in the Fano case can be written as

$$\frac{\partial g}{\partial t} = -\mathrm{Ric}(g) + g, \quad g(0) = g_0. \tag{1-5}$$

One important point about this flow is that it preserves the Kähler class. A foundational result about this flow, due to [Cao 1985], is that it exists for all time. The convergence of it is an extremely subtle issue because the complex structure can jump in the limit at infinity. Hence the type of convergence one expects is rather weak. This is illustrated by the following example:

Theorem 1.7 [Tian and Zhu 2007]. *Let* M *be a compact manifold which admits a Kähler–Ricci soliton* (g_{KRS}, f). *Then any solution of* (1-5) *will converge to* g_{KRS} *in the sense of Cheeger–Gromov if the initial metric* g_0 *is invariant under the maximal compact subset of the automorphism group of* M.

The unstable perturbations in Theorem 1.6 do not preserve the canonical class. Therefore, from the point of view of the Kähler–Ricci flow it is natural to consider perturbations which fix the Kähler class but allow the complex structure of the manifold to vary. This was initiated by Tian and Zhu [2008].

Definition 1.8. Let (M, g_{KRS}) be a Kähler–Ricci soliton with complex structure J_{KRS}. The space of perturbations is defined as

$$\mathcal{W}(g_{KRS}) = \big\{h \in s^2(TM^*) \;\big|\; \text{there is a family of Kähler metrics } (g_t, J_t) \\ \text{with } \partial g_t/\partial t|_{t=0} = h, \, [g_t(J_t\,\cdot\,, \cdot\,)] = c_1(M, J_{KRS}), \\ (g_0, J_0) = (g_{KRS}, J_{KRS})\big\}.$$

The following result was our main motivation for considering this space of perturbations:

Theorem 1.9 [Tian and Zhu 2008]. *Let (M, g_{KE}) be a Kähler–Einstein metric and let $h \in \mathcal{W}(g_{KE})$. Then*

$$\langle N(h), h\rangle_f \le 0.$$

Tian and Zhu then conjectured that a similar result should be true for Ricci solitons. Our formula in Theorem 0.1 shows that this might not be true in general. The integral in the main theorem does not seem to have a sign in general. However, the examples we calculate in Section 4 do all have $\langle N(h), h\rangle_f = 0$; this seems be an artefact of their construction rather than a manifestation of some result in complex differential geometry.

We mention here the related study of stability by Dai, Wang, and Wei [Dai et al. 2007]. They prove that Kähler–Einstein metrics with negative scalar curvature are stable. There is also the recent work of Nefton Pali [2012] in this area. He considers a related functional known in the literature as the W-functional (here one is free to pick a volume form whereas in the definition of the ν-functional one is determined by the metric).

Notation and convention. We use the curvature convention that $\mathrm{Rm}(X, Y)Z = \nabla_Y\nabla_X Z - \nabla_X\nabla_Y Z + \nabla_{[X,Y]}Z$. The convention for divergence that we adopt is $\mathrm{div}(h) = tr_{12}(\nabla h)$. The rough Laplacian

$$\Delta h = \mathrm{div}(\nabla h) = -\nabla^*\nabla h$$

is then negative definite. Set

$$\langle\,\cdot\,,\cdot\,\rangle_f = \int_M \langle\,\cdot\,,\cdot\,\rangle e^{-f}\, dV_g$$

to be the twisted inner product on tensors at a Ricci soliton (M, g, f). We will denote pointwise inner products induced on tensor bundles by g with round brackets $(\,\cdot\,,\cdot\,)$. The adjoint of a differential operator (such as ∇) with respect to this inner product will be denoted with a subscript f (for example, div_f) throughout.

2. Background on variations of complex structure

Variations of complex structure. We recall that an almost complex structure on a manifold M is a section J of the endomorphism bundle $\mathrm{End}(TM)$ satisfying $J^2 = -\,\mathrm{id}$. For M to be a complex manifold we require that the complex structure is *integrable*. By the Newlander–Nirenberg theorem we may take integrable to mean that the Nijenhuis tensor $\mathcal{N}(J) = 0$. We will be concerned with infinitesimal variations of complex structure that are modelled on those coming from a one-parameter family of complex structures J_t. As we are only working at an infinitesimal level, we don't actually mind if our variations are induced by such a family.

Definition 2.1 (Infinitesimal variation of complex structure). Let (M, g, J) be a Kähler manifold. A tensor $\zeta \in \mathrm{End}(TM)$ is called an infinitesimal variation of complex structure if it satisfies the two equations

$$\zeta J + J\zeta = 0, \tag{2-1}$$

$$\dot{\mathcal{N}}(\zeta) = 0, \tag{2-2}$$

where $\dot{\mathcal{N}}(\zeta)$ is the infinitesimal variation in the Nijenhuis tensors $\mathcal{N}(J + t\zeta)$.

Equation (2-1) simply says that the J_t are almost complex structures, and (2-2) comes from requiring that they are integrable. In the above definition we are viewing ζ as a section of the bundle $\mathrm{End}(TM)$ which is defined for any manifold. Switching in the usual manner to the complex viewpoint, (2-1) can be thought of as saying that ζ is a section of the bundle $\Lambda^{(0,1)} \otimes TM^{(1,0)}$. We will variously view the variation as an element of the real bundle $\mathrm{End}(TM)$, a section of the bundle $\Lambda^{(0,1)} \otimes TM^{(1,0)}$, and, using the metric to lower indices, as a section of $TM^* \otimes TM^*$ and $\Lambda^{(0,1)} \otimes \Lambda^{(0,1)}$. We note that in complex coordinates Equations (2-1) and (2-2) become

$$\zeta_\alpha^\beta = 0 \quad \text{and} \quad \nabla_\alpha \zeta_{\beta\gamma} = \nabla_\beta \zeta_{\alpha\gamma}.$$

The bundle $\Lambda^{(0,1)} \otimes TM^{(0,1)}$ is an element of the Dolbeault complex

$$TM^{(1,0)} \xrightarrow{\bar\partial} \Lambda^{(0,1)} \otimes TM^{(1,0)} \xrightarrow{\bar\partial} \Lambda^{(0,2)} \otimes TM^{(1,0)} \xrightarrow{\bar\partial} \cdots,$$

where $\bar\partial$ is the usual d-bar operator associated to a holomorphic vector bundle over a complex manifold. Equation (2-2) is equivalent to requiring that $\bar\partial\zeta = 0$.

Analogous to [Tian and Zhu 2008] and following [Koiso 1983], we will decompose the space of infinitesimal variations into *trivial* variations and *f-essential* variations.

By analogy with the twisted inner product, set

$$\Delta_{\bar\partial, f} := \bar\partial\bar\partial_f^* + \bar\partial_f^*\bar\partial$$

to be the twisted $\bar\partial$-Laplacian.

Definition 2.2 (f-essential variation). Let ζ be an infinitesimal variation of the complex structure J. We say ζ is trivial if $\zeta = L_Z J$ for a smooth vector field $Z \in TM$. A variation ζ is said to be f-essential if

$$\int_M \langle \zeta, L_Z J\rangle e^{-f}\, dV_g = 0$$

for all $Z \in \Gamma(TM)$.

The following lemma gives a useful characterisation of f-essential variations:

Lemma 2.3 [Koiso 1983, Lemma 6.4]. *Let ζ be an f-essential variation and let $h(\cdot\,,\cdot) = \omega(\cdot\,,\zeta\,\cdot)$. If h is symmetric then*

(1) $\bar\partial_f^*\zeta = 0$, *and*

(2) $\operatorname{div}_f h = 0$.

In particular, an f-essential variation is $\Delta_{\bar\partial,f}$-harmonic.

Proof. (1) As ζ is f-essential,

$$\int_M \langle L_Z J, \zeta\rangle e^{-f} = 0$$

for all $Z \in \Gamma(TM)$. The Lie derivative of the complex stucture is related to the $\bar\partial$-operator by

$$\bar\partial.\, Z = -\tfrac{1}{2} J L_Z J(\cdot).$$

Hence, up to a constant, $\langle L_Z J, \zeta\rangle_f = \langle \bar\partial Z, \zeta\rangle_f$ and $\bar\partial_f^*\zeta = 0$, as claimed.

(2) We begin by noting that ζ being f-essential means that

$$\langle L_Z J, \zeta\rangle_f = \langle \omega(\cdot\,, L_Z J(\cdot)), h\rangle_f = 0.$$

Rewriting and using the Cartan formula we have

$$\omega(\cdot\,, L_Z J(\cdot)) = L_Z g(\cdot\,,\cdot) - L_Z\omega(\cdot\,,\cdot) = 2\operatorname{div}^* Z^\flat(\cdot\,,\cdot) - (d\circ\iota_Z\omega)(\cdot\,, J\cdot).$$

The result follows by noting that

$$\langle (d\circ\iota_Z\omega)(\cdot\,, J\cdot), h\rangle_f = -\langle (d\circ\iota_Z\omega)(\cdot\,,\cdot), h(\cdot\,, J\cdot)\rangle_f,$$

and that $h(\cdot\,, J\cdot)$ is symmetric. □

In the previous lemma we have assumed that h is symmetric. This is not strictly necessary by the following argument: If there existed an antisymmetric, $\Delta_{\bar\partial,f}$-harmonic section of $\Lambda^{(0,1)}\otimes TM^{(1,0)}$ then there would have to exist an antisymmetric $\Delta_{\bar\partial}$-harmonic section of $\Lambda^{(0,1)}\otimes TM^{(1,0)}$ as

$$\mathcal{H}^{p,q}(E) \equiv H^q(M, E\otimes\Lambda^{(p,0)}) \equiv \mathcal{H}_f^{p,q}(E)$$

for any holomorphic vector bundle E. The Dai–Wei–Wang Weitzenböck formula (Lemma 3.3) and Lemma 3.4 then imply that the associated (0, 2)-form is parallel. This would imply that $h^{0,2}(M) > 0$. One can then appeal to a classical result of Bochner to show that on a Fano manifold such a holomorphic form cannot exist (see [Besse 1987, 11.24]). Tian and Zhu [2008] give a straightforward proof of this fact in the case one is at a Kähler–Einstein metric.

Tian and Zhu decompose the space $\mathscr{W}(g)$ modulo the action of the diffeomorphism group. They show that

$$\mathscr{W}(g)/\mathscr{D}(M) = \mathscr{A}^{(1,1)} \oplus H^1(M, TM),$$

where $\mathscr{A}^{(1,1)}$ is the space of $\partial\bar{\partial}$-exact (1,1)-forms and $H^1(M, TM)$ is the usual cohomology for the holomorphic vector bundle TM. Tian and Zhu then show that for a general Kähler–Ricci soliton, $N|_{\mathscr{A}^{(1,1)}} \leq 0$ so that potentially destabilising elements of $\mathscr{W}$ actually lie in $H^1(M, TM)$ (they then show that N vanishes on this space when g is an Einstein metric). Hence we will only consider perturbations in $H^1(M, TM)$ and we will use the special representatives given by f-essential perturbations. Formally:

Proposition 2.4 [Tian and Zhu 2008]. *Let (M, g_{KRS}, J) be a Kähler–Ricci soliton. Then we have the following decomposition*:

$$\mathscr{W}(g_{KRS})/\mathscr{D}(M) \cong \mathscr{A}^{(1,1)}(M, J) \oplus H^1(M, TM),$$

where $\mathscr{D}(M)$ is the diffeomorphism group of M. The operator N is nonpositive when restricted to $\mathscr{A}^{(1,1)}(M, J)$.

3. Proof of Main Theorem

Consider an f-essential variation of the complex structure $h \in H^1(M, TM)$. Firstly, as h is J-anti-invariant it is apparent that $C(h, g) = 0$. Thus

$$\langle N(h), h\rangle_f = \langle \tfrac{1}{2}\Delta_f h + \mathrm{Rm}(h, \cdot), h\rangle_f .$$

In order to evaluate the above we will use a Weitzenböck formula. In order to explain the formula we will digress briefly into the spinorial construction used in [Dai et al. 2007]. This is a powerful generalisation of the techniques used by Koiso [1983].

As M is Fano it has a canonical spin^c structure and parallel spinor $\sigma_0 \in \Gamma(\mathscr{S}^c)$, where $\mathscr{S}^c \to M$ is the spin^c spinor bundle. This induces a map

$$\Phi : s^2(TM^*) \to \mathscr{S}^c \otimes TM^*,$$
$$\Phi(h) = h_{ij} e_i \cdot \sigma_0 \otimes e^j,$$

where $\{e_i\}$ is an orthonormal basis of TM and $e_i \cdot \sigma_0$ denotes Clifford multiplication in $\mathcal{S}^c$.

For $1 \le i \le m$, following [Dai et al. 2007], choose

$$X_i = \frac{e_i - \sqrt{-1}Je_i}{\sqrt{2}} \quad \text{and} \quad \bar{X}_i = \frac{e_i + \sqrt{-1}Je_i}{\sqrt{2}}.$$

Then $\{X_1, \dots, X_m\}$ is a local unitary frame for $T^{1,0}M$. Set $\{\theta^1, \dots, \theta^m\}$ to be its dual frame. Then

$$\Phi(h) = h(\bar{X}_i, \bar{X}_j)\bar{\theta^i} \otimes \bar{\theta^j}.$$

This can be identified with

$$\Psi(h) = h(\bar{X}_i, \bar{X}_j)\bar{\theta^i} \otimes X_j \in \textstyle\bigwedge^{0,1}(TM^{1,0}),$$

where $TM^{1,0}$ is the holomorphic tangent bundle.

Lemma 3.1 [Dai et al. 2005, Lemma 2.3]. *For $h, \tilde{h} \in s^2(TM^*)$,*

$$\operatorname{Re}(\Phi(h), \Phi(\tilde{h})) = (h, \tilde{h}).$$

We will also need the following, which is a result of the calculations on page 680 of [Dai et al. 2007]:

Lemma 3.2. *Let (M, g) be a Fano manifold with canonical spin^c spinor bundle $\mathcal{S}^c$ and Dirac operator D. Let Φ and Ψ be defined as above. Then*

$$D\Phi(h) = \sqrt{2}(\bar{\partial} - \bar{\partial}^*)\Psi(h).$$

The main result we need is the following Weitzenböck formula:

Lemma 3.3 [Dai et al. 2007, Lemma 2.3]. *Let $h \in s^2(TM^*)$ and let D be the Dirac operator. Then*

$$D^*D(\Phi(h)) = \Phi(\nabla^*\nabla h - 2\operatorname{Rm}(h, \cdot\,) + \operatorname{Ric} \circ h - h \circ i\rho), \tag{3-1}$$

where ρ is the Ricci form.

In order to deal with the Ricci curvature terms we use the following lemma, which is implicit in the proof of Theorem 2.5 in [Dai et al. 2007]:

Lemma 3.4. *Let h be a skew-hermitian section of $s^2(TM^*)$. Then*

$$(\operatorname{Ric} \circ h - h \circ i\rho, h) = 0.$$

Proof. This is a pointwise calculation. Choose normal coordinates at $p \in M$, $\{e_1, \dots, e_{2m}\}$, where $e_{m+i} = Je_i$ for $1 \le i \le m$. We can also choose this basis so that the Ricci tensor is diagonalised; that is, $\operatorname{Ric}(e_i, e_j) = c_i\delta_{ij}$, where $c_{m+i} = c_i$. We have

$$\mathrm{Re}(\Phi(\mathrm{Ric}\circ h),\,\Phi(h)) = \sum_{i,j=1}^{2m} c_i h_{ij}^2,$$

$$-\mathrm{Re}(\Phi(h\circ i\rho),\,\Phi(h)) = -2\sum_{j=1}^{m}\sum_{i=1}^{m} c_j (h_{(i+m)j}h_{i(j+m)} - h_{ij}h_{(i+m)(j+m)}).$$

If h is skew-Hermitian then

$$h_{ij} = -h_{(i+m)(j+m)} \quad \text{and} \quad h_{i(j+m)} = h_{(i+m)j}.$$

Hence

$$-\mathrm{Re}(\Phi(h\circ i\rho),\,\Phi(h)) = -2\sum_{i=1}^{m}\sum_{j=1}^{2m} c_j (h_{ij}^2) = -\sum_{i,j=1}^{2m} c_i h_{ij}^2,$$

and the result follows. □

The final lemma we need to prove the main result in this section is a technical lemma to deal with the extra term one obtains by using the rescaled volume form $e^{-f}\,dV_g$.

Lemma 3.5. *Let $A \in \Omega^1(M)$ be a one-form and $B \in \bigotimes^k TM^*$. Then*

(1) $\mathrm{div}(A\otimes B) = \mathrm{div}(A)\otimes B + \nabla_{A^\sharp}B$,

(2) $\mathrm{div}(df\otimes h) = (\Delta f)h + \nabla_{\nabla f}h$, *and*

(3) $-\langle \nabla_{\nabla f}h, h\rangle_f = \frac{1}{2}\int_M \Delta_f f\,\|h\|^2 e^{-f}\,dV_g$.

Proof. (1) We calculate using a normal, orthonormal basis $\{e_i\}$,

$$\mathrm{div}(A\otimes B) = \nabla_{e_i}(A\otimes B)(e_i,\cdot\,) = \mathrm{div}(A)\otimes B + \nabla_{A^\sharp}B.$$

(2) We use $A = df$, $B = h$ in (1).

(3) We note that

$$\langle \nabla_{\nabla f}h, h\rangle_f = \langle \iota_{\nabla f}\nabla h, h\rangle_f = \langle \nabla h, df\otimes h\rangle_f = -\langle h, \mathrm{div}_f(df\otimes h)\rangle_f.$$

Now using (2) we have

$$\begin{aligned}\langle \nabla_{\nabla f}h, h\rangle_f &= \int_M |\nabla f|^2\|h\|^2 e^{-f}\,dV_g - \langle h, \mathrm{div}(df\otimes h)\rangle_f \\ &= -\int_M (\Delta_f f)\|h\|^2 e^{-f}\,dV_g - \langle \nabla_{\nabla f}h, h\rangle_f.\end{aligned}$$ □

As is well known, the soliton potential function of a normalised gradient Ricci soliton solves the equation

$$\Delta_f f = -2f.$$

Proof of Main Theorem. Lemmas 3.1, 3.2, and 3.3 yield that pointwise

$$\begin{aligned}\left(\tfrac{1}{2}\Delta h+\mathrm{Rm}(h,\cdot),h\right)&=\mathrm{Re}\left(\Phi(\tfrac{1}{2}\Delta h+\mathrm{Rm}(h,\cdot)),\Phi(h)\right)\\&=\mathrm{Re}\left((D^*D\Phi(h),\Phi(h))\right)\\&=\mathrm{Re}\left(-2\Delta_{\bar\partial}\Psi(h),\Psi(h)\right).\end{aligned}$$

However, as h is f-essential then $\Psi(h)$ is orthogonal to the image of $\Delta_{\bar\partial}$ with respect to the global inner product. Hence

$$\int_M \left(\tfrac{1}{2}\Delta h+\mathrm{Rm}(h,\cdot),h\right)e^{-f}\,dV_g=0. \qquad \square$$

4. Examples and applications

Setup. As mentioned in the introduction, there are three main sources for concrete examples of Kähler–Ricci solitons: the Dancer–Wang, Podestà–Spiro, and the Wang–Zhu examples. The Wang–Zhu solitons exist on toric-Kähler manifolds and are nontrivial precisely when the Futaki invariant is nonzero. Unfortunately, this class of manifold does not admit any nontrivial deformations of complex structure:

Theorem 4.1 [Bien and Brion 1996, Theorem 3.2]. *Every Fano toric-Kähler manifold M has $H^1(M,TM)=0$.*

Similarly, one can see the Podestà–Spiro examples are rigid. The next class of examples to investigate are provided by the Dancer–Wang solitons. These solitons are generalisations of the soliton on $\mathbb{CP}^2\,\sharp\,\overline{\mathbb{CP}}^2$ constructed by Koiso [1990] and Cao [1996]. We begin by reviewing their construction.

Let (V_i,r_i,J_i), $1\le i\le r$ be Fano Kähler–Einstein manifolds with first Chern class $c_1(V_i,J_i)=p_ia_i$, where p_i are positive integers and $a_i\in H^2(V_i;\mathbb{Z})$ are indivisible classes. The Kähler–Einstein metrics r_i are normalised so that $\mathrm{Ric}(r_i)=p_ir_i$. For $q=(q_1,\dots,q_r)$ with $q_i\in\mathbb{Z}-\{0\}$, let P_q be the total space of the principal $U(1)$-bundle over $B:=V_1\times V_2\times\cdots\times V_r$ with Euler class $\sum_1^r q_i\pi_i^*a_i$, where

$$\pi_i: V_1\times\cdots\times V_r\to V_i$$

is the projection onto the i-th factor. Denote by M_0 the product $I\times P_q$ for the unit interval I. We denote by θ the principal $U(1)$-connection on P_q with curvature

$$\Omega:=\sum_{i=1}^r q_i\pi_i^*\eta_i,$$

where η_i is the Kähler form of r_i. There is a one-parameter family of metrics on P_q given by

$$g_t:=f^2(t)\theta\otimes\theta+\sum_{i=1}^r l_i^2(t)\pi_i^*r_i,$$

where f and l_i are smooth functions on I with prescribed boundary behaviour. Finally, consider the metric on M_0 given by

$$g = dt^2 + g_t,$$

with the correct boundary behaviour of f and the l_i. This metric then extends to a metric on a compactification of M_0, which we denote M.

The complex structure on this manifold can be described explicitly by lifting the complex structure on the base and requiring that $J(N) = -f(t)^{-1}Z$, where $N = \partial_t$ is normal to the hypersurfaces, and Z is the Killing vector that generates the isometric $U(1)$ action on P_q.

Deformations of Dancer–Wang solitons. The Ricci soliton equations in this setting reduce to a system of ODEs. We have the following existence theorem:

Theorem 4.2 [Dancer and Wang 2011, Theorem 4.30]. *Let M denote the compactification of M_0 as above. Then M admits a Kähler–Ricci soliton (M, g, u) which is Einstein if and only if the associated Futaki invariant vanishes.*

We refer to [Dancer and Wang 2011] for details of the constructions. If one chooses the components V_i to be homogeneous Kähler–Einstein manifolds then the resulting M is toric. However, by choosing the components V_i to be nonhomogeneous, Fano and Kähler–Einstein, and calculating the Futaki invariant, they give examples of nontoric Kähler–Ricci solitons. It is these that may admit complex deformations.

Suppose that V_i is a Fano, Kähler–Einstein manifold admitting deformations of its complex structure J_i. We consider an essential variation h_i in the Kähler metric r_i such that the Kähler form $\eta_i = r_i(J_i\,\cdot\,,\,\cdot\,)$ remains in the class $c_1(V_i, J_0)$. This induces a variation in the metric on the whole space given by

$$h = l_i^2(t)\pi^* h_i.$$

Clearly the same procedure works for any product of Kähler–Einstein manifolds with some (or all) of the factors admitting complex deformations. Here it is simply stated for one factor for simplicity. Let us state our final result:

Theorem 4.3. *For this perturbation h, one has $N(h) = 0$.*

Proof. It follows from the construction of h that the pointwise norm $\|h\|$ is independent of t. It also follows that if h_i is essential then h is u-essential. We see now that

$$\langle Nh, h\rangle = \int_M u\|h\|^2 e^{-u}\,dV_g = \|h\|^2_{L^2}(V_i)\int_I ue^{-u}\,dt = 0. \qquad \square$$

Remark 4.4. The significance of this result is that it verifies Tian–Zhu's conjecture for every obvious example of a complex deformation of the known Kähler–Ricci solitons. We do not know of any explicit deformations beyond these.

It is notable that for all f-essential perturbations h known to us, one has $N(h)=0$. Understanding if this is always the case would involve calculating $H^1(M, TM)$, which is not easy to calculate in general.

Acknowledgements

It is a pleasure to thank Joel Fine for useful conversations and Nefton Pali for helpful comments on the paper. We also thank the anonymous referees for corrections and suggestions.

References

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Bien and Brion 1996] F. Bien and M. Brion, "Automorphisms and local rigidity of regular varieties", *Compositio Math.* **104**:1 (1996), 1–26. MR 97h:14034 Zbl 0910.14004

[Cao 1985] H.-D. Cao, "Deformation of Kähler metrics to Kähler–Einstein metrics on compact Kähler manifolds", *Invent. Math.* **81**:2 (1985), 359–372. MR 87d:58051 Zbl 0574.53042

[Cao 1996] H.-D. Cao, "Existence of gradient Kähler–Ricci solitons", pp. 1–16 in *Elliptic and parabolic methods in geometry* (Minneapolis, MN, 1994), edited by B. Chow et al., A K Peters, Wellesley, MA, 1996. MR 98a:53058 Zbl 0868.58047 arXiv 1203.4794

[Cao and Zhu 2012] H.-D. Cao and M. Zhu, "On second variation of Perelman's Ricci shrinker entropy", *Math. Ann.* **353**:3 (2012), 747–763. MR 2923948 Zbl 1252.53074

[Cao et al. 2004] H.-D. Cao, R. S. Hamilton, and T. Ilmanen, "Gaussian densities and stability for some Ricci solitons", preprint, 2004. arXiv math/0404165

[Dai et al. 2005] X. Dai, X. Wang, and G. Wei, "On the stability of Riemannian manifold with parallel spinors", *Invent. Math.* **161**:1 (2005), 151–176. MR 2006h:53041 Zbl 1075.53042

[Dai et al. 2007] X. Dai, X. Wang, and G. Wei, "On the variational stability of Kähler–Einstein metrics", *Comm. Anal. Geom.* **15**:4 (2007), 669–693. MR 2009a:53078 Zbl 1149.53027

[Dancer and Wang 2011] A. S. Dancer and M. Y. Wang, "On Ricci solitons of cohomogeneity one", *Ann. Global Anal. Geom.* **39**:3 (2011), 259–292. MR 2012a:53124 Zbl 1215.53040

[Hall and Murphy 2011] S. J. Hall and T. Murphy, "On the linear stability of Kähler–Ricci solitons", *Proc. Amer. Math. Soc.* **139**:9 (2011), 3327–3337. MR 2012g:53135 Zbl 1231.53056

[Hamilton 1995] R. S. Hamilton, "The formation of singularities in the Ricci flow", pp. 7–136 in *Proceedings of the Conference on Geometry and Topology* (Cambridge, MA, 1993), edited by C.-C. Hsiung and S.-T. Yau, Surveys in Differential Geometry **2**, Int. Press, Cambridge, MA, 1995. MR 97e:53075 Zbl 0867.53030

[Koiso 1983] N. Koiso, "Einstein metrics and complex structures", *Invent. Math.* **73**:1 (1983), 71–106. MR 85d:58018 Zbl 0515.53040

[Koiso 1990] N. Koiso, "On rotationally symmetric Hamilton's equation for Kähler–Einstein metrics", pp. 327–337 in *Recent topics in differential and analytic geometry*, edited by T. Ochiai, Adv. Stud. Pure Math. **18**, Academic Press, Boston, 1990. MR 93d:53057 Zbl 0739.53052

[Pali 2012] N. Pali, "The total second variation of Perelman's $\mathcal{W}$-functional", preprint, 2012. arXiv 1201.0969

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv math/0211159

[Podestà and Spiro 2010] F. Podestà and A. Spiro, "Kähler–Ricci solitons on homogeneous toric bundles", *J. Reine Angew. Math.* **642** (2010), 109–127. MR 2011f:32049 Zbl 1210.14057

[Sesum 2006] N. Sesum, "Linear and dynamical stability of Ricci-flat metrics", *Duke Math. J.* **133**:1 (2006), 1–26. MR 2007c:53089 Zbl 1103.53040

[Tian and Zhu 2007] G. Tian and X. Zhu, "Convergence of Kähler–Ricci flow", *J. Amer. Math. Soc.* **20**:3 (2007), 675–699. MR 2007k:53107 Zbl 1185.53078

[Tian and Zhu 2008] G. Tian and X. Zhu, "Perelman's $\mathcal{W}$-functional and stability of Kähler–Ricci flow", preprint, 2008. arXiv 0801.3504

[Wang 2011] Y. Wang, *On Ricci solitons and Ricci flows*, Ph.D. thesis, University of Wisconsin, Madison, WI, 2011, Available at http://search.proquest.com/docview/910543286. MR 2982322

[Wang and Zhu 2004] X.-J. Wang and X. Zhu, "Kähler–Ricci solitons on toric manifolds with positive first Chern class", *Adv. Math.* **188**:1 (2004), 87–103. MR 2005d:53074 Zbl 1086.53067

Received July 6, 2012. Revised April 16, 2013.

STUART J. HALL
DEPARTMENT OF APPLIED COMPUTING
UNIVERSITY OF BUCKINGHAM
HUNTER STREET
BUCKINGHAM, MK18 1G
UNITED KINGDOM
stuart.hall@buckingham.ac.uk

THOMAS MURPHY
DÉPARTMENT DE MATHÉMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
CP 218, BOULEVARD DU TRIOMPHE
1050 BRUXELLES
BELGIUM
thomas.murphy@ulb.ac.be

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.455

ON CROSSED HOMOMORPHISMS OF THE VOLUME PRESERVING DIFFEOMORPHISM GROUPS

RYOJI KASAGAWA

We construct crossed homomorphisms of the groups of volume preserving diffeomorphisms of closed manifolds with nontrivial first cohomology groups and give their applications to the volume flux groups. Moreover, we see that they descend to crossed homomorphisms of their isotopy groups. In the two dimensional case, we show that their restrictions to Torelli groups are the first Johnson homomorphisms.

1. Introduction

In this paper we construct crossed homomorphisms J of the group $\mathcal{D}_{\text{vol}}$ of volume preserving diffeomorphisms of a closed oriented smooth manifold M of dimension n with volume form vol. Each J is related with a Pontryagin class p of degree k and takes values in $\mathcal{H} = \text{Hom}\big(\bigwedge^{n-4k} H^1(M;\mathbb{R}), H^{n-1}(M;\mathbb{R})\big)$. This construction is an analogy of that for the symplectomorphism groups of symplectic manifolds in [Kasagawa 2008]. In it, crossed homomorphisms are constructed from certain relations of Chern classes and the cohomology class of the symplectic form. In this volume case, we use relations such as $p(M) \cup a_1 \cup \cdots \cup a_{n-4k} = \kappa[\text{vol}]$, where $a_i \in H^1(M;\mathbb{R})$ and $\kappa \in \mathbb{R}$. But there are usually many such relations, and the domains of crossed homomorphisms constructed from such relations need to be restricted to certain subgroups, so we consider them all together. This is the reason why the target of J is the space of homomorphisms between cohomology groups as above. The crossed homomorphism J, which is a 1-cocycle in terms of group cohomology theory, depends on the choice of the ingredients used in the construction, but we can show that its cohomology class does only on the Pontryagin class p, not on the other ingredients. Some cohomology classes on groups of volume preserving diffeomorphisms were studied by McDuff [1983], but they are defined only on the identity component. A significant point of our construction is that J's and their cohomology classes are defined on the whole group of volume preserving diffeomorphisms.

MSC2010: primary 57R50, 58D05; secondary 22E65, 57R15.
Keywords: crossed homomorphisms, volume preserving diffeomorphisms, volume flux groups.

We calculate the derivative of J along smooth curves in $\mathcal{D}_{\mathrm{vol}}$. Its formula contains the derivative of the volume flux homomorphism of (M, vol) as a term, so we have an application of our crossed homomorphisms to the volume flux ones. As a corollary of it, we obtain some conditions for which the volume flux groups vanish, some of which have been obtained by Kędra, Kotschick, and Morita [2006]. They studied more properties of flux groups not only for the volume case, but also for the symplectic case and others. In the two dimensional case, Kotschick and Morita also studied cohomology classes of the symplectomorphism groups of surfaces related with the extensions of the flux homomorphisms in [Kotschick and Morita 2005; 2007]. Their work suggests some applications of the crossed homomorphisms J.

The derivative formula for J tells us that the image of the identity component of $\mathcal{D}_{\mathrm{vol}}$ under J is easily understood. It turns out that J descends to a crossed homomorphism $\mathcal{J}$, from $\pi_0\mathcal{D}_{\mathrm{vol}}$ to a quotient of $\mathcal{H}$. It can be considered a crossed homomorphism of the group $\pi_0\mathcal{D}$ of path components of the diffeomorphism group $\mathcal{D}$ of the oriented manifold M since it is isomorphic to $\pi_0\mathcal{D}_{\mathrm{vol}}$ by the induced homomorphism of the standard inclusion $\mathcal{D}_{\mathrm{vol}} \hookrightarrow \mathcal{D}$ by Moser [1965]. The first nontrivial example of $\mathcal{J}$ is the two dimensional case. Let M be a closed oriented surface Σ_g of genus $g \geqq 3$. The group $\pi_0\mathcal{D}$ is called the mapping class group of Σ_g. The standard action of it on $H_{\mathbb{Z}} = H_1(\Sigma_g; \mathbb{Z})$ gives the well-known representation $\pi_0\mathcal{D} \to \mathrm{Sp}(2g; \mathbb{Z})$, whose kernel $\mathcal{I}_g$ is called the Torelli group. Johnson [1980] defined a surjective homomorphism $\tau_g \colon \mathcal{I}_g \to \bigwedge^3 H_{\mathbb{Z}}/H_{\mathbb{Z}}$, which is called the first Johnson homomorphism. If we take $p = 1$ as a Pontryagin class of degree 0, the target of the crossed homomorphism $\mathcal{J}$ is $\mathcal{H}/\!\sim\, = \mathrm{Hom}\bigl(\bigwedge^2 H^1(\Sigma_g; \mathbb{R}), H^1(\Sigma_g; \mathbb{R})\bigr)/\!\sim$. Using the Poincaré duality, we have a natural injection $\bigwedge^3 H_{\mathbb{Z}} \to \mathcal{H}$. We can see that it induces an injective homomorphism from the target of τ_g to that of $\mathcal{J}$. Thus, we can compare τ_g with $\mathcal{J}$ on the Torelli group $\mathcal{I}_g$ and can show that $\mathcal{J}$ coincides with $-\tau_g/2$ on $\mathcal{I}_g$. In other words, $\mathcal{J}$ is an extension of $-\tau_g/2$ as a crossed homomorphism of the whole mapping class group with larger target. Thus, the $\mathcal{J}$'s in dimension greater than 2 are considered analogues of the first Johnson homomorphisms for higher dimensions. On the other hand, Morita [1993] had already extended τ_g to a crossed homomorphism of $\pi_0\mathcal{D}$ with target $\frac{1}{2}\bigwedge^3 H_{\mathbb{Z}}/H_{\mathbb{Z}}$. The target of Morita's crossed homomorphism is also contained in that of $\mathcal{J}$ in the same way as above. We do not know if they are essentially the same, but in this paper we don't pursue this problem.

We note that there are many ways of extending Johnson homomorphisms, such as [Day 2007; Morita 1993; Morita and Penner 2008; Perron 2004] and others, and also that some relations between first Johnson homomorphisms and flux homomorphisms in the two dimensional case were obtained [Day 2011]. So, if we consider high dimensional analogues of these works, our crossed homomorphisms give ingredients for them.

This paper is organized as follows: In Section 2, we define the crossed homomorphisms J and $\mathscr{J}$ and state the main results. In Sections 3 and 4, we show the existence of main ingredients in the construction of J and some lemmas related to them. In Section 5, we recall properties of Bott homomorphisms. In Section 6, we prove Theorem 2.3, which states that the crossed homomorphism J is well defined and defines a unique cohomology class. In Section 7, we calculate the derivative of J along a smooth path in $\mathscr{D}_{\mathrm{vol}}$ and show some results for the volume flux homomorphisms and groups. In Section 8, we see that J descends to the crossed homomorphism $\mathscr{J}$. In Section 9, we recall the first Johnson homomorphism τ_g of Σ_g. In Sections 10, 11, and 12, we explicitly describe ingredients needed to calculate $\mathscr{J}$ in the 2 dimensional case. In Sections 13 and 14 we compute $\mathscr{J}$, which shows that it coincides with $-\frac{1}{2}\tau_g$ on the Torelli group $\mathscr{I}_g$.

2. Definitions and main results

Let M be a closed oriented smooth manifold of dimension n with a volume form vol. Let $\mathscr{D} = \mathrm{Diff}_+(M)$ be the group of orientation preserving diffeomorphisms of M and $\mathscr{D}_{\mathrm{vol}} = \{\varphi \in \mathscr{D} \mid \varphi^* \mathrm{vol} = \mathrm{vol}\}$ its subgroup of volume preserving ones. Put $C_0^\infty(M) = \{g \in C^\infty(M) \mid \int_M g \,\mathrm{vol} = 0\}$, $\mathscr{D}_{\mathrm{vol}}$ then acts on it by $\varphi_* g = (\varphi^{-1})^* g$ for $\varphi \in \mathscr{D}_{\mathrm{vol}}$ and $g \in C_0^\infty(M)$. Put $\varphi_\sharp h(a) = \varphi_*(h(\varphi^* a))$ for $a \in H^1(M)$ and $h \in \mathrm{Hom}(H^1(M), C_0^\infty(M))$, $\mathscr{D}_{\mathrm{vol}}$ then acts on $\mathrm{Hom}(H^1(M), C_0^\infty(M))$ by $(\varphi, h) \mapsto \varphi_\sharp h$. Here, $H^1(M)$ is the first cohomology group of M with real coefficients, and hereafter we use cohomology groups with real coefficients if not mentioned explicitly. In this paper, actions of $\mathscr{D}_{\mathrm{vol}}$ on similar spaces of homomorphisms, such as $\mathrm{Hom}\big(\bigwedge^l H^1(M), V\big)$, $(V = \Omega^{n-1}(M), H^{n-1}(M), \dots)$, are given by the same formula, $\varphi_\sharp g(a) = \varphi_*(g(\varphi^* a))$, for g considered there. Take a linear section $r\colon H^1(M) \to Z^1(M)$ of the projection $Z^1(M) \to H^1(M)$, where $Z^1(M)$ is the space of closed 1-forms on M.

Lemma 2.1. *There exists a unique crossed homomorphism*

$$f\colon \mathscr{D}_{\mathrm{vol}} \ni \varphi \mapsto f_\varphi \in \mathrm{Hom}(H^1(M), C_0^\infty(M))$$

such that $df_\varphi(a) = \varphi_\sharp r(a) - r(a)$ *for each* $\varphi \in \mathscr{D}_{\mathrm{vol}}$ *and* $a \in H^1(M)$.

Here, $\varphi_\sharp r(a) = \varphi_*\{r(\varphi^* a)\}$ as mentioned above, and by definition, f is a crossed homomorphism if and only if the equality $f_{\varphi\psi} = f_\varphi + \varphi_\sharp f_\psi$ holds for all $\varphi, \psi \in \mathscr{D}_{\mathrm{vol}}$.

For any $a = a_1 \wedge \cdots \wedge a_h \in \bigwedge^h H^1(M)$, put $\bigwedge_r a = r(a_1) \wedge \cdots \wedge r(a_h)$ and $\bigcup a = a_1 \cup \cdots \cup a_h$, we then have the homomorphisms

$$\textstyle\bigwedge_r\colon \bigwedge^h H^1(M) \to \Omega^h(M) \quad \text{and} \quad \bigcup\colon \bigwedge^h H^1(M) \to H^h(M)$$

by linear extension for each h. We use the same symbols $\bigwedge_r$ and $\bigcup$ for different h's.

Let I_n^{2k} be the set of invariant, symmetric, multilinear functions on $\mathrm{gl}(n, \mathbb{R})$ of degree $2k$ with values in $\mathbb{R}$. Take $p \in I_n^{2k}$. We have the Pontryagin class $p(M) \in H^{4k}(M)$ of M corresponding to it. Let A be a $\mathrm{GL}(n, \mathbb{R})$-connection on TM, and $F_A \in \Omega^2(M, \mathrm{End}(TM))$ its curvature form. We then have $p(M) = [\Delta(A)p] \in H^{4k}(M)$ by Chern–Weil theory, where $\Delta(A)p := p(F_A^{(2k)}) = p(F_A, F_A, \dots, F_A)$. We also need $\Delta(A, B)p = 2k \int_0^1 p(B - A, F_{A+t(B-A)}^{(2k-1)})\, dt$, which is the Chern–Simons–Bott form, where B is another $\mathrm{GL}(n, \mathbb{R})$-connection. This introduction of Pontryagin classes, more generally of primary and secondary characteristic classes, is referred to in Chapter 4 of [Vaisman 1987]. It is also helpful to compute them in this paper.

Let $l = n - 4k$ and $\kappa \colon \bigwedge^l H^1(M) \to \mathbb{R}$, the homomorphism defined by

$$\kappa(a) = \big\langle p(M) \cup \bigcup a, [M] \big\rangle \Big/ \int_M \mathrm{vol} \quad \text{for } a \in \textstyle\bigwedge^l H^1(M).$$

Lemma 2.2. *There exists a homomorphism*

$$\mu \colon \textstyle\bigwedge^l H^1(M) \to \Omega^{n-1}(M)$$

such that $d\mu(a) = \kappa(a)\,\mathrm{vol} - \Delta(A)p \wedge \bigwedge_r a$ *for all* $a \in \bigwedge^l H^1(M)$.

For $\varphi \in \mathfrak{D}_{\mathrm{vol}}$ and $a = a_1 \wedge \cdots \wedge a_l \in \bigwedge^l H^1(M)$, put

$$f_\varphi(a) = \sum_{m=1}^{l} (-1)^{m-1} \bigwedge_{j=1}^{m-1} \varphi_\sharp r(a_j) \wedge f_\varphi(a_m) \bigwedge_{j=m+1}^{l} r(a_j)$$

in $\Omega^{l-1}(M)$ and extend it by linear combination for any $a \in \bigwedge^l H^1(M)$. We use the same symbol f_φ for all l, but there is no confusion. Put

$$J_\varphi(a) = \big[\Delta(A, \varphi_* A)p \wedge \textstyle\bigwedge_r a + \Delta(\varphi_* A)p \wedge f_\varphi(a) + \varphi_\sharp \mu(a) - \mu(a)\big].$$

Then it is a well-defined element of $H^{n-1}(M)$. Here, $\varphi_* A$ is the pushforward connection of A by $\varphi_* \colon TM \to TM$. Put

$$\mathcal{H} = \mathrm{Hom}\big(\textstyle\bigwedge^l H^1(M), H^{n-1}(M)\big).$$

We can show $J_\varphi \in \mathcal{H}$ and the following theorem:

Theorem 2.3. *The map*

$$J \colon \mathfrak{D}_{\mathrm{vol}} \ni \varphi \mapsto J_\varphi \in \mathcal{H}$$

is a well-defined crossed homomorphism depending on the choice of p, A, r, *and* μ. *Its cohomology class* $[J] \in H^1(\mathfrak{D}_{\mathrm{vol}}, \mathcal{H})$ *in group cohomology depends only on* p, *not on the choice of* A, r, *and* μ.

Here the action of $\mathscr{D}_{\mathrm{vol}}$ on $\mathscr{H}$ is given by $(\varphi, h) \mapsto \varphi_\sharp h$ as mentioned before. By definition, the map J is a crossed homomorphism if and only if the equality $J_{\varphi\psi} = J_\varphi + \varphi_\sharp J_\psi$ holds for all $\varphi, \psi \in \mathscr{D}_{\mathrm{vol}}$. See [Brown 1982] for group cohomology.

In general, the cohomology classes $[J]$ are nontrivial. Proposition 7.8 gives a condition for them to be nontrivial, and Corollary 7.7 is a simple example of them. These are corollaries of Proposition 2.4 below. Since the target $\mathscr{H}$ of J is the same for all $p \in I_n^{2k}$, we obtain the map from I_n^{2k} to $H^1(\mathscr{D}_{\mathrm{vol}}, \mathscr{H})$ by Theorem 2.3, which is easily checked to be linear, but we don't study this homomorphism in this paper.

In order to state Proposition 2.4, we introduce the volume flux homomorphism. Let $\mathscr{D}_{\mathrm{vol},0}$ be the identity component of $\mathscr{D}_{\mathrm{vol}}$, and $\pi \colon \widetilde{\mathscr{D}}_{\mathrm{vol},0} \to \mathscr{D}_{\mathrm{vol},0}$ its universal covering. Each element of $\widetilde{\mathscr{D}}_{\mathrm{vol},0}$ is represented by a smooth curve $\{\varphi_s\}_{s\in[0,1]} \subset \mathscr{D}_{\mathrm{vol},0}$ with $\varphi_0 = \mathrm{id}_M$, which is denoted by $[\varphi_s] \in \widetilde{\mathscr{D}}_{\mathrm{vol},0}$. The volume flux homomorphism

$$\mathrm{Flux}^\sim \colon \widetilde{\mathscr{D}}_{\mathrm{vol},0} \to H^{n-1}(M), \quad \mathrm{Flux}^\sim([\varphi_s]) = \int_0^1 [\iota(X_s)\,\mathrm{vol}]\,ds$$

with respect to vol is a well-defined surjective homomorphism [Banyaga 1997]. Here X_s is the time-dependent vector field given by $d\varphi_s/ds = X_s \circ \varphi_s$, and $\iota(X_s)$ denotes the interior product by X_s. The image $\mathrm{Flux}^\sim(\pi^{-1}(\mathrm{id}))$ of the fiber $\pi^{-1}(\mathrm{id}) \subset \widetilde{\mathscr{D}}_{\mathrm{vol},0}$ at the identity under $\mathrm{Flux}^\sim$ is called the volume flux group, which is denoted by $\Gamma_{\mathrm{vol}}(M)$.

In order to state the relation of J with $\mathrm{Flux}^\sim$, we define the homomorphisms

$$L, L_+ \colon H^{n-1}(M) \to \mathscr{H} \tag{2-1}$$

as the linear extensions of

$$L(w)(a) = p(M) \cup \sum_{m=1}^{l} (-1)^{m-1} \langle a_m \cup w, [M]\rangle \bigcup_{j=1}^{m-1} a_j \cup \bigcup_{j=m+1}^{l} a_j \Big/ \int_M \mathrm{vol}$$

and $L_+(w)(a) = -\kappa(a)w + L(w)(a)$ for $w \in H^{n-1}(M)$ and $a = a_1 \wedge \cdots \wedge a_l \in \bigwedge^l H^1(M)$.

Proposition 2.4. *The equality $J \circ \pi = L_+ \circ \mathrm{Flux}^\sim$ holds on $\widetilde{\mathscr{D}}_{\mathrm{vol},0}$; that is,*

$$J_{\varphi_1}(a) = -\kappa(a)\,\mathrm{Flux}^\sim([\varphi_s]) + L(\mathrm{Flux}^\sim([\varphi_s]))(a)$$

holds for any $[\varphi_s] \in \widetilde{\mathscr{D}}_{\mathrm{vol},0}$ and $a \in \bigwedge^l H^1(M)$.

Let L_p, L_{p+} and κ_p be the homomorphisms L, L_+, and κ with respect to $p \in I_n^{2k}$ respectively. This proposition implies that the volume flux group $\Gamma_{\mathrm{vol}}(M)$ is contained in the kernel of L_{p+} for all p, since $\Gamma_{\mathrm{vol}}(M)$ is defined independently of p. In particular, if there exists an L_p whose kernel is zero, we have $\Gamma_{\mathrm{vol}}(M) = \{0\}$. The next theorem restates this consequence in terms of L_p, and its examples are given in Corollary 7.6. Let $\mathcal{P}$ be the set of pairs (p, a) of $p \in I_n^{2k}$ and $a \in \bigwedge^{n-4k} H^1(M)$ for some k, and $\mathcal{P}_0 = \{(p, a) \in \mathcal{P} \mid \kappa_p(a) = 0\}$. For each $(p, a) \in \mathcal{P}$, we have the homomorphism $L_p(\,\cdot\,)(a)\colon H^{n-1}(M) \to H^{n-1}(M)$.

Theorem 2.5. *The volume flux group $\Gamma_{\mathrm{vol}}(M)$ of any closed oriented smooth manifold M of dimension n with a volume form* vol *satisfies*

$$\Gamma_{\mathrm{vol}}(M) \subset \Big\{ \bigcap_{(p,a)\in\mathcal{P}_0} \ker L_p(\,\cdot\,)(a) \Big\} \cap \Big\{ \bigcap_{(p,a)\in\mathcal{P}\setminus P_0} \operatorname{Im} L_p(\,\cdot\,)(a) \Big\}.$$

Proposition 2.4 tells us that the image of $\mathcal{D}_{\mathrm{vol},0}$ by J is contained in $\operatorname{Im} L_+$. On the other hand, we know from Moser's result [1965] that the inclusion $\mathcal{D}_{\mathrm{vol}} \hookrightarrow \mathcal{D}$ is a weak homotopy equivalence. Using these, we can show the following theorem:

Theorem 2.6. *The crossed homomorphism J descends to a well-defined one*;

$$\mathcal{J}\colon \pi_0\mathcal{D} \to \mathcal{H}/\operatorname{Im} L_+.$$

Its cohomology class $[\mathcal{J}] \in H^1(\pi_0\mathcal{D}, \mathcal{H}/\operatorname{Im} L_+)$ depends only on p, not on the choice of vol, A, r *and* μ.

As mentioned above, Corollary 7.7 gives a nontrivial example of Theorem 2.3. But since the image of J in this corollary is contained in $\operatorname{Im} L_+$, it is a trivial case of Theorem 2.6, so we need a nontrivial example of the cohomology class $[\mathcal{J}]$.

Let $M = \Sigma_g$ be a closed oriented surface of genus $g \geqq 3$. The isotopy group $\pi_0\mathcal{D}$ is called the mapping class group of Σ_g. We take $p = 1 \in I_2^0$. Then, our crossed homomorphism $J\colon \mathcal{D}_{\mathrm{vol}} \to \mathcal{H} = \operatorname{Hom}\big(\bigwedge^2 H^1(\Sigma_g), H^1(\Sigma_g)\big)$ is simply given by $J_\varphi(a) = [f_\varphi(a) + \varphi_\sharp\mu(a) - \mu(a)]$ for an area form $\omega (= \mathrm{vol})$ on Σ_g, which descends to $\mathcal{J}$ on $\pi_0\mathcal{D}$. The subgroup $\mathcal{I}_g = \{\varphi \in \pi_0\mathcal{D} \mid \varphi_* = \mathrm{id} \text{ on } H_1(\Sigma_g;\mathbb{Z})\} \subset \pi_0\mathcal{D}$ is called the Torelli group. Let $\tau_g\colon \mathcal{I}_g \to \bigwedge^3 H_1(\Sigma_g;\mathbb{Z})/H_1(\Sigma_g;\mathbb{Z})$ be the first Johnson homomorphism [Johnson 1980]. Using Poincaré duality, we have a natural homomorphism $\bigwedge^3 H_1(\Sigma_g;\mathbb{Z}) \hookrightarrow \mathcal{H}$. Moreover, we can see that it descends to an injective homomorphism

$$j\colon \textstyle\bigwedge^3 H_1(\Sigma_g;\mathbb{Z})/H_1(\Sigma_g;\mathbb{Z}) \hookrightarrow \mathcal{H}/\operatorname{Im} L_+. \tag{2-2}$$

Theorem 2.7. *For any closed oriented surface Σ_g of genus $g \geqq 3$, the following equality holds*:

$$\mathcal{J}|_{\mathcal{I}_g} = -\tfrac{1}{2} j \circ \tau_g.$$

We remark that $\mathcal{J}|_{\mathcal{I}_g}$ is a usual homomorphism because the subgroup $\mathcal{I}_g$ of $\pi_0\mathcal{D}$ acts trivially on the target of $\mathcal{J}$. This theorem implies that the restriction $\mathcal{J}|_{\mathcal{I}_g}$ is a nontrivial homomorphism, which defines a nontrivial first cohomology class. Thus, the class $[\mathcal{J}]$ is also nontrivial. This gives a nontrivial example of Theorem 2.6.

3. Lemmas

In this section we will prepare the lemmas which are needed to show that the homomorphism J_φ is defined on the exterior product $\bigwedge^l H^1(M)$ and that the cohomology class of J is unique.

Let $\alpha_j \in Z^1(M)$ and $\zeta_j \in C^\infty(M)$ with $1 \leq j \leq h$. We expand the exterior product $\bigwedge_{j=1}^h(\alpha_j + d\zeta_j)$, we then have

$$\bigwedge_{j=1}^{h}(\alpha_j + d\zeta_j) - \bigwedge_{j=1}^{h}\alpha_j = \sum_{m=1}^{h} \sum_{1\leq k_1<\cdots<k_m\leq h} \bigwedge_{p=1}^{m}\left\{\bigwedge_{j=k_{p-1}+1}^{k_p-1}\alpha_j \wedge d\zeta_{k_p}\right\} \wedge \bigwedge_{j=k_m+1}^{h}\alpha_j,$$

where $k_0 = 0$. Set $\tilde{\alpha} = \alpha_1 \otimes \cdots \otimes \alpha_h$, $\tilde{\zeta} = \zeta_1 \otimes \cdots \otimes \zeta_h$, and

$$K_h(\tilde{\alpha}, \tilde{\zeta}) = \sum_{m=1}^{h} \sum_{1\leq k_1<\cdots<k_m\leq h} (-1)^{k_1-1} \bigwedge_{j=1}^{k_1-1}\alpha_j \wedge \zeta_{k_1} \bigwedge_{p=2}^{m}\left\{\bigwedge_{j=k_{p-1}+1}^{k_p-1}\alpha_j \wedge d\zeta_{k_p}\right\} \wedge \bigwedge_{j=k_m+1}^{h}\alpha_j,$$

we then have $K_h(\tilde{\alpha}, \tilde{\zeta}) \in \Omega^{h-1}(M)$ and

$$\bigwedge_{j=1}^{h}(\alpha_j + d\zeta_j) - \bigwedge_{j=1}^{h}\alpha_j = dK_h(\tilde{\alpha}, \tilde{\zeta}). \tag{3-1}$$

This equality is the only requirement for K_h, so the choice of it is not unique. The following lemma is an easy consequence of the definition of K_h, so we omit the proof:

Lemma 3.1. *The following equalities hold*:

(i) $K_h(\tilde{\alpha}, \tilde{\zeta}) \wedge \beta = K_{h+1}(\tilde{\alpha} \otimes \beta, \tilde{\zeta} \otimes \eta) - (-1)^h \bigwedge\tilde{\alpha} \cdot \eta - K_h(\tilde{\alpha}, \tilde{\zeta}) \wedge d\eta.$

(ii) $\beta \wedge K_h(\tilde{\alpha}, \tilde{\zeta}) = -K_{h+1}(\beta \otimes \tilde{\alpha}, \eta \otimes \tilde{\zeta}) + \eta\bigwedge\tilde{\alpha} + \eta dK_h(\tilde{\alpha}, \tilde{\zeta}).$

where $\beta \in Z^1(M)$, $\eta \in C^\infty(M)$, *and* $\bigwedge\tilde{\alpha} = \alpha_1 \wedge \cdots \wedge \alpha_h$.

Let $\mathscr{S}_h$ be the h-th symmetric group. Put $\tilde{\alpha}_\sigma = \alpha_{\sigma(1)} \otimes \cdots \otimes \alpha_{\sigma(h)}$ and $\tilde{\zeta}_\sigma = \zeta_{\sigma(1)} \otimes \cdots \otimes \zeta_{\sigma(h)}$ for $\sigma \in \mathscr{S}_h$.

Lemma 3.2. *The equality* $K_h(\tilde{\alpha}_\sigma, \tilde{\zeta}_\sigma) \equiv \operatorname{sgn}(\sigma) K_h(\tilde{\alpha}, \tilde{\zeta}) (\bmod\ d\Omega^{h-2}(M))$ *holds for each* $\sigma \in \mathscr{S}_h$.

Proof. It is sufficient to show the equality for any transposition $\sigma = (s, t) \in \mathscr{S}_h$ with $s < t$. For each $1 \leq m \leq h$, put $I_m = \{(k_1, \ldots, k_m) \mid 1 \leq k_1 < \cdots < k_m \leq h\}$, and define the map $\sigma : I_m \to I_m$ by

$$\sigma(u)=\begin{cases} u & \text{if } s,t\notin u,\\ u & \text{if } s,t\in u,\\ \{k_1,\dots,k_m,t\}\setminus\{s\} \text{ reordered in ascending order} & \text{if } s\in u,\ t\notin u,\\ \{k_1,\dots,k_m,s\}\setminus\{t\} \text{ reordered in ascending order} & \text{if } s\notin u,\ t\in u,\end{cases}$$

where $u=(k_1,\dots,k_m)\in I_m$, and $s\in u$ means $s\in\{k_1,\dots,k_m\}$ by abuse of notation. Put

$$A_u=(-1)^{k_1-1}\bigwedge_{j=1}^{k_1-1}\alpha_j\wedge\zeta_{k_1}\bigwedge_{p=2}^{m}\Bigl\{\bigwedge_{j=k_{p-1}+1}^{k_p-1}\alpha_j\wedge d\zeta_{k_p}\Bigr\}\wedge\bigwedge_{j=k_m+1}^{h}\alpha_j$$

and

$$A_u^\sigma=(-1)^{k_1-1}\bigwedge_{j=1}^{k_1-1}\alpha_{\sigma(j)}\wedge\zeta_{\sigma(k_1)}\bigwedge_{p=2}^{m}\Bigl\{\bigwedge_{j=k_{p-1}+1}^{k_p-1}\alpha_{\sigma(j)}\wedge d\zeta_{\sigma(k_p)}\Bigr\}\wedge\bigwedge_{j=k_m+1}^{h}\alpha_{\sigma(j)}$$

for $u=(k_1,\dots,k_m)\in I_m$. It is easy to see the following equalities:

(i) If $s,t\notin u$, then $A_u^\sigma=-A_{\sigma(u)}$.

(ii) If $s,t\in u$, then $A_u^\sigma=-A_{\sigma(u)}+\begin{cases}\text{an exact form} & \text{if } s=k_1,\\ 0 & \text{if } s\neq k_1.\end{cases}$

(iii) If $s\in u,\ t\notin u$, then $A_u^\sigma=-A_{\sigma(u)}+\begin{cases}\text{an exact form} & \text{if } s=k_1<k_2<t,\\ 0 & \text{otherwise.}\end{cases}$

(iv) If $s\notin u,\ t\in u$, then $A_u^\sigma=-A_{\sigma(u)}+\begin{cases}\text{an exact form} & \text{if } s<k_1<k_2\le t,\\ 0 & \text{otherwise.}\end{cases}$

Thus, we have $A_u^\sigma\equiv -A_{\sigma(u)}(\operatorname{mod} d\Omega^{h-2}(M))$. Since the map σ on I_m is bijective, we have

$$\begin{aligned}K_h(\tilde\alpha_\sigma,\tilde\zeta_\sigma)&=\sum_{m=1}^{h}\sum_{u\in I_m}A_u^\sigma\equiv-\sum_{m=1}^{h}\sum_{u\in I_m}A_{\sigma(u)}(\operatorname{mod} d\Omega^{h-2}(M))\\&=-\sum_{m=1}^{h}\sum_{u\in I_m}A_u=\operatorname{sgn}(\sigma)K_h(\tilde\alpha,\tilde\zeta).\end{aligned}$$

□

4. The proofs of Lemmas 2.1 and 2.2

In this section we will prove Lemmas 2.1 and 2.2 and show some properties of the homomorphisms in these lemmas.

We preserve the notation in Section 2.

Proof of Lemma 2.1. Let $B^1(M)$ be the space of smooth exact 1-forms on M. We have $\varphi_\sharp r(a)-r(a)\in B^1(M)$ for any $\varphi\in\mathscr{D}_{\mathrm{vol}}$ and $a\in H^1(M)$ since $[\varphi_\sharp r(a)]=a=[r(a)]\in H^1(M)$. Since the exterior derivative d gives an isomorphism from $C_0^\infty(M)$

to $B^1(M)$, we have a unique $f_\varphi(a) \in C_0^\infty(M)$ satisfying $df_\varphi(a) = \varphi_\sharp r(a) - r(a)$. The equality

$$(\varphi\psi)_\sharp r(a) - r(a) = \varphi_\sharp r(a) - r(a) + \varphi_*(\psi_\sharp r - r)(\varphi^* a)$$

for $\varphi, \psi \in \mathscr{D}_{\mathrm{vol}}$ and $a \in H^1(M)$, and the injectivity of d on $C_0^\infty(M)$ imply $f_{\varphi\psi}(a) = f_\varphi(a) + (\varphi_\sharp f_\psi)(a)$. $\square$

Let $u\colon H^1(M) \to Z^1(M)$ be another linear section of the projection $Z^1(M) \to H^1(M)$. We have two crossed homomorphisms f_r and f_u in Lemma 2.1 with respect to r and u respectively. The proof of the following lemma is almost the same as that of Lemma 2.1, so we omit it:

Lemma 4.1. *There is a unique homomorphism*

$$q = q_{ru}\colon H^1(M) \to C_0^\infty(M)$$

such that $dq(a) = u(a) - r(a)$ for all $a \in H^1(M)$. Moreover, the equality

$$f_{u,\varphi} - f_{r,\varphi} = \varphi_\sharp q - q$$

holds in $\mathrm{Hom}(H^1(M), C_0^\infty(M))$.

Proof of Lemma 2.2. Note that $\kappa(a)\,\mathrm{vol} - \Delta(A)p \wedge \bigwedge_r a \in \Omega^n(M)$ is exact for any $a \in \bigwedge^l H^1(M)$ by the definition of κ. Fix a basis $\{e_i\} \subset \bigwedge^l H^1(M)$ and take $\mu(e_i) \in \Omega^{n-1}(M)$ arbitrarily such that $d\mu(e_i) = \kappa(e_i)\,\mathrm{vol} - \Delta(A) \wedge \bigwedge_r e_i$ for all i. The linear combination of $\mu(e_i)$'s give the required homomorphism. $\square$

For each $\tilde{a} = a_1 \otimes \cdots \otimes a_h \in \bigotimes^h H^1(M)$, set

$$\mathfrak{K}_h(\tilde{a}) = K_h\big(r(a_1) \otimes \cdots \otimes r(a_h), q(a_1) \otimes \cdots \otimes q(a_h)\big),$$

where r and q are the homomorphisms in Lemma 2.1 and 4.1 respectively. Then the linear extension defines a homomorphism

$$\mathfrak{K}_h\colon \bigotimes^h H^1(M) \to \Omega^{h-1}(M)$$

for each h. The composition $\bigotimes^h H^1(M) \to \bigwedge^h H^1(M) \xrightarrow{\bigwedge_r} \Omega^h(M)$ is also denoted by the same symbol, $\bigwedge_r$, where the first homomorphism is the projection $\bigotimes^h H^1(M) \to \bigwedge^h H^1(M)$ given by $\tilde{a} = a_1 \otimes \cdots \otimes a_h \mapsto a = a_1 \wedge \cdots \wedge a_h$.

Similarly, we use the same symbol for a homomorphism on $\bigwedge^h H^1(M)$ and the composition of it with the projection $\bigotimes^h H^1(M) \to \bigwedge^h H^1(M)$. For example, the image of $\tilde{a}$ by the composition $\bigotimes^l H^1(M) \to \bigwedge^l H^1(M) \xrightarrow{\mu} \Omega^{n-1}(M)$ is denoted by $\mu(\tilde{a}) := \mu(a)$. There is no confusion since we can distinguish them by a or $\tilde{a}$.

For each permutation $\sigma \in \mathscr{S}_h$, let $\bigotimes^h H^1(M) \ni \tilde{a} \mapsto \tilde{a}_\sigma \in \bigotimes^h H^1(M)$ be the linear extension of $a_1 \otimes \cdots \otimes a_h \mapsto a_{\sigma(1)} \otimes \cdots \otimes a_{\sigma(h)}$.

Lemma 4.2. *Let $\tilde{a} \in \bigotimes^h H^1(M)$, $b \in H^1(M)$, and $\varphi \in \mathscr{D}_{\mathrm{vol}}$. Then*:

(i) $\bigwedge_u \tilde{a} - \bigwedge_r \tilde{a} = d\mathfrak{K}_h(\tilde{a})$.

(ii) $\varphi_\sharp \mathfrak{K}_{ru,h} = \mathfrak{K}_{\varphi_\sharp r \varphi_\sharp u,h}$, *where* $\mathfrak{K}_{ru,h}$ *and* $\mathfrak{K}_{\varphi_\sharp r \varphi_\sharp u,h}$ *are* $\mathfrak{K}_h$ *with respect to* r, u, *and* $\varphi_\sharp r$, $\varphi_\sharp u$ *respectively.*

(iii) $r(b) \wedge \mathfrak{K}_h(\tilde{a}) = -\mathfrak{K}_{h+1}(b \otimes \tilde{a}) + q(b)\bigwedge_r \tilde{a} + q(b) d\mathfrak{K}_h(\tilde{a})$.

(iv) $\mathfrak{K}_h(\tilde{a}) \wedge r(b) = \mathfrak{K}_{h+1}(\tilde{a} \otimes b) - (-1)^h \bigwedge_r \tilde{a} \cdot q(b) - \mathfrak{K}_h(\tilde{a}) \wedge dq(b)$.

(v) $\varphi_\sharp \mathfrak{K}_h(\tilde{a}) \wedge \varphi_\sharp r(b) = \varphi_\sharp \mathfrak{K}_{h+1}(\tilde{a} \otimes b) - (-1)^h \bigwedge_{\varphi_\sharp r} \tilde{a} \cdot \varphi_\sharp q(b) - \varphi_\sharp \mathfrak{K}_h(\tilde{a}) \wedge d\varphi_\sharp q(b)$.

(vi) $\mathfrak{K}_h(\tilde{a}_\sigma) \equiv \mathrm{sgn}(\sigma) \mathfrak{K}_h(\tilde{a}) (\bmod d\Omega^{h-2}(M))$ *for each* $\sigma \in \mathscr{S}_h$.

Proof. (i) and (ii) follow immediately from the definition of $\mathfrak{K}_h$. (iii), (iv), (v), and (vi) are direct consequences of Lemmas 3.1 and 3.2. □

5. Bott homomorphisms

In this section we shall recall Bott homomorphisms [Vaisman 1987], which are useful for our computation.

Let G be a Lie group, but we consider only $G = \mathrm{GL}(n, \mathbb{R})$ in this paper. Let $\pi : P \to M$ be a principal G-bundle over a manifold M. Let A_h with $h = 0, 1, \dots, r$ be $r + 1$ connection forms on P, and

$$\Delta^r := \left\{ (t_0, t_1, \dots, t_r) \in \mathbb{R}^{r+1} \;\middle|\; t_h \geqq 0, \sum_{h=0}^{r} t_h = 1 \right\}$$

the standard r-simplex. Then we have the average connection $\tilde{A} = \sum_{h=0}^r t_h A_h$ on the product bundle $\pi \times \mathrm{id} \colon P \times \Delta^r \to M \times \Delta^r$. Let $I^k(G)(= I_n^k)$ be the vector space of invariant, symmetric, multilinear functions on the k-th product $\mathfrak{g}^k$ of the Lie algebra $\mathfrak{g}$ of G with values in $\mathbb{R}$. For each $p \in I^k(G)$, put $\Delta(A_0, \dots, A_r)p = (-1)^{\frac{r+1}{2}} \int_{\Delta^r} p(F_{\tilde{A}}^{(k)})$, where $F_{\tilde{A}}$ is the curvature form of $\tilde{A}$, the orientation of Δ^r is given by $dt_1 \wedge \cdots \wedge dt_r$ with $t_0 = 1 - \sum_{h=1}^r t_h$, and $p(F_{\tilde{A}}, F_{\tilde{A}}, \dots, F_{\tilde{A}})$ is denoted by $p(F_{\tilde{A}}^{(k)})$. Then we have Bott homomorphisms $\Delta(A_0, \dots, A_r) \colon I^k(G) \to \Omega^{2k-r}(M)$. They have the following properties:

(i) $d\Delta(A_0, \dots, A_r)p = \sum_{h=0}^r (-1)^h \Delta(A_0, \dots, A_{h-1}, A_{h+1}, \dots, A_r)p$, in particular, $d\Delta(A_0)p = 0$ for $r = 0$.

(ii) $\Delta(A_0) \colon I^k(G) \to \Omega^*(M)$ is the Chern–Weil homomorphism; that is, the equality $\Delta(A_0)p = p(F_{A_0}^{(k)})$ holds.

(iii) $\Delta(A_0, A_1)p = k \int_0^1 p(\alpha, F_{A_t}^{(k-1)})\, dt$, where $\alpha = A_1 - A_0$ and $A_t = A_0 + t\alpha$. Let Q be another principal G-bundle over M, and $\tilde{\varphi} \colon Q \to P$ a G-bundle isomorphism over a diffeomorphism φ of M. The pushforward connection of A by $\tilde{\varphi}$ is denoted by $\tilde{\varphi}_* A := (\tilde{\varphi}^{-1})^* A$.

(iv) $\Delta(\tilde{\varphi}_* A_0, \tilde{\varphi}_* A_1, \dots, \tilde{\varphi}_* A_r)p = \varphi_* \Delta(A_0, A_1, \dots, A_r)p$.

6. Proof of Theorem 2.3

Recall that for any integer h and any homomorphism $v\colon H^1(M) \to Z^1(M)$, $\bigwedge_v\colon \bigwedge^h H^1(M) \to \Omega^h(M)$ is defined as the linear extension of $a_1 \wedge \cdots \wedge a_h \mapsto v(a_1) \wedge \cdots \wedge v(a_h)$. We use the same symbol $\bigwedge_v$ for the composition $\bigotimes^h H^1(M) \to \bigwedge^h H^1(M) \to \Omega^h(M)$ as mentioned above. We consider mainly the cases of $v = r, u, \varphi_\sharp r, \ldots$.

For any $\varphi, \psi \in \mathfrak{D}_{\mathrm{vol}}$ and $\tilde{a} = a_1 \otimes \cdots \otimes a_l \in \bigotimes^l H^1(M)$, put

$$f_\varphi(\tilde{a}) = \sum_{m=1}^{l} (-1)^{m-1} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m-1} \wedge f_\varphi(a_m) \bigwedge_r \tilde{a}_{m+1,l}$$

and

$$\begin{aligned} &f_{\varphi,\psi}(\tilde{a}) \\ &\quad = -\sum_{1 \le k < m \le l} (-1)^{m+k} \textstyle\bigwedge_{(\varphi\psi)_\sharp r} \tilde{a}_{1,k-1} \wedge \varphi_\sharp(f_\psi)(a_k) \bigwedge_{\varphi_\sharp r} \tilde{a}_{k+1,m-1} \wedge f_\varphi(a_m) \bigwedge_r \tilde{a}_{m+1,l}, \end{aligned}$$

where $\tilde{a}_{i,j} = a_i \otimes \cdots \otimes a_j$ with $i < j$ and $\bigwedge_* a_{i+1,i} = 1$. Linear extension defines the maps

$$\mathfrak{D}_{\mathrm{vol}} \ni \varphi \mapsto f_\varphi \in \operatorname{Hom}\bigl(\textstyle\bigotimes^l H^1(M), \Omega^{l-1}(M)\bigr)$$

and

$$\mathfrak{D}_{\mathrm{vol}}^2 \ni (\varphi, \psi) \mapsto f_{\varphi,\psi} \in \operatorname{Hom}\bigl(\textstyle\bigotimes^l H^1(M), \Omega^{l-2}(M)\bigr).$$

Lemma 6.1. *For any $\varphi, \psi \in \mathfrak{D}_{\mathrm{vol}}$, and $\tilde{a} \in \bigotimes^l H^1(M)$, the following equalities hold*:

(i) $df_\varphi(\tilde{a}) = \bigwedge_{\varphi_\sharp r} \tilde{a} - \bigwedge_r \tilde{a}$.

(ii) $d\varphi_\sharp(f_\psi)(\tilde{a}) = \bigwedge_{(\varphi\psi)_\sharp r} \tilde{a} - \bigwedge_{\varphi_\sharp r} \tilde{a}$.

(iii) $\varphi_\sharp f_\psi(\tilde{a}) - f_{\varphi\psi}(\tilde{a}) + f_\varphi(\tilde{a}) = df_{\varphi,\psi}(\tilde{a})$.

Proof. For (i), it is sufficient to show the equality for $\tilde{a} = a_1 \otimes \cdots \otimes a_l \in \bigotimes^l H^1(M)$. Direct computation shows

$$\begin{aligned} df_\varphi(\tilde{a}) &= \sum_{m=1}^{l} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m-1} \wedge df_\varphi(a_m) \wedge \bigwedge_r \tilde{a}_{m+1,l} \\ &= \sum_{m=1}^{l} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m-1} \wedge \{\varphi_\sharp r(a_m) - r(a_m)\} \wedge \bigwedge_r \tilde{a}_{m+1,l} \\ &= \sum_{m=1}^{l} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m} \wedge \bigwedge_r \tilde{a}_{m+1,l} - \sum_{m=1}^{l} \bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m-1} \wedge \bigwedge_r \tilde{a}_{m,l} \\ &= \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a} - \bigwedge_r \tilde{a}. \end{aligned}$$

A similar computation to (i) shows (ii) and (iii), so we omit it. □

Lemma 6.2. *For any $\varphi \in \mathfrak{D}_{\mathrm{vol}}$ and $\tilde{a} = a_1 \otimes \cdots \otimes a_l \in \bigotimes^l H^1(M)$, the following equalities hold*:

(i) $f_\varphi(\tilde{a}) = f_\varphi(\tilde{a}_{1,l-1}) \wedge r(a_l) + (-1)^{l-1} \bigwedge_{\varphi_\sharp r} \tilde{a}_{1,l-1} \wedge f_\varphi(a_l)$.

(ii) $f_\varphi(\tilde{a}) = f_\varphi(a_1) \bigwedge_r \tilde{a}_{2,l} - \varphi_\sharp r(a_1) \wedge f_\varphi(\tilde{a}_{2,l})$.

(iii) $f_\varphi(\tilde{a}_\sigma) = \mathrm{sgn}(\sigma) f_\varphi(\tilde{a}) +$ an exact form, *where $\sigma \in \mathcal{S}_l$ is any permutation of degree l and $\tilde{a}_\sigma = a_{\sigma(1)} \otimes \cdots \otimes a_{\sigma(l)}$*.

Proof. (i) and (ii) directly follow from the definition of f.

For (iii), we proceed by induction on l. It is trivial for $l = 1$. For $l = 2$ and $\sigma = (1, 2) \in \mathcal{S}_2$, the equality $f_\varphi(a_1 \otimes a_2) + f_\varphi(a_2 \otimes a_1) = d\{-f_\varphi(a_1) f_\varphi(a_2)\}$ is a desired one. Assume that the statement holds for $l - 1 (\geqq 1)$. Let $\sigma \in \mathcal{S}_l$. To begin with, we consider the case of $\sigma(l) = l$. We can consider σ as an element of $\mathcal{S}_{l-1}$. By assumption, there exists $h \in \Omega^{l-3}(M)$ such that $f_\varphi(\tilde{a}_{\sigma\{1,l-1\}}) = \mathrm{sgn}(\sigma) f_\varphi(\tilde{a}_{1,l-1}) + dh$, where $\tilde{a}_{\sigma\{1,l-1\}} = a_{\sigma(1)} \otimes \cdots \otimes a_{\sigma(l-1)}$. Using (i), we have

$$\begin{aligned} f_\varphi(\tilde{a}_\sigma) &= f_\varphi(\tilde{a}_{\sigma\{1,l-1\}} \otimes a_l) \\ &= f_\varphi(\tilde{a}_{\sigma\{1,l-1\}}) \wedge r(a_l) + (-1)^{l-1} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{\sigma\{1,l-1\}} \wedge f_\varphi(a_l) \\ &= \{\mathrm{sgn}(\sigma) f_\varphi(\tilde{a}_{1,l-1}) + dh\} \wedge r(a_l) + (-1)^{l-1} \mathrm{sgn}(\sigma) \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,l-1} \wedge f_\varphi(a_l) \\ &= \mathrm{sgn}(\sigma) f_\varphi(\tilde{a}) + d\{h \wedge r(a_l)\}. \end{aligned}$$

In the other cases we consider that σ can be given as the product of a transposition and a permutation of essentially lower degree such as the case above. Using this and induction hypothesis we can show the l case. By induction, we complete the proof of (iii). □

For each $\varphi \in \mathfrak{D}_{\mathrm{vol}}$, we define the map

$$\tilde{J}_\varphi : \bigotimes^l H^1(M) \to \Omega^{n-1}(M)$$

by

$$\tilde{J}_\varphi(\tilde{a}) = \Delta(A, \varphi_* A) p \wedge \textstyle\bigwedge_r \tilde{a} + \Delta(\varphi_* A) p \wedge f_\varphi(\tilde{a}) + \varphi_\sharp \mu(\tilde{a}) - \mu(\tilde{a})$$

for each $\tilde{a} \in \bigotimes^l H^1(M)$.

Lemma 6.3. (i) $d\tilde{J}_\varphi(\tilde{a}) = 0$ *for any $\varphi \in \mathfrak{D}_{\mathrm{vol}}$ and $\tilde{a} \in \bigotimes^l H^1(M)$. Put $J_\varphi(\tilde{a}) = [\tilde{J}_\varphi(\tilde{a})]$.*

(ii) *The map $J_\varphi : \bigotimes^l H^1(M) \to H^{n-1}(M)$ is a homomorphism.*

(iii) *The map $J : \mathfrak{D}_{\mathrm{vol}} \ni \varphi \mapsto J_\varphi \in \mathrm{Hom}\left(\bigotimes^l H^1(M), H^{n-1}(M)\right)$ is a crossed homomorphism.*

Proof. (i) Using the definitions, Lemma 2.2, Lemma 6.1, and properties of Bott homomorphisms, we have

$$\begin{aligned} d\tilde{J}_\varphi(\tilde{a}) &= \{\Delta(\varphi_* A)p - \Delta(A)p\} \wedge \textstyle\bigwedge_r \tilde{a} + \Delta(\varphi_* A)p \wedge \left\{\textstyle\bigwedge_{\varphi_\sharp r}\tilde{a} - \textstyle\bigwedge_r \tilde{a}\right\} \\ &\quad + \varphi_*\left\{\kappa(\varphi^*\tilde{a})\,\mathrm{vol} - \Delta(A)p \wedge \textstyle\bigwedge_r \varphi^*\tilde{a}\right\} - \left\{\kappa(\tilde{a})\,\mathrm{vol} - \Delta(A)p \wedge \textstyle\bigwedge_r \tilde{a}\right\} \\ &= 0. \end{aligned}$$

(ii) By (i), $\tilde{a} \mapsto J_\varphi(\tilde{a})$ is well-defined as a map. Its linearity is obvious.

(iii) Direct computation using Lemma 6.1 and properties of Bott homomorphisms shows the equality

$$\begin{aligned} &\varphi_\sharp \tilde{J}_\psi(\tilde{a}) - \tilde{J}_{\varphi\psi}(\tilde{a}) + \tilde{J}_\varphi(\tilde{a}) \\ &\quad = d\big[\Delta(A, \varphi_* A, (\varphi\psi)_* A)p \wedge \textstyle\bigwedge_r \tilde{a} + \Delta((\varphi\psi)_* A)p \wedge f_{\varphi,\psi}(\tilde{a}) \\ &\qquad - \Delta(\varphi_* A, (\varphi\psi)_* A)p \wedge f_\varphi(\tilde{a})\big] \end{aligned}$$

in $\Omega^{n-1}(M)$. This implies that the map J is a crossed homomorphism. □

Let B be another $\mathrm{GL}(n, \mathbb{R})$-connection on TM and $u\colon H^1(M) \to Z^1(M)$ another linear section of the projection $Z^1(M) \to H^1(M)$. Let

$$\tilde{\nu}\colon \textstyle\bigotimes^l H^1(M) \to \Omega^{n-1}(M)$$

be the homomorphism defined by

$$\tilde{\nu}(\tilde{a}) = \mu_T(\tilde{a}) - \mu_S(\tilde{a}) + \Delta(A, B)p \wedge \textstyle\bigwedge_u \tilde{a} + \Delta(A)p \wedge \mathfrak{K}_l(\tilde{a})$$

for each $\tilde{a} \in \bigotimes^l H^1(M)$. Here μ_T and μ_S are μ's in Lemma 2.2 with respect to $T = \{p, B, u\}$ and $S = \{p, A, r\}$ respectively.

Lemma 6.4. (i) $d\tilde{\nu}(\tilde{a}) = 0$ *for any* $\tilde{a} \in \bigotimes^l H^1(M)$.

(ii) *The map*

$$\nu\colon \textstyle\bigotimes^l H^1(M) \ni \tilde{a} \mapsto [\tilde{\nu}(\tilde{a})] \in H^{n-1}(M)$$

is a well-defined homomorphism.

Proof. (i) Using Lemma 2.2, Lemma 4.2, and properties of the Bott homomorphisms, we have

$$\begin{aligned} d\tilde{\nu}(\tilde{a}) &= \kappa(\tilde{a})\,\mathrm{vol} - \Delta(B)p \wedge \textstyle\bigwedge_u \tilde{a} - \left\{\kappa(\tilde{a})\,\mathrm{vol} - \Delta(A)p \wedge \textstyle\bigwedge_r \tilde{a}\right\} \\ &\quad + \left\{\Delta(B)p - \Delta(A)p\right\} \wedge \textstyle\bigwedge_u \tilde{a} + \Delta(A)p \wedge \left\{\textstyle\bigwedge_u \tilde{a} - \textstyle\bigwedge_r \tilde{a}\right\} \\ &= 0. \end{aligned}$$

(ii) By (i), ν is defined as a map and its linearity is clear. □

$\tilde{J}_\varphi(\tilde{a})$ is written as

$$\tilde{J}_{S,\varphi}(\tilde{a}) = \Delta(A, \varphi_* A)p \wedge \textstyle\bigwedge_r \tilde{a} + \Delta(\varphi_* A)p \wedge f_{r,\varphi}(\tilde{a}) + \varphi_\sharp \mu_S(\tilde{a}) - \mu_S(\tilde{a})$$

for the choice $S = \{p, A, r, \mu_S\}$ of the ingredients to define it. S is doubly used, but there is no confusion, and f_r denotes the f in Lemma 2.1 with respect to r. We consider also

$$\tilde{J}_{T,\varphi}(\tilde{a}) = \Delta(B, \varphi_* B)p \wedge \textstyle\bigwedge_u \tilde{a} + \Delta(\varphi_* B)p \wedge f_{u,\varphi}(\tilde{a}) + \varphi_\sharp \mu_T(\tilde{a}) - \mu_T(\tilde{a})$$

with respect to another choice $T = \{p, B, u, \mu_T\}$. We note that p in S and T is common.

Lemma 6.5. *For any $\varphi \in \mathscr{D}_{\mathrm{vol}}$ and $\tilde{a} \in \bigotimes^l H^1(M)$, the following equality holds*:

$$\tilde{J}_{T,\varphi}(\tilde{a}) - \tilde{J}_{S,\varphi}(\tilde{a}) = \varphi_\sharp \tilde{\nu}(\tilde{a}) - \tilde{\nu}(\tilde{a}) + \text{an exact form.}$$

Proof. It is sufficient to show the following equality for $\tilde{a} = a_1 \otimes \cdots \otimes a_l$ with $a_i \in H^1(M)$:

$$\begin{aligned}
&\tilde{J}_{T,\varphi}(\tilde{a}) - \tilde{J}_{S,\varphi}(\tilde{a}) - \bigl\{\varphi_\sharp \tilde{\nu}(\tilde{a}) - \tilde{\nu}(\tilde{a})\bigr\} \\
&\quad = d\biggl(\{\Delta(A, B, \varphi_* B)p - \Delta(A, \varphi_* A, \varphi_* B)p\} \wedge \textstyle\bigwedge_u \tilde{a} \\
&\qquad\quad + (-1)^{4k-1}\Delta(A, \varphi_* A)p \wedge \mathfrak{K}_l(\tilde{a}) - (-1)^{4k-1}\{\varphi_* \Delta(A, B)p\} \wedge f_{u,\varphi}(\tilde{a}) \\
&\qquad\qquad + \Delta(\varphi_* A)p \wedge \biggl[\sum_{m=1}^{l} (-1)^{m-1}\bigl\{\varphi_\sharp \mathfrak{K}_{m-1}(\tilde{a}_{1,m-1}) \wedge f_{u,\varphi}(a_m) \textstyle\bigwedge_u \tilde{a}_{m+1,l} \\
&\qquad\qquad\qquad + (-1)^{m-1} \textstyle\bigwedge_{\varphi_\sharp r} \tilde{a}_{1,m-1} \wedge f_{r,\varphi}(a_m) \mathfrak{K}_{l-m}(\tilde{a}_{m+1,l})\bigr\}\biggr]\biggr).
\end{aligned}$$

Here we note $\mathfrak{K}_0(*) = 0$ and $\bigwedge_{**} \tilde{a}_{*+1,*} = 1$.

It is easy to show that the difference of both sides of the expression is equal to 0. The computation is carried out by using Lemmas 2.1, 4.1, 4.2, and 6.1. But it is standard, so we omit it. □

Proof of Theorem 2.3. For any $\varphi \in \mathscr{D}_{\mathrm{vol}}$, the homomorphism J_φ defined in Lemma 6.3 descends to the homomorphism $J_\varphi : \bigwedge^l H^1(M) \to H^{n-1}(M)$ by Lemma 6.2(iii) and the closedness of $\Delta(A)p$. Here we use the same symbol J_φ for different domains $\bigotimes^l H^1(M)$ and $\bigwedge^l H^1(M)$ as mentioned before. Lemma 6.3 implies that the map $J : \mathscr{D}_{\mathrm{vol}} \to \mathscr{H}$ is a crossed homomorphism. Since a crossed homomorphism is a 1-cocycle in group cohomology theory, we have the cohomology class $[J] \in H^1(\mathscr{D}_{\mathrm{vol}}, \mathscr{H})$ in group cohomology [Brown 1982]. Similarly, the homomorphism ν defined in Lemma 6.4(ii) descends to a homomorphism $\nu \in \mathscr{H}$ by Lemma 4.2(vi). For any $S = \{p, A, r, \mu_S\}$ and $T = \{p, B, u, \mu_T\}$ with common p, Lemma 6.5

implies $J_{T,\varphi} - J_{S,\varphi} = \varphi_\sharp \nu - \nu$; hence, $J_T - J_S = \delta\nu$ as 1-cocycles in $C^1(\mathcal{D}_{\mathrm{vol}}, \mathcal{H})$, where $\mathcal{H} = C^0(\mathcal{D}_{\mathrm{vol}}, \mathcal{H}) \xrightarrow{\delta} C^1(\mathcal{D}_{\mathrm{vol}}, \mathcal{H}) \xrightarrow{\delta} \cdots$ is the cochain complex. Thus, we have the independence of $[J]$ from the choice of S except for $p \in S$. $\square$

7. The derivative of J

In this section we will compute the derivative of J along curves in $\mathcal{D}_{\mathrm{vol}}$ and show some results related to volume flux homomorphisms and groups.

Let $\{\varphi_s\}_{s\in[0,1]} \subset \mathcal{D}_{\mathrm{vol}}$ be any smooth curve and X_s the time-dependent vector field on M defined by $d\varphi_s/ds = X_s \circ \varphi_s$. Let

$$I : C^\infty(M) \to \mathbb{R}$$

be the homomorphism defined by $I(h) = \int_M h \,\mathrm{vol} / \int_M \mathrm{vol}$ for $h \in C^\infty(M)$. Recall that f is the crossed homomorphism defined in Lemma 2.1.

Lemma 7.1. *$df_{\varphi_s}(a)/ds = -\iota(X_s)\varphi_{s\sharp} r(a) + I(\iota(X_s)\varphi_{s\sharp} r(a))$ for any $a \in H^1(M)$.*

Proof. Put $Y_s := -(\varphi_s^{-1})_* X_s$, where $(\,\cdot\,)_*$ denotes the push-forward of vector fields. It satisfies $d\varphi_s^{-1}/ds = Y_s \circ \varphi_s^{-1}$. Set $b := \varphi_s^* a \in H^1(M)$. It is constant with respect to s and we have $df_{\varphi_s}(a) = (\varphi_s^{-1})^* r(b) - r(a)$. Using Cartan's formula for the Lie derivative $\mathcal{L}$ of differential forms, we have

$$\begin{aligned} d\left\{\frac{d}{ds} f_{\varphi_s}(a)\right\} &= \frac{d}{ds}\{(\varphi_s^{-1})^* r(b)\} = (\varphi_s^{-1})^* \mathcal{L}_{Y_s} r(b) \\ &= (\varphi_s^{-1})^*\{d\iota(Y_s) r(b) + \iota(Y_s) dr(b)\} = -d\{\iota(X_s)\varphi_{s*} r(b)\} \\ &= d\{-\iota(X_s)\varphi_{s\sharp} r(a) + I(\iota(X_s)\varphi_{s\sharp} r(a))\}. \end{aligned}$$

Since $(df_{\varphi_s}/ds)(a)$ and $-\iota(X_s)\varphi_{s\sharp} r(a) + I(\iota(X_s)\varphi_{s\sharp} r(a))$ belong to $C_0^\infty(M)$, we obtain the lemma. $\square$

In the following lemma, f_φ and μ are the homomorphisms from $\bigotimes^l H^1(M)$ to $\Omega^{n-1}(M)$:

Lemma 7.2. *For any $\tilde{a}_{1,l} = a_1 \otimes \cdots \otimes a_l$ and $\tilde{a} \in \bigotimes^l H^1(M)$, the following equalities hold*:

(i) $$\begin{aligned} \frac{d}{ds} f_{\varphi_s}(\tilde{a}_{1,l}) = d\Big\{ &\textstyle\sum_{m=1}^{l} (-1)^m \iota(X_s)\big(\bigwedge_{\varphi_{s\sharp} r} \tilde{a}_{1,m-1}\big) \wedge f_{\varphi_s}(a_m) \bigwedge_r \tilde{a}_{m+1,l}\Big\} \\ &+ \textstyle\sum_{m=1}^{l} (-1)^{m-1} \bigwedge_{\varphi_{s\sharp} r} \tilde{a}_{1,m-1} \wedge I(\iota(X_s)\varphi_{s\sharp} r(a_m)) \bigwedge_r \tilde{a}_{m+1,l} \\ &- \iota(X_s)\big(\textstyle\bigwedge_{\varphi_{s\sharp} r} \tilde{a}_{1,l}\big). \end{aligned}$$

(ii) $$\begin{aligned} \frac{d}{ds}\{\varphi_{s\sharp}\mu(\tilde{a}) - \mu(\tilde{a})\} = &-\kappa(\tilde{a})\iota(X_s)\,\mathrm{vol} + \{\iota(X_s)\Delta(\varphi_{s*}A)p\} \wedge \textstyle\bigwedge_{\varphi_{s\sharp} r} \tilde{a} \\ &+ \Delta(\varphi_{s*}A)p \wedge \iota(X_s)\big(\textstyle\bigwedge_{\varphi_{s\sharp} r} \tilde{a}\big) - d\{\iota(X_s)\varphi_{s\sharp}\mu(\tilde{a})\}. \end{aligned}$$

Proof. We note that $\varphi_s^* a_j \in H^1(M)$ is constant with respect to s.

(i) Similar to the proof of Lemma 7.1, we get

$$\frac{d}{ds}\big(\textstyle\bigwedge_{\varphi_{s\sharp}r}\tilde{a}_{1,m-1}\big) = -d\big\{\iota(X_s)\big(\textstyle\bigwedge_{\varphi_{s\sharp}r}\tilde{a}_{1,m-1}\big)\big\},$$

and using Lemma 7.1, we have

$$\begin{aligned}\frac{d}{ds} f_{\varphi_s}(\tilde{a}_{1,l}) &= \sum_{m=1}^{l}(-1)^m d\big\{\iota(X_s)\big(\textstyle\bigwedge_{\varphi_{s\sharp}r}\tilde{a}_{1,m-1}\big)\big\} \wedge f_{\varphi_s}(a_m)\textstyle\bigwedge_r \tilde{a}_{m+1,l}\\ &\quad - \sum_{m=1}^{l}(-1)^{m-1}\textstyle\bigwedge_{\varphi_{s\sharp}r}\tilde{a}_{1,m-1} \wedge \big\{\iota(X_s)\varphi_{s\sharp}r(a_m)\big\}\textstyle\bigwedge_r \tilde{a}_{m+1,l}\\ &\quad + \sum_{m=1}^{l}(-1)^{m-1}\textstyle\bigwedge_{\varphi_{s\sharp}r}\tilde{a}_{1,m-1} \wedge I(\iota(X_s)\varphi_{s\sharp}r(a_m))\textstyle\bigwedge_r \tilde{a}_{m+1,l}.\end{aligned}$$

Computing the first term $d\{\dots\}(= RHS_1)$ in the right hand side of (i), and comparing this result with the last equality, we have the desired equality. In the computation, we note that $\sum_{m=1}^{l} = \sum_{m=2}^{l}$ in RHS_1 and that the interior product is an antiderivation of degree -1.

(ii) Using Cartan's formula for the Lie derivative, we can show (ii) by direct computation. □

Lemma 7.3. *The following equalities hold*:

(i) $\frac{d}{ds}(\varphi_{s*}A - A) = d_{\varphi_{s*}A}\beta_s - \iota(X_s)F_{\varphi_{s*}A}$ *for some* $\beta_s \in \Gamma(End(TM))$, *where* $d_{\varphi_{s*}A}$ *denotes the covariant exterior derivative with respect to* $\varphi_{s*}A$.

(ii) $\frac{d}{ds}\Delta(\varphi_{s*}A)p = -d\,\{\iota(X_s)\Delta(\varphi_{s*}A)p\}$.

(iii) $\frac{d}{ds}\Delta(A, \varphi_{s*}A)p = -\iota(X_s)\Delta(\varphi_{s*}A)p + dR_s$ *for some* $R_s \in \Omega^{4k-1}(M)$.

Proof. For (ii), since $\Delta(\varphi_{s*}A)p = \varphi_{s*}\Delta(A)p$ is a closed form, we can show the equality by using Cartan's formula.

Equalities (i) and (iii) can be obtained in the same way as the proofs of Lemma 4.2 in [Kasagawa 2008] and the equality below (5.1) in the same paper, so we omit the proofs. □

For each $p \in I_n^{2k}$, let

$$\tilde{L}, \tilde{L}_+ : H^{n-1}(M) \to \mathrm{Hom}\big(\textstyle\bigotimes^l H^1(M), H^{n-1}(M)\big) \tag{7-1}$$

be the homomorphisms defined as the linear extensions of

$$\tilde{L}(w)(\tilde{a}) = p(M) \cup \sum_{m=1}^{l}(-1)^{m-1}\langle a_m \cup w, [M]\rangle \bigcup_{j=1}^{m-1} a_j \cup \bigcup_{j=m+1}^{l} a_j \Big/ \int_M \mathrm{vol}$$

and

$$\tilde{L}_+(w)(\tilde{a}) = -\kappa(\tilde{a})w + \tilde{L}(w)(\tilde{a})$$

for $\tilde{a} = a_1 \otimes \cdots \otimes a_l \in \bigotimes^l H^1(M)$.

The following lemma can be easily checked, so we omit the proof.

Lemma 7.4. *For any $w \in H^{n-1}(M)$, $\tilde{a} \in \otimes^l H^1(M)$, and $\sigma \in \mathfrak{S}_l$, the equalities $\tilde{L}(w)(\tilde{a}_\sigma) = \operatorname{sgn}(\sigma)\tilde{L}(w)(\tilde{a})$ and $\tilde{L}_+(w)(\tilde{a}_\sigma) = \operatorname{sgn}(\sigma)\tilde{L}_+(w)(\tilde{a})$ hold. Hence, $\tilde{L}$ and $\tilde{L}_+$ induce the homomorphisms L and L_+ in (2-1) respectively.*

Proposition 7.5. *For any smooth curve $\{\varphi_s\} \subset \mathcal{D}_{\mathrm{vol}}$, the equality*

$$\frac{d}{ds} J_{\varphi_s} = L_+([\iota(X_s)\,\mathrm{vol}]) \in \mathcal{H} \tag{7-2}$$

holds. Here, X_s is the time-dependent vector field on M given by $d\varphi_s/ds = X_s \circ \varphi_s$.

Proof. To begin with, we compute the derivative $d\tilde{J}_{\varphi_s}(\tilde{a})/ds$ for each

$$\tilde{a} = a_1 \otimes \cdots \otimes a_l \in \bigotimes^l H^1(M).$$

By using Lemmas 7.1, 7.2, and 7.3, we have

$$\begin{aligned}
\frac{d}{ds}&\tilde{J}_{\varphi_s}(\tilde{a}) \\
&= \left\{\frac{d}{ds}\Delta(A, \varphi_{s*}A)p\right\} \wedge \textstyle\bigwedge_r \tilde{a} + \left\{\frac{d}{ds}\Delta(\varphi_{s*}A)p\right\} \wedge f_{\varphi_s}(\tilde{a}) \\
&\qquad\qquad + \Delta(\varphi_{s*}A)p \wedge \frac{d}{ds} f_{\varphi_s}(\tilde{a}) + \frac{d}{ds}\{\varphi_{s\sharp}\mu(\tilde{a}) - \mu(\tilde{a})\} \\
&= -\kappa(\tilde{a})\iota(X_s)\,\mathrm{vol} \\
&\quad + \Delta(\varphi_{s*}A)p \wedge \left\{\sum_{m=1}^{l}(-1)^{m-1} \textstyle\bigwedge_{\varphi_{s\sharp}r} \tilde{a}_{1,m-1} \wedge I(\iota(X_s)\varphi_{s\sharp}r(a_m)) \textstyle\bigwedge_r \tilde{a}_{m+1,l}\right\} \\
&\quad + d\Big[R_s \wedge \textstyle\bigwedge_r \tilde{a} - \{\iota(X_s)\Delta(\varphi_{s*}A)p\} \wedge f_{\varphi_s}(\tilde{a}) \\
&\qquad + \Delta(\varphi_{s*}A)p \wedge \left\{\sum_{m=1}^{l}(-1)^m \iota(X_s)\big(\textstyle\bigwedge_{\varphi_{s\sharp}r} \tilde{a}_{1,m-1}\big) \wedge f_{\varphi_s}(a_m) \textstyle\bigwedge_r a_{m+1,l}\right\} \\
&\qquad\qquad - \iota(X_s)\varphi_{s\sharp}\mu(\tilde{a})\Big].
\end{aligned}$$

Since $\varphi_{s\sharp}r(a_m) \wedge \mathrm{vol} = 0$, we get $\{\iota(X_s)\varphi_{s\sharp}r(a_m)\}\,\mathrm{vol} = \varphi_{s\sharp}r(a_m) \wedge \iota(X_s)\,\mathrm{vol}$. By integrating this on M, we obtain

$$I(\iota(X_s)\varphi_{s\sharp}r(a_m)) \int_M \mathrm{vol} = \langle [a_m] \cup [\iota(X_s)\,\mathrm{vol}], [M]\rangle.$$

Using these equalities and linearity, we have

$$d\tilde{J}_{\varphi_s}(\tilde{a})/ds = [dJ_{\varphi_s}(\tilde{a})/ds] = \tilde{L}_+([\iota(X_s)\,\mathrm{vol}])(\tilde{a})$$

for each $\tilde{a}$; therefore, we obtain $dJ_{\varphi_s}/ds = \tilde{L}_+([\iota(X_s)\,\mathrm{vol}])$ as homomorphisms from $\bigotimes^l H^1(M)$ to $H^{n-1}(M)$. By Lemma 7.4, we have the proposition. □

Proposition 2.4 and Theorem 2.5, which are results related with volume flux homomorphisms and groups, easily follow from Proposition 7.5, so we show them and their corollaries here.

Proof of Proposition 2.4. We integrate the equality in Proposition 7.5 with respect to s for $[\varphi_s] \in \tilde{\mathfrak{D}}_{\mathrm{vol},0}$. Since the map L_+ is linear and $J_{\mathrm{id}} = 0$, we have

$$J_{\varphi_1} = L_+\left(\int_0^1 [\iota(X_s)\,\mathrm{vol}]\,ds\right) = L_+(\mathrm{Flux}^{\sim}([\varphi_s])).$$

This implies the proposition. □

Proof of Theorem 2.5. Take $[\varphi_s] \in \pi^{-1}(\mathrm{id}) \subset \tilde{\mathfrak{D}}_{\mathrm{vol},0}$. Since $\varphi_0 = \varphi_1 = \mathrm{id}$ and $J_{p,\mathrm{id}}(a) = 0$ for any $(p, a) \in \mathcal{P}$, by Proposition 2.4 we have $-\kappa_p(a)\,\mathrm{Flux}^{\sim}([\varphi_s]) + L_p(\mathrm{Flux}^{\sim}([\varphi_s]))(a) = 0$; therefore, $\mathrm{Flux}^{\sim}([\varphi_s]) \in \ker L_p(\,\cdot\,)(a)$ if $\kappa_p(a) = 0$, otherwise $\mathrm{Flux}^{\sim}([\varphi_s]) \in \mathrm{Im}\, L_p(\,\cdot\,)(a)$. Since the flux homomorphism is defined independently of $(p, a) \in \mathcal{P}$, the theorem follows. □

Let $\mathrm{Pont}_k(M)$ be the space of the Pontryagin classes lying in $H^{4k}(M)$. As an application of Theorem 2.5, we can show the following corollary:

Corollary 7.6. *Let M be a closed oriented smooth manifold of dimension n with a volume form* vol *satisfying one of the following conditions*:

(i) $n = 4k$ *and* $\dim \mathrm{Pont}_k(M) = 1$,

(ii) $n = 4k + 1$ *and* $\dim \mathrm{Pont}_k(M) \geqq 2$, *or*

(iii) *there exists a rational homology $n-1$ sphere $N \subset M$ separating M into two connected submanifolds M_1 and M_2 with boundaries N and $-N$ satisfying* $\bigcup^n H^1(M_i, N) = H^n(M_i, N)$ *with* $i = 1, 2$.

Then the volume flux group of (M, vol) *is trivial*; *that is*, $\Gamma_{\mathrm{vol}}(M) = \{0\}$.

Proof. (i) By the assumption, there is a nonzero Pontryagin class $p(M) \in H^n(M)$. Take $a = 1 \in \bigwedge^0 H^1(M) \cong \mathbb{R}$. The pair $(p, 1)$ belongs to $\mathcal{P} \setminus \mathcal{P}_0$ in Theorem 2.5. Since the map L_p is the zero map, we have $\mathrm{Im}\, L_p(\,\cdot\,)(1) = \{0\}$. By Theorem 2.5, we have the result.

(ii) By the assumption, there are Pontryagin classes $p(M)$ and $q(M)$ which are linearly independent in $H^{n-1}(M)$. By Poincaré duality, we can take $a, b \in H^1(M)$ satisfying $p(M)\cup a \neq 0$ and $q(M)\cup b \neq 0$ in $H^n(M)$. Then the pairs (p, a) and (q, b) belong to $\mathcal{P} \setminus \mathcal{P}_0$, and we have $\mathrm{Im}\, L_p(\,\cdot\,)(a) = \mathbb{R}p(M)$ and $\mathrm{Im}\, L_q(\,\cdot\,)(b) = \mathbb{R}q(M)$. Since their intersection is $\{0\}$, the result follows.

(iii) Let $\iota_i\colon M_i \hookrightarrow M$ with $i = 1, 2$ be the inclusions. Then we can show the induced homomorphism $\iota_1^* + \iota_2^*\colon H^j(M_1, N) \oplus H^j(M_2, N) \to H^j(M)$ is an isomorphism for each $1 \le j \le n-1$. By the assumption, there exist $e_{i1}, \dots, e_{in} \in H^1(M_i, N)$ with $\bigcup_{j=1}^n e_{ij} \neq 0$ in $H^n(M_i, N)$ for $i = 1, 2$. Put $a_i = \iota_i^*\bigl(\bigcup_{j=1}^n e_{ij}\bigr) \in H^n(M)$ with $i = 1, 2$, the pairs $(1, a_1)$ and $(1, a_2)$ then belong to $\mathcal{P} \setminus \mathcal{P}_0$, where $p = 1 \in I_{2n}^0$. By the definition of the homomorphism $L = L_{p=1}$, we have $\operatorname{Im} L(\cdot)(a_i) \subset \iota_i^*(H^{n-1}(M_i, N))$ for $i = 1, 2$. Using Theorem 2.5, we have $\Gamma_{\mathrm{vol}}(M) \subset \iota_1^*(H^{n-1}(M_1, N)) \cap \iota_2^*(H^{n-1}(M_2, N)) = \{0\}$. $\square$

We remark that Kędra, Kotschick, and Morita [2006] have obtained various conditions for the volume flux group $\Gamma_{\mathrm{vol}}(M)$ to vanish. (i) in Corollary 7.6 is one of them. The case of a connected sum in (iii) can also be obtained from their stronger results for volume flux groups. Moreover, in the case of (i) in Corollary 7.6, the volume flux homomorphism $\mathrm{Flux}^\sim$ descends to a homomorphism $\mathrm{Flux}\colon \mathcal{D}_{\mathrm{vol},0} \to H^{n-1}(M)/\Gamma_{\mathrm{vol}}(M) = H^{n-1}(M)$, which is also called the volume flux homomorphism. In this case they also proved that Flux extends to a crossed homomorphism on the whole group $\mathcal{D}_{\mathrm{vol}}$, but they didn't give an explicit formula of it.

Corollary 7.7. *In the case of* (*i*) *in Corollary 7.6, take a $p \in I_n^{2k}$ such that $\langle p(M), [M]\rangle = \int_M \mathrm{vol}$. Then the crossed homomorphism J with respect to p is rewritten as*

$$J\colon \mathcal{D}_{\mathrm{vol}} \to H^{n-1}(M),\ \ J_\varphi = [\Delta(A, \varphi_* A)p + \varphi_*\mu - \mu],$$

and it is an extension of the flux homomorphism Flux *as a crossed one.*

Proof. Since $l = n - 4k = 0$, we have $\bigwedge^l H^1(M) \cong \mathbb{R}$. Then the target of J can be considered as $H^{n-1}(M)$. By the same reason, we have the simple form of J as above. In particular, μ is an $(n-1)$-form satisfying $d\mu = \mathrm{vol} - \Delta(A)p$. Since $L = 0$, Proposition 2.4 and $\Gamma_{\mathrm{vol}}(M) = \{0\}$ imply that the restriction of J to $\mathcal{D}_{\mathrm{vol},0}$ is equal to $\mathrm{Flux}^\sim = \mathrm{Flux}$. $\square$

This corollary also gives an example of the next proposition. We return to the situation of Theorem 2.3.

Proposition 7.8. *If the homomorphism L_+ is nontrivial, so is the cohomology class* $[J]$ *in Theorem 2.3.*

Proof. Since $\mathcal{D}_{\mathrm{vol},0}$ acts trivially on the target of J, the restriction $J|_{\mathcal{D}_{\mathrm{vol},0}}$ is a usual homomorphism. Moreover, the assumption and Proposition 2.4 imply that it is nonzero since the flux homomorphism $\mathrm{Flux}^\sim$ is surjective. The cohomology class of a nontrivial homomorphism is also nontrivial. Thus, the restriction $[J|_{\mathcal{D}_{\mathrm{vol},0}}]$ of the class $[J]$ in Theorem 2.3 is nonzero; hence, so is $[J]$. $\square$

Since the flux homomorphism $\mathrm{Flux}^{\sim}$ is surjective, the image $J(\mathscr{D}_{\mathrm{vol},0})$ of $\mathscr{D}_{\mathrm{vol},0}$ by J coincides with $\operatorname{Im} L_+$ by Proposition 2.4. This implies that J descends to a map

$$\mathscr{J}\colon \pi_0\mathscr{D}_{\mathrm{vol}} \to \mathscr{H}/\operatorname{Im} L_+.$$

The following lemma follows from Theorem 2.3:

Lemma 7.9. *The map $\mathscr{J}$ is a crossed homomorphism and its cohomology class*

$$[\mathscr{J}] \in H^1(\pi_0\mathscr{D}_{\mathrm{vol}}, \mathscr{H}/\operatorname{Im} L_+)$$

depends only on p, not on the choice of A, r and μ.

Here we give a trivial example of J.

Example 7.10. We consider the standard n-dimensional torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$. Let $(x_1, \dots, x_n) \in \mathbb{R}^n$ be the standard coordinates of $\mathbb{R}^n$. The volume form $dx^1 \wedge \cdots \wedge dx^n$ also gives the standard one vol on T^n with $\int_{T^n} \mathrm{vol} = 1$. One-forms $dx^1, \dots, dx^n$ give a basis $[dx^1], \dots, [dx^n]$ of $H^1(T^n)$. Put $Y = [dx^1] \wedge \cdots \wedge [dx^n]$. Then this is the base of $\bigwedge^n H^1(T^n)$.

We consider the case of $p = 1$; hence, $k = 0$, $l = n - 4k = n$, and a connection A is not needed. We take the section $r\colon H^1(T^n) \to Z^1(T^n)$ given by $r([dx^i]) = dx^i$ with $i = 1, \dots, n$. Since $\kappa(Y)\,\mathrm{vol} - \bigwedge_r Y = 0$, we can take $\mu = 0$. Thus, we have the crossed homomorphism $J\colon \mathscr{D}_{\mathrm{vol}} \to \mathscr{H} = \operatorname{Hom}\bigl(\bigwedge^n H^1(T^n), H^{n-1}(T^n)\bigr)$. In this case we have $J_\varphi(a) = [f_\varphi(a)]$ because $\mu = 0$ and $p = 1$. Since $\kappa(Y) = 1$ and $L(w)(Y) = w$ for any $w \in H^{n-1}(T^n)$, which are easily checked, we have $J_\varphi(Y) = 0$ for any $\varphi \in \mathscr{D}_{\mathrm{vol},0}$ by Proposition 2.4. Let $\varphi \in \mathrm{SL}(n, \mathbb{Z}) \subset \mathscr{D}_{\mathrm{vol}}$. We have $\varphi_\sharp r(a) - r(a) = 0$ for all $a \in H^1(T^n)$. This implies $f_\varphi(a) = 0$ hence $J_\varphi = 0$. Thus, J descends to a crossed homomorphism $\mathscr{J}^{\sim}\colon \pi_0\mathscr{D}_{\mathrm{vol}} \to \mathscr{H}$, whose image of $\mathrm{SL}(n, \mathbb{Z}) (\subset \pi_0\mathscr{D}_{\mathrm{vol}})$ is $\{0\}$.

Let $n \geqq 5$. There is a split exact sequence $0 \to K \to \pi_0 \operatorname{Diff}(T^n) \to \mathrm{GL}(n, \mathbb{Z}) \to 0$, where $K = \mathbb{Z}_2^\infty \oplus \binom{n}{2}\mathbb{Z}_2 \oplus \sum_{i=0}^n \binom{n}{i}\Gamma_{i+1}$, by [Hatcher 1978] for $n \geqq 5$ and [Hsiang and Sharpe 1976] for $n \geqq 6$. So we have a split one $0 \to K \to \pi_0\mathscr{D}_{\mathrm{vol}} \to \mathrm{SL}(n, \mathbb{Z}) \to 0$. Here the action of K on $H^*(T^n)$ is trivial. The groups Γ_{i+1} of twisted spheres are finite abelian groups (see [Milnor 2011] and its references). Hence, every element of K is of finite order. This shows $\mathscr{J}^{\sim}(K) = \{0\}$; hence, $\operatorname{Im} \mathscr{J}^{\sim} = \{0\}$, so we have $J = 0$ for the choice of r and μ as above. Its cohomology class $[J]$ is also zero.

8. Proof of Theorem 2.6

In this section we will prove Theorem 2.6, whose main part is that the crossed homomorphism $\mathscr{J}$ is essentially independent of the choice of volume form on M.

Let vol and vol$'$ be two positive volume forms on M. By Moser [1965], there exist $\epsilon > 0$ and $\xi \in \mathrm{Diff}_+(M)$ such that $\xi_* \,\mathrm{vol} = \epsilon \,\mathrm{vol}'$. We consider the isomorphism

$$c_\xi : \mathcal{D}_{\mathrm{vol}} \to \mathcal{D}_{\mathrm{vol}'} \quad \text{with} \quad c_\xi(\varphi) = \xi\varphi\xi^{-1}$$

given by the conjugation by ξ. Let $C_0^\infty(M)'$ be the vector space of smooth functions on M with integral 0 with respect to vol$'$. It is easy to see that the map $\xi_* : C_0^\infty(M) \to C_0^\infty(M)'$ given by $\xi_* h := (\xi^{-1})^* h$ for $h \in C_0^\infty(M)$ is a well-defined isomorphism. Let

$$\xi_\sharp : \mathrm{Hom}(H^1(M), C_0^\infty(M)) \to \mathrm{Hom}(H^1(M), C_0^\infty(M)')$$

be the homomorphism defined by $(\xi_\sharp h)(a) = \xi_*(h(\xi^*(a)))$ for $a \in H^1(M)$ and $h \in \mathrm{Hom}(H^1(M), C_0^\infty(M))$. We also need the homomorphism

$$\xi_\sharp : \mathcal{H} \to \mathcal{H}$$

defined in the same way as before. Here we use the same symbol $\xi_\sharp$ as above.

We recall that $r : H^1(M) \to Z^1(M)$ is an injective linear section of the projection $Z^1(M) \to H^1(M)$ and so is $\xi_\sharp r$. Let f be the crossed homomorphism in Lemma 2.1 with respect to the volume form vol and r, and f' with respect to vol$'$ and $\xi_\sharp r$. Let J be the crossed homomorphism in Theorem 2.3 with respect to volume forms vol and $S = \{p, A, r, \mu(=\mu_S)\}$, and J' with respect to vol$'$ and $S' = \{p, \xi_* A, \xi_\sharp r, \xi_\sharp \mu\}$.

Lemma 8.1. *The following diagrams commute*:

$$\text{(i)} \quad \begin{tikzcd} \mathcal{D}_{\mathrm{vol}} \arrow[r, "f"] \arrow[d, "c_\xi"'] & \mathrm{Hom}(H^1(M), C_0^\infty(M)) \arrow[d, "\xi_\sharp"] \\ \mathcal{D}_{\mathrm{vol}'} \arrow[r, "f'"] & \mathrm{Hom}(H^1(M), C_0^\infty(M)') \end{tikzcd} \quad \textit{and} \quad \text{(ii)} \quad \begin{tikzcd} \mathcal{D}_{\mathrm{vol}} \arrow[r, "J"] \arrow[d, "c_\xi"'] & \mathcal{H} \arrow[d, "\xi_\sharp"] \\ \mathcal{D}_{\mathrm{vol}'} \arrow[r, "J'"] & \mathcal{H} \end{tikzcd}.$$

Proof. Let $\varphi \in \mathcal{D}_{\mathrm{vol}}$.

(i) For any $a \in H^1(M)$, we have

$$d\{(\xi_\sharp f_\varphi)(a)\} = \{\xi_\sharp \varphi_\sharp r\}(a) - (\xi_\sharp r)(a) = \{c_\xi(\varphi)_\sharp \xi_\sharp r\}(a) - (\xi_\sharp r)(a) = d\{f'_{c_\xi(\varphi)}(a)\};$$

then $(\xi_\sharp f_\varphi)(a) = f'_{c_\xi(\varphi)}(a)$ because $(\xi_\sharp f_\varphi)(a),\ f'_{c_\xi(\varphi)}(a) \in C_0^\infty(M)'$. This implies (i).

(ii) For any $a \in \bigwedge^l H^1(M)$, we can show

$$d\{\xi_\sharp \mu(a)\} = \kappa'(a)\,\mathrm{vol}' - \Delta(\xi_* A)\, p \wedge \textstyle\bigwedge_{\xi_\sharp r} a,$$

where κ' is the κ with respect to vol$'$. Thus, we can take $\xi_\sharp \mu$ as μ in Lemma 2.2 with respect to vol$'$ and $\{p, \xi_* A, \xi_\sharp r\}$. Using (i), we can find $(\xi_\sharp J_\varphi)(a) = J'_{c_\xi(\varphi)}(a)$ by direct computation. This implies that the diagram commutes. $\square$

Let $L_+, L'_+ : H^{n-1}(M) \to \mathcal{H}$ be the homomorphisms in (2-1) with respect to vol and vol$'$ respectively.

Lemma 8.2. (i) $\operatorname{Im} L_+ = \operatorname{Im} L'_+$.

(ii) $\xi_\sharp(\operatorname{Im} L_+) = \operatorname{Im} L'_+$.

Proof. Since $\xi_* \operatorname{vol} = \epsilon \operatorname{vol}'$, we have $\int_M \operatorname{vol} = \epsilon \int_M \operatorname{vol}'$ and $\kappa' = \epsilon\kappa$. Those imply $L'_+ = \epsilon L_+$ and $L'_+(\xi_* w) = \epsilon \xi_\sharp\{L_+(w)\}$ for all $w \in H^{n-1}(M)$, which show (i) and (ii). □

By Lemmas 8.1 and 8.2, we have the following commutative diagram:

$$\begin{tikzcd}
\pi_0\mathcal{D}_{\mathrm{vol}} \arrow[r, "\mathcal{J}"] \arrow[d, "c_{\xi*}"'] & \mathcal{H}/\operatorname{Im} L_+ \arrow[d, "\xi_\sharp"] \\
\pi_0\mathcal{D}_{\mathrm{vol}'} \arrow[r, "\mathcal{J}'"] & \mathcal{H}/\operatorname{Im} L_+,
\end{tikzcd}$$

where $\mathcal{J}$ and $\mathcal{J}'$ are induced from J and J' respectively. By Moser, the inclusion $\mathcal{D}_{\mathrm{vol}} \hookrightarrow \mathcal{D}$ is a weak homotopy equivalence. Hence, it induces an isomorphism $\pi_0\mathcal{D}_{\mathrm{vol}} \cong \pi_0\mathcal{D}$.

Hereafter, we assume that ξ is isotopic to the identity $\xi \simeq \mathrm{id}$, which is always possible in Moser's method. We have the following commutative diagram:

$$\begin{tikzcd}
\pi_0\mathcal{D} \arrow[r, "\cong"] \arrow[d, equal] & \pi_0\mathcal{D}_{\mathrm{vol}} \arrow[r, "\mathcal{J}"] \arrow[d, "c_{\xi*}=\mathrm{id}"'] & \mathcal{H}/\operatorname{Im} L_+ \arrow[d, "\xi_\sharp=\mathrm{id}"] \\
\pi_0\mathcal{D} \arrow[r, "\cong"] & \pi_0\mathcal{D}_{\mathrm{vol}'} \arrow[r, "\mathcal{J}'"] & \mathcal{H}/\operatorname{Im} L_+,
\end{tikzcd}$$

which implies that any $\mathcal{J}$ with respect to vol coincides with some $\mathcal{J}'$ with respect to vol'. The remaining part of Theorem 2.6 can be shown in the same way as Theorem 2.3. Thus, we complete the proof of Theorem 2.6.

9. Johnson homomorphisms

Let Σ_g be a closed oriented surface of genus $g \geqq 3$, and $H_{\mathbb{Z}} := H_1(\Sigma_g; \mathbb{Z})$ its first homology group with coefficients in $\mathbb{Z}$. Let $\mathrm{Diff}_+(\Sigma_g)$ be the group of orientation preserving diffeomorphisms of Σ_g with the C^∞ topology. The mapping class group $\mathcal{M}_g$ of Σ_g is the group of path components of $\mathrm{Diff}_+(\Sigma_g)$. The standard action of $\mathcal{M}_g$ on $H_{\mathbb{Z}}$ induces a well-known representation $\mathcal{M}_g \to \mathrm{Aut}(H_{\mathbb{Z}}, \cdot) \cong \mathrm{Sp}(2g, \mathbb{Z})$, where $\cdot$ denotes the intersection pairing on $H_{\mathbb{Z}}$. The kernel $\mathcal{I}_g$ of the representation is called the Torelli group. Take a base point $* \in \Sigma_g$ as depicted in Figure 1, and fix it. We can consider the mapping class group $\mathcal{M}_{g,*}$ of $(\Sigma_g, *)$, which is the group of path components of the subgroup $\mathrm{Diff}_+(\Sigma_g, *) \subset \mathrm{Diff}_+(\Sigma_g)$ of diffeomorphisms preserving the base point. The kernel $\mathcal{I}_{g,*}$ of the composition $\mathcal{M}_{g,*} \to \mathcal{M}_g \to \mathrm{Sp}(2g, \mathbb{Z})$ is also called the Torelli group.

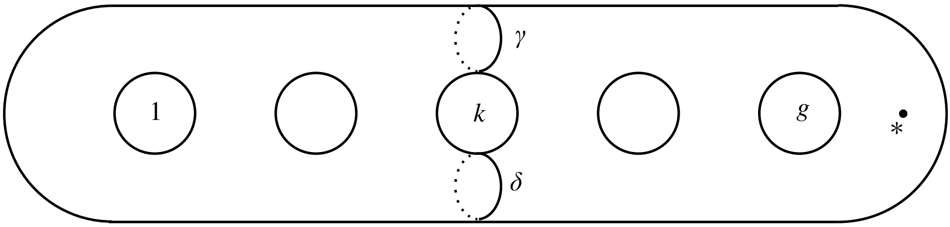


Figure 1. A base point $*$ and a bounding pair (γ, δ).

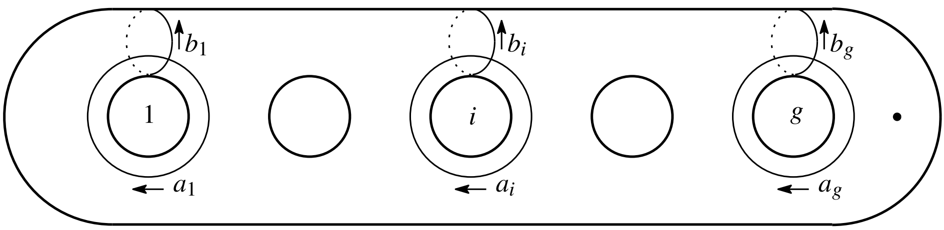


Figure 2. A basis a_i, b_i of $H_{\mathbb{Z}}$.

Let T_γ be a Dehn twist along a simple closed curve (SCC) γ in Σ_g. Let (γ, δ) be a bounding pair (BP); that is, a pair of disjoint homologous SCC's which are not homologically trivial. A BP map is given by $T_\gamma T_\delta^{-1}$. D. Johnson [1979] showed that $\mathcal{I}_{g,*}$ are generated by all BP maps. He also defined the (first) Johnson homomorphism $\tau = \tau_g : \mathcal{I}_{g,*} \to \bigwedge^3 H_{\mathbb{Z}}$. Let

$$\varphi_k = T_\gamma T_\delta^{-1} \in \mathrm{Diff}_+(\Sigma_g, *)$$

be the BP map for the BP (γ, δ) as depicted in Figure 1. Johnson [1980] calculated the value of τ at $\varphi_k \in \mathcal{M}_{g,*}$, which is $\tau(\varphi_k) = \left(\sum_{i=1}^{k-1} a_i \wedge b_i\right) \wedge b_k$. Here $\{a_i, b_i\}_{i=1}^{g}$ is the symplectic basis for $H_{\mathbb{Z}}$ as depicted in Figure 2, and the mapping class $[\varphi_k]$ of φ_k is also denoted by the same symbol φ_k. Hereafter we use this symbol for a diffeomorphism and its mapping class.

The Johnson homomorphism τ descends to $\tau : \mathcal{I}_g \to \bigwedge^3 H_{\mathbb{Z}}/H_{\mathbb{Z}}$, which is also a Johnson homomorphism, and is denoted by the same letter τ. Here, $H_{\mathbb{Z}}$ is considered a subgroup of $\bigwedge^3 H_{\mathbb{Z}}$ by the injection $H_{\mathbb{Z}} \ni x \mapsto x \wedge \left(\sum_{i=1}^{g} a_i \wedge b_i\right) \in \bigwedge^3 H_{\mathbb{Z}}$. Let $\{a_i^*, b_i^*\}$ be the dual basis of $H^1(\Sigma_g; \mathbb{Z}) = H_{\mathbb{Z}}^*$ to $\{a_i, b_i\}$. Poincaré duality gives the identification $H_{\mathbb{Z}} \cong H_{\mathbb{Z}}^*$ by $a_i \mapsto -b_i^*$, $b_i \mapsto a_i^*$. Using it, we have

$$\bigwedge^3 H_{\mathbb{Z}} \subset H_{\mathbb{Z}} \otimes \bigwedge^2 H_{\mathbb{Z}} \cong \mathrm{Hom}(H_{\mathbb{Z}}, \bigwedge^2 H_{\mathbb{Z}}).$$

The image $\tau(\varphi_k) = \left(\sum_{i=1}^{k-1} a_i \wedge b_i\right) \wedge b_k \in \bigwedge^3 H_{\mathbb{Z}}$ of φ_k by τ is given, as an element of $\mathrm{Hom}(H_{\mathbb{Z}}\bigwedge^2 H_{\mathbb{Z}})$, by

$$\text{(9-1)}\qquad \tau(\varphi_k)\colon \begin{cases} a_i \mapsto b_k \wedge a_i & \text{for } 1 \le i \le k-1, \qquad a_k \mapsto \sum_{i=1}^{k-1} a_i \wedge b_i, \\ b_i \mapsto b_k \wedge b_i & \text{for } 1 \le i \le k-1, \qquad c \mapsto 0, \end{cases}$$

where c denotes the remaining base elements, and it is given as an element of $H_{\mathbb{Z}}^* \otimes \bigwedge^2 H_{\mathbb{Z}}$ by

$$\text{(9-2)}\qquad \tau(\varphi_k) = -\sum_{i=1}^{k-1}\{a_i^* \otimes (a_i \wedge b_k) + b_i^* \otimes (b_i \wedge b_k) - a_k^* \otimes (a_i \wedge b_i)\}.$$

10. The homomorphism L_+ in the 2-dimensional case

In this section we will compute the homomorphism L_+ defined in Section 2 in the 2-dimensional case with $p = 1$.

Let $M = \Sigma_g$ be a closed oriented surface of genus $g \geqq 3$ and ω an area form with area $A = \int_{\Sigma_g} \omega$. Let $\{a_i, b_i\}$ and $\{a_i^*, b_i^*\}$ be the dual bases of $H_{\mathbb{Z}}$ and $H_{\mathbb{Z}}^*$ to each other in Section 9. Set $H = H_1(\Sigma_g)$ and $H^* = H^1(\Sigma_g)$. The bases $\{a_i, b_i\}$ and $\{a_i^*, b_i^*\}$ can also be considered bases of H and H^* respectively.

In this case, since $n = l = 2$ and $p = 1$, the homomorphism $L_+\colon H^* \to \mathrm{Hom}\big(\bigwedge^2 H^*, H^*\big)$ of (2-1) is given by

$$L_+(w) = \Big[a \mapsto \frac{1}{A}\big\{\langle c_1 \cup w, [\Sigma_g]\rangle c_2 - \langle c_2 \cup w, [\Sigma_g]\rangle c_1 - \langle c_1 \cup c_2, [\Sigma_g]\rangle w\big\}\Big]$$

for $w \in H^*$ and $c = c_1 \wedge c_2 \in \bigwedge^2 H^*$. In particular, for $w = a_l^*, b_l^* \in H^*$ with $1 \le l \le g$, we have

$$L_+(a_l^*)\colon \begin{cases} a_i^* \wedge a_j^* \mapsto 0, \\ b_i^* \wedge b_j^* \mapsto \frac{1}{A}(-\delta_{il} b_j^* + b_i^* \delta_{jl}), \\ a_i^* \wedge b_j^* \mapsto \frac{1}{A}(a_i^* \delta_{jl} - \delta_{ij} a_l^*), \end{cases}$$

and

$$L_+(b_l^*)\colon \begin{cases} a_i^* \wedge a_j^* \mapsto \frac{1}{A}(\delta_{il} a_j^* - a_i^* \delta_{jl}), \\ b_i^* \wedge b_j^* \mapsto 0, \\ a_i^* \wedge b_j^* \mapsto \frac{1}{A}(\delta_{il} b_j^* - \delta_{ij} b_l^*), \end{cases}$$

for all $1 \le i, j \le g$, where δ_{ij} denotes Kronecker's delta.

We can represent $L_+(a_l^*)$ and $L_+(b_l^*)$ as elements of $H^* \otimes \bigwedge^2 H$ as follows:

$$\text{(10-1)}\qquad L_+(a_l^*) = \frac{1}{2A}\sum_{i=1}^{g}\{a_i^* \otimes (a_i \wedge b_l) + b_i^* \otimes (b_i \wedge b_l) - a_l^* \otimes (a_i \wedge b_i)\},$$

$$\text{(10-2)}\qquad L_+(b_l^*) = \frac{1}{2A}\sum_{j=1}^{g}\{a_j^* \otimes (a_l \wedge a_j) + b_j^* \otimes (a_l \wedge b_j) - b_l^* \otimes (a_j \wedge b_j)\}.$$

We remark that under the identification $\left(\bigwedge^2 H^*\right)^* \cong \bigwedge^2 H$, the dual basis of $\bigwedge^2 H^*$ to $\{a_i^* \wedge a_j^* (i < j), b_i^* \wedge b_j^* (i < j), a_i^* \wedge b_j^* (\forall i, j)\}$ is given by

$$\{\tfrac{1}{2}a_i \wedge a_j \ (i < j), \tfrac{1}{2}b_i \wedge b_j \ (i < j), \tfrac{1}{2}a_i \wedge b_j \ (\forall i, j)\} \subset \textstyle\bigwedge^2 H$$

by our convention.

11. A section *r*

In this section we will explicitly give a section r as in Section 2, which is needed in order to define our crossed homomorphism.

Let T_i $(1 \le i \le g)$ be compact submanifolds of Σ_g — as depicted in Figure 3 — which are diffeomorphic to a 2-torus with two open disks deleted. We consider each T_i as a submanifold of $\mathbb{R}^2/(2\pi\mathbb{Z})^2$ and use the induced coordinates $(x, y) \in T_i \subset \mathbb{R}^2/(2\pi\mathbb{Z})^2$. But we mainly take $(x, y) \in (-\pi, \pi] \times (-\pi, \pi] \setminus (\operatorname{int} D^2 \amalg \operatorname{int} D^2) \subset T_i$ as depicted in Figure 4. We assume

$$(-\pi, \pi] \times [-1, 1] \cup [-1, 1] \times (-\pi, \pi] \subset (-\pi, \pi] \times (-\pi, \pi] \setminus (\operatorname{int} D^2 \amalg \operatorname{int} D^2).$$

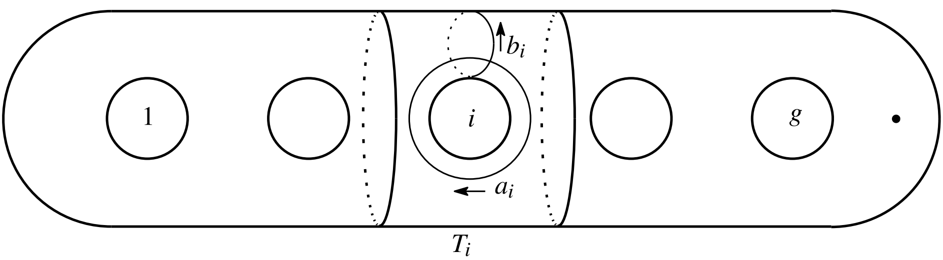


Figure 3. Compact submanifolds T_i of Σ_g.

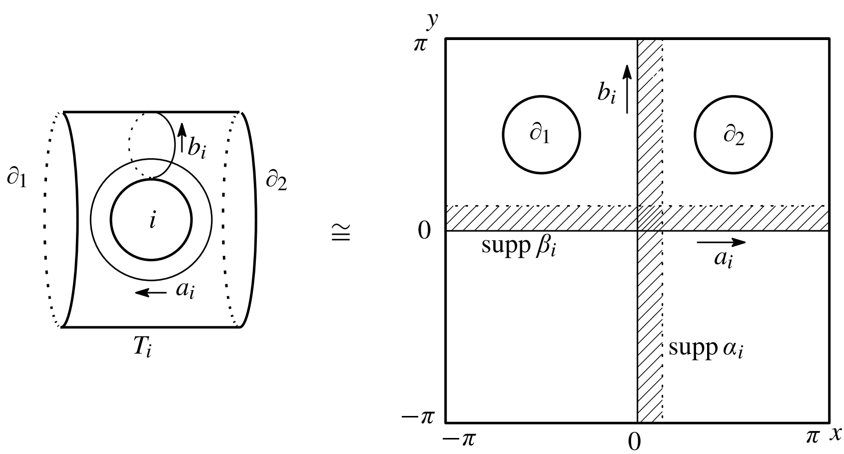


Figure 4. Coordinates of T_i.

Let $\rho\colon \mathbb{R} \to \mathbb{R}$ be a smooth function such that

$$\rho(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \text{monotone increasing} & \text{if } 0 < x < \varepsilon, \\ 1 & \text{if } \varepsilon \le x, \end{cases}$$

for a sufficiently small $\varepsilon > 0$. We define closed 1-forms $\alpha_i, \beta_i \in Z^1(\Sigma_g)$ with $1 \le i \le g$ by

$$\alpha_i(p) = \begin{cases} d\rho(x) & \text{if } p = (x, y) \in T_i, \\ 0 & \text{if } p \in \Sigma_g \setminus T_i, \end{cases} \qquad \beta_i(p) = \begin{cases} d\rho(y) & \text{if } p = (x, y) \in T_i, \\ 0 & \text{if } p \in \Sigma_g \setminus T_i, . \end{cases}$$

It is easy to see that $\{\alpha_i, \beta_i\}$ represents the basis $\{a_i^*, b_i^*\}$ of H^*.

We use the section $r\colon H^* \to Z^1(\Sigma_g)$ of the projection $Z^1(\Sigma_g) \to H^*$ defined as the linear extension of $r(a_i^*) = \alpha_i$ and $r(b_i^*) = \beta_i$ with $1 \le i \le g$.

12. BP maps φ_k

In this section, we will define BP maps φ_k with $1 < k < g$ as ω-preserving diffeomorphisms. We will compute $\mathcal{J}_{\varphi_k}$ in later sections.

Let $(x, y) \in T_k$ be the local coordinates of T_k given in Section 11. We explicitly give simple closed curves γ and δ on T_k by

$$\left\{\left(-\tfrac{5}{2}\varepsilon, y\right) \in T_k \mid y \in \mathbb{R}/2\pi\mathbb{Z}\right\}$$

and

$$\left\{\left(\pi - \tfrac{5}{2}\varepsilon, y\right) \in T_k \mid y \in \mathbb{R}/2\pi\mathbb{Z}\right\}$$

respectively. A BP map $\varphi_k\colon \Sigma_g \to \Sigma_g$ is given by

$$\varphi_k(p) = \begin{cases} (x, a(x) + y) & \text{if } p = (x, y) \in [-3\varepsilon, -2\varepsilon] \times (-\pi, \pi] \subset T_k \\ (x, -a(x - \pi) + y) & \text{if } p = (x, y) \in [\pi - 3\varepsilon, \pi - 2\varepsilon] \times (-\pi, \pi] \subset T_k \\ p & \text{otherwise,} \end{cases}$$

where $a\colon [-3\varepsilon, -2\varepsilon] \to \mathbb{R}$ is a nonincreasing smooth function satisfying

$$a(t) = \begin{cases} 0 & \text{near } t = -3\varepsilon \\ -2\pi & \text{near } t = -2\varepsilon. \end{cases}$$

See Figure 5 for the support of φ_k. It is easy to check $\varphi_k \in \mathcal{D}_\omega = \{\varphi \in \mathrm{Diff}(\Sigma_g) \mid \varphi^*\omega = \omega\}$ and

$$\varphi_k^{-1}(p) = \begin{cases} (x, -a(x) + y) & \text{if } p = (x, y) \in [-3\varepsilon, -2\varepsilon] \times (-\pi, \pi] \subset T_k, \\ (x, a(x - \pi) + y) & \text{if } p = (x, y) \in [\pi - 3\varepsilon, \pi - 2\varepsilon] \times (-\pi, \pi] \subset T_k, \\ p & \text{otherwise.} \end{cases}$$

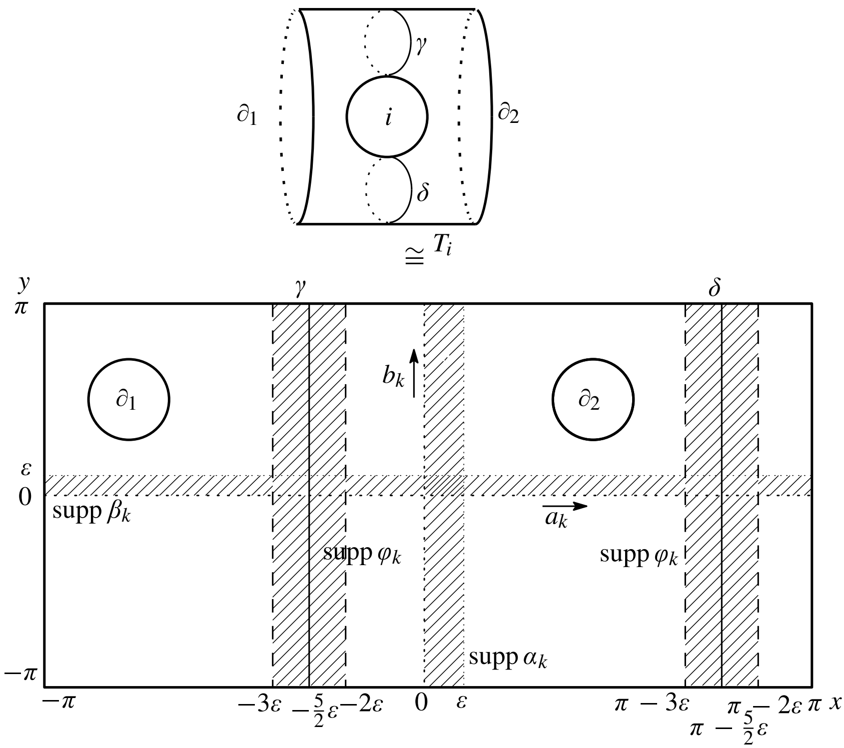


Figure 5. Simple closed curves γ and δ and $\operatorname{supp}\varphi_k$.

13. J_{φ_k}

In this section we will compute J_{φ_k} and prove Theorem 2.7 without showing some lemmas, whose proofs are given in the next section.

In the case of $(M, \mathrm{vol}) = (\Sigma_g, \omega)$, the crossed homomorphism J is written as $J : \mathcal{D}_\omega \ni \varphi \mapsto J_\varphi \in \mathcal{H} = \mathrm{Hom}\big(\bigwedge^2 H^*, H^*\big)$, where $J_\varphi(a) = [\tilde{J}_\varphi(a)] \in H^*$ and

$$\tilde{J}_\varphi(a) = f_\varphi(c_1)r(c_2) - \varphi_\sharp r(c_1)\cdot f_\varphi(c_2) + \varphi_\sharp\mu(a) - \mu(a)$$

for $a = c_1 \wedge c_2 \in \bigwedge^2 H^*$.

Hereafter, we fix k as $1 < k < g$ and write $\varphi = \varphi_k$ for simplicity.

Let $* \in \Sigma_g$ be the base point depicted in Figure 1 as before. Since $\varphi_*\beta_k - \beta_k$ is exact, we have a unique function $h \in C^\infty(\Sigma_g)$ satisfying $dh = \varphi_*\beta_k - \beta_k$ and $h(*) = 0$. Let Σ_{k-} and Σ_{k+} be the connected components of $\Sigma_g \setminus T_k$ such that $* \in \Sigma_{k+}$. The following lemma is easily checked, so its proof is omitted:

Lemma 13.1. (i) $h \equiv -1$ *on* Σ_{k-}.

(ii) $h \equiv 0$ *on* Σ_{k+}.

Set $h_0 := \int_{\Sigma_g} h\omega$.

Lemma 13.2. *Assume $\varphi = \varphi_k$ with $1 < k < g$.*

(i) $J_\varphi(a) = 0$ *for* $a = a_i^* \wedge a_j^*$ $(\forall i, j)$ *or* $a = b_i^* \wedge b_j^*$ $(i \neq k, j \neq k)$ *or* $a = a_i^* \wedge b_j^*$ $(i \neq j, i \neq k, j \neq k)$.

(ii) *For* $a = b_i^* \wedge b_k^*$ $(i \neq k)$, *we have* $J_\varphi(a) = \begin{cases} (1 + h_0/A) b_i^* & \text{if } i < k, \\ (h_0/A) b_i^* & \text{if } i > k. \end{cases}$

(iii) *For* $a = a_i^* \wedge b_k^*$ $(i \neq k)$, *we have* $J_\varphi(a) = \begin{cases} (1 + h_0/A) a_i^* & \text{if } i < k, \\ (h_0/A) a_i^* & \text{if } i > k, \end{cases}$

(iv) *For* $a = a_k^* \wedge b_j^*$ *with* $j \neq k$, *we have* $J_\varphi(a) = 0$.

(v) *For* $a = a_i^* \wedge b_i^*$, *we have* $J_\varphi(a) = \begin{cases} -(1 + h_0/A) a_k^* & \text{if } i < k, \\ 0 & \text{if } i = k, \\ -(h_0/A) a_k^* & \text{if } i > k, \end{cases}$

For a while we admit this lemma. It implies that J_φ is given by

$$\begin{aligned} J_\varphi = & \sum_{i<k} (b_i^* \wedge b_k^*)^* \otimes \Big(1 + \frac{h_0}{A}\Big) b_i^* + \sum_{i>k} (b_i^* \wedge b_k^*)^* \otimes \frac{h_0}{A} b_i^* \\ & + \sum_{i<k} (a_i^* \wedge b_k^*)^* \otimes \Big(1 + \frac{h_0}{A}\Big) a_i^* + \sum_{i>k} (a_i^* \wedge b_k^*)^* \otimes \frac{h_0}{A} a_i^* \\ & + \sum_{i<k} (a_i^* \wedge b_i^*)^* \otimes \Big(-1 - \frac{h_0}{A}\Big) a_k^* + \sum_{i>k} (a_i^* \wedge b_i^*)^* \otimes \Big(-\frac{h_0}{A}\Big) a_k^* \end{aligned}$$

as an element of $\big(\bigwedge^2 H^*\big)^* \otimes H^*$. Under the identification $\big(\bigwedge^2 H^*\big)^* \cong \bigwedge^2 H$, which is given by $(a_i^* \wedge a_j^*)^* \mapsto \frac{1}{2} a_i \wedge a_j$, $(b_i^* \wedge b_j^*)^* \mapsto \frac{1}{2} b_i \wedge b_j$, and $(a_i^* \wedge b_j^*)^* \mapsto \frac{1}{2} a_i \wedge b_j$ by our convention as remarked before, we have

$$\begin{aligned} J_\varphi = & \frac{1}{2} \sum_{i<k} \{ b_i^* \otimes (b_i \wedge b_k) + a_i^* \otimes (a_i \wedge b_k) - a_k^* \otimes (a_i \wedge b_i) \} \\ & + \frac{h_0}{2A} \sum_{i=1}^{g} \{ b_i^* \otimes (b_i \wedge b_k) + a_i^* \otimes (a_i \wedge b_k) - a_k^* \otimes (a_i \wedge b_i) \} \end{aligned}$$

as an element of $H^* \otimes \bigwedge^2 H$; therefore, by (9-2) and (10-1), we obtain

$$J_\varphi = -\tfrac{1}{2} \tau(\varphi) + h_0 L_+(a_k^*).$$

Proof of Theorem 2.7. As mentioned before, we use the same symbol φ for the mapping class $[\varphi] \in \pi_0 \mathfrak{D} \cong \pi_0 \mathfrak{D}_\omega$ of $\varphi \in \mathfrak{D}_\omega$. The computation above implies $\mathcal{J}_\varphi = -\frac{1}{2} j \circ \tau(\varphi)$ for $\varphi = \varphi_k$ $(1 < \forall k < g)$, where j is the homomorphism (2-2). Johnson [1980] showed that τ is $\pi_0 \mathfrak{D}$-equivariant, which means $\tau(\psi \varphi \psi^{-1}) = \psi_* \{\tau(\varphi)\}$ for any $\varphi \in \mathcal{I}_g$ and $\psi \in \pi_0 \mathfrak{D}$. $\mathcal{J}$ is also $\pi_0 \mathfrak{D}$-equivariant on $\mathcal{I}_g$; in fact, since $\mathcal{J}$ is a crossed homomorphism on $\pi_0 \mathfrak{D}$, we have $\mathcal{J}_{\psi\varphi\psi^{-1}} = \mathcal{J}_\psi + \psi_\sharp \mathcal{J}_\varphi - (\psi \varphi \psi^{-1})_\sharp \mathcal{J}_\psi =$

$\psi_\sharp \mathcal{J}_\varphi$, where we used $(\psi\varphi\psi^{-1})_\sharp = \mathrm{id}$ because $\psi\varphi\psi^{-1} \in \mathcal{I}_g$. Clearly j is also $\pi_0\mathfrak{D}$-equivariant, which means $j(\varphi_* t) = \varphi_\sharp\{j(t)\}$ for any $t \in \bigwedge^3 H_{\mathbb{Z}}/H_{\mathbb{Z}}$. Since all BP maps generate $\mathcal{I}_g$ by [Johnson 1979] and are conjugate to some φ_k by an element of $\pi_0\mathfrak{D}$. Thus, $\mathcal{J}$ and $-\frac{1}{2} j \circ \tau$ coincide on all BP maps. This implies $\mathcal{J} = -\frac{1}{2} j \circ \tau$ on $\mathcal{I}_g$. □

To show Lemma 13.2, note that we can retake homomorphisms r and μ for every $\varphi \in \mathcal{I}_g$ so as to compute J_φ easily; in fact, we have shown $J_{T,\varphi} - J_{S,\varphi} = \varphi_\sharp \nu - \nu$ for different choices of S and T with common p in the proofs of Theorem 2.3 and Lemma 6.5. Since $\varphi \in \mathcal{I}_g$ acts trivially on the cohomology group of Σ_g, we obtain $J_{T,\varphi} - J_{S,\varphi} = 0$. In particular, the value J_φ is independent of the choice of r and μ. So we can use r defined in Section 11, namely $r(a_i^*) = \alpha_i$ and $r(b_j^*) = \beta_j$, in the computation below. With regard to μ, for any φ_k and $a_0 \in \bigwedge^2 H^1(\Sigma_g)$ we take $\mu(a_0)$ and extend it linearly to μ on whole $\bigwedge^2 H^1(\Sigma_g)$. A connection A is not needed since we compute J with respect to $p = 1$. The computation below is carried out using such r and μ.

Recall that $f_\varphi \in \mathrm{Hom}(H^1(\Sigma_g), C_0^\infty(\Sigma_g))$ in the following lemma is the image of $\varphi = \varphi_k$ under the crossed homomorphism f in Lemma 2.1.

Lemma 13.3. *We have $f_\varphi(a_i^*) = 0$ with $1 \le i \le g$, and*

$$f_\varphi(b_j^*) = \begin{cases} 0 & \text{if } 1 \le j \le g,\ j \ne k, \\ h - h_0/A & \text{if } j = k. \end{cases}$$

Proof. Since $\mathrm{supp}(\varphi) \cap \mathrm{supp}(\alpha_i) = \varnothing$ (see Figure 5), we have $df_\varphi(a_i^*) = \varphi_*\alpha_i - \alpha_i = 0$. The condition $f_\varphi(a_i^*) \in C_0^\infty(\Sigma_g)$ implies $f_\varphi(a_i^*) = 0$. For the same reason, we have $f_\varphi(b_j^*) = 0$ for $j \ne k$. Since $df_\varphi(b_k^*) = \varphi_*\beta_k - \beta_k$, $f_\varphi(b_k^*)$ is equal to h up to a constant. The result follows from the condition $f_\varphi(b_k^*) \in C_0^\infty(\Sigma_g)$. □

Next we prove Lemma 13.2 using lemmas in Section 14, which are needed only for (v) in Lemma 13.2 and are shown there.

Proof of Lemma 13.2. (i) For $a = a_i^* \wedge a_j^*$, we have $\bigwedge_r a = \alpha_i \wedge \alpha_j = 0$ and $\kappa(a) = 0$. Hence, we can take $\mu(a) = 0$. Using Lemma 13.3, we have

$$\tilde{J}_\varphi(a) = f_\varphi(a_i^*)\alpha_j - \varphi_*\alpha_i \cdot f_\varphi(a_j^*) + \varphi_\sharp \mu(a) - \mu(a) = 0.$$

Similarly we obtain the equality for the other cases since all terms in $\tilde{J}_\varphi(a)$ are zero.

(ii) Since $\bigwedge_r a = 0$ and $\kappa(a) = 0$, we can take $\mu(a) = 0$. Using Lemma 13.3, we obtain $\tilde{J}_\varphi(a) = -(h - h_0/A)\varphi_*\beta_i$. Since $\varphi = \mathrm{id}$ and $h \equiv -1$ on $\mathrm{supp}\,\beta_i$ for $i < k$, and $h \equiv 0$ for $i > k$ by Lemma 13.1, we have (ii).

(iii) and (iv) These items are shown in the same way as (i) and (ii), so we omit the proofs.

(v) In order to compute $J_\varphi(a)$ for $a = a_i^* \wedge b_i^*$, we can take μ satisfying $\mu(a) = \mu_- + \mu_+ + \frac{1}{A}\lambda x dy - \tau_i$ by Lemma 14.2 and use the notation there.

Let $i \neq k$. By Lemma 13.3 we have

$$\tilde{J}_\varphi(a) = \varphi_*\mu(a) - \mu(a) = \frac{1}{A}\{\varphi_*(\lambda x dy) - \lambda x dy\} - (\varphi_*\tau_i - \tau_i).$$

By Lemmas 14.3 and 14.4, we obtain, for $1 \leq l, m \leq g$,

$$\int_{a_l} \tilde{J}_\varphi(a) = \begin{cases} 0 & \text{if } l \neq k, \\ -(2\pi^2 + A_+)/A & \text{if } l = k, i < k, \\ -(2\pi^2 - A_-)/A & \text{if } l = k, i > k, \end{cases}$$

$$\int_{b_m} \tilde{J}_\varphi(a) = 0.$$

This case follows from $A_+ + A_- = A$, $h_0 = 2\pi^2 - A_-$ (which is Lemma 14.5) and the fact that $\tilde{J}_\varphi(a)$ is a closed form.

Let $i = k$. We have

$$\tilde{J}_\varphi(a) = -\varphi_*\alpha_k \cdot f_\varphi(b_k^*) + \varphi_*\mu(a) - \mu(a) = \frac{h_0}{A}\alpha_k + \varphi_*\mu(a) - \mu(a),$$

where we use $\varphi = \mathrm{id}$ and $h \equiv 0$ on $\operatorname{supp}\alpha_k$. By a similar computation of $\int_c \tilde{J}_\varphi(a)$ with $c = a_l, b_m$ as above, we obtain

$$J_\varphi(a) = \frac{h_0}{A}a_k^* + \frac{-2\pi^2 - (-A_-)}{A}a_k^* = 0$$

by Lemma 14.5. This implies (v). □

14. The main part of the computation

In this section we will show the lemmas needed to prove Lemma 13.2. These are the main parts of the computation of $J_\varphi(a)$. Throughout this section, we fix the integer k as $1 < k < g$.

Let $\lambda\colon \Sigma_g \to [0, 1]$ be a smooth function with support as depicted in Figure 6 such that

$$\begin{cases} \operatorname{supp}\lambda \subset T_k \\ \partial(\operatorname{supp}\lambda) \cong \coprod^3 S^1 \\ \operatorname{supp} d\lambda \subset \text{small neighborhood of } \partial\ (\operatorname{supp}\lambda) \\ \lambda \equiv 1 \text{ on } \operatorname{supp}\lambda \setminus \operatorname{supp} d\lambda. \end{cases}$$

Since $\varepsilon > 0$ is sufficiently small, we can also assume

$$T_k' := T_k \cap \{(x, y) \mid -4\varepsilon \leq x \leq 0\} \subset \lambda^{-1}(1).$$

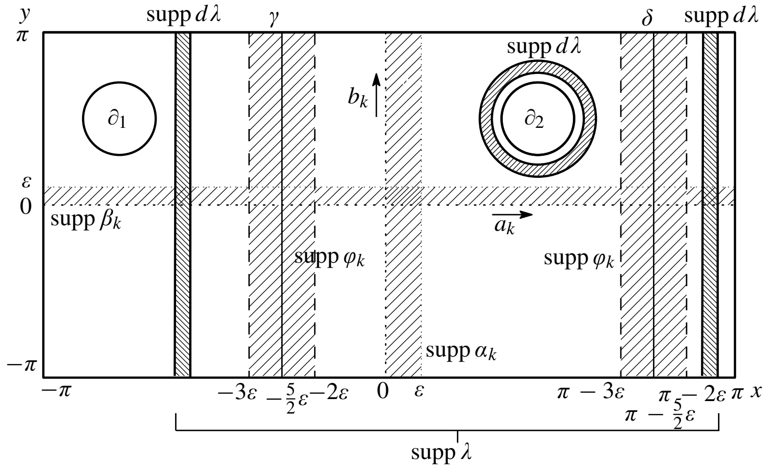


Figure 6. Support of λ.

Let $\lambda_\pm \colon \Sigma_g \to [0, 1]$ be two smooth functions uniquely defined by

$$\begin{cases} \lambda_- + \lambda + \lambda_+ = 1, \\ \operatorname{supp} \lambda_- \subset \Sigma_{k-} \cup T_k, \\ \operatorname{supp} \lambda_+ \subset \Sigma_{k+} \cup T_k, \\ \operatorname{supp} \lambda_- \cap \operatorname{supp} \lambda_+ = \varnothing. \end{cases}$$

As depicted in Figure 6, $\operatorname{supp} \lambda_+$ and $\operatorname{supp} \lambda_-$ are closed subsurfaces of Σ_g with one circle boundary and two respectively.

Hereafter, we assume $\omega|_{T_k} = dx \wedge dy$ since the crossed homomorphism $\mathscr{J}$ is independent of the choice of ω.

We consider the 2-form $d(\lambda x\, dy)$ on T_k as that on Σ_g by extending it by 0 on $\Sigma_g \setminus T_k$. Let $\omega_\pm \in \Omega^2(\Sigma_g)$ be the two closed 2-forms defined by $\operatorname{supp} \omega_\pm \subset \operatorname{supp} \lambda_\pm$ and

$$\omega_- + d(\lambda x\, dy) + \omega_+ = \omega. \tag{14-1}$$

Set $A_\pm = \int_{\Sigma_g} \omega_\pm$. We have $A_- + A_+ = A$.

Lemma 14.1. *For any i with $1 \le i \le g$ and $p, q \in \mathbb{R}$ with $p + q = 1$, there exists $\tau_i \in \Omega^1(\Sigma_g)$ satisfying*

(i) $\alpha_i \wedge \beta_i = p\alpha_1 \wedge \beta_1 + q\alpha_g \wedge \beta_g + d\tau_i$,

(ii) $\operatorname{supp} \tau_i \subset$ *the image of an embedding of a rectangle*,

(iii) $\operatorname{supp} \tau_i \cap \operatorname{supp} \varphi \subset T'_k$, *and*

(iv) $\tau_i|_{T'_k} = s\, d\rho(y)$ *with* $s = \begin{cases} q & \textit{if } i < k, \\ -p & \textit{if } i \geqq k. \end{cases}$

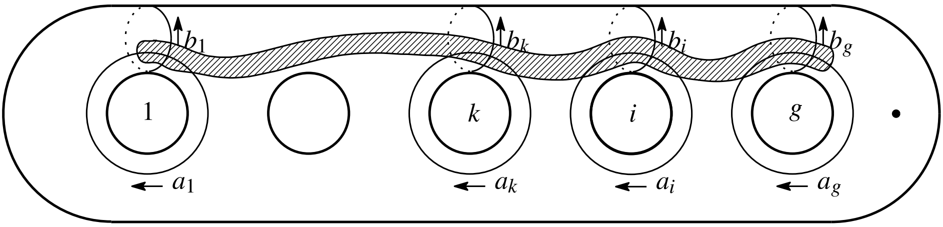


Figure 7. Support of τ_i.

Proof. Let ρ be the smooth function on $\mathbb{R}$ in Section 11. For any $t_-, t_0, t_+ \in \mathbb{R}$ satisfying $t_- + \varepsilon \le t_0 + \varepsilon \le t_+$, we define the smooth function $\rho_W : \mathbb{R} \to [0, 1]$, where $W = \{t_-, t_0, t_+, p, q\}$, by

$$\rho_W(t) = \begin{cases} 0 & \text{if } t < t_-, \\ -p\rho(t - t_0) & \text{if } t_- \le t < t_- + \varepsilon, \\ -p & \text{if } t_1 + \varepsilon \le t < t_0 \\ -p + \rho(t - t_0, & \text{if } t_0 \le t < t_0 + \varepsilon, \\ 1 - p \ (= q) & \text{if } t_0 + \varepsilon \le t < t_+, \\ q\{1 - \rho(t - t_+)\} & \text{if } t_+ \le t < t_+ + \varepsilon, \\ 0 & \text{if } t_+ + \varepsilon \le t, \end{cases}$$

for $t \in \mathbb{R}$. Let $\tilde{\tau}_W \in \Omega^1(\mathbb{R} \times [-\varepsilon, 2\varepsilon])$ be the 1-form defined by $\tilde{\tau}_W(x, y) = \rho_W(x)\, d\rho(y)$ for $(x, y) \in \mathbb{R} \times [-\varepsilon, 2\varepsilon]$. We have

$$d\tilde{\tau}_W(x, y) = \begin{cases} 0 & \text{if } x < t_-, \\ -p\, d\rho(x - t_0) \wedge d\rho(y) & \text{if } t_- \le x < t_- + \varepsilon, \\ 0 & \text{if } t_- + \varepsilon \le x < t_0, \\ d\rho(x - t_0) \wedge d\rho(y) & \text{if } t_0 \le x < t_0 + \varepsilon, \\ 0 & \text{if } t_0 + \varepsilon \le x < t_+, \\ -q\, d\rho(x - t_+) \wedge d\rho(y) & \text{if } t_+ \le x < t_+ + \varepsilon, \\ 0 & \text{if } t_+ + \varepsilon \le x. \end{cases}$$

By the definition of α_i and β_i in Section 11, we have $\alpha_i \wedge \beta_i = d\rho(x) \wedge d\rho(y)$ on the local coordinates $(x, y) \in T_i$ for each $1 \le i \le g$. So, we can appropriately choose $t_+, t_0, t_- \in \mathbb{R}$ and an embedding $[t_-, t_+ + \varepsilon] \times [-\varepsilon, 2\varepsilon] \hookrightarrow \Sigma_g$ as depicted in Figure 7 such that the extension τ_i of $\tilde{\tau}_W$ by 0 on the complement of the image of the embedding satisfies the required properties. $\square$

We apply Lemma 14.1 for $p = A_-/A$ and $q = A_+/A$, we then have $\tau_i \in \Omega^1(\Sigma_g)$. Let $a = a_i^* \wedge b_i^*$; then $\kappa(a) = 1/A$. Using (14-1) and (i) in Lemma 14.1, we obtain

$$\kappa(a)\omega - \alpha_i \wedge \beta_i = \frac{1}{A}\omega_- - p\alpha_1 \wedge \beta_1 + \frac{1}{A}\omega_+ - q\alpha_g \wedge \beta_g + d\left\{\frac{1}{A}\lambda x dy - \tau_i\right\}.$$

Since

$$\operatorname{supp}\Big(\frac{1}{A}\omega_- - p\alpha_1\wedge\beta_1\Big)\subset\operatorname{supp}\lambda_-,\quad \operatorname{supp}\Big(\frac{1}{A}\omega_+ - q\alpha_g\wedge\beta_g\Big)\subset\operatorname{supp}\lambda_+,$$

and

$$\int_{\operatorname{supp}\lambda_-}\Big(\frac{1}{A}\omega_- - p\alpha_1\wedge\beta_1\Big)=\int_{\operatorname{supp}\lambda_+}\Big(\frac{1}{A}\omega_+ - q\alpha_g\wedge\beta_g\Big)=0,$$

there exist $\mu_-,\mu_+\in\Omega^1(\Sigma_g)$ such that $\operatorname{supp}\mu_-\subset\operatorname{supp}\lambda_-$, $\operatorname{supp}\mu_+\subset\lambda_+$, $d\mu_-=\frac{1}{A}\omega_- - p\alpha_1\wedge\beta_1$ and $d\mu_+=\frac{1}{A}\omega_+ - q\alpha_g\wedge\beta_g$. Thus, we have

$$\kappa(a)\omega-\alpha_i\wedge\beta_i=d\Big(\mu_-+\mu_++\frac{1}{A}\lambda x dy-\tau_i\Big).$$

Hence, we can take

$$\mu(a)=\mu_-+\mu_++\frac{1}{A}\lambda x\,dy-\tau_i. \tag{14-2}$$

Thus, we have:

Lemma 14.2. *In the situation above, for every $\varphi=\varphi_k$ and $a=a_i^*\wedge b_i^*$, there exists a homomorphism $\mu\colon\bigwedge^2H^1(\Sigma_g)\to\Omega^1(\Sigma_g)$ in Lemma 2.2 satisfying (14-2).*

Lemma 14.3. (i) $\int_{a_j}\{\varphi_*(\lambda x\,dy)-\lambda x\,dy\}=\begin{cases}0 & \text{if } j\neq k,\\ -2\pi^2 & \text{if } j=k,\end{cases}$

(ii) $\int_{b_j}\{\varphi_*(\lambda x\,dy)-\lambda x\,dy\}=0$ *for all* j.

Proof. Since $\operatorname{supp}\{\varphi_*(\lambda xdy)-\lambda xdy\}\subset\operatorname{supp}\varphi$ is disjoint from a_j $(j\neq k)$ and b_j for all j, the integrals along them are equal to 0. So, we have only to compute the integral along a_k. We recall that λ is equal to 1 on $\operatorname{supp}\varphi$. On T_k, we have

$$\begin{aligned}\varphi_*(\lambda x\,dy)-\lambda x\,dy&=(\varphi^{-1})^*(x\,dy)-x\,dy\\&=\begin{cases}-xa'(x)\,dx & (x,y)\in[-3\varepsilon,-2\varepsilon]\times(-\pi,\pi]\\ xa'(x-\pi)\,dx & (x,y)\in[\pi-3\varepsilon,\pi-2\varepsilon]\times(-\pi,\pi]\\ 0 & \text{otherwise;}\end{cases}\end{aligned}$$

hence,

$$\begin{aligned}\int_{a_k}\{\varphi_*(\lambda x\,dy)-\lambda x\,dy\}&=\int_{-3\varepsilon}^{-2\varepsilon}\{-xa'(x)\}\,dx+\int_{\pi-3\varepsilon}^{\pi-2\varepsilon}xa'(x-\pi)\,dx\\&=\pi\{a(-2\varepsilon)-a(-3\varepsilon)\}=-2\pi^2.\end{aligned}$$ □

Lemma 14.4. (i) $\displaystyle\int_{a_j}(\varphi_*\tau_i-\tau_i)=\begin{cases}0 & \text{if } j\neq k,\\ q=A_+/A & \text{if } j=k,\ i<k,\\ -p=-A_-/A & \text{if } j=k,\ i\geqq k.\end{cases}$

(ii) $\displaystyle\int_{b_j}(\varphi_*\tau_i-\tau_i)=0$ *for all* j.

Proof. We have only to compute the integral along a_k since the others are clear by considering the support of the integrand. Let $\tilde{\rho}$ be the smooth function on $\mathbb{R}$ with $\tilde{\rho}(0)=0$ whose differential $d\tilde{\rho}$ agrees with the pullback of $d\rho$ by the projection $\mathbb{R}\to\mathbb{R}/2\pi\mathbb{Z}$. By Lemma 14.1, on T_k' we have

$$\begin{aligned}\varphi_*\tau_i-\tau_i&=(\varphi^{-1})^*\{s\,d\rho(y)\}-s\,d\rho(y)\\&=-s\tilde{\rho}'(-a(x)+y)a'(x)\,dx+s\{\tilde{\rho}'(-a(x)+y)-\tilde{\rho}'(y)\}\,dy\end{aligned}$$

for $(x,y)\in[-3\varepsilon,-2\varepsilon]\times(-\pi,\pi]$, and $\varphi_*\tau_i-\tau_i=0$ otherwise. Since

$$\operatorname{supp}(\varphi_*\tau_i-\tau_i)\subset T_k',$$

we have

$$\int_{a_k}(\varphi_*\tau_i-\tau_i)=-s\int_{-3\varepsilon}^{-2\varepsilon}\tilde{\rho}'(-a(x)+0)a'(x)\,dx=s[\tilde{\rho}(-a(x))]_{-3\varepsilon}^{-2\varepsilon}=s.\qquad\square$$

Finally, we prove the following lemma:

Lemma 14.5. $$h_0=2\pi^2-A_-.$$

Proof. Since $\operatorname{supp}h\cap\operatorname{supp}\lambda_+=\varnothing$ and $h\equiv-1$ on $\operatorname{supp}\lambda_-$, we have

$$h=(\lambda_-+\lambda+\lambda_+)h=-\lambda_-+\lambda h$$

and

$$h_0=\int_{\Sigma_g}h\omega=-\int_{\Sigma_g}\lambda_-\omega+\int_{\Sigma_g}\lambda h\omega. \tag{14-3}$$

Since $dh=\varphi_*\beta_k-\beta_k$ by definition and $\omega|_{T_k}=dx\wedge dy$ by assumption, we obtain

$$d(\lambda hx\,dy)=d\lambda\wedge hx\,dy+\lambda(\varphi_*\beta_k-\beta_k)\wedge x\,dy+\lambda h\omega.$$

Then we have

$$\int_{\Sigma_g}\lambda h\omega=-\int_{\Sigma_g}d\lambda_-\wedge x\,dy-\int_{\Sigma_g}(\varphi_*\beta_k-\beta_k)\wedge x\,dy \tag{14-4}$$

by Stokes' formula and $h\equiv-1$ on $\operatorname{supp}d\lambda\cap\operatorname{supp}h\subset\operatorname{supp}d\lambda_-$.

On the other hand, since $\operatorname{supp}\lambda\subset T_k$, we have

$$\lambda\omega=d(\lambda x\,dy)-d\lambda\wedge x\,dy=d\lambda_-\wedge x\,dy+d(\lambda x\,dy)+d\lambda_+\wedge x\,dy$$

and

$$\omega=\lambda_-\omega+d\lambda_-\wedge x\,dy+d(\lambda x\,dy)+d\lambda_+\wedge x\,dy+\lambda_+\omega.$$

By considering supports, we get $\omega_-=\lambda_-\omega+d\lambda_-\wedge x\,dy$ and

$$-\int_{\Sigma_g}\lambda_-\omega=-\int_{\Sigma_g}\omega_-+\int_{\Sigma_g}d\lambda_-\wedge x\,dy. \tag{14-5}$$

Summing up equalities (14-3), (14-4), and (14-5), we have

$$h_0 = -\int_{\Sigma_g} \omega_- - \int_{\Sigma_g} (\varphi_*\beta_k - \beta_k) \wedge x\,dy.$$

Since $\beta_k = d\rho(y) = d\tilde{\rho}(y)$ on T_k, we have

$$(\varphi_*\beta_k - \beta_k) \wedge x\,dy = d[\varphi_*\{\tilde{\rho(y)}\}] \wedge x\,dy =$$
$$\begin{cases} -\tilde{\rho}'(-a(x)+y)a'(x)\,dx \wedge x\,dy & \text{if } (x,y) \in [-3\varepsilon, -2\varepsilon] \times (-\pi, \pi], \\ \tilde{\rho}'(a(x-\pi)+y)a'(x-\pi)\,dx \wedge x\,dy & \text{if } (x,y) \in [\pi-3\varepsilon, \pi-2\varepsilon] \times (-\pi, \pi], \\ 0 & \text{otherwise,} \end{cases}$$

and then we obtain

$$\begin{aligned} \int_{\Sigma_g} (\varphi_*\beta_k - \beta_k) \wedge x\,dy &= -\int_{-3\varepsilon}^{-2\varepsilon} a'(x)x\,dx + \int_{\pi-3\varepsilon}^{\pi-2\varepsilon} a'(x-\pi)x\,dx \\ &= \pi \int_{-3\varepsilon}^{-2\varepsilon} a'(x)\,dx = \pi[a(x)]_{-3\varepsilon}^{-2\varepsilon} = -2\pi^2. \end{aligned}$$

Since $\int_{\Sigma_g} \omega_- = A_-$, we have $h_0 = -A_- + 2\pi^2$. □

References

[Banyaga 1997] A. Banyaga, *The structure of classical diffeomorphism groups*, Mathematics and its Applications **400**, Kluwer, Dordrecht, 1997. MR 98h:22024 Zbl 0874.58005

[Brown 1982] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1982. MR 83k:20002 Zbl 0584.20036

[Day 2007] M. B. Day, "Extending Johnson's and Morita's homomorphisms to the mapping class group", *Algebr. Geom. Topol.* **7** (2007), 1297–1326. MR 2008j:57033 Zbl 1181.57025

[Day 2011] M. B. Day, "Extended flux maps on surfaces and the contracted Johnson homomorphism", *J. Symplectic Geom.* **9**:4 (2011), 445–482. MR 2900787 Zbl 06049656

[Hatcher 1978] A. E. Hatcher, "Concordance spaces, higher simple-homotopy theory, and applications", pp. 3–21 in *Algebraic and geometric topology, I* (Stanford, CA, 1976), edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, R.I., 1978. MR 80f:57014 Zbl 0406.57031

[Hsiang and Sharpe 1976] W. C. Hsiang and R. W. Sharpe, "Parametrized surgery and isotopy", *Pacific J. Math.* **67**:2 (1976), 401–459. MR 58 #13091 Zbl 0348.57015

[Johnson 1979] D. L. Johnson, "Homeomorphisms of a surface which act trivially on homology", *Proc. Amer. Math. Soc.* **75**:1 (1979), 119–125. MR 80h:57008 Zbl 0407.57003

[Johnson 1980] D. Johnson, "An abelian quotient of the mapping class group $\mathcal{I}_g$", *Math. Ann.* **249**:3 (1980), 225–242. MR 82a:57008 Zbl 0409.57009

[Kasagawa 2008] R. Kasagawa, "On crossed homomorphisms on symplectic mapping class groups", *Trans. Amer. Math. Soc.* **360**:9 (2008), 4815–4839. MR 2009f:57040 Zbl 1158.53056

[Kędra et al. 2006] J. Kędra, D. Kotschick, and S. Morita, "Crossed flux homomorphisms and vanishing theorems for flux groups", *Geom. Funct. Anal.* **16**:6 (2006), 1246–1273. MR 2008c:53091 Zbl 1114.57031

[Kotschick and Morita 2005] D. Kotschick and S. Morita, "Signatures of foliated surface bundles and the symplectomorphism groups of surfaces", *Topology* **44**:1 (2005), 131–149. MR 2005k:57050 Zbl 1163.57304

[Kotschick and Morita 2007] D. Kotschick and S. Morita, "Characteristic classes of foliated surface bundles with area-preserving holonomy", *J. Diff. Geom.* **75**:2 (2007), 273–302. MR 2008k:57046 Zbl 1115.57016

[McDuff 1983] D. McDuff, "Some canonical cohomology classes on groups of volume preserving diffeomorphisms", *Trans. Amer. Math. Soc.* **275**:1 (1983), 345–356. MR 84h:58154 Zbl 0522.57029

[Milnor 2011] J. Milnor, "Differential topology forty-six years later", *Notices Amer. Math. Soc.* **58**:6 (2011), 804–809. MR 2012i:57001 Zbl 1225.01040

[Morita 1993] S. Morita, "The extension of Johnson's homomorphism from the Torelli group to the mapping class group", *Invent. Math.* **111**:1 (1993), 197–224. MR 93j:57001 Zbl 0787.57008

[Morita and Penner 2008] S. Morita and R. C. Penner, "Torelli groups, extended Johnson homomorphisms, and new cycles on the moduli space of curves", *Math. Proc. Cambridge Philos. Soc.* **144**:3 (2008), 651–671. MR 2009c:32028 Zbl 1157.32015

[Moser 1965] J. Moser, "On the volume elements on a manifold", *Trans. Amer. Math. Soc.* **120** (1965), 286–294. MR 32 #409 Zbl 0141.19407

[Perron 2004] B. Perron, "Homomorphic extensions of Johnson homomorphisms via Fox calculus", *Ann. Inst. Fourier* (*Grenoble*) **54**:4 (2004), 1073–1106. MR 2005k:57034 Zbl 1109.57013

[Vaisman 1987] I. Vaisman, *Symplectic geometry and secondary characteristic classes*, Progress in Mathematics **72**, Birkhäuser, Boston, MA, 1987. MR 89f:58062 Zbl 0629.53002

Received June 7, 2012. Revised January 22, 2013.

RYOJI KASAGAWA
DEPARTMENT OF MATHEMATICS, COLLEGE OF SCIENCE AND TECHNOLOGY
NIHON UNIVERSITY
1-8 KANDA, SURUGADAI, CHIYODA-KU
TOKYO 101-8308
JAPAN

kasagawa@math.cst.nihon-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.491

REGULARITY AT THE BOUNDARY AND TANGENTIAL REGULARITY OF SOLUTIONS OF THE CAUCHY–RIEMANN SYSTEM

TRAN VU KHANH AND GIUSEPPE ZAMPIERI

For a pseudoconvex domain $D \subset \mathbb{C}^n$, we prove the equivalence of the local hypoellipticity of the system $(\bar\partial, \bar\partial^*)$ with the system $(\bar\partial_b, \bar\partial_b^*)$ induced at the boundary. This develops a former result of ours in which the theory of harmonic extension by Kohn was used. This technique is inadequate for the purpose of the present paper and must be replaced by that of the holomorphic extension.

Let D be a pseudoconvex domain of $\mathbb{C}^n$ defined by $r < 0$ with C^∞ boundary bD. We use the standard notation $\Box = \bar\partial\bar\partial^* + \bar\partial^*\bar\partial$ for the complex Laplacian, $Q(u, u) = \|\bar\partial u\|^2 + \|\bar\partial^* u\|^2$ for the energy form, and some variants such as $Q_{\mathrm{Op}}(u, u) = \|\mathrm{Op}\,\bar\partial u\|^2 + \|\mathrm{Op}\,\bar\partial^* u\|^2$ for an operator Op. Here u is a $(0, k)$ form belonging to $D_{\bar\partial^*}$. We similarly define the tangential versions as $\Box_b$, $\bar\partial_b$, $\bar\partial_b^*$, and Q^b_{Op}. We take local coordinates (x, r) in $\mathbb{C}^n$, with $x \in \mathbb{R}^{2n-1}$ being the tangential coordinates and r, the equation of bD, serving as the last coordinate. We define the tangential s-Sobolev norm by $|||u|||_s := \|\Lambda^s u\|_0$, where Λ^s is the standard tangential pseudodifferential operator with symbol $\Lambda^s_\xi = (1 + |\xi|^2)^{s/2}$. We note that

$$\begin{cases} \|\bar\partial u\|_s^2 + \|\bar\partial^* u\|_s^2 = \sum_{j\le s} Q_{\Lambda^{s-j}\partial_r^j}(u, u), \\ |||\bar\partial u|||_s^2 + |||\bar\partial^* u|||_s^2 = Q_{\Lambda^s}(u, u), \\ \|\bar\partial_b u_b\|_s^2 + \|\bar\partial_b^* u_b\|_s^2 = Q^b_{\Lambda^s}(u_b, u_b). \end{cases} \tag{1-1}$$

We decompose u into a tangential and normal component; that is,

$$u = u^\tau + u^\nu,$$

and further decompose into microlocal components (see [Kohn 2002])

$$u^\tau = u^{\tau+} + u^{\tau-} + u^{\tau 0}.$$

We similarly decompose u_b as $u_b^+ + u_b^- + u_b^0$. We use the notation $\bar L_n$ for the normal $(0, 1)$-vector field and $\bar L_1, \dots, \bar L_{n-1}$ for the tangential ones. Therefore we have the

MSC2010: 32F10.
Keywords: $\bar\partial$-Neumann problem, tangential $\bar\partial$ system.

description for the totally real tangential and normal vector fields, denoted by T and ∂_r respectively:

$$\begin{cases} T = i(L_n - \bar{L}_n), \\ \partial_r = L_n + \bar{L}_n. \end{cases}$$

From this, we get back $\bar{L}_n = \frac{1}{2}(\partial_r + iT)$. We denote the symbol of a (pseudo)-differential operator by σ and the partial tangential Fourier transform of u by $\tilde{u}$. We define a holomorphic extension (see [Khanh and Zampieri 2011]) $u^{\tau+(H)}$ of $u^{\tau+}|_{bD}$ by

$$u^{\tau+(H)} = (2\pi)^{-2n+1} \int_{\mathbb{R}^{2n-1}} e^{ix\xi} e^{r\sigma(\dot{T})} \psi^+(\xi) \tilde{u}(\xi, 0)\, d\xi, \tag{1-2}$$

where $\dot{T} := T(x, 0)$. Note that $\sigma(T) \gtrsim (1 + |\xi|^2)^{\frac{1}{2}}$ for ξ in $\operatorname{supp}\psi^+$ and (x, r) in a local patch; thus in the integral the exponential is dominated by $e^{-|r|(1+|\xi|^2)^{1/2}}$ for $r < 0$. Differently from the harmonic extension by Kohn, the present one is well defined only in positive microlocalization. We can think of $u^{\tau+(H)}$ in two different ways: as a modification of $u^{\tau+}$, or as an extension of u_b^+. The property which motivates the terminology of holomorphic extension is

$$\|\bar{L}_n u^{\tau+(H)}\| = \|r \operatorname{Tan} u^{\tau+(H)}\| \le \|u_b^{\tau+}\|_{-\frac{1}{2}}. \tag{1-3}$$

This follows from the relationships $\bar{L}_n = \frac{1}{2}(\partial_r + iT)$ and $T - \dot{T} = r \operatorname{Tan}$. We have our first relationship between a trace v_b and the general extension v ([Kohn 2002] p. 241); for any ϵ and suitable c_ϵ,

$$\|v_b\|_s \lesssim c_\epsilon |||v|||_{s+\frac{1}{2}} + \epsilon |||\partial_r v|||_{s-\frac{1}{2}}. \tag{1-4}$$

This is also seen in [Khanh and Zampieri 2011] as the small/large constant argument. As a specific property of our extension we have the reciprocal relation to (1-4):

$$\|r^k u^{\tau+(H)}\|_s \lesssim \|u_b^+\|_{s-k-\frac{1}{2}}. \tag{1-5}$$

This is readily checked; see [Khanh and Zampieri 2011, (1.12)].

A combination of (1-3) and (1-4) shows that $\bar{L}_n$ acts on $u^{\tau+(H)}$ as an operator of order 0. On the other hand, on the straightening of $b\Omega$ in which $r = x_n$, we have that $J\partial_r$ — i.e., T — coincides with ∂_{y_n}, and therefore $\bar{L}_n$ is the Cauchy–Riemann operator $\partial_{\bar{z}_n}$. A reference to the related literature is in order. The extension of generalized functions to half-spaces or wedges of $\mathbb{C}^n$ using the decomposition of the δ-function in plane waves as in (1-2) was introduced by Sato, Kashiwara, and Kawai in [Sato et al. 1973] as a general method for microlocal decomposition of the singularities. It has been used, among others, by Boutet de Monvel and Sjöstrand [1976] and by Hsiao [2010] in the study of the singularities of Szegő and Bergman kernels.

We denote by the symbol $\bar\partial^\tau$ the extension of the $\bar\partial_b$ from $b\Omega$ to Ω, which stays tangential to the level surfaces $r \equiv \text{const}$. It acts on tangential forms u^τ and its action is $\bar\partial^\tau u^\tau = (\bar\partial u^\tau)^\tau$. We denote its adjoint by $\bar\partial^{\tau\,*}$; thus $\bar\partial^{\tau\,*} u^\tau = \bar\partial^*(u^\tau)$. We use the notations $\Box^\tau$ and Q^τ for the corresponding Laplacian and energy forms. We notice that

$$Q(u^{\tau+(H)}, u^{\tau+(H)}) = Q^\tau(u^{\tau+(H)}, u^{\tau+(H)}) + \|\bar L_n u^{\tau+(H)}\|_0^2. \tag{1-6}$$

We have to describe how (1-4) and (1-5) are affected by $\bar\partial$ and $\bar\partial^*$.

Proposition 1.1. *We have for any extension v of v_b that*

$$Q^b(v_b, v_b) \lesssim Q^\tau_{\Lambda^{\frac12}}(v, v) + Q^\tau_{\partial_r \Lambda^{-\frac12}}(v, v), \tag{1-7}$$

and specifically for $u^{\tau+(H)}$,

$$Q^\tau(u^{\tau+(H)}, u^{\tau+(H)}) \lesssim Q^b_{\Lambda^{-\frac12}}(u_b^+, u_b^+) + \|u_b^+\|^2_{-\frac12}. \tag{1-8}$$

Proof. We have

$$\bar\partial^\tau v|_{bD} = \bar\partial_b v_b, \qquad \bar\partial^{\tau *} v|_{bD} = \bar\partial_b^* v_b.$$

Then, (1-7) follows from (1-4).

We proceed to prove (1-8). We have $\bar\partial^\tau = \bar\partial_b + r\,\mathrm{Tan}$, and $\bar\partial^{\tau *} = \bar\partial_b^* + r\,\mathrm{Tan}$, which yields

$$\begin{aligned} \bar\partial^\tau u^{\tau+(H)} &= (\bar\partial_b u_b)^{\tau+(H)} + r\,\mathrm{Tan}\; u^{\tau+(H)}, \\ \bar\partial^{\tau *} u^{\tau+(H)} &= (\bar\partial_b^* u_b)^{\tau+(H)} + r\,\mathrm{Tan}\; u^{\tau+(H)}. \end{aligned} \tag{1-9}$$

Application of (1-5) yields

$$\begin{aligned} \|\bar\partial^\tau u^{\tau+(H)}\|^2 + \|\bar\partial^{\tau *} u^{\tau+(H)}\|^2 & \\ &= \|(\bar\partial_b u_b)^{\tau+(H)}\|^2 + \|(\bar\partial_b^* u_b)^{\tau+(H)}\|^2 + \|r\,\mathrm{Tan}\; u^{\tau+(H)}\|^2 \\ &\lesssim \|\bar\partial_b u_b^+\|^2_{-\frac12} + \|\bar\partial_b^* u_b^+\|^2_{-\frac12} + \|u_b^+\|^2_{-\frac12}. \end{aligned}$$ □

We decompose $u^{\tau+}$ as $u^{\tau+(H)} + u^{\tau+(0)}$, which also serves as a definition of $u^{\tau+(0)}$. Let ζ and ζ' be cut-offs with $\zeta \prec \zeta'$ in the sense that $\zeta'|_{\mathrm{supp}\,\zeta} \equiv 1$.

Proposition 1.2. *Each of the forms $u^\# = u^\nu, u^{\tau\,-}, u^{\tau\,0}, u^{\tau+(0)}, u_b^-$, and u_b^0 enjoy elliptic estimates*; *that is*,

$$\|\zeta u^\#\|_s \lesssim \|\zeta' \bar\partial u^\#\|_{s-1} + \|\zeta' \bar\partial^* u^\#\|_{s-1} + \|u^\#\|_0, \quad s \geq 2. \tag{1-10}$$

Proof. Estimate (1-10) follows by iteration from

$$\|\zeta u^\#\|_s \lesssim \|\zeta\, \bar\partial u^\#\|_{s-1} + \|\zeta\, \bar\partial^* u^\#\|_{s-1} + \|\zeta' u^\#\|_{s-1}. \tag{1-11}$$

As for u^ν and $u^{\tau+(0)}$, this latter follows from $u^\nu|_{bD}\equiv 0$ and $u^{\tau+(0)}|_{bD}\equiv 0$. For the terms with $-$ and 0, this follows from the fact that $|\sigma(T)|\lesssim|\sigma(\bar\partial)|$ in the region of 0-microlocalization, and from $\sigma[\bar\partial,\bar\partial^*]\le 0$ and $\sigma(T)<0$ in the negative microlocalization. We refer to (1) in the Main Theorem of [Folland and Kohn 1972] as a general reference, but also give an outline of the proof. We start from

$$|||\zeta u^\#|||_1^2\lesssim Q\big(\zeta u^\#,\zeta u^\#\big)+\|\zeta' u^\#\|_0^2; \tag{1-12}$$

this is the basic estimate in the case of u^ν and $u^{\tau+(0)}$ (which vanish at bD), and it is Lemma 8.6 of [Kohn 2002] for $u^{\tau\,-}$, $u^{\tau\,0}$ and u_b^-, u_b^0. Applying (1-12) to $\zeta\Lambda^{s-1}\zeta u^\#$ one gets the estimate of tangential norms for any s; that is, (1-11) with the usual norm replaced by the triplet norm. Finally, by noncharacteristicity of $(\bar\partial,\bar\partial^*)$, one passes from tangential to full norms along the guidelines of [Zampieri 2008, Theorem 1.9.7]. The version of this argument for $\Box$ can be found in [Kohn 2002, second part of p. 245]. $\square$

Let s and l be indices.

Theorem 1.3. *Consider the estimates*

$$\|\zeta u_b\|_s\lesssim\|\zeta'\bar\partial_b u_b\|_{s+l}+\|\zeta'\bar\partial_b^* u_b\|_{s+l}+\|u_b\|_0\quad \textit{for any } u_b\in C^\infty(b\Omega), \tag{1-13}$$

$$\|\zeta u\|_s\lesssim\|\zeta'\bar\partial u\|_{s+l}+\|\zeta'\bar\partial^* u\|_{s+l}+\|u\|_0\quad \textit{for any } u\in D_{\bar\partial^*}\cap C^\infty(\bar\Omega), \tag{1-14}$$

$$\|\zeta u\|_s\le\epsilon(\|\zeta\bar\partial u\|_s+\|\zeta\bar\partial^* u\|_s)+c_\epsilon\|u\| \tag{1-15}$$

$$\textit{for any } \epsilon, \textit{ for suitable } c_\epsilon, \textit{ and for any } u\in D_{\bar\partial^*}\cap C^\infty(\bar\Omega).$$

Then (1-13) *implies* (1-14) *and* (1-15) *implies* (1-13) *for* $l=0$.

Remark 1.4. (i) The above estimates (1-13) and (1-14) for any s, ζ, ζ' and for suitable l, characterize the local hypoellipticity of the system $(\bar\partial_b,\bar\partial_b^*)$ and $(\bar\partial,\bar\partial^*)$ respectively (see [Kohn 2005]). When $l>0$, one says that the system has a loss of l derivatives; when $l<0$, one says that it has a gain of $-l$ derivatives.

(ii) The point in (1-15), as opposed to (1-13) and (1-14), is that we have the same cut-off ζ in both sides, and also that there is a factor ϵ of compactness. Though (1-15) is stronger than (1-14), there are wide classes of domains Ω for which it holds, including all domains of infraexponential type, for which a superlogarithmic estimate holds (see [Baracco et al. 2014]). Indeed, let R^s be the pseudodifferential operator defined by $\widetilde{R^s u}=\Lambda_\xi^{s\sigma(x)}\tilde u$ (see [Kohn 2002, p. 234]). On one hand, we have $R^s\sim\Lambda^s$ modulo operators of order $-\infty$ over u such that $\sigma|_{\operatorname{supp} u}\equiv 1$. On the other, we have that $[R^s,\zeta']$ has order $-\infty$ if $\zeta'|_{\operatorname{supp}\sigma}\equiv 1$ and hence the supports of σ and $\dot\zeta'$ are disjoint. Finally, we have

$$|\zeta''[\bar\partial,R^s]\zeta'|\lesssim\log\Lambda R^s\zeta'$$

in the sense of operators when $\sigma \prec \zeta' \prec \zeta''$. Using R^s as a substitute for Λ^s, we can prove (1-15) whenever a superlogarithmic estimate holds (see [Kohn 2002, § 7]).

Proof. First, it is clearly not restrictive that u and u_b have compact support. Because of Proposition 1.2, it suffices to prove (1-13) for u_b^+ and (1-14) for $u^{\tau+}$. It is also obvious that we can consider cut-off functions ζ and ζ' only in tangential coordinates, not in r. We start by proving that (1-13) implies (1-14). We recall the decomposition

$$u^{\tau+} = u^{\tau+(H)} + u^{\tau+(0)}$$

and begin by estimating $u^{\tau+(H)}$. We then have

$$\begin{aligned} (1\text{-}16) \quad |||\zeta u^{\tau+(H)}|||_s^2 &\underset{(1\text{-}5)}{\lesssim} \|\zeta u_b^+\|_{s-\frac{1}{2}}^2 \\ &\underset{(1\text{-}13)}{\lesssim} Q^b_{\Lambda^{s+l-\frac{1}{2}}\zeta'}(u_b^+, u_b^+) + \|u_b^+\|_{-\frac{1}{2}}^2 \\ &\underset{(1\text{-}7)}{\lesssim} Q^\tau_{\Lambda^{s+l}\zeta'}(u^{\tau+}, u^{\tau+}) + Q^\tau_{\partial_r\Lambda^{s+l-1}\zeta'}(u^{\tau+}, u^{\tau+}) + \|u^{\tau+}\|_0^2. \end{aligned}$$

It remains to estimate $u^{\tau+(0)}$. Since $u^{\tau+(0)}|_{bD} \equiv 0$, then by 1-elliptic estimates

$$\begin{aligned} (1\text{-}17) \quad & |||\zeta u^{\tau+(0)}|||_s^2 \\ &\underset{(1\text{-}11)}{\lesssim} Q_{\Lambda^{s-1}\zeta}(u^{\tau+(0)}, u^{\tau+(0)}) + |||\zeta' u^{\tau+(0)}|||_{s-1}^2 \\ &\lesssim Q_{\Lambda^{s-1}\zeta}(u^{\tau+}, u^{\tau+}) + Q^\tau_{\Lambda^{s-1}\zeta}(u^{\tau+(H)}, u^{\tau+(H)}) + |||r\zeta u^{\tau+(H)}|||_s^2 + |||\zeta' u^{\tau+(0)}|||_{s-1}^2 \\ &\lesssim Q_{\Lambda^{s-1}\zeta}(u^{\tau+}, u^{\tau+}) + |||\zeta u^{\tau+(H)}|||_s^2 + |||\zeta' u^{\tau+(H)}|||_{s-1}^2 + |||\zeta' u^{\tau+(0)}|||_{s-1}^2, \end{aligned}$$

where we have used $Q = Q^\tau + O(r)\Lambda$ over $h^{\tau+(H)}$; that is, (1-6) in addition to (1-3) in the second inequality, together with the estimate

$$Q^\tau_{\Lambda^{s-1}} \lesssim \Lambda^s$$

in the third. We estimate terms in the last line. First, the term $|||\zeta u^{\tau+(H)}|||_s^2$ is estimated by means of (1-16). Next, the terms in $(s-1)$-norm can be brought to 0-norm by combined inductive use of (1-16) and (1-17), and eventually their sum is controlled by $\|u^{\tau+}\|_0^2$. We put together (1-16) and (1-17) (with the above further reductions), recall the first part of (1-1) in order to estimate $Q^\tau_{\Lambda^{s+l}\zeta'} + Q^\tau_{\partial_r\Lambda^{s+l-1}\zeta'}$ in the right side of (1-16), and end up with

$$(1\text{-}18) \qquad |||\zeta u^{\tau+}|||_s \lesssim \|\zeta'\bar\partial u^{\tau+}\|_{s+l} + \|\zeta'\bar\partial^* u^{\tau+}\|_{s+l} + \|u^{\tau+}\|_0.$$

Finally, by noncharacteristicity of $(\bar\partial, \bar\partial^*)$, one passes from tangential to full norms in the left side of (1-18) along the guidelines of [Zampieri 2008, Theorem 1.9.7]. The version of this argument for $\Box$ can be found in [Kohn 2002] in the second part of p. 245. Thus we get (1-14).

We prove that (1-15) implies (1-13) for $l=0$. Thanks to $\partial_r=\bar L_n+\text{Tan}$ and to (1-3), we have

$$\partial_r u^{\tau+(H)}=\text{Tan}\,u^{\tau+(H)}\quad\text{and}\quad \bar L_n u^{\tau+(H)}=r\,\text{Tan}\,u^{\tau+(H)}.$$

It follows that

$$\begin{aligned}
&\|\zeta u_b^+\|_s^2\\
&\underset{(1\text{-}4)}{\lesssim} |||\zeta u^{\tau+(H)}|||^2_{s+\frac12}+|||\partial_r\zeta u^{\tau+(H)}|||^2_{s-\frac12}\\
&\lesssim |||\zeta u^{\tau+(H)}|||^2_{s+\frac12}+\|\bar L_n\zeta u^{\tau+(H)}\|^2_{s-\frac12}\\
&\underset{(1\text{-}15)}{\lesssim} \epsilon\big(Q^\tau_{\Lambda^{s+\frac12}\zeta}(u^{\tau+(H)},u^{\tau+(H)})+|||\zeta\bar L_n u^{\tau+(H)}|||^2_{s+\frac12}\big)\\
&\qquad\qquad+c_\epsilon\big(|||\zeta' u^{\tau+(H)}|||^2_{s-\frac12}+|||u^{\tau+(H)}|||^2_0\big)\\
&\underset{(1\text{-}8)}{\lesssim} \epsilon\big(Q^b_{\Lambda^s\zeta}(u_b^+,u_b^+)+\|\zeta u_b^+\|_s^2\big)+c_\epsilon\big(Q^b_{\Lambda^{s-1}\zeta'}(u_b^+,u_b^+)+\|\zeta' u_b^+\|^2_{s-1}+\|u_b^{\tau+}\|^2_{-\frac12}\big)\\
&\lesssim Q^b_{\Lambda^s\zeta'}(u_b^+,u_b^+)+\epsilon\|\zeta u_b^+\|_s^2+c_\epsilon\big(\|\zeta' u_b^+\|^2_{s-1}+\|u_b^{\tau+}\|^2_{-\frac12}\big),
\end{aligned}\tag{1-19}$$

where in the second-to-last line we have calculated $[\zeta,\#^{(H)}]$, which yields

$$|||\zeta u^{\tau+(H)}|||_{s+\frac12}\lesssim\|\zeta u_b^+\|_s+\|\zeta' u_b^+\|_{s-1}$$

(and similarly for $[\zeta,Q^{(H)}]$). We absorb the term with ϵ and get (1-13). $\square$

Since on a pseudoconvex domain the H^0-ranges of $\Box$ and $\Box_b$ are closed by basic estimates and by [Kohn 1986] respectively, then there are well defined H^0-inverses denoted by N and G, and named the Neumann and Green operators.

Remark 1.5. Equations (1-13) and (1-14) imply local regularity in degree ≥ 2 of G and N respectively. We first prove regularity for N. We start by remarking that

$$\begin{aligned}
&\bar\partial^*N_q \text{ is regular over Ker}\,\bar\partial\quad \text{if } q\geq 2,\\
&\bar\partial N_q \text{ is regular over Ker}\,\bar\partial^*\quad \text{if } q\geq 0.
\end{aligned}\tag{1-20}$$

In the first case, we set $u=\bar\partial^*Nf$ for $f\in\text{Ker}\,\bar\partial$. We have $(\bar\partial u=f,\ \bar\partial^*u=0)$, and hence by (1-14)

$$\|\zeta u\|_s\lesssim\|\zeta' f\|_{s+l}+\|u\|_0.$$

To prove the second case, we simply set $u=\bar\partial Nf$ for $f\in\text{Ker}\,\bar\partial^*$ and reason likewise. It follows from (1-20) that the Bergman projection B_q is regular in any degree $q\geq 0$. (Notice that even if one started from exact regularity by assuming (1-15), this is perhaps lost by taking the additional $\bar\partial$ in $B:=\text{Id}-\bar\partial^*N\bar\partial$.) Finally, we exploit formula (5.36) in [Straube 2010] in unweighted norms; that is, for $t = 0$:

$$N_q = B_q(N_q\bar\partial)(\mathrm{Id}-B_{q-1})(\bar\partial^* N_q)B_q + (\mathrm{Id}-B_q)(\bar\partial^* N_{q+1})B_{q+1}(N_{q+1}\bar\partial)(\mathrm{Id}-B_q). \tag{1-21}$$

Now, in the right side, the $\bar\partial N$'s and $\bar\partial^* N$'s are evaluated over $\operatorname{Ker}\bar\partial^*$ and $\operatorname{Ker}\bar\partial$ respectively; thus they are regular for $q \geq 2$. The B's are also regular and therefore such is N. This concludes the proof of the regularity of N. The proof of the regularity of G is similar, apart from replacing (1-21) by its version for the Green operator G stated in Section 5 of [Khanh 2010].

References

[Baracco et al. 2014] L. Baracco, T. V. Khanh, and G. Zampieri, "Hypoellipticity of the $\bar\partial_b$-Neumann problem at exponentially degenerate points", *Asian J. Math.* (2014), to appear.

[Folland and Kohn 1972] G. B. Folland and J. J. Kohn, *The Neumann problem for the Cauchy–Riemann complex*, Annals of Mathematics Studies **75**, Princeton University Press, 1972. MR 57 #1573 Zbl 0247.35093

[Hsiao 2010] C.-Y. Hsiao, *Projections in several complex variables*, Mém. Soc. Math. Fr. (N.S.) **123**, Société Mathématique de France, Paris, 2010. MR 2011m:32004 Zbl 1229.32002

[Khanh 2010] T. V. Khanh, "Global hypoellipticity of the Kohn–Laplacian $\Box_b$ on pseudoconvex CR manifolds", preprint, 2010. arXiv 1012.5906

[Khanh and Zampieri 2011] T. V. Khanh and G. Zampieri, "Estimates for regularity of the tangential $\bar\partial$-system", *Math. Nachr.* **284**:17-18 (2011), 2212–2224. MR 2859760 Zbl 1229.32013

[Kohn 1986] J. J. Kohn, "The range of the tangential Cauchy–Riemann operator", *Duke Math. J.* **53**:2 (1986), 525–545. MR 87m:32041 Zbl 0609.32015

[Kohn 2002] J. J. Kohn, "Superlogarithmic estimates on pseudoconvex domains and CR manifolds", *Ann. of Math.* (2) **156**:1 (2002), 213–248. MR 2003i:32059 Zbl 1037.32032

[Kohn 2005] J. J. Kohn, "Hypoellipticity and loss of derivatives", *Ann. of Math.* (2) **162**:2 (2005), 943–986. MR 2006k:35036 Zbl 1107.35044

[Boutet de Monvel and Sjöstrand 1976] L. Boutet de Monvel and J. Sjöstrand, "Sur la singularité des noyaux de Bergman et de Szego", pp. 123–164 in *Journées: Équations aux Dérivées Partielles de Rennes* (Rennes, 1975), edited by J. Camus, Astérisque **34–35**, Société Mathématique de France, Paris, 1976. MR 58 #28684 Zbl 0344.32010

[Sato et al. 1973] M. Sato, T. Kawai, and M. Kashiwara, "Microfunctions and pseudo-differential equations", pp. 265–529 in *Hyperfunctions and pseudo-differential equations* (Katata, 1971), edited by H. Komatsu, Lecture Notes in Math. **287**, Springer, Berlin, 1973. MR 54 #8747 Zbl 0277.46039

[Straube 2010] E. J. Straube, *Lectures on the $\mathscr{L}^2$-Sobolev theory of the $\bar\partial$-Neumann problem*, ESI Lectures in Mathematics and Physics **7**, European Mathematical Society (EMS), Zurich, 2010. MR 2011b:35004 Zbl 1247.32003

[Zampieri 2008] G. Zampieri, *Complex analysis and CR geometry*, University Lecture Series **43**, American Mathematical Society, Providence, RI, 2008. MR 2009f:32052 Zbl 1160.32001

Received August 15, 2012. Revised November 26, 2012.

TRAN VU KHANH
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF PADOVA
VIA TRIESTE 63,
I-35121 PADOVA
ITALY
khanh@math.unipd.it

GIUSEPPE ZAMPIERI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF PADOVA
VIA TRIESTE, 63
I-35121 PADOVA
ITALY
zampieri@math.unipd.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 265, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.265.499

ON THE STEINBERG CHARACTER OF A SEMISIMPLE p-ADIC GROUP

JU-LEE KIM AND GEORGE LUSZTIG

Dedicated to Robert Steinberg on the occasion of his ninetieth birthday.

We show that the character of the Steinberg representation of a split semisimple p-adic group at a very regular element is given (up to sign) by a power of q, the number of elements in the residue field. We also show that (under an assumption on the characteristic) the character of an Iwahori-spherical representation at a split very regular element is given by a trace in the corresponding Hecke algebra module.

1. Introduction

1.1. Let K be a nonarchimedean local field and let $\underline{K}$ be a maximal unramified field extension of K. Let $\mathcal{O}$ be the ring of integers of K and let $\mathfrak{p}$ be the maximal ideal of $\mathcal{O}$; the counterparts for $\underline{K}$ are denoted by $\underline{\mathcal{O}}$ and $\underline{\mathfrak{p}}$. Let $\underline{K}^* = \underline{K} - \{0\}$. We write $\mathcal{O}/\mathfrak{p} = F_q$, a finite field with q elements of characteristic p.

Let G be a semisimple almost simple algebraic group defined and split over K with a given $\mathcal{O}$-structure compatible with the K-structure.

If V is an admissible representation of $G(K)$ of finite length, we denote by ϕ_V the character of V in the sense of Harish-Chandra, viewed as a $\mathbb{C}$-valued function on the set $G(K)_{rs} := G_{rs} \cap G(K)$. (Here, G_{rs} is the set of regular semisimple elements of G, and $\mathbb{C}$ is the field of complex numbers.)

In this paper we study the restriction of the function ϕ_V to:

(a) a certain subset $G(K)_{vr}$ of $G(K)_{rs}$, namely, the set of very regular elements in $G(K)$ (see 1.2) in the case where V is the Steinberg representation of $G(K)$, and

(b) a certain subset $G(K)_{svr}$ of $G(K)_{vr}$, namely, the set of split very regular elements in $G(K)$ (see 1.2) in the case where V is an irreducible admissible representation of $G(K)$ with nonzero vectors fixed by an Iwahori subgroup.

Both authors are supported in part by the National Science Foundation.
MSC2010: 20G99.
Keywords: p-adic group, character, unipotent representation.

In case (a), we show that $\phi_V(g)$ with $g \in G(K)_{rs}$ is of the form $\pm q^n$ with $n \in \{0, -1, -2, \dots\}$ (see Corollary 3.4) with more precise information when $g \in G(K)_{svr}$ (see Theorem 2.2) or when $g \in G(K)_{cvr}$ (see Theorem 3.2). In case (b) we show (with some restriction on characteristic) that $\phi_V(g)$ with $g \in G(K)_{svr}$ can be expressed as the trace of a certain element of an affine Hecke algebra on an irreducible module (see Theorem 4.3).

Note that the Steinberg representation **S** is an irreducible admissible representation of $G(K)$ with a one-dimensional subspace invariant under an Iwahori subgroup on which the corresponding affine Hecke algebra acts through the "sign" representation; see [Matsumoto 1969; Shalika 1970]. This is a p-adic analogue of the Steinberg representation [Steinberg 1951] of a reductive group over F_q. In [Rodier 1986], it is proven that $\phi_{\mathbf{S}}(g) \neq 0$ for any $g \in G(K)_{rs}$.

1.2. Let $g \in G_{rs} \cap G(\underline{K})$. Let $T' = T'_g$ be the maximal torus of G that contains g. We say that g is *very regular* if T' is split over $\underline{K}$ and for any root α with respect to T' viewed as a homomorphism $T'(\underline{K}) \to \underline{K}^*$ we have $\alpha(g) \notin (1 + \underline{\mathfrak{p}})$. If, in addition, $\alpha(g) \in \underline{\mathcal{O}}$, we say that g is *compact very regular*.

Let $G(\underline{K})_{vr}$ be the set of elements in $G(\underline{K})$ that are very regular, and $G(\underline{K})_{cvr}$ the set of compact very regular ones. We write $G(K)_{vr} = G(\underline{K})_{vr} \cap G(K)$ and $G(K)_{cvr} = G(\underline{K})_{cvr} \cap G(K)$. Let $G(K)_{svr}$ be the set of all $g \in G(K)_{vr}$ such that T'_g is split over K.

1.3. ***Notation.*** Let $K^* = K - \{0\}$, and let $v : K^* \to \mathbb{Z}$ be the unique (surjective) homomorphism such that $v(\mathfrak{p}^n - \mathfrak{p}^{n+1}) = n$ for any $n \in \mathbb{N}$. For $a \in K^*$ we set $|a| = q^{-v(a)}$.

We fix a maximal torus T of G defined and split over K. Let Y (resp. X) be the group of cocharacters (resp. characters) of the algebraic group T. Let $\langle\ ,\ \rangle : Y \times X \to \mathbb{Z}$ be the obvious pairing. Let $R \subset X$ be the set of roots of G with respect to T, let R^+ be a set of positive roots for R, and let Π be the set of simple roots of R determined by R^+. We write $\Pi = \{\alpha_i : i \in I\}$. Let $R^- = R - R^+$. Let Y^+ (resp. Y^{++}) be the set of all $y \in Y$ such that $\langle y, \alpha\rangle \geq 0$ (resp. $\langle y, \alpha\rangle > 0$) for all $\alpha \in R^+$. We define $2\rho \in X$ by $2\rho = \sum_{a \in R^+} \alpha$.

We have canonically $T(K) = K^* \otimes Y$; we define a homomorphism $\chi : T(K) \to Y$ by $\chi(\lambda \otimes y) = v(\lambda) y$ for any $\lambda \in K^*, y \in Y$. For any $y \in Y$, we set $T(K)_y = \chi^{-1}(y)$. For $y \in Y$, let $T(K)^{\spadesuit}_y = T(K)_y \cap G(K)_{svr}$. Note that if $y \in Y^{++}$ then $T(K)^{\spadesuit}_y = T(K)_y$.

For each $\alpha \in R$ let U_α be the corresponding root subgroup of G.

2. Calculation of $\phi_{\mathbf{S}}$ on $G(K)_{svr}$

2.1. Let $\mathcal{W} \subset \mathrm{Aut}(T)$ be the Weyl group of G regarded as a Coxeter group; for $i \in I$, let s_i be the simple reflection in $\mathcal{W}$ determined by α_i. We can also view

$\mathcal{W}$ as a subgroup of $\mathrm{Aut}(Y)$ or $\mathrm{Aut}(X)$. Let $w = w_0$ be the longest element of $\mathcal{W}$. For any $J \subset I$, let $\mathcal{W}_J$ be the subgroup of $\mathcal{W}$ generated by $\{s_i : i \in J\}$ and let $R_J = R \cap \sum_{i \in J} \mathbb{Z}\alpha_i$. Let

$$R_J^+ = R_J \cap R^+ \quad \text{and} \quad R_J^- = R_J - R_J^+.$$

Let $\mathfrak{g}$ be the Lie algebra of G, and let $\mathfrak{t} \subset \mathfrak{g}$ be the Lie algebra of T. For any $J \subset I$, let $\mathfrak{l}_J$ be the Lie subalgebra of $\mathfrak{g}$ spanned by $\mathfrak{t}$ and the root spaces corresponding to the roots in R_J. Let $\mathfrak{n}_J$ be the Lie subalgebra of $\mathfrak{g}$ spanned by the root spaces corresponding to roots in $R^+ - R_J^+$.

According to [Casselman 1973], $\phi_{\mathbf{S}}$ is an alternating sum of characters of representations induced from one-dimensional representations of various parabolic subgroups of G defined over K. From this, one can deduce that if $t \in T(K) \cap G(K)_{rs}$ then

$$\phi_{\mathbf{S}}(t) = \sum_{J \subset I} (-1)^{\sharp J} \sum_{w \in {}^J\mathcal{W}} \delta_J(w(t))^{1/2} D_{I,J}(w(t))^{-1/2},$$

where for any $J \subset I$ and $t' \in T(K) \cap G(K)_{rs}$ we set

$$D_{I,J}(t') = \left|\det(1 - \mathrm{Ad}(t')|_{\mathfrak{g}/\mathfrak{l}_J})\right|,$$
$$\delta_J(t') = \left|\det(\mathrm{Ad}(t')|_{\mathfrak{n}_J})\right|,$$

and ${}^J\mathcal{W}$ is the set of representatives of minimal length for the cosets $\mathcal{W}_J \backslash \mathcal{W}$. Here for a real number $a \geq 0$ we denote by $a^{1/2}$ or $\sqrt{a}$ the nonnegative square root of a. Writing ϕ instead of $\phi_{\mathbf{S}}$, we have:

Theorem 2.2. *Let $y \in Y^+$ and let $t \in T(K)_y^\spadesuit$. Then $\phi(t) = q^{-\langle y, 2\rho \rangle}$.*

2.3. More generally, let $t \in T(K)_y^\spadesuit$, where $y \in Y$. By a standard property of Weyl chambers, there exists $w \in \mathcal{W}$ such that $w(y) \in Y^+$. Let $t_1 = w(t)$. Then the theorem is applicable to t_1, and we have $\phi(t) = \phi(t_1) = q^{-\langle w(y), 2\rho \rangle}$.

2.4. Let $y' = w_0(y)$, $t' = w_0(t)$. We have $\phi_{\mathbf{S}}(t) = \phi_{\mathbf{S}}(t')$, $t' \in T(K)_{y'}^\spadesuit$, $-y' \in Y^+$. We show that

$$v(1 - \beta(t')) = \begin{cases} v(\beta(t')) & \text{if } \beta \in R^+, \\ 0 & \text{if } \beta \in R^-. \end{cases} \tag{1}$$

Assume first that $\beta \in R^+$. If $v(\beta(t')) \neq 0$ then $v(\beta(t')) < 0$ (since $\langle y', \beta \rangle \neq 0$ and $\langle y', \beta \rangle \leq 0$); hence, $v(1 - \beta(t')) = v(\beta(t'))$. If $v(\beta(t')) = 0$ then $\beta(t') - 1 \in \mathcal{O} - \mathfrak{p}$; hence, $v(1 - \beta(t')) = 0 = v(\beta(t'))$ as required.

Assume next that $\beta \in R^-$. If $v(\beta(t')) \neq 0$ then $v(\beta(t')) > 0$ (since $\langle y', \beta \rangle \neq 0$ and $\langle y', \beta \rangle \geq 0$); hence, $v(1 - \beta(t')) = 0$. If $v(\beta(t')) = 0$ then $\beta(t') - 1 \in \mathcal{O} - \mathfrak{p}$; hence, $v(1 - \beta(t')) = 0$ as required.

For any $w \in \mathcal{W}$, $J \subset I$ we have

$$D_{I,J}(w(t')) = \prod_{\alpha \in R - R_J} q^{-v(1-\alpha(w(t')))} = \prod_{\substack{\alpha \in R - R_J \\ w^{-1}\alpha \in R^+}} q^{-v(\alpha(w(t')))} = \prod_{\substack{\alpha \in R - R_J \\ w^{-1}\alpha \in R^+}} q^{-\langle y', w^{-1}\alpha \rangle}$$

and

$$\delta_J(w(t')) = \prod_{\alpha \in R^+ - R_J^+} q^{-v(\alpha(w(t')))} = \prod_{\alpha \in R^+ - R_J^+} q^{-\langle y', w^{-1}\alpha \rangle}.$$

(We have used (1) with $\beta = w^{-1}(\alpha)$.) We see that

$$\phi(t) = \phi(t') = \sum_{J \subset I} (-1)^{\sharp J} \sum_{w \in {}^J\mathcal{W}} \sqrt{q}^{-\langle y', x_{w,J} \rangle},$$

where for $w \in {}^J\mathcal{W}$ we have

$$\begin{aligned} x_{w,J} &= \sum_{\alpha \in R^+ - R_J^+} w^{-1}\alpha - \sum_{\substack{\alpha \in R - R_J \\ w^{-1}\alpha \in R^+}} w^{-1}\alpha \\ &= \sum_{\substack{\alpha \in R^+ - R_J^+ \\ w^{-1}(\alpha) \in R^-}} w^{-1}\alpha - \sum_{\substack{\alpha \in R^- - R_J^- \\ w^{-1}(\alpha) \in R^+}} w^{-1}\alpha \\ &= 2 \sum_{\substack{\alpha \in R^+ - R_J^+ \\ w^{-1}\alpha \in R^-}} w^{-1}\alpha \in X. \end{aligned}$$

For $w \in {}^J\mathcal{W}$, we have $\alpha \in R_J^+ \implies w^{-1}\alpha \in R^+$; hence,

$$\sum_{\substack{\alpha \in R^+ - R_J^+ \\ w^{-1}\alpha \in R^-}} w^{-1}\alpha = \sum_{\substack{\alpha \in R^+ \\ w^{-1}\alpha \in R^-}} w^{-1}\alpha,$$

so that $x_{w,J} = x_w$, where

$$x_w = 2 \sum_{\substack{\alpha \in R^+ \\ w^{-1}\alpha \in R^-}} w^{-1}\alpha \in X.$$

Thus, we have

$$\phi(t) = \sum_{J \subset I} (-1)^{\sharp J} \sum_{w \in {}^J\mathcal{W}} \sqrt{q}^{-\langle y', x_w \rangle} = \sum_{w \in \mathcal{W}} c_w \sqrt{q}^{-\langle y', x_w \rangle},$$

where for $w \in \mathcal{W}$ we set

$$c_w = \sum_{\substack{J \subset I \\ w \in {}^J\mathcal{W}}} (-1)^{\sharp J}.$$

For $w \in \mathcal{W}$, let $\mathcal{L}(w) = \{i \in I : s_i w > w\}$, where $>$ refers to the standard partial order on $\mathcal{W}$. For $J \subset I$, we have $w \in {}^J\mathcal{W}$ if and only if $J \subset \mathcal{L}(w)$; thus,

$$c_w = \sum_{J \subset \mathcal{L}(w)} (-1)^{\sharp J},$$

and this is 0 unless $\mathcal{L}(w) = \varnothing$ (that is $w = w_0$), in which case $c_w = 1$. Note also that $x_{w_0} = -4\rho$; thus, we have

$$\phi(t) = c_{w_0} \sqrt{q}^{-\langle y', x_{w_0} \rangle} = q^{\langle y', 2\rho \rangle} = q^{-\langle y, 2\rho \rangle}.$$

Theorem 2.2 is proved. □

2.5. Assume now that $\tau \in T(K)$ satisfies the following condition: for any $\alpha \in R$ we have $\alpha(\tau) - 1 \in \mathfrak{p} - \{0\}$ so that $\alpha(\tau) - 1 \in \mathfrak{p}^{n_\alpha} - \mathfrak{p}^{n_\alpha + 1}$ for a well defined integer $n_\alpha \geq 1$. Note that $n_{-\alpha} = n_\alpha$ and $v(1 - \alpha(\tau)) = n_\alpha \geq 1$ for all $\alpha \in R$; hence,

$$\phi(\tau) = \sum_{J \subset I} (-1)^{\sharp J} \sum_{w \in {}^J\mathcal{W}} q^{\sum_{\alpha \in R} n_\alpha/2 - \sum_{\alpha \in R_J} n_{w^{-1}(\alpha)}/2}.$$

Thus,

$$\phi(\tau) = \sharp(\mathcal{W}) q^{\sum_{\alpha \in R} n_\alpha/2} + \text{strictly smaller powers of } q. \tag{2}$$

In the case where K is the field of power series over F_q, the leading term in (2) is equal to $\sharp(\mathcal{W}) q^m$, where m is the dimension of the "variety" of Iwahori subgroups of $G(\underline{K})$ that contain the topologically unipotent element τ (see [Kazhdan and Lusztig 1988]).

3. Calculation of $\phi_{\mathbf{S}}$ on $G(K)_{vr}$

3.1. We will again write ϕ instead of $\phi_{\mathbf{S}}$. In this section we assume that we are given $\gamma \in G(K)_{vr}$. Let $T' = T'_\gamma$. Note that T' is defined over K; let A' be the largest K-split torus of T'. For any parabolic subgroup P of G defined over K such that $\gamma \in P$, we set $\delta_P(\gamma) = \left|\det(\mathrm{Ad}(\gamma)|_{\mathfrak{n}})\right|$, where $\mathfrak{n}$ is the Lie algebra of the unipotent radical of P.

Let $\mathcal{X}$ be the set of all pairs (P, A), where P is a parabolic subgroup of G defined over K and A is the unique maximal K-split torus in the center of some Levi subgroup of P defined over K. Then that Levi subgroup is uniquely determined by A and is denoted by M_A. Let $\mathcal{X}' = \{(P, A) \in \mathcal{X} : A \subset A'\}$. According to [Harish-Chandra 1973], we have

$$\phi(\gamma) = (-1)^{\dim T} \sum_{(P,A) \in \mathcal{X}'} (-1)^{\dim A} \delta_P(\gamma)^{1/2} D_{G/M_A}(\gamma)^{-1/2}, \tag{3}$$

where $D_{G/M_A}(\gamma) = \left|\det(1 - \mathrm{Ad}(\gamma)|_{\mathfrak{g}/\mathfrak{l}})\right|$ (we denote by $\mathfrak{l}$ the Lie algebra of M_A).

Theorem 3.2. *Assume in addition that $\gamma \in G(K)_{cvr}$. Then $\phi(\gamma) = (-1)^{\dim T - \dim A'}$.*

Proof. From our assumptions we see that $\delta_P(\gamma)=1=D_{G/M_A}(\gamma)$ for all $(P, A)\in\mathcal{X}'$; hence, (3) becomes

$$\phi(\gamma)=(-1)^{\dim T}\sum_{(P,A)\in\mathcal{X}'}(-1)^{\dim A}.$$

Let $\mathcal{Y}$ be the group of cocharacters of A' and let $\mathfrak{H}=\mathcal{Y}\otimes\mathbb{R}$. The real vector space $\mathfrak{H}$ can be partitioned into facets $F_{P,A}$ indexed by $(P, A)\in\mathcal{X}'$ such that $F_{P,A}$ is homeomorphic to $\mathbb{R}^{\dim A}$. Note that the Euler characteristic with compact support of $F_{P,A}$ is $(-1)^{\dim A}$, and the Euler characteristic with compact support of $\mathfrak{H}$ is $(-1)^{\dim_{\mathbb{R}}\mathfrak{H}}=(-1)^{\dim A'}$. Using the additivity of the Euler characteristic with compact support we see that $\sum_{(P,A)\in\mathcal{X}'}(-1)^{\dim A}=(-1)^{\dim A'}$; thus, $\phi(\gamma)=(-1)^{\dim T-\dim A'}$, as required. □

3.3. In the setup of 3.1, let P_γ be the parabolic subgroup of G associated to γ as in [Casselman 1977]. Note that P_γ is defined over K. The following result can be deduced by combining Theorem 3.2 with the results in [Casselman 1977] and with Proposition 2 in [Rodier 1986].

Corollary 3.4. *We have* $\phi(\gamma)=(-1)^{\dim T-\dim A'}\delta_{P_\gamma}(\gamma)$.

The corollary provides another proof of Theorem 2.2.

4. Iwahori spherical representations: split elements

4.1. Let B be the subgroup of $G(K)$ generated by

$$\{U_\alpha(\mathcal{O}):\alpha\in R^+\}\cup\{U_\alpha(\mathfrak{p}):\alpha\in R^-\}\cup T(K)_0.$$

(The subgroups $U_\alpha(\mathcal{O})$, $U_\alpha(\mathfrak{p})$ of U_α are defined by the $\mathcal{O}$-structure of G.) Then B is an Iwahori subgroup of $G(K)$. For any $\alpha\in R$ we choose an isomorphism $x_\alpha: K\xrightarrow{\sim}U_\alpha(K)$ (the restriction of an isomorphism of algebraic groups from the additive group to U_α), which carries $\mathcal{O}$ onto $U_\alpha(\mathcal{O})$ and $\mathfrak{p}$ onto $U_\alpha(\mathfrak{p})$. We set $W:=Y\cdot\mathcal{W}$ with Y normal in W (recall that $\mathcal{W}$ acts naturally on Y). Let Y' be the subgroup of Y generated by the coroots. Then $W':=Y'\cdot\mathcal{W}$ is naturally a subgroup of W. According to [Iwahori and Matsumoto 1965], W is an extended Coxeter group (the semidirect product of the Coxeter group W' with the finite abelian group Y/Y') with length function

$$l(yw)=\sum_{\substack{\alpha\in R^+\\ w^{-1}(\alpha)\in R^+}}\|\langle y,\alpha\rangle\| + \sum_{\substack{\alpha\in R^+\\ w^{-1}(\alpha)\in R^-}}\|\langle y,\alpha\rangle-1\|,$$

where $\|a\|=a$ if $a\geq 0$ and $\|a\|=-a$ if $a<0$. From the same reference we know that the set of double cosets $B\backslash G(K)/B$ is in bijection with W; to yw (where $y\in Y$, $w\in\mathcal{W}$) corresponds the double coset Ω_{yw} containing $T(K)_y\dot{w}$ (here $\dot{w}$ is an element in $G(\mathcal{O})$ which normalizes $T(K)_0$ and acts on it in the same way as w); moreover, $\sharp(\Omega_{yw}/B) = \sharp(B\backslash\Omega_{yw}) = q^{l(yw)}$ for any $y \in Y$, $w \in \mathcal{W}$. For example, if $y \in Y^{++}$ then $l(y) = \langle y, 2\rho\rangle$.

Let H be the algebra of B-biinvariant functions $G(K) \to \mathbb{C}$ with compact support with respect to convolution (we use the Haar measure dg on $G(K)$ for which $\mathrm{vol}(B) = 1$). For y, w as above, let $\mathfrak{T}_{yw} \in H$ be the characteristic function of Ω_{yw}. Then the functions $\mathfrak{T}_{\underline{w}}$, $\underline{w} \in W$ form a $\mathbb{C}$-basis of H, and according to [Iwahori and Matsumoto 1965], we have

$$\begin{aligned} \mathfrak{T}_{\underline{w}}\mathfrak{T}_{\underline{w}'} &= \mathfrak{T}_{\underline{w}\underline{w}'} \quad \text{for } \underline{w}, \underline{w}' \in W \text{ with } l(\underline{w}\underline{w}') = l(\underline{w}) + l(\underline{w}'), \\ (\mathfrak{T}_{\underline{w}} + 1)(\mathfrak{T}_{\underline{w}} - q) &= 0 \quad \text{for } \underline{w} \in W' \text{ with } l(\underline{w}) = 1. \end{aligned}$$

In other words, H is what one now calls the Iwahori–Hecke algebra of the (extended) Coxeter group W with parameter q.

4.2. Let $\mathcal{C}_0^\infty(G(K))$ be the vector space of locally constant functions with compact support from $G(K)$ to $\mathbb{C}$. Let (V, σ) be an irreducible admissible representation of $G(K)$ such that the space V^B of B-invariant vectors in V is nonzero. If $f \in \mathcal{C}_0^\infty(G(K))$ then there is a well defined linear map $\sigma_f : V \to V$ such that for any $x \in V$ we have $\sigma_f(x) = \int_G f(g)\sigma(g)(x)\,dg$. This linear map has finite rank; hence, it has a well defined trace $\mathrm{tr}(\sigma_f) \in \mathbb{C}$. From the definitions we see that for $f, f' \in \mathcal{C}_0^\infty(G(K))$ we have $\sigma_{f*f'} = \sigma_f\sigma_{f'} : V \to V$ where $*$ denotes convolution. If $f \in H$ then σ_f maps V into V^B and $\mathrm{tr}(\sigma_f) = \mathrm{tr}(\sigma_f|_{V^B})$. (Recall that $\dim V^B < \infty$.) We see that the maps $\sigma_f|_{V^B}$ define a (unital) H-module structure on V^B. It is known that the H-module V^B is irreducible [Borel 1976]. Moreover, for $\underline{w} \in W$ we have $\mathrm{tr}(\sigma_{\mathfrak{T}_{\underline{w}}}) = \mathrm{tr}(\mathfrak{T}_{\underline{w}})$, where the trace in the right side is taken in the H-module V^B.

Theorem 4.3. *Assume that K has characteristic zero and that p is sufficiently large. Let $y \in Y^+$ and $t \in T(K)^\spadesuit_y$. We have*

$$\phi_V(t) = q^{-\langle y, 2\rho\rangle}\mathrm{tr}(\mathfrak{T}_y),$$

where the trace in the right side is taken in the irreducible H-module V^B.

An equivalent statement is that

$$\phi_V(t) = \mathrm{tr}(\sigma_{\mathfrak{T}_y})/\,\mathrm{vol}(\Omega_y).$$

(Recall that $\mathfrak{T}_y$ on the right side is the characteristic function of $\Omega_y = BT(K)_yB$.)

The assumption on characteristic in the theorem is needed only to be able to use a result from [Adler and Korman 2007]; see (5) below. We expect that the theorem holds without that assumption.

In the case where $y = 0$, the theorem becomes

$$t \in T(K) \cap G_{cvr} \implies \phi_V(t) = \dim(V^B). \tag{4}$$

As pointed out to us by R. Bezrukavnikov and S. Varma, in the special case where $y \in Y^{++}$, Theorem 4.3 can be deduced from results in [Casselman 1977].

4.4. In the case where $V = \mathbf{S}$ (see 1.1), for any $y \in Y^+$, $\mathfrak{T}_y$ acts on the one-dimensional vector space V^B as the identity map, so that $\phi_V(t) = q^{-\langle y, 2\rho \rangle}$ for all $t \in T(K)^\spadesuit_y$. We thus recover Theorem 2.2 (which holds in any characteristic).

5. Proof of Theorem 4.3

5.1. Let $B = B_0, B_1, B_2, \ldots$ be the strictly decreasing Moy–Prasad [1994] filtration of B. This is a sequence associated to a point x in the building such that $B = G_{x,0}$. Each B_i/B_{i+1} is abelian. Let $T_n := T(K) \cap B_n$. Applying [Adler and Korman 2007, Corollary 12.11] to ϕ_V, we conclude that

$$\phi_V \text{ is constant on the } \mathrm{Ad}(G)\text{-orbit } {}^G(tT_1) \text{ of } tT_1. \tag{5}$$

Lemma 5.2. *Let $n \geq 1$. For any $t' \in T(K)^\spadesuit_y$ and $z \in B_n$, there exist $g \in B_n$, $t'' \in T_n$, and $z' \in B_{n+1}$ such that* $\mathrm{Ad}(g)(t'z) = t't''z'$.

Proof. Let $Z = \{\alpha \in R : U_\alpha \cap B_n \supsetneq U_\alpha \cap B_{n+1}\}$. If $Z = \varnothing$ then $B_n = T_n B_{n+1}$; hence, $z = t''z'$ for some $t'' \in T_n$ and $z' \in B_{n+1}$, and one can take $g = 1$. If $Z \neq \varnothing$ then we can find $a_\alpha \in K$ for each $\alpha \in Z$ such that $x_\alpha(a_\alpha) \in B_n$ and $z \equiv \prod_{\alpha \in Z} x_\alpha(a_\alpha) \pmod{T_n B_{n+1}}$. Such a_α can be chosen independent of the order of the product since $B_n/T_n B_{n+1}$ is abelian. Take $g = \prod_{\alpha \in Z} x_\alpha((1-\alpha(t'^{-1}))^{-1} a_\alpha)$. Then $g \in B_n$ since $|1-\alpha(t'^{-1})| \geq 1$ for $y \in Y^+$. (To show $|1-\alpha(t'^{-1})| \geq 1$ for $y \in Y^+$, we argue as for (1). Assume first that $\alpha \in R^+$. If $v(\alpha(t'^{-1})) \neq 0$ then $v(\alpha(t'^{-1})) < 0$ (since $\langle y, \alpha \rangle \neq 0$, $\langle y, \alpha \rangle \geq 0$); therefore, $v(1-\alpha(t'^{-1})) = v(\alpha(t'^{-1})) < 0$ and $|1-\alpha(t'^{-1})| > 1$. If $v(\alpha(t'^{-1})) = 0$ then $\alpha(t'^{-1}) - 1 \in \mathcal{O} - \mathfrak{p}$; hence, $v(1-\alpha(t'^{-1})) = 0$ and $|1-\alpha(t'^{-1})| = 1$ as required. Assume next that $\alpha \in R^-$. If $v(\alpha(t'^{-1})) \neq 0$ then $v(\alpha(t'^{-1})) > 0$ (since $\langle y, \alpha \rangle \neq 0$, $\langle y, \alpha \rangle \leq 0$); hence, $v(1-\alpha(t'^{-1})) = 0$ and $|1-\alpha(t'^{-1})| = 1$ as required. If $v(\alpha(t'^{-1})) = 0$ then $\alpha(t'^{-1}) - 1 \in \mathcal{O} - \mathfrak{p}$; hence, $v(1-\alpha(t'^{-1})) = 0$ and $|1-\alpha(t'^{-1})| = 1$ as required.) Now, we have $t'^{-1} g t' g^{-1} \equiv z^{-1} \pmod{T_n B_{n+1}}$.

Writing $\mathrm{Ad}(g)(t'z) = t' \cdot (t'^{-1} g t' g^{-1}) \cdot (g z g^{-1})$, we observe that $g z g^{-1} \equiv z \pmod{B_{n+1}}$ and $t'^{-1} g t' g^{-1} z \in T_n B_{n+1}$; hence, $\mathrm{Ad}(g)(t'z)$ can be written as $t't''z'$ with $t'' \in T_n$ and $z' \in B_{n+1}$. □

Lemma 5.3. $B_1 t B_1 \subset {}^{B_1}(tT_1)$.

Proof. It is enough to show that $tB_1 \subset {}^{B_1}(tT_1)$. Let $t_0 z_1 \in tB_1$ with $t_0 = t$ and $z_1 \in B_1$. We will construct inductively sequences $g_1, g_2, \ldots,$ $t_1, t_2, \ldots,$ and $z_1, z_2, \ldots$ such that $\mathrm{Ad}(g_k \cdots g_2 g_1)(t_0 z_1) = \mathrm{Ad}(g_k)(t_0 t_1 \cdots t_{k-1} z_k) = (t_0 t_1 \cdots t_k) z_{k+1}$ with $g_i \in B_i$, $t_i \in T_i$, and $z_i \in B_i$.

Applying Lemma 5.2 to $n = 1$, $t' = t_0$, and $z = z_1$, we find $t_1 \in T_1$ and $z_2 \in B_2$ such that $g_1 t_0 z_1 g_1^{-1} = t_0 t_1 z_2$ with $t_1 \in T_1$ and $z_2 \in B_2$. Suppose we found $g_i \in B_i$, $z_{i+1} \in B_{i+1}$, and $t_i \in T_i$ for $i = 1, \dots, k$ where $k \geq 1$. Applying Lemma 5.2 to $n = k+1$, $t' = t_0 t_1 \cdots t_k$, and $z = z_{k+1}$, we find $g_{k+1} \in B_{k+1}$, $t_{k+1} \in T_{k+1}$, and $z_{k+2} \in B_{k+2}$ so that $g_{k+1} t_0 t_1 \cdots t_k z_{k+1} g_{k+1}^{-1} = \mathrm{Ad}(g_{k+1} \cdots g_2 g_1)(t_0 z_1) = t_0 t_1 t_2 \cdots t_{k+1} z_{k+2}$. (To apply Lemma 5.2 we note that $t' \in T(K)_y^{\spadesuit}$ since $t_0 \in T(K)_y^{\spadesuit}$ and $t_1 \cdots t_k \in T_1$, so that for any $\alpha \in R$ we have $\alpha(t_1 \cdots t_k) \in 1 + \mathfrak{p}$.) Taking $g \in B_1$ to be the limit of $g_k \cdots g_2 g_1$ as $k \to \infty$, we have $\mathrm{Ad}(g)(t_0 z_1) \in t T_1$. □

5.4. Continuing with the proof of Theorem 4.3, we note that by Lemma 5.3 and (5), for the characteristic function f_t of $B_1 t B_1$, we have

$$\mathrm{tr}(\sigma_{f_t}) = \int_G f_t(g) \phi_V(g)\, dg = \int_{B_1 t B_1} \phi_V(t)\, dg = \mathrm{vol}(B_1 t B_1) \phi_V(t).$$

Thus it remains to show

$$\frac{\mathrm{tr}(\sigma_{f_t})}{\mathrm{vol}(B_1 t B_1)} = \frac{\mathrm{tr}(\sigma_{\mathfrak{T}_y})}{\mathrm{vol}(BtB)}.$$

Since B_1 is normalized by B, B acts on V^{B_1}; moreover, since V is irreducible and $V^B \neq 0$, B acts trivially on V^{B_1}. (Otherwise, there would exist a nonzero subspace of V on which B acts through a nontrivial character of B/B_1; since $V^B \neq 0$, we see that (V, σ) would have two distinct cuspidal supports, a contradiction.) Thus we have $V^{B_1} = V^B$. Since σ_{f_t} and $\sigma_{\mathfrak{T}_y}$ have images contained in $V^{B_1} = V^B$, it is enough to show

$$\frac{\mathrm{tr}(\sigma_{f_t}|_{V^B})}{\mathrm{vol}(B_1 t B_1)} = \frac{\mathrm{tr}(\sigma_{\mathfrak{T}_y}|_{V^B})}{\mathrm{vol}(BtB)}. \tag{6}$$

We can find a finite subset L of $T(K)_0$ such that $BtB = \bigsqcup_{\tau \in L} B_1 t B_1 \tau$. It follows that

$$\mathrm{vol}(BtB) = \mathrm{vol}(B_1 t B_1) \sharp(L) \tag{7}$$

and $\sigma_{\mathfrak{T}_y} = \sum_{\tau \in L} \sigma_{f_t} \sigma(\tau)$ as linear maps $V \to V$. Restricting this equality to V^B and using the fact that $\sigma(\tau)$ acts as identity on V^B, we obtain

$$\sigma_{\mathfrak{T}_y}|_{V^B} = \sharp(L) \sigma_{f_t}|_{V^B} \tag{8}$$

as linear maps $V^B \to V^B$. Clearly, (6) follows from (7) and This completes the proof of Theorem 4.3. □

The following result will not be used in the rest of the paper:

Proposition 5.5. *If $y \in Y^{++}$ and $t \in T(K)_y$ then $BtB \subset {}^{B_1}T(K)_y$.*

Proof. It is enough to show that $tz \subset {}^{B_1}T(K)_y$ for any $z \in B$. We can write $z = t_0 z'$, where $t_0 \in T(K)_0$, $z' \in B_1$. We have $tz = t t_0 z'$, where $t t_0 \in T(K)_y = T(K)_y^{\spadesuit}$ (here we use that $y \in Y^{++}$). Using Lemma 5.3, we have $t t_0 z' \in {}^{B_1}(t t_0 T_1) \subset {}^{B_1}T(K)_y$. □

5.6. In the remainder of this section we assume that G is adjoint. In this case, the irreducible representations (V, σ) as in 4.2 (up to isomorphism) are known to be in bijection with the irreducible finite-dimensional representations of the Hecke algebra H (see [Borel 1976]) by $(V, \sigma) \mapsto V^B$. The irreducible finite-dimensional representations of H have been classified in [Kazhdan and Lusztig 1987] in terms of geometric data; moreover, in [Lusztig 2010], an algorithm to compute the dimensions of the (generalized) weight spaces of the action of the commutative semigroup $\{\mathfrak{T}_y : y \in Y^+\}$ on any tempered H module is given. In particular the right hand side of the equality in Theorem 4.3 (hence also $\phi_V(t)$ in that theorem) is computable when V is tempered.

References

[Adler and Korman 2007] J. D. Adler and J. Korman, "The local character expansion near a tame, semisimple element", *Amer. J. Math.* **129**:2 (2007), 381–403. MR 2008a:22020 Zbl 1128.22008

[Borel 1976] A. Borel, "Admissible representations of a semi-simple group over a local field with vectors fixed under an Iwahori subgroup", *Invent. Math.* **35** (1976), 233–259. MR 56 #3196 Zbl 0334.22012

[Casselman 1973] W. Casselman, "The Steinberg character as a true character", pp. 413–417 in *Harmonic analysis on homogeneous spaces* (Williamstown, MA, 1972), edited by C. C. Moore, Amer. Math. Soc., Providence, R.I., 1973. MR 49 #3039 Zbl 0289.22017

[Casselman 1977] W. Casselman, "Characters and Jacquet modules", *Math. Ann.* **230**:2 (1977), 101–105. MR 58 #11237 Zbl 0337.22019

[Harish-Chandra 1973] Harish-Chandra, "Harmonic analysis on reductive p-adic groups", pp. 167–192 in *Harmonic analysis on homogeneous spaces* (Williamstown, MA, 1972), edited by C. C. Moore, Amer. Math. Soc., Providence, R.I., 1973. MR 49 #5238 Zbl 0289.22018

[Iwahori and Matsumoto 1965] N. Iwahori and H. Matsumoto, "On some Bruhat decomposition and the structure of the Hecke rings of $\mathfrak{p}$-adic Chevalley groups", *Inst. Hautes Études Sci. Publ. Math.* **25** (1965), 5–48. MR 32 #2486 Zbl 0228.20015

[Kazhdan and Lusztig 1987] D. Kazhdan and G. Lusztig, "Proof of the Deligne–Langlands conjecture for Hecke algebras", *Invent. Math.* **87**:1 (1987), 153–215. MR 88d:11121 Zbl 0613.22004

[Kazhdan and Lusztig 1988] D. Kazhdan and G. Lusztig, "Fixed point varieties on affine flag manifolds", *Israel J. Math.* **62**:2 (1988), 129–168. MR 89m:14025 Zbl 0658.22005

[Lusztig 2010] G. Lusztig, "Graded Lie algebras and intersection cohomology", pp. 191–224 in *Representation theory of algebraic groups and quantum groups*, edited by A. Gyoja et al., Progr. Math. **284**, Birkhäuser, New York, 2010. MR 2011m:20106 Zbl 1231.20044

[Matsumoto 1969] H. Matsumoto, "Fonctions sphériques sur un groupe semi-simple p-adique", *C. R. Acad. Sci. Paris Sér. A-B* **269** (1969), 829–832. In French. MR 41 #8576 Zbl 0189.44802

[Moy and Prasad 1994] A. Moy and G. Prasad, "Unrefined minimal K-types for p-adic groups", *Invent. Math.* **116**:1-3 (1994), 393–408. MR 95f:22023 Zbl 0804.22008

[Rodier 1986] F. Rodier, "Sur le caractère de Steinberg", *Compositio Math.* **59**:2 (1986), 147–149. In French. MR 87m:22045 Zbl 0609.20025

[Shalika 1970] J. A. Shalika, "On the space of cusp forms of a $\mathcal{P}$-adic Chevalley group", *Ann. of Math.* (2) **92** (1970), 262–278. MR 42 #423 Zbl 0217.37001

[Steinberg 1951] R. Steinberg, "A geometric approach to the representations of the full linear group over a Galois field", *Trans. Amer. Math. Soc.* **71** (1951), 274–282. MR 13,317d Zbl 0045.30201

Received August 2, 2012. Revised November 26, 2012.

JU-LEE KIM
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA 02139
UNITED STATES
julee@math.mit.edu

GEORGE LUSZTIG
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA 02139-4307
UNITED STATES
gyuri@math.mit.edu

CONTENTS

Volume 265, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 265 No. 2 October 2013
