Pacific Journal of Mathematics

Volume 249 No. 2 February 2011

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

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

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

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

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

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

A GLUING CONSTRUCTION FOR PRESCRIBED MEAN CURVATURE

ADRIAN BUTSCHER

The gluing technique is used to construct hypersurfaces in Euclidean space having approximately constant prescribed mean curvature. These surfaces are perturbations of unions of finitely many spheres of the same radius assembled end-to-end along a line segment. The condition on the existence of these hypersurfaces is the vanishing of the sum of certain integral moments of the spheres with respect to the prescribed mean curvature function.

1. Introduction

In [Butscher and Mazzeo 2008] we have constructed examples of constant mean curvature (CMC) hypersurfaces in a Riemannian manifold M with axial symmetry by gluing together small spheres positioned end-to-end along a geodesic γ. These examples have very large mean curvature $2/r$ and lie within a distance $\mathcal{O}(r)$ of either a segment or a ray of γ; hence we say that these surfaces *condense* to the appropriate subset of γ. Such surfaces cannot exist in Euclidean space, and their existence relies on the fact that the gradient of the ambient scalar curvature of M acts as a "friction term" that permits the usual analytic gluing construction (akin to the classical gluing constructions pioneered by Kapouleas [1990a; 1991]) to be carried out. The purpose of this paper is to show the same techniques used in [Butscher and Mazzeo 2008] can be adapted in a straightforward manner to show that a similar construction is possible in a much simpler yet fairly general context: that of hypersurfaces having *prescribed* near-constant mean curvature in Euclidean space, in a certain sense to be explained forthwith. The essence of the gluing construction carried out herein therefore lies in identifying and appropriately exploiting the analogous friction term appearing in this setting.

Let $F : \mathbb{R}^{n+1} \times T\mathbb{R}^{n+1} \to \mathbb{R}$ be a given, fixed smooth function. For simplicity and to maintain the parallel with the earlier paper, we will assume that F has cylindrical symmetry in the following sense. Endow $\mathbb{R}^{n+1}$ with coordinates $(x^0, x^1, \dots, x^n)$ and let $G \subseteq O(n+1)$ be the set of orthogonal transformations that fix the x^0-axis. Each rotation $R \in G$ acts on $T\mathbb{R}^{n+1}$ via the differential $R_* : T\mathbb{R}^{n+1} \to T\mathbb{R}^{n+1}$. We

MSC2000: 53A10, 58J05.
Keywords: gluing constructions, prescribed mean curvature.

will now demand that $F(R(p), R_* V_p) = F(p, V_p)$ for all $(p, V_p) \in \mathbb{R}^{n+1} \times T\mathbb{R}^{n+1}$. The prescribed mean curvature problem that will be solved in this paper is to find, for every sufficiently small $r \in \mathbb{R}_+$, a G-invariant hypersurface Σ_r which satisfies

$$H[\Sigma_r](p) = 2 + r^2 F(p, N_{\Sigma_r}(p)) \quad \text{for all } p \in \Sigma_r, \tag{1-1}$$

where $H[\Sigma_r]$ is the mean curvature of Σ_r and N_{Σ_r} is the unit normal vector field of Σ_r. Note that we are not "prescribing" mean curvature in the usual sense; i.e., we don't have an *a priori* curvature function in mind that should equal the mean curvature of the hypersurfaces we construct. Instead, we should understand "prescribed mean curvature" to mean that a fixed external quantity (the function F) imposes an extra condition on the geometry of the hypersurface, which must adjust itself in $\mathbb{R}^3$ in order to satisfy this condition. Consequently, we won't know exactly the value of the mean curvature function, but we will know that it is near-constant and that the external condition is satisfied.

The prescribed mean curvature hypersurfaces of this paper will be built by gluing together a finite number K of spheres of radius one (and thus of mean curvature exactly equal to two) whose centers lie on the x^0-axis using small catenoidal necks having the x^0-axis as their axes of symmetry. In order to properly state the Main Theorem, we must make the following definition, which is meant to capture the most important effect of the prescribed mean curvature function F on the surface whose construction is accomplished in this paper.

Definition 1.1. Let S be a compact surface in $\mathbb{R}^{n+1}$. The *F-moment* of S is the quantity

$$\mu_F(S) := \int_S F(x, N_S(x))\, J \, \mathrm{dVol}_S$$

where N_S is the unit normal vector field of S and dVol_S is the induced volume form of S, while $J : S \to \mathbb{R}$ is defined by $J(x) := \langle \partial/\partial x^0, N_S(x) \rangle$ for $x \in S$.

Now let $p_k^0(s) := (s + 2(k-1), 0, \dots, 0)$ and consider the spheres $S_k(s) := \partial B_1(p_k^0(s))$. These spheres are positioned along the x^0-axis in such a way that each $S_k(s)$ makes tangential contact with $S_{k\pm1}(s)$. The following theorem will be proved in this paper.

Main Theorem. *Suppose that there is $s_0 \in \mathbb{R}$ such that*

- *the F-moments of the spheres $S_k(s_0)$ satisfy $\sum_{k=1}^K \mu_F(S_k(s_0)) = 0$, and*
- *the function $s \mapsto \sum_{k=1}^K \mu_F(S_k(s))$ has nonvanishing derivative at $s = s_0$,*

then for all sufficiently small $r > 0$, there is a smooth, embedded hypersurface Σ_r which is a small perturbation of $\bigcup_{k=1}^K S_k(s_0)$ that satisfies the prescribed mean curvature equation (1-1).

It is easy to find a situation in which the conditions of the Main Theorem hold. For example: if $F(\,\cdot\,,\cdot\,)$ is such that $\mu_F(\partial B_1(x^0, x^1, \ldots, x^n))$ is negative whenever x^0 is sufficiently negative and positive whenever x^0 is sufficiently positive, the mean value theorem asserts that the function $s \mapsto \sum_{k=1}^{K} \mu_F(S_k(s))$ has a zero. And if also $F(x,\cdot\,)$ is monotone as a function of x^0, this function will have nonzero derivative.

An application of the Main Theorem, and indeed an inspiration for it, is the earlier work by Kapouleas [1990b] on slowly rotating assemblies of water droplets. In this case, the prescribed mean curvature function $F : \mathbb{R}^{n+1} \times T\mathbb{R}^{n+1} \to \mathbb{R}$ takes the form $F(p, N_{\Sigma_r}(p)) := C(\omega)(p^0)^2$ where $p := (p^0, p^1, \ldots p^n)$ and $C(\omega)$ depends on the angular velocity ω. The prescribed mean curvature equation now approximates the effect of centrifugal force on the surface Σ_r when ω is small. One of the assemblies of water droplets that Kapouleas constructs is exactly as described in the Main Theorem. (He constructs many other, more complex, and less symmetrical assemblies as well.)

Another application of the Main Theorem is for understanding the possible shapes an electrically charged soap film can adopt in the presence of a weak, axially symmetric electric field. In this case, the equation satisfied by the surface adopted by the soap film is exactly (1-1), where the prescribed mean curvature function $F : \mathbb{R}^{n+1} \times T\mathbb{R}^{n+1} \to \mathbb{R}$ takes the form $F(p, N_{\Sigma_r}(p)) := -C\langle \nabla\phi(p), N_{\Sigma_r}(p)\rangle$ and $\phi : \mathbb{R}^{n+1} \to \mathbb{R}$ is the electric potential and C is a constant. We can see why this is so by writing the total energy of the soap film as the sum of a surface area term and a term proportional to the surface integral of ϕ, and then computing the Euler-Lagrange equation for the variation of this energy subject to the constraint that the volume enclosed by the surface remains constant. If we now assume that ϕ is such that the existence conditions of the Main Theorem hold, then the Main Theorem asserts that K spherical, electrically charged soap films connected by small catenoidal necks can be held in equilibrium at special points in space by the electric field.

2. The approximate solution

To construct an approximate solution for the Main Theorem, we use essentially exactly the same procedure as in [Butscher and Mazzeo 2008, §3.1]. This will be outlined here very briefly for the convenience of the reader. The presentation is given for the dimension $n = 2$ for simplicity; everything that follows can be easily adapted to the $(n+1)$-dimensional setting.

Endow $\mathbb{R}^3$ with coordinates (x^0, x^1, x^2), and let γ be the x^0-axis and $\gamma(t)$ be the arc-length parametrization of the x^0-axis with $\gamma(0) = (0, 0, 0)$. We will construct an approximate solution for the Main Theorem out of K spheres of radius one as follows. Choose a localization parameter $s \in \mathbb{R}$ and small separation parameters $\sigma_1, \dots, \sigma_{K-1} \in \mathbb{R}_+$. Define $s_1 := s$ and $s_k := s + 2(k-1) + \sum_{l=1}^{k-1} \sigma_l$ for $k = 2, \dots, K$ and set $p_k := \gamma(s_k)$ and $p_k^\pm := \gamma(s_k \pm 1)$. Define the spheres $S_k := \partial B_1(p_k)$. These spheres will now be joined together according to the following three steps.

Step 1. The first step is to replace each S_k with the surface $\tilde{S}_k$ obtained by taking the normal graph of a specially chosen function G_k over $S_k \setminus [B_{\rho_k}(p_k^+) \cup B_{\rho_k}(p_k^-)]$ where $\rho_k \in (0, 1)$ is a small radius as yet to be determined. The functions we use for this purpose can be defined as follows. Let $\mathcal{L}_{\mathbb{S}^2} := \Delta_{\mathbb{S}^2} + 2$ be the linearized mean curvature operator of the unit sphere, let $\varepsilon_k^\pm$ be yet-to-be-determined small scale parameters and let $J_k := \langle \partial/\partial x^0, N_{S_k} \rangle$ be the sole G-invariant function in the kernel of $\mathcal{L}_{\mathbb{S}^2}$ normalized to have unit L^2-norm. Then the functions G_k should satisfy the equations

$$\begin{aligned} &\mathcal{L}_{\mathbb{S}^2}(G_k) = \varepsilon_k^+ \delta(p_k^+) + \varepsilon_k^- \delta(p_k^-) + A_k J_k && \text{if } k = 2, \dots K-1, \\ &\mathcal{L}_{\mathbb{S}^2}(G_1) = \varepsilon_1^+ \delta(p_1^+) + A_1 J_1 && \text{if } k = 1, \\ &\mathcal{L}_{\mathbb{S}^2}(G_K) = \varepsilon_K^- \delta(p_K^-) + A_K J_K && \text{if } k = K, \end{aligned}$$

where $\delta(q)$ is the Dirac δ-function centered at q and A_k is chosen to ensure L^2-orthogonality to J_k. (Of course $J_k = x^0|_{S_k}$, the restriction of the x^0 coordinate function to S_k). Furthermore, G_k should be chosen L^2-orthogonal to J_k, normalized to have unit L^2-norm, and to be positive in a neighborhood of p_k^+.

Step 2. Let Ξ be the catenoid, i.e., the unique complete minimal surface of revolution whose axis of symmetry is γ and whose waist lies in the (x^1, x^2)-plane. The next step is to find the truncated and rescaled catenoidal neck of the form $\Xi_k := B_{\rho_k'}(p_k^\flat) \cap [\varepsilon_k \Xi + p_k^\flat + (\delta_k, 0, 0)]$ that fits optimally in the space between $\tilde{S}_k$ and $\tilde{S}_{k+1}$ for $k = 2, \dots, K-1$. Here $\varepsilon_k > 0$ is a small scale parameter and $p_k^\flat$ is a point between p_k^+ and p_{k+1}^- that are determined by the optimal fitting procedure while δ_k is a small vertical displacement parameter that takes Ξ_k away from its optimal location and ρ_k' is a small radius as yet to be determined. The optimal fit is obtained by matching the asymptotic expansions of the functions giving $\tilde{S}_k \cap B_{\rho_k'}(p_k^\flat)$ and $\tilde{S}_{k+1} \cap B_{\rho_k'}(p_k^\flat)$ and Ξ_k as graphs over the translate of the (x^1, x^2)-plane passing through $p_k^\flat$ exactly as in [Butscher and Mazzeo 2008, §3.1]. One particularly important outcome of the matching is that ε_k from the previous step, as well as $\varepsilon_k^\pm$ and $p_k^\flat$ are all uniquely determined by σ_k. In fact, an invertible relationship of the form $\sigma_k := \Lambda_k(\varepsilon_k)$ holds, with $\Lambda_k(\varepsilon_k) = \mathcal{O}(\varepsilon_k |\log(\varepsilon_k)|)$. Finally, we find that we must choose $\rho_k, \rho_k' = \mathcal{O}(\varepsilon_k^{3/4})$ to ensure the optimal fit between the necks and the perturbed spheres.

Step 3. The final step is to use cut-off functions to smoothly glue the neck Ξ_k into the space between $\tilde{S}_k$ and $\tilde{S}_{k+1}$. In this way we obtain a family of surfaces depending on the σ, δ and s parameters. Denote the neck modified by the cut-off functions by $\tilde{\Xi}_k$. The interpolating region is the annulus $B_{\rho'_k}(p^\flat_k) \setminus B_{\rho'_k/2}(p^\flat_k)$.

Definition 2.1. Let K be given. The approximate solution with parameters $\sigma := \{\sigma_1, \dots, \sigma_{K-1}\}$ and $\delta := \{\delta_1, \dots, \delta_{K-1}\}$ and s is the surface given by

$$\tilde{\Sigma}(\sigma, \delta, s) := \bigcup_{k=1}^{K} \tilde{S}_k \cup \bigcup_{k=1}^{K-1} \tilde{\Xi}_k.$$

3. Solving the projected problem

We now proceed to solve (1-1) up to a finite-dimensional error term by perturbing the approximate solution constructed in the previous section. The required analysis is in most respects identical to or less involved than the analysis found in [Butscher and Mazzeo 2008, SS4–6] and will thus again only be abbreviated here for the sake of the reader. The outcome will be a surface $\Sigma^\sharp_r(\sigma, \delta, s)$ satisfying $H[\Sigma^\sharp_r(\sigma, \delta, s)] - 2 - r^2 F|_{\Sigma^\sharp_r(\sigma,\delta,s)} \in \tilde{\mathcal{W}}$, where $\tilde{\mathcal{W}}$ is a finite-dimensional space of functions that will be defined precisely below. It arises because the linearized mean curvature operator, which governs the solvability of (3-1), possesses a finite-dimensional *approximate kernel* consisting of eigenfunctions corresponding to small eigenvalues. These small eigenvalues make it impossible to implement a convergent algorithm for prescribing the components of the mean curvature of the approximate solution lying in $\tilde{\mathcal{W}}$.

Function spaces. We first define the weighted Hölder spaces in which the analysis will be carried out. These are essentially the same weighted spaces as in [Butscher and Mazzeo 2008, §4], namely the spaces $C^{k,\alpha}_\nu(\tilde{\Sigma}(\sigma, \delta, s))$ consisting of all $C^{k,\alpha}_{\mathrm{loc}}$ functions on $\tilde{\Sigma}(\sigma, \delta, s)$ where the rate of growth in the neck regions of $\tilde{\Sigma}(\sigma, \delta, s)$ is controlled by the parameter ν. Choose some fixed, small $0 < R \ll 1$ and define a weight function $\zeta : \tilde{\Sigma}(\sigma, \delta, s) \to \mathbb{R}$ as

$$\zeta(p) := \begin{cases} \|x\| & \text{for } p = (x^0, x) \in \bar{B}_{R/2}(p^\flat_k) \text{ for some } k, \\ \text{interpolation} & \text{for } p \in \bar{B}_R(p^\flat_k) \setminus B_{R/2}(p^\flat_k) \text{ for some } k, \\ 1 & \text{elsewhere,} \end{cases}$$

where the interpolation is such that ζ is smooth and monotone in the region of interpolation, has appropriately bounded derivatives, and is G-invariant. Now, for any open set $\mathcal{U} \subseteq \tilde{\Sigma}(\sigma, \delta, s)$, define

$$|f|_{C^{k,\alpha}_\nu(\mathcal{U})} := \sum_{i=0}^{k} |\zeta^{i-\nu} \nabla^i f|_{0,\mathcal{U}} + [\zeta^{k+\alpha-\nu} \nabla^k f]_{\alpha,\mathcal{U}},$$

where $|\cdot|_{0,\mathcal{U}}$ is the supremum norm on $\mathcal{U}$ and $[\cdot]_{\alpha,\mathcal{U}}$ is the α-Hölder coefficient on $\mathcal{U}$. This is the norm that will be used in the $C^{k,\alpha}_\nu(\tilde{\Sigma}(\sigma, \delta, s))$ spaces.

The equation to solve. Let $\mu : C^{2,\alpha}_\nu(\tilde\Sigma(\sigma,\delta,s)) \to \mathrm{Emb}(\tilde\Sigma(\sigma,\delta,s),\mathbb{R}^{n+1})$ be the exponential map of $\tilde\Sigma(\sigma,\delta,s)$ in the direction of the unit normal vector field of $\tilde\Sigma(\sigma,\delta,s)$. Hence $\mu_f(\tilde\Sigma(\sigma,\delta,s))$ is the normal deformation of $\tilde\Sigma(\sigma,\delta,s)$ generated by $f \in C^{2,\alpha}_\nu(\tilde\Sigma(\sigma,\delta,s))$. The equation

(3-1) $$H\big[\mu_f(\tilde\Sigma(\sigma,\delta,s))\big] = 2 + r^2\, F \circ \big(\mu_f \times N_{\mu_f(\tilde\Sigma(\sigma,\delta,s))}\big)$$

selects $f \in C^{2,\alpha}_\nu(\tilde\Sigma(\sigma,\delta,s))$ so that $\mu_f(\tilde\Sigma(\sigma,\delta,s))$ satisfies (1-1). In addition, the function f will be assumed G-invariant. Define the operator

$$\Phi_{r,\sigma,\delta,s} : C^{2,\alpha}_\nu(\tilde\Sigma(\sigma,\delta,s)) \to C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma,\delta,s))$$

by

$$\Phi_{r,\sigma,\delta,s}(f) := H\big[\mu_f(\tilde\Sigma(\sigma,\delta,s))\big] - 2 - r^2\, \mathcal{F}(f),$$

where $\mathcal{F}(f) := F \circ (\mu_f \times N_{\mu_f(\tilde\Sigma(\sigma,\delta,s))})$. The linearization of $\Phi_{r,\sigma,\delta,s}$ at zero is given by

$$\begin{aligned}\mathcal{L} &:= \mathrm{D}\Phi_{r,\sigma,\delta,s}(0)\\ &= \Delta + \|B\|^2 + r^2\big(\mathrm{D}_1 F(\mu_0, N_{\tilde\Sigma(\sigma,\delta,s)})\cdot f N_{\tilde\Sigma(\sigma,\delta,s)} - \mathrm{D}_2 F(\mu_0, N_{\tilde\Sigma(\sigma,\delta,s)})\cdot \nabla f\big),\end{aligned}$$

where $\mathrm{D}_1 F$ and $\mathrm{D}_2 F$ are the derivatives of F in its first and second slots and $B := B[\tilde\Sigma(\sigma,\delta,s)]$ is the second fundamental form of $\tilde\Sigma(\sigma,\delta,s)$.

The space $\tilde{\mathcal{W}}$ is defined as follows. On the k-th spherical part of $\tilde\Sigma(\sigma,\delta,s)$, the operator $\mathcal{L}$ is a small perturbation of $\mathcal{L}_k := \Delta_{S_k} + 2$ which is the linearized mean curvature operator of the sphere S_k. Let J_k once again be the G-invariant function in its kernel. Now let $\Pi_{\mathrm{ext},k} : \tilde S_k \to S_k \setminus [B_{\rho^k}(p_k^+) \cup B_{\rho^k}(p_k^-)]$ for $k = 1, \dots, K-1$ and also $\Pi_{\mathrm{ext},1} : \tilde S_1 \to S_1 \setminus B_{\rho^k}(p_k^+)$ and $\Pi_{\mathrm{ext},K} : \tilde S_K \to S_K \setminus B_{\rho^k}(p_K^-)$ be the nearest-point projection mappings and define $\tilde J_k := J_k \circ \Pi_{\mathrm{ext},k}$. Finally, let $\chi_{\mathrm{ext},k}$ be a smooth cut-off function supported on $\tilde S_k$ and let η_k be a smooth cut-off function supported on the transition region between the k-th neck and $\tilde S_k$ with the property that the support of $\nabla\eta_k$ and $\nabla\chi_{\mathrm{ext},k}$ do not overlap (this technical assumption is needed in the fine details of the analysis carried out in [Butscher and Mazzeo 2008]).

Definition 3.1. The space $\tilde{\mathcal{W}}$ is defined as

$$\tilde{\mathcal{W}} := \mathrm{span}\{\chi_{\mathrm{ext},k}\tilde J_k : k = 1, \dots, K\} \cup \{\chi_{\mathrm{ext},k}\mathcal{L}_k(\eta_k) : k = 1, \dots, K-1\}.$$

We now prove the following theorem. Let $\varepsilon := \max\{\varepsilon_1, \dots, \varepsilon_{K-1}\}$ and $\delta := \max\{\delta_1, \dots, \delta_{K-1}\}$ and we will assume that $\varepsilon = \mathcal{O}(r^2)$ and $\delta = \mathcal{O}(r)$, which will be justified *a posteriori*.

Theorem 3.2. *If $r > 0$ is sufficiently small, then there exists $f := f_r(\sigma,\delta,s) \in C^{2,\alpha}_\nu(\tilde\Sigma(\sigma,\delta,s))$ with $\nu \in (1,2)$ so that*

(3-2) $$\Phi_{r,\sigma,\delta,s}(f) \in \tilde{\mathcal{W}}.$$

The estimate $|f|_{C^{2,\alpha}_\nu} \le Cr^2$ *holds for the function* f, *where the constant* C *is independent of* r. *Finally, the mapping* $(\sigma, \delta, s) \mapsto f_r(\sigma, \delta, s)$ *is smooth in the sense of Banach spaces.*

Proof. As in [Butscher and Mazzeo 2008], we will use a fixed-point argument to solve the equation $\Phi_{r,\sigma,\delta,s}(f) \in \tilde{\mathcal{W}}$ for a function $f \in C^{2,\alpha}_\nu(\tilde\Sigma(\sigma, \delta, s))$ with $\nu \in (1, 2)$. The fixed-point argument follows from three steps: an estimate of the size of $\Phi_{r,s,\sigma,\delta,s}(0)$; the construction of a bounded parametrix $\mathcal{R}$ satisfying $\mathcal{L} \circ \mathcal{R} = \mathrm{id} + \mathcal{E}$ where $\mathcal{E} : C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma, \delta, s)) \to \tilde{\mathcal{W}}$; and an estimate of the nonlinear part of the operator $\Phi_{r,\sigma,\delta,s}$. Each of these steps is given in great detail in the paper cited, so we just point out how the analysis there applies to the present situation.

Step 1. We begin with the estimate of $|\Phi_{r,\sigma,\delta,s}(0)|_{C^{0,\alpha}_{\nu-2}}$, the amount that the approximate solution $\tilde\Sigma(\sigma, \delta, s)$ deviates from being an actual solution of (3-1). This is done by adapting [Butscher and Mazzeo 2008, Proposition 13]. In fact, by using that proposition's steps 1, 2 and 4 in the estimate of $H[\tilde\Sigma(\sigma, \delta, s)] - 2$ in the $C^{0,\alpha}_{\nu-2}$ norm for $\nu \in (1, 2)$, together with a straightforward estimate for the $C^{0,\alpha}_{\nu-2}$ norm of the $r^2\mathcal{F}$ term, we find that

$$\left|\Phi_{r,\sigma,\delta,s}(0)\right|_{C^{0,\alpha}_{\nu-2}} \le C \max\{r^2,\ \varepsilon^{3/2-3\nu/4}, \delta\varepsilon^{1-3\nu/4}\} \le Cr^2$$

for some constant C independent of r.

Step 2. We now find a parametrix

$$\mathcal{R} : C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma, \delta, s)) \to C^{2,\alpha}_\nu(\tilde\Sigma(\sigma, \delta, s))$$

satisfying $\mathcal{L} \circ \mathcal{R} = \mathrm{id} + \mathcal{E}$, where $\mathcal{E} : C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma, \delta, s)) \to \tilde{\mathcal{W}}$. As in [Butscher and Mazzeo 2008, Proposition 15], this is done by first constructing an approximate parametrix by patching together parametrices for the linearized mean curvature operator of each sphere with parametrices for the linearized mean curvature operator of each neck; and then iterating to produce an exact parametrix plus an error term in $\tilde{\mathcal{W}}$ in the limit. The difference here is that the terms coming from the noneuclidean background metric in the result just cited must be replaced by the $r^2\mathcal{F}$ term. The same result holds because this term can easily be shown to satisfy the right estimates. In fact, $\mathcal{R}$ and $\mathcal{E}$ satisfy the estimate $|\mathcal{R}(w)|_{C^{2,\alpha}_\nu} + |\mathcal{E}(w)|_{C^{2,\alpha}_0} \le C|w|_{C^{0,\alpha}_{\nu-2}}$ for all $w \in C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma, \delta, s))$, where C is a constant independent of r.

Step 3. We define

$$\mathcal{Q} : C^{2,\alpha}_\nu(\tilde\Sigma(\sigma, \delta, s)) \to C^{0,\alpha}_{\nu-2}(\tilde\Sigma(\sigma, \delta, s)),$$

the quadratic (and higher) remainder term of the operator $\Phi_{r,\sigma,\delta,s}$, by

$$\mathcal{Q}(f) := \Phi_{r,\sigma,\delta,s}(f) - \Phi_{r,\sigma,\delta,s}(0) - \mathcal{L}(f).$$

The estimates for the $C^{0,\alpha}_{\nu}$ norm of $\mathcal{Q}$ can be found exactly as in [Butscher and Mazzeo 2008, Proposition 18] with the terms coming from the noneuclidean background metric replaced by the $r^2\mathcal{F}$ term. Then there exists $C_0 > 0$ so that if $f_1, f_2 \in C^{2,\alpha}_{\nu}(\tilde{\Sigma}(\sigma,\delta,s))$ for $\nu \in (1,2)$ and satisfying $|f_1|_{C^{2,\alpha}_{\nu}} + |f_2|_{C^{2,\alpha}_{\nu}} \le C_0$, then

$$|\mathcal{Q}(f_1) - \mathcal{Q}(f_2)|_{C^{0,\alpha}_{\nu-2}} \le C|f_1 - f_2|_{C^{2,\alpha}_{\nu}} \max\left\{|f_1|_{C^{2,\alpha}_{\nu}}, |f_2|_{C^{2,\alpha}_{\nu}}\right\},$$

where C is a constant independent of r. Once again, this works because the $r^2\mathcal{F}$ term can easily be shown to satisfy the right estimates.

Step 4. We can now solve the CMC equation up to a finite-dimensional error term by implementing a fixed-point argument based on the parametrix constructed in Step 2 as well as the estimates we have computed so far. Let $E := \Phi_{r,\sigma,\delta,s}(0)$ and use the *Ansatz* $f := \mathcal{R}(w - E)$ to convert the equation $\Phi_{r,\sigma,\delta,s}(f) \in \tilde{\mathcal{W}}$ into the fixed point problem $w - \mathcal{N}_{r,\sigma,\delta,s}(w) \in \tilde{\mathcal{W}}$, where

$$\mathcal{N}_{r,\sigma,\delta,s} : C^{0,\alpha}_{\nu-2}(\tilde{\Sigma}(\sigma,\delta,s)) \to C^{0,\alpha}_{\nu-2}(\tilde{\Sigma}(\sigma,\delta,s))$$

is defined by

$$\mathcal{N}_{r,\sigma,\delta,s}(w) := -\mathcal{Q} \circ \mathcal{R}(w - E).$$

The estimates established up to now give us

$$|\mathcal{N}_r(w_1) - \mathcal{N}_r(w_2)|_{C^{0,\alpha}_{\nu-2}} \le Cr^2|w_1 - w_2|_{C^{0,\alpha}_{\nu-2}}$$

for w in a ball of radius $\mathcal{O}(r^2)$ about zero in $C^{0,\alpha}_{\nu-2}(\tilde{\Sigma}(\sigma,\delta,s))$, where C is independent of r. Hence $\mathcal{N}_r$ is a contraction mapping on this ball if r is sufficiently small, and a solution of (3-2) satisfying the desired estimate can be found. The smooth dependence of this solution on the parameters (σ,δ,s) is a consequence of the fixed-point process. □

4. Force balancing arguments and the proof of the Main Theorem

When r is sufficiently small, we have now found a function

$$f_r(\sigma,\delta) \in C^{2,\alpha}_{*}(\tilde{\Sigma}(\sigma,\delta,s))$$

for each (σ,δ,s) such that

$$H\left[\mu_{f_r(\sigma,\delta)}(\tilde{\Sigma}(\sigma,\delta,s))\right] - 2 - r^2\mathcal{F}(f_r(\sigma,\delta,s)) = \mathcal{E}_r(\sigma,\delta,s),$$

where $\mathcal{E}_r(\sigma,\delta,s)$ is an error term belonging to the finite-dimensional space $\tilde{\mathcal{W}}$ depending on the free parameters (σ,δ,s). The corresponding surface that satisfies the prescribed mean curvature condition up to finite-dimensional error is $\Sigma^{\sharp}_r(\sigma,\delta,s) := \mu_{f_r(\sigma,\delta,s)}(\tilde{\Sigma}(\sigma,\delta,s))$.

To complete the proof of the Main Theorem, we must show that it is possible to find a value of (σ, δ, s) for which these error terms vanish identically. As in [Butscher and Mazzeo 2008, §7.2], we take cut-off functions $\chi'_{\text{ext},k}$ and $\chi'_{\text{neck},k}$ supported on the k-th spherical region and the k-th neck and transition region, respectively, and consider the *balancing map* $B_r : \mathbb{R}^{2K-1} \to \mathbb{R}^{2K-1}$ defined by

$$B_r(\sigma, \delta, s) := \big(\pi_1(\mathscr{E}_r(\sigma, \delta, s)), \ldots, \pi_{2K-1}(\mathscr{E}_r(\sigma, \delta, s))\big), \tag{4-1}$$

where $\pi_{2k+1} : \tilde{\mathscr{W}} \to \mathbb{R}$ and $\pi_{2k} : \tilde{\mathscr{W}} \to \mathbb{R}$ are the L^2-projection operators given by

$$\pi_{2k}(e) := \int_{\tilde{\Sigma}(\sigma,\delta,s)} e \cdot \chi'_{\text{neck},k} \tilde{I}_k, \qquad \pi_{2k+1}(e) := \int_{\tilde{\Sigma}(\sigma,\delta,s)} e \cdot \chi'_{\text{ext},k} \tilde{J}_k.$$

Here $\tilde{I}_k := I_k \circ \Pi_{\text{neck},k}$ where $\Pi_{\text{neck},k}$ is the nearest-point projection mapping of the perturbed k-th neck region onto the unperturbed k-th neck, and I_k is the Jacobi field of the k-th neck coming from translation along the neck axis. This is an odd, bounded function with respect to the center of the neck. Note that B_r is a smooth map between finite-dimensional vector spaces by virtue of the fact that the dependence of the solution $f_r(\sigma, \delta, s)$ on (σ, δ, s) is smooth and the mean curvature operator is a smooth map of the Banach spaces upon which it is defined. The following lemma proves that $\pi(e) = 0$ implies that $e = 0$.

Lemma 4.1. *Choose* $e \in \tilde{\mathscr{W}}$ *as* $e = \sum_{k=1}^{K} a_k \chi_{\text{ext},k} \tilde{J}_k + \sum_{k=1}^{K-1} b_k \mathscr{L}_k(\eta_k)$ *for* $a_k, b_k \in \mathbb{R}$. *Then*

$$\pi_{2k}(e) = C_1 b_k - C'_1 \varepsilon_k^{3/2} a_k, \quad \pi_{2k+1}(e) = C_2 a_k,$$

where C_1, C'_1, C_2 *are positive constants independent of* r *and* (σ, δ, s).

Proof. In the integral

$$\int e \cdot \chi'_{\text{ext},k} \tilde{J}_k = a_k \int \chi_{\text{ext},k} \chi'_{\text{ext},k} \tilde{J}_k^2 + \sum_{\ell=k-1}^{k} b_\ell \int \chi'_{\text{ext},k} \chi_{\text{ext},\ell} \mathscr{L}_\ell(\eta_\ell) \tilde{J}_k,$$

the second two terms can be made to vanish by choosing the supports of χ_{ext}, χ'_{ext} and η_k appropriately. The remaining term has large integral because $\tilde{J}_k = J_k \circ \Pi_{\text{ext},k}$ and J_k has unit L^2 norm as a function of the sphere. In the integral $\int e \cdot \chi'_{\text{neck},k} \tilde{I}_k = \sum_{\ell=k}^{k+1} a_\ell \int \chi_{\text{ext},\ell} \chi'_{\text{neck},k} \tilde{J}_\ell \tilde{I}_k + b_k \int \chi'_{\text{neck},k} \chi_{\text{ext},k} \mathscr{L}_k(\eta_k) \tilde{I}_k$ the first two terms contribute quantities proportional to the volume of the transition regions surrounding the k-th neck where $\chi_{\text{ext},k} \chi'_{\text{neck},k}$ is supported. The remaining term can be made to have large, positive and negative values (depending on the sign of $\tilde{I}_k$) by choosing the supports of χ_{ext}, χ'_{ext} to fall where the quantity $\mathscr{L}_k(\eta_k)$ is largest. □

We must now show that $B_r(\sigma, \delta, s)$ can be controlled by the initial geometry of $\tilde{\Sigma}(\sigma, \delta, s)$, at least to lowest order in r. The calculations are similar to those found in [Butscher and Mazzeo 2008, §7.2] except with the contributions from the ambient background geometry replaced by a contribution from the prescribed mean curvature in the form of the F-moments of the spheres making up $\tilde{\Sigma}(\sigma, \delta, s)$.

The highest-order part of $\mathscr{E}_r(\sigma, \delta, s)$ involves the F-moments of the spherical constituents S_k of $\tilde{\Sigma}(\sigma, \delta, s)$ as follows. Set $\mu_k(\sigma, s) := \mu_F(S_k)$ — this depends on s and $\sigma_1, \dots, \sigma_k$ because the location of the center of S_k is determined by these parameters. Let us continue to assume that $\varepsilon_k = \mathbb{O}(r^2)$ and $\delta_k = \mathbb{O}(r)$ for each k. This will be justified shortly.

Lemma 4.2. *The quantity $\mathscr{E}_r(\sigma, \delta, s)$ satisfies*

$$\pi_{2k}\big(\mathscr{E}_r(\sigma, \delta, s)\big) = C_1 \delta_k \varepsilon_k^{3/2} + \mathbb{O}(r^{2+2\nu}) \tag{4-2}$$

and

$$\begin{aligned} \pi_{2k+1}\big(\mathscr{E}_r(\sigma, \delta, s)\big) \\ = \begin{cases} C_2\varepsilon_1 - r^2\mu_1(\sigma, s) + \mathbb{O}(r^4) & \text{if } k = 0, \\ C_2(\varepsilon_{k+1} - \varepsilon_k) - r^2\mu_{k+1}(\sigma, s) + \mathbb{O}(r^4) & \text{if } 0 < k < K-1, \\ -C_2\varepsilon_K - r^2\mu_K(\sigma, s) + \mathbb{O}(r^4) & \text{if } k = K-1, \end{cases} \end{aligned} \tag{4-3}$$

where C_1, C_2 are constants independent of r, σ, δ, s.

Proof. Set $\Sigma_r^\sharp := \Sigma_r^\sharp(\sigma, \delta, s)$ and $\Sigma := \tilde{\Sigma}(\sigma, \delta, s)$ for convenience. Consider first (4-3) with $0 < k < K-1$. By the first variation formula and estimates of the size of the perturbation generating $\Sigma_r^\sharp$ from $\tilde{\Sigma}(\sigma, \delta, s)$, and calculating as in [Butscher and Mazzeo 2008, Proposition 27], we have

$$\begin{aligned} \pi_{2k+1}\big(\mathscr{E}_r(\sigma, \delta, s)\big) &= \int_{\Sigma_r^\sharp} \big(H[\Sigma_r^\sharp] - 2 - r^2\mathscr{F}(f_r(\sigma, \delta, s))\big)\chi_{\text{ext},k} J_k \\ &= \int_{\partial\Sigma^\sharp \cap \operatorname{supp} \chi_{\text{ext},k}} \Big\langle \frac{\partial}{\partial x^0}, \nu_k \Big\rangle - r^2 \int_{S_k} F(x, N_{S_k}(x)) J_k + \mathbb{O}(r^4) \\ &= C_2\big(\varepsilon_{k+1} - \varepsilon_k\big) - r^2\mu_k(s, \sigma) + \mathbb{O}(r^4), \end{aligned}$$

where ν_k is the unit normal vector field of $\partial\Sigma^\sharp \cap \operatorname{supp} \chi_{\text{ext},k}$ in $\Sigma^\sharp$.

Now consider (4-2). In the neck we have $H[\tilde{\Sigma}(\sigma, \delta, s)] = 0$. Using similar estimates, we get

$$\begin{aligned} \pi_{2k}\big(\mathscr{E}_r(\sigma, \delta, s)\big) &= \int_{\Sigma_r^\sharp} \big(H[\Sigma_r^\sharp] - 2 - r^2\mathscr{F}(f_r(\sigma, \delta, s))\big)\chi_{\text{neck},k} I_k \\ &= -2\int_{\Sigma \cap \operatorname{supp} \chi_{\text{neck},k}} \chi_{\text{neck},k} I_k + \mathbb{O}(r^{2+2\nu}) = C_1\delta_k\varepsilon_k^{3/2} + O(r^{2+2\nu}), \end{aligned}$$

where δ_k is the displacement parameter of the k-th neck. This is because I_k is an odd function with respect to the neck having $\delta_k = 0$, whereas the integral is being taken over the neck with $\delta_k \neq 0$. Hence the integral $\int_{\Sigma\cap\operatorname{supp}\chi_{\mathrm{neck},k}} \chi_{\mathrm{neck},k} I_k$ picks up the displacement of the k-th neck from its position at $\delta_k = 0$. This same phenomenon arises in [Butscher and Mazzeo 2008, Proposition 27]. □

4.1. ***Proof of the Main Theorem.*** It remains to find a value of the parameters (σ, δ, s) so that $\mathscr{E}_r(\sigma, \delta, s) = 0$. As shown in Lemma 4.1, this is equivalent to find a solution of the equation $B_r(\sigma, \delta, s) = 0$. In what follows, we will continue to assume that $\varepsilon = \mathbb{O}(r^2)$ and $\delta = \mathbb{O}(r)$ and this will be justified shortly. As a consequence of Lemma 4.2, the equations that we must solve are as follows:

$$\begin{aligned} C_1\delta_1 &= E_1(\sigma, \delta, s),\\ &\vdots\\ C_1\delta_{K-1} &= E_{K-1}(\sigma, \delta, s),\\ C_2\varepsilon_1 &= r^2\mu_1(\sigma, s) + E_1'(\sigma, \delta, s),\\ C_2(\varepsilon_2 - \varepsilon_1) &= r^2\mu_2(\sigma, s) + E_2'(\sigma, \delta, s),\\ &\vdots\\ C_2(\varepsilon_{K-1} - \varepsilon_{K-2}) &= r^2\mu_{K-1}(\sigma, s) + E_{K-1}'(\sigma, \delta, s),\\ -C_2\varepsilon_{K-1} &= r^2\mu_K(\sigma, s) + E_K'(\sigma, \delta, s), \end{aligned}$$

where ε_k depends on σ_k in an invertible manner as indicated in Step 2 of the construction of the approximate solution, and E_k, E_k' are error quantities satisfying the bounds $|E_k| = \mathbb{O}(r^{-1+2\nu})$ and $|E_k'| = \mathbb{O}(r^4)$. We can abbreviate these equations by introducing the matrix $M := \left(\begin{smallmatrix} I & 0\\ 0 & J\end{smallmatrix}\right)$, where I is the $(K-1)\times(K-1)$ identity matrix and J is the $K\times(K-1)$ matrix

$$J := \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & & \ddots & \\ & & -1 & 1\\ & & & -1 \end{pmatrix}.$$

The equations become

$$M(C_1\delta, C_2\varepsilon)^t = (E, r^2\mu + E')^t, \tag{4-4}$$

where $\delta := (\delta_1, \dots, \delta_{K-1})$, $\varepsilon := (\varepsilon_1, \dots, \varepsilon_{K-1})$ and so on for E, E' and μ.

We will solve these equations in two steps as follows. Note first that the matrix M is injective but not surjective, with vectors in the image of M satisfying the relation $(0, e)\cdot M(v, w) = 0$ for all $(v, w) \in \mathbb{R}^{2K-2}$, where $e := (1, 1, \dots, 1) \in \mathbb{R}^K$. Let $\rho : \mathbb{R}^{2K-1} \to \mathbb{R}^{2K-2}$ be the orthogonal projection onto the image of M. The equation

$$\rho M(\varepsilon,\delta) - \rho(E, r^2\mu + E') = 0 \tag{4-5}$$

can now be solved using the implicit function theorem when $r > 0$ is sufficiently small if the derivative matrix in (ε,δ) of the mapping on the left hand side of (4-5) above is nonsingular when $r = 0$. But this holds because the matrix $\rho M : \mathbb{R}^{2K-2} \to \mathbb{R}^{2K-2}$ is nonsingular and the contribution to the derivative matrix coming from the error term $\rho(E, r^2\mu + E')$ vanishes when $r = 0$.

We thus now have a solution $\varepsilon := \varepsilon_r(s)$ and $\delta_r(s)$ of (4-4) for all sufficiently small r and depending implicitly on the one remaining free parameter s. Moreover, we see that $\varepsilon = \mathbb{O}(r^2)$ and $\delta = \mathbb{O}(r^{-1+2\nu}) = \mathbb{O}(r)$ since $\nu \in (1,2)$. It remains to solve (4-5) and we proceed as follows. Once (ε,δ) satisfy (4-5), then (4-4) becomes equivalent to $0 = (0,e)\cdot M(\varepsilon,\delta) = r^2 e\cdot\mu + e\cdot E'$, or simply

$$\sum_{k=1}^{K} \mu_k(\sigma_r(s), s) + E''(\sigma_r(s), \delta_r(s), s) = 0, \tag{4-6}$$

where the error quantity satisfies the estimate $|E''| = \mathbb{O}(r^2)$.

Equation (4-6) may or may not have a solution, depending on the nature of the function $\sum_k \mu_k$, which in turn depends on the specific nature of the prescribed mean curvature function F. However, if the following two conditions are met, then the implicit function theorem guarantees the existence of a solution. First, it must be the case that the equation at $r = 0$ has a solution, in other words if the F-moments of the spheres $S_1, \dots, S_K$ satisfy

$$\sum_{k=1}^{K} \mu_F(\partial B_1(p_k^0(s))) = 0$$

for some s, where $p_k^0(s) := (s + 2(k-1), 0, \dots, 0)$. Second, if s_0 is the solution of this equation, then it must also be the case that the mapping

$$s \mapsto \sum_{k=1}^{K} \mu_F(\partial B_1(p_k^0(s)))$$

has nonvanishing derivative at $s = s_0$. If these conditions are satisfied, then the implicit function theorem implies that for r sufficiently small, there is a solution $s(r)$ of (4-6). This completes the proof. □

References

[Butscher and Mazzeo 2008] A. Butscher and R. Mazzeo, "Constant mean curvature hypersurfaces condensing to geodesic segments and rays in Riemannian manifolds", preprint, 2008. arXiv 0812.3133

[Kapouleas 1990a] N. Kapouleas, "Complete constant mean curvature surfaces in Euclidean three-space", *Ann. of Math.* (2) **131**:2 (1990), 239–330. MR 1043269 (93a:53007a

[Kapouleas 1990b] N. Kapouleas, "Slowly rotating drops", *Comm. Math. Phys.* **129**:1 (1990), 139–159. MR 1046281 (91c:76024

[Kapouleas 1991] N. Kapouleas, "Compact constant mean curvature surfaces in Euclidean three-space", *J. Differential Geom.* **33**:3 (1991), 683–715. MR 1100207 (93a:53007b

Received May 28, 2009. Revised September 7, 2010.

ADRIAN BUTSCHER
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CA 94305
UNITED STATES
butscher@stanford.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

LARGE EIGENVALUES AND CONCENTRATION

BRUNO COLBOIS AND ALESSANDRO SAVO

Let $M^n = (M, g)$ be a compact, connected, Riemannian manifold of dimension n. Let μ be the measure $\mu = \sigma\,\mathrm{dvol}_g$, where $\sigma \in C^\infty(M)$ is a nonnegative density. We first show that, under some mild metric conditions that do not involve the curvature, the presence of a large eigenvalue (or more precisely of a large gap in the spectrum) for the Laplacian associated to the density σ on M implies a strong concentration phenomenon for the measure μ. When the density is positive, we show that our result is optimal. Then we investigate the case of a Laplace-type operator $D = \nabla^*\nabla + T$ on a vector bundle E over M, and show that the presence of a large gap between the $(k+1)$-st eigenvalue λ_{k+1} and the k-th eigenvalue λ_k implies a concentration phenomenon for the eigensections associated to the eigenvalues $\lambda_1, \dots, \lambda_k$ of the operator D.

1. Introduction

The goal of this paper is to show that, under some mild metric conditions, the presence of a large eigenvalue of the Laplacian Δ on a compact Riemannian manifold M implies that the Riemannian volume concentrates around a finite set of points. Actually, we show that a similar phenomenon holds for any Laplace-type operator D acting on sections of a vector bundle on M, if one replaces the Riemannian volume by the squared norm of a first eigensection of D.

Let us recall briefly the main known facts about concentration and the spectrum of the Laplace operator. In what follows, we number the eigenvalues of Δ so that $\lambda_1(M) = 0$ and $\lambda_2(M)$ is the first positive eigenvalue.

For a closed Riemannian manifold of dimension n whose Ricci curvature is bounded below, that is, $\mathrm{Ric} \geq -(n-1)a^2$, we have the following well-known inequality due to Cheng [1975]:

$$\lambda_{k+1}(M) \leq \frac{(n-1)^2 a^2}{4} + \frac{c(n)k^2}{\mathrm{diam}(M)^2},$$

MSC2000: primary 58J50; secondary 35P15.
Keywords: eigenvalues, upper bounds, Laplace-type operators, concentration.

where $c(n)$ is a constant depending only on n. This shows that when the k-th eigenvalue is very large, the whole manifold is contained in a small neighborhood of any of its points and so we have a strong concentration phenomenon.

At the other extreme, if we make no assumption other than compactness we still have a concentration phenomenon, first observed by Gromov and Milman [1983, Theorem 4.1]. It says that if A is a closed subset with *positive* normalized measure $\mu(A) = \alpha$ and $r > 0$, then

$$\mu(A^r) \geq 1 - (1-\alpha^2)\exp(-r\sqrt{\lambda_2(M)}\ln(1+\alpha)), \tag{1}$$

where $A^r = \{x \in M : d(A, x) < r\}$.

So, when the first (positive) eigenvalue is large, almost all relative volume of M lies in a small neighborhood of any set of fixed positive measure.

However, we stress that $\mu(A)$ being positive is essential in the estimate; the sole assumption that $\lambda_2(M)$ is large does not guarantee that the volume concentrates around, say, a finite set of points. For example, take M_n to be the n-dimensional unit sphere. Then $\lambda_2(M_n)$ (which is equal to n) tends to infinity with n; we have concentration in the sense of Gromov and Milman, and yet the volume of M_n is uniformly distributed and cannot concentrate around any finite set. In Section A.4 we will give another counterexample in which the dimension is fixed.

Inequality (1) can be generalized to the other eigenvalues using an interesting upper bound of $\lambda_k(M)$ due to Chung, Grigor'yan and Yau; the upper bound is given in terms of the least distance between k mutually disjoint subsets of fixed positive measure; see [Chung et al. 1997] and also [Friedman and Tillich 2000] for a sharp estimate.

This paper deals with concentration around a finite number of points, and with a simple metric condition that will imply this phenomenon. Namely, we require that the number of balls of radius r needed to cover a ball of radius $4r$ is uniformly bounded above by a constant C for $r \leq 1$. We then prove the following fact:

If the $(k+1)$-st eigenvalue of the Laplacian of M is large, then most of the volume of M concentrates near (at most) k points of the manifold.

However, we will prove a result (Theorem 4) that is much more general; in particular, it will imply the following fact. Consider a Laplace-type operator D acting on the sections of a smooth vector bundle on M (for example, the Laplacian on forms, the square of the Dirac operator or the Schroedinger operator). Then:

If the gap between the $(k+1)$-st and the k-th eigenvalue of D is large, then any eigensection associated to the first k eigenvalues concentrates its L^2-norm near (at most) k points of the manifold.

Both the above estimates depend explicitly on the constant C.

In the rest of the introduction we state the precise results: Theorems 1, 2 and 3.

1.1. *Some definitions.* We will consider metric measure spaces (M, μ, d) of the following type:

- $M = (M^n, g)$ is a compact, connected Riemannian manifold of dimension n, possibly with nonempty boundary.
- μ is the measure $\mu = \sigma \, d\mathrm{vol}_g$, where $\sigma \in C^\infty(M)$ is a nonnegative density. We will also assume, without loss of generality, that μ is a probability measure, that is, $\int_M \sigma \, d\mathrm{vol}_g = 1$.
- d is a distance function that is assumed to be Lipschitz, that is, $|\nabla d| \leq 1$ almost everywhere with respect to μ.

For $r > 0$, define $C_d(M, r)$ to be the minimal number of balls of radius r in (M, d) needed to cover a ball of radius $4r$. Then $C_d(M, r)$ is finite for all r.

We will set

$$C_d(M) = \sup_{r \in (0,1]} C_d(M, r), \tag{2}$$

and call it the *packing constant* of the pair (M, d). It is a metric invariant (it does not depend on the measure μ).

The packing constant is often used in similar contexts (it is used extensively in the survey [Grigor'yan et al. 2004]). By the compactness of M, $C_d(M)$ is well-defined.

Note that d is not necessarily the Riemannian distance. In fact, here are three typical situations in which it is easy to control the packing constant:

(I) (M^n, g) is a closed Riemannian manifold and d is the intrinsic distance on M associated to the Riemannian metric g.

(II) M^n is an immersed submanifold of another manifold X (for example, hyperbolic or Euclidean space) and $d = d_{\mathrm{ext}}$ is the extrinsic distance, that is, the restriction to M of the Riemannian distance on X.

(III) M^n is a bounded domain with smooth boundary in a complete Riemannian manifold X and again $d = d_{\mathrm{ext}}$ is the extrinsic distance.

In the first case we can easily estimate the packing constant in terms of a lower bound of the Ricci curvature and the dimension, using the Bishop–Gromov inequality; see [Colbois and Maerten 2008, Example 2.1]. In cases (II) and (III), a simple argument shows that $C_d(M) \leq C_d(X)^2$, and so the packing constant of an immersed submanifold of Euclidean space (or of a manifold with nonnegative Ricci curvature) is bounded above by an absolute constant depending only on the dimension of X; in particular, it is independent on the Ricci curvature of M. For example, if M is any submanifold of $\mathbb{R}^m$ then $C_d(M) \leq (1 + 3^{2m})^2$. Here d is the extrinsic distance; for the intrinsic distance this is no longer true in general.

1.2. ***Estimates for the Laplacian on functions.*** When the density σ is positive, we can consider the following operator L acting on any $u \in C^\infty(M)$:

$$Lu = \Delta u - \frac{1}{\sigma}\langle \nabla u, \nabla \sigma \rangle. \tag{3}$$

If $\partial M \neq \varnothing$, we assume Neumann boundary conditions. L is self-adjoint when acting on $L^2(M,\mu)$, where $\mu = \sigma\, \mathrm{dvol}_g$, and is associated to the quadratic form

$$u \mapsto \int_M |\nabla u|^2 \sigma\, \mathrm{dvol}_g\,.$$

The spectrum of L is discrete and will be denoted by $\{\lambda_k(L)\}_{k=1}^{\infty}$. Note that $\lambda_1(L) = 0$ and $\lambda_2(L) > 0$. If σ is constant (that is, μ is just a multiple of the Riemannian measure) one recovers the eigenvalues of the ordinary Laplacian on M. However, the generalization to Laplace-type operators will force us to consider nonconstant densities.

Theorem 1. *Suppose $\mathcal{M} = (M,\mu,d)$ is a metric measured space as defined in Section 1.1 and assume that $\mu = \sigma\, \mathrm{dvol}_g$, with $\sigma > 0$ everywhere on M. Let L be the operator defined in* (3). *Then, for all $k \geq 1$, there exists a set S of k points $x_1, \dots, x_k \in M$ such that*

$$r = 8(k+1)C_d(M)^2 \cdot \frac{\log \lambda_{k+1}(L)}{\sqrt{\lambda_{k+1}(L)}} \quad \textit{implies} \quad \mu(S^r) \geq 1 - r,$$

provided that $\lambda_{k+1}(L) \geq e$. Here $C_d(M)$ is the packing constant defined in (2).

Remarks. The estimate is sharp, in the sense that the decay $\log \lambda/\sqrt{\lambda}$ is optimal as $\lambda = \lambda_{k+1}(L)$ tends to infinity, and cannot be replaced by a function with a faster rate of decrease. We refer to Section A.2 for an explicit example.

If the eigenvalue $\lambda_{k+1}(L)$ is large (so that r is small), then almost all of the measure μ is in the r-neighborhhood of k suitable points: This is the concentration property that we want to emphasize.

There is an equivalent formulation of our estimate in terms of the so-called Lévy–Prokhorov distance between probability measures. If (X,d) is a metric space, $\mathcal{B}(X)$ the borelian σ-algebra and $\mathcal{P}(X)$ the set of the probability measures on X, the *Lévy–Prokhorov distance d_P* between two elements ν_1 and ν_2 of $\mathcal{P}(X)$ is defined as

$$d_P(\nu_1,\nu_2) = \inf\{r > 0 : \nu_1(C) \leq \nu_2(C^r) + r \text{ and } \nu_2(C) \leq \nu_1(C^r) + r \text{ for all } C \in \mathcal{B}(X)\}.$$

See for example [Villani 2009, (6.5), page 97].

The following result is an equivalent formulation of Theorem 1.

Theorem 2. *In the hypothesis of Theorem 1, there exist k points* $x_1, \dots, x_k \in M$ *and weights* $p_1, \dots, p_k \in [0, 1)$ *such that* $\sum p_j = 1$ *and*

$$d_P(\mu, \delta_S) \le 8(k+1)C_d(M)^2 \cdot \frac{\log \lambda_{k+1}(L)}{\sqrt{\lambda_{k+1}(L)}},$$

where $\delta_S = \sum_{i=1}^k p_i \delta_{x_i}$ *and* δ_{x_i} *is the Dirac measure concentrated at the point* x_i.

In particular, for $k = 1$ *there exists a point* $x_1 \in M$ *such that*

$$d_P(\mu, \delta_{x_1}) \le 16 C_d(M)^2 \cdot \frac{\log \lambda_2(L)}{\sqrt{\lambda_2(L)}}.$$

The estimate is sharp: *see Section A.2.*

In other words, when the eigenvalue is large, the measure μ is close, in the Lévy–Prokhorov sense, to a weighted linear combination of the Dirac measures at the points $x_1, \dots, x_k$.

The equivalence between the formulations in Theorem 1 and Theorem 2 will be proved in Section A.1.

Note that Theorems 1 and 2 apply obviously to the Laplacian acting on functions: it suffices to choose $\sigma = 1/\mathrm{Vol}(M)$. In that case the concentration is relative to the (normalized) Riemannian volume.

1.3. ***Estimates for vector bundle Laplacians.*** The next task will be to generalize Theorem 1 when the density σ is only assumed to be nonnegative. For that purpose we introduce, in Section 2, a weaker notion of spectrum and prove the relevant Theorem 4. Besides being interesting in itself, Theorem 4 will lead to a concentration phenomenon of eigensections in the context of Laplacians acting on sections of a vector bundle.

So, consider a vector bundle E over a compact Riemannian manifold (M^n, g) with empty boundary, and denote by ∇ a connection on E that is compatible with the metric g (see [Bérard 1988] for details). An operator D acting on sections of the bundle is said to be of *Laplace-type* if it can be written $D = \nabla^* \nabla + T$, where T is a symmetric endomorphism of the fiber. Then, D is self-adjoint and elliptic. We list its eigenvalues as

$$\lambda_1(D) \le \lambda_2(D) \le \cdots \le \lambda_k(D) \le \cdots$$

and denote by $\{\psi_1, \psi_2, \dots\}$ a corresponding orthonormal basis of eigensections.

Important examples of Laplace-type operators are given by the Laplacian acting on differential forms, by the square of the Dirac operator and by a Schrödinger operator acting on functions. In the first case, T is the curvature term in the classical Bochner–Weitzenböck formula, in the second case it is multiplication by a constant multiple of the scalar curvature, and in the third case T is just the potential.

In the second main theorem we assume a large gap in the spectrum of D and prove that eigensections concentrate their norms near a finite set of points.

Theorem 3. *For each positive integer k there is a set S of k points $x_1, \dots, x_k \in M$ with the following property. Let ψ be any unit L^2-norm linear combination of the first k eigensections of D, and $\mu = |\psi|^2 \, d\mathrm{vol}_g$. Then*

$$r = 25k\left(\frac{k^2(k+1)C_d(M)^2}{\lambda_{k+1}(D) - \lambda_k(D)}\right)^{1/3} \quad \textit{implies} \quad \mu(S^r) \geq 1 - r.$$

Equivalently, the Lévy–Prokhorov distance between μ and a suitable linear combination of the Dirac measures at $x_1, \dots, x_k$ is bounded above by r.

Example. We take D to be the ordinary Laplacian on functions and assume that λ_{k+1} tends to infinity while λ_k is uniformly bounded. Then by Theorem 1 the Riemannian volume concentrates around k suitable points $x_1, \dots, x_k$. Theorem 3 then says that any eigenfunction associated to eigenvalues less than λ_{k+1} will also concentrate its L^2-norm around $x_1, \dots, x_k$.

Example. We take D to be the Laplacian acting on p-forms and assume that the p-th Betti number of M is positive, say $b_p(M) = k > 0$. Then $\lambda_k(D) = 0$ and $\lambda = \lambda_{k+1}(D)$ is the first positive eigenvalue of D. Assume that λ is very large. Then the theorem gives the existence of $b_p(M)$ points such that all harmonic p-forms must concentrate their L^2-norms in a small neighborhood of the union of these points.

We also observe that, in general, a large gap in the spectrum of D does not necessarily imply concentration of the Riemannian volume unless, of course, D is the ordinary Laplacian, or there exist parallel sections (so that the density $\sigma = |\psi|^2$ is constant). We refer to Section A.3 for an explicit example.

The paper is structured as follows: In Section 2 we will prove Theorem 1 and a more general version of it, Theorem 4. In Section 3 we will establish the results for vector bundle Laplacians and prove Theorem 3. The appendix is devoted to the examples, in particular, the sharpness of the estimate given in Theorem 1 and 2.

2. Estimates for functions

2.1. ***A general estimate when the density is only nonnegative.*** We consider a compact manifold M (with or without boundary) endowed with a distance function d and a measure $\mu = \sigma \, d\mathrm{vol}_g$ as in Section 1.1. We first consider the general case in which $\sigma \geq 0$. This will be needed to treat Laplace-type operators, where the density σ will be the squared norm of an eigensection, which can vanish at some points of M. However it is well known from elliptic theory that eigensections can vanish only on sets of measure zero.

Let us then introduce the *weak spectrum* of the metric measured space $\mathcal{M} = (M, \mu, d)$ as follows. First, define the following Rayleigh quotient of the Lipschitz function f (such that $\int_M f^2\mu > 0$):

$$R(f) = \int_M |\nabla f|^2 \mu \Big/ \int_M f^2 \mu.$$

Denote by W_k a vector space of Lipschitz functions on M of finite dimension k. Then, for all integers $k \geq 0$ we define

$$\lambda_{k+1}(\mathcal{M}) \doteq \sup_{W_k} \inf\{R(f) : f \perp W_k\}.$$

It is clear that $\lambda_1(\mathcal{M}) = 0$. It is easy to check that the sequence $\lambda_j(\mathcal{M})$ is non-decreasing.

Having said that, we state the main theorem of this section.

Theorem 4. *Let $\mathcal{M} = (M, \mu, d)$ be as above, with $\mu = \sigma \,\mathrm{dvol}_g$ and $\sigma \geq 0$. Then, for each $k = 1, 2, \ldots$ we can find a set S of k points $x_1, \ldots, x_k \in M$ such that*

$$r = 5\left(\frac{(k+1)C_d(M)^2}{\lambda_{k+1}(\mathcal{M})}\right)^{1/3} \quad \textit{implies} \quad \mu(S^r) \geq 1 - r.$$

Remark. If the density σ is strictly positive on M, then it is clear by the max-min principle that the weak spectrum of $\mathcal{M}$ is equal to the spectrum of the self-adjoint elliptic operator L acting on $L^2(M, \sigma \cdot \mathrm{dvol}_g)$ and already defined in (3). That is, $\lambda_k(\mathcal{M}) = \lambda_k(L)$ for all k. In this case, using an upper bound of [Chung et al. 1997] and an additional measure theoretic lemma proved in [Colbois and Maerten 2008] we can prove Theorem 1, which is an improvement of Theorem 4 for large $\lambda = \lambda_{k+1}$ because $\log \lambda/\sqrt{\lambda}$ decays faster than $\lambda^{-1/3}$.

2.2. ***Preparatory results.*** In the next lemma we estimate the eigenvalues of $\mathcal{M}$ as defined in the previous section. The first part follows from a standard argument involving plateau functions, which applies to our case. The second part is an estimate due to Chung, Grigor'yan and Yau.

Lemma 5. (a) *Let $\mathcal{M} = (M, \mu, d)$ and assume that $\mu = \sigma \cdot \mathrm{dvol}_g$ with $\sigma \geq 0$. Assume that there exist $k+1$ subsets of M, each of measure at least $\alpha > 0$, which are $2r$-separated (meaning that the distance between any two of the given sets is at least $2r$). Then $\lambda_{k+1}(\mathcal{M}) \leq 1/\alpha r^2$.*

(b) *If the density σ is strictly positive on M, then*

$$\lambda_{k+1}(\mathcal{M}) = \lambda_{k+1}(L) \leq \log^2(2/\alpha)/r^2,$$

where L is the operator $Lu = \Delta u - \langle \nabla u, \nabla \sigma \rangle/\sigma$ defined in (3).

Proof. (a) Fix a subspace W of the space of Lipschitz functions on M, of finite dimension k. Let $A_1, \dots, A_{k+1}$ be the subsets satisfying the assumptions, that is, $\int_{A_j} \mu = \int_{A_j} \sigma \,\mathrm{dvol}_g \geq \alpha$ and $d(A_i, A_j) \geq 2r$ if $i \neq j$. For each $j = 1, \dots, k+1$, let ϕ_j be the plateau function

$$\phi_j(x) = \begin{cases} 1 & \text{on } A_j, \\ 1 - d(x, A_j)/r & \text{on } \Omega_j = A_j^r \setminus A_j, \\ 0 & \text{on the complement of } A_j^r. \end{cases}$$

Note that the ϕ_j are disjointly supported. Linear algebra shows that we can find numbers $a_1, \dots, a_{k+1}$ such that the function $\phi = \sum_{j=1}^{k+1} a_j \phi_j$ is Lipschitz, $L^2(\mu)$-orthogonal to W and nonzero. We can also assume that $\sum a_j^2 = 1$. The gradient of ϕ is supported on the union of the Ω_j, and on Ω_j one has $|\nabla\phi| \leq |a_j|/r$ almost everywhere. Then

$$\int_M |\nabla\phi|^2 \mu \leq \frac{1}{r^2} \int_M \mu = \frac{1}{r^2}$$

On the other hand,

$$\int_M \phi^2 \mu \geq \sum_j a_j^2 \int_{A_j} \mu \geq \alpha.$$

Therefore $R(\phi) \leq 1/(\alpha r^2)$. Since ϕ was orthogonal to W, we get

$$\inf\{R(f) : f \perp W\} \leq 1/(\alpha r^2).$$

The right side is independent of the subspace W; hence taking the supremum over all k-dimensional subspaces W does not change the upper bound. Recalling the definition of λ_{k+1}, one obtains the first part of the lemma.

(b) If the density σ is positive, we can use an estimate of Chung, Grigor'yan and Yau [1996]. It says that, if the subsets $A_1, \dots, A_{k+1}$ are at distance at least s from each other, then

$$\lambda_{k+1}(L) \leq \frac{4}{s^2} \cdot \max_{i \neq j} \left(\log \frac{2}{\sqrt{\mu(A_i)\mu(A_j)}} \right)^2.$$

The second inequality is now immediate by taking $s = 2r$. □

We will use [Colbois and Maerten 2008, Corollary 2.3], which we state in a way more convenient to our purposes. Consider our metric space (M, d) and recall the packing constant $C_d(M)$. Let ν be any measure on M.

Proposition 6. *Let N be a positive integer. Suppose that for a given $s > 0$, we have for each $x \in M$*

$$\nu(B(x, s)) \leq \frac{\nu(M)}{4C_d(M)^2 N}.$$

Then, there exist N subsets $A_1, \dots, A_N$ of M such that $\nu(A_i) \geq \nu(M)/(2C_d(M)N)$ for each i and $d(A_i, A_j) \geq 3s$ for each $i \neq j$.

We will use the proposition in the proof of Theorem 4 for ν given by the restriction of μ to a closed subset.

Proof of Theorem 4. Let $\lambda_{k+1}(\mathcal{M}) = \lambda$ and assume that it is positive. Let

$$r = 5\left(\frac{(k+1)C_d(M)^2}{\lambda}\right)^{1/3}.$$

We will prove that there exist a set S of suitably chosen points $x_1, \ldots, x_k$ (not necessarily distinct) such that

$$\mu(S^r) \geq 1 - r. \tag{4}$$

We can suppose $r < 1$.

Let $\alpha = r/(4(k+1)C_d(M)^2)$. By the definitions of r and α one has

$$\lambda = \frac{125}{4\alpha r^2}. \tag{5}$$

Step 1 (construction of the points). Choose x_1 so that $\mu(B(x_1, \frac{1}{4}r)) \geq \mu(B(x, \frac{1}{4}r))$ for all $x \in M$, and set

$$X_1 = B(x_1, r)^c.$$

Next, choose $x_2 \in X_1$ so that $\mu(B(x_2, \frac{1}{4}r)) \geq \mu(B(x, \frac{1}{4}r))$ for all $x \in X_1$, and set

$$X_2 = (B(x_1, r) \cup B(x_2, r))^c\,.$$

We continue in this way until we obtain k points $x_1, \ldots, x_k$: To construct the j-th point $x_j \in X_{j-1}$, we demand that $\mu(B(x_j, \frac{1}{4}r)) \geq \mu(B(x, \frac{1}{4}r))$ for all $x \in X_{j-1}$ and define

$$X_j = (B(x_1, r) \cup \cdots \cup B(x_j, r))^c.$$

Note that if X_j is empty for some $j \leq k$, then $\mu(B(x_1, r) \cup \cdots \cup B(x_j, r)) = 1 > 1 - r$, so we can take $S = \{x_1, \ldots, x_{j-1}\}$. We have $\mu(S^r) \geq 1 - r$ and the theorem is proved. So we can assume that

$$X_k = (B(x_1, r) \cup \cdots \cup B(x_k, r))^c$$

is nonempty. Inequality (4) (and the theorem) follows if we show that $\mu(X_k) \leq r$.

Step 2 (proof that $\mu(X_k) \leq r$). We argue by contradiction and show that the inequality

$$\mu(X_k) > r \tag{6}$$

cannot occur. Let us then assume (6) and denote by B_i the ball $B(x_i, \frac{1}{4}r)$. By construction, the sets $B_1, \ldots, B_k$ and X_k are $\frac{1}{2}r$-separated and

$$\mu(B_1) \geq \mu(B_2) \geq \cdots \geq \mu(B_k).$$

First case. Assume $\mu(B_k) \geq \alpha$. Then $\mu(B_j) \geq \alpha$ for all j; moreover

$$\mu(X_k) \geq r > \frac{r}{4(k+1)C_d(M)^2} = \alpha$$

simply because $C_d(M) \geq 1$. Therefore the sets $B_1, \dots, B_k, X_k$ are $\frac{1}{2}r$-separated and each of them has measure at least α. By Lemma 5,

$$\lambda = \lambda_{k+1}(\mathcal{M}) \leq 16/(\alpha r^2), \tag{7}$$

which contradicts (5). Then the first case does not occur.

Second case. Assume $\mu(B_k) < \alpha$. Consider the closed subset $X = X_{k-1}$. By the definition of x_k, one has

$$\mu(B(x, \tfrac{1}{4}r)) \leq \mu(B_k) \leq \alpha \quad \text{for all } x \in X.$$

Recall that $X_k \subseteq X_{k-1} = X$.

We now consider the metric space (M, d) with the measure ν given by the restriction of μ to the closed subspace X, that is, $\nu(A) = \mu(A \cap X)$. By (6) we have $r < \mu(X_k) \leq \mu(X) = \nu(M)$ and therefore

$$\nu(B(x, \tfrac{1}{4}r) \leq \mu(B(x, \tfrac{1}{4}r) \leq \alpha = \frac{r}{4(k+1)C_d(M)^2} \leq \frac{\nu(M)}{4(k+1)C_d(M)^2}.$$

By Proposition 6 applied for $s = \frac{1}{4}r$ and $N = k+1$, we conclude there exist $k+1$ subsets $A_1, \dots, A_k$ that are $\frac{3}{4}r$-separated and satisfy

$$\nu(A_i) \geq \frac{\nu(M)}{2C_d(M)(k+1)} > \frac{r}{2C_d(M)(k+1)} \geq 2C_d(M)\alpha \geq 2\alpha \quad \text{for all } i.$$

Then $\mu(A_i) \geq 2\alpha$ for all i. Applying Lemma 5, one would obtain

$$\lambda = \lambda_{k+1}(\mathcal{M}) \leq \frac{32}{9\alpha r^2}, \tag{8}$$

which again contradicts (5). The proof of Theorem 4 is now complete. □

Proof of Theorem 1. Set $\lambda_{k+1}(\mathcal{M}) = \lambda$ and assume $\lambda \geq e$. Let

$$r = \frac{\beta \log \lambda}{\sqrt{\lambda}}, \tag{9}$$

where $\beta = 8(k+1)C_d(M)^2$. We will find a set S of k points $x_1, \dots, x_k$ such that

$$\mu(S^r) \geq 1 - r, \tag{10}$$

which is the statement of the theorem.

Set $\alpha = r/(4(k+1)C_d(M)^2)$. We first observe that

$$\lambda > \frac{256}{r^2} \log^2(2/\alpha). \tag{11}$$

In fact (9) gives $\lambda = \beta^2 \log^2 \lambda^2/r^2 \geq \beta^2/r^2$, and substituting inside $\log \lambda$ we get (11) because $\beta/r = 2/\alpha$ by the definitions of α and β and the fact that $\beta \geq 8$.

To show (10) we follow Step 1 and Step 2 exactly as in the proof of the previous theorem: We construct the points $x_1, \dots, x_k$ as before and show that the inequality $\mu(X_k) > r$ leads to a contradiction with the inequality (11). The only change is to use the second inequality of Lemma 5 instead of the first, so that (7) and (8) respectively become

$$\lambda \leq \frac{16}{r^2} \log^2(2/\alpha) \quad \text{and} \quad \lambda \leq \frac{64}{9r^2} \log^2(2/\alpha),$$

both of which contradict (11). □

Remark. It is not possible to replace the constant β in (9) by $\beta(\lambda)$ for a function $\beta(\lambda) \to 0$ as $\lambda \to \infty$. In fact, taking $\beta =$ constant is the optimal choice for the radius r; see Section A.2.

3. The estimate for Laplace-type operators

In this section we prove Theorem 3.

Theorem 7. *Let M^n be a compact Riemannian manifold without boundary and D any Laplace-type operator on M. Fix integers i and k with $i \leq k$ and consider the m-m-space (M, μ_i, d), where $\mu_i = |\psi_i|^2 \cdot \mathrm{d}\mathrm{vol}_g$ and ψ_i is a unit norm eigensection associated to $\lambda_i(D)$. Then there exists a set S_i of k points $x_1^i, \dots, x_k^i \in M$ such that*

$$r = 5\left(\frac{k(k+1)C_d(M)^2}{\lambda_{k+1}(D) - \lambda_i(D)}\right)^{1/3} \quad \textit{implies} \quad \mu_i(S_i^r) \geq 1 - r.$$

Of course, the result is significant only when the gap $\lambda_{k+1}(D) - \lambda_i(D)$ is large enough. As the gap $\lambda_{k+1}(D) - \lambda_k(D)$ increases to ∞, we see that any eigensection associated to $\lambda_i(D)$, with $i \leq k$, tends to concentrate its norm around at most k points $x_1^i, \dots, x_k^i$, a priori depending on i. It is natural to ask if there is a relation between all these points for different eigenvalues. We can in fact show that, as the gap tends to infinity, all squared norms $|\psi_1|^2, \dots, |\psi_k|^2$ will concentrate around a *common* set of k points. Actually, we will show that this also happens for the squared norm of any section in the direct sum of the first k eigenspaces; this is the statement of Theorem 3.

Proof of Theorem 7. The proof depends on the following two lemmas, in which we bound the gaps in the spectrum of D by the weak spectrum of the m-m-spaces $\mathcal{M}$ corresponding to the densities $\sigma = |\psi|^2$, where ψ is an eigensection of D. We then apply Theorem 4 to conclude.

Recall that $D = \nabla^*\nabla + T$, where T is a symmetric endomorphism of the fiber.

So the quadratic form associated to D is

$$\mathfrak{Q}(\psi)=\int_M |\nabla\psi|^2+\langle T\psi,\psi\rangle,$$

which is defined on the space of H^1-sections of the bundle (here integration is with respect to the Riemannian measure $d\mathrm{vol}_g$). We fix an orthonormal basis of eigensections of D and denote it by $(\psi_1,\psi_2,\dots)$.

Lemma 8. *Let f be a Lipschitz function on M and ψ a smooth section of the bundle. Then*

$$\mathfrak{Q}(f\psi)=\int_M f^2\langle D\psi,\psi\rangle+|\nabla f|^2|\psi|^2.$$

Lemma 9. *Fix a positive integer k and let $i\le k$. Let ψ_i be an eigensection associated to $\lambda_i(D)$, of unit L^2-norm, and consider the m-m-space $\mathcal{M}_i=(M,\mu_i,d)$ where $\mu_i=|\psi_i|^2\,d\mathrm{vol}_g$. Then*

$$\lambda_{k+1}(D)-\lambda_i(D)\le k\lambda_{k+1}(\mathcal{M}_i).$$

Theorem 7 now follows immediately from Lemma 9 and Theorem 4 applied with the density $\sigma=|\psi_i|^2$. □

Proof of Lemma 8. On the subset where ∇f exists (hence almost everywhere on M), one has

$$|\nabla(f\psi)|^2=|\nabla f|^2|\psi|^2+f^2|\nabla\psi|^2+2f\langle\nabla_{\nabla f}\psi,\psi\rangle.$$

Now

$$\int_M 2f\langle\nabla_{\nabla f}\psi,\psi\rangle=\int_M \tfrac12\langle\nabla f^2,\nabla|\psi|^2\rangle=\int_M \tfrac12 f^2\Delta|\psi|^2,$$

and hence

$$\begin{aligned}\mathfrak{Q}(f\psi)&=\int_M|\nabla(f\psi)|^2+\langle T(f\psi),f\psi\rangle\\&=\int_M f^2(|\nabla\psi|^2+\tfrac12\Delta|\psi|^2+\langle T\psi,\psi\rangle)+|\nabla f|^2|\psi|^2.\end{aligned}$$

Now recall the identity (Bochner formula) $\langle D\psi,\psi\rangle=|\nabla\psi|^2+\frac12\Delta|\psi|^2+\langle T\psi,\psi\rangle$. The lemma follows. □

Proof of Lemma 9. Given the metric-measure space $\mathcal{M}=(M,\mu,d)$, recall the definition of weak spectrum:

$$\lambda_{h+1}(\mathcal{M})=\sup_{W_h}\inf\{R(f):f\perp W_h\},\quad \text{where } R(f)=\int_M|\nabla f|^2\mu\Big/\int_M f^2\mu,$$

and W_h denotes a vector subspace of Lipschitz functions having dimension h. We will write for brevity $\lambda_i(\mathcal{M})=\lambda_i$.

Fix $\epsilon > 0$. Then, for all integers $k \in \mathbb{N}$ we construct a $(k+1)$-dimensional subspace W_{k+1} of the space of Lipschitz functions on M such that, for all $f \in W_{k+1}$,

$$R(f) \leq k(\lambda_{k+1} + \epsilon). \tag{12}$$

Set $W_1 = \text{span}(f_1)$, where f_1 is the constant function 1. By definition, there exists a nonvanishing smooth function f_2 that is orthogonal to W_1 and satisfies

$$R(f_2) \leq \lambda_2 + \epsilon.$$

Set $W_2 = \text{span}(f_1, f_2)$. We can assume that f_2 has unit L^2-norm. Continuing this process, we get $W_{k+1} = \text{span}(f_1, \ldots, f_{k+1})$, where $(f_1, \ldots, f_{k+1})$ is an orthonormal set and, for all $j = 1, \ldots, k+1$,

$$R(f_j) \leq \lambda_j + \epsilon \leq \lambda_{k+1} + \epsilon. \tag{13}$$

Let us prove (12). Let $f = \sum_{i=1}^{k+1} a_i f_i$ be a function in W_{k+1}. We can assume that it has unit norm, so that $\sum_i a_i^2 = 1$. By the triangle inequality, since $\nabla f_1 = 0$, one has $|\nabla f| \leq \sum_{i=2}^{k+1} |a_i||\nabla f_i|$. By the Schwarz inequality, $|\nabla f|^2 \leq \sum_{i=2}^{k+1} |\nabla f_i|^2$ and therefore, by (13),

$$R(f) \leq \sum_{i=2}^{k+1} R(f_i) \leq k(\lambda_{k+1} + \epsilon).$$

We can now prove the lemma. Fix $\epsilon > 0$ and consider the m-m-space $\mathcal{M}_i$ with measure $\mu_i = |\psi_i|^2 \, d\text{vol}_g$, as in the statement of the lemma. Let W_{k+1} be the subspace satisfying (12). By linear algebra, we can find a nonvanishing $f \in W_{k+1}$ such that the section $f\psi_i$ has unit norm and is orthogonal to the first k eigensections $\psi_1, \ldots, \psi_k$ of the spectrum of D. Using $f\psi_i$ as test-section for the eigenvalue $\lambda_{k+1}(D)$, we obtain by Lemma 8

$$\lambda_{k+1}(D) \leq \mathfrak{Q}(f\psi_i) = \int_M f^2 \langle D\psi_i, \psi_i \rangle + |\nabla f|^2 |\psi_i|^2.$$

Since $\langle D\psi_i, \psi_i \rangle = \lambda_i(D)|\psi_i|^2$, this becomes

$$\lambda_{k+1}(D) - \lambda_i(D) \leq R(f) \leq k(\lambda_{k+1}(\mathcal{M}_i) + \epsilon),$$

by (12). Letting $\epsilon \to 0$ we obtain the assertion. □

Proof of Theorem 3. Let us start with the formal proof by considering an orthonormal basis $(\psi_1, \ldots, \psi_k)$ of the direct sum of the first k eigenspaces of D. Given $\mu_j = |\psi_j|^2 \cdot d\text{vol}_g$, let us introduce the following auxiliary measure, which is just the average of the μ_j:

$$\tilde{\mu} = \frac{1}{k} \sum_{j=1}^{k} \mu_j.$$

We also fix the radius

$$r = 5\left(\frac{k^2(k+1)C_d(M)^2}{\lambda_{k+1}(D)-\lambda_k(D)}\right)^{1/3}. \tag{14}$$

The theorem follows from two claims.

Claim 1. *There exists a set of points $Q = \{y_1, \ldots, y_l\}$ with the property that*

$$\tilde{\mu}(B(y_j, r)) \geq r/k^2 \quad \textit{for all } j \textit{ and } \tilde{\mu}(Q^r) \geq 1-2r.$$

Claim 2. *There exists a subset $T = \{x_1, \ldots, x_m\}$ of Q, with $m \leq k$, such that*

$$\tilde{\mu}(T^{5r}) \geq 1-5r.$$

(This gives a concentration result for the averaged measure $\tilde{\mu}$).

Thanks to Claims 1 and 2, we can conclude as follows. Let $\psi = \sum_{i=1}^k a_i\psi_i$ be any unit norm section in the direct sum of the first k eigenspaces of D (so that $\sum_i a_i^2 = 1$), and let $\mu = |\psi|^2\,d\mathrm{vol}_g$. By the Schwarz inequality we have, at any point,

$$|\psi|^2 \leq \left(\textstyle\sum_i |a_i||\psi_i|\right)^2 \leq \textstyle\sum_i |\psi_i|^2,$$

that is, $\mu \leq k\tilde{\mu}$. We deduce $\mu((T^{5kr})^c) \leq \mu((T^{5r})^c) \leq k\tilde{\mu}((T^{5r})^c \leq 5kr$ by Claim 2. We now take $S = T$. Then $\mu(S^{5kr}) \geq 1-5kr$ and the theorem follows. □

For the proof of the two claims we need a lemma. We can assume $r < 1/5$.

Lemma 10. *Assume there exist $k+1$ subsets $A_1, \ldots, A_{k+1}$ that are $2r$-separated and have $\tilde{\mu}$-measure at least β. Then*

$$\lambda_{k+1}(D) - \lambda_k(D) \leq \frac{k}{\beta r^2}.$$

Proof. As in the proof of Lemma 5, we can construct $k+1$ disjointly supported, plateau functions $f_1, \ldots, f_{k+1}$ with $R_{\tilde{\mu}}(f_j) \leq 1/(\beta r^2)$ for each j, where $R_{\tilde{\mu}}$ is the Rayleigh quotient relative to the measure $\tilde{\mu}$. Since $\tilde{\mu}$ is the average of the μ_j, we see that for any nonnegative function f there is an index i (depending on f) such that $\int_M f\tilde{\mu} \leq \int_M f\mu_i$. Therefore, for each $j = 1, \ldots, k+1$ there is an index $\alpha(j) = 1, \ldots, k$ such that

$$R_{\tilde{\mu}}(f_j) = \frac{\int_M |\nabla f_j|^2\tilde{\mu}}{\int_M f_j^2\tilde{\mu}} \geq \frac{1}{k}\frac{\int_M |\nabla f_j|^2\mu_{\alpha(j)}}{\int_M f_j^2\mu_{\alpha(j)}} \geq \frac{1}{k}\, R_{\mu_{\alpha(j)}}(f_j)$$

and then $R_{\mu_{\alpha(j)}}(f_j) \leq k/(\beta r^2)$ for all j. We consider the sections $s_j = f_j\psi_{\alpha(j)}$ for $j = 1, \ldots, k+1$; they are disjointly supported and we can use them as test-sections for the eigenvalue $\lambda_{k+1}(D)$. Using Lemma 8 one sees that

$$\lambda_{k+1}(D) - \lambda_k(D) \leq \sup_j\{R_{\mu_{\alpha(j)}}(f_j)\} \leq k/(\beta r^2).$$ □

Proof of Claim 1. For all $j \le k$ we observe from (14) that

$$r \ge 5\Big(\frac{k(k+1)C_d(M)^2}{\lambda_{k+1}(D)-\lambda_j(D)}\Big)^{1/3}.$$

So, by Theorem 7, there exist finite subsets $S_1, \dots, S_k \subseteq M$ of cardinality less than or equal to k such that $\mu_j(S_j^r) \ge 1-r$ for all j. We set $P = S_1 \cup \dots \cup S_k$ and observe that, by the definition of $\tilde{\mu}$,

$$\tilde{\mu}(P^r) \ge 1-r. \tag{15}$$

We now consider the subset $Q = \{y_1, \dots, y_l\}$ formed by all points $y_j \in P$ such that $\tilde{\mu}(B(y_j, r)) \ge r/k^2$. Let $Q' = P \setminus Q$. Then by definition $\tilde{\mu}((Q')^r) \le r$. Since $\tilde{\mu}((Q')^r) + \tilde{\mu}(Q^r) \ge 1-r$ by (15), we obtain

$$\tilde{\mu}(Q^r) \ge 1-2r \tag{16}$$

as claimed. Note that Q is not empty because $r < 1/5$ by assumption. □

Proof of Claim 2. We construct the subset $T = \{x_1, \dots, x_m\}$ of Q as follows. Set $x_1 = y_1$. If there exists some point $y_j \in Q$ in the complement of $B(x_1, 4r)$, we select it and denote it by x_2. Next, if there exists a point of Q in the complement of $B(x_1, 4r) \cup B(x_2, 4r)$, we select it and denote it by x_3, and so on. We iterate the process until it is possible, and obtain after $m \le l$ steps the required subset T.

Assume that $m \ge k+1$. Then the balls $A_j = B(x_j, r)$ with $j = 1, \dots, k+1$ are $2r$-separated by construction, and have $\tilde{\mu}$-measure at least equal to $\beta = r/k^2$. By Lemma 10 we see that

$$\lambda_{k+1}(D) - \lambda_k(D) \le k^3/r^3. \tag{17}$$

However, the definition (14) of r gives $\lambda_{k+1}(D) - \lambda_k(D) = c/r^3$ with the constant $c = 125k^2(k+1)C_d(M)^2 > k^3$ and we get a contradiction with (17).

Therefore $m \le k$.

By the construction of T, every point $y_j \in Q$ is at distance not greater than $4r$ to some point of T, that is, $Q \subseteq T^{4r}$. By the triangle inequality $Q^r \subseteq T^{5r}$ and therefore, by (16)

$$\tilde{\mu}(T^{5r}) \ge \tilde{\mu}(Q^r) \ge 1-2r > 1-5r,$$

and Claim 2 follows. □

Appendix

A.1. ***Facts about the Lévy–Prokhorov distance.*** Recall that the Lévy–Prokhorov distance d_P between two probability measures defined on the same metric space (M, d) is

$$d_P(\nu_1, \nu_2) = \inf\{r > 0 : \nu_1(C) \le \nu_2(C^r) + r \text{ and } \nu_2(C) \le \nu_1(C^r) + r \text{ for all } C\}.$$

Proposition 11. *Let (M, μ, d) be an m-m-space, and let $S = \{x_1, \dots, x_k\}$ be a set of k points in M and $r > 0$. Then $\mu(S^r) \geq 1 - r$ if and only if there exist weights $p_1, \dots, p_k \in [0, 1)$ such that $\sum p_j = 1$ and $d_P(\mu, \delta) \leq r$, where $\delta = \sum_{i=1}^k p_i \delta_{x_i}$ and δ_{x_i} is the Dirac measure concentrated at the point x_i.*

Proof. Suppose first that $d_P(\mu, \delta) \leq r$. Then, choosing $C = S$ in the definition of d_P, we have $1 = \delta(S) \leq \mu(S^r) + r$ and therefore $\mu(S^r) \geq 1 - r$.

To prove the converse, we assume $\mu(S^r) \geq 1 - r$. We first define the weights p_i. Denote by B_i the ball $B(x_i, r)$ and consider the sets $\{A_i\}_{i=1}^k$ defined by

$$\begin{cases} A_1 = B_1, \\ A_i = B_i \cap (B_1 \cup \dots \cup B_{i-1})^c & \text{for } i \geq 2. \end{cases}$$

Then $A_i \subseteq B_i$ and $A_i \cap A_j = \varnothing$ if $i \neq j$. Set $A = A_1 \cup \dots \cup A_k$. Then $A = B_1 \cup \dots \cup B_k = S^r$, so that $\mu(A) = \mu(S^r) \geq 1 - r$.

We now choose the weights $p_i = \mu(A_i)/\mu(A)$.

The proof is complete if we show that, for each Borel subset C, we have

$$\begin{cases} \delta(C) \leq \mu(C^r) + r, \\ \mu(C) \leq \delta(C^r) + r. \end{cases} \tag{18}$$

We can order the points so that $x_1, \dots, x_t \in C$ and $x_j \notin C$ for $j = t+1, \dots, k$. Then $\delta(C) = p_1 + \dots + p_t$. Now $B_1 \cup \dots \cup B_t \subseteq C^r$; since $A_i \subseteq B_i$ and the A_i are pairwise disjoint, we have

$$\mu(A_1) + \dots + \mu(A_t) \leq \mu(B_1 \cup \dots \cup B_t) \leq \mu(C^r).$$

Then

$$\begin{aligned} \delta(C) = p_1 + \dots + p_t &= \frac{\mu(A_1) + \dots + \mu(A_t)}{\mu(A)} \\ &= \mu(A_1) + \dots + \mu(A_t) + \frac{\mu(A_1) + \dots + \mu(A_t)}{\mu(A)}(1 - \mu(A)) \\ &\leq \mu(C^r) + 1 - \mu(A) \\ &\leq \mu(C^r) + r, \end{aligned}$$

which proves the first inequality in (18).

For the second, write

$$\mu(C) = \mu(C \cap A_1) + \dots + \mu(C \cap A_k) + \mu(C \cap A^c)$$

and note that $x_i \in C^r$ if $C \cap A_i \neq \varnothing$. Since $\mu(C \cap A_i) \leq \mu(A_i) = p_i \mu(A) \leq p_i$ and $\mu(C \cap A^c) \leq \mu(A^c) \leq r$, we have

$$\mu(C) \leq \sum_{i : x_i \in C^r} p_i + r \leq \delta(C^r) + r.$$

□

A.2. ***Theorem 1 is sharp.*** For $R > 0$, let M_R be the surface of revolution in $\mathbb{R}^3$:

$$M_R = \{(x, y, z) \in \mathbb{R}^3 : y^2 + z^2 = e^{-2Rx}/R^2, x \in [0, 1]\},$$

and consider the metric measure space (M_R, μ, d), where μ is the normalized Riemannian measure and d is the extrinsic distance inherited from $\mathbb{R}^3$. By a calculation in [Friedman and Tillich 2000], one knows that

$$\lambda_2(M_R) \geq \tfrac{1}{8} R^2 \tag{19}$$

(we take the Neumann boundary conditions). By the equivalent formulation of Theorem 1, given in Theorem 2, for each R there exists a point $p \in M_R$ such that

$$d_P(\mu, \delta_p) \leq \gamma_R \frac{\log \lambda_R}{\sqrt{\lambda_R}}$$

for the constant $\gamma_R = 16 C_d(M_R)^2$, where we set $\lambda_R = \lambda_2(M_R)$. However, since we use the extrinsic distance, the constant γ_R admits a uniform upper bound by the packing constant of $\mathbb{R}^3$ (see Section 1.1); hence

$$d_P(\mu, \delta_p) \leq \gamma \frac{\log \lambda_R}{\sqrt{\lambda_R}} \tag{20}$$

for some $p \in M_R$ and an absolute constant γ (we can take in fact $\gamma = 16(1+3^6)^2$).

Now, when R goes to ∞ the first positive eigenvalue λ_R goes to ∞ by (19). Therefore, by (20), the normalized Riemannian measure μ concentrates at some point of M_R: This is quite evident and can be verified directly from the definition of M_R, because the limit metric measure space as $R \to \infty$ (in any reasonable sense) is the unit interval [0, 1] endowed with its canonical distance and the Dirac measure supported at 0. In fact, one can check that the relative measure of a set at positive distance α from the circle $\{x = 0\}$ decreases to zero like $e^{-\alpha R}$.

In this section we show that, apart from the constant γ, the inequality (20) is actually sharp.

Theorem 12. *Let M_R and λ_R be as above. Then there exists R_0 such that, for all $R \geq R_0$ and for all $q \in M_R$, one has*

$$d_P(\mu, \delta_q) \geq \frac{1}{48} \frac{\log \lambda_R}{\sqrt{\lambda_R}}.$$

Lemma 13. *Assume that there exist two subsets A and B with relative volume at least s, and such that $d(A, B) \geq 2s$. Then $d_P(\mu, \delta_q) \geq s$ for all $q \in M_R$.*

Proof. Assume that there exists $q \in M_R$ such that $d_P(\mu, \delta_q) < s$. One sees from the definition of d_P that $\mu(B(q, s)) > 1 - s$ and therefore $\mu(B(q, s)) + \mu(A) > 1$. So A must intersect $B(q, s)$ and there exists $a \in A$ such that $d(a, q) < s$. Similarly, there exists $b \in B$ with $d(b,q) < s$. Applying the triangle inequality we get a contradiction with the assumption $d(A,B) \geq 2s$. □

Proof of Theorem 12. By (19) one has $\lambda_R > \frac{1}{9}R^2$; hence, for R large,

$$\frac{1}{48}\frac{\log\lambda_R}{\sqrt{\lambda_R}} \leq \frac{1}{8}\frac{\log R}{R}.$$

So, it is enough to show that

$$d_P(\mu,\delta_q) \geq \frac{1}{8}\frac{\log R}{R} \quad \text{for } R \text{ large and for all } q \in M_R.$$

For $L < L'$ in the interval $[0,1]$, consider the strip

$$M_{[L,L']} = \{(x,y,z) \in M_R : L \leq x \leq L'\}.$$

We will apply the lemma, taking

$$A = M_{[0,1/R]}, \quad B = M_{[(1/2)(\log R)/R,1]}, \quad s = \tfrac{1}{8}(\log R)/R.$$

We need the simple volume estimate

$$\mu(M_{[L,L']}) \geq \frac{e^{-LR}-e^{-L'R}}{2(1-e^{-R})}. \tag{21}$$

In fact, observe that M_R is obtained by rotating the curve $y = e^{-Rx}/R$ around the x-axis. Then

$$\mathrm{Vol}(M_{[L,L']}) = \frac{2\pi}{R}\int_L^{L'} e^{-Rx}ds, \quad \text{with } ds = \sqrt{1+e^{-2Rx}}dx.$$

Inequality (21) now follows from observing that $dx \leq ds < 2dx$ and recalling that $\mu(M_{[L,L']}) = \mathrm{Vol}(M_{[L,L']})/\mathrm{Vol}(M_{[0,1]})$.

By the volume estimate in (21),

$$\mu(A) \geq \frac{1-e^{-1}}{2(1-e^{-R})}, \quad \mu(B) \geq \frac{R^{-1/2}-e^{-R}}{2(1-e^{-R})}, \quad d(A,B) \geq \frac{1}{2}\frac{\log R}{R} - \frac{1}{R}.$$

It is now clear that, for $R \geq R_0$ sufficiently large, one has $\mu(A) \geq s$, $\mu(B) \geq s$ and $d(A,B) \geq 2s$. The lemma gives $d_P(\mu,\delta_q) \geq s = \frac{1}{8}(\log R)/R$ and the theorem is proved. □

A.3. ***Example for differential forms.*** We will now construct an example with a large gap on the spectrum of the Laplacian on p-forms, but in which there is no concentration of the Riemannian volume.

Indeed, the construction of large eigenvalues for p-forms is well known; see [Gentile and Pagliara 1995; Guerini 2004; Colbois and El Soufi 2006]. We can easily adapt the construction of Gentile and Pagliara for an hypersurface in $\mathbb{R}^{n+1}$, and we will only briefly sketch it.

We begin with a hypersurface $M_0 \subset \mathbb{R}^{n+1}$, with p-th De Rham cohomology space of a given positive dimension. Then we deform M_0 by adding a long cylinder $[0, L] \times S^{n-1}$ closed by a hemisphere. We denote by M_L this family of manifolds, whose volume is of the order of L as $L \to \infty$. Gentile and Pagliara showed that, for $2 \le p \le n-2$, the nonzero p-forms spectrum of M_L is bounded below by a positive constant C not depending on L.

After renormalisation by a factor of order $L^{-1/n}$, we get a family of constant volume 1, with first nonzero eigenvalue for p-forms going to ∞ with L. Using the extrinsic Euclidean distance, we see that the packing constant is uniformly bounded, and we can conclude that the L^2-norms of the harmonic forms have to concentrate, indeed on the part corresponding to M_0.

However, there is no concentration of the volume; the part M_0 concentrates to a point and the cylinder looks like a homogeneous 1-dimensional cylinder of length $L^{1-1/n}$.

A.4. ***Expanders.*** In this section we construct a family of manifolds $\overline{M}_i$ of fixed dimension n such that $\lambda_2(\overline{M}_i) \to \infty$ but for which there is no concentration of the volume around any point.

We start from an n-dimensional compact, hyperbolic manifold M_i such that $\mathrm{Vol}(M_i) \to \infty$ as $i \to \infty$ and $\lambda_2(M_i) \ge C(n) > 0$, where $C(n)$ is a constant not depending on i. Such examples do exist (see for example [Brooks 1986]), even if their construction, related to the concept of expanders, is not easy. The M_i can be realized as coverings of a fixed manifold. The diameter of M_i is proportional to $\ln \mathrm{Vol}(M_i)$, and hence tends to infinity as $i \to \infty$.

So, if we multiply the metric of M_i by $(\mathrm{diam}(M_i))^{-1}$, and denote by $\overline{M}_i$ the new family of Riemannian manifolds, it is clear that $\lambda_2(\overline{M}_i) \to \infty$ but $\mathrm{diam}\, \overline{M}_i = 1$. Since $\overline{M}_i$ is a covering, the distribution of the volume is uniform, and we see that it cannot concentrate in a neighborhood of a single point. It concentrates however in the sense described in [Chung et al. 1996]: Two sets $A_i, B_i \subset \overline{M}_i$ of volume no less than $\kappa\, \mathrm{Vol}(\overline{M}_i)$ (with a fixed $\kappa > 0$) have to be very close to each other, even if κ is small.

References

[Bérard 1988] P. H. Bérard, "From vanishing theorems to estimating theorems: the Bochner technique revisited", *Bull. Amer. Math. Soc.* (*N.S.*) **19** (1988), 371–406. MR 89i:58152 Zbl 0662.53037

[Brooks 1986] R. Brooks, "The spectral geometry of a tower of coverings", *J. Differential Geom.* **23**:1 (1986), 97–107. MR 87j:58095 Zbl 0576.58033

[Cheng 1975] S. Y. Cheng, "Eigenvalue comparison theorems and its geometric applications", *Math. Z.* **143**:3 (1975), 289–297. MR 51 #14170 Zbl 0329.53035

[Chung et al. 1996] F. R. K. Chung, A. Grigor'yan, and S.-T. Yau, "Upper bounds for eigenvalues of the discrete and continuous Laplace operators", *Adv. Math.* **117**:2 (1996), 165–178. MR 96m:58254 Zbl 0844.53029

[Chung et al. 1997] F. R. K. Chung, A. Grigor'yan, and S.-T. Yau, "Eigenvalues and diameters for manifolds and graphs", pp. 79–105 in *Tsing Hua lectures on geometry & analysis* (Hsinchu, 1990–1991), edited by S.-T. Yau, Int. Press, Cambridge, MA, 1997. MR 98h:58206 Zbl 0890.58093

[Colbois and El Soufi 2006] B. Colbois and A. El Soufi, "Eigenvalues of the Laplacian acting on p-forms and metric conformal deformations", *Proc. Amer. Math. Soc.* **134**:3 (2006), 715–721. MR 2007e:58049 Zbl 1095.35013

[Colbois and Maerten 2008] B. Colbois and D. Maerten, "Eigenvalues estimate for the Neumann problem of a bounded domain", *J. Geom. Anal.* **18**:4 (2008), 1022–1032. MR 2009f:35243 Zbl 1158.58014

[Friedman and Tillich 2000] J. Friedman and J.-P. Tillich, "Laplacian eigenvalues and distances between subsets of a manifold", *J. Differential Geom.* **56**:2 (2000), 285–299. MR 2002g:58050 Zbl 1032.58015

[Gentile and Pagliara 1995] G. Gentile and V. Pagliara, "Riemannian metrics with large first eigenvalue on forms of degree p", *Proc. Amer. Math. Soc.* **123**:12 (1995), 3855–3858. MR 96b:58115 Zbl 0848.53022

[Grigor'yan et al. 2004] A. Grigor'yan, Y. Netrusov, and S.-T. Yau, "Eigenvalues of elliptic operators and geometric applications", pp. 147–217 in *Surveys in differential geometry*, vol. 9, edited by A. Grigor'yan and S. T. Yau, Int. Press, Somerville, MA, 2004. MR 2007f:58039 Zbl 1061.58027

[Gromov and Milman 1983] M. Gromov and V. D. Milman, "A topological application of the isoperimetric inequality", *Amer. J. Math.* **105**:4 (1983), 843–854. MR 84k:28012 Zbl 0522.53039

[Guerini 2004] P. Guerini, "Prescription du spectre du laplacien de Hodge-de Rham", *Ann. Sci. École Norm. Sup.* (4) **37**:2 (2004), 270–303. MR 2005k:58063 Zbl 1068.58016

[Villani 2009] C. Villani, *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften **338**, Springer, Berlin, 2009. MR 2010f:49001 Zbl 1156.53003

Received December 14, 2009.

BRUNO COLBOIS
UNIVERSITÉ DE NEUCHÂTEL
INSTITUT DE MATHÉMATIQUES
RUE EMILE ARGAND 11
CH-2007, NEUCHÂTEL
SWITZERLAND
bruno.colbois@unine.ch

ALESSANDRO SAVO
DIPARTIMENTO DI SCIENZE DI BASE E APPLICATE PER L'INGEGNERIA
SAPIENZA UNIVERSITÀ DI ROMA
VIA ANTONIO SCARPA 16
00161 ROMA
ITALY
savo@dmmm.uniroma1.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

SUR LES CONDITIONS D'EXISTENCE DES FAISCEAUX SEMI-STABLES SUR LES COURBES MULTIPLES PRIMITIVES

JEAN-MARC DRÉZET

On donne des conditions suffisantes pour la (semi-)stabilité des faisceaux sans torsion sur une curve multiple primitive. Ces conditions sont utilisées pour démontrer que certaines variétés de modules de faisceaux stables sont non vides. On étudie surtout les *faisceaux quasi localement libres de type générique*, y inclus les faisceaux localement libres. Ces faisceaux sont *génériques*, c'est-à-dire pour chaque variété de modules de faisceaux sans torsion, les faisceaux de ce type correspondent à un ouvert de la variété.

We give sufficient conditions for the (semi-)stability of torsion free sheaves on a primitive multiple curve. These conditions are used to prove that some moduli spaces of stable sheaves are not empty. We study mainly the *quasi locally free sheaves of generic type* (this includes the locally free sheaves). These sheaves are *generic*, i.e. for every moduli space of torsion free sheaves, the sheaves of this type correspond to an open subset of the moduli space.

1. Introduction

Une *courbe multiple primitive* est une variété algébrique complexe de Cohen–Macaulay qui peut localement être plongée dans une surface lisse, et dont la sous-variété réduite associée est une courbe lisse. Les courbes projectives multiples primitives ont été définies et étudiées pour la première fois par C. Bănică et O. Forster [1986]. Leur classification a été faite dans [Bayer et Eisenbud 1995] pour

MSC2000: primary 14D20; secondary 14H60.
Mots-clefs: multiple curves, moduli spaces, semi-stable sheaves.

les courbes doubles, et dans [Drézet 2007] dans le cas général. Les faisceaux semi-stables sur des variétés non lisses ont déjà été étudiés [Seshadri 1982; Bhosle 1992; 1999; Teixidor i Bigas 1991; 1995; 1998; Inaba 2004; 2002].

On peut espérer en trouver des applications concernant les fibrés vectoriels ou leurs variétés de modules sur les courbes lisses [Eisenbud et Green 1995; Sun 2000; 2002] en faisant dégénérer des courbes lisses vers une courbe multiple primitive. Le problème de la dégénération des courbes lisses en courbes primitives doubles est évoqué dans [González 2006].

Les articles [Drézet 2006; 2009] sont consacrés à l'étude des faisceaux cohérents et de leurs variétés de modules sur les courbes multiples primitives. On donne ici des critères de (semi-)stabilité et des conditions suffisantes d'existence des faisceaux semi-stables sur ces courbes. On appliquera ensuite ces critères à des faisceaux sans torsion génériques. Les conditions d'existence des faisceaux (semi-) stables s'expriment en fonction d'*invariants* de ces faisceaux, parmi lesquels se trouvent le rang et le degré généralisés.

Le cas des faisceaux localement libres est traité. Dans ce cas les seuls invariants sont le rang et le degré généralisés. Les variétés de modules obtenues sont irréductibles.

On considère aussi des faisceaux plus compliqués, les *faisceaux quasi localement libres de type rigide* non localement libres, où il y a deux invariants supplémentaires. Dans ce cas les variétés de modules de faisceaux de rang et degré généralisés fixés peuvent avoir de multiples composantes.

Pour finir on traitera des exemples simples de faisceaux sans torsion non quasi localement libres.

1.1. ***Faisceaux cohérents sur les courbes multiples primitives.*** Soit C une courbe projective lisse irréductible. Soient n un entier tel que $n \geq 2$ et Y une courbe multiple primitive de multiplicité n et de courbe réduite associée C. Si $\mathcal{I}_C$ est le faisceau d'idéaux de C dans Y,

$$L = \mathcal{I}_C / \mathcal{I}_C^2$$

est un fibré en droites sur C, dit *associé* à Y. Dans cet article on supposera que $\deg(L) < 0$. Le cas où $\deg(L) \geq 0$ est moins intéressant car les seuls faisceaux stables sont alors les fibrés vectoriels stables sur C.

Pour $1 \leq i \leq n$ on note C_i le sous-schéma de Y défini par le faisceau d'idéaux $\mathcal{I}_C^i$. C'est une courbe multiple primitive de multiplicité i et on a une filtration

$$C = C_1 \subset \cdots \subset C_n = Y.$$

On notera $\mathcal{O}_i = \mathcal{O}_{C_i}$.

Le faisceau $\mathcal{I}_C$ est localement libre de rang 1 sur C_{n-1}. Il existe un fibré en droites $\mathbb{L}$ sur C_n tel que $\mathbb{L}_{|C_{n-1}} = \mathcal{I}_C$. Pour tout faisceau cohérent $\mathcal{E}$ sur C_n on a donc un morphisme canonique

$$\mathcal{E} \otimes \mathbb{L} \longrightarrow \mathcal{E}$$

dont le noyau et le conoyau sont indépendants du choix de $\mathbb{L}$.

Si $\mathcal{F}$ est un faisceau cohérent sur Y on note $\mathcal{F}_i$ le noyau de la restriction $\mathcal{F} \to \mathcal{F}_{|C_i}$, $\mathcal{F}^{(i)}$ celui du morphisme canonique $\mathcal{F} \to \mathcal{F} \otimes \mathbb{L}^{-i}$. On a des suites exactes canoniques

$$0 \longrightarrow \mathcal{F}_i \longrightarrow \mathcal{F} \longrightarrow \mathcal{F}_{|C_i} \longrightarrow 0,$$
$$0 \longrightarrow \mathcal{F}^{(i)} \longrightarrow \mathcal{F} \longrightarrow \mathcal{F}_i \otimes \mathbb{L}^{-i} \longrightarrow 0.$$

Les quotients $G_i(\mathcal{F}) = \mathcal{F}_i/\mathcal{F}_{i+1}$, $0 \leq i < n$, sont des faisceaux sur C. Ils permettent de définir le *rang généralisé* et le *degré généralisé* de $\mathcal{F}$:

$$R(\mathcal{F}) = \sum_{i=0}^{n-1} \operatorname{rg} G_i(\mathcal{F}), \qquad \operatorname{Deg}(\mathcal{F}) = \sum_{i=0}^{n-1} \deg(G_i(\mathcal{F})).$$

Ce sont des invariants par déformation ; voir le section 2.3 et [Drézet 2006; 2009]. Si $R(\mathcal{F}) > 0$, le nombre rationnel

$$\mu(\mathcal{F}) = \frac{\operatorname{Deg}(\mathcal{F})}{R(\mathcal{F})}$$

s'appelle la *pente* de $\mathcal{F}$.

Pour $1 \leq i < n$, on note $\mathcal{F}[i]$ le noyau du morphisme canonique surjectif

$$\mathcal{F} \longrightarrow\!\!\!\!\!\rightarrow \mathcal{F}_{|C_i} \longrightarrow\!\!\!\!\!\rightarrow (\mathcal{F}_{|C_i})^{\vee\vee}.$$

1.1.1. *Faisceaux quasi localement libres.* On dit qu'un faisceau cohérent $\mathcal{E}$ sur Y est *quasi localement libre* s'il existe des entiers $m_1, \ldots, m_n$ non négatifs tels que $\mathcal{E}$ soit localement isomorphe à

$$\bigoplus_{i=1}^{n} m_i \mathcal{O}_i.$$

Les entiers m_i sont alors uniquement déterminés.

1.1.2. *Faisceaux quasi localement libres de type rigide.* On renvoie le lecteur à [Drézet 2009]. Si $\mathcal{E}$ est quasi localement libre on dit qu'il est *de type rigide* s'il est localement libre, ou s'il existe un entier k, $1 \leq k \leq n-1$, tel que $m_k = 1$ et $m_j = 0$ pour $j \neq k$. Donc un faisceau quasi localement libre de type rigide non localement libre est localement isomorphe à un faisceau du type $a\mathcal{O}_n \oplus \mathcal{O}_k$, avec $1 \leq k \leq n-1$. L'intérêt de ces faisceaux est que la propriété pour un faisceau d'être quasi localement libre de type rigide est une *propriété ouverte*. En particulier les faisceaux stables localement libres de type rigide de rang généralisé R et de degré généralisé d constituent un ouvert de la variété de modules des faisceaux stables de rang généralisé R et de degré généralisé d sur C_n.

Soit $\mathcal{E}$ un faisceau quasi localement libre de type rigide localement isomorphe à $a\mathbb{O}_n \oplus \mathbb{O}_k$, avec $a \geq 1$, $1 \leq k < n$. Alors les faisceaux $\mathcal{E}_k$ et $\mathcal{E}^{(k)}$ sont localement libres sur C_{n-k} et C_k respectivement. On pose

$$E_{\mathcal{E}} = \mathcal{E}_{|C}, \qquad F_{\mathcal{E}} = \mathcal{E}_{k|C}, \ V_{\mathcal{E}} = (\mathcal{E}^{(k)})_{|C}.$$

Ce sont des fibrés vectoriels sur C de rang $a+1$, a, $a+1$ respectivement. On montre en 3.1 qu'on a une suite exacte canonique

$$(*)_{\mathcal{E}} \qquad 0 \longrightarrow F_{\mathcal{E}} \otimes L^{n-k} \longrightarrow V_{\mathcal{E}} \otimes L^{n-k} \longrightarrow E_{\mathcal{E}} \longrightarrow F_{\mathcal{E}} \longrightarrow 0.$$

Les rangs et degrés des fibrés $E_{\mathcal{E}}$ et $F_{\mathcal{E}}$ (et donc aussi $V_{\mathcal{E}}$) sont invariants par déformation.

1.1.3. *Construction des faisceaux quasi localement libres de type rigide.* Elle est faite par récurrence sur n dans 3.1.2, 3.2, 3.3 et 3.4. On construit le faisceau $\mathcal{E}$ sur C_n connaissant $\mathcal{E}_1$, dont le support est C_{n-1}, et $\mathcal{E}_{|C}$. A priori il semble plus naturel de construire $\mathcal{E}$ connaissant $\mathcal{E}_{|C_{n-1}}$. On montre dans 3.5 que cela est impossible car les faisceaux sur C_{n-1} qui sont des restrictions de faisceaux quasi localement libres de type rigide sur C_n sont *spéciaux*.

Cette méthode de construction devrait rendre possible la description précise d'ouverts des variétés de modules de faisceaux stables qui contiennent de tels faisceaux.

1.2. ***Variétés de modules de faisceaux stables.*** La stabilité ou semi-stabilité, au sens de [Simpson 1994], des faisceaux sans torsion sur C_n ne dépend pas du choix d'un fibré en droites très ample sur C_n. Elle est analogue à celle des fibrés (semi-) stables sur les courbes projectives lisses (cf. [Drézet 2006; 2009]) : un faisceau sans torsion $\mathcal{E}$ sur C_n est *semi-stable* si pour tout sous-faisceau propre $\mathcal{F}$ de $\mathcal{E}$ on a $\mu(\mathcal{F}) \leq \mu(\mathcal{E})$. Si l'on a $\mu(\mathcal{F}) < \mu(\mathcal{E})$, on dit que $\mathcal{E}$ est *stable*.

L'hypothèse $\deg(L) < 0$ est justifiée par le fait que dans le cas contraire les seuls faisceaux sans torsion stables sur C_n sont les fibrés vectoriels stables sur C.

Soient R, d des entiers, avec $R \geq 1$. On note $\mathcal{M}(R, d)$ la variété de modules des faisceaux stables de rang généralisé R et de degré généralisé d sur C_n.

Soient a, k, ϵ, δ des entiers, avec $a \geq 1$ et $1 \leq k < n$. Soient

$$R = an + k, d = k\epsilon + (n-k)\delta + \tfrac{1}{2}\big(n(n-1)a + k(k-1)\big)\deg(L).$$

Les faisceaux quasi localement libres $\mathcal{E}$ de type générique stables localement isomorphes à $a\mathbb{O}_n \oplus \mathbb{O}_k$ et tels que $E_{\mathcal{E}}$ et $F_{\mathcal{E}}$ soient de rang $a+1$ et a (respectivement) et de degré ϵ et δ constituent un ouvert irréductible de $\mathcal{M}(R, d)$, dont la sous-variété réduite sous-jacente est notée $\mathcal{N}(a, k, \delta, \epsilon)$ [Drézet 2009]. A priori $\mathcal{M}(R, d)$ a donc plusieurs composantes irréductibles.

1.3. ***Principaux résultats.*** On démontre dans 5.1 le résultat suivant :

Théorème 5.1.2. *Soient $\mathcal{E}$ un faisceau cohérent sans torsion sur C_n et k un entier tel que $1 \leq k < n$ et que $\mathcal{E}_k \neq 0$. On suppose que*

$$\mu(\mathcal{E}^{(k)}) \leq \mu(\mathcal{E}), \quad \mu((\mathcal{E}^\vee)^{(k)}) \leq \mu(\mathcal{E}^\vee). \tag{1-1}$$

Si $\mathcal{E}[k]$, $(\mathcal{E}_{|C_k})^{\vee\vee}$, $(\mathcal{E}^\vee)[k]$ et $((\mathcal{E}^\vee)_{|C_k})^{\vee\vee}$ sont semi-stables il en est de même de $\mathcal{E}$.

Si de plus les inégalités de (1-1) *sont strictes, et si $\mathcal{E}[k]$ ou $(\mathcal{E}_{|C_k})^{\vee\vee}$, ainsi que $(\mathcal{E}^\vee)[k]$ ou $((\mathcal{E}^\vee)_{|C_k})^{\vee\vee}$, sont stables, alors $\mathcal{E}$ est stable.*

Même si on se limitait aux faisceaux quasi localement libres il serait nécessaire de faire intervenir des sous-faisceaux non quasi localement libres : on donne en 2.6 des exemples de fibrés vectoriels sur C_2 dont la filtration de Harder–Narasimhan comporte des faisceaux non quasi localement libres.

Dans tout ce qui suit on suppose que C est de genre $g \geq 2$. On applique d'abord le théorème précédent aux fibrés vectoriels :

Théorème 5.2.1. *Soit $\mathbb{E}$ un fibré vectoriel sur C_n. Alors, si $\mathbb{E}_{|C}$ est semi-stable* (*ou stable*), *il en est de même de $\mathbb{E}$.*

On en déduit que les variétés de modules de fibrés vectoriels stables sur C_n sont non vides, pourvu qu'il n'y ait pas d'incompatibilité au niveau du rang et du degré généralisés. Soient r, δ des entiers avec $r \geq 1$. Alors le rang généralisé R et le degré généralisé d d'un fibré vectoriel $\mathbb{E}$ sur C_n tel que $\mathbb{E}_{|C}$ soit de rang r et de degré δ sont

$$R = nr, \quad d = n\delta + \tfrac{1}{2}n(n-1)r\deg(L).$$

L'ouvert $U(R, d)$ de $\mathcal{M}(R, d)$ correspondant aux fibrés vectoriels stables est non vide, lisse et irréductible, de dimension

$$1 + nr^2(g-1) - \tfrac{1}{2}n(n-1)r^2\deg(L).$$

On s'intéresse ensuite aux faisceaux quasi localement libres de type rigide non localement libres :

Théorème 5.3.1. *Soient a, k des entiers tels que $a > 0$ et $1 \leq k < n$. Soit $\mathcal{E}$ un faisceau quasi localement libre de type rigide, localement isomorphe à $a\mathcal{O}_n \oplus \mathcal{O}_k$ et tel que*

$$\mu(V_{\mathcal{E}}) + \tfrac{1}{2}n\deg(L) \leq \mu(F_{\mathcal{E}}) \leq \mu(E_{\mathcal{E}}) - \tfrac{1}{2}n\deg(L).$$

Alors si $E_{\mathcal{E}}$, $F_{\mathcal{E}}$ et $V_{\mathcal{E}}$ sont semi-stables, il en est de même de $\mathcal{E}$.

Si les inégalités précédentes sont strictes, et si $E_{\mathcal{E}}$, $F_{\mathcal{E}}$ et $V_{\mathcal{E}}$ sont stables, il en est de même de $\mathcal{E}$.

Le problème de l'*existence* des faisceaux quasi localement libres de type rigide (semi-)stables est plus compliqué que celui de l'existence des fibrés vectoriels (semi-)stables, car si $\mathcal{E}$ en est un, la (semi-)stabilité de $E_{\mathcal{E}}$, $F_{\mathcal{E}}$ et $V_{\mathcal{E}}$ impose des conditions supplémentaires sur les invariants de ces faisceaux, à cause de la suite exacte $(*)_{\mathcal{E}}$.

Avec les notations de 1.2, on a :

Théorème 5.3.3. *Si on a*

$$\frac{\epsilon}{a+1} < \frac{\delta}{a} < \frac{\epsilon-(n-k)\deg(L)}{a+1},$$

alors $\mathcal{N}(a,k,\delta,\epsilon)$ *est non vide.*

Ce résultat généralise la proposition 9.2.1 de [Drézet 2006], où le cas des faisceaux de rang généralisé 3 sur C_2 localement isomorphes à $\mathcal{O}_2\oplus\mathcal{O}_C$ était traité. La démonstration du théorème précédent utilise la résolution de la *conjecture de Lange* [Russo et Teixidor i Bigas 1999].

D'après [Drézet 2009, proposition 6.12], la variété $\mathcal{N}(a,k,\delta,\epsilon)$ est irréductible et lisse, et on a

$$\dim\mathcal{N}(a,k,\delta,\epsilon) \\ =1-\Big(\frac{n(n-1)}{2}a^2+k(n-1)a+\frac{k(k-1)}{2}\Big)\deg(L)+(g-1)\big(na^2+k(2a+1)\big).$$

On termine par donner des applications du premier des théorèmes précédents à des faisceaux non quasi localement libres.

Soient $\mathbb{E}$ un fibré vectoriel sur C_n, $E=\mathbb{E}_{|C}$ et Z un ensemble fini de points de C. On pose $z=h^0(\mathcal{O}_Z)$. Soient $\phi:\mathbb{E}\to\mathcal{O}_Z$ un morphisme surjectif, $\mathcal{E}_\phi=\ker\phi$ et E_ϕ le noyau du morphisme induit $E\to\mathcal{O}_Z$.

Théorème 5.4.2. *Si on a* $z\leq -\operatorname{rg}E\deg(L)$ *et si* E *et* E_ϕ *sont semi-stables, alors* $\mathcal{E}_\phi$ *est semi-stable. Si l'inégalité est stricte et si* E *et* E_ϕ *sont stables, il en est de même de* $\mathcal{E}_\phi$*.*

1.4. ***Plan des sections suivantes.*** La section 2 contient des rappels sur les courbes multiples primitives et les propriétés élémentaires des faisceaux cohérents sur ces courbes. On décrit dans 2.5 la méthode de construction d'un faisceau cohérent $\mathcal{E}$ sur C_n connaissant le faisceau $\mathcal{E}_1$ sur C_{n-1} et $\mathcal{E}_{|C}$. Elle sera employée aussi bien pour les faisceaux localement libres que pour les faisceaux quasi localement libres de type rigide. On donne dans 2.6 des exemples de fibrés vectoriels instables sur une courbe double primitive dont la filtration de Harder–Narasimhan n'est pas constituée de faisceaux quasi localement libres. Cela rend nécessaire, dans l'étude de la (semi-)stabilité d'un faisceau, la considération de sous-faisceaux sans torsion généraux dont les filtrations canoniques peuvent comporter des faisceaux ayant de la torsion.

La section 3 est une étude des faisceaux quasi localement libres de type rigide et de leur construction.

La section 4 traite de la dualité des faisceaux cohérents sur C_n et des faisceaux de torsion.

Dans la section 5 sont démontrés les résultats énoncés dans 1.3.

2. Préliminaires

2.1. ***Définition des courbes multiples primitives et notations.*** Une *courbe primitive* est une variété lisse Y de Cohen–Macaulay telle que la sous-variété réduite associée $C = Y_{\text{red}}$ soit une courbe lisse irréductible, et que tout point fermé de Y possède un voisinage pouvant être plongé dans une surface lisse.

Soient P un point fermé de Y, et U un voisinage de P pouvant être plongé dans une surface lisse S. Soit z un élément de l'idéal maximal de l'anneau local $\mathcal{O}_{S,P}$ de S en P engendrant l'idéal de C dans cet anneau. Il existe alors un unique entier n, indépendant de P, tel que l'idéal de Y dans $\mathcal{O}_{S,P}$ soit engendré par (z^n). Cet entier n s'appelle la *multiplicité* de Y. Si $n = 2$ on dit que Y est une *courbe double*. D'après [Drézet 2007, théorème 5.2.1], l'anneau $\mathcal{O}_{YP}$ est isomorphe à $\mathcal{O}_{CP} \otimes (\mathbb{C}[t]/(t^n))$.

Soit $\mathcal{I}_C$ le faisceau d'idéaux de C dans Y. Alors le faisceau conormal de C, $L = \mathcal{I}_C/\mathcal{I}_C^2$, est un fibré en droites sur C, dit *associé* à Y. Il existe une filtration canonique

$$C = C_1 \subset \cdots \subset C_n = Y,$$

où au voisinage de chaque point P l'idéal de C_i dans $\mathcal{O}_{S,P}$ est (z^i). On notera $\mathcal{O}_i = \mathcal{O}_{C_i}$ pour $1 \le i \le n$.

Le faisceau $\mathcal{I}_C$ est un fibré en droites sur C_{n-1}. Il existe d'après [Drézet 2006, théorème 3.1.1], un fibré en droites $\mathbb{L}$ sur C_n dont la restriction à C_{n-1} est $\mathcal{I}_C$. On a alors, pour tout faisceau de $\mathcal{O}_n$-modules $\mathcal{E}$, un morphisme canonique

$$\mathcal{E} \otimes \mathbb{L} \longrightarrow \mathcal{E}$$

qui en chaque point fermé P de C est la multiplication par z.

2.2. ***Filtrations canoniques.*** Dans toute la suite de la section 2 on considère une courbe multiple primitive C_n de courbe réduite associée C. On utilise les notations de 2.1.

Soient P un point fermé de C, M un $\mathcal{O}_{nP}$-module de type fini. Soit $\mathcal{E}$ un faisceau cohérent sur C_n.

Définition 2.2.1. La *première filtration canonique de M* est la filtration

$$M_n = \{0\} \subset M_{n-1} \subset \cdots \subset M_1 \subset M_0 = M$$

telle que pour $0 \le i < n$, M_{i+1} soit le noyau du morphisme canonique surjectif $M_i \to M_i \otimes_{\mathcal{O}_{n,P}} \mathcal{O}_{C,P}$. On a donc

$$M_i/M_{i+1} = M_i \otimes_{\mathcal{O}_{n,P}} \mathcal{O}_{C,P}, \quad M/M_i \simeq M \otimes_{\mathcal{O}_{n,P}} \mathcal{O}_{i,P}, \quad M_i = z^i M.$$

On pose, si $i > 0$, $G_i(M) = M_i/M_{i+1}$. Le gradué

$$\operatorname{Gr} M = \bigoplus_{i=0}^{n-1} G_i(M) = \bigoplus_{i=0}^{n-1} z^i M/z^{i+1}M$$

est un $\mathcal{O}_{C,P}$-module.

On définit de même la *première filtration canonique de* $\mathcal{E}$: c'est la filtration

$$\mathcal{E}_n = 0 \subset \mathcal{E}_{n-1} \subset \cdots \subset \mathcal{E}_1 \subset \mathcal{E}_0 = \mathcal{E}$$

telle que pour $0 \le i < n$, $\mathcal{E}_{i+1}$ soit le noyau du morphisme canonique surjectif $\mathcal{E}_i \to \mathcal{E}_{i|C}$. On a donc $\mathcal{E}_i/\mathcal{E}_{i+1} = \mathcal{E}_{i|C}$, $\mathcal{E}/\mathcal{E}_i = \mathcal{E}_{|C_i}$. On pose, si $i \ge 0$,

$$G_i(\mathcal{E}) = \mathcal{E}_i/\mathcal{E}_{i+1}.$$

Le gradué $\operatorname{Gr}\mathcal{E}$ est un $\mathcal{O}_C$-module.

Définition 2.2.2. La *seconde filtration canonique de M* est la filtration

$$M^{(0)} = \{0\} \subset M^{(1)} \subset \cdots \subset M^{(n-1)} \subset M^{(n)} = M$$

avec $M^{(i)} = \{u \in M; z^i u = 0\}$. Si $M_n = \{0\} \subset M_{n-1} \subset \cdots \subset M_1 \subset M_0 = M$ est la (première) filtration canonique de M on a $M_i \subset M^{(n-i)}$ pour $0 \le i \le n$. On pose, si $i > 0$, $G^{(i)}(M) = M^{(i)}/M^{(i-1)}$. Le gradué $\operatorname{Gr}_2 M = \bigoplus_{i=1}^n G^{(i)}(M)$ est un $\mathcal{O}_{C,P}$-module.

On définit de même la *seconde filtration canonique de* $\mathcal{E}$:

$$\mathcal{E}^{(0)} = \{0\} \subset \mathcal{E}^{(1)} \subset \cdots \subset \mathcal{E}^{(n-1)} \subset \mathcal{E}^{(n)} = \mathcal{E}.$$

On pose, si $i > 0$,

$$G^{(i)}(\mathcal{E}) = \mathcal{E}^{(i)}/\mathcal{E}^{(i-1)}.$$

Le gradué $\operatorname{Gr}_2\mathcal{E}$ est un $\mathcal{O}_C$-module.

2.3. *Invariants.*

Définition 2.3.1. L'entier $R(M) = \operatorname{rg}(\operatorname{Gr} M)$ s'appelle le *rang généralisé* de M. L'entier $R(\mathcal{E}) = \operatorname{rg}(\operatorname{Gr}\mathcal{E})$ s'appelle le *rang généralisé* de $\mathcal{E}$. On a donc $R(\mathcal{E}) = R(\mathcal{E}_P)$ pour tout $P \in C$.

Définition 2.3.2. L'entier $\operatorname{Deg}(\mathcal{E}) = \deg(\operatorname{Gr}\mathcal{E})$ s'appelle le *degré généralisé de* $\mathcal{E}$. Si $R(\mathcal{E}) > 0$ on pose

$$\mu(\mathcal{E}) = \frac{\operatorname{Deg}(\mathcal{E})}{R(\mathcal{E})}$$

et on appelle ce nombre la *pente* de $\mathcal{E}$.

Le rang et le degré généralisés sont *additifs*, c'est-à-dire que si $0 \to \mathcal{E}' \to \mathcal{E} \to \mathcal{E}'' \to 0$ est une suite exacte de faisceaux cohérents sur C_n alors on a

$$R(\mathcal{E}) = R(\mathcal{E}') + R(\mathcal{E}''), \quad \mathrm{Deg}(\mathcal{E}) = \mathrm{Deg}(\mathcal{E}') + \mathrm{Deg}(\mathcal{E}'').$$

Il sont également invariants par déformation.

2.4. ***Faisceaux quasi localement libres.*** Soit P un point fermé de C. Soit M un $\mathcal{O}_{n,P}$-module de type fini. On dit que M est *quasi libre* s'il existe des entiers $m_1, \ldots, m_n$ non négatifs et un isomorphisme $M \simeq \bigoplus_{i=1}^n m_i \mathcal{O}_{i,P}$. Les entiers m_1, $\ldots, m_n$ sont uniquement déterminés. On dit que M est *de type* $(m_1, \ldots, m_n)$. On a $R(M) = \sum_{i=1}^n i.m_i$.

Soit $\mathcal{E}$ un faisceau cohérent sur C_n. On dit que $\mathcal{E}$ est *quasi localement libre* en un point P de C s'il existe un ouvert U de C_n contenant P et des entiers non négatifs $m_1, \ldots, m_n$ tels que pour tout point Q de U, $\mathcal{E}_{n,Q}$ soit quasi localement libre de type $m_1, \ldots, m_n$. Les entiers $m_1, \ldots, m_n$ sont uniquement déterminés et ne dépendent que de $\mathcal{E}$, et on dit que $(m_1, \ldots, m_n)$ est le *type de* $\mathcal{E}$. Sur un voisinage de P, $\mathcal{E}$ est alors isomorphe à $\bigoplus_{i=1}^n m_i \mathcal{O}_i$.

On dit que $\mathcal{E}$ est *quasi localement libre* s'il l'est en tout point de C_n.

D'après [Drézet 2006, théorème 5.1.3] $\mathcal{E}$ est quasi localement libre en P si et seulement si pour $0 \le i < n$, $G_i(\mathcal{E})$ est libre en P.

Il en découle que $\mathcal{E}$ est quasi localement libre si et seulement si pour $0 \le i < n$, $G_i(\mathcal{E})$ est localement libre sur C.

2.5. ***Construction des faisceaux cohérents.***

2.5.1. On décrit ici le moyen de construire un faisceau cohérent $\mathcal{E}$ sur C_n, connaissant $\mathcal{E}_{|C}$ et $\mathcal{E}_1$, qui sont des faisceaux sur C et C_{n-1} respectivement.

Soient $\mathcal{F}$ un faisceau cohérent sur C_{n-1} et E un fibré vectoriel sur C. On s'intéresse aux faisceaux cohérents $\mathcal{E}$ sur C_n tels que $\mathcal{E}_{|C} = E$ et $\mathcal{E}_1 = \mathcal{F}$. Soit $0 \to \mathcal{F} \to \mathcal{E} \to E \to 0$ une suite exacte, associée à $\sigma \in \mathrm{Ext}^1_{\mathcal{O}_n}(E, \mathcal{F})$. On voit aisément que le morphisme canonique $\pi_{\mathcal{E}} : \mathcal{E} \otimes \mathcal{I}_C \to \mathcal{E}$ induit un morphisme

$$\Phi_{\mathcal{F},E}(\sigma) : E \otimes L \longrightarrow \mathcal{F}_{|C}.$$

On a $\mathcal{E}_{|C} = E$ et $\mathcal{E}_1 = \mathcal{F}$ si et seulement si $\Phi_{\mathcal{F},E}(\sigma)$ est surjectif [Drézet 2009, lemme 3.13].

D'après la proposition 3.14 de [Drézet 2009], on a une suite exacte canonique

$$(2\text{-}1) \quad 0 \longrightarrow \mathrm{Ext}^1_{\mathcal{O}_C}(E, \mathcal{F}^{(1)}) \longrightarrow \mathrm{Ext}^1_{\mathcal{O}_n}(E, \mathcal{F}) \xrightarrow{\Phi_{\mathcal{F},E}} \mathrm{Hom}(E \otimes L, \mathcal{F}_{|C}) \longrightarrow 0.$$

2.5.2. On suppose que $n \geq 3$. On s'intéresse maintenant aux extensions $0 \to \mathcal{F} \to \mathcal{E} \to E \to 0$ associées aux éléments $\sigma \in \operatorname{Ext}^1_{\mathcal{O}_n}(E, \mathcal{F})$ tels que $\Phi_{\mathcal{F},E}(\sigma) = 0$ $\big($donc $\sigma \in \operatorname{Ext}^1_{\mathcal{O}_C}(E, \mathcal{F}^{(1)})\big)$. Dans ce cas $\mathcal{E}$ est localement isomorphe à $\mathcal{F} \oplus E$, d'après [Drézet 2009, 2.4]. Plus précisément dans la suite exacte (2-1) le terme $\operatorname{Ext}^1_{\mathcal{O}_C}(E, \mathcal{F}^{(1)})$ est en fait $H^1(\mathcal{H}om(E, \mathcal{F}))$. On peut donc représenter σ par un cocycle (f_{ij}) relativement à un recouvrement ouvert (U_i) de C_n, f_{ij} étant un morphisme $E_{|U_{ij}} \to \mathcal{F}_{|U_{ij}}$. D'après la proposition 2.2 de [Drézet 2009], le faisceau $\mathcal{E}$ est obtenu en recollant les $(\mathcal{F} \oplus E)_{|U_i}$ au moyen des morphismes

$$\begin{pmatrix} I_{\mathcal{F}} & f_{ij} \\ 0 & I_E \end{pmatrix}.$$

On suppose maintenant que $\mathcal{F}$ est localement libre sur C_{n-1}. Soit $F = \mathcal{F}_{|C} \otimes L^{-1}$, on a donc $\mathcal{F}^{(1)} = F \otimes L^{n-1}$. En utilisant la construction précédente de $\mathcal{E}$ au moyen d'un cocycle on voit aisément que $\mathcal{E}_{|C} \simeq (F \otimes L) \oplus E$, et qu'on a une suite exacte

$$0 \longrightarrow F \otimes L^{n-1} \longrightarrow \mathcal{E}^{(1)} \longrightarrow E \longrightarrow 0,$$

qui est associée à σ.

2.5.3. *Construction des fibrés vectoriels.* On suppose que $\mathcal{F}$ est un fibré vectoriel sur C_{n-1}. On veut construire et paramétrer les fibrés vectoriels $\mathbb{E}$ sur C_n tels que $\mathbb{E}_1 = \mathcal{F}$. Il convient donc de prendre $E = \mathcal{F}_{|C} \otimes L^{-1}$ et de considérer les extensions $0 \to \mathcal{F} \to \mathbb{E} \to E \to 0$ telles que l'élément associé σ de $\operatorname{Ext}^1_{\mathcal{O}_n}(E, \mathcal{F})$ soit tel que $\Phi_{\mathcal{F},E}(\sigma) : E \otimes L \to E \otimes L$ soit l'identité de $E \otimes L$. Si E est simple on montre aisément, en utilisant le fait que $\deg(L) < 0$, que deux éléments σ, σ' de $\Phi^{-1}_{\mathcal{F},E}(I_{E \otimes L})$ définissent des fibrés vectoriels $\mathbb{E}$ isomorphes si et seulement si $\sigma = \sigma'$. Dans ce cas les fibrés vectoriels recherchés sont donc paramétrés par l'espace affine $\Phi^{-1}_{\mathcal{F},E}(I_{E \otimes L}) \simeq \operatorname{Ext}^1_{\mathcal{O}_C}(E, E \otimes L^{n-1})$.

2.6. ***Filtration de Harder–Narasimhan.*** Nous supposons encore que $\deg(L) < 0$. On montre ici que la filtration de Harder–Narasimhan d'un fibré vectoriel sur C_2 n'est pas nécessairement constituée de faisceaux quasi localement libres. Cela entraine que dans l'étude de la (semi-)stabilité des faisceaux localement libres (ou a fortiori quasi localement libres) il faut aussi considérer des sous-faisceaux sans torsion non nécessairement quasi localement libres.

Soient P un point fermé de C_2 et $\mathcal{I}_P$ son faisceau d'idéaux. Soient $z \in \mathcal{O}_{2,P}$ un générateur de l'idéal de C et $x \in \mathcal{O}_{2,P}$ au dessus d'un générateur de l'idéal de P dans $\mathcal{O}_{C,P}$. On a donc $\mathcal{I}_{P,P} = (x, z)$. On a une suite exacte de $\mathcal{O}_{2,P}$-modules

$$0 \longrightarrow (x, z) \xrightarrow{\alpha} 2\mathcal{O}_{2,P} \xrightarrow{\beta} (x, z) \longrightarrow 0, \tag{2-2}$$

où pour tous $a, b \in \mathbb{O}_{2,P}$

$$\alpha(ax+bz) = (-az, ax+bz),$$
$$\beta(a, b) = ax+bz.$$

On va globaliser cette suite exacte afin d'obtenir des suites exactes

$$0 \longrightarrow \mathcal{I}_P \otimes \mathbb{D} \longrightarrow \mathbb{E} \longrightarrow \mathcal{I}_P \longrightarrow 0, \tag{2-3}$$

où $\mathbb{D}$ est fibré en droites sur C_2 et $\mathbb{E}$ un fibré vectoriel de rang 2 sur C_2. Le faisceau $\mathcal{E}xt^1_{\mathbb{O}_2}(\mathcal{I}_P, \mathcal{I}_P \otimes \mathbb{D})$ est concentré au point P. On en déduit qu'il existe une section s de $\mathcal{E}xt^1_{\mathbb{O}_2}(\mathcal{I}_P, \mathcal{I}_P \otimes \mathbb{D})$ dont la valeur en P correspond à l'extension (2-2).

On a un morphisme surjectif canonique

$$\Psi : \mathrm{Ext}^1_{\mathbb{O}_2}(\mathcal{I}_P, \mathcal{I}_P \otimes \mathbb{D}) \longrightarrow H^0(\mathcal{E}xt^1_{\mathbb{O}_2}(\mathcal{I}_P, \mathcal{I}_P \otimes \mathbb{D})).$$

Donc $\Psi^{-1}(s)$ est non vide. Si $0 \to \mathcal{I}_P \otimes \mathbb{D} \to \mathcal{E} \to \mathcal{I}_P \to 0$ est une extension associée à un élément de $\Psi^{-1}(s)$, le faisceau $\mathcal{E}$ est localement libre. L'existence des extensions (2-3) est donc prouvée.

Proposition 2.6.1. *Soit* $0 \to \mathcal{I}_P \otimes \mathbb{D} \to \mathbb{E} \to \mathcal{I}_P \to 0$ *une extension, où* $\mathbb{D}$ *est un fibré en droites sur* C_2 *et* $\mathbb{E}$ *un fibré vectoriel de rang* 2 *sur* C_2*. Alors si* $\deg(\mathbb{D}_{|C}) > 0$, *le faisceau* $\mathcal{I}_P \otimes \mathbb{D}$ *est le sous-faisceau semi-stable maximal de* $\mathbb{E}$*.*

Démonstration. Soit $\mathcal{H} \subset \mathbb{E}$ le sous-faisceau semi-stable maximal de $\mathbb{E}$. On a $R(\mathcal{H}) = 1, 2$ ou 3.

On note $\mathbb{L}_x$ le faisceau d'idéaux égal à $\mathbb{O}_2$ sur $C_2 \backslash P$ et à (x) au point P. C'est un fibré en droites sur C_2 et on a $\mathcal{I}_P^\vee \simeq \mathcal{I}_P \otimes \mathbb{L}_x^{-1}$. On a donc une suite exacte

$$0 \longrightarrow \mathcal{I}_P \otimes \mathbb{D} \longrightarrow \mathbb{E}^\vee \otimes \mathbb{L}_x \otimes \mathbb{D} \longrightarrow \mathcal{I}_P \longrightarrow 0$$

En considérant cette suite exacte on se ramène au cas où $R(\mathcal{H}) = 1$ ou 2.

On montre d'abord que $\mathcal{I}_P$ est semi-stable. Soit $\mathcal{F} \subset \mathcal{I}_P$ un sous-faisceau propre tel que $\mathcal{I}_P/\mathcal{F}$ soit sans torsion. On a alors $R(\mathcal{F}) = 1$, donc $\mathcal{F}$ est concentré sur C, et est donc contenu dans $(\mathcal{I}_P)^{(1)} = L$. Donc

$$\mu(\mathcal{F}) \leq \deg(L) \leq \mu(\mathcal{I}_P) = \tfrac{1}{2}(\deg(L) - 1),$$

car $\deg(L) < 0$.

Supposons d'abord que $R(\mathcal{H}) = 1$. Si $\mathcal{H} \subset \mathcal{I}_P \otimes \mathbb{D}$ on a $\mu(\mathcal{H}) \leq \mu(\mathcal{I}_P \otimes \mathbb{D})$ (car $\mathcal{I}_P \otimes \mathbb{D}$ est semi-stable), ce qui contredit la maximalité de $\mathcal{H}$. Si $\mathcal{H} \not\subset \mathcal{I}_P \otimes \mathbb{D}$, on peut voir $\mathcal{H}$ comme un sous-faisceau de $\mathcal{I}_P$, donc $\mu(\mathcal{H}) \leq \mu(\mathcal{I}_P)$, donc $\mu(\mathcal{H}) < \mu(\mathcal{I}_P \otimes \mathbb{D})$, ce qui est absurde.

On a donc $R(\mathcal{H}) = 2$. Soit r le rang généralisé de l'image $\mathcal{U}$ de $\mathcal{H}$ dans $\mathcal{I}_P$. Si $r = 0$ on a $\mathcal{H} = \mathcal{I}_P \otimes \mathbb{D}$, ce qu'il fallait démontrer. Si $r = 2$ on peut voir $\mathcal{H}$ comme un sous-faisceau de $\mathcal{I}_P$ et on a alors encore $\mu(\mathcal{H}) \leq \mu(\mathcal{I}_P)$, ce qui est impossible.

Il reste à traiter le cas où $r = 1$ en montrant qu'il est impossible. Soit d le degré de $\mathcal{U}$ (qui est concentré sur C). On a, puisque $\mathcal{H} \cap (\mathcal{I}_P \otimes \mathbb{D})$ est aussi de rang généralisé 1,

$$\deg(\mathcal{U}) \leq \deg(L) \quad \text{et} \quad \deg(\mathcal{H} \cap (\mathcal{I}_P \otimes \mathbb{D})) \leq \deg(L) + \deg(\mathbb{D}_{|C}).$$

Donc

$$\mu(\mathcal{H}) \leq \deg(L) + \tfrac{1}{2}\deg(\mathbb{D}_{|C}) < \tfrac{1}{2}(\deg(L) - 1) + \deg(\mathbb{D}_{|C}) = \mu(\mathcal{I}_P \otimes \mathbb{D}),$$

ce qui contredit la définition de $\mathcal{H}$. □

Remarque 2.6.2. Si on suppose que $\deg(\mathbb{D}_{|C}) = 0$ on obtient des fibrés vectoriels $\mathbb{E}$ semi-stables de rang 2 sur C_2 dont la filtration de Jordan–Hölder n'est pas constituée de faisceaux quasi localement libres.

3. Faisceaux quasi localement libres de type rigide

Dans toute la suite de cette section on considère une courbe multiple primitive C_n de courbe réduite associée C. On utilise les notations de 2.1, et on suppose que $\deg(L) < 0$.

3.1. ***Définitions.*** Soit $\mathcal{E}$ un faisceau cohérent quasi localement libre sur C_n. Soient $a = [R(\mathcal{E})/n]$ et $k = R(\mathcal{E}) - an$. On a donc $R(\mathcal{E}) = an + k$. On dit que $\mathcal{E}$ est de *type rigide* s'il est localement libre si $k = 0$, et localement isomorphe à $a\mathcal{O}_n \oplus \mathcal{O}_k$ si $k > 0$. Si $k > 0$ cela revient à dire que $\mathcal{E}$ est de type $(m_1, \ldots, m_n)$, avec $m_i = 0$ si $i \neq k, n$ et $m_k = 0$ ou 1.

Le fait d'être quasi localement libre de type rigide est une *propriété ouverte* : autrement dit si Y une variété algébrique intègre et $\mathcal{F}$ une famille plate de faisceaux cohérents sur C_n paramétrée par Y, alors l'ensemble des points $y \in Y$ tels que $\mathcal{E}_y$ soit quasi localement libre de type rigide est un ouvert de Y [Drézet 2009, proposition 6.9].

Supposons que $\mathcal{E}$ soit quasi localement libre de type rigide et que $k > 0$. Alors $\mathbb{E} = \mathcal{E}_{|C_k}$ est un fibré vectoriel de rang $a+1$ sur C_k, et $\mathbb{F} = \mathcal{E}_k$ est un fibré vectoriel de rang a sur C_{n-k}. Donc $\mathcal{E}$ est une extension

$$0 \longrightarrow \mathbb{F} \longrightarrow \mathcal{E} \longrightarrow \mathbb{E} \longrightarrow 0$$

d'un fibré vectoriel de rang $a+1$ sur C_k par un fibré vectoriel de rang a sur C_{n-k}. De même $\mathbb{V} = \mathcal{E}^{(k)}$ est un fibré vectoriel sur C_k et on a une suite exacte

$$0 \longrightarrow \mathbb{V} \longrightarrow \mathcal{E} \longrightarrow \mathbb{F} \otimes \mathbb{L}^{-k} \longrightarrow 0.$$

Posons $E = \mathscr{E}_{|C} = \mathbb{E}_{|C}$, $F = G_k(\mathscr{E}) \otimes L^{-k} = \mathbb{F}_{|C} \otimes L^{-k}$. Alors on a rg $E = a+1$, rg $F = a$, et

$$\big(G_0(\mathscr{E}), G_1(\mathscr{E}), \ldots, G_{n-1}(\mathscr{E})\big) = (E, E\otimes L, \ldots, E\otimes L^{k-1}, F\otimes L^k, \ldots, F\otimes L^{n-1}).$$

Donc

$$\mathrm{Deg}(\mathscr{E}) = k\deg(E) + (n-k)\deg(F) + \tfrac{1}{2}\big(n(n-1)a + k(k-1)\big)\deg(L).$$

On a

$$G^{(n)}(\mathscr{E}) = \mathscr{E}/\mathscr{E}^{(n-1)} = \mathscr{E}_{n-1} \otimes L^{1-n} = G_{n-1}(\mathscr{E}) \otimes L^{1-n} = F.$$

Posons $V = G^{(k)}(\mathscr{E}) \otimes L^{k-n} = \mathbb{V}_{|C} \otimes L^{k-n}$. On a rg $V = a+1$, $\deg(V) = \deg(E) - (n-k)\deg(L)$, et

$$\big(G^{(n)}(\mathscr{E}), G^{(n-1)}(\mathscr{E}), \ldots, G^{(1)}(\mathscr{E})\big) = (F, F\otimes L, \ldots, F\otimes L^{n-k-1}, V\otimes L^{n-k}, \ldots, V\otimes L^{n-1}).$$

Les morphismes canoniques

$$G_i(\mathscr{E}) \otimes L \hookrightarrow G_{i+1}(\mathscr{E}), \qquad G^{(i+1)} \otimes L \twoheadrightarrow G^{(i)}(\mathscr{E})$$

définissent un morphisme surjectif $\phi : E \to F$ et un morphisme injectif $\psi : F \to V$. D'après [Drézet 2009, corollaire 3.4], on a un isomorphisme canonique

$$\ker\phi \simeq (\operatorname{coker}\psi) \otimes L^{n-k}.$$

Posons $D = \ker\phi$. C'est un fibré en droites sur C. On a des suites exactes

$$0 \longrightarrow D \longrightarrow E \xrightarrow{\ \phi\ } F \longrightarrow 0,$$

$$0 \longrightarrow F \xrightarrow{\ \psi\ } V \longrightarrow D \otimes L^{k-n} \longrightarrow 0.$$

3.1.1. *Notations.* On pose $E_{\mathscr{E}} = E$, $F_{\mathscr{E}} = F$, $V_{\mathscr{E}} = V$, $D_{\mathscr{E}} = D$,

$$\phi_{\mathscr{E}} = \phi : E_{\mathscr{E}} \longrightarrow F_{\mathscr{E}} \quad \text{et} \quad \psi_{\mathscr{E}} = \phi : F_{\mathscr{E}} \longrightarrow V_{\mathscr{E}}.$$

On a une suite exacte canonique

$$(*)_{\mathscr{E}} \qquad 0 \longrightarrow F_{\mathscr{E}} \otimes L^{n-k} \longrightarrow V_{\mathscr{E}} \otimes L^{n-k} \longrightarrow E_{\mathscr{E}} \longrightarrow F_{\mathscr{E}} \longrightarrow 0.$$

3.1.2. *Construction et paramétrisation.* On cherche ici à décrire comment on peut obtenir les faisceaux quasi localement libres de type rigide $\mathscr{E}$ précédents. On part d'abord d'un fibré vectoriel $\mathbb{F}$ sur C_{n-k} de rang $a \geq 1$ (voir 2.5.3 pour la construction et la paramétrisation des fibrés vectoriels) qui sera $\mathscr{E}_k$. On construira ensuite successivement $\mathscr{E}_{k-1}, \ldots, \mathscr{E}_1, \mathscr{E}$. Il y a deux cas différents : le passage de $\mathbb{F}$ à $\mathscr{E}_{k-1}$, et celui de $\mathscr{E}_i$ à $\mathscr{E}_{i-1}$ si $1 \leq i < k$. On va donc étudier dans les sections suivantes les deux étapes suivantes :

La première étape consiste à étudier les extensions

$$0 \longrightarrow \mathbb{F} \longrightarrow \mathcal{E} \longrightarrow H \longrightarrow 0$$

sur C_{n-k+1}, où H est un fibré vectoriel de rang $a+1$ sur C, telles que le morphisme induit $\Phi_{\mathbb{F},H} : H \to \mathbb{F}_{|C}$ soit surjectif (voir 2.5). Le faisceau $\mathcal{E}$ est alors quasi localement libre de type rigide, et localement isomorphe à $a\mathbb{O}_{n-k+1} \oplus \mathbb{O}_C$. On a $\mathcal{E}_{|C} = H$ et $\mathcal{E}_1 = \mathbb{F}$.

Dans la seconde étape on part d'un faisceau quasi localement libre de type rigide $\mathcal{G}$ sur C_{n-k+i}, $1 \leq i < k$, localement isomorphe à $a\mathbb{O}_{n-k+i} \oplus \mathbb{O}_i$. Soit $H = \mathcal{G}_{\mathbb{C}} \otimes L^{-1}$. On s'intéresse alors aux extensions

$$0 \longrightarrow \mathcal{G} \longrightarrow \mathcal{E} \longrightarrow H \longrightarrow 0$$

sur $C_{n-k+i+1}$ telles que le morphisme induit $\Phi_{\mathcal{G},H} : H \otimes L \to H \otimes L$ soit l'identité de $H \otimes L$. Le faisceau $\mathcal{E}$ est alors quasi localement libre de type rigide, et localement isomorphe à $a\mathbb{O}_{n-k+i+1} \oplus \mathbb{O}_{i+1}$. On a $\mathcal{E}_{|C} = H$ et $\mathcal{E}_1 = \mathcal{G}$.

3.2. ***Construction et paramétrisation – première étape.*** On décrit ici la première étape évoquée dans 3.1.2, dont on conserve les notations.

On pose $F = \mathbb{F}_{|C} \otimes L^{-1}$. Soient $\sigma \in \mathrm{Ext}^1_{\mathbb{O}_{n-k}}(H, \mathbb{F})$ et

$$0 \longrightarrow \mathbb{F} \longrightarrow \mathcal{E}_\sigma \longrightarrow H \longrightarrow 0$$

l'extension correspondante. On suppose que $\phi = \Phi_{\mathbb{F},H}(\sigma) \otimes I_{L^{-1}} : H \to F$ est surjectif. Soit $D = \ker \phi$. On a $E_{\mathcal{E}_\sigma} = H$, $F_{\mathcal{E}_\sigma} = F$, et une suite exacte

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow V_{\mathcal{E}_\sigma} \otimes L^{n-k} \longrightarrow D \longrightarrow 0. \tag{3-1}$$

On a d'après 2.5.1 une suite exacte

$$0 \longrightarrow \mathrm{Ext}^1_{\mathbb{O}_C}(H, F \otimes L^{n-k}) \longrightarrow \mathrm{Ext}^1_{\mathbb{O}_{n-k+1}}(H, \mathbb{F}) \xrightarrow{\Phi_{\mathbb{F},H}} \mathrm{Hom}(H, F) \longrightarrow 0.$$

Lemme 3.2.1. *L'image de σ dans $\mathrm{Ext}^1_{\mathbb{O}_{n-k+1}}(D, \mathbb{F})$ est contenue dans*

$$\mathrm{Ext}^1_{\mathbb{O}_C}(D, F \otimes L^{n-k}),$$

et c'est l'élément associé à la suite exacte (3-1).

Démonstration. Soit σ' l'image de σ dans $\mathrm{Ext}^1_{\mathbb{O}_{n-k+1}}(D, \mathbb{F})$. La fonctorialité de $\phi_{\mathbb{F},H}$ par rapport à $\mathbb{F}$ et H entraine que $\Phi_{\mathbb{F},D}(\sigma') = 0$. On a donc bien d'après 2.5.2 $\sigma' \in \mathrm{Ext}^1_{\mathbb{O}_C}(D, F \otimes L^{n-k})$. D'après [Drézet 2005, proposition 4.3.1] on a un diagramme commutatif avec lignes exactes

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbb{F} & \longrightarrow & \mathcal{H} & \longrightarrow & D & \longrightarrow & 0 \\ & & \| & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathbb{F} & \longrightarrow & \mathcal{E}_\sigma & \longrightarrow & H & \longrightarrow & 0 \end{array}$$

l'extension du haut étant associée à σ'. On a $\mathcal{H}^{(1)} \subset \mathcal{E}_\sigma^{(1)}$, et d'après 2.5.2 on a une suite exacte

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow \mathcal{H}^{(1)} \otimes L^{n-k} \longrightarrow D \longrightarrow 0.$$

Il en découle que $\mathcal{H}^{(1)} = \mathcal{E}_\sigma^{(1)} = V_{\mathcal{E}_\sigma}$. D'après 2.5.2 σ' correspond bien à l'extension (3-1). □

Proposition 3.2.2. *Pour toute extension*

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow W \otimes L^{n-k} \longrightarrow D \longrightarrow 0$$

sur C il existe $\sigma_0 \in \Phi_{H,\mathbb{F}}^{-1}(\phi \otimes I_L)$ tel que l'extension précédente soit isomorphe à l'extension

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow V_{\mathcal{E}_{\sigma_0}} \otimes L^{n-k} \longrightarrow D \longrightarrow 0.$$

Démonstration. Cela découle du lemme 3.2.1, du carré commutatif

$$\begin{array}{ccc} \mathrm{Ext}^1_{\mathcal{O}_C}(H, F \otimes L^{n-k}) & \hookrightarrow & \mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(H, \mathbb{F}) \\ \downarrow & & \downarrow \\ \mathrm{Ext}^1_{\mathcal{O}_C}(D, F \otimes L^{n-k}) & \hookrightarrow & \mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(D, \mathbb{F}) \end{array}$$

et de la surjectivité du morphisme de gauche. □

Soient $\phi : H \to \mathbb{F}_{|C}$ un morphisme surjectif et $\eta \in \mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(D, \mathbb{F})$. Alors on a $\Phi_{\mathbb{F},H}(\sigma) = \phi \otimes I_L$ et la suite exacte

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow V_{\mathcal{E}_{\sigma_0}} \otimes L^{n-k} \longrightarrow D \longrightarrow 0$$

est associée à η si et seulement si η appartient au sous-espace affine

$$\Phi_{\mathbb{F},H}^{-1}(\phi \otimes I_L) \cap \psi^{-1}(\eta)$$

de $\mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(H, \mathbb{F})$. Dans cette expression ψ désigne l'application canonique

$$\mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(H, \mathbb{F}) \to \mathrm{Ext}^1_{\mathcal{O}_{n-k+1}}(D, \mathbb{F}).$$

3.3. ***Construction et paramétrisation – seconde étape.*** On décrit ici la seconde étape évoquée dans 3.1.2, dont on conserve les notations.

On suppose que $H = \mathcal{G}_{|C} \otimes L^{-1}$. Soient $\sigma \in \mathrm{Ext}^1_{\mathcal{O}_{n-k+i+1}}(H, \mathcal{G})$ tel que $\Phi_{\mathcal{G},H}(\sigma)$ soit l'identité de $H \otimes L$ et $0 \to \mathcal{G} \to \mathcal{E}_\sigma \to H \to 0$ l'extension correspondante.

Proposition 3.3.1. *On a* $E_{\mathcal{E}_\sigma} = E_{\mathcal{G}} \otimes L^{-1}$, $F_{\mathcal{E}_\sigma} = F_{\mathcal{G}} \otimes L^{-1}$, $V_{\mathcal{E}_\sigma} = V_{\mathcal{G}} \otimes L^{-1}$ *et* $(*)_{\mathcal{E}_\sigma} = (*)_{\mathcal{G}} \otimes L^{-1}$.

Démonstration. Il suffit de le faire avec $a\mathcal{O}_{n-k+i+1} \oplus \mathcal{O}_{i+1}$ à la place de $\mathcal{E}_\sigma$ en utilisant les isomorphismes locaux $\mathcal{E}_\sigma \simeq a\mathcal{O}_{n-k+i+1} \oplus \mathcal{O}_{i+1}$ et la fonctorialité de $(*)_{\mathcal{E}_\sigma}$, ce qui est immédiat. □

On a d'après 2.5.1 une suite exacte

$$0 \longrightarrow \mathrm{Ext}^1_{\mathcal{O}_C}(H, \mathcal{G}^{(1)}) \longrightarrow \mathrm{Ext}^1_{\mathcal{O}_{n-k+i+1}}(H, \mathcal{G}) \xrightarrow{\Phi_{\mathcal{G},H}} \mathrm{Hom}(H \otimes L, \mathcal{G}_{|C}) \longrightarrow 0.$$

Les faisceaux $\mathcal{E}_\sigma$ considérés ici sont donc indexés par le sous-espace affine $\Phi^{-1}_{\mathcal{G},H}(I_{H\otimes L})$ de $\mathrm{Ext}^1_{\mathcal{O}_{n-k+i+1}}(H, \mathcal{G})$.

3.4. ***Construction et paramétrisation – conclusion.***

Proposition 3.4.1. *Soient* k, a *des entiers tels que* $1 \leq k < n$, $a > 0$. *Soient* E, F, V *des fibrés vectoriels sur* C *de rangs* $a+1$, a, $a+1$ *respectivement, et*

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow V \otimes L^{n-k} \longrightarrow E \longrightarrow F \longrightarrow 0 \tag{3-2}$$

une suite exacte. Alors il existe un faisceau quasi localement libre de type rigide $\mathcal{E}$, *localement isomorphe à* $a\mathcal{O}_n \oplus \mathcal{O}_k$ *et tel que* $(*)_{\mathcal{E}}$ *soit isomorphe à* (3-2).

Cela signifie qu'il existe un diagramme commutatif reliant les suite exactes $(*)_{\mathcal{E}}$ *et* (3-2) :

$$\begin{array}{ccccccc} F \otimes L^{n-k} & \longrightarrow & V \otimes L^{n-k} & \longrightarrow & E & \longrightarrow & F \\ \downarrow \simeq & & \downarrow \simeq & & \downarrow \simeq & & \downarrow \simeq \\ F_{\mathcal{E}} \otimes L^{n-k} & \longrightarrow & V_{\mathcal{E}} \otimes L^{n-k} & \longrightarrow & E_{\mathcal{E}} & \longrightarrow & F_{\mathcal{E}} \end{array}$$

3.5. ***Restrictions des faisceaux quasi localement libres de type rigide.*** Les méthodes précédentes de construction de faisceaux quasi localement libres de type rigide se font sur le principe suivant : on part d'un tel faisceau $\mathcal{F}$ sur C_{n-1} et on en construit un $\mathcal{E}$ sur C_n tel que $\mathcal{E}_1 = \mathcal{F}$.

A priori il semblerait plus naturel de chercher un faisceau $\mathcal{E}$ tel que $\mathcal{E}_{|C_{n-1}} = \mathcal{F}$. Mais c'est impossible car un faisceau quasi localement libre de type rigide sur C_{n-1}, non localement libre, n'est pas nécessairement la restriction d'un faisceau du même type sur C_n :

Proposition 3.5.1. *Soit $\mathcal{E}$ un faisceau quasi localement libre de type rigide non localement libre sur C_n localement isomorphe à $a\mathbb{O}_n \oplus \mathbb{O}_k$, avec $a \geq 1$ et $1 \leq k < n-1$. Alors $(\mathcal{E}_{|C_{n-1}})^{(1)}$ est scindé.*

Démonstration. Soient P un point fermé de C_n et $z \in \mathbb{O}_{n,P}$ un générateur de l'idéal de C. On fixe un isomorphisme $\mathcal{E}_P \simeq a\mathbb{O}_{n,P} \oplus \mathbb{O}_{k,P}$. On a alors $(\mathcal{E}_{|C_{n-1}})_P = a\mathbb{O}_{n-1,P} \oplus \mathbb{O}_{k,P}$, et

$$(\mathcal{E}^{(1)})_P = a(z^{n-1}) \oplus (z^{k-1}), \quad ((\mathcal{E}_{|C_{n-1}})^{(1)})_P = a\big((z^{n-2})/(z^{n-1})\big) \oplus (z^{k-1}).$$

L'image du morphisme canonique $\lambda : \mathcal{E}^{(1)} \to (\mathcal{E}_{|C_{n-1}})^{(1)}$ au point P est (z^{k-1}). L'autre facteur $a(z^{n-2})/(z^{n-1})$ est $((\mathcal{E}_{|C_{n-1}})_{n-2})_P$. On a donc

$$(\mathcal{E}_{|C_{n-1}})^{(1)} = (\operatorname{im}\lambda) \oplus (\mathcal{E}_{|C_{n-1}})_{n-2}. \qquad \square$$

4. Dualité et torsion

On considère dans cette section une courbe multiple primitive C_n de courbe réduite associée C. On utilise les notations de 2.1.

4.1. ***Généralités sur la dualité des faisceaux cohérents sur C_n.*** Soient $P \in C$ et M un $\mathbb{O}_{n,P}$-module de type fini. On note $M^{\vee_n}$ le *dual* de M :

$$M^{\vee_n} = \operatorname{Hom}(M, \mathbb{O}_{n,P}).$$

Si aucune confusion n'est à craindre on notera $M^{\vee} = M^{\vee_n}$. Si N est un $\mathbb{O}_{C,P}$-module, on note N^* le dual de N : $N^* = \operatorname{Hom}(N, \mathbb{O}_{C,P})$.

Soit $\mathcal{E}$ un faisceau cohérent sur C_n. On note $\mathcal{E}^{\vee_n}$ le *dual* de $\mathcal{E}$:

$$\mathcal{E}^{\vee_n} = \mathcal{H}om(\mathcal{E}, \mathbb{O}_n).$$

Si aucune confusion n'est à craindre on notera $\mathcal{E}^{\vee} = \mathcal{E}^{\vee_n}$. Si E est un faisceau cohérent sur C, on note E^* le dual de E : $E^* = \mathcal{H}om(E, \mathbb{O}_C)$. Ces notations sont justifiées par le fait que $E^{\vee} \neq E^*$. Plus généralement on a, si i un entier tel que $1 \leq i \leq n$ et $\mathcal{E}$ un faisceau cohérent sur C_i, un isomorphisme canonique

$$\mathcal{E}^{\vee_n} \simeq \mathcal{E}^{\vee_i} \otimes \mathcal{I}_C^{n-i},$$

($\mathcal{I}_C$ désignant le faisceau d'idéaux de C, qui est un fibré en droites sur C_{n-1}). En particulier, pour tout faisceau cohérent E sur C, on a $E^{\vee_n} \simeq E^* \otimes L^{n-1}$ [Drézet 2009, lemme 4.1].

Pour tout entier i tel que $1 \leq i < n$, on a $(\mathcal{E}^{\vee})^{(i)} = (\mathcal{E}_{|C_i})^{\vee}$ [Drézet 2009, proposition 4.2].

4.1.1. *Sous-faisceau de torsion d'un faisceau cohérent sur C_n.* Soient P un point fermé de C et $x \in \mathcal{O}_{nP}$ un élément au dessus d'un générateur de l'idéal maximal de $\mathcal{O}_{CP}$. Soit M un $\mathcal{O}_{nP}$-module de type fini. Le *sous-module de torsion $T(M)$ de M* est constitué des éléments annulés par une puissance de x. On dit que M est *sans torsion* si ce sous-module est nul. C'est donc le cas si et seulement si pour tout $m \in M$ non nul et tout entier $p > 0$ on a $x^p m \neq 0$.

Soit $\mathcal{E}$ un faisceau cohérent sur C_n. Le *sous-faisceau de torsion $T(\mathcal{E})$ de $\mathcal{E}$* est le sous-faisceau maximal de $\mathcal{E}$ dont le support est fini. Pour tout point fermé P de C on a $T(\mathcal{E})_P = T(\mathcal{E}_P)$. On a donc une suite exacte canonique

$$0 \longrightarrow T(\mathcal{E}) \longrightarrow \mathcal{E} \longrightarrow \mathcal{E}^{\vee\vee} \longrightarrow 0.$$

4.1.2. *Faisceaux réflexifs.* Un faisceau cohérent $\mathcal{E}$ sur C_n est réflexif si et seulement si il est sans torsion [Drézet 2009, théorème 4.4], si et seulement si $\mathcal{E}^{(1)}$ est localement libre sur C [Drézet 2009, proposition 3.8].

4.2. *Dualité des faisceaux de torsion.*

Soit $\mathbb{T}$ un faisceau de torsion sur C_n. Alors on a évidemment $\mathbb{T}^\vee = 0$. On appelle *dual* de $\mathbb{T}$ le faisceau

$$D_n(\mathbb{T}) = \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n).$$

S'il n'y a pas d'ambiguïté sur n, on notera plus simplement $\tilde{\mathbb{T}} = D_n(\mathbb{T})$. Rappelons que $\mathcal{E}xt^i_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n) = 0$ pour tout $i \geq 2$, d'après le corollaire 4.6 de [Drézet 2009].

Proposition 4.2.1. *Soit $\mathbb{T}$ un faisceau de torsion sur C_i, $1 \leq i < n$. Alors on a un isomorphisme canonique*

$$D_n(\mathbb{T}) \simeq D_i(\mathbb{T}) \otimes \mathbb{L}^{n-i}.$$

Bien sûr on a $D_i(\mathbb{T}) \otimes \mathbb{L}^{n-i} \simeq D_i(\mathbb{T})$.

Démonstration. D'après la proposition 2.1 de [Drézet 2009] on a un isomorphisme $D_i(\mathbb{T}) \simeq \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_i)$. On considère la suite exacte

$$0 \longrightarrow \mathcal{O}_i \otimes \mathbb{L}^{n-i} \longrightarrow \mathcal{O}_n \xrightarrow{\ r\ } \mathcal{O}_{n-i} \longrightarrow 0.$$

Il suffit de montrer que le morphisme induit par r

$$\Phi : \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n) \longrightarrow \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_{n-i})$$

est nul.

On considère une résolution localement libre de $\mathbb{T}$:

$$\cdots \mathbb{E}_2 \xrightarrow{f_2} \mathbb{E}_1 \xrightarrow{f_1} \mathbb{E}_0 \longrightarrow \mathbb{T} \longrightarrow 0.$$

Soient P un point fermé de C et $z \in \mathcal{O}_{nP}$ une équation de C. Alors $\mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n)$ est isomorphe à la cohomologie de degré 1 du complexe dual

$$\mathbb{E}_0^\vee \xrightarrow{\ {}^tf_1\ } \mathbb{E}_1^\vee \xrightarrow{\ {}^tf_2\ } \mathbb{E}_2^\vee \cdots$$

et $\mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_{n-i})$ est isomorphe à la cohomologie de degré 1 du complexe obtenu en restreignant le précédent à C_{n-i}. Le morphisme Φ provient du morphisme de complexes

$$\begin{array}{ccccc} \mathbb{E}_0^\vee & \xrightarrow{\ {}^tf_1\ } & \mathbb{E}_1^\vee & \xrightarrow{\ {}^tf_2\ } & \mathbb{E}_2^\vee \cdots \\ \downarrow{\scriptstyle \pi_0} & & \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} \\ (\mathbb{E}_0^\vee)_{|C_{n-i}} & \xrightarrow{\ {}^tf_1\ } & (\mathbb{E}_1^\vee)_{|C_{n-i}} & \xrightarrow{\ {}^tf_2\ } & (\mathbb{E}_2^\vee)_{|C_{n-i}} \cdots \end{array}$$

(les flèches verticales étant les restrictions).

Soient P un point du support de $\mathbb{T}$ et $z \in \mathcal{O}_{nP}$ une équation de C. Soient $\alpha \in \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n)_P$ et $u \in \ker {}^tf_2$ au dessus de α. Puisque $\mathbb{T}$ est concentré sur C_i, la multiplication par $z^i : \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n)_P \to \mathcal{E}xt^1_{\mathcal{O}_n}(\mathbb{T}, \mathcal{O}_n)_P$ est nulle. Donc $z^i u \in \operatorname{im} {}^tf_1$, et on peut écrire $z^i u = {}^tf_1(\theta)$, avec $\theta \in (\mathbb{E}_0^\vee)_P$. On va montrer que θ est multiple de z^i. Pour cela on suppose que ce n'est pas le cas, et on va aboutir à une contradiction. On a donc $\theta = z^k\theta'$, avec $0 \leq k < i$ et θ' non multiple de z. On a ${}^tf_1(z^{n-i+k}\theta') = z^n u = 0$, et puisque ${}^tf_1$ est injectif, on a $z^{n-i+k}\theta' = 0$. Puisque $n - i + k < n$, il en découle que θ' est multiple de z, ce qui est la contradiction recherchée. On peut donc écrire $\theta = z^i\theta'$, d'où $z^i(u - {}^tf_1(\theta')) = 0$, et il en découle qu'on peut écrire u sous la forme $u = {}^tf_1(\theta') + z^{n-i}\rho$. Il en découle que $\pi_1(u) = {}^tf_1(\pi_0(\theta'))$. On a donc $\Phi_P(\alpha) = 0$. □

Corollaire 4.2.2. *Soit $\mathbb{T}$ un faisceau de torsion sur C_n. Alors on a $h^0(\mathbb{T}) = h^0(\tilde{\mathbb{T}})$.*

Démonstration. D'après la proposition 4.2.1, on a, pour tout faisceau de torsion T sur C, $D_n(T) \simeq T$. Le corollaire en découle, en utilisant par exemple la première filtration canonique de $\mathbb{T}$. □

Les faisceaux de torsion sur C_n et les morphismes entre eux constituent une catégorie abélienne et noethérienne $\mathcal{T}_n(C_n)$, qui est évidemment une sous-catégorie pleine de celle des faisceaux cohérents sur C_n. La dualité définit un foncteur contravariant exact

$$D_n : \mathcal{T}_n(C_n) \longrightarrow \mathcal{T}_n(C_n).$$

Proposition 4.2.3. *Le foncteur D_n est une involution. Donc si $\mathbb{T}$ est un faisceau de torsion sur C_n, il existe un isomorphisme canonique*

$$\tilde{\tilde{\mathbb{T}}} \simeq \mathbb{T}.$$

Démonstration. Il existe un fibré vectoriel $\mathbb{E}$ et un morphisme surjectif $f : \mathbb{E} \to \mathbb{T}$. Alors $\mathcal{E} = \ker f$ est un faisceau sans torsion, donc réflexif. On obtient donc en dualisant la suite exacte $0 \to \mathcal{E} \to \mathbb{E} \to \mathbb{T} \to 0$ les suivantes :

$$0 \longrightarrow \mathbb{E}^\vee \longrightarrow \mathcal{E}^\vee \longrightarrow \tilde{\mathbb{T}} \longrightarrow 0, \quad 0 \longrightarrow \mathcal{E} \longrightarrow \mathbb{E} \longrightarrow \tilde{\tilde{\mathbb{T}}} \longrightarrow 0.$$

Le résultat en découle aisément. □

Si $\mathbb{T}$ est un faisceau de torsion sur C_n, l'entier $h^0(\mathbb{T})$ s'appelle la *longueur* de T. On a

$$h^0(\mathbb{T}) = \sum_{P \in C} \dim_{\mathbb{C}}(\mathbb{T}_P).$$

Lemme 4.2.4. *Soit* $\mathbb{T}$ *un faisceau de torsion sur* C_n. *Alors on a* $h^0(G_i(\mathbb{T})) = h^0(G^{(i+1)}(\mathbb{T}))$ *pour* $0 \le i < n$.

Démonstration. Découle aisément du [Drézet 2009, corollaire 3.4]. □

Corollaire 4.2.5. *Soit* $\mathbb{T}$ *un faisceau de torsion sur* C_n. *Alors on a, pour* $1 \le i \le n$, *des isomorphismes canoniques*

$$[\tilde{\mathbb{T}}]_i \simeq [\widetilde{\mathbb{T}_i}] \otimes \mathbb{L}^i, \qquad (\tilde{\mathbb{T}})^{(i)} \simeq \widetilde{\mathbb{T}/\mathbb{T}_i}, \qquad G^{(i+1)}(\tilde{\mathbb{T}}) \simeq \widetilde{G_i(\mathbb{T})}.$$

Démonstration. De la suite exacte $0 \to \mathbb{T}_i \to \mathbb{T} \to \mathbb{T}/\mathbb{T}_i \to 0$ on déduit la suivante :

$$0 \longrightarrow \widetilde{\mathbb{T}/\mathbb{T}_i} \longrightarrow \tilde{\mathbb{T}} \longrightarrow [\widetilde{\mathbb{T}_i}] \longrightarrow 0.$$

D'après la proposition 4.2.1, $\widetilde{\mathbb{T}/\mathbb{T}_i}$ est concentré sur C_i. On a donc $\widetilde{\mathbb{T}/\mathbb{T}_i} \subset (\tilde{\mathbb{T}})^{(i)}$. Mais le lemme 4.2.4 entraine que $h^0(\widetilde{\mathbb{T}/\mathbb{T}_i}) = h^0((\tilde{\mathbb{T}})^{(i)})$, donc on a en fait l'égalité. Il en découle que $[\widetilde{\mathbb{T}_i}] \simeq [\tilde{\mathbb{T}}]_i \otimes \mathbb{L}^{-i}$.

Le dernier isomorphisme découle de la suite exacte

$$0 \longrightarrow G^{(i+1)}(\tilde{\mathbb{T}}) \longrightarrow [\tilde{\mathbb{T}}]_i \otimes \mathbb{L} \longrightarrow [\tilde{\mathbb{T}}]_{i+1} \longrightarrow 0$$

(voir [Drézet 2009, lemme 3.2]), du fait que par définition on a $G_i(\mathbb{T}) = \mathbb{T}_i/\mathbb{T}_{i+1}$, et du premier isomorphisme. □

4.3. *Dualité des faisceaux sans torsion.* Soit $\mathcal{E}$ un faisceau cohérent sans torsion sur C_n. Il est donc réflexif (voir 4.1.2). Les faisceaux $\mathcal{E}_i$, $\mathcal{E}^{(i)}$ le sont donc aussi, étant des sous-faisceaux de $\mathcal{E}$. Mais les faisceaux $\mathcal{E}/\mathcal{E}_i$ ne le sont pas en général. On note $\Sigma_i(\mathcal{E})$ le sous-faisceau de torsion de $\mathcal{E}/\mathcal{E}_i$, et $T_i(\mathcal{E})$ celui de $G_i(\mathcal{E})$.

Pour $1 \le i < n$, on note $\mathcal{E}[i]$ le noyau du morphisme canonique surjectif

$$\mathcal{E} \longrightarrow\!\!\!\!\!\rightarrow \mathcal{E}_{|C_i} \longrightarrow\!\!\!\!\!\rightarrow (\mathcal{E}_{|C_i})^{\vee\vee}.$$

Proposition 4.3.1. *Soit* $\mathcal{E}$ *un faisceau cohérent sans torsion sur* C_n. *Alors, pour* $1 \le i < n$:

(i) *On a un isomorphisme* $\Sigma_i(\mathcal{E}^\vee) \simeq \widetilde{\Sigma_i(\mathcal{E})} \otimes \mathbb{L}^i$, *et une suite exacte*

$$0 \longrightarrow (\mathcal{E}^\vee)_i \longrightarrow (\mathcal{E}_i)^\vee \otimes \mathbb{L}^i \longrightarrow \Sigma_i(\mathcal{E}^\vee) \longrightarrow 0$$

canoniques.

(ii) *On a un isomorphisme canonique* $\mathcal{E}[i]^\vee \simeq (\mathcal{E}^\vee)_i \otimes \mathbb{L}^{-i}$.

(iii) *Il existe un morphisme canonique* $\phi_i(\mathcal{E}) : \Sigma_{i+1}(\mathcal{E}) \to \Sigma_i(\mathcal{E})$ *tel que* $\ker \phi_i(\mathcal{E}) \simeq T_i(\mathcal{E})$, *et que* $\operatorname{coker} \phi_i(\mathcal{E}) = R_i(\mathcal{E})$ *soit concentré sur* C.

(iv) *Il existe une inclusion canonique*

$$G^{(i+1)}(\mathcal{E}^\vee) \hookrightarrow G_i(\mathcal{E})^* \otimes L^{n-1}$$

telle que le quotient soit isomorphe à $R_i(\mathcal{E})$.

Démonstration. En dualisant la suite exacte $0 \to \mathcal{E}_i \to \mathcal{E} \to \mathcal{E}/\mathcal{E}_i \to 0$, on obtient la suite exacte

$$0 \longrightarrow (\mathcal{E}/\mathcal{E}_i)^\vee \longrightarrow \mathcal{E}^\vee \longrightarrow (\mathcal{E}_i)^\vee \longrightarrow \mathcal{E}xt^1_{\mathcal{O}_n}(\mathcal{E}/\mathcal{E}_i, \mathcal{O}_n) = \widetilde{\Sigma_i(\mathcal{E})} \longrightarrow 0.$$

D'après la proposition 4.2 de [Drézet 2009] on a $(\mathcal{E}/\mathcal{E}_i)^\vee = (\mathcal{E}^\vee)^{(i)}$. On en déduit la suite exacte

$$0 \longrightarrow (\mathcal{E}^\vee)_i \otimes \mathbb{L}^{-i} \longrightarrow (\mathcal{E}_i)^\vee \longrightarrow \widetilde{\Sigma_i(\mathcal{E})} \longrightarrow 0. \tag{4-1}$$

En la dualisant et tensorisant par $\mathbb{L}^{-i}$, et en utilisant la proposition 4.2.3 on obtient la suite exacte

$$0 \longrightarrow \mathcal{E}_i \otimes \mathbb{L}^{-i} \longrightarrow ((\mathcal{E}^\vee)_i)^\vee \longrightarrow \Sigma_i(\mathcal{E}) \otimes \mathbb{L}^{-i} \longrightarrow 0, \tag{4-2}$$

qui est (4-1) avec $\mathcal{E}^\vee$ à la place de $\mathcal{E}$. On obtient donc l'isomorphisme canonique de (i). On en déduit (ii) en dualisant la suite exacte $0 \to \mathcal{E}_i \to \mathcal{E}[i] \to \Sigma_i(\mathcal{E}) \to 0$.

On a un diagramme commutatif avec lignes et colonnes exactes

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \longrightarrow & \mathcal{E}_{i+1} & \longrightarrow & ((\mathcal{E}^\vee)_{i+1})^\vee \otimes \mathbb{L}^{i+1} & \longrightarrow & \Sigma_{i+1}(\mathcal{E}) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & & & \\
0 & \longrightarrow & \mathcal{E}_i & \longrightarrow & ((\mathcal{E}^\vee)_i)^\vee \otimes \mathbb{L}^i & \longrightarrow & \Sigma_i(\mathcal{E}) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & & & \\
 & & G_i(\mathcal{E}) & & G^{(i+1)}(\mathcal{E}^\vee)^\vee & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

où les suites horizontales proviennent de (4-2) et la suite verticale du milieu du [Drézet 2009, lemme 3.2]. On en déduit aisément (iii) et (iv). □

4.4. *Invariants du dual d'un faisceau cohérent.*

Proposition 4.4.1. *Soit $\mathcal{E}$ un faisceau cohérent. Alors on a*

$$R(\mathcal{E}^\vee) = R(\mathcal{E}), \quad \mathrm{Deg}(\mathcal{E}^\vee) = -\mathrm{Deg}(\mathcal{E}) + R(\mathcal{E})(n-1)\deg(L) + h^0(T(\mathcal{E})).$$

Démonstration. La première assertion concernant les rangs est immédiate, par exemple en se plaçant sur l'ouvert où $\mathcal{E}$ est quasi localement libre. Démontrons la seconde. Soit $\mathcal{F} = \mathcal{E}/T(\mathcal{E})$, qui est un faisceau sans torsion. On a $\mathrm{Deg}(\mathcal{E}) = \mathrm{Deg}(\mathcal{F})+h^0(T(\mathcal{E}))$, $R(\mathcal{E}) = R(\mathcal{F})$ et $\mathcal{E}^\vee = \mathcal{F}^\vee$, donc la seconde assertion équivaut à

$$\mathrm{Deg}(\mathcal{F}^\vee) = -\mathrm{Deg}(\mathcal{F}) + R(\mathcal{F})(n-1)\deg(L).$$

On peut donc supposer que $\mathcal{E}$ est sans torsion. On va montrer que

$$\mathrm{Deg}(\mathcal{E}^\vee) = -\mathrm{Deg}(\mathcal{E}) + R(\mathcal{E})(n-1)\deg(L) \tag{4-3}$$

par récurrence sur n. Si $n = 1$ c'est évident. Supposons que $n > 1$ et que (4-3) soit vraie pour $n-1$. On a donc

$$\mathrm{Deg}((\mathcal{E}_1)^{\vee_{n-1}}) = -\mathrm{Deg}(\mathcal{E}_1) + R(\mathcal{E}_1)(n-2)\deg(L).$$

Mais d'après 4.1 on a $(\mathcal{E}_1)^\vee = (\mathcal{E}_1)^{\vee_{n-1}} \otimes \mathcal{I}_C$, donc

$$\mathrm{Deg}((\mathcal{E}_1)^\vee) = \mathrm{Deg}((\mathcal{E}_1)^{\vee_{n-1}}) + R(\mathcal{E}_1)\deg(L),$$

d'où

$$\mathrm{Deg}((\mathcal{E}_1)^\vee) = -\mathrm{Deg}(\mathcal{E}_1) + R(\mathcal{E}_1)(n-1)\deg(L) \tag{4-4}$$

(c'est-à-dire que (4-3) est vraie pour $\mathcal{E}_1$). D'après la suite exacte

$$0 \longrightarrow \mathcal{E}_1 \longrightarrow \mathcal{E} \longrightarrow \mathcal{E}_{|C} \longrightarrow 0$$

on a $\mathrm{Deg}(\mathcal{E}) = \mathrm{Deg}(\mathcal{E}_1) + \mathrm{Deg}(\mathcal{E}_{|C})$. Soit $T = T(\mathcal{E}_{|C})$. On a une suite exacte

$$0 \longrightarrow (\mathcal{E}_{|C})^\vee \longrightarrow \mathcal{E}^\vee \longrightarrow (\mathcal{E}_1)^\vee \longrightarrow \tilde{T} \longrightarrow 0,$$

donc

$$\mathrm{Deg}(\mathcal{E}^\vee) = \mathrm{Deg}((\mathcal{E}_{|C})^\vee) + \mathrm{Deg}((\mathcal{E}_1)^\vee) - h^0(T).$$

Mais

$$\mathrm{Deg}((\mathcal{E}_{|C})^\vee) - h^0(T) = -\mathrm{Deg}(\mathcal{E}_{|C}) + (n-1)R(\mathcal{E}_{|C})\deg(L),$$

car $(\mathscr{E}_{|C})^\vee = (\mathscr{E}_{|C})^* \otimes L^{n-1}$. Donc

$$\begin{aligned}\deg(\mathscr{E}^\vee) &= \mathrm{Deg}((\mathscr{E}_1)^\vee) - \mathrm{Deg}(\mathscr{E}_{|C}) + (n-1)R(\mathscr{E}_{|C})\deg(L)\\ &= -\mathrm{Deg}(\mathscr{E}) + R(\mathscr{E})(n-1)\deg(L)\end{aligned}$$

d'après (4-4). □

Corollaire 4.4.2. *Soit $\mathscr{E}$ un faisceau cohérent réflexif sur C_n. Alors, pour $1 \le i < n$, on a*

$$\begin{gathered}R((\mathscr{E}^\vee)_i) = R(\mathscr{E}_i), \quad R((\mathscr{E}^\vee)^{(i)}) = R(\mathscr{E}^{(i)}), \quad R((\mathscr{E}^\vee)_{|C_i}) = R(\mathscr{E}_{|C_i}),\\ \mathrm{Deg}((\mathscr{E}^\vee)_i) = -\mathrm{Deg}(\mathscr{E}_i) + (n+i-1)R(\mathscr{E}_i)\deg(L) - h^0(\Sigma_i(\mathscr{E})),\\ \mathrm{Deg}((\mathscr{E}^\vee)_{|C_i}) = \mathrm{Deg}((\mathscr{E}_{|C_i})^\vee) - iR(\mathscr{E}_i)\deg(L).\end{gathered}$$

Démonstration. Découle aisément des propositions 4.3.1 et 4.4.1. □

Corollaire 4.4.3. *Soient $\mathscr{E}$ un faisceau cohérent réflexif sur C_n et i un entier tel que $1 \le i < n$ et $R(\mathscr{E}_i) > 0$. Alors on a*

$$\mu((\mathscr{E}^\vee)_{|C_i}) - \mu((\mathscr{E}^\vee)_i) = \mu(\mathscr{E}_i \otimes \mathbb{L}^{-i}) - \mu(\mathscr{E}^{(i)}) + h^0(\Sigma_i(\mathscr{E}))\Big(\frac{1}{R(\mathscr{E}^{(i)})} + \frac{1}{R(\mathscr{E}_i)}\Big).$$

5. Conditions d'existence des faisceaux (semi-)stables

Dans toute la suite de l'article on considère une courbe multiple primitive C_n de courbe réduite associée C. On utilise les notations de 2.1, et on suppose que $\deg(L) < 0$.

5.1. *Critères de (semi-)stabilité.*

Lemme 5.1.1. *Soient A, A'', B, B'', E, E'' des faisceaux cohérents de rang positif sur C_n, tels que*

$$\begin{aligned}R(E) &= R(A) + R(B), & R(E'') &= R(A'') + R(B''),\\ \mathrm{Deg}(E) &= \mathrm{Deg}(A) + \mathrm{Deg}(B), & \mathrm{Deg}(E'') &= \mathrm{Deg}(A'') + \mathrm{Deg}(B'').\end{aligned}$$

On suppose qu'on a $\mu(B) \ge \mu(A)$, $\mu(A'') \ge \mu(A)$, $\mu(B'') \ge \mu(B)$, et que

$$\frac{R(E'')}{R(E)} \ge \frac{R(A'')}{R(A)}.$$

Alors on a $\mu(E'') \ge \mu(E)$. Si de plus $\mu(A'') > \mu(A)$ ou $\mu(B'') > \mu(B)$, alors on a $\mu(E'') > \mu(E)$.

Démonstration. D'après les hypothèses, $R(E'')/R(E) \ge R(A'')/R(A)$ équivaut à $R(B'')/R(B) \ge R(A'')/R(A)$, et $\mu(E'') - \mu(E) = \Delta/R(E)R(E'')$, avec

$$\Delta = (\mathrm{Deg}(A'')+\mathrm{Deg}(B''))(R(A)+R(B)) - (\mathrm{Deg}(A)+\mathrm{Deg}(B))(R(A'')+R(B'')).$$

On a

$$\operatorname{Deg}(A'') \geq \operatorname{Deg}(A)\frac{R(A'')}{R(A)}, \quad \operatorname{Deg}(B'') \geq \operatorname{Deg}(B)\frac{R(B'')}{R(B)},$$

Donc $\Delta \geq \Delta'$, avec

$$\begin{aligned}\Delta' = \Big(\operatorname{Deg}(A)\frac{R(A'')}{R(A)} + \operatorname{Deg}(B)\frac{R(B'')}{R(B)}\Big)&(R(A)+R(B)) \\ &- (\operatorname{Deg}(A)+\operatorname{Deg}(B))(R(A'')+R(B'')) \\ = (\mu(B)-\mu(A))(R(B'')R(A) - R(A'')R(B)).&\end{aligned}$$

Le résultat en découle immédiatement. □

Théorème 5.1.2. *Soient $\mathscr{E}$ un faisceau cohérent sans torsion sur C_n et k un entier tel que $1 \leq k < n$ et que $\mathscr{E}_k \neq 0$. On suppose que*

$$\mu(\mathscr{E}^{(k)}) \leq \mu(\mathscr{E}), \quad \mu((\mathscr{E}^\vee)^{(k)}) \leq \mu(\mathscr{E}^\vee). \tag{5-1}$$

Alors, si $\mathscr{E}[k]$, $(\mathscr{E}_{|C_k})^{\vee\vee}$, $(\mathscr{E}^\vee)[k]$ et $((\mathscr{E}^\vee)_{|C_k})^{\vee\vee}$ sont semi-stables il en est de même de $\mathscr{E}$.

Si de plus les inégalités de (5-1) *sont strictes, et si $\mathscr{E}[k]$ ou $(\mathscr{E}_{|C_k})^{\vee\vee}$, ainsi que $(\mathscr{E}^\vee)[k]$ ou $((\mathscr{E}^\vee)_{|C_k})^{\vee\vee}$, sont stables, alors $\mathscr{E}$ est stable.*

Démonstration. Supposons que les hypothèses du théorème soient vérifiées. Soit

$$\mathscr{E} \twoheadrightarrow \mathscr{E}''$$

un quotient de $\mathscr{E}$. Il faut montrer que $\mu(\mathscr{E}'') \geq \mu(\mathscr{E})$. On peut supposer que $\mathscr{E}''$ est sans torsion. On a un diagramme commutatif avec lignes exactes

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathscr{E}[k] & \longrightarrow & \mathscr{E} & \longrightarrow & (\mathscr{E}_{|C_k})^{\vee\vee} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathscr{E}''[k] & \longrightarrow & \mathscr{E}'' & \longrightarrow & (\mathscr{E}''_{|C_k})^{\vee\vee} & \longrightarrow & 0 \end{array}$$

où les deux flèches verticales de droite sont surjectives. Le cas où $\mathscr{E}''[k] = 0$ est évident. On supposera donc que $\mathscr{E}''[k] \neq 0$. Remarquons que les inégalités (5-1) équivalent à

$$\mu((\mathscr{E}_{|C_k})^{\vee\vee}) \geq \mu(\mathscr{E}[k]), \quad \mu(((\mathscr{E}^\vee)_{|C_k})^{\vee\vee}) \geq \mu(\mathscr{E}^\vee[k]),$$

car $(\mathscr{E}^\vee)^{(k)} = (\mathscr{E}_{|C_k})^\vee$. Le morphisme vertical de droite du diagramme précédent est surjectif, donc on a $\mu((\mathscr{E}''_{|C_k})^{\vee\vee}) \geq \mu((\mathscr{E}_{|C_k})^{\vee\vee})$ d'après la semi-stabilité de $(\mathscr{E}_{|C_k})^{\vee\vee}$. Le conoyau du morphisme vertical de gauche est de torsion, donc on a $\mu(\mathscr{E}''[k]) \geq \mu(\mathscr{E}[k])$ d'après la semi-stabilité de $\mathscr{E}[k]$. D'après le lemme 5.1.1 on a $\mu(\mathscr{E}'') \geq \mu(\mathscr{E})$ si

$$\frac{R(\mathscr{E}'')}{R(\mathscr{E})} \geq \frac{R(\mathscr{E}''[k])}{R(\mathscr{E}[k])}.$$

On peut donc supposer que

$$\frac{R(\mathscr{E}'')}{R(\mathscr{E})} < \frac{R(\mathscr{E}''[k])}{R(\mathscr{E}[k])}. \tag{5-2}$$

On utilise maintenant la suite exacte

$$0 \longrightarrow \mathscr{E}''^{\vee} \longrightarrow \mathscr{E}^{\vee} \longrightarrow \mathscr{E}'^{\vee} \longrightarrow 0$$

obtenue en utilisant le fait que $\mathscr{E}''$ est réflexif. D'après la proposition 4.4.1, $\mu(\mathscr{E}'') \geq \mu(\mathscr{E})$ équivaut à $\mu(\mathscr{E}'^{\vee}) \geq \mu(\mathscr{E}^{\vee})$, et d'après le lemme 5.1.1, cette inégalité est vérifiée si

$$\frac{R(\mathscr{E}')}{R(\mathscr{E})} = \frac{R(\mathscr{E}'^{\vee})}{R(\mathscr{E}^{\vee})} \geq \frac{R(\mathscr{E}'^{\vee}[k])}{R(\mathscr{E}^{\vee}[k])}. \tag{5-3}$$

D'après 4.1.1, on a $R(\mathscr{E}'^{\vee}[k]) = R(\mathscr{E}'[k])$ et $R(\mathscr{E}^{\vee}[k]) = R(\mathscr{E}[k])$. Donc (5-3) équivaut à

$$\frac{R(\mathscr{E}')}{R(\mathscr{E})} \geq \frac{R(\mathscr{E}'[k])}{R(\mathscr{E}[k])}. \tag{5-4}$$

Puisque $\mathscr{E}'_k$ est contenu dans le noyau du morphisme canonique surjectif $\mathscr{E}_k \to \mathscr{E}''_k$, on a

$$R(\mathscr{E}'[k]) = R(\mathscr{E}'_k) \leq R(\mathscr{E}_k) - R(\mathscr{E}''_k) = R(\mathscr{E}[k]) - R(\mathscr{E}''[k]),$$

donc on peut écrire

$$R(\mathscr{E}'[k]) = R(\mathscr{E}[k]) - R(\mathscr{E}''[k]) - \eta,$$

avec $\eta \geq 0$. Donc (5-4) s'écrit

$$\frac{R(\mathscr{E}'')}{R(\mathscr{E})} \leq \frac{R(\mathscr{E}''[k]) + \eta}{R(\mathscr{E}[k])}.$$

L'inégalité précédente est vraie d'après (5-2). On a donc bien $\mu(\mathscr{E}'') \geq \mu(\mathscr{E})$.

L'assertion concernant la stabilité se démontre de manière analogue. □

5.2. ***Le cas des fibrés vectoriels.*** On suppose que le genre de C est $g \geq 2$.

Théorème 5.2.1. *Soit $\mathbb{E}$ un fibré vectoriel sur C_n. Alors, si $\mathbb{E}_{|C}$ est semi-stable* (*ou stable*), *il en est de même de $\mathbb{E}$.*

Démonstration. Posons $E = \mathbb{E}_{|C}$. Les filtrations canoniques de $\mathbb{E}$ sont identiques, et leurs gradués sont

$$\big(G_0(\mathscr{E}), G_1(\mathscr{E}), \ldots, G_{n-1}(\mathscr{E})\big) = (E, E \otimes L, \ldots, E \otimes L^{n-1}).$$

Les inégalités (5-1) sont trivialement vérifiées (car $\deg(L) < 0$) pour tout entier k tel que $1 \leq k < n$.

Le théorème 5.2.1 se démontre par récurrence sur n : pour $n = 1$ c'est évident. Supposons que ce soit vrai pour $n - 1 \geq 1$. Alors $\mathbb{E}_1$ est semi-stable (ou stable). Le théorème 5.1.2 permet alors de conclure qu'il en est de même de $\mathbb{E}$. □

5.2.2. *Variétés de modules.* Soient r, δ des entiers tels que $r \geq 1$. Posons

$$R = nr, \quad d = n\delta + \tfrac{1}{2}n(n-1)r\deg(L).$$

Pour tout fibré vectoriel $\mathbb{E}$ sur C_n tel que $\mathbb{E}_{|C}$ soit de rang r et de degré δ, on a $R(\mathbb{E}) = R$ et $\mathrm{Deg}(\mathcal{E}) = d$. Il découle de 2.5.3 que la variété de modules $\mathbb{M}(R, d)$ des fibrés vectoriels stables de rang généralisé R et de degré généralisé d sur C_n est non vide. C'est un ouvert irréductible et lisse de la variété de modules $\mathcal{M}(R, d)$ des faisceaux stables de rang généralisé R et de degré généralisé d. Pour calculer sa dimension on considère un fibré stable $\mathbb{E}$ tel que $R(\mathbb{E}) = R$ et $\mathrm{Deg}(\mathbb{E}) = d$. On a

$$\dim(\mathbb{M}(R, d)) = 1 - \chi(\mathbb{E}^\vee \otimes \mathbb{E}) = 1 + nr^2(g-1) - \tfrac{1}{2}n(n-1)r^2\deg(L).$$

5.3. ***Le cas des faisceaux quasi localement libres de type générique.*** On suppose que le genre de C est $g \geq 2$.

Théorème 5.3.1. *Soient a, k des entiers tels que $a > 0$ et $1 \leq k < n$. Soit $\mathcal{E}$ un faisceau quasi localement libre de type rigide, localement isomorphe à $a\mathcal{O}_n \oplus \mathcal{O}_k$ et tel que*

$$\mu(V_\mathcal{E}) + \tfrac{1}{2}n\deg(L) \leq \mu(F_\mathcal{E}) \leq \mu(E_\mathcal{E}) - \tfrac{1}{2}n\deg(L). \tag{5-5}$$

Alors si $E_\mathcal{E}$, $F_\mathcal{E}$ et $V_\mathcal{E}$ sont semi-stables, il en est de même de $\mathcal{E}$.

Si les inégalités précédentes sont strictes, et si $E_\mathcal{E}$, $F_\mathcal{E}$ et $V_\mathcal{E}$ sont stables, il en est de même de $\mathcal{E}$.

Démonstration. On ne démontrera que la première assertion, la seconde étant analogue. On utilise les notations de 3.1. Supposons les inégalités (5-5) vérifiées et $E_\mathcal{E}$, $F_\mathcal{E}$ et $V_\mathcal{E}$ semi-stables. Alors on a

$$\mathcal{E}[k] = \mathcal{E}_k = \mathbb{F}, \quad \mathcal{E}_{|C_k} = \mathbb{E}, \quad \mathcal{E}^\vee[k] = (\mathcal{E}^\vee)_k = \mathbb{F} \otimes \mathbb{L}^k, \quad (\mathcal{E}^\vee)_{|C_k} = \mathbb{V}^\vee.$$

Donc d'après le théorème 5.2.1, $\mathcal{E}[k]$, $\mathcal{E}_{|C_k}$, $\mathcal{E}^\vee[k]$ et $(\mathcal{E}^\vee)_{|C_k}$ sont semi-stables. Un calcul simple montre que les inégalités (5-5) équivalent aux inégalités (5-1). La semi-stabilité de $\mathcal{E}$ découle donc du théorème 5.1.2. □

La semi-stabilité de $E_\mathcal{E}$, $F_\mathcal{E}$ et $V_\mathcal{E}$ entraine d'autres inégalités :

$$\mu(E_\mathcal{E}) \leq \mu(F_\mathcal{E}), \qquad \mu(F_\mathcal{E}) \leq \mu(V_\mathcal{E})$$

(car il existe un morphisme surjectif $E_{\mathcal{E}} \to F_{\mathcal{E}}$ et un morphisme injectif $F_{\mathcal{E}} \to V_{\mathcal{E}}$). Les inégalités précédentes et (5-5) équivalent aux inégalités

$$\mu(E_{\mathcal{E}}) \leq \mu(F_{\mathcal{E}}) \leq \mu(E_{\mathcal{E}}) - \frac{n-k}{a+1}\deg(L).$$

5.3.2. *Variétés de modules de faisceaux stables.* Soient a, k, ϵ, δ des entiers, avec $a \geq 1$ et $1 \leq k < n$. Soient

$$R = an + k, \qquad d = k\epsilon + (n-k)\delta + \tfrac{1}{2}\big(n(n-1)a + k(k-1)\big)\deg(L).$$

On note $\mathcal{M}(R, d)$ la variété de modules des faisceaux stables de rang généralisé R et de degré généralisé d sur C_n. Les faisceaux quasi localement libres $\mathcal{E}$ de type générique stables localement isomorphes à $a\mathcal{O}_n \oplus \mathcal{O}_k$ et tels que $E_{\mathcal{E}}$ et $F_{\mathcal{E}}$ soient de rang $a+1$ et a (respectivement) et de degré ϵ et δ constituent un ouvert irréductible de $\mathcal{M}(R, d)$, dont la sous-variété réduite associée est notée $\mathcal{N}(a, k, \delta, \epsilon)$.

Théorème 5.3.3. *Si on a*

$$\frac{\epsilon}{a+1} < \frac{\delta}{a} < \frac{\epsilon - (n-k)\deg(L)}{a+1}$$

$\mathcal{N}(a, k, \delta, \epsilon)$ *est non vide.*

Démonstration. Les hypothèses et les résultats de [Russo et Teixidor i Bigas 1999] impliquent qu'il existe des fibrés stables E, F, V sur C, tels que

$$\begin{aligned} &\operatorname{rg} E = a+1, \quad \deg(E) = \epsilon, \quad \operatorname{rg} F = a, \quad \deg(F) = \delta, \\ &\operatorname{rg} V = a+1, \quad \deg(V) = \epsilon - (n-k)\deg(L), \end{aligned}$$

et tels qu'il existe une suite exacte

$$0 \longrightarrow F \otimes L^{n-k} \longrightarrow V \otimes L^{n-k} \longrightarrow E \longrightarrow F \longrightarrow 0.$$

D'après la proposition 3.4.1 il existe un faisceau quasi localement libre de type rigide $\mathcal{E}$, localement isomorphe à $a\mathcal{O}_n \oplus \mathcal{O}_k$ et tel que $(*)_{\mathcal{E}}$ soit isomorphe à la suite exacte précédente. D'après le théorème 5.3.1, $\mathcal{E}$ est stable, et définit donc un point de $\mathcal{N}(a, k, \delta, \epsilon)$. □

5.4. ***Exemple d'application à des faisceaux non quasi localement libres.*** Soient $\mathbb{E}$ un fibré vectoriel sur C_n, $E = \mathbb{E}_{|C}$ et Z un ensemble fini de points de C. On pose $z = h^0(\mathcal{O}_Z)$. Soient $\phi : \mathbb{E} \to \mathcal{O}_Z$ un morphisme surjectif, et $\mathcal{E}_\phi = \ker\phi$. On a deux suites exactes

$$0 \longrightarrow \mathcal{E}_\phi \longrightarrow \mathbb{E} \longrightarrow \mathcal{O}_Z \longrightarrow 0, \quad 0 \longrightarrow \mathbb{E}^\vee \longrightarrow \mathcal{E}_\phi^\vee \longrightarrow \mathcal{O}_Z \longrightarrow 0.$$

Le morphisme ϕ se factorise par E. On note E_ϕ le noyau du morphisme induit $E \to \mathcal{O}_Z$. On note $\mathcal{E}'_\phi$ le noyau du morphisme induit $\mathbb{E}_{|C_{n-1}} \to \mathcal{O}_Z$.

Lemme 5.4.1. *On a* $\mathscr{E}_\phi[1] = \mathbb{E}_1$,

$$(\mathscr{E}_{\phi|C})^{\vee\vee} = E_\phi, \quad \mathscr{E}_\phi^\vee[1] = (\mathscr{E}'_\phi)^\vee, \quad \big(((\mathscr{E}_\phi)^\vee)_{|C}\big)^{\vee\vee} = E^*.$$

Démonstration. Il suffit de le démontrer en un point P de Z. Soit $z \in \mathcal{O}_{n,P}$ une équation de C et $x \in \mathcal{O}_{n,P}$ au dessus d'un générateur de l'idéal maximal de P dans $\mathcal{O}_C$. Si $r = \operatorname{rg} \mathbb{E}_{|C}$, on a $\mathscr{E}_{\phi,P} \simeq r\mathcal{O}_{n,P} \oplus (x,z)$. On peut donc supposer que $\mathscr{E}_{\phi,P} = (x,z)$. Il faut montrer que $\mathscr{E}_\phi[1]_P = (z)$. On a $(x,z)_{|C} = (x)/(xz) \oplus (z)/(xz,z^2)$. Le premier facteur est isomorphe à $\mathcal{O}_{C,P}$ et le second à $\mathbb{C}$. Donc $\mathscr{E}_\phi[1]_P$ est le noyau du morphisme

$$(x,z) \to \mathcal{O}_{C,P}, \quad ax + bz \mapsto \bar{a}$$

($\bar{a}$ désignant l'image de a dans $\mathcal{O}_{C,P}$). On a donc $\mathscr{E}_\phi[1]_P = (z) = \mathbb{E}_{1,P}$. On a $\mathscr{E}_{\phi|C} = E_\phi \oplus \mathcal{O}_Z$, donc $(\mathscr{E}_{\phi|C})^{\vee\vee} = E_\phi$.

On a $\mathscr{E}_\phi^\vee[1] = ((\mathscr{E}_\phi)_1)^\vee \otimes \mathbb{L}$ d'après la proposition 4.3.1(ii). On a un diagramme commutatif avec lignes exactes

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E \otimes L^{n-1} = (\mathscr{E}_\phi)^{(1)} & \longrightarrow & \mathscr{E}_\phi & \longrightarrow & (\mathscr{E}_\phi)_1 \otimes \mathbb{L}^{-1} & \longrightarrow & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & E \otimes L^{n-1} = \mathbb{E}^{(1)} & \longrightarrow & \mathbb{E} & \longrightarrow & \mathbb{E}_{|C_{n-1}} & \longrightarrow & 0
\end{array}$$

On en déduit immédiatement la troisième égalité. On a enfin

$$((\mathscr{E}^\vee)_{|C})^{\vee\vee} = (\mathscr{E}^{(1)})^\vee = (E \otimes L^{n-1})^\vee = E^*. \qquad \square$$

Théorème 5.4.2. *Si on a* $z \le -\operatorname{rg} E \deg(L)$ *et si* E *et* E_ϕ *sont semi-stables, alors* $\mathscr{E}_\phi$ *est semi-stable. Si l'inégalité est stricte et si* E *et* E_ϕ *sont stables, il en est de même de* $\mathscr{E}_\phi$.

Démonstration. Cela se démontre aisément par récurrence sur n, en utilisant le lemme 5.4.1 et les théorèmes 5.1.2 et 5.2.1. $\square$

Bibliographie

[Bănică et Forster 1986] C. Bănică et O. Forster, "Multiplicity structures on space curves", pp. 47–64 dans *The Lefschetz centennial conference, I* (Ciudad de Mexico, 1984), édité par D. Sundararaman, Contemp. Math. **58**, Amer. Math. Soc., Providence, RI, 1986. MR 88c:32018 Zbl 0605.14026

[Bayer et Eisenbud 1995] D. Bayer et D. Eisenbud, "Ribbons and their canonical embeddings", *Trans. Amer. Math. Soc.* **347**:3 (1995), 719–756. MR 95g:14032 Zbl 0853.14016

[Bhosle 1992] U. Bhosle, "Generalised parabolic bundles and applications to torsionfree sheaves on nodal curves", *Ark. Mat.* **30**:2 (1992), 187–215. MR 95g:14022 Zbl 0773.14006

[Bhosle 1999] U. N. Bhosle, "Picard groups of the moduli spaces of vector bundles", *Math. Ann.* **314**:2 (1999), 245–263. MR 2000g:14016 Zbl 0979.14003

[Drézet 2005] J.-M. Drézet, "Déformations des extensions larges de faisceaux", *Pacific J. Math.* **220**:2 (2005), 201–297. MR 2007b:14022 Zbl 1106.14005

[Drézet 2006] J.-M. Drézet, "Faisceaux cohérents sur les courbes multiples", *Collect. Math.* **57**:2 (2006), 121–171. MR 2007b:14077 Zbl 1106.14019

[Drézet 2007] J.-M. Drézet, "Paramétrisation des courbes multiples primitives", *Adv. Geom.* **7**:4 (2007), 559–612. MR 2008j:14051 Zbl 1135.14017

[Drézet 2009] J.-M. Drézet, "Faisceaux sans torsion et faisceaux quasi localement libres sur les courbes multiples primitives", *Math. Nachr.* **282**:7 (2009), 919–952. MR 2010j:14027 Zbl 1171.14010

[Eisenbud et Green 1995] D. Eisenbud et M. Green, "Clifford indices of ribbons", *Trans. Amer. Math. Soc.* **347**:3 (1995), 757–765. MR 95g:14033 Zbl 0854.14016

[González 2006] M. González, "Smoothing of ribbons over curves", *J. Reine Angew. Math.* **591** (2006), 201–235. MR 2007d:14008 Zbl 1094.14016

[Inaba 2002] M.-A. Inaba, "On the moduli of stable sheaves on a reducible projective scheme and examples on a reducible quadric surface", *Nagoya Math. J.* **166** (2002), 135–181. MR 2003d:14014 Zbl 1056.14014

[Inaba 2004] M.-A. Inaba, "On the moduli of stable sheaves on some nonreduced projective schemes", *J. Algebraic Geom.* **13**:1 (2004), 1–27. MR 2004h:14020 Zbl 1061.14013

[Russo et Teixidor i Bigas 1999] B. Russo et M. Teixidor i Bigas, "On a conjecture of Lange", *J. Algebraic Geom.* **8**:3 (1999), 483–496. MR 2000d:14039 Zbl 0942.14013

[Seshadri 1982] C. S. Seshadri, *Fibrés vectoriels sur les courbes algébriques*, Astérisque **96**, Société Mathématique de France, Paris, 1982. MR 85b:14023 Zbl 0517.14008

[Simpson 1994] C. T. Simpson, "Moduli of representations of the fundamental group of a smooth projective variety, I", *Inst. Hautes Études Sci. Publ. Math.* **79** (1994), 47–129. MR 96e:14012 Zbl 0891.14005

[Sun 2000] X. Sun, "Degeneration of moduli spaces and generalized theta functions", *J. Algebraic Geom.* **9**:3 (2000), 459–527. MR 2001h:14040 Zbl 0971.14030

[Sun 2002] X. Sun, "Degeneration of $\mathrm{SL}(n)$-bundles on a reducible curve", pp. 229–243 dans *Algebraic geometry in East Asia* (Kyoto, 2001), édité par A. Ohbuchi et al., World Scientific, River Edge, NJ, 2002. Disponible aussi sur l'arXiv: math.AG/0112072. MR 2005a:14043 Zbl 1080.14529

[Teixidor i Bigas 1991] M. Teixidor i Bigas, "Moduli spaces of (semi)stable vector bundles on tree-like curves", *Math. Ann.* **290**:2 (1991), 341–348. MR 92c:14014 Zbl 0719.14015

[Teixidor i Bigas 1995] M. Teixidor i Bigas, "Moduli spaces of vector bundles on reducible curves", *Amer. J. Math.* **117**:1 (1995), 125–139. MR 96e:14014 Zbl 0836.14012

[Teixidor i Bigas 1998] M. Teixidor i Bigas, "Compactifications of moduli spaces of (semi)stable bundles on singular curves: two points of view", *Collect. Math.* **49**:2-3 (1998), 527–548. MR 20003:14050 Zbl 0932.14015

Received September 4, 2009.

JEAN-MARC DRÉZET
INSTITUT DE MATHÉMATIQUES DE JUSSIEU
UMR 7586 DU CNRS
175 RUE DU CHEVALERET
F-75013 PARIS
FRANCE
drezet@math.jussieu.fr
http://people.math.jussieu.fr/~drezet/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

A QUANTITATIVE ESTIMATE FOR QUASIINTEGRAL POINTS IN ORBITS

LIANG-CHUNG HSIA AND JOSEPH H. SILVERMAN

Let $\varphi(z) \in K(z)$ be a rational function of degree $d \geq 2$ defined over a number field whose second iterate $\varphi^2(z)$ is not a polynomial, and let $\alpha \in K$. The second author previously proved that the forward orbit $\mathcal{O}_\varphi(\alpha)$ contains only finitely many quasi-S-integral points. We give an explicit upper bound for the number of such points.

Introduction

Let $K/\mathbb{Q}$ be a number field, let S be a finite set of places of K, and let $1 \geq \varepsilon > 0$. An element $x \in K$ is said to be *quasi-*(S, ε)*-integral* if

$$\sum_{v \in S} \frac{[K_v : \mathbb{Q}_v]}{[K : \mathbb{Q}]} \log^+ |x|_v \geq \varepsilon h(x). \tag{1}$$

We observe that x is in the ring of S-integers of K if and only if it is quasi-$(S, 1)$-integral, in which case (1) is an equality by definition of the height.

Let $\varphi(z) \in K(z)$ be a rational function of degree $d \geq 2$, let $\alpha \in K$ be a point, and let

$$\mathcal{O}_\varphi(\alpha) = \{\alpha, \varphi(\alpha), \varphi^2(\alpha), \dots\}$$

denote the forward orbit of α under iteration of φ. Silverman [1993] proved that if $\varphi^2(z)$ is not a polynomial, then the orbit $\mathcal{O}_\varphi(\alpha)$ contains only finitely many quasi-(S, ε)-integral points. More generally, if $\#\mathcal{O}_\varphi(\alpha) = \infty$ and if β is not an exceptional point for φ, then there are only finitely many $n \geq 1$ such that

$$\frac{1}{\varphi^n(\alpha) - \beta}$$

is quasi-(S, ε)-integral. In this note we give an upper bound for the number of such n, making explicit the dependence on S, φ, α, and β. More precisely, we

Hsia is supported by NSC-97-2918-I-008-005 and NSC-96-2115-008-012-MY3. Silverman is supported by NSF DMS-0650017 and DMS-0854755.
MSC2010: primary 37P15; secondary 11B37, 11G99, 14G99.
Keywords: arithmetic dynamics, integral points.

prove that the number of elements in the set

$$\{n \geq 0 : (\varphi^n(\alpha) - \beta)^{-1} \text{ is quasi-}(S, \varepsilon)\text{-integral}\} \tag{2}$$

is smaller than

$$4^{\#S}\gamma + \log_d^+\left(\frac{h(\varphi) + \hat{h}_\varphi(\beta)}{\hat{h}_\varphi(\alpha)}\right), \tag{3}$$

where γ depends only on d, ε, and $[K : \mathbb{Q}]$. (See Section 2 for the definitions of the height $h(\varphi)$ of the map φ and the canonical height $\hat{h}_\varphi$.) Our main result, Theorem 11 in Section 5, is a strengthened version of this statement.

The specific form of the upper bound in (3) is interesting, especially the dependence on the wandering point α and the target point β. For example, if $\hat{h}_\varphi(\alpha)$ is sufficiently large (depending on β and φ), then the bound is independent of α, β, and φ. It is also interesting to ask whether it is possible, for a given φ and α, to make the set (2) arbitrarily large by varying β. We discuss this question further in Remark 14.

We briefly describe the organization of the paper. We start in Section 1 by setting notation and proving an elementary estimate for the chordal metric. Section 2 is devoted to height functions, both the canonical height associated to a rational map and various results relating heights and polynomials. In Section 3, we prove a uniform version of the inverse function theorem for rational maps of degree d. Section 4 states an estimate for the ramification of the iterate of a rational function, taken from [Silverman 1993; 2007], and a quantitative version of Roth's theorem, taken from [Silverman 1987b]. In Section 5 we combine the preliminary material to prove our main theorem. Finally, in Section 6, we use the main theorem to give an explicit upper bound for the number of S-integral points in an orbit.

Remark 1. Silverman's paper [1993] on finiteness of quasi-S-integral points in orbits has been used by Patrick Ingram and Silverman [2009] to prove a dynamical version of the classical Bang–Zsigmondy theorem on primitive divisors [Bang 1886; Zsigmondy 1892]. It has also been used by Felipe Voloch and Silverman [2009] to prove a local-global criterion for dynamics on $\mathbb{P}^1$. The quantitative results proved here should enable one to prove quantitative versions of the papers with Ingram and Voloch, but we have not included these applications in this paper in order to keep it to a manageable length.

Remark 2. Quantitative estimates similar to those in this paper have been proved for the number of integral points on elliptic curves and on certain other types of curves. See for example [Gross and Silverman 1995] and [Silverman 1987b].

1. Preliminary material and notation

We set the following notation:

K	a number field.
M_K	the set of places of K.
M_K^∞	the set of archimedean (infinite) places of K.
M_K^0	the set of nonarchimedean (finite) places of K.
$\log^+(x)$	the maximum of $\log(x)$ and 0. We write $\log_d^+$ for log base d.

For each $v \in M_K$, we let $|\cdot|_v$ denote the corresponding normalized absolute value on K whose restriction to $\mathbb{Q}$ gives the usual v-adic absolute value on $\mathbb{Q}$. That is, if $v \in M_K^\infty$, then $|x|_v$ is the usual archimedean absolute value, and if $v \in M_K^0$, then $|x|_v = |x|_p$ is the usual p-adic absolute value for a unique prime p. We also write K_v for the completion of K with respect to $|\cdot|_v$, and we let $\mathbb{C}_v$ denote the completion of an algebraic closure of K_v.

For each $v \in M_K$, we let ρ_v denote the *chordal metric* defined on $\mathbb{P}^1(\mathbb{C}_v)$, where we recall that for $[x_1, y_1], [x_2, y_2] \in \mathbb{P}^1(\mathbb{C}_v)$,

$$\rho_v([x_1, y_1], [x_2, y_2]) = \begin{cases} \dfrac{|x_1 y_2 - x_2 y_1|_v}{\sqrt{|x_1|_v^2 + |y_1|_v^2}\sqrt{|x_2|_v^2 + |y_2|_v^2}} & \text{if } v \in M_K^\infty, \\ \dfrac{|x_1 y_2 - x_2 y_1|_v}{\max\{|x_1|_v, |y_1|_v\}\max\{|x_2|_v, |y_2|_v\}} & \text{if } v \in M_K^0. \end{cases}$$

In this paper, we use the logarithmic version of the chordal metric to measure the distance between points in $\mathbb{P}^1(\mathbb{C}_v)$.

Definition. The *logarithmic chordal metric function*

$$\lambda_v : \mathbb{P}^1(\mathbb{C}_v) \times \mathbb{P}^1(\mathbb{C}_v) \to \mathbb{R} \cup \{\infty\}$$

is defined by

$$\lambda_v([x_1, y_1], [x_2, y_2]) = -\log \rho_v([x_1, y_1], [x_2, y_2]).$$

Note that $\lambda_v(P, Q) \geq 0$ for all $P, Q \in \mathbb{P}^1(\mathbb{C}_v)$, and that two points $P, Q \in \mathbb{P}^1(\mathbb{C}_v)$ are close if and only if $\lambda_v(P, Q)$ is large. We also note that λ_v is a particular choice of an *arithmetic distance function* as defined in [Silverman 1987a, Section 3], that is, it is a local height function $\lambda_{\mathbb{P}^1 \times \mathbb{P}^1, \Delta}$, where Δ is the diagonal of $\mathbb{P}^1 \times \mathbb{P}^1$.

The next lemma relates the logarithmic chordal metric $\lambda_v(x, y)$ to the usual metric $|x - y|_v$ arising from the absolute value v.

Lemma 3. *Let $v \in M_K$ and let λ_v be the logarithmic chordal metric on $\mathbb{P}^1(\mathbb{C}_v)$. Define $\ell_v = 2$ if v is archimedean, and $\ell_v = 1$ if v is nonarchimedean. Then for*

$x, y \in \mathbb{C}_v$ *the inequality* $\lambda_v(x, y) > \lambda_v(y, \infty) + \log \ell_v$ *implies*

$$\lambda_v(y, \infty) \le \lambda_v(x, y) + \log|x - y|_v \le 2\lambda_v(y, \infty) + \log \ell_v.$$

Proof. Notice that by the definition of chordal metric,

$$\lambda_v(x, y) = \lambda_v(x, \infty) + \lambda_v(y, \infty) - \log|x - y|_v.$$

Therefore

$$\lambda_v(x, y) + \log|x - y|_v = \lambda_v(x, \infty) + \lambda_v(y, \infty) \ge \lambda_v(y, \infty).$$

This gives the lower bound for the sum $\lambda_v(x, y) + \log|x - y|_v$.

For the upper bound, if v is an archimedean place, then the assertion is the same as [Silverman 2007, Lemma 3.53]. We will not repeat the proof here. For the case where v is nonarchimedean, notice that λ_v satisfies the strong triangle inequality,

$$\lambda_v(x, y) \ge \min\left(\lambda_v(x, z), \lambda_v(y, z)\right),$$

and that this inequality is an equality if $\lambda_v(x, z) \ne \lambda_v(y, z)$. Suppose that x and y satisfy the condition required in the lemma, that is, $\lambda_v(x, y) > \lambda_v(y, \infty)$. (In this case, $\ell_v = 1$.) We claim that $\lambda_v(x, \infty) \le \lambda_v(y, \infty)$. Assume to the contrary that $\lambda_v(x, \infty) > \lambda_v(y, \infty)$. Then by the strong triangle inequality for λ_v, we have

$$\lambda_v(x, y) = \min\left(\lambda_v(x, \infty), \lambda_v(y, \infty)\right) = \lambda_v(y, \infty).$$

But this contradicts the assumption that $\lambda_v(x, y) > \lambda_v(y, \infty)$. Now

$$\begin{aligned} \lambda_v(x, y) + \log|x - y|_v &= \lambda_v(x, \infty) + \lambda_v(y, \infty) \\ &\le 2\lambda_v(y, \infty) \quad \text{by the claim,} \end{aligned}$$

which completes the proof of the lemma. □

2. Dynamics and height functions

Let $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ be a rational map on $\mathbb{P}^1$ of degree $d \ge 2$ defined over the number field K. We identify $K \cup \{\infty\} \simeq \mathbb{P}^1(K)$ by fixing an affine coordinate z on $\mathbb{P}^1$, so $\alpha \in K$ equals $[\alpha, 1] \in \mathbb{P}^1(K)$, and the point at infinity is $[1, 0]$. With respect to this affine coordinate, we identity rational maps $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ with rational functions $\varphi(z) \in K(z)$.

Let $P \in \mathbb{P}^1$. Then the (*forward*) *orbit* of P under iteration of φ is the set

$$\mathbb{O}_\varphi(P) = \{\varphi^n(P) : n = 0, 1, 2, \dots\}.$$

The point P is called a *wandering point* of φ if $\mathbb{O}_\varphi(P)$ is an infinite set; otherwise P is called a *preperiodic point* of φ. The set of preperiodic points of φ is denoted by PrePer(φ). We say that a point $A \in \mathbb{P}^1$ is an *exceptional point* if it is preperiodic and $\varphi^{-1}(\mathbb{O}_\varphi(A)) = \mathbb{O}_\varphi(A)$, which is equivalent to the assumption that the complete (forward and backward) φ-orbit of A is a finite set. It is a standard fact that A is an exceptional point for φ if and only if A a totally ramified fixed point of φ^2. (One direction is clear, and the other follows from the fact [Silverman 2007, Theorem 1.6] that if A is an exceptional point, then $\mathbb{O}_\varphi(A)$ consists of at most two points.)

For a point $P = [x_0, x_1] \in \mathbb{P}^1(K)$, the *height* of P is

$$h(P) = \sum_{v \in M_K} \frac{[K_v : \mathbb{Q}_v]}{[K : \mathbb{Q}]} \log \max(|x_0|_v, |x_1|_v).$$

Then the *canonical height* of P relative to the rational map φ is given by the limit

$$\hat{h}_\varphi(P) = \lim_{n \to \infty} h(\varphi^n P)/d^n.$$

To simplify notation, we let $d_v = [K_v : \mathbb{Q}_v]/[K : \mathbb{Q}]$.

Using the definition of λ_v, we see that

$$h(P) = \sum_{v \in M_K} d_v \lambda_v(P, \infty) + O(1).$$

More precisely, writing $P = [x_0, x_1]$ and $\infty = [1, 0]$, we have

$$h(P) = \sum_{v \in M_K^0} d_v \lambda_v(P, \infty) + \sum_{v \in M_K^\infty} d_v \log\left(\frac{\max\{|x_0|_v, |x_1|_v\}}{\sqrt{|x_0|_v^2 + |x_1|_v^2}} \right).$$

The quantity $\max\{a, b\}/\sqrt{a^2 + b^2}$ is between $1/\sqrt{2}$ and 1 for all nonnegative $a, b \in \mathbb{R}$, so

$$-\tfrac{1}{2} \log 2 \le h(P) - \sum_{v \in M_K} d_v \lambda_v(P, \infty) \le 0.$$

For further material and basic properties of height functions, see [Silverman 2007, Sections 3.1–3.5].

For a polynomial $f = \sum a_i z^i \in K[z]$ and absolute value $v \in M_K$, we define

$$|f|_v = \max\{|a_i|_v\} \quad \text{and} \quad h(f) = h([\ldots, a_i, \ldots]) = \sum_{v \in M_K} d_v \log |f|_v.$$

We say that a rational function $\varphi(z) = f(z)/g(z) \in K(z)$ of degree d is written in normalized form if

$$f(z) = \sum_{i=0}^{d} a_i z^i \quad \text{and} \quad g(z) = \sum_{i=0}^{d} b_i z^i \quad \text{with } a_i, b_i \in K,$$

if a_d and b_d are not both zero, and if f and g are relatively prime in $K[z]$. For $v \in M_K$, we set $|\varphi|_v = \max\{|f|_v, |g|_v\}$, and then the height of φ is defined by

$$h(\varphi) = h([a_0, \dots, a_d, b_0, \dots, b_d]) = \sum_{v \in M_K} d_v \log|\varphi|_v.$$

Directly from the definitions, we have

$$\max(h(f), h(g)) \le h(\varphi). \tag{4}$$

The following basic properties of absolute values of polynomials will be useful.

Lemma 4. *Let $v \in M_K$ and let $f, g \in K[x]$ be polynomials with coefficients in K.*

(a) $|f + g|_v \le \begin{cases} |f|_v + |g|_v & \text{if } v \text{ is archimedean,} \\ \max\{|f|_v, |g|_v\} & \text{if } v \text{ is nonarchimedean.} \end{cases}$

(b) Gauss's lemma. *If v is nonarchimedean, then $|fg|_v = |f|_v|g|_v$.*

(c) *If v is archimedean and* $\deg f + \deg g < d$, *then*

$$4^{-d}|fg|_v \le |f|_v|g|_v \le 4^d|fg|_v.$$

Proof. (a) follows from the definition. For (b) and (c), see for example [Lang 1983, Chapter 3, Propositions 2.1 and 2.3]. □

Proposition 5. *Let $\{f_1, \dots, f_r\}$ be a collection of polynomials in the ring $K[x]$.*

(a) $$h(f_1 f_2 \cdots f_r) \le \sum_{i=1}^{r} (h(f_i) + (\deg f_i + 1)\log 2)$$
$$\le r \max_{1 \le i \le r} \{h(f_i) + (\deg f_i + 1)\log 2\}.$$

(b) $h(f_1 + f_2 + \cdots + f_r) \le \sum_{i=1}^{r} h(f_i) + \log r.$

(c) *Let $\varphi(z), \psi(z) \in K(z)$ be rational functions. Then*

$$h(\varphi \circ \psi) \le h(\varphi) + (\deg \varphi)h(\psi) + (\deg \varphi)(\deg \psi)\log 8.$$

(d) *Let $\varphi(z) \in K(z)$ be a rational function of degree $d \ge 2$. Then for all $n \ge 1$, we have*

$$h(\varphi^n) \le \left(\frac{d^n - 1}{d - 1}\right)h(\varphi) + d^2\left(\frac{d^{n-1} - 1}{d - 1}\right)\log 8.$$

Proof. The proofs of (a) and (b) can be found in [Hindry and Silverman 2000, Proposition B.7.2], where the proposition is stated for multivariable polynomials. Since we'll use the arguments in (a) for the proof of (c), we repeat the proof of (a) for the one-variable case. (Also, our situation is slightly different from that in

[Hindry and Silverman 2000], since we are using a projective height instead of an affine height.) Writing $f_i = \sum_E a_{iE} X^E$, we have

$$f_1 \cdots f_r = \sum_E \Biggl(\sum_{e_1+\cdots+e_r=E} a_{1e_1} \cdots a_{re_r} \Biggr) X^E,$$

and hence for $v \in M_K$,

$$|f_1 \cdots f_r|_v = \max_E \Biggl| \sum_{e_1+\cdots+e_r=E} a_{1e_1} \cdots a_{re_r} \Biggr|_v \tag{5}$$

and $h(f_1 \cdots f_r) = \sum_{v \in M_K} d_v \log|f_1 \cdots f_r|_v$. If v is nonarchimedean, then by Gauss's lemma, Lemma 4(b), we have

$$|f_1 \cdots f_r|_v = \prod_{i=1}^r |f_i|_v.$$

It remains to deal with an archimedean place v. We note that the number of terms in the sum appearing in the right side of (5) is $\binom{E+r-1}{E}$. Hence

$$\begin{aligned} |f_1 \cdots f_r|_v &\le \max_E \biggl(\binom{E+r-1}{E} \max_{e_1+\cdots+e_r=E} |a_{1e_1} \cdots a_{re_r}|_v \biggr) \\ &\le \max_E \Bigl(2^{E+r-1} \max_{e_1+\cdots+e_r=E} |a_{1e_1} \cdots a_{re_r}|_v \Bigr). \end{aligned}$$

Further, if $E > \deg(f_1 \ldots f_r)$, then the product $a_{1e_1} \cdots a_{re_r}$ is zero, since in that case at least one of the a_{ij} is zero. Hence

$$|f_1 \cdots f_r|_v \le 2^{\deg(f_1 \cdots f_r)+r-1} \prod_{i=1}^r |f_i|_v. \tag{6}$$

Let $N_v = 2^{\sum_i (\deg f_i + 1)}$ if v is archimedean, and $N_v = 1$ if v is nonarchimedean. Then we compute

$$\begin{aligned} h(f_1 \cdots f_r) &= \sum_{v \in M_K} d_v \log|f_1 \cdots f_r|_v \\ &\le \sum_{v \in M_K} d_v \biggl(\log N_v + \log \prod_{i=1}^r |f_i|_v \biggr) \\ &\le \sum_{i=1}^r (h(f_i) + (\deg f_i + 1) \log 2) \\ &\le r \max_{1 \le i \le r} \{h(f_i) + (\deg f_i + 1) \log 2\}, \end{aligned}$$

which completes the proof of (a).

Next we give a proof of (c). Write $\psi = \psi_0/\psi_1 \in K(z)$ in normalized form, so in particular ψ_0 and ψ_1 are relatively prime polynomials. Then

$$(\varphi \circ \psi)(z) = \frac{\sum a_i \psi_0^i \psi_1^{d-i}}{\sum b_i \psi_0^i \psi_1^{d-i}},$$

so by definition of the height of a rational function, we have

$$h(\varphi \circ \psi) \le \sum_{v \in M_K} d_v \log \max\Big\{ \Big|\sum a_i \psi_0^i \psi_1^{d-i}\Big|_v, \Big|\sum b_i \psi_0^i \psi_1^{d-i}\Big|_v \Big\}.$$

For the right side of this inequality, if v is nonarchimedean, then by Gauss's lemma again we have

$$\Big|\sum a_i \psi_0^i \psi_1^{d-i}\Big|_v \le \max\big(|f|_v |\psi_0|_v^i |\psi_1|_v^{d-i}\big) \le |\varphi|_v |\psi|_v^d.$$

Similarly,

$$\Big|\sum b_i \psi_0^i \psi^{d-i}\Big|_v \le |\varphi|_v |\psi|_v^d.$$

Hence for v nonarchimedean, $|\varphi \circ \psi|_v \le |\varphi|_v |\psi|_v^d$.

Next let v be an archimedean place of K. Then the triangle inequality gives

$$\Big|\sum a_i \psi_0^i \psi_1^{d-i}\Big|_v \le (d+1)|f|_v \max_i\big\{|\psi_0^i \psi_1^{d-i}|_v\big\}.$$

Applying the estimate (6) to the product $\psi_0^i \psi_1^{d-i}$ yields

$$|\psi_0^i \psi_1^{d-i}|_v \le 2^{d(\deg\psi+1)} |\psi_0|_v^i |\psi_1|_v^{d-i} \le 2^{d(\deg\psi+1)} |\psi|_v^d.$$

Therefore,

$$\Big|\sum a_i \psi_0^i \psi_1^{d-1}\Big|_v \le (d+1) 2^{d(\deg\psi+1)} |f|_v |\psi|_v^d \le (d+1) 2^{d(\deg\psi+1)} |\varphi|_v |\psi|_v^d.$$

Similarly,

$$\Big|\sum b_i \psi_0^i \psi_1^{d-1}\Big|_v \le (d+1) 2^{d(\deg\psi+1)} |\varphi|_v |\psi|_v^d.$$

We combine these estimates. To ease notation, we let $N_v = 1$ for v nonarchimedean and $N_v = (d+1)2^{2d\deg\psi} = (d+1)4^{\deg\varphi \deg\psi}$ for v archimedean. Then

$$\begin{aligned} h(\varphi \circ \psi) &\le \sum_{v \in M_K} d_v \log \max\Big\{ \Big|\sum a_i \psi_0^i \psi_1^{d-1}\Big|_v, \Big|\sum b_i \psi_0^i \psi_1^{d-1}\Big|_v \Big\} \\ &\le \sum_{v \in M_K} d_v (\log|\varphi|_v + d \log|\psi|_v + \log N_v) \\ &\le h(\varphi) + d h(\psi) + (\deg\varphi)(\deg\psi) \log 4 + \log(d+1) \\ &\le h(\varphi) + d h(\psi) + (\deg\varphi)(\deg\psi) \log 8, \end{aligned}$$

since $d+1 \le 2^d \le 2^{d \deg \psi}$. This completes the proof of (c).

Finally, we prove (d) by induction on n. The stated inequality is clearly true for $n=1$. Assume now it is true for n. Then

$$\begin{aligned} h(\varphi^{n+1}) &\le h(\varphi^n) + d^n h(\varphi) + d^{n+1}\log 8 \quad \text{from (c) applied to } \varphi^n \text{ and } \varphi \\ &\le \Big(\frac{d^n-1}{d-1} h(\varphi) + d^2 \frac{d^{n-1}-1}{d-1}\log 8\Big) + d^n h(\varphi) + d^{n+1}\log 8 \\ &\qquad\qquad\qquad\qquad \text{from the induction hypothesis} \\ &= \Big(\frac{d^{n+1}-1}{d-1}\Big) h(\varphi) + d^2\Big(\frac{d^n-1}{d-1}\Big)\log 8. \qquad \square \end{aligned}$$

The following facts about height functions are well known.

Proposition 6. *Let $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ be a rational map of degree $d \ge 2$ defined over K. There are constants $c_1, c_2, c_3,$ and c_4, depending only on d, such that the following estimates hold for all $P \in \mathbb{P}^1(\bar{K})$.*

(a) $|h(\varphi(P)) - dh(P)| \le c_1 h(\varphi) + c_2$.

(b) $|\hat{h}_\varphi(P) - h(P)| \le c_3 h(\varphi) + c_4$.

(c) $\hat{h}_\varphi(\varphi(P)) = d\hat{h}_\varphi(P)$.

(d) $P \in \operatorname{PrePer}(\varphi)$ *if and only if* $\hat{h}_\varphi(P) = 0$.

Proof. See, for example, [Hindry and Silverman 2000, Sections B.2 and B.4] or [Silverman 2007, Section 3.4]. □

3. A distance estimate

Our goal in this section is a version of the inverse function theorem that gives explicit estimates for the dependence on the (local) heights of both the points and the function. It is undoubtedly possible to give a direct, albeit long and messy, proof of the desired result. We instead give a proof using universal families of maps and arithmetic distance functions. Before stating our result, we set notation for the universal family of degree d rational maps on $\mathbb{P}^1$.

We write $\operatorname{Rat}_d \subset \mathbb{P}^{2d+1}$ for the space of rational maps of degree d, where we identify a rational map $\varphi = f/g = \sum a_i z^i \big/ \sum b_i z^i$ with the point

$$[\varphi] = [f, g] = [a_0, \dots, a_d, b_0, \dots, b_d] \in \mathbb{P}^{2d+1}.$$

If $\varphi \in \operatorname{Rat}_d(\bar{\mathbb{Q}})$ is defined over $\bar{\mathbb{Q}}$, we define the height of φ as in Section 2 to be the height of the associated point in $\mathbb{P}^{2d+1}(\bar{\mathbb{Q}})$:

$$h(\varphi) = h([a_0, \dots, a_d, b_0, \dots, b_d]).$$

Over Rat_d, there is a universal family of degree d maps, which we denote by

$$\Psi : \mathbb{P}^1 \times \operatorname{Rat}_d \to \mathbb{P}^1 \times \operatorname{Rat}_d, \quad (P, \psi) \mapsto (\psi(P), \psi).$$

We note that Rat_d is the complement in $\mathbb{P}^{2d+1}$ of a hypersurface, which we denote by $\partial\mathrm{Rat}_d$. (The set $\partial\mathrm{Rat}_d$ is given by the resultant $\mathrm{Res}(f, g) = 0$, so $\partial\mathrm{Rat}_d$ is a hypersurface of degree $2d$.) Since $\mathbb{P}^1$ is complete, we have

$$\partial(\mathbb{P}^1 \times \mathrm{Rat}_d) = \mathbb{P}^1 \times \partial\mathrm{Rat}_d.$$

The map Ψ is a finite map of degree d. Let $R(\Psi)$ denote its ramification locus. Looking at the behavior of Ψ in a neighborhood of a point (P, ψ), it is easy to see that the restriction of $R(\Psi)$ to a fiber $\mathbb{P}^1_\psi = \mathbb{P}^1 \times \{\psi\}$ is the ramification divisor $R(\Psi)|_{\mathbb{P}^1_\psi} = R(\psi)$ of ψ. So the ramification indices of the universal map Ψ and a particular map ψ are related by

$$e_{(P,\psi)}(\Psi) = e_P(\psi). \tag{7}$$

Proposition 7. *Let $\psi \in K(z)$ be a nontrivial rational function, let $S \subset M_K$ be a finite set of absolute values on K, each extended in some way to $\overline{K}$, and let $A, P \in \mathbb{P}^1(K)$. Then*

$$\sum_{v\in S} \max_{A'\in\psi^{-1}(A)} e_{A'}(\psi) d_v \lambda_v(P, A') \geq \sum_{v\in S} d_v \lambda_v(\psi(P), A) + O(h(A) + h(\psi) + 1),$$

where the implied constant depends only on the degree of the map ψ.

Proof. The statement and proof of Proposition 7 use the machinery of arithmetic distance functions and local height functions on quasiprojective varieties, as described in [Silverman 1987a], to which we refer for definitions, notation, and basic properties. We begin with the distribution relation for finite maps of smooth quasiprojective varieties [Silverman 1987a, Proposition 6.2(b)]. Applying this relation to the map Ψ and points $x, y \in \mathbb{P}^1 \times \mathrm{Rat}_d$ yields

$$\delta\bigl(\Psi(x), y; v\bigr) = \sum_{y'\in\Psi^{-1}(y)} e_{y'}(\Psi)\delta(x, y'; v) + O(\lambda_{\partial(\mathbb{P}^1\times\mathrm{Rat}_d)^2}(x, y; v)), \tag{8}$$

where $\delta(\,\cdot\,, \,\cdot\,; v)$ is a v-adic arithmetic distance function on $\mathbb{P}^1 \times \mathrm{Rat}_d$, and where $\lambda_{\partial(\mathbb{P}^1\times\mathrm{Rat}_d)^2}$ is a local height function for the indicated divisor. In particular, if we take $x = (P, \psi)$ and $y = (A, \psi)$, then the arithmetic distance function δ and the chordal metric λ_v defined in Section 1 satisfy

$$\begin{aligned} \delta(\Psi(x), y; v) = \delta(\Psi(P, \psi), (A, \psi); v) &= \delta((\psi(P), \psi), (A, \psi); v) \\ &= \lambda_v(\psi(P), A). \end{aligned} \tag{9}$$

Similarly, if $y' = (A', \psi) \in \Psi^{-1}(y)$, then

$$\delta(x, y'; v) = \delta((P, \psi), (A', \psi); v) = \lambda_v(P, A').$$

Further, since $\partial(\mathbb{P}^1 \times \mathrm{Rat}_d) = \mathbb{P}^1 \times \partial\mathrm{Rat}_d$ is the pull-back of a divisor on Rat_d and

$$\partial(\mathbb{P}^1 \times \mathrm{Rat}_d)^2 = (\mathbb{P}^1 \times \partial\mathrm{Rat}_d) \times (\mathbb{P}^1 \times \mathrm{Rat}_d) + (\mathbb{P}^1 \times \mathrm{Rat}_d) \times (\mathbb{P}^1 \times \partial\mathrm{Rat}_d),$$

applying [Silverman 1987a, Proposition 5.3(a)] gives

$$\begin{aligned} \lambda_{\partial(\mathbb{P}^1\times\mathrm{Rat}_d)^2}(x, y; v) &\gg\ll \lambda_{\mathbb{P}^1\times\partial\mathrm{Rat}_d}((P, \psi); v) + \lambda_{\mathbb{P}^1\times\partial\mathrm{Rat}_d}((A, \psi); v) \\ &\gg\ll \lambda_{\partial\mathrm{Rat}_d}(\psi; v). \end{aligned} \tag{10}$$

Substituting (7), (9), and (10) into the distribution relation (8) yields

$$\lambda_v(\psi(P), A) = \sum_{A'\in\psi^{-1}(A)} e_{A'}(\psi)\lambda_v(P, A') + O(\lambda_{\partial\mathrm{Rat}_d}(\psi; v)). \tag{11}$$

To ease notation, let $A'_v \in \psi^{-1}(A)$ be a point satisfying

$$e_{A'_v}(\psi)\lambda_v(P, A'_v) = \max_{A'\in\psi^{-1}(A)} e_{A'}\lambda_v(P, A').$$

Then for any $A' \in \psi^{-1}(A)$ we have

(12)

$$\begin{aligned} e_{A'}(\psi)\lambda_v(P, A') &= \min\{e_{A'_v}(\psi)\lambda_v(P, A'_v), e_{A'}(\psi)\lambda_v(P, A')\} && \text{from the choice of } A'_v \\ &\le d \min\{\lambda_v(P, A'_v), \lambda_v(P, A')\} && \text{since } \psi \text{ has degree } d \\ &\le d\lambda_v(A'_v, A') + O(1) && \text{from the triangle inequality.} \end{aligned}$$

This is a nontrivial estimate for $A' \ne A'_v$, so in (11) we pull off the A'_v term and use (12) for the other terms to obtain

$$\lambda_v(\psi(P), A) \le e_{A'_v}(\psi)\lambda_v(P, A'_v) + d \sum_{\substack{A'\in\psi^{-1}(A) \\ A'\ne A'_v}} \lambda_v(A'_v, A') + O(\lambda_{\partial\mathrm{Rat}_d}(\psi; v)). \tag{13}$$

The next lemma gives an upper bound for $\lambda_v(A'_v, A')$.

Lemma 8. *There is a constant $C = C(d)$ such that the following holds. Let $\psi \in \mathrm{Rat}_d(\overline{\mathbb{Q}})$, let $A \in \mathbb{P}^1(\overline{\mathbb{Q}})$, and let $A', A'' \in \psi^{-1}(A)$ be* distinct *points. Then*

$$\sum_{v\in M_K} d_v\lambda_v(A', A'') \le C(h(A) + h(\psi) + 1).$$

Proof. In the notation of [Silverman 1987a], we have

$$\begin{aligned} \lambda_v(A', A'') &= \delta_{\mathbb{P}^1\times\mathrm{Rat}_d}((A', \psi), (A'', \psi); v) \\ &= \lambda_{(\mathbb{P}^1\times\mathrm{Rat}_d)^2,\Delta}((A', \psi), (A'', \psi); v), \end{aligned}$$

where Δ is the diagonal of $(\mathbb{P}^1 \times \mathrm{Rat}_d)^2$. Summing over v gives height functions

$$\sum_{v\in M_K} \lambda_v(A', A'') = h_{(\mathbb{P}^1\times\mathrm{Rat}_d)^2,\Delta}((A', \psi), (A'', \psi)) + O\big(h_{\partial(\mathbb{P}^1\times\mathrm{Rat}_d)^2}((A', \psi), (A'', \psi))\big) + 1.$$

Choosing an ample divisor H on $\mathbb{P}^1 \times \mathrm{Rat}_d$, we use the fact that heights with respect to a subscheme are dominated by ample heights away from the support of the subscheme [Silverman 1987a, Proposition 4.2]. (This is where we use the assumption that $A' \neq A''$, which ensures that the point $((A', \psi), (A'', \psi))$ is not on the diagonal.) This yields

$$\begin{aligned}\sum_{v\in M_K} \lambda_v(A', A'') &\ll h_{\mathbb{P}^1\times\mathrm{Rat}_d,H}(A', \psi) + h_{\mathbb{P}^1\times\mathrm{Rat}_d,H}(A'', \psi) + 1\\ &\ll h(A') + h(A'') + h(\psi) + 1.\end{aligned} \tag{14}$$

We now use [Silverman 2009, Theorem 2], which says that there are positive constants C_1, C_2, C_3, depending only on the degree of ψ, such that

$$h(\psi(P)) \geq C_1 h(P) - C_2 h(\psi) - C_3. \tag{15}$$

(The paper [Silverman 2009] deals with general rational maps $\mathbb{P}^n \dashrightarrow \mathbb{P}^n$. In our case with $n = 1$, it would be a tedious, but not difficult, calculation to give explicit values for the C_i, including of course $C_1 = \deg\psi$.) Applying (15) with $P = A'$ and $P = A''$, we substitute into (14) to obtain

$$\sum_{v\in M_K} \lambda_v(A', A'') \ll h(A) + h(\psi) + 1,$$

which completes the proof of Lemma 8. □

We use Lemma 8 to bound the sum in the right side of the inequality (13). We note that $\lambda_v(A', A'') \geq 0$ for all points, so the lemma implies in particular that $\sum_{v\in S} d_v\lambda_v(A', A'') \ll h(A) + h(\psi) + 1$ for any set of places S. Further, the sum in (13) has at most $d - 1$ terms. Hence we obtain

$$\sum_{v\in S} d_v\lambda_v(\psi(P), A) \leq \sum_{v\in S} e_{A'_v}(\psi) d_v\lambda_v(P, A'_v) + O(h(A) + h(\psi) + 1).$$

In this inequality, the $O(h(\psi))$ term comes from two places, Lemma 8 and

$$\sum_{v\in S} d_v\lambda_{\partial\mathrm{Rat}_d}(\psi; v) \leq \sum_{v\in M_K} d_v\lambda_{\partial\mathrm{Rat}_d}(\psi; v) = h_{\partial\mathrm{Rat}_d}(\psi) = O(h(\psi) + 1),$$

where the last equality comes from the fact that $\partial\mathrm{Rat}_d$ is a hypersurface of degree $2d$ in $\mathbb{P}^{2d+1}$. This completes the proof of Proposition 7. □

4. A ramification estimate and a quantitative version of Roth's theorem

In this section we state two known results that will be needed to prove our main theorem. The first says that away from exceptional points, the ramification of φ^m tends to spread out as m increases.

Lemma 9. *Fix an integer $d \geq 2$. There exist constants κ_1 and $\kappa_2 < 1$, depending only on d, such that for all degree d rational maps $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$, all points $Q \in \mathbb{P}^1$ that are not exceptional for φ, all integers $m \geq 1$, and all $P \in \varphi^{-m}(Q)$, we have*

$$e_P(\varphi^m) \leq \kappa_1(\kappa_2 d)^m.$$

Proof. This is [Silverman 2007, Lemma 3.52]; see in particular the last paragraph of the proof. It is not difficult to give explicit values for the constants. If Q is not preperiodic, then the stronger estimate $e_P(\varphi^m) \leq e^{2d-2}$ is true for all m. □

The second result is the following quantitative version of Roth's theorem.

Theorem 10. *Let S be a finite subset of M_K that contains all infinite places. We assume that each place in S is extended to $\overline{K}$ in some fashion. Set the following notation.*

s	*the cardinality of S.*
Υ	*a finite, $G_{\overline{K}/K}$-invariant subset of $\overline{K}$.*
β	*a map $S \to \Upsilon$.*
$\mu > 2$	*a constant.*
$M \geq 0$	*a constant.*

There are constants r_1 and r_2, depending only on $[K : \mathbb{Q}]$, $\#\Upsilon$, and μ, such that there are at most $4^s r_1$ elements $x \in K$ satisfying both of the following conditions:

$$\sum_{v \in S} d_v \log^+ |x - \beta_v|_v^{-1} \geq \mu h(x) - M. \tag{16}$$

$$h(x) \geq r_2 \max_{v \in S}\{h(\beta_v), M, 1\}. \tag{17}$$

Proof. This is [Silverman 1987b, Theorem 2.1], with a small change of notation. For explicit values of the constants, see [Gross 1990]. □

5. A bound for the number of quasiintegral points in an orbit

In this section we prove our main result, which is an explicit upper bound for the number of iterates $\varphi^n(P)$ that are close to a given base point A in any one of a fixed finite number of v-adic topologies. Here is the precise statement.

Theorem 11. *Let $\varphi \in K(z)$ be a rational map of degree $d \geq 2$. Fix a point $A \in \mathbb{P}^1(K)$ that is not an exceptional point for φ, and let $P \in \mathbb{P}^1(K)$ be a wandering* *point for φ. For any finite set of places $S \subset M_K$ and any constant $1 \ge \varepsilon > 0$, define a set of nonnegative integers by*

$$\Gamma_{\varphi,S}(A, P, \varepsilon) = \bigl\{n \ge 0 : \textstyle\sum_{v\in S} d_v \lambda_v(\varphi^n P, A) \ge \varepsilon \hat{h}_\varphi(\varphi^n P)\bigr\}.$$

(a) *There exist constants*

$$\gamma_1 = \gamma_1(d, \varepsilon, [K : \mathbb{Q}]) \quad \textit{and} \quad \gamma_2 = \gamma_2(d, \varepsilon, [K : \mathbb{Q}])$$

such that

$$\#\left\{n \in \Gamma_{\varphi,S}(A, P, \varepsilon) : n > \gamma_1 + \log_d^+\left(\frac{h(\varphi) + \hat{h}_\varphi(A)}{\hat{h}_\varphi(P)}\right)\right\} \le 4^{\#S}\gamma_2. \tag{18}$$

(b) *In particular, there is a constant $\gamma_3 = \gamma_3(d, \varepsilon, [K : \mathbb{Q}])$ such that*

$$\#\Gamma_{\varphi,S}(A, P, \varepsilon) \le 4^{\#S}\gamma_3 + \log_d^+\left(\frac{h(\varphi) + \hat{h}_\varphi(A)}{\hat{h}_\varphi(P)}\right). \tag{19}$$

(c) *There is a constant $\gamma_4 = \gamma_4(K, S, \varphi, A, \epsilon)$ that is independent of P such that*

$$\max \Gamma_{\varphi,S}(A, P, \varepsilon) \le \gamma_4.$$

Before giving the proof of Theorem 11, we make a number of remarks.

Remark 12. Note that as a consequence of Proposition 6(d), we have $\hat{h}_\varphi(P) > 0$ if P is wandering point for φ. Hence the right side of (19) is well-defined.

Remark 13. If we take $\varepsilon = 1$, then the set $\Gamma_{\varphi,S}(A, P, \varepsilon)$ more or less coincides with the set of points in the orbit $\mathcal{O}_\varphi(P)$ that are S-integral with respect to A. We say "more or less" because $\Gamma_{\varphi,S}(A, P, \varepsilon)$ is defined using the canonical height of $\varphi^n(P)$, rather than the naive height. But using the inequality $|\hat{h}_\varphi(P) - h(P)| \ll h(\varphi) + 1$ from Proposition 6 and adjusting the constants, it is not hard to see that the estimate (19) remains true for the set

$$\Gamma^{\text{naive}}_{\varphi,S}(A, P, \varepsilon) = \bigl\{n \ge 0 : \textstyle\sum_{v\in S} d_v \lambda_v(\varphi^n P, A) \ge \varepsilon h(\varphi^n P)\bigr\}.$$

(See the proof of Corollary 17.) For example, taking $A = \infty$, the set $\Gamma^{\text{naive}}_{\varphi,S}(A, P, \varepsilon)$ consists of the points $\varphi^n(P)$ such that $z(\varphi^n(P))$ is (S, ε_0)-integral for some ε_0. This is the motivation for saying that the points in $\Gamma_{\varphi,S}(A, P, \varepsilon)$ are quasi-(S, ε)-integral with respect to A, where ε measures the degree of S-integrality.

Remark 14. The dependence of the bounds (18) and (19) on $h(\varphi)$, $\hat{h}_\varphi(A)$, and $\hat{h}_v(P)$ are quite interesting. A dynamical analogue of a conjecture of Lang asserts that the ratio $h(\varphi)/\hat{h}_\varphi(P)$ is bounded, independently of φ and P, provided that φ is suitably minimal with respect to $\mathrm{PGL}_2(K)$-conjugation. See [Silverman 2007, Conjecture 4.98].

On the other hand, there cannot be a uniform bound for the ratio $\hat{h}_\varphi(A)/\hat{h}_\varphi(P)$, since A and P may be chosen arbitrarily and independent of one another. This raises the interesting question of whether the bound for $\#\Gamma_{\varphi,S}(A, P, \varepsilon)$ actually needs to depend on A. Even in very simple situations, it appears difficult to answer this question. For example, consider the map $\varphi(z) = z^2$, the initial point $P = 2$, and the set of primes $S = \{\infty, 3, 5\}$. As $A \in \mathbb{Q}^*$ varies, is it possible for the orbit $\mathbb{O}_\varphi(P)$ to contain more and more points that are S-integral with respect to A? Writing $A = x/y$, we are asking if

$$\sup_{x,y\in\mathbb{Z}} \#\bigl\{(n, i, j) \in \mathbb{N}^3 : y \cdot 2^{2^n} - x = 3^i 5^j\bigr\} = \infty.$$

Remark 15. We observe that $\#\Gamma_{\varphi,S}(A, P, \varepsilon)$ can grow as fast as $\log(\varepsilon^{-1})$ as $\varepsilon \to 0^+$. For example, consider the map $\varphi(z) = z^d + z^{d-1}$, the points $A = 0$ and $P = p$, and the set of primes $S = \{p\}$. Since $\varphi^n(z) = z^{(d-1)^n} +$ higher order terms, we have $|\varphi^n(p)|_p = p^{-(d-1)^n}$, so

$$\lambda_p(\varphi^n P, A) = \lambda_p(\varphi^n(p), 0) = -\log|\varphi^n(p)|_p = (d-1)^n \log p.$$

Thus $\Gamma_{\varphi,S}(A, P, \varepsilon)$ consists of all $n \ge 0$ satisfying

$$(d-1)^n \log p \ge \varepsilon \hat{h}_\varphi(\varphi^n P) = \varepsilon d^n \hat{h}_\varphi(P).$$

Hence

$$\begin{aligned}\#\Gamma_{\varphi,S}(A, P, \varepsilon) &= \left\lfloor \log\left(\frac{\log p}{\varepsilon \hat{h}_\varphi(P)}\right) \Big/ \log\left(\frac{d}{d-1}\right) \right\rfloor + 1 \\ &= \frac{\log(\varepsilon^{-1})}{\log(d/(d-1))} + o(\log \varepsilon^{-1}) \qquad \text{as } \epsilon \to 0^+.\end{aligned}$$

In particular, if ε is small and d is large, so $\log(d/(d-1)) \approx 1/(d-1)$, then we have

$$\#\Gamma_{\varphi,S}(A, P, \varepsilon) \approx (d-1)\log(\epsilon^{-1}).$$

Remark 16. See [Gross and Silverman 1995; Silverman 1987b] for a version of Theorem 11 for elliptic curves. These papers deal with points on an elliptic curve E that are quasi-(S, ϵ)-integral with respect to O, the zero point of E. It is also of interest to study points that are integral with respect to some other point A, and in particular to see how the bound depends on A. The distance function on E is translation invariant up to $O\bigl(h(E)\bigr)$, so we want to estimate the size of the set

$$\bigl\{P \in E(K) : \textstyle\sum_{v\in S} d_v \lambda_v(P - A) \ge \varepsilon \hat{h}_E(P)\bigr\}. \tag{20}$$

Translating the points in (20) by A, we want to count points satisfying $\sum d_v \lambda_v(P) \ge \varepsilon \hat{h}_E(P + A) + O(h(E))$. The canonical height on an elliptic curve is a quadratic form, so $\hat{h}_E(P + A) \le 2\hat{h}_E(P) + 2\hat{h}_E(A)$. Using the results in [Silverman 1987b], this leads to a bound for the set (20) in which the dependence on A appears as the ratio $\hat{h}_E(A)/\hat{h}_E(P_{\min})$, where $P_{\min}$ is the point of smallest nonzero height in $E(K)$. This is analogous to the dependence on A in (19).

Proof of Theorem 11. For brevity, we will write $\Gamma_S(\varepsilon)$ in place of $\Gamma_{\varphi,S}(A, P, \varepsilon)$. For the given ε, we set $m \ge 1$ to be the smallest integer satisfying $\kappa_2^m \le \varepsilon/5\kappa_1$, where κ_1 and κ_2 are the positive constants appearing in Lemma 9. Since $\kappa_2 < 1$, there exists such an integer m. Notice that κ_1 and κ_2 depend only on d; consequently m depends only on d and ε. More precisely, if we assume (without loss of generality) that $\varepsilon < 1/2$, then $m \ll \log(\varepsilon^{-1})$, where the implied constant depends only on d.

Put

$$\mathbf{e}_m = \max_{A' \in \varphi^{-m}(A)} e_{A'}(\varphi^m).$$

Then Lemma 9 and our choice of m imply that

$$\mathbf{e}_m \le \kappa_1(\kappa_2 d)^m \le \varepsilon d^m/5. \tag{21}$$

Further, Proposition 7 says that for all $Q \in \mathbb{P}^1(K)$ we have

$$\mathbf{e}_m \sum_{v\in S} \max_{A' \in \varphi^{-m}(A)} d_v \lambda_v(Q, A') \ge \sum_{v\in S} d_v \lambda_v(\varphi^m Q, A) - O(h(A) + h(\varphi^m) + 1), \tag{22}$$

where the implied constant depends on $\deg(\varphi^m)$.

Suppose first that $n \le m$ for all $n \in \Gamma_S(\varepsilon)$. Then clearly $\#\Gamma_S(\varepsilon) \le m$, and from our choice of m we have

$$\#\Gamma_S(\varepsilon) \le m \le \frac{\log(5\kappa_1) + \log(\varepsilon^{-1})}{\log(\kappa_2^{-1})} + 1.$$

This upper bound has the desired form, since $\kappa_1 > 0$ and $1 > \kappa_2 > 0$ depend only on d.

We may thus assume that there exists an $n \in \Gamma_S(\varepsilon)$ such that $n > m$, and we fix such an n. By the definition of $\Gamma_S(\varepsilon)$ we have

$$\varepsilon \hat{h}_\varphi(\varphi^n P) \le \sum_{v\in S} d_v \, \lambda_v(\varphi^n P, A).$$

Applying (22) to the point $Q = \varphi^{n-m}(P)$ yields

$$\varepsilon \hat{h}_\varphi(\varphi^n P) \le \mathbf{e}_m \sum_{v\in S} d_v \max_{A' \in \varphi^{-m}(A)} \lambda_v(\varphi^{n-m} P, A') + O(h(A) + h(\varphi^m) + 1), \tag{23}$$

where the big-O constant depends on $\deg \varphi^m = d^m$, and so on d and ε.

For each $v \in S$ we choose an $A'_v \in \varphi^{-m}(A)$ satisfying

$$\lambda_v(\varphi^{n-m} P, A'_v) = \max_{A' \in \varphi^{-m} A} \lambda_v(\varphi^{n-m} P, A').$$

(For ease of exposition, we will assume that $z(A') \neq \infty$ for all $A' \in \varphi^{-m}A$. If this is not the case, then we use z for some of the A', and we use z^{-1} for the others.)

Let $S' \subset S$ be the set of places in S defined by

$$S' = \{v \in S : \lambda_v(\varphi^{n-m}(P), A'_v) > \lambda_v(A'_v, \infty) + \log \ell_v\},$$

where we recall that $\ell_v = 2$ if v is archimedean and $\ell_v = 1$ otherwise. Set $S'' = S \setminus S'$. Applying Lemma 3 to the places in S' and using the definition of S'' for the places in S'', we find that

$$\begin{aligned}
&\varepsilon \hat{h}_\varphi(\varphi^n(P)) \\
&\quad \le \Bigl(\sum_{v\in S'} + \sum_{v \in S''}\Bigr) d_v \lambda_v(\varphi^n P, A) && \text{since } n \in \Gamma_S(A, P, \varepsilon) \\
&\quad \le \mathbf{e}_m \Bigl(\sum_{v\in S'} + \sum_{v \in S''}\Bigr) d_v \lambda_v(\varphi^{n-m}(P), A'_v) + O(h(A) + h(\varphi^m) + 1) \\
&&& \text{from the definition of } A'_v \text{ and (23)} \\
&\quad \le \mathbf{e}_m \sum_{v \in S'} d_v \bigl(2\lambda_v(A'_v, \infty) - \log\bigl|z(\varphi^{n-m}(P)) - z(A'_v)\bigr| + \log \ell_v\bigr) \\
&\qquad\qquad + \mathbf{e}_m \sum_{v \in S''} d_v \bigl(\lambda_v(A'_v, \infty) + \log \ell_v\bigr) + O(h(A) + h(\varphi^m) + 1) \\
&&& \text{from Lemma 3} \\
&\quad \le \mathbf{e}_m \sum_{v \in S'} d_v \log\bigl|z(\varphi^{n-m}(P)) - z(A'_v)\bigr|^{-1} \\
&\qquad\qquad + \mathbf{e}_m \sum_{v \in S} d_v (2\lambda_v(A'_v, \infty) + \log \ell_v) + O(h(A) + h(\varphi^m) + 1).
\end{aligned}$$

We now use (b) and (c) of Proposition 6 to observe that

$$\begin{aligned}
\sum_{v \in S} d_v \lambda_v(A'_v, \infty) &\le \sum_{A' \in \varphi^{-m}(A)} \sum_{v \in S} d_v \lambda_v(A', \infty) \le \sum_{A' \in \varphi^{-m}(A)} h(A') \\
&\le \sum_{A' \in \varphi^{-m}(A)} \bigl(\hat{h}_\varphi(A') + O(h(\varphi) + 1)\bigr) \\
&\le \hat{h}_\varphi(A) + O(h(\varphi) + 1),
\end{aligned}$$

Here the last line follows because there are at most d^m terms in the sum, and $\hat{h}_\varphi(A') = d^{-m}\hat{h}_\varphi(A)$. The constants depend only on m and d, and so on ε and d. Further, from the definition of ℓ_v, we have

$$\sum_{v \in S} d_v \log \ell_v \le \log 2.$$

We also note from Proposition 5(d) that $h(\varphi^m) \ll h(\varphi)+1$, with the implied constant depending only on d and m. Hence

$$\varepsilon \hat{h}_\varphi(\varphi^n(P)) \le \mathbf{e}_m \sum_{v\in S'} d_v \log^+ \bigl|z(\varphi^{n-m}(P)) - z(A'_v)\bigr|^{-1} + O(\hat{h}_\varphi(A)+h(\varphi)+1). \tag{24}$$

We are going to apply Roth's theorem (Theorem 10) to the set

$$\Upsilon = \{z(A') : A' \in \varphi^{-m}(A)\} \subset \overline{K},$$

the map $\beta : S' \to \Upsilon$ given by $\beta_v = A'_v$, and the points $x = \varphi^{n-m}(P)$ for $n \in \Gamma_S(\epsilon)$. We note that Υ is a $G_{\overline{K}/K}$-invariant set and that $\#\Upsilon \le d^m$. We apply Theorem 10 to the set of places S', taking $M = 0$ and $\mu = 5/2$. This gives constants r_1 and r_2, depending only on $[K : \mathbb{Q}]$, d, and ε, such that the set of $n \in \Gamma_S(\epsilon)$ with $n > m$ can be written as a union

$$\{n \in \Gamma_S(\epsilon) : n > m\} = T_1 \cup T_2 \cup T_3,$$

whose three sets are characterized as follows:

$$\begin{aligned}
&\#T_1 \le 4^{\#S'} r_1,\\
&T_2 = \bigl\{n > m : \textstyle\sum_{v\in S'} d_v \log^+ |z\bigl(\varphi^{n-m}(P)\bigr) - z(A'_v)|^{-1} \le \frac{5}{2} h(\varphi^{n-m}(P))\bigr\},\\
&T_3 = \bigl\{n > m : h(\varphi^{n-m}(P)) \le r_2 \max_{v\in S'}\{h(A'_v), 1\}\bigr\}.
\end{aligned}$$

We already have a bound for the size of T_1, so we look at T_2 and T_3. We start with T_3 and use (b) and (c) of Proposition 6 to estimate

$$\begin{aligned}
h(A'_v) &\le \hat{h}_\varphi(A') + c_3 h(\varphi) + c_4\\
&= d^{-m}\hat{h}_\varphi(A) + c_3 h(\varphi) + c_4,\\
h(\varphi^{n-m}(P)) &\ge \hat{h}_\varphi(\varphi^{n-m}(P)) - c_3 h(\varphi) - c_4\\
&= d^{n-m}\hat{h}_\varphi(P) - c_3 h(\varphi) - c_4.
\end{aligned}$$

Hence

$$T_3 \subset \bigl\{n > m : d^{n-m}\hat{h}_\varphi(P) \le c_5\hat{h}_\varphi(A) + c_6 h(\varphi) + c_7\bigr\},$$

so every $n \in T_3$ satisfies

$$\begin{aligned}
n &\le m + \log_d^+\left(\frac{c_5\hat{h}_\varphi(A) + c_6 h(\varphi) + c_7}{\hat{h}_\varphi(P)}\right)\\
&\le c_8 + \log_d^+\left(\frac{\hat{h}_\varphi(A) + h(\varphi)}{\hat{h}_\varphi(P)}\right). \qquad (25)
\end{aligned}$$

Finally, we consider the set T_2. Again using (b) and (c) of Proposition 6 to relate $h(\varphi^{n-m}(P))$ to $d^{n-m}\hat{h}_\varphi(P)$, we find that every $n \in T_2$ satisfies

$$\sum_{v\in S'} d_v \log^+\left|z(\varphi^{n-m}(P)) - z(A'_v)\right|^{-1} \le \tfrac{5}{2}d^{n-m}\hat{h}_\varphi(P) + c_3 h(\varphi) + c_4.$$

We substitute this estimate into (24) to obtain

$$\varepsilon\hat{h}_\varphi(\varphi^n(P)) \le \mathbf{e}_m \tfrac{5}{2}d^{n-m}\hat{h}_\varphi(P) + c_9(\hat{h}_\varphi(A) + h(\varphi) + 1).$$

We know from (21) that $\mathbf{e}_m \le \varepsilon d^m/5$, and also $\hat{h}_\varphi(\varphi^n(P)) = d^n\hat{h}_\varphi(P)$, which yields

$$\varepsilon d^n\hat{h}_\varphi(P) \le \left(\frac{\varepsilon}{5}d^m\right)\frac{5}{2}d^{n-m}\hat{h}_\varphi(P) + c_9(\hat{h}_\varphi(A) + h(\varphi) + 1).$$

A little bit of algebra gives the inequality

$$\begin{aligned} n &\le \log_d\left(2c_9\frac{\hat{h}_\varphi(A) + h(\varphi) + 1}{\varepsilon\hat{h}_\varphi(P)}\right) \\ &\le c_{10} + \log_d^+\left(\frac{\hat{h}_\varphi(A) + h(\varphi)}{\hat{h}_\varphi(P)}\right). \end{aligned} \tag{26}$$

Combining the estimate for $\#T_1$ with the bounds (25) and (26) for the largest elements in T_2 and T_3 completes the proof of (a).

We note that (b) follows immediately from (a).

Finally, we prove (c). Our first observation is that the set $\Upsilon = z(\varphi^{-m}(A))$ used in the application of Roth's theorem does not depend on the point P. So the largest element in the finite set T_1 is bounded independently of P. (Of course, since Roth's theorem is not effective, we do not have an explicit bound for $\max\Upsilon$ in terms K, S, ε, φ and A, but that is not relevant.)

Our second observation is to note that the quantity

$$\hat{h}^{\min}_{\varphi,K} \stackrel{\text{def}}{=} \inf\left\{\hat{h}_\varphi(P) : P \in \mathbb{P}^1(K) \text{ wandering for } \varphi\right\}$$

is strictly positive. To see this, let $P_0 \in \mathbb{P}^1(K)$ be any φ-wandering point. Then

$$\hat{h}^{\min}_{\varphi,K} = \inf\left\{\hat{h}_\varphi(P) : P \in \mathbb{P}^1(K) \text{ and } 0 < \hat{h}_\varphi(P) \le \hat{h}_\varphi(P_0)\right\}.$$

This last set is finite, so the infimum is over a finite set of positive numbers, and hence is strictly positive. Therefore in the upper bounds (25) and (26) for $\max T_2$ and $\max T_3$, we may replace $\hat{h}_\varphi(P)$ with $\hat{h}^{\min}_{\varphi,K}$ to get upper bounds independent of P. This proves that $\max(T_1 \cup T_2 \cup T_3)$ may be bounded independently of P, which completes the proof of (c). □

6. A bound for the number of integral points in an orbit

In this section, we use Theorem 11 to give a uniform upper bound for the number of S-integral points in an orbit.

Corollary 17. *Let K be a number field, let $S \subset M_K$ be a finite set of places that includes all archimedean places, let R_S be the ring of S-integers of K, and let $d \ge 2$. There is a constant $\gamma = \gamma(d, [K : \mathbb{Q}])$ such that for all rational maps $\varphi \in K(z)$ of degree d satisfying $\varphi^2(z) \notin K[z]$ and all φ-wandering points $P \in \mathbb{P}^1(K)$, the number of S-integral points in the orbit of P is bounded by*

$$\#\bigl\{n \ge 1 : z(\varphi^n(P)) \in R_S\bigr\} \le 4^{\#S}\gamma + \log_d^+\left(\frac{h(\varphi)}{\hat{h}_\varphi(P)}\right).$$

Proof. By definition, an element $\alpha \in K$ is in R_S if and only if $|\alpha|_v \le 1$ for all $v \notin S$, or equivalently, if and only if

$$h(\alpha) = \sum_{v \in S} d_v \log \max\{|\alpha|_v, 1\}.$$

We note that for $v \in M_K^0$ we have

$$\lambda_v(\alpha, \infty) = \lambda_v([\alpha, 1], [1, 0]) = \log \max\{|\alpha|_v, 1\}.$$

The formula for λ_v when v is archimedean is slightly different, but the trivial inequality $\max\{t, 1\} \le \sqrt{t^2+1}$ shows that for $v \in M_K^\infty$ we have

$$\log \max\{|\alpha|_v, 1\} \le \lambda_v(\alpha, \infty).$$

Hence $\alpha \in R_S$ implies $h(\alpha) \le \sum_{v \in S} d_v \lambda_v(\alpha, \infty)$.

Let $n \ge 1$ satisfy $z(\varphi^n(P)) \in R_S$. Then

$$h(\varphi^n(P)) \le \sum_{v \in S} d_v \lambda_v(\varphi^n(P), \infty). \tag{27}$$

Proposition 6 tells us that

$$h(\varphi^n(P)) \ge \hat{h}_\varphi(\varphi^n(P)) - c_3 h(\varphi) - c_4 = d^n \hat{h}_\varphi(P) - c_3 h(\varphi) - c_4, \tag{28}$$

where c_3 and c_4 depend only on d. Combining (27) and (28) gives

$$\sum_{v \in S} d_v \lambda_v(\varphi^n(P), \infty) \ge d^n \hat{h}_\varphi(P) - c_3 h(\varphi) - c_4. \tag{29}$$

We consider two cases. First, if

$$d^n \hat{h}_\varphi(P) \le 2c_3 h(\varphi) + 2c_4,$$

then the number of possible values of n is at most

$$\log_d^+\left(\frac{2c_3h(\varphi)+2c_4}{\hat{h}_\varphi(P)}\right),$$

which has the desired form. Second, if

$$d^n\hat{h}_\varphi(P)\ge 2c_3h(\varphi)+2c_4,$$

then (29) implies that

$$\sum_{v\in S} d_v\lambda_v(\varphi^n(P),\infty)\ge \tfrac{1}{2}d^n\hat{h}_\varphi(P)=\tfrac{1}{2}\hat{h}_\varphi(\varphi^n(P)). \tag{30}$$

Now Theorem 11(b) with $\varepsilon = 1/2$ and $A = \infty$ tells us that the number of n satisfying (30) is at most

$$4^{\#S}\gamma_3+\log_d^+\left(\frac{h(\varphi)+\hat{h}_\varphi(\infty)}{\hat{h}_\varphi(P)}\right), \tag{31}$$

where γ_3 depends only on $[K:\mathbb{Q}]$ and d. (Note that our assumption that $\varphi^2(z)$ is not a polynomial is equivalent to the assertion that ∞ is not an exceptional point for φ. This is needed in order to apply Theorem 11.) It only remains to observe that

$$\hat{h}_\varphi(\infty)\le h(\infty)+c_3h(\varphi)+c_4 \quad\text{and}\quad h(\infty)=h([0,1])=0$$

to see that the bound (31) has the desired form. $\square$

Acknowledgement

Hsia would like to thank his coauthor and the Department of Mathematics at Brown University for their hospitality during his visit when this work was initiated, and NCTS for his visit during spring 2009. Silverman would like to thank Microsoft Research New England for inviting him to be a visiting researcher during fall 2009.

References

[Bang 1886] A. S. Bang, "Taltheoretiske Undersogelser", *Tidsskrift Mat.* **4**:5 (1886), 70–80, 130–137. JFM 19.0168.02

[Gross 1990] R. Gross, "A note on Roth's theorem", *J. Number Theory* **36**:1 (1990), 127–132. MR 91i:11076 Zbl 0722.11034

[Gross and Silverman 1995] R. Gross and J. Silverman, "*S*-integer points on elliptic curves", *Pacific J. Math.* **167**:2 (1995), 263–288. MR 96c:11057 Zbl 0824.11038

[Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry: An introduction*, Graduate Texts in Math. **201**, Springer, New York, 2000. MR 2001e:11058 Zbl 0948.11023

[Ingram and Silverman 2009] P. Ingram and J. H. Silverman, "Primitive divisors in arithmetic dynamics", *Math. Proc. Cambridge Philosophical Soc.* **146**:2 (2009), 289–302. MR 2010a:11023 Zbl 05532375

[Lang 1983] S. Lang, *Fundamentals of Diophantine geometry*, Springer, New York, 1983. MR 85j:11005 Zbl 0528.14013

[Silverman 1987a] J. H. Silverman, "Arithmetic distance functions and height functions in Diophantine geometry", *Math. Ann.* **279**:2 (1987), 193–216. MR 89a:11066 Zbl 0607.14013

[Silverman 1987b] J. H. Silverman, "A quantitative version of Siegel's theorem: Integral points on elliptic curves and Catalan curves", *J. Reine Angew. Math.* **378** (1987), 60–100. MR 89g:11047 Zbl 0608.14021

[Silverman 1993] J. H. Silverman, "Integer points, Diophantine approximation, and iteration of rational maps", *Duke Math. J.* **71**:3 (1993), 793–829. MR 95e:11070 Zbl 0811.11052

[Silverman 2007] J. H. Silverman, *The arithmetic of dynamical systems*, Graduate Texts in Math. **241**, Springer, New York, 2007. MR 2008c:11002 Zbl 1130.37001

[Silverman 2009] J. H. Silverman, "Height estimates for equidimensional dominant rational maps", preprint, 2009. To appear in *J. Ramanujan Math. Soc.* arXiv 0908.3835

[Silverman and Voloch 2009] J. H. Silverman and J. F. Voloch, "A local-global criterion for dynamics on $\mathbb{P}^1$", *Acta Arith.* **137**:3 (2009), 285–294. MR 2010g:37173 Zbl 05538712

[Zsigmondy 1892] K. Zsigmondy, "Zur Theorie der Potenzreste", *Monatsh. Math. Phys.* **3**:1 (1892), 265–284. MR 1546236 JFM 24.0176.02

Received October 22, 2009.

LIANG-CHUNG HSIA
DEPARTMENT OF MATHEMATICS
NATIONAL CENTRAL UNIVERSITY
CHUNG-LI 32054
TAIWAN
hsia@math.ncu.edu.tw

JOSEPH H. SILVERMAN
DEPARTMENT OF MATHEMATICS, BOX 1917
BROWN UNIVERSITY
151 THAYER STREET
PROVIDENCE, RI 02912
UNITED STATES
jhs@math.brown.edu
http://www.math.brown.edu/~jhs

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

MÖBIUS ISOPARAMETRIC HYPERSURFACES WITH THREE DISTINCT PRINCIPAL CURVATURES, II

ZEJUN HU AND SHUJIE ZHAI

Using the method of moving frames and the algebraic techniques of T. E. Cecil and G. R. Jensen that were developed while they classified the Dupin hypersurfaces with three principal curvatures, we extend Hu and Li's main theorem in *Pacific J. Math.* 232:2 (2007), 289–311 by giving a complete classification for all Möbius isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ with three distinct principal curvatures.

1. Introduction

Let $x : M^n \to \mathbb{S}^{n+1}$ be a connected smooth hypersurface in the $(n+1)$-dimensional unit sphere $\mathbb{S}^{n+1}$ without umbilic point. We choose a local orthonormal basis $\{e_1, \ldots, e_n\}$ with respect to the induced metric $I = dx \cdot dx$, and let $\{\theta_1, \ldots, \theta_n\}$ be the dual basis. Let $h = \sum_{i,j} h_{ij}\theta_i \otimes \theta_j$ be the second fundamental form of x, with squared length $\|h\|^2 = \sum_{i,j}(h_{ij})^2$ and mean curvature $H = (1/n)\sum_i h_{ii}$. Define $\rho^2 = n/(n-1)\cdot(\|h\|^2 - nH^2)$. Then the positive definite form $g = \rho^2 dx \cdot dx$ is Möbius invariant and is called the Möbius metric of $x : M^n \to \mathbb{S}^{n+1}$. The Möbius second fundamental form $\mathbf{B}$, another basic Möbius invariant of x, together with g determine completely a hypersurface of $\mathbb{S}^{n+1}$ up to Möbius equivalence; see Theorem 2.2 below.

An important class of hypersurfaces for Möbius differential geometry is the so-called Möbius isoparametric hypersurfaces in $\mathbb{S}^{n+1}$. According to [Li et al. 2002], a Möbius isoparametric hypersurface of $\mathbb{S}^{n+1}$ is an umbilic-free hypersurface of $\mathbb{S}^{n+1}$ such whose Möbius-invariant 1-form

$$\Phi = -\rho^{-1}\sum_i\Big(e_i(H) + \sum_j(h_{ij} - H\delta_{ij})e_j(\log\rho)\Big)\theta_i$$

vanishes and whose Möbius principal curvatures are all constant. These curvatures are the eigenvalues of the Möbius shape operator $\Psi := \rho^{-1}(\boldsymbol{S} - H\,\mathrm{id})$ with respect to g, where $\boldsymbol{S}$ denotes the shape operator of $x : M^n \to \mathbb{S}^{n+1}$. This definition

Supported by the Natural Sciences Foundation of China, grant numbers 10671181 and 11071225.
MSC2000: primary 53A30; secondary 53B25.
Keywords: Möbius isoparametric hypersurfaces, Möbius second fundamental form, Möbius metric, Möbius form, Möbius equivalence.

of Möbius isoparametric hypersurfaces is meaningful. Indeed, comparing it with that of (Euclidean) isoparametric hypersurfaces in $\mathbb{S}^{n+1}$, we see that the images of all hypersurfaces of the sphere with constant mean curvature and constant scalar curvature under the Möbius transformation satisfy $\Phi \equiv 0$, and the Möbius-invariant operator Ψ plays the role in Möbius geometry that $\boldsymbol{S}$ does in Euclidean geometry; see Theorem 2.2 below. The two conditions of a Möbius isoparametric hypersurface, namely, that it has vanishing Möbius form and has constant Möbius principal curvatures, are independent and also closely related; for detailed discussion, see [Hu and Tian 2009]. Standard examples of Möbius isoparametric hypersurfaces are the images of (Euclidean) isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ under Möbius transformations. But there are other examples which cannot be obtained by this way; for example, one occurs in our classification for hypersurfaces of $\mathbb{S}^{n+1}$ with *parallel* Möbius second fundamental form, that is, those whose Möbius second fundamental form is parallel with respect to the Levi-Civita connection of the Möbius metric g; see [Hu and Li 2004; Li et al. 2002] for details. On the other hand, it was proved in [Li et al. 2002] that any Möbius isoparametric hypersurface is in particular a Dupin hypersurface, which implies from [Thorbergsson 1983] that for a compact Möbius isoparametric hypersurface embedded in $\mathbb{S}^{n+1}$, the number γ of distinct principal curvatures can only take the values $\gamma = 2, 3, 4, 6$. A characterization of Möbius isoparametric hypersurfaces in terms of Dupin hypersurfaces was given in [Li et al. 2002] and was obtained very recently also by L. A. Rodriques and K. Tenenblat [2009]; this characterization states that a Möbius isoparametric hypersurface is either a cyclide of Dupin or a Dupin hypersurface whose Möbius curvatures are constant. Hence the problem of investigating Möbius isoparametric hypersurfaces reduces to that of investigating Dupin hypersurfaces with constant Möbius curvatures.

In [Li et al. 2002], the authors classified locally all Möbius isoparametric hypersurfaces of $\mathbb{S}^{n+1}$ with $\gamma = 2$. By relaxing the restriction that $\gamma = 2$, local classifications for all Möbius isoparametric hypersurfaces in $\mathbb{S}^4$, $\mathbb{S}^5$ and $\mathbb{S}^6$ were established in [Hu and Li 2005], [Hu et al. 2007] and [Hu and Zhai 2008], respectively. It was shown that a Möbius isoparametric hypersurface in $\mathbb{S}^4$ is either of parallel Möbius second fundamental form or Möbius equivalent to the Euclidean isoparametric hypersurface in $\mathbb{S}^4$ with three distinct principal curvatures, that is, a tube of constant radius over a standard Veronese embedding of $\mathbb{R}P^2$ into $\mathbb{S}^4$. Similarly, a hypersurface in $\mathbb{S}^5$ is Möbius isoparametric if and only if either it has parallel Möbius second fundamental form; or it is Möbius equivalent to the preimage of the stereographic projection of the cone $\tilde{x} : N^3 \times \mathbb{R}^+ \to \mathbb{R}^5$ defined by $\tilde{x}(x, t) = tx$, where $t \in \mathbb{R}^+$ and $x : N^3 \to \mathbb{S}^4 \hookrightarrow \mathbb{R}^5$ is the Cartan isoparametric immersion in $\mathbb{S}^4$ with three principal curvatures; or it is Möbius equivalent to the Euclidean isoparametric hypersurfaces in $\mathbb{S}^5$ with four distinct principal curvatures.

All these results remind us of their counterparts in Dupin hypersurfaces; see [Cecil and Jensen 1998; 2000; Cecil et al. 2007; Miyaoka and Ozawa 1989; Niebergall 1991; Pinkall 1985; Pinkall and Thorbergsson 1989].

Hence, the classification of Möbius isoparametric hypersurfaces by Möbius transformation group equivalence can be compared with that of the Dupin hypersurfaces by Lie sphere transformation group equivalence. Note that the Lie sphere transformation group contains the Möbius transformation group in $\mathbb{S}^{n+1}$ as a subgroup and the dimension difference is $n+3$. Thus, Möbius differential geometry for hypersurfaces in sphere should, in some sense, be very different from Lie sphere geometry in many respects, and therefore is worthwhile to pay more attention.

Inspired by the close similarity between Dupin hypersurfaces under the Lie sphere transformation group and Möbius isoparametric hypersurfaces under the Möbius transformation group, and by T. E. Cecil and G. R. Jensen's result [1998] that any locally irreducible Dupin hypersurface in $\mathbb{S}^n$ with three distinct principal curvatures is equivalent by Lie sphere transformation to an isoparametric hypersurface in $\mathbb{S}^n$, we started in [Hu and Li 2007] a program of classifying all Möbius isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ with three distinct Möbius principal curvatures. There, we were able to obtain the classification under the additional condition that one of the Möbius principal curvatures is of multiplicity one. The purpose of this paper is to extend that result to the general case:

Classification theorem. *Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures. Then x is Möbius equivalent to an open part of one of the following hypersurfaces in $\mathbb{S}^{n+1}$:*

(i) *The preimage of the stereographic projection of the warped product embedding*
$$\tilde{x} : \mathbb{S}^p(a) \times \mathbb{S}^q(\sqrt{1-a^2}) \times \mathbb{R}^+ \times \mathbb{R}^{n-p-q-1} \to \mathbb{R}^{n+1}$$
with $p \geq 1,\ q \geq 1,\ p+q \leq n-1$ and $0 < a < 1$, defined by
$$\tilde{x}(u', u'', t, u''') = (tu', tu'', u'''),$$
where $u' \in \mathbb{S}^p(a),\ u'' \in \mathbb{S}^q(\sqrt{1-a^2}),\ t \in \mathbb{R}^+$ and $u''' \in \mathbb{R}^{n-p-q-1}$.

(ii) *The Euclidean isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ with three distinct principal curvatures. Thus all the principal curvatures must have the same multiplicity $m \in \{1, 2, 4, 8\}$, and the isoparametric hypersurface must be a tube of constant radius over a standard Veronese embedding of a projective plane $\mathbb{F}P^2$ into $\mathbb{S}^{3m+1}$, where $\mathbb{F}$ is the division algebra $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ (the quaternions), $\mathbb{O}$ (the Cayley numbers) for $m = 1, 2, 4, 8$, respectively.*

(iii) *The minimal hypersurfaces defined by*
$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : \tilde{M}^n = N^{3m} \times \mathbb{H}^{n-3m}\Big(-\frac{n-1}{6mn}\Big) \to \mathbb{S}^{n+1},$$

with

$$\tilde{x}_1 = y_1/y_0, \quad \tilde{x}_2 = y_2/y_0, \quad y_0 \in \mathbb{R}^+, \quad y_1 \in \mathbb{R}^{3m+2}, \quad y_2 \in \mathbb{R}^{n-3m},$$

where $y_1 : N^{3m} \to \mathbb{S}^{3m+1}(\sqrt{6mn/(n-1)}) \hookrightarrow \mathbb{R}^{3m+2}$ *is Cartan's minimal isoparametric hypersurface with scalar curvature* $\tilde{R}_1 = 3(m-1)(n-1)/2n$ *and principal curvatures*

$$\sqrt{\frac{n-1}{2mn}}, \qquad 0, \qquad -\sqrt{\frac{n-1}{2mn}} \tag{1-1}$$

which have the same multiplicity m, *where* $m = 1, 2, 4$ *or* 8, *and*

$$(y_0, y_2) : \mathbb{H}^{n-3m}\left(-\frac{n-1}{6mn}\right) \hookrightarrow \mathbb{L}^{n-3m+1}$$

is the standard embedding of the hyperbolic space of sectional curvature $-(n-1)/(6mn)$ *into the* $(n-3m+1)$*-dimensional Lorentz space with*

$$-y_0^2 + y_2^2 = -\frac{6mn}{n-1}.$$

Remark 1.1. All hypersurfaces in (i) are of parallel Möbius second fundamental form and have three distinct Möbius principal curvatures with arbitrary multiplicities p, q and $n-p-q$, respectively. The hypersurfaces in (ii) and (iii) are of nonparallel Möbius second fundamental form. For hypersurfaces in (iii), the multiplicities of the three Möbius principal curvatures are m, m and $n-2m > m$.

Remark 1.2. In the cases that $n = 3$, 4 and 5, the classification theorem was proved in [Hu and Li 2005; Hu et al. 2007; Hu and Li 2007], respectively. The theorem extends the main theorem of [Hu and Li 2007], where it was assumed that the Möbius isoparametric hypersurface M^n for $n \geq 5$ has three distinct Möbius principal curvatures and one of which is simple. The extension is successfully achieved by using the wonderful techniques developed by T. E. Cecil and G. R. Jensen [1998] in their classification of Dupin hypersurfaces with three principal curvatures.

Remark 1.3. As a counterpart to the Cecil–Ryan conjecture for Dupin hypersurfaces, which states that a compact embedded Dupin hypersurface in a space form is Lie equivalent to an Euclidean isoparametric hypersurface, C. P. Wang conjectured that *any compact embedded Möbius isoparametric hypersurface in* $\mathbb{S}^{n+1}$ *is Möbius equivalent to an Euclidean isoparametric hypersurface*. Pinkall and Thorbergsson [1989] and Miyaoka and Ozawa [1989], have constructed counterexamples to the Cecil–Ryan conjecture, but we point out that the classifications of Möbius isoparametric hypersurfaces in [Hu and Li 2007; 2005; Hu et al. 2007; Hu and Zhai 2008; Li et al. 2002] and this paper strengthen Wang's conjecture.

This paper consists of six sections. In Section 2, we first review the elementary facts of Möbius geometry for hypersurfaces in $\mathbb{S}^{n+1}$, and then we recall the classification for hypersurfaces of $\mathbb{S}^{n+1}$ with parallel Möbius second fundamental form [Hu and Li 2004] and the classification for hypersurfaces of $\mathbb{S}^{n+1}$ with two distinct constant Blaschke eigenvalues [Li and Zhang 2007]. In Section 3, we treat the Möbius isoparametric hypersurfaces of $\mathbb{S}^{n+1}$ with nonparallel Möbius second fundamental form and three distinct Möbius principal curvatures. We first present several important properties of the Möbius second fundamental form, and then we divide the discussion into two cases and state the main results, Theorem 3.1 and Theorem 3.2. We prove Theorem 3.1 in Section 4. In Section 5, we prove Theorem 5.1, which gives a preliminary classification for Möbius isoparametric hypersurfaces with three distinct Möbius principal curvatures whose multiplicities are not equal. By the analysis of the Möbius invariants of the hypersurfaces that appear in Theorem 5.1 we obtain Propositions 5.3 — 5.5, from which Theorem 3.2 follows. In Section 6, we complete the proof of the classification theorem.

2. Möbius invariants for hypersurfaces in $\mathbb{S}^{n+1}$

In this section we define the Möbius invariants and recall the structure equations for hypersurfaces in the unit sphere $\mathbb{S}^{n+1}$. We refer to [Wang 1998] for more details. Let $\mathbb{L}^{n+3}$ be the Lorentz space, namely $\mathbb{R}^{n+3}$ with inner product $\langle\cdot,\cdot\rangle_1$ defined by

$$\langle x, w\rangle_1 = -x_0w_0 + x_1w_1 + \cdots + x_{n+2}w_{n+2}$$

for $x = (x_0, x_1, \dots, x_{n+2})$, $w = (w_0, w_1, \dots, w_{n+2}) \in \mathbb{R}^{n+3}$.

Let $x : M^n \to \mathbb{S}^{n+1} \hookrightarrow \mathbb{R}^{n+2}$ be an immersed hypersurface of $\mathbb{S}^{n+1}$ without umbilics. We define the Möbius position vector $Y : M^n \to \mathbb{L}^{n+3}$ of x by

$$Y = \rho(1, x) \quad \text{and} \quad \rho^2 = \frac{n}{n-1}(\|h\|^2 - nH^2) > 0. \tag{2-1}$$

Theorem 2.1 [Wang 1998]. *Two hypersurfaces $x, \tilde{x} : M^n \to \mathbb{S}^{n+1}$ are Möbius equivalent if and only if there exists T in the Lorentz group $O(n+2,1)$ such that $Y = \tilde{Y}T$ on M^n.*

It follows immediately that $g = \langle dY, dY\rangle_1 = \rho^2 dx\cdot dx$ is a Möbius invariant, which is defined as the *Möbius metric* of $x : M^n \to \mathbb{S}^{n+1}$. Let Δ be the Beltrami–Laplace operator of g. Define $N = -\triangle Y/n - \langle\Delta Y, \Delta Y\rangle_1 Y/(2n^2)$. Then one can show that

$$\langle\Delta Y, Y\rangle_1 = -n, \qquad \langle\Delta Y, dY\rangle_1 = 0, \qquad \langle\Delta Y, \Delta Y\rangle_1 = 1 + n^2R, \tag{2-2}$$

$$\langle Y, Y\rangle_1 = 0, \qquad \langle N, Y\rangle_1 = 1, \qquad \langle N, N\rangle_1 = 0, \tag{2-3}$$

where R is the normalized scalar curvature of g and is called the normalized Möbius scalar curvature of $x : M^n \to \mathbb{S}^{n+1}$.

Let $\{E_1, \dots, E_n\}$ be a local orthonormal basis for (M^n, g), and let $\{\omega_1, \dots, \omega_n\}$ be the dual basis. Write $Y_i = E_i(Y)$, then it follows from (2-1), (2-2) and (2-3) that

$$\langle Y_i, Y\rangle_1 = \langle Y_i, N\rangle_1 = 0, \quad \langle Y_i, Y_j\rangle_1 = \delta_{ij} \quad \text{for } 1 \le i, j \le n.$$

Let V be the orthogonal complement to the subspace $\mathrm{Span}\{Y, N, Y_1, \dots, Y_n\}$ in $\mathbb{L}^{n+3}$. Then along M we have the orthogonal decomposition

$$\mathbb{L}^{n+3} = \mathrm{Span}\{Y, N\} \oplus \mathrm{Span}\{Y_1, \dots, Y_n\} \oplus V.$$

V is called the Möbius normal bundle of $x : M^n \to \mathbb{S}^{n+1}$. A local unit vector basis $E = E_{n+1}$ for V can be written as

$$E = E_{n+1} := (H, Hx + e_{n+1}).$$

Then, $\{Y, N, Y_1, \dots, Y_n, E\}$ forms a moving frame along M^n in $\mathbb{L}^{n+3}$.

In the rest of this paper, we will use the range $1 \le i, j, k, l, t \le n$ of indices.

We can write the structure equations as

$$dY = \sum_i Y_i\omega_i, \quad dY_i = -\sum_j A_{ij}\omega_j Y - \omega_i N + \sum_j \omega_{ij} Y_j + \sum_i B_{ij}\omega_j E, \tag{2-4}$$

$$dN = \sum_{i,j} A_{ij}\omega_j Y_i + \sum_i C_i\omega_i E, \quad dE = -\sum_i C_i\omega_i Y - \sum_{i,j} B_{ij}\omega_j Y_i, \tag{2-5}$$

where ω_{ij} is the connection form of the Möbius metric g and is defined by the structure equations $d\omega_i = \sum_j \omega_{ij} \wedge \omega_j$ and $\omega_{ij} + \omega_{ji} = 0$. The tensors $\boldsymbol{A} = \sum_{i,j} A_{ij}\omega_i \otimes \omega_j$, $\Phi = \sum_i C_i\omega_i$ and $\boldsymbol{B} = \sum_{i,j} B_{ij}\,\omega_i \otimes \omega_j$ are called the Blaschke tensor, the Möbius form and the Möbius second fundamental form of $x : M^n \to \mathbb{S}^{n+1}$, respectively. The relations between Φ, $\boldsymbol{B}$, $\boldsymbol{A}$ and the Euclidean invariants of x are given by [Wang 1998]

$$C_i = -\rho^{-2}\big(e_i(H) + \textstyle\sum_j (h_{ij} - H\delta_{ij})e_j(\log\rho)\big),$$

$$B_{ij} = \rho^{-1}(h_{ij} - H\delta_{ij}), \tag{2-6}$$

$$A_{ij} = -\rho^{-2}\big(\mathrm{Hess}_{ij}(\log\rho) - e_i(\log\rho)e_j(\log\rho) - Hh_{ij}\big) - \tfrac{1}{2}\rho^{-2}(|\nabla\log\rho|^2 - 1 + H^2)\delta_{ij}, \tag{2-7}$$

where Hess_{ij} and ∇ are the Hessian matrix and the gradient with respect to the orthonormal basis $\{e_i\}$ of $dx \cdot dx$.

The covariant derivatives of C_i, A_{ij}, B_{ij} are defined by

$$\textstyle\sum_j C_{i,j}\omega_j = dC_i + \sum_j C_j\omega_{ji}, \tag{2-8}$$

$$\textstyle\sum_k A_{ij,k}\omega_k = dA_{ij} + \sum_k A_{ik}\omega_{kj} + \sum_k A_{kj}\omega_{ki}, \tag{2-9}$$

$$\textstyle\sum_k B_{ij,k}\omega_k = dB_{ij} + \sum_k B_{ik}\omega_{kj} + \sum_k B_{kj}\omega_{ki}. \tag{2-10}$$

The integrability conditions for the structure equations (2-4) and (2-5) are

$$A_{ij,k} - A_{ik,j} = B_{ik}C_j - B_{ij}C_k, \tag{2-11}$$

$$C_{i,j} - C_{j,i} = \textstyle\sum_k (B_{ik}A_{kj} - A_{ik}B_{kj}), \tag{2-12}$$

$$B_{ij,k} - B_{ik,j} = \delta_{ij}C_k - \delta_{ik}C_j, \tag{2-13}$$

and

$$R_{ijkl} = B_{ik}B_{jl} - B_{il}B_{jk} + \delta_{ik}A_{jl} + \delta_{jl}A_{ik} - \delta_{il}A_{jk} - \delta_{jk}A_{il}, \tag{2-14}$$

$$\textstyle\sum_i B_{ii} = 0, \quad \sum_{i,j}(B_{ij})^2 = \dfrac{n-1}{n}, \quad \operatorname{tr} \boldsymbol{A} = \sum_i A_{ii} = \dfrac{1}{2n}(1+n^2R). \tag{2-15}$$

Here R_{ijkl} denote the components of the curvature tensor of g, which are defined by the structure equations

$$d\omega_{ij} - \textstyle\sum_k \omega_{ik} \wedge \omega_{kj} = -\frac{1}{2}\sum_{k,l} R_{ijkl}\omega_k \wedge \omega_l, \quad R_{ijkl} = -R_{ijlk}. \tag{2-16}$$

The normalized Möbius scalar curvature of $x : M^n \to \mathbb{S}^{n+1}$ is

$$R = \frac{1}{n(n-1)} \textstyle\sum_{i,j} R_{ijij}.$$

The second covariant derivative of B_{ij} is defined by

$$\textstyle\sum_l B_{ij,kl}\,\omega_l = dB_{ij,k} + \sum_l B_{lj,k}\,\omega_{li} + \sum_l B_{il,k}\,\omega_{lj} + \sum_l B_{ij,l}\,\omega_{lk}. \tag{2-17}$$

From exterior differentiation of (2-10), we have the Ricci identity

$$B_{ij,kl} - B_{ij,lk} = \textstyle\sum_t B_{tj}R_{tikl} + \sum_t B_{it}R_{tjkl}. \tag{2-18}$$

From (2-6), we see that the Möbius shape operator of $x : M^n \to \mathbb{S}^{n+1}$ takes the form $\Psi = \rho^{-1}(\boldsymbol{S} - H\,\mathrm{id}) = \sum_{i,j} B_{ij}\omega_i E_j$, which implies that for an umbilic-free hypersurface in $\mathbb{S}^{n+1}$, the number of distinct Möbius principal curvatures is the same as that of its distinct Euclidean principal curvatures.

One can easily show that all coefficients in (2-4) and (2-5) are determined by $\{g, \Psi\}$. Thus:

Theorem 2.2 [Wang 1998; Akivis and Goldberg 1997]. *For $n \geq 3$, two hypersurfaces $x : M^n \to \mathbb{S}^{n+1}$ and $\tilde{x} : \tilde{M}^n \to \mathbb{S}^{n+1}$ are Möbius equivalent if and only if there exists a diffeomorphism $F : M^n \to \tilde{M}^n$ that preserves the Möbius metric and the Möbius shape operator.*

An umbilic-free hypersurface $x : M^n \to \mathbb{S}^{n+1}$ is said to have parallel Möbius second fundamental form if $B_{ij,k} = 0$ for all i, j, k. Hypersurfaces of $\mathbb{S}^{n+1}$ with parallel Möbius second fundamental form have now been completely classified. A special case of the classification can be stated as follows.

Theorem 2.3 [Hu and Li 2004]. *For $n \geq 2$, let $x : M^n \to \mathbb{S}^{n+1}$ be an immersed umbilic-free hypersurface with parallel Möbius second fundamental form and with three distinct Möbius principal curvatures. Then x is Möbius equivalent to an open part of the image of σ of the warped product embedding*

$$\tilde{x} : \mathbb{S}^p(a) \times \mathbb{S}^q(\sqrt{1-a^2}) \times \mathbb{R}^+ \times \mathbb{R}^{n-p-q-1} \to \mathbb{R}^{n+1}$$

with $p \geq 1,\ q \geq 1,\ p+q \leq n-1$ and $0 < a < 1$, defined by

$$\tilde{x}(u', u'', t, u''') = (tu', tu'', u'''),$$

for

$$u' \in \mathbb{S}^p(a), \quad u'' \in \mathbb{S}^q(\sqrt{1-a^2}), \quad t \in \mathbb{R}^+, \quad u''' \in \mathbb{R}^{n-p-q-1},$$

where the conformal diffeomorphism $\sigma : \mathbb{R}^{n+1} \to \mathbb{S}^{n+1} \setminus \{(-1, 0, \dots, 0)\}$ is the inverse of the stereographic projection and is defined by

$$\sigma(u) = \left(\frac{1-|u|^2}{1+|u|^2}, \frac{2u}{1+|u|^2} \right) \quad \text{for } u \in \mathbb{R}^{n+1}.$$

To prove our main theorem, we also need the following partial classification for umbilic-free hypersurfaces in $\mathbb{S}^{n+1}$ with two distinct Blaschke eigenvalues, due to Li and Zhang [2007]; see also [Hu and Li 2007]

Theorem 2.4. *For $n \geq 3$, let $x : M^n \to \mathbb{S}^{n+1}$ be an immersed umbilic-free hypersurface with two distinct constant Blaschke eigenvalues and vanishing Möbius form. If x has three distinct Möbius principal curvatures, then it is locally Möbius equivalent to either of the following two families of hypersurfaces in $\mathbb{S}^{n+1}$:*

(1) *Minimal hypersurfaces defined by*

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : \tilde{M}^n = N^p \times \mathbb{H}^{n-p}(-r^{-2}) \to \mathbb{S}^{n+1}$$

with $r > 0$ and

$$\tilde{x}_1 = y_1/y_0, \quad \tilde{x}_2 = y_2/y_0,$$
$$y_0 \in \mathbb{R}^+, \qquad y_1 \in \mathbb{R}^{p+2}, \quad y_2 \in \mathbb{R}^{n-p} \quad \text{for } 2 \leq p \leq n-1,$$

where $y_1 : N^p \to \mathbb{S}^{p+1}(r) \hookrightarrow \mathbb{R}^{p+2}$ is an umbilic-free minimal hypersurface immersed into the $(p+1)$-dimensional sphere of radius r and constant scalar curvature

$$\tilde{R}_1 = \frac{np(p-1)-(n-1)r^2}{nr^2},$$

and $(y_0, y_2) : \mathbb{H}^{n-p}(-r^{-2}) \to \mathbb{L}^{n-p+1}$ is the standard embedding of hyperbolic space of sectional curvature $-r^{-2}$ into the $(n-p+1)$-dimensional Lorentz space with $-y_0^2 + y_2^2 = -r^2$.

(2) *Nonminimal hypersurfaces defined by*

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : \tilde{M}^n = N^p \times \mathbb{S}^{n-p}(r) \to \mathbb{S}^{n+1}$$

with $r > 0$ *and*

$$\tilde{x}_1 = y_1/y_0, \quad \tilde{x}_2 = y_2/y_0,$$
$$y_0 \in \mathbb{R}^+, \qquad y_1 \in \mathbb{R}^{p+1}, \quad y_2 \in \mathbb{R}^{n-p+1} \quad \text{for } 2 \le p \le n-1,$$

where $(y_0, y_1) : N^p \to \mathbb{H}^{p+1}(-r^{-2}) \hookrightarrow \mathbb{L}^{p+2}$, *with* $-y_0^2 + y_1^2 = -r^2$, *is an umbilic free minimal hypersurface immersed into* $(p+1)$*-dimensional hyperbolic space of sectional curvature* $-r^{-2}$ *and constant scalar curvature*

$$\tilde{R}_1 = -\frac{np(p-1)+(n-1)r^2}{nr^2},$$

and $y_2 : \mathbb{S}^{n-p}(r) \to \mathbb{R}^{n-p+1}$ *is the standard embedding of the* $(n-p)$*-sphere of radius* r.

3. Möbius isoparametric hypersurfaces with $\gamma = 3$

Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct principal curvatures B_1, B_2, B_3 of multiplicities m_1, m_2, m_3, respectively. Without loss of generality, we assume that $m_1 \ge m_2 \ge m_3 \ge 1$.

Since x has constant Möbius principal curvatures, we can choose, around each point of M, a local frame field $\{E_i\}_{1\le i\le n}$ orthonormal with respect to the Möbius metric g such that the matrix (B_{ij}) is diagonalized. Let us write

$$(B_{ij}) = \operatorname{diag}(b_1, \dots, b_n), \tag{3-1}$$

where $\{b_i\}$ are all constants. From the assumption, we can assume without loss of generality that

$$b_1 = \cdots = b_{m_1} = B_1, \quad b_{m_1+1} = \cdots = b_{m_1+m_2} = B_2, \quad b_{m_1+m_2+1} = \cdots = b_n = B_3.$$

Here B_1, B_2 and B_3 are distinct and, by (2-15), they satisfy the conditions

$$m_1B_1 + m_2B_2 + m_3B_3 = 0, \quad m_1B_1^2 + m_2B_2^2 + m_3B_3^2 = \frac{n-1}{n}. \tag{3-2}$$

From now on, unless stated otherwise we impose the additional index conventions

$$\begin{aligned} 1 \le a, b, c, d &\le m_1, \\ m_1 + 1 \le \quad p, q \quad &\le m_1 + m_2, \\ m_1 + m_2 + 1 \le \quad \alpha, \beta \quad &\le m_1 + m_2 + m_3 = n. \end{aligned} \tag{3-3}$$

With respect to the local frame field $\{E_i\}$, we write the Blaschke tensor as $\boldsymbol{A} = \sum_{i,j} A_{ij}\, \omega_i \otimes \omega_j$. Since the Möbius form Φ vanishes, we see from (2-12) that $\boldsymbol{A}$

and $\boldsymbol{B}$ commute, which implies that $A_{pa} = A_{a\alpha} = A_{p\alpha} = 0$. Moreover, for any fixed point $\xi \in M$, we can choose the local frame field $\{E_i\}$ to guarantee that, in addition to (3-1) around ξ, we have at the pont ξ

$$(A_{ij}) = \operatorname{diag}(A_1, \dots, A_n). \tag{3-4}$$

Here $\{A_i\}_{1\le i\le n}$ are the eigenvalues of the Blaschke tensor $\boldsymbol{A}$. Obviously, we can further arrange the local frame field $\{E_i\}$ around ξ so that, in addition to (3-1) around ξ, these eigenvalues are ordered at ξ as

$$\begin{aligned} A_1(\xi) \le A_2(\xi) \le \cdots \le A_{m_1}(\xi), \\ A_{m_1+1}(\xi) \le \cdots \le A_{m_1+m_2}(\xi), \\ A_{m_1+m_2+1}(\xi) \le \cdots \le A_n(\xi). \end{aligned} \tag{3-5}$$

In this way, we see that $A_1, \dots, A_n$ are well-defined continuous functions on M. Denote by M^* the set of all such points $\xi \in M$: Around ξ there exists an orthonormal frame field $\{E_i\}$ with respect to which (3-1) and (3-4) hold. Obviously, M^* is an open subset of M. In the computation that follows, we will fix a point $\xi \in M^*$ and then take an open set $U \subset M^*$ containing ξ such that over U there exists an orthonormal frame field $\{E_i\}$ for which (3-1) and (3-4) hold.

Applying the condition to (2-11) and (2-13), we see that both $A_{ij,k}$ and $B_{ij,k}$ are totally symmetric tensors. As usual we define

$$\omega_{ij} = \sum_k \Gamma^i_{kj}\omega_k \quad \text{and} \quad \Gamma^i_{kj} = -\Gamma^j_{ki}. \tag{3-6}$$

From (2-10), (3-1) and (3-6) and that $\{b_i\}_{1\le i\le n}$ consists of constants, we get

$$B_{ij,k} = (b_i - b_j)\Gamma^i_{kj} = (b_j - b_k)\Gamma^j_{ik} = (b_k - b_i)\Gamma^k_{ji} \quad \text{for all } i, j, k. \tag{3-7}$$

Hence we see that

$$B_{ii,j} = B_{ij,i} = B_{ab,j} = B_{pq,j} = B_{\alpha\beta,j} = 0 \quad \text{for all } i, j, a, b, p, q, \alpha, \beta, \tag{3-8}$$

and the only possible nonzero elements in $\{B_{ij,k}\}$ are of the form $B_{pa,\alpha}$.

For the rest of this section, we assume that $B_{ij,k} \not\equiv 0$. We define the nonnegative smooth function f by

$$f = \tfrac{1}{6}|\nabla \boldsymbol{B}|^2 = \frac{1}{6}\sum_{i,j,k} B^2_{ij,k} = \sum_{p,a,\alpha} B^2_{pa,\alpha}.$$

Moreover, we define three arrays of vectors, an $m_2 \times m_3$ array $(\vec{v}_{p\alpha})$ of vectors in $\mathbb{R}^{m_1}$, an $m_1 \times m_3$ array $(\vec{v}_{a\alpha})$ of vectors in $\mathbb{R}^{m_2}$, and an $m_2 \times m_1$ array $(\vec{v}_{pa})$ of vectors in $\mathbb{R}^{m_3}$, by

$$\vec{v}_{p\alpha} = (B_{p\alpha,1}, B_{p\alpha,2}, \dots, B_{p\alpha,m_1}),$$
$$\vec{v}_{a\alpha} = (B_{a\alpha,m_1+1}, B_{a\alpha,m_1+2}, \dots, B_{a\alpha,m_1+m_2}),$$
$$\vec{v}_{pa} = (B_{pa,m_1+m_2+1}, B_{pa,m_1+m_2+2}, \dots, B_{pa,n}).$$

Lemma 3.1. *Let U be an open set of M^* as stated above. Then at each point of U, the arrays $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ and $(\vec{v}_{pa})$ satisfy*

$$\text{(3-9)}\qquad \begin{cases} \vec{v}_{p\alpha}\cdot\vec{v}_{p\beta} = 0 = \vec{v}_{a\alpha}\cdot\vec{v}_{a\beta} & \text{for all } p, a \text{ and any } \alpha\neq\beta,\\ \vec{v}_{p\alpha}\cdot\vec{v}_{q\alpha} = 0 = \vec{v}_{pa}\cdot\vec{v}_{qa} & \text{for all } \alpha, a \text{ and any } p\neq q,\\ \vec{v}_{a\alpha}\cdot\vec{v}_{b\alpha} = 0 = \vec{v}_{pa}\cdot\vec{v}_{pb} & \text{for all } p, \alpha \text{ and } a\neq b; \end{cases}$$

$$\text{(3-10)}\qquad \begin{cases} \vec{v}_{p\alpha}\cdot\vec{v}_{q\beta} + \vec{v}_{q\alpha}\cdot\vec{v}_{p\beta} = 0 & \text{if } \alpha\neq\beta \text{ and } p\neq q,\\ \vec{v}_{a\alpha}\cdot\vec{v}_{b\beta} + \vec{v}_{b\alpha}\cdot\vec{v}_{a\beta} = 0 & \text{if } \alpha\neq\beta \text{ and } a\neq b,\\ \vec{v}_{pa}\cdot\vec{v}_{qb} + \vec{v}_{qa}\cdot\vec{v}_{pb} = 0 & \text{if } a\neq b \text{ and } p\neq q; \end{cases}$$

$$\text{(3-11)}\qquad \begin{cases} |\vec{v}_{p\alpha}|^2 + |\vec{v}_{q\beta}|^2 = |\vec{v}_{q\alpha}|^2 + |\vec{v}_{p\beta}|^2 & \text{if } \alpha\neq\beta \text{ and } p\neq q,\\ |\vec{v}_{a\alpha}|^2 + |\vec{v}_{b\beta}|^2 = |\vec{v}_{b\alpha}|^2 + |\vec{v}_{a\beta}|^2 & \text{if } \alpha\neq\beta \text{ and } a\neq b,\\ |\vec{v}_{pa}|^2 + |\vec{v}_{qb}|^2 = |\vec{v}_{qa}|^2 + |\vec{v}_{pb}|^2 & \text{if } a\neq b \text{ and } p\neq q, \end{cases}$$

where the dot denotes the standard product in $\mathbb{R}^{m_1}$, $\mathbb{R}^{m_2}$ and $\mathbb{R}^{m_3}$, respectively.

Proof. From (2-10) and (3-8), we have

$$\text{(3-12)}\qquad \textstyle\sum_a B_{p\alpha,a}\omega_a = (B_2 - B_3)\omega_{p\alpha},$$
$$\text{(3-13)}\qquad \textstyle\sum_p B_{a\alpha,p}\omega_p = (B_1 - B_3)\omega_{a\alpha},$$
$$\text{(3-14)}\qquad \textstyle\sum_\alpha B_{pa,\alpha}\omega_\alpha = (B_2 - B_1)\omega_{pa}.$$

Differentiating (3-12) and then using (3-6) and (3-7), we get

$$\text{(3-15)}\qquad \frac{\sum_{a,q,\beta} B_{p\alpha,a}B_{q\beta,a}(B_3 - B_2)}{(B_1 - B_2)(B_1 - B_3)}\omega_q\wedge\omega_\beta + \textstyle\sum_{a,b} B_{p\alpha,a}\omega_{ab}\wedge\omega_b + \sum_a dB_{p\alpha,a}\wedge\omega_a$$
$$= (B_2 - B_3)\Bigg(\frac{\sum_{a,q,\beta} B_{p\beta,a}B_{q\alpha,a}}{(B_1 - B_2)(B_1 - B_3)}\omega_q\wedge\omega_\beta + \textstyle\sum_q \omega_{pq}\wedge\omega_{q\alpha} + \sum_\beta \omega_{p\beta}\wedge\omega_{\beta\alpha} - R_{p\alpha p\alpha}\omega_p\wedge\omega_\alpha\Bigg).$$

Comparing the coefficients of $\omega_q\wedge\omega_\beta$ on both sides of (3-15), we obtain

$$\text{(3-16)}\qquad \textstyle\sum_a B_{p\alpha,a}B_{q\beta,a} + \sum_a B_{p\beta,a}B_{q\alpha,a} = (B_1 - B_2)(B_1 - B_3)R_{p\alpha p\alpha}\delta_{pq}\delta_{\alpha\beta}.$$

Similarly, by differentiating (3-13) and (3-14), we get

$$\sum_p B_{a\alpha,p}B_{b\beta,p}+\sum_p B_{a\beta,p}B_{b\alpha,p}=(B_2-B_1)(B_2-B_3)R_{a\alpha a\alpha}\delta_{ab}\delta_{\alpha\beta}, \tag{3-17}$$

$$\sum_\alpha B_{pa,\alpha}B_{qb,\alpha}+\sum_\alpha B_{pb,\alpha}B_{qa,\alpha}=(B_3-B_2)(B_3-B_1)R_{papa}\delta_{pq}\delta_{ab}. \tag{3-18}$$

From (3-16), (3-17) and (3-18), the relations in (3-9) and (3-10) immediately follow.

Moreover, from (3-16)–(3-18) and (2-14), we get

$$2|\vec{v}_{p\alpha}|^2=(B_1-B_2)(B_1-B_3)(B_2B_3+A_p+A_\alpha), \tag{3-19}$$

$$2|\vec{v}_{a\alpha}|^2=(B_2-B_1)(B_2-B_3)(B_1B_3+A_a+A_\alpha), \tag{3-20}$$

$$2|\vec{v}_{pa}|^2=(B_3-B_2)(B_3-B_1)(B_1B_2+A_p+A_a). \tag{3-21}$$

Then the relations in (3-11) also immediately follow. □

Lemma 3.2. *If, on some open set, the array $(\vec{v}_{p\alpha})$ contains a zero vector, then all the vectors in either the whole row or in the whole column where the zero vector is located must be zero.*

Proof. For simplicity of notation, in this proof we denote the $m_2\times m_1$ array $(\vec{v}_{p\alpha})$ by $(\vec{v}_{ij})$ for $1\le i\le m_2$ and $1\le j\le m_1$, where $\vec{v}_{ij}\in\mathbb{R}^{m_3}$. By Lemma 3.1, the array has the following properties:

(P1) The vectors of any row form an orthogonal set.

(P2) The vectors of any column form an orthogonal set.

For any 2×2 minor $\begin{pmatrix}\vec{v}_{ik} & \vec{v}_{il}\\ \vec{v}_{jk} & \vec{v}_{jl}\end{pmatrix}$,

(P3) $\vec{v}_{ik}\cdot\vec{v}_{jl}+\vec{v}_{il}\cdot\vec{v}_{jk}=0$, and

(P4) $|\vec{v}_{ik}|^2+|\vec{v}_{jl}|^2=|\vec{v}_{il}|^2+|\vec{v}_{jk}|^2$.

Obviously, all these four properties will remain unchanged if either the rows or the columns of the array are permuted.

Suppose that a vector in the array is zero on an open set $U\subset M^*$. Permuting rows and columns, if necessary, we may assume that $\vec{v}_{11}=0$ on U. Then (P1), (P2) and (P3) imply that at each point of U, the remaining vectors

$$\vec{v}_{12},\dots,\vec{v}_{1m_1}\quad\text{and}\quad\vec{v}_{21},\dots,\vec{v}_{m_21}$$

in the first row and the first column form a mutually orthogonal set of m_1+m_2-2 vectors in $\mathbb{R}^{m_3}$, and at most m_3 vectors of which can be nonzero at any point. Let ξ_0 be a point where a maximal number of these vectors is nonzero. By continuity, the nonzero vectors at ξ_0 will remain nonzero in some open subset $V\subset U$ containing ξ_0. By maximality, the vectors that are zero at ξ_0 must remain zero on V.

By permuting rows and columns if necessary, we may assume that

$$\vec{v}_{11} = \cdots = \vec{v}_{1j} = 0 \quad \text{and} \quad \vec{v}_{11} = \cdots = \vec{v}_{i1} = 0$$

for some $i \in \{1, \ldots, m_2\}$ and $j \in \{1, \ldots, m_1\}$. The remaining vectors of the first column and the first row are all nonzero at each point of V, so the array has first row $(0, 0, \ldots, 0, \vec{v}_{1(j+1)}, \ldots, \vec{v}_{1m_1})$ and first column $(0, \ldots, 0, \vec{v}_{(i+1)1}, \ldots, \vec{v}_{m_2 1})$ and (P4) implies that $\vec{v}_{kl} = 0$ for $1 \le k \le i$ and $1 \le l \le j$. Hence all elements in the upper left $i \times j$ block of the array should be zero vectors on V,

If the first row of the array is zero on V, then we are done. If otherwise, we have $j < m_1$ and $\vec{v}_{1l} \neq 0$ for all $l \ge j+1$. Let us fix an arbitrary $k \in \{i+1, \ldots, m_2\}$ and $l \in \{j+1, \ldots, m_1\}$. Then property (P4) easily implies that

$$|\vec{v}_{k1}| = \cdots = |\vec{v}_{kj}| \quad \text{and} \quad |\vec{v}_{1l}| = \cdots = |\vec{v}_{il}| \neq 0. \tag{3-22}$$

Also by using (P4) with the minor $\begin{pmatrix} 0 & \vec{v}_{1l} \\ \vec{v}_{kj} & \vec{v}_{kl} \end{pmatrix}$, we get

$$|\vec{v}_{kl}|^2 = |\vec{v}_{kj}|^2 + |\vec{v}_{1l}|^2 \neq 0. \tag{3-23}$$

On the other hand, the properties (P1), (P2) and (P3) imply that

$$\vec{v}_{k1}, \ldots, \vec{v}_{kj}, \quad \vec{v}_{1l}, \ldots, \vec{v}_{il}, \quad \vec{v}_{kl} \tag{3-24}$$

form an orthogonal set of $i+j+1$ vectors in $\mathbb{R}^{m_3}$. But, the nonzero vectors in the first column and the first row together form an orthogonal set of $(m_1-j)+(m_2-i)$ nonzero vectors. Hence, $m_1+m_2-i-j \le m_3$ and thus $i+j+1 \ge m_1+m_2-m_3+1 \ge m_1+1 > m_3$, so some of the vectors in (3-24) must be zero. By (3-22) and (3-23), it must be the case that $\vec{v}_{k1} = \cdots = \vec{v}_{kj} = 0$. As this is true for $k = i+1, \ldots, m_2$, it follows that the first j columns of the array are all zero on the open set V. □

Lemma 3.3. *If* $\nabla \boldsymbol{B} \neq 0$, *then for any one of the three arrays* $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$, $(\vec{v}_{pa})$, *it cannot happen that there exists both a row and a column whose elements are all zero vectors on some open set* $U \subset M^*$.

Proof. Suppose to the contrary that we have such an array $(\vec{v}_{ij})$ for which each element of the $\bar{i}$-th row and the $\bar{j}$-th column is zero on an open set $U \subset M^*$. Then for any $k \neq \bar{i}$ and $l \neq \bar{j}$, the property (P4) gives that

$$|\vec{v}_{kl}|^2 = |\vec{v}_{\bar{i}l}|^2 + |\vec{v}_{k\bar{j}}|^2 - |\vec{v}_{\bar{i}\bar{j}}|^2 = 0.$$

Thus all elements of $(\vec{v}_{ij})$ are zero vectors on U, which contradicts $\nabla \boldsymbol{B} \neq 0$. □

Now we can divide our discussions into two cases:

Case I. $m_1 = m_2 = m_3$.

Case II. $m_1 \ge m_2 \ge m_3$ and $m_1 > m_3$.

Each case corresponds to a main result of this paper:

Theorem 3.1. *Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures of multiplicities $m_1 = m_2 = m_3$. If the Möbius second fundamental form is not parallel, then x is locally Möbius equivalent to the Euclidean isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ with three distinct principal curvatures.*

Theorem 3.2. *Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures of multiplicities m_1, m_2 and m_3 satisfying $m_1 \geq m_2 \geq m_3$ and $m_1 > m_3$. If the Möbius second fundamental form is not parallel, then $m_2 = m_3 := m$ and x is locally Möbius equivalent one of the minimal hypersurfaces as given by part* (iii) *of the classification theorem.*

The proofs of these two theorems are quite involved and will be given separately in the next two sections.

4. Möbius isoparametric hypersurfaces with $m_1 = m_2 = m_3$

This section is devoted to Case I and giving a proof of Theorem 3.1. Assume that $m_1 = m_2 = m_3 := m$ and $\nabla \boldsymbol{B} \neq 0$.

Proposition 4.1. *Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures of the same multiplicity m. If the Möbius second fundamental form $\boldsymbol{B}$ is not parallel, then every vector in each of the three $m \times m$ arrays $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ and $(\vec{v}_{pa})$ has length equal to $\sqrt{f}/m$, where $f = \sum_{p,a,\alpha} B^2_{pa,\alpha}$ is a constant function.*

To prove the proposition, we first establish two lemmas whose proofs can be given by the crucial algebraic techniques that were essentially discovered by Cecil and Jensen [1998]; we present the proofs here for the reader's convenience.

Lemma 4.1. *There is an open subset $U \subset M^*$ on which every vector is nonzero in each of the three $m \times m$ arrays $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ and $(\vec{v}_{pa})$.*

Proof. Suppose to the contrary and without loss of generality that $\vec{v}_{(m+1)1} = 0$ on some open set U. Then by Lemma 3.2, one of two cases must occur:

- $\vec{v}_{(m+1)a} = 0$ for $1 \leq a \leq m$, or
- $\vec{v}_{p1} = 0$ for $m+1 \leq p \leq 2m$.

In the first case, the first component of each vector of $(\vec{v}_{a\alpha})$ is zero. Hence $\vec{v}_{a\alpha}$ can be looked at as if it were in $\mathbb{R}^{m-1}$. By using (P1) and (P2), we see that at least one element both in each row and in each column of the array $(\vec{v}_{a\alpha})$ is zero. Then by using (P4), Lemma 3.2 and Lemma 3.3, we easily get $(\vec{v}_{a\alpha}) = 0$ on U. This contradicts that $\nabla \boldsymbol{B} \neq 0$, so this case does not occur.

In the second case, we can show as above that $(\vec{v}_{p\alpha}) = 0$, also a contradiction. Hence this case cannot occur either. □

Lemma 4.2. *Suppose that every vector in the arrays* $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ *and* $(\vec{v}_{pa})$ *is nonzero on* $U \subset M^*$. *Then, for each array, all vectors either in each row or in each column have the same length.*

Proof. Consider one of the arrays and denote its first row by $\vec{v}_1, \dots, \vec{v}_m$. By property (P1) and the assumption that none of these vectors is zero, it follows that this is an orthogonal basis of $\mathbb{R}^m$. Thus, there exist linear operators T_j of $\mathbb{R}^m$ for $j = 2, \dots, m$, such that the j-th row of the array is given by $T_j\vec{v}_1, \dots, T_j\vec{v}_m$. For each of these operators, the properties (P1)–(P4) imply also that

(O1) T_j is skew-symmetric for $j = 2, \dots, m$,

(O2) each of the vectors $\vec{v}_1, \dots, \vec{v}_m$ is an eigenvector of T_j^2 for $j = 2, \dots, m$, and

(O3) the relation $|T_j\vec{v}_i|^2 + |\vec{v}_k|^2 = |\vec{v}_i|^2 + |T_j\vec{v}_k|^2$ holds for any $j = 2, \dots, m$ and $i \neq k$, where $1 \le i, k \le m$.

In fact, from (P2) we can see that $T_j\vec{v}_i \cdot \vec{v}_i = 0$ holds for all $i = 1, \dots, m$ and $j = 2, \dots, m$. Similarly, $T_j\vec{v}_i \cdot \vec{v}_k + \vec{v}_i \cdot T_j\vec{v}_k = 0$ follows from (P3). Thus, (P2) and (P3) imply (O1). In addition, (P1) implies that $T_j\vec{v}_i \cdot T_j\vec{v}_k = 0$ whenever $i \neq k$, and thus $T_j^2\vec{v}_i \cdot \vec{v}_k = 0$ by (O1). It follows that $\vec{v}_i$ must be an eigenvector of T_j^2. Property (O3) follows immediately from (P4).

Having seen that each $\vec{v}_i$ is an eigenvector of T_j^2, the correspondent eigenvalue is easily seen to be given by

$$T_j^2\vec{v}_i = -\frac{|T_j\vec{v}_i|^2}{|\vec{v}_i|^2}\vec{v}_i. \tag{4-1}$$

This follows from the fact that $a|\vec{v}_i|^2 = a\vec{v}_i \cdot \vec{v}_i = T_j^2\vec{v}_i \cdot \vec{v}_i = -T_j\vec{v}_i \cdot T_j\vec{v}_i$ if $T_j^2\vec{v}_i = a\vec{v}_i$.

Fix any $j \in \{2, \dots, m\}$. Let $T = T_j$ and denote by $a_1, \dots, a_m$ the eigenvalues of T^2. Then property (O3) implies the relation

$$(1 + a_i)|\vec{v}_i|^2 = (1 + a_k)|\vec{v}_k|^2 \quad \text{for all } i,\ k \in \{1, \dots, m\}. \tag{4-2}$$

Consequently, if some eigenvalue a_i is equal to -1, then so are all the others, and thus $T^2 = -I$.

If none of the eigenvalues equals -1, then $a_i = a_k$ if and only if $|\vec{v}_i| = |\vec{v}_k|$.

Suppose that, for some row of the array, the vectors do not have the same length, and suppose likewise for some column. Relabeling if necessary, we may suppose that $\vec{v}_1, \dots, \vec{v}_m$ do not have the same length. Then there must be some vector $\vec{v}_i$ such that $|\vec{v}_i|$ is not equal to $|\vec{v}_k|$ for at least $m - \lfloor m/2 \rfloor$ vectors $\vec{v}_k$, where $\lfloor z \rfloor$

denotes the greatest integer less than or equal to z. Permute the columns so that

$$|\vec{v}_1| \neq |\vec{v}_k| \quad \text{for } \lfloor m/2 \rfloor + 1 \leq k \leq m. \tag{4-3}$$

From (3-19), (3-20) and (3-21), we have

$$\begin{aligned}
|\vec{v}_{p\alpha}|^2 - |\vec{v}_{p\beta}|^2 &= \tfrac{1}{2}(B_1 - B_2)(B_1 - B_3)(A_\alpha - A_\beta),\\
|\vec{v}_{p\alpha}|^2 - |\vec{v}_{q\alpha}|^2 &= \tfrac{1}{2}(B_1 - B_2)(B_1 - B_3)(A_p - A_q),\\
|\vec{v}_{a\alpha}|^2 - |\vec{v}_{a\beta}|^2 &= \tfrac{1}{2}(B_2 - B_1)(B_2 - B_3)(A_\alpha - A_\beta),\\
|\vec{v}_{a\alpha}|^2 - |\vec{v}_{b\alpha}|^2 &= \tfrac{1}{2}(B_2 - B_1)(B_2 - B_3)(A_a - A_b),\\
|\vec{v}_{pa}|^2 - |\vec{v}_{pb}|^2 &= \tfrac{1}{2}(B_3 - B_2)(B_3 - B_1)(A_a - A_b),\\
|\vec{v}_{pa}|^2 - |\vec{v}_{qa}|^2 &= \tfrac{1}{2}(B_3 - B_2)(B_3 - B_1)(A_p - A_q).
\end{aligned} \tag{4-4}$$

Consequently, if $(\vec{v}_{ij})$ denotes any one of the arrays, then there exist numbers $c_{ij} = -c_{ji}$ and $d_{ij} = -d_{ji}$ such that

$$|\vec{v}_{ij}|^2 - |\vec{v}_{ik}|^2 = c_{jk} \quad \text{for all } i, \tag{4-5}$$

$$|\vec{v}_{ik}|^2 - |\vec{v}_{jk}|^2 = d_{ij} \quad \text{for all } k. \tag{4-6}$$

Now (4-5) implies that (4-3) must hold for every row in our array. Thus (4-3) continues to hold after permuting the rows. We may thus assume that for some i,

$$|\vec{v}_i| \neq |T_j \vec{v}_i| \quad \text{for } \lfloor m/2 \rfloor + 1 \leq j \leq m. \tag{4-7}$$

Then (4-6) implies that (4-7) holds for every column of the array, and in particular for the first column.

In summary, we can conclude that

$$|\vec{v}_1| \neq |\vec{v}_j| \quad \text{and} \quad |\vec{v}_1| \neq |T_j \vec{v}_1| \quad \text{for } \lfloor m/2 \rfloor + 1 \leq j \leq m.$$

Now we fix $j,\ k \in \{\lfloor m/2 \rfloor + 1, \dots, m\}$. Then we claim that $\vec{v}_1$ and $\vec{v}_j$ must be in different eigenspaces of T_k^2. In fact, by (4-1) and (4-4), we see that none of the eigenvalues of T_k^2 is -1. But then by (4-2) and the first part of (4-4), the eigenvalues of T_k^2 associated to the eigenvectors $\vec{v}_1$ and $\vec{v}_j$ must be different.

On the other hand, $\vec{v}_1$ and $T_k \vec{v}_1$ are in the same eigenspace of T_k^2. In fact, if $T_k^2 \vec{v}_1 = a\vec{v}_1$, then $T_k^2 T_k \vec{v}_1 = T_k T_k^2 \vec{v}_1 = a T_k \vec{v}_1$. Thus, $\vec{v}_j$ and $T_k \vec{v}_1$ are in different eigenspaces of T_k^2. Since T_k^2 is symmetric, we have

$$\vec{v}_j \cdot T_k \vec{v}_1 = 0 \quad \text{for } \lfloor m/2 \rfloor + 1 \leq j, k \leq m.$$

By (P1), we also have $\vec{v}_1 \cdot \vec{v}_j = 0$ for $\lfloor m/2 \rfloor + 1 \leq j, k \leq m$. Thus, the $m - \lfloor m/2 \rfloor$ nonzero orthogonal vectors $\vec{v}_{\lfloor m/2 \rfloor + 1}, \dots, \vec{v}_m$ lie in the orthogonal complement of the $(m - \lfloor m/2 \rfloor + 1)$-dimensional space spanned by $\vec{v}_1, T_{\lfloor m/2 \rfloor + 1}\vec{v}_1, \dots, T_m \vec{v}_1$. This is impossible, which implies the impossibility of the assumption above that some row and some column of the array have vectors of unequal length. □

Proof of Proposition 4.1. According to Lemmas 4.1 and 4.2, we may assume that all vectors in each row of array $(\vec{v}_{pa})$ have the same length, that is,

$$|\vec{v}_{p1}|^2 = |\vec{v}_{p2}|^2 = \cdots = |\vec{v}_{pm}|^2 \quad \text{for all } p \in \{m+1, \dots, 2m\}. \tag{4-8}$$

Consider the $m \times m$ matrix

$$F = \begin{pmatrix} B_{p1,2m+1} & B_{p2,2m+1} & \cdots & B_{pm,2m+1} \\ B_{p1,2m+2} & B_{p2,2m+2} & \cdots & B_{pm,2m+2} \\ \vdots & \vdots & \ddots & \vdots \\ B_{p1,n} & B_{p2,n} & \cdots & B_{pm,n} \end{pmatrix},$$

whose i-th row is exactly the components of $\vec{v}_{p(2m+i)}$, and whose j-th column is exactly the components of $\vec{v}_{pj}$, where $1 \le i,\ j \le m$. Using properties (P1) and (P2), we have

$${}^tFF = |\vec{v}_{p1}|^2 I_m, \tag{4-9}$$

$$F^tF = \begin{pmatrix} |\vec{v}_{p(2m+1)}|^2 & 0 & \cdots & 0 \\ 0 & |\vec{v}_{p(2m+2)}|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & |\vec{v}_{pn}|^2 \end{pmatrix}. \tag{4-10}$$

From (4-9), we see that $F^tF = {}^tFF$. Then we compare (4-9) with (4-10) to obtain

$$|\vec{v}_{p(2m+1)}|^2 = \cdots = |\vec{v}_{pn}|^2 = |\vec{v}_{p1}|^2 \quad \text{for all } p \in \{m+1, \dots, 2m\}. \tag{4-11}$$

Now, from (3-21) and (4-8), we get $A_a = A_b$ for all $a \neq b$. Similarly, from (3-19) and (4-11) we get $A_\alpha = A_\beta$ for all $\alpha \neq \beta$. These facts together with (3-20) give

$$|\vec{v}_{a\alpha}|^2 = \frac{1}{m^2} \sum_{b,\beta,p} B^2_{b\beta,p} = \frac{1}{m^2} f \quad \text{for all } a, \alpha.$$

Proceeding as in the proof of (4-11), we get

$$|\vec{v}_{(m+1)a}|^2 = \cdots = |\vec{v}_{(2m)a}|^2 = |\vec{v}_{a(2m+1)}|^2 = \cdots = |\vec{v}_{an}|^2 \quad \text{for all } a \in \{1, \dots, m\}, \tag{4-12}$$

$$|\vec{v}_{(m+1)\alpha}|^2 = \cdots = |\vec{v}_{(2m)\alpha}|^2 = |\vec{v}_{1\alpha}|^2 = \cdots = |\vec{v}_{m\alpha}|^2 \quad \text{for all } \alpha \in \{2m+1, \dots, n\}. \tag{4-13}$$

Then (4-11)–(4-13) imply that every vector in each of the three arrays $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ and $(\vec{v}_{pa})$ has length equal to $\sqrt{f}/m$.

Next, we will show that f is constant. Using (2-17), (3-6) and (3-8), we get

$$\sum_i B_{ab,pi}\omega_i=\sum_\alpha B_{a\alpha,p}\omega_{\alpha b}+\sum_\alpha B_{\alpha b,p}\omega_{\alpha a}=\sum_{\alpha,q} B_{a\alpha,p}\Gamma^\alpha_{qb}\omega_q+\sum_{\alpha,q} B_{\alpha b,p}\Gamma^\alpha_{qa}\omega_q.$$

Comparing two sides of this, we obtain $B_{ab,p\alpha}=0$. A similar argument gives $B_{pq,a\alpha}=0$ and $B_{\alpha\beta,ap}=0$. By (2-14), (2-18), (3-1) and (3-4), we easily see that the four indices in $B_{pa,\alpha i}$ for $1\le i\le n$ are totally symmetric. Hence we get

$$0=\sum_i B_{pa,\alpha i}\omega_i=dB_{pa,\alpha}+\sum_b B_{pb,\alpha}\omega_{ba}+\sum_q B_{qa,\alpha}\omega_{qp}+\sum_\beta B_{pa,\beta}\omega_{\beta\alpha}.$$

Multiplying this equation by $B_{pa,\alpha}$ and summing, we get

$$0=\sum_{p,a,\alpha} B_{pa,\alpha}dB_{pa,\alpha}+\sum_{p,a,b,\alpha} B_{pa,\alpha}B_{pb,\alpha}\omega_{ba}+\sum_{p,a,q,\alpha} B_{pa,\alpha}B_{qa,\alpha}\omega_{qp}+\sum_{p,a,\alpha,\beta} B_{pa,\alpha}B_{pa,\beta}\omega_{\beta\alpha},$$

or, equivalently,

$$0=\tfrac{1}{2}df+\sum_{p,a,b}(\vec{v}_{pa}\cdot\vec{v}_{pb})\,\omega_{ba}+\sum_{p,q,a}(\vec{v}_{pa}\cdot\vec{v}_{qa})\,\omega_{qp}+\sum_{p,\alpha,\beta}(\vec{v}_{p\alpha}\cdot\vec{v}_{p\beta})\,\omega_{\beta\alpha}. \tag{4-14}$$

Lemma 3.1 and (4-14) imply that $df=0$, showing that f is constant. □

Lemma 4.3. *The eigenvalues of the Blaschke tensor* **A** *are all constant on* M.

Proof. By (2-14) and (3-19)–(3-21), we get

$$R_{apap}=\frac{2|\vec{v}_{pa}|^2}{(B_3-B_1)(B_3-B_2)}=B_1B_2+A_a+A_p, \tag{4-15}$$

$$R_{a\alpha a\alpha}=\frac{2|\vec{v}_{a\alpha}|^2}{(B_2-B_1)(B_2-B_3)}=B_1B_3+A_a+A_\alpha, \tag{4-16}$$

$$R_{p\alpha p\alpha}=\frac{2|\vec{v}_{p\alpha}|^2}{(B_1-B_2)(B_1-B_3)}=B_2B_3+A_p+A_\alpha. \tag{4-17}$$

Using Proposition 4.1 and adding (4-15), (4-16) and (4-17), we have

$$B_1B_2+B_1B_3+B_2B_3+2(A_a+A_p+A_\alpha)=0. \tag{4-18}$$

From (4-15) up to (4-18) we get

$$\begin{aligned} A_a&=\tfrac{1}{2}(B_2B_3-B_1B_2-B_1B_3)-\frac{2f}{m^2(B_1-B_2)(B_1-B_3)},\\ A_p&=\tfrac{1}{2}(B_1B_3-B_1B_2-B_2B_3)-\frac{2f}{m^2(B_2-B_1)(B_2-B_3)},\\ A_\alpha&=\tfrac{1}{2}(B_1B_2-B_1B_3-B_2B_3)-\frac{2f}{m^2(B_3-B_1)(B_3-B_2)}.\end{aligned} \tag{4-19}$$

Therefore all the eigenvalues of $\boldsymbol{A}$ are constant on M^*. On the other hand, the well-defined continuous functions $A_1, A_2, \ldots, A_n$ satisfy (3-5). Thus we can indeed choose a frame field $\{E_i\}$ around each point of M so that (3-1) and (3-4) hold identically. This fact and the argument above show that the open set M^* is also closed in M. By connectedness, we know that $M^* = M$. □

Remark 4.1. Now that the Blaschke eigenvalues $A_1, A_2, \ldots, A_n$ are constant, we can find everywhere local frame fields $\{E_i\}$ such that (3-1) and (3-4) hold at the same time.

Proof of Theorem 3.1. From Proposition 4.1 and (4-19), we get

$$A_1 = \cdots = A_m,\ A_{m+1} = \cdots = A_{2m},\ A_{2m+1} = \cdots = A_n. \tag{4-20}$$

From Lemma 4.1 we know that $\vec{v}_{pa} \neq 0$; thus there exist α such that $B_{pa,\alpha} \neq 0$. From (3-6), (3-7), (2-9) and that both $A_{ij,k}$ and $B_{ij,k}$ are totally symmetric, we get

$$A_{pa,\alpha} = (A_p - A_a)\Gamma^p_{\alpha a} = (A_a - A_\alpha)\Gamma^a_{p\alpha} = (A_\alpha - A_p)\Gamma^\alpha_{ap}, \tag{4-21}$$

$$B_{pa,\alpha} = (B_2 - B_1)\Gamma^p_{\alpha a} = (B_1 - B_3)\Gamma^a_{p\alpha} = (B_3 - B_2)\Gamma^\alpha_{ap}. \tag{4-22}$$

From (4-21) and (4-22), we derive

$$\frac{A_{pa,\alpha}}{B_{pa,\alpha}} = \frac{A_p - A_a}{B_2 - B_1} = \frac{A_a - A_\alpha}{B_1 - B_3} = \frac{A_\alpha - A_p}{B_3 - B_2},$$

which together with (4-20) implies the existence of constant functions λ and μ with the property

$$\begin{aligned} A_1 + \lambda B_1 = \cdots = A_m + \lambda B_1 = A_{m+1} + \lambda B_2 = \cdots = A_{2m} + \lambda B_2 \\ = A_{2m+1} + \lambda B_3 = \cdots = A_n + \lambda B_3 = \mu. \end{aligned}$$

Hence we have $\boldsymbol{A} + \lambda \boldsymbol{B} - \mu g = 0$, and by it we can apply the result of Li and Wang [2003] to conclude that $x : M \to \mathbb{S}^{n+1}$ is locally Möbius equivalent to one of the following hypersurfaces:

- a hypersurface $\tilde{x} : \tilde{M} \to \mathbb{S}^{n+1}$ with constant mean curvature and constant scalar curvature;
- the image under σ of a hypersurface $\tilde{x} : \tilde{M} \to \mathbb{R}^{n+1}$ with constant mean curvature and constant scalar curvature;
- the image under τ of a hypersurface $\tilde{x} : \tilde{M} \to \mathbb{H}^{n+1}$ with constant mean curvature and constant scalar curvature. Here, we recall that we have defined the conformal diffeomorphism $\tau : \mathbb{H}^{n+1} \to \mathbb{S}^{n+1}_+$, $y \mapsto (1, y')/y_0$, where

$$\mathbb{H}^{n+1} = \{(y_0, y_1, \ldots, y_{n+1}) \in \mathbb{L}^{n+2} \mid \langle y, y\rangle_1 = -1, y_0 \geq 1)\},$$
$$\mathbb{S}^{n+1}_+ = \{(x_1, \ldots, x_{n+2}) \in \mathbb{S}^{n+1} \mid x_1 > 0)\},$$

and $y' = (y_1, \dots, y_{n+1})$.

For each of these possibilities, from [Hu et al. 2007, Propositions 3.1 and 3.2], and because the B_i are all constant, we see that $\tilde{x} : \tilde{M} \to \mathbb{S}^{n+1}$, or $\tilde{x} : \tilde{M} \to \mathbb{R}^{n+1}$, or $\tilde{x} : \tilde{M} \to \mathbb{H}^{n+1}$, respectively, are all Euclidean isoparametric hypersurfaces with three distinct principal curvatures. From the classical result that isoparametric hypersurfaces in $\mathbb{R}^{n+1}$ and $\mathbb{H}^{n+1}$ can have at most two distinct principal curvatures, we finally see that x is Möbius equivalent to an open part of some isoparametric hypersurface in $\mathbb{S}^{n+1}$ with three distinct principal curvatures. □

5. Möbius isoparametric hypersurfaces with $m_1 > m_3$

This section is devoted to Case II and proving Theorem 3.2. Assume that

$$\nabla \boldsymbol{B} \neq 0 \quad \text{and} \quad m_1 \geq m_2 \geq m_3 \text{ such that } m_1 > m_3. \tag{5-1}$$

To add to the index conventions (3-3), we introduce the notation

$$\mathcal{I}_1 = \{1, 2, \dots, m_1\},$$
$$\mathcal{I}_2 = \{m_1 + 1, m_1 + 2, \dots, m_1 + m_2\},$$
$$\mathcal{I}_3 = \{m_1 + m_2 + 1, m_1 + m_2 + 2, \dots, n\}.$$

In follows, we will concentrate on the $m_2 \times m_1$ array $(\vec{v}_{pa})$ of vectors in $\mathbb{R}^{m_3}$.

Lemma 5.1. *There exists an integer m_1', where $0 < m_1 - m_3 \leq m_1' < m_1$, such that exactly m_1' columns of the $m_2 \times m_1$ array $(\vec{v}_{pa})$ are identically zero on an open set $U \subset M^*$. Explicitly, there exists a subset $\mathfrak{D}_0 \subset \mathcal{I}_1$ of m_1' elements, with complement $\mathfrak{D}_1$ in $\mathcal{I}_1$, such that*

$$\vec{v}_{pa} = 0 \quad \textit{for all } a \in \mathfrak{D}_0 \textit{ and } p \in \mathcal{I}_2, \tag{5-2}$$

$$\vec{v}_{pc} \neq 0 \quad \textit{for all } c \in \mathfrak{D}_1 \textit{ and } p \in \mathcal{I}_2. \tag{5-3}$$

Proof. By Lemma 3.1, for each $\bar{p} \in \mathcal{I}_2$, the vectors in row $\bar{p}$ of the array $(\vec{v}_{pa})$ constitute a set of m_1 mutually orthogonal vectors in $\mathbb{R}^{m_3}$. Thus, at least $m_1 - m_3$ vectors in row $\bar{p}$ must be zero at any point of M^*. On the other hand, by Lemmas 3.2 and 3.3 we know that it is impossible that a whole row is zero in the array $(\vec{v}_{pa})$. Permute the columns of $(\vec{v}_{pa})$, so that row $\bar{p}$ has all its nonzero vectors occurring first (left to right). Let $\vec{v}_{\bar{p}\tilde{m}_1}$ denote the last nonzero vector in this row. Then $1 < \tilde{m}_1 \leq m_3 < m_1$. Thus we have

$$\vec{v}_{\bar{p}c} \neq 0 \quad \text{if } 1 \leq c \leq \tilde{m}_1 \quad \text{and} \quad \vec{v}_{\bar{p}a} = 0 \quad \text{if } \tilde{m}_1 + 1 \leq a \leq m_1.$$

Since at least one vector is nonzero in row $\bar{p}$, by Lemma 3.2 the last $m_1 - \tilde{m}_1$ columns of array $(\vec{v}_{pa})$ are all zero on an open set $U \subset M^*$. That is,

$$\text{if } \tilde{m}_1 + 1 \leq a \leq m_1, \quad \text{then } \vec{v}_{pa} = 0 \quad \text{for all } p \in \mathcal{I}_2.$$

Now we apply property (P4) to the minor

$$\begin{pmatrix} \vec{v}_{\bar{p}c} & \vec{v}_{\bar{p}a} \\ \vec{v}_{pc} & \vec{v}_{pa} \end{pmatrix} \quad \text{with } 1 \le c \le \tilde{m}_1,\ \tilde{m}_1 + 1 \le a \le m_1 \text{ and any } p \in \mathcal{I}_2,$$

to obtain

$$|\vec{v}_{(m_1+1)c}| = \cdots = |\vec{v}_{(m_1+m_2)c}| = |\vec{v}_{\bar{p}c}| \neq 0 \quad \text{for all } 1 \le c \le \tilde{m}_1. \tag{5-4}$$

Let $m_1' = m_1 - \tilde{m}_1$. Then $0 < m_1 - m_3 \le m_1' < m_1$ and the assertion follows by setting

$$\mathfrak{D}_0 = \{\tilde{m}_1 + 1, \tilde{m}_1 + 2, \dots, m_1\} \quad \text{and} \quad \mathfrak{D}_1 = \{1, 2, \dots, \tilde{m}_1\}. \qquad \square$$

Lemma 5.2. *Assume that $\nabla \boldsymbol{B} \neq 0$ and $m_1 \ge m_2 \ge m_3$. If $m_1 > m_3$, then $m_2 = m_3$.*

Proof. By (5-3) and Lemma 3.1, for each $c \in \mathfrak{D}_1$ the vectors in column c of the array constitute a set of m_2 mutually orthogonal nonzero vectors in $\mathbb{R}^{m_3}$; hence we have $m_2 \le m_3$. By the assumption $m_2 \ge m_3$, we get $m_2 = m_3$. □

Lemma 5.3. *For all $a, b \in \mathfrak{D}_0$, $c \in \mathfrak{D}_1$, $p, q \in \mathcal{I}_2$ and $\alpha, \beta \in \mathcal{I}_3$, we have*

$$A_a = A_b \neq A_c, \quad A_p = A_q, \quad A_\alpha = A_\beta.$$

Proof. From (5-2) and (5-3), we get that, for all $a, b \in \mathfrak{D}_0$, $c \in \mathfrak{D}_1$ and $p, q \in \mathcal{I}_2$,

$$|\vec{v}_{pa}| = |\vec{v}_{pb}| = |\vec{v}_{qa}| = 0 \neq |\vec{v}_{pc}|.$$

This combined with (3-21) gives $A_a = A_b \neq A_c$ and $A_p = A_q$.

From (5-2) we have

$$B_{pa,\alpha} = 0 \quad \text{for all } a \in \mathfrak{D}_0,\ p \in \mathcal{I}_2,\ \alpha \in \mathcal{I}_3. \tag{5-5}$$

The fact that $B_{ij,k}$ is totally symmetric and (5-5) implies that $\vec{v}_{a\alpha} = 0$ for all $a \in \mathfrak{D}_0$ and $\alpha \in \mathcal{I}_3$. Combining this with (3-20), we get $A_\alpha = A_\beta$. □

Lemma 5.4. $\tilde{m}_1 = m_3 = m_2$.

Proof. By Lemma 5.3, we get $A_p = A_q$ and $A_\alpha = A_\beta$. Combining (3-19) with (5-1), we obtain

$$|\vec{v}_{p\alpha}|^2 = \frac{1}{m_2^2} \sum_{q,\beta,c} B_{q\beta,c}^2 = \frac{1}{m_2^2} f \neq 0 \quad \text{for all } p, \alpha. \tag{5-6}$$

From (5-5) we know that the last $m_1 - \tilde{m}_1$ components of each vector $\vec{v}_{p\alpha}$ are zero on the open set U as we stated in Lemma 5.1; thus $\vec{v}_{p\alpha}$ can be regarded as an element of $\mathbb{R}^{\tilde{m}_1}$. By Lemma 3.1, for each p the vectors in row p of the array $(\vec{v}_{p\alpha})$ constitute a set of m_3 mutually orthogonal nonzero vectors in $\mathbb{R}^{\tilde{m}_1}$. Hence $m_3 \le \tilde{m}_1$, while Lemma 5.1 tells that $\tilde{m}_1 \le m_3$. Hence $\tilde{m}_1 = m_3 = m_2$. □

Next, by using (5-4), (5-6) and Lemma 3.1, we get the following by adapting the proof of Proposition 4.1.

Proposition 5.1. *All the nonzero vectors of the arrays* $(\vec{v}_{p\alpha})$, $(\vec{v}_{a\alpha})$ *and* $(\vec{v}_{pa})$ *have constant length equal to* $\sqrt{f}/m_2$. *That is, we have*

$$|\vec{v}_{cp}|^2 = |\vec{v}_{d\alpha}|^2 = |\vec{v}_{q\beta}|^2 = f/m_2^2 := L^2 = \text{const} \tag{5-7}$$

for any $c, d \in \mathcal{D}_1$, $p, q \in \mathcal{I}_2$ *and* $\alpha, \beta \in \mathcal{I}_3$.

Now, we are ready to prove one of the main results in this section.

Proposition 5.2. *Let* $x : M^n \to \mathbb{S}^{n+1}$ *be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures of multiplicities* $m_1 \geq m_2 \geq m_3$ *and* $m_1 > m_3$. *If the Möbius second fundamental form is not parallel, then it must be the case that* $m_2 = m_3 := m$ *and that the Möbius principal curvatures satisfy* $B_1 = 0$ *and* $B_2 = -B_3 = \pm\sqrt{(n-1)/(2mn)}$.

Proof. By Lemma 5.2 we may assume that $m_2 = m_3 := m$. Let us take $a \in \mathcal{D}_0$, $c \in \mathcal{D}_1$, $p \in \mathcal{I}_2$ and $\alpha \in \mathcal{I}_3$. Then by the proof of Lemma 5.3, we have $\vec{v}_{a\alpha} = 0$. By using (2-14), (3-19)–(3-21) and Lemma 5.1, we obtain

$$R_{apap} = B_1B_2 + A_a + A_p = \frac{2|\vec{v}_{pa}|^2}{(B_3 - B_1)(B_3 - B_2)} = 0, \tag{5-8}$$

$$R_{a\alpha a\alpha} = B_1B_3 + A_a + A_\alpha = \frac{2|\vec{v}_{a\alpha}|^2}{(B_2 - B_1)(B_2 - B_3)} = 0, \tag{5-9}$$

$$R_{cpcp} = B_1B_2 + A_c + A_p = \frac{2|\vec{v}_{pc}|^2}{(B_3 - B_1)(B_3 - B_2)}, \tag{5-10}$$

$$R_{c\alpha c\alpha} = B_1B_3 + A_c + A_\alpha = \frac{2|\vec{v}_{c\alpha}|^2}{(B_2 - B_1)(B_2 - B_3)}, \tag{5-11}$$

$$R_{p\alpha p\alpha} = B_2B_3 + A_p + A_\alpha = \frac{2|\vec{v}_{p\alpha}|^2}{(B_1 - B_2)(B_1 - B_3)}. \tag{5-12}$$

With the summation (5-9) + (5-10) − (5-8) − (5-11), we get

$$\frac{2|\vec{v}_{pc}|^2}{(B_3 - B_1)(B_3 - B_2)} - \frac{2|\vec{v}_{c\alpha}|^2}{(B_2 - B_1)(B_2 - B_3)} = 0.$$

This equation and (5-7) imply that $B_2 + B_3 - 2B_1 = 0$. Combining this with (3-2), we obtain $B_1 = 0$ and $B_2 = -B_3 = \pm\sqrt{(n-1)/(2mn)}$. □

Without loss of generality, in what follows we may assume that

$$B_1 = 0, \quad B_2 = \sqrt{\frac{n-1}{2mn}}, \quad B_3 = -\sqrt{\frac{n-1}{2mn}}. \tag{5-13}$$

Lemma 5.5. *For all $a \in \mathscr{D}_0$, $c \in \mathscr{D}_1$, $p \in \mathscr{I}_2$ and $\alpha \in \mathscr{I}_3$, we have*

$$\omega_{ac} = \omega_{ap} = \omega_{a\alpha} = 0, \qquad \omega_{cp} = \frac{1}{B_1 - B_2} \sum\nolimits_\alpha B_{cp,\alpha}\omega_\alpha,$$
$$\omega_{c\alpha} = \frac{1}{B_1 - B_3} \sum\nolimits_p B_{cp,\alpha}\omega_p, \qquad \omega_{p\alpha} = \frac{1}{B_2 - B_3} \sum\nolimits_c B_{cp,\alpha}\omega_c,$$
$$R_{apap} = R_{a\alpha a\alpha} = 0, \qquad R_{cpcp} = \frac{2}{(B_3 - B_1)(B_3 - B_2)} |\vec{v}_{cp}|^2,$$
$$R_{c\alpha c\alpha} = \frac{2}{(B_2 - B_1)(B_2 - B_3)} |\vec{v}_{c\alpha}|^2, \quad R_{p\alpha p\alpha} = \frac{2}{(B_1 - B_2)(B_1 - B_3)} |\vec{v}_{p\alpha}|^2.$$

Proof. The formulas follow directly from (2-14), (3-6)–(3-8) and (3-19)–(3-21). First of all, from (5-5) we get $\omega_{ap} = \omega_{a\alpha} = 0$. The remaining formulas in Lemma 5.5 except $\omega_{ac} = 0$ can be easily obtained.

To show that $\omega_{ac} = 0$ holds for any $a \in \mathscr{D}_0$ and $c \in \mathscr{D}_1$, we use the following two equations for any $p \in \mathscr{I}_2$ and $\alpha \in \mathscr{I}_3$:

$$0 = -R_{apap}\omega_a \wedge \omega_p = d\omega_{ap} - \sum_i \omega_{ai} \wedge \omega_{ip} = - \sum_{\beta \in \mathscr{I}_3,\, c \in \mathscr{D}_1} \Gamma^c_{\beta p}\omega_{ac} \wedge \omega_\beta, \tag{5-14}$$

$$0 = -R_{a\alpha a\alpha}\omega_a \wedge \omega_\alpha = d\omega_{a\alpha} - \sum_i \omega_{ai} \wedge \omega_{i\alpha} = - \sum_{q \in \mathscr{I}_2,\, c \in \mathscr{D}_1} \Gamma^c_{q\alpha}\omega_{ac} \wedge \omega_q. \tag{5-15}$$

Let us write

$$\omega_{ac} = \sum_{b \in \mathscr{D}_0} \Gamma^a_{bc}\omega_b + \sum_{d \in \mathscr{D}_1} \Gamma^a_{dc}\omega_d + \sum_{q \in \mathscr{I}_2} \Gamma^a_{qc}\omega_q + \sum_{\beta \in \mathscr{I}_3} \Gamma^a_{\beta c}\omega_\beta.$$

Then the two equations above give that

$$\sum_{c \in \mathscr{D}_1} \Gamma^a_{bc}\Gamma^c_{\alpha p} = 0 \quad \text{for all } a, b \in \mathscr{D}_0,\ p \in \mathscr{I}_2,\ \alpha \in \mathscr{I}_3, \tag{5-16}$$

$$\sum_{c \in \mathscr{D}_1} \Gamma^a_{dc}\Gamma^c_{\alpha p} = 0 \quad \text{for all } a \in \mathscr{D}_0,\ d \in \mathscr{D}_1,\ p \in \mathscr{I}_2,\ \alpha \in \mathscr{I}_3, \tag{5-17}$$

$$\sum_{c \in \mathscr{D}_1} \Gamma^a_{qc}\Gamma^c_{\alpha p} = 0 \quad \text{for all } a \in \mathscr{D}_0,\ p, q \in \mathscr{I}_2,\ \alpha \in \mathscr{I}_3, \tag{5-18}$$

$$\sum_{c \in \mathscr{D}_1} \Gamma^a_{\beta c}\Gamma^c_{\alpha p} = 0 \quad \text{for all } a \in \mathscr{D}_0,\ p \in \mathscr{I}_2,\ \alpha, \beta \in \mathscr{I}_3. \tag{5-19}$$

From (5-16), we get for any $b \in \mathscr{D}_0$ a linear system of equations on $\{\Gamma^a_{bc}\}_{1 \le c \le m}$:

$$\begin{cases} B_{p(m_1+m+1),1}\Gamma^a_{b1} + B_{p(m_1+m+1),2}\Gamma^a_{b2} + \cdots + B_{p(m_1+m+1),m}\Gamma^a_{bm} = 0, \\ B_{p(m_1+m+2),1}\Gamma^a_{b1} + B_{p(m_1+m+2),2}\Gamma^a_{b2} + \cdots + B_{p(m_1+m+2),m}\Gamma^a_{bm} = 0, \\ \quad \vdots \\ B_{pn,1}\Gamma^a_{b1} \ + B_{pn,2}\Gamma^a_{b2} \ + \cdots + B_{pn,m}\Gamma^a_{bm} = 0. \end{cases} \tag{5-20}$$

By using (P1), (P2) and Proposition 5.1, we see that the coefficient matrix F of (5-20) satisfies ${}^tFF = \mathrm{diag}(|\vec{v}_{p1}|^2, |\vec{v}_{p2}|^2, \dots, |\vec{v}_{pm}|^2) = |\vec{v}_{p1}|^2 I_m$. Hence we have $|F| \neq 0$, and then (5-20) implies that $\Gamma^a_{b1} = \Gamma^a_{b2} = \cdots = \Gamma^a_{bm} = 0$ for all $b \in \mathfrak{D}_0$, that is,

$$\Gamma^a_{bc} = 0 \quad \text{for all } b \in \mathfrak{D}_0.$$

Analogously, from (5-17), (5-18) and (5-19), respectively, we can show that

$$\Gamma^a_{dc} = \Gamma^a_{qc} = \Gamma^a_{\beta c} = 0 \quad \text{for all } d \in \mathfrak{D}_1,\ q \in \mathcal{I}_2 \text{ and } \beta \in \mathcal{I}_3.$$

Hence $\Gamma^a_{ic} = 0$ for all i, and $\omega_{ac} = 0$ follows. □

Lemma 5.6. *For all $p \in \mathcal{I}_2,\ \alpha \in \mathcal{I}_3$ and $a \in \mathfrak{D}_0,\ c \in \mathfrak{D}_1$,*

$$A_a = -A_c = -A_p = -A_\alpha = -\frac{n-1}{12mn}.$$

Proof. Lemma 5.5 and (2-16) imply that $R_{acij} = 0$ and thus we have $R_{acac} = 0$. On the other hand, (2-14) gives that $R_{acac} = B_1^2 + A_a + A_c$. It follows that $A_a = -A_c$. From (5-8), (5-9) and (5-13), we further get $A_a = -A_p = -A_\alpha$ and hence

$$A_a = -A_c = -A_p = -A_\alpha. \tag{5-21}$$

These together with (5-10), (5-12) and (5-13) give that

$$A_c = \frac{L^2}{2{B_2}^2} = A_p = \frac{B_2^2}{2} - \frac{L^2}{B_2^2}.$$

It follows that $L^2 = \frac{1}{3}B_2^4$ and $A_c = \frac{1}{6}B_2^2$. Then our conclusions follow immediately from (5-13) and (5-21). □

Remark 5.1. Because all the Blaschke eigenvalues $A_1, A_2, \dots, A_n$ are constant on M^*, the reasoning of the proof of Lemma 4.3 shows that $M = M^*$. Hence we can find everywhere local frame fields $\{E_i\}$, such that (3-1) and (3-4) hold simultaneously in Case II.

Lemma 5.6 shows that the Blaschke tensor has exactly two distinct constant eigenvalues. Then applying Theorem 2.4 we immediately get the following result.

Theorem 5.1. *Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with nonparallel Möbius second fundamental form and three distinct Möbius principal curvatures whose multiplicities are not equal. Then there is an $\tilde{n}$ with $2 \le \tilde{n} \le n-1$, and locally x is Möbius equivalent to one of the following two families of hypersurfaces in $\mathbb{S}^{n+1}$:*

($\mathfrak{C}$1) *Minimal hypersurfaces defined by*

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : \tilde{M}^n = N^{\tilde{n}} \times \mathbb{H}^{n-\tilde{n}}(-r^{-2}) \to \mathbb{S}^{n+1},$$

with $r > 0$ *and*

$$\tilde{x}_1 = y_1/y_0, \quad \tilde{x}_2 = y_2/y_0, \quad y_0 \in \mathbb{R}^+, \quad y_1 \in \mathbb{R}^{\tilde{n}+2}, \quad y_2 \in \mathbb{R}^{n-\tilde{n}},$$

where $y_1 : N^{\tilde{n}} \to \mathbb{S}^{\tilde{n}+1}(r) \hookrightarrow \mathbb{R}^{\tilde{n}+2}$ *is an umbilic-free minimal hypersurface immersed into the* $(\tilde{n}+1)$*-dimensional sphere of radius* r *and constant scalar curvature*

$$\tilde{R}_1 = \frac{n\tilde{n}(\tilde{n}-1) - (n-1)r^2}{nr^2},$$

and $(y_0, y_2) : \mathbb{H}^{n-\tilde{n}}(-r^{-2}) \to \mathbb{L}^{n-\tilde{n}+1}$ *is the standard embedding of hyperbolic space of sectional curvature* $-r^{-2}$ *into the* $(n-\tilde{n}+1)$*-dimensional Lorentz space with* $-y_0^2 + y_2^2 = -r^2$.

($\mathfrak{C}2$) *Nonminimal hypersurfaces defined by*

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : \tilde{M}^n = N^{\tilde{n}} \times \mathbb{S}^{n-\tilde{n}}(r) \to \mathbb{S}^{n+1},$$

with $r > 0$ *and*

$$\tilde{x}_1 = y_1/y_0, \quad \tilde{x}_2 = y_2/y_0, \quad y_0 \in \mathbb{R}^+, \quad y_1 \in \mathbb{R}^{\tilde{n}+1}, \quad y_2 \in \mathbb{R}^{n-\tilde{n}+1},$$

where $(y_0, y_1) : N^{\tilde{n}} \to \mathbb{H}^{\tilde{n}+1}(-r^{-2}) \hookrightarrow \mathbb{L}^{\tilde{n}+2}$, *with* $-y_0^2 + y_1^2 = -r^2$, *is an umbilic-free minimal hypersurface immersed into* $(\tilde{n}+1)$*-dimensional hyperbolic space of sectional curvature* $-r^{-2}$ *and constant scalar curvature*

$$\tilde{R}_1 = -\frac{n\tilde{n}(\tilde{n}-1) + (n-1)r^2}{nr^2},$$

and $y_2 : \mathbb{S}^{n-\tilde{n}}(r) \to \mathbb{R}^{n-\tilde{n}+1}$ *is the standard embedding of the* $(n-\tilde{n})$*-sphere of radius* r.

Determining which of the hypersurfaces ($\mathfrak{C}1$) and ($\mathfrak{C}2$) is Möbius isoparametric requires knowing their Möbius invariants — but this was done in [Hu and Li 2007, Section 4]. For simplicity we will not repeat this calculation here. With the omitted calculations and Lemma 5.6, we immediately get the following results.

Proposition 5.3. *A hypersurface* $\tilde{x}$ *in* ($\mathfrak{C}_1$) *is Möbius isoparametric if and only if it satisfies*

(1) $\tilde{n} = 3m$;

(2) $r = \sqrt{6mn/(n-1)}$;

(3) $y_1 : N^{3m} \to \mathbb{S}^{3m+1}(\sqrt{6mn/(n-1)})$ *is a minimal isoparametric hypersurface with constant scalar curvature* $\tilde{R}_1 = 3(m-1)(n-1)/(2n)$; *moreover, it has three distinct principal curvatures with values given by* (1-1), *each of them with the same multiplicity* m.

Remark 5.2. Cartan [1939] proved that minimal isoparametric hypersurfaces in $\mathbb{S}^{3m+1}(\sqrt{6mn/(n-1)})$ with three distinct principal curvatures do exist and are unique with principal curvatures having the same multiplicities $m \in \{1, 2, 4, 8\}$. More precisely, it is the tube of constant radius over a standard Veronese embedding of a projective plane $\mathbb{F}P^2$ into $\mathbb{S}^{3m+1}(\sqrt{6mn/(n-1)})$ with principal curvatures of (1-1) where $m = 1, 2, 4$ or 8, and $\mathbb{F}$ is the division algebra $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ (quaternions) or $\mathbb{O}$ (Cayley numbers), respectively.

Proposition 5.4. *If a hypersurface $\tilde{x}$ in* ($\mathfrak{C}_2$) *is Möbius isoparametric*, *then it must satisfy the following three conditions*:

(1) $\tilde{n} = n - 3m$;

(2) $r = \sqrt{6mn/(n-1)}$;

(3) $\tilde{y} = (y_0, y_1) : N^{n-3m} \to \mathbb{H}^{n-3m+1}(-(n-1)/(6mn))$ *is a minimal isoparametric hypersurface with the principal curvatures of* (1-1).

On the other hand, by Cartan's theorem [1938], an isoparametric hypersurface M^n in the hyperbolic space $\mathbb{H}^{n+1}$ can have at most two distinct principal curvatures, which can only be either totally umbilic or else an open subset of a standard product $\mathbb{S}^k \times \mathbb{H}^{n-k}$ in $\mathbb{H}^{n+1}$. Moreover, the latter must be nonminimal. From this fact and Proposition 5.4, we immediately get the following:

Proposition 5.5. *There is no Möbius isoparametric hypersurface in* ($\mathfrak{C}2$) *that has three distinct Möbius principal curvatures.*

Proof of Theorem 3.2. This is an immediate consequence of the Theorem 5.1, Remark 5.1 and Propositions 5.3 and 5.5. □

6. Completion of the proof of the classification theorem

Let $x : M^n \to \mathbb{S}^{n+1}$ be a Möbius isoparametric hypersurface with three distinct Möbius principal curvatures whose multiplicities satisfy $m_1 \geq m_2 \geq m_3$.

If x has parallel Möbius second fundamental form, then we apply Theorem 2.3 to obtain that it is locally Möbius equivalent to a hypersurface in part (i) of the classification theorem.

If x has nonparallel Möbius second fundamental form, then we have exactly two cases as we stated in section three:

For Case I, we apply Theorem 3.1 and Cartan's theorem to obtain that it is locally Möbius equivalent to a hypersurface in (ii). For Case II, we can apply Theorem 3.2 and Cartan's theorem to conclude that it is locally Möbius equivalent to the hypersurface in (iii). □

Final remarks. For the general theory (see [Wang 1998]) of Möbius submanifolds in $\mathbb{S}^{n+p}$, the Möbius form Φ is an important invariant. Closely related to Möbius isoparametric hypersurfaces is the concept of Blaschke isoparametric hypersurfaces in spheres. It is interesting to mention a conjecture by X. X. Li [Li and Zhang 2009; Li and Peng 2010]: *A Blaschke isoparametric hypersurfaces with more than two distinct Blaschke eigenvalues is Möbius isoparametric*. For definitions and some recent progress on Blaschke isoparametric hypersurfaces, see [Li and Peng 2010; Li and Zhang 2006; 2007; 2009].

Acknowledgements

We would like to thank Professors H. Li, C. P. Wang and X. X. Li for useful comments and help. We are also greatly indebted to the paper of T. E. Cecil and G. R. Jensen [1998], the crucial algebraic techniques of which make this paper more easily accessible.

References

[Akivis and Goldberg 1997] M. A. Akivis and V. V. Goldberg, "A conformal differential invariant and the conformal rigidity of hypersurfaces", *Proc. Amer. Math. Soc.* **125**:8 (1997), 2415–2424. MR 97j:53017 Zbl 0887.53030

[Cartan 1938] É. Cartan, "Familles de surfaces isoparamétriques dans les espaces à courbure constante", *Ann. Mat. Pura Appl.* **17**:1 (1938), 177–191. MR 1553310 Zbl 0020.06505

[Cartan 1939] E. Cartan, "Sur des familles remarquables d'hypersurfaces isoparamétriques dans les espaces sphériques", *Math. Z.* **45** (1939), 335–367. MR 1,28f Zbl 0021.15603

[Cecil and Jensen 1998] T. E. Cecil and G. R. Jensen, "Dupin hypersurfaces with three principal curvatures", *Invent. Math.* **132**:1 (1998), 121–178. MR 2000k:53051 Zbl 0908.53007

[Cecil and Jensen 2000] T. E. Cecil and G. R. Jensen, "Dupin hypersurfaces with four principal curvatures", *Geom. Dedicata* **79**:1 (2000), 1–49. MR 2001g:53103 Zbl 0965.53039

[Cecil et al. 2007] T. E. Cecil, Q.-S. Chi, and G. R. Jensen, "Dupin hypersurfaces with four principal curvatures, II", *Geom. Dedicata* **128** (2007), 55–95. MR 2009k:53116 Zbl 1144.53067

[Hu and Li 2004] Z. Hu and H. Li, "Classification of hypersurfaces with parallel Möbius second fundamental form in $\mathbb{S}^{n+1}$", *Sci. China Ser. A* **47**:3 (2004), 417–430. MR 2005c:53066 Zbl 1082.53016

[Hu and Li 2005] Z. Hu and H. Li, "Classification of Möbius isoparametric hypersurfaces in $\mathbb{S}^4$", *Nagoya Math. J.* **179** (2005), 147–162. MR 2006e:53098 Zbl 1110.53010

[Hu and Li 2007] Z. Hu and D. Li, "Möbius isoparametric hypersurfaces with three distinct principal curvatures", *Pacific J. Math.* **232**:2 (2007), 289–311. MR 2008m:53131 Zbl 1154.53011

[Hu and Tian 2009] Z. J. Hu and X. L. Tian, "On Möbius form and Möbius isoparametric hypersurfaces", *Acta Math. Sin.* (*Engl. Ser.*) **25**:12 (2009), 2077–2092. MR 2010m:53076 Zbl 1191.53014

[Hu and Zhai 2008] Z. Hu and S. Zhai, "Classification of Möbius isoparametric hypersurfaces in the unit six-sphere", *Tohoku Math. J.* (2) **60**:4 (2008), 499–526. MR 2010f:53022 Zbl 1165.53008

[Hu et al. 2007] Z. Hu, H. Li, and C. Wang, "Classification of Möbius isoparametric hypersurfaces in $\mathbb{S}^5$", *Monatsh. Math.* **151**:3 (2007), 201–222. MR 2009g:53091 Zbl 1144.53021

[Li and Peng 2010] X. Li and Y. Peng, "Classification of the Blaschke isoparametric hypersurfaces with three distinct Blaschke eigenvalues", *Results Math.* **58**:1-2 (2010), 145–172. MR 2672631 Zbl 1202.53016

[Li and Wang 2003] H. Li and C. Wang, "Möbius geometry of hypersurfaces with constant mean curvature and scalar curvature", *Manuscripta Math.* **112**:1 (2003), 1–13. MR 2004e:53092 Zbl 1041.53008

[Li and Zhang 2006] X. Li and F. Zhang, "A classification of immersed hypersurfaces in spheres with parallel Blaschke tensors", *Tohoku Math. J.* (2) **58**:4 (2006), 581–597. MR 2008d:53077 Zbl 1135.53309

[Li and Zhang 2007] X. X. Li and F. Y. Zhang, "Immersed hypersurfaces in the unit sphere $\mathbb{S}^{m+1}$ with constant Blaschke eigenvalues", *Acta Math. Sin.* (*Engl. Ser.*) **23**:3 (2007), 533–548. MR 2008d:53078 Zbl 1151.53013

[Li and Zhang 2009] X. X. Li and F. Y. Zhang, "On the Blaschke isoparametric hypersurfaces in the unit sphere", *Acta Math. Sin.* (*Engl. Ser.*) **25**:4 (2009), 657–678. MR 2010c:53089 Zbl 1177.53016

[Li et al. 2002] H. Z. Li, H. L. Liu, C. P. Wang, and G. S. Zhao, "Möbius isoparametric hypersurfaces in $\mathbb{S}^{n+1}$ with two distinct principal curvatures", *Acta Math. Sin.* (*Engl. Ser.*) **18**:3 (2002), 437–446. MR 2003h:53079 Zbl 1030.53017

[Miyaoka and Ozawa 1989] R. Miyaoka and T. Ozawa, "Construction of taut embeddings and Cecil–Ryan conjecture", pp. 181–189 in *Geometry of manifolds* (Matsumoto, 1988), edited by K. Shiohama, Perspect. Math. **8**, Academic Press, Boston, MA, 1989. MR 92f:53071 Zbl 0697.53055

[Niebergall 1991] R. Niebergall, "Dupin hypersurfaces in $\mathbb{R}^5$, I", *Geom. Dedicata* **40**:1 (1991), 1–22. MR 92k:53106 Zbl 0733.53031

[Pinkall 1985] U. Pinkall, "Dupinsche Hyperflächen in $\mathbb{E}^4$", *Manuscripta Math.* **51**:1-3 (1985), 89–119. MR 86m:53010

[Pinkall and Thorbergsson 1989] U. Pinkall and G. Thorbergsson, "Deformations of Dupin hypersurfaces", *Proc. Amer. Math. Soc.* **107**:4 (1989), 1037–1043. MR 90c:53145 Zbl 0682.53061

[Rodrigues and Tenenblat 2009] L. A. Rodrigues and K. Tenenblat, "A characterization of Moebius isoparametric hypersurfaces of the sphere", *Monatsh. Math.* **158**:3 (2009), 321–327. MR 2552098 Zbl 1190.53008

[Thorbergsson 1983] G. Thorbergsson, "Dupin hypersurfaces", *Bull. London Math. Soc.* **15** (1983), 493–498. MR 85b:53066 Zbl 0529.53044

[Wang 1998] C. Wang, "Moebius geometry of submanifolds in $\mathbb{S}^n$", *Manuscripta Math.* **96**:4 (1998), 517–534. MR 2000a:53019 Zbl 0912.53012

Received January 1, 2010. Revised March 13, 2010.

ZEJUN HU
DEPARTMENT OF MATHEMATICS
ZHENGZHOU UNIVERSITY
ZHENGZHOU 450052
CHINA
huzj@zzu.edu.cn

SHUJIE ZHAI
DEPARTMENT OF MATHEMATICS
ZHENGZHOU UNIVERSITY
ZHENGZHOU 450052
CHINA
zhaishujie@zzu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

DISCRETE MORSE THEORY AND HOPF BUNDLES

DMITRY N. KOZLOV

We use Hopf bundles to give an example of a regular CW complex X and an acyclic matching M on the face poset of X, such that there are no critical cells in neighboring dimensions but the complex X is not homotopy equivalent to the corresponding wedge of spheres. The key fact here is that the higher homotopy groups of spheres are nontrivial. We also give a sufficient condition on an acyclic matching M for concluding that X is homotopy equivalent to a wedge of spheres indexed by the critical cells.

1. Introduction

Discrete Morse theory, introduced by Robin Forman [1998], has become quite a useful tool for doing specific computations in combinatorial algebraic topology; see [Kozlov 2008] for the general framework, and [Clark and Ehrenborg $\geq$ 2011] for an interesting recent application in case of the Frobenius complex.

Let us briefly describe how the computational model provided by discrete Morse theory works. Given a regular CW complex X, let $\mathcal{F}(X)$ denote the poset of all nonempty cells of X. This poset is ranked by the dimensions of the cells. A partial matching on the Hasse diagram of $\mathcal{F}(X)$ is a bijection $M : U \to D$, where U and D are (possibly empty) disjoint sets of elements of $\mathcal{F}(X)$ such that $\dim(\sigma) = \dim(M(\sigma)) + 1$, and $M(\sigma)$ lies on the boundary of σ for all $\sigma \in U$. A partial matching M is called *acyclic* if there do not exist $\sigma_1, \dots, \sigma_t \in \mathcal{F}(X)$ such that $\sigma_1 \neq \cdots \neq \sigma_t$, and $\sigma_{i+1} > M(\sigma_i)$ for all $i = 1, \dots, t$, where as usual we set $\sigma_{t+1} := \sigma_1$. We set $C_M := \mathcal{F}(X) \setminus (U \cup D)$ and call the elements of C *critical*. For all i, let $f_i(C_M)$ denote the number of critical cells of dimension i. The main theorem in [Forman 1998] states that whenever M is an acyclic matching, there exists a CW complex $\widetilde{X}$, called the *critical Morse complex*, with $f_i(C_M)$ cells of dimension i, for all i, such that $\widetilde{X}$ is homotopy equivalent to X.

Frequently, the actual goal of applying discrete Morse theory is to prove that X is in fact homotopy equivalent to a wedge of spheres, or at least to compute the

This research was supported by University of Bremen, as part of AG CALTOP.
MSC2000: primary 57Q05; secondary 54G20.
Keywords: homotopy group, fibrations, gluing map, acyclic matching, long exact sequence for homotopy.

homology groups of X. In a very fortunate situation, one might be able to produce an acyclic matching M such that for all i we have $f_i(C_M)f_{i+1}(C_M) = 0$, that is, there are no critical cells in neighboring dimensions. This would settle the question of computing the homology groups of the space X. However, in order to determine the homotopy type of X, one would want to conclude that the critical cells are somehow independent of each other, and so we have the homotopy equivalence

$$X \simeq \bigvee_i \underbrace{S^i \vee \cdots \vee S^i}_{\tilde{f}_i(C_M)}, \tag{1-1}$$

where $\tilde{f}_i(C_M) = f_i(C_M)$ for $i \geq 1$ and $\tilde{f}_0(C_M) = f_0(C_M) - 1$, and we use the convention that the empty wedge is a point.

We will use the fact that higher homotopy groups of spheres are nontrivial to give an example showing that just assuming that there are no critical cells in neighboring dimensions is not enough to conclude that the space is homotopy equivalent to a wedge of spheres.

But first, on the positive side, we give a sufficient condition on acyclic matching that lets us conclude that the space is homotopy equivalent to a wedge of spheres indexed by the critical cells. In fact, for this result we will not need the condition that there are no critical cells in neighboring dimensions; see also [Clark and Ehrenborg $\geq$ 2011, Theorem 2.5].

2. Acyclic matchings yielding a wedge of spheres

Here we are interested in acyclic matchings that allow us to conclude that the considered complex is in fact homotopy equivalent to a wedge of spheres that are enumerated by the critical cells. First we need some terminology.

Definition 2.1. Let P be a partially ordered set and M a partial matching on P.

(1) A *generalized alternating path* is a sequence $x_1 > x_2 < x_3 > \cdots < x_{2t+1}$ or a sequence $x_1 > x_2 < x_3 > \cdots > x_{2t+2}$, where $t \geq 0$, such that $M(x_{2k+1}) = x_{2k}$ for all $k = 1, \ldots, t$.

(2) Let x be an element of P. We set $F(x)$ to be the set of the endpoints of all generalized alternating paths starting at x, and call $F(x)$ the *feasibility domain* of x.

Note that in a generalized alternating path, we require that x_{2k+1} covers x_{2k} for all $k = 1, \ldots, t$, but we obviously do not require that x_{2k-1} covers x_{2k} for all such k.

It is easy to see that $F(x)$ shall always contain a critical cell of dimension 0. Let A denote the set of 0-dimensional cells in $F(x)$. If none of them is critical, then there exists the set of 1-dimensional cells $B \subset F(x)$ such that $M : B \to A$ is a bijection. Since every $y \in B$ covers two elements, the graph with the vertex set

$A \cup B$ and the covering relations as edges cannot be a forest, so it contains cycles, which contradicts the assumption that the matching M is acyclic.

The following theorem gives a sufficient condition on an acyclic matching for the critical Morse complex to be homotopy equivalent to a wedge of spheres enumerated by critical cells.

Theorem 2.2. *Let X be a connected regular CW complex, and let M be an acyclic partial matching on $\mathcal{F}(X)$. Assume that for every critical cell c of dimension larger than* 0, *its feasibility domain $F(c)$ contains precisely two critical cells*: *c itself and one critical cell of dimension* 0. *Then X is homotopy equivalent to a wedge of spheres enumerated by critical cells*, *that is*, (1-1) *is true.*

Proof. For this argument, we adopt the point of view of [Kozlov 2008] and follow the proof of its Theorem 11.13(b). There the main theorem of discrete Morse theory for CW complexes is proved by a stepwise attachment of either a critical cell or of a pair of cells matched by M, with a parallel explicit construction of a Morse homotopy map. This stepwise attachment is done along a certain linear extension of the face poset of X, which we denote by l. When a pair of matched cells is attached, we simply have a strong deformation retraction of the obtained complex to what we have had before that attachment, so we just need to understand the case of attaching a critical cell.

Assume that a critical cell c of dimension at least 1 is being attached. The cells in $F(c) \setminus \{c\}$ form a subcomplex C of X. The assumption of the theorem implies that C is collapsible along the matching M. It means that prior to the attachment of c, the Morse homotopy has already shrunk the complex C to a point a, where a is the critical 0-dimensional cell of $F(c) \setminus \{c\}$. Since the image of the attaching map of the cell c lies inside C, we conclude that in the critical Morse complex, the attaching map of c will simply map everything to the point a. Thus we can conclude that all the attaching maps in the critical Morse complex are trivial.

Finally, we need to see that the matching M has exactly one 0-dimensional critical cell, which will imply that all the critical cells will be attached to the same vertex. Assume we have another critical 0-dimensional cell b, and assume that b occurs after a in the linear extension l. Then, when b is added, it will form a new connected component. So, since the total complex X is connected, at some point in the inductive process of adding critical cells and pairs of matched cells we will have to connect that connected component to the connected component containing a. This can only be achieved by adding a critical 1-dimensional cell, which we call e. The set $F(e)$ cannot contain any critical 0-dimensional cells other than b. Let v_1 denote the vertex of e that does not lie in the same connected component as b. The vertex v_1 is not critical, and we set $e_1 := M^{-1}(v_1)$. Both v_1 and e_1 were added before e. We now proceed, starting with $k = 1$ be letting

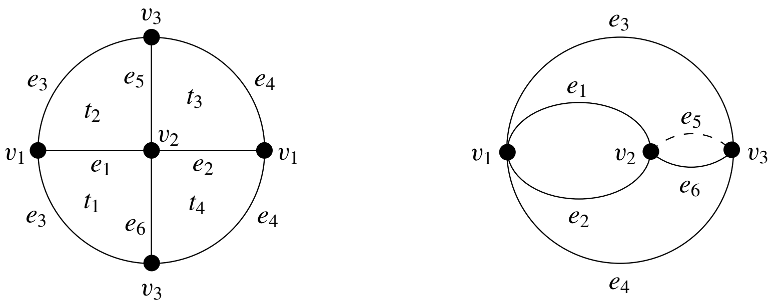


Figure 2.1. Two presentations of the simplicial complex L.

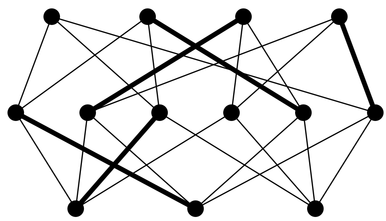


Figure 2.2. The face poset of L. The vertices, edges, and triangles are shown with the index increasing from left to right. The bold edges indicate an acyclic matching.

v_{k+1} be the vertex of e_k other than v_k. Since $v_{k+1} \in F(e)$, we see that v_{k+1} is not critical, and set $e_{k+1} := M^{-1}(v_{k+1})$. Both v_{k+1} and e_{k+1} were added before e. Eventually we will have to conclude that for some $k \geq 1$ the vertex v_{k+1} lies in the same connected component as b. But this means that b was connected to the vertex v_1 even before adding e, yielding a contradiction to the choice of e. $\square$

The condition of Theorem 2.2 is not necessary for getting a wedge of spheres enumerated by critical cells. For example, let L be the simplicial complex shown on Figure 2.1. A direct examination yields that the matching shown on Figure 2.2 is acyclic with one critical cell in each of the dimensions 0, 1, and 2. The condition of Theorem 2.2 is not satisfied, but the space L is homotopy equivalent to $S^1 \vee S^2$.

3. Hopf fiber bundles

We now give an example that simply having an acyclic matching with no critical cells in neighboring dimensions is not sufficient to conclude that the space X is homotopy equivalent to a wedge of spheres. Our example exploits the fact that the higher homotopy groups (unlike the homology groups) of spheres are nontrivial. The first such nontrivial group is $\pi_3(S^2) = \mathbb{Z}$, and it is this one which we use for our construction.

Consider the set $A := \{(z_1, z_2) \mid z_1, z_2 \in \mathbb{C}, |z_1|^2 + |z_2|^2 = 1\}$, and let the multiplicative group $G = \{z \mid |z| = 1\} \subseteq \mathbb{C}$ act on A diagonally by multiplication: $z : (z_1, z_2) \mapsto (zz_1, zz_2)$. The quotient A/G can be viewed as a complex projective line $\mathbb{CP}^1$, with the quotient map $q : A \to A/G$ simply being $q : (z_1, z_2) \mapsto (z_1 : z_2)$. Note that topologically $A \cong S^3$, $G \cong S^1$, and $A/G \cong S^2$. The obtained fiber bundle $S^1 \to S^3 \to S^2$ is the first example of a Hopf bundle; see [Hatcher 2002, Example 4.45].

Consider the CW structure on A obtained by intersecting with the real coordinate hyperplanes $\operatorname{Re} z_1 = 0$, $\operatorname{Im} z_1 = 0$, $\operatorname{Re} z_2 = 0$, and $\operatorname{Im} z_2 = 0$. Then A is a regular CW complex with face vector (8, 24, 32, 16). Furthermore, consider the CW structure on A/G consisting of the two vertices $v_1 = (1 : 0)$ and $v_2 = (0 : 1)$, four edges $e_1 = \{(1 : r) \mid r > 0\}$, $e_2 = \{(1 : ir) \mid r > 0\}$, $e_3 = \{(1 : -r) \mid r > 0\}$, and $e_4 = \{(1 : -ir) \mid r > 0\}$, and four 2-cells denoted s_1, s_2, s_3, s_4, where s_i is bound by e_i and e_{i+1} for $i = 1, 2, 3$, and s_4 is bound by e_1 and e_4. Again A/G is a regular CW complex with the face vector (2, 4, 4), and one sees that q is a cellular map.

Set $C := ((A \times [0, 1]) \coprod (A/G))/\sim$ to be the mapping cylinder of q, that is $\sim$ is given by $(a, 1) \sim q(a)$, for all $a \in A$. We choose a CW structure on C by taking all the cells of $A/G \subseteq C$, subdividing $A \times \{0\}$, the top copy of A, as described above, and taking the open cells $\tilde{\sigma} := \operatorname{int} \sigma \times (0, 1)$ for all cells σ of A. Here we write $\operatorname{int} \sigma = \sigma \setminus \partial\sigma$ to denote the relative interiors of cells. Finally, let X be the regular CW complex obtained from C by attaching a 4-cell k along $A \times \{0\} \cong S^3$.

Consider the following acyclic matching: $M(\tilde{\sigma}) = \sigma$ whenever σ is a cell of $A \times \{0\}$, $M(s_i) = e_i$ for $i = 1, 2, 3$, and $M(e_4) = v_2$. The partial matching M has three critical cells: v_1, k, and s_4, in dimensions 0, 2, and 4. It is easily verified directly that all the matched pairs are regular in the sense of [Cohen 1973] and [Kozlov 2008, Definition 11.12]; in particular the main theorem of discrete Morse theory (see [Forman 1998]) can be applied and we can conclude that X is homotopy equivalent to a CW complex with one cell in each of the dimensions 0, 2, and 4.

However, the space X is not homotopy equivalent to $S^2 \vee S^4$. For example, these two spaces have different π_3 groups.[1] Namely $\pi_3(X) = 0$, while $\pi_3(S^2 \vee S^4) = \mathbb{Z}$. Both of these statements can be seen using the long exact sequence for relative homotopy; see [Hatcher 2002, page 344]. Indeed, when a space $\widetilde{Y}$ is obtained from a space $Y \cong S^2$ by attaching a 4-cell along some continuous map $\varphi : S^3 \to S^2$, the relevant part of the long exact sequence for homotopy of the pair $(\widetilde{Y}, Y)$ is

$$\cdots \to \pi_4(\widetilde{Y}, Y, y) \xrightarrow{\partial} \pi_3(Y, y) \xrightarrow{i_*} \pi_3(\widetilde{Y}, y) \to \pi_3(\widetilde{Y}, Y, y) \to \cdots, \tag{3-1}$$

[1] Serge Ochanine pointed out to the author in 2009 that the constructed space X is actually the complex projective plane. We use the long exact sequence of the homotopy of a pair to show how we came up with the example.

where $y \in Y$ is a base point, the map ∂ comes from restricting maps $(D^n, S^{n-1}, s) \to (\widetilde{Y}, Y, y)$ to S^{n-1}, and i_* is the map between homotopy groups induced by the inclusion map $i : Y \hookrightarrow \widetilde{Y}$. Since $(Y, y) \cong (S^2, x_2)$, and $(\widetilde{Y}, Y, y) \simeq (D^4, S^3, x_3)$, where $x_2 \in S^2$ and $x_3 \in S^3$ are corresponding base points, the sequence (3-1) translates to

$$\cdots \to \pi_3(S^3, x_3) \xrightarrow{\varphi_*} \pi_3(S^2, x_2) \to \pi_3(\widetilde{Y}, y) \to 0 \to \cdots . \tag{3-2}$$

For the space $S^2 \vee S^4$ the map φ_* is trivial, and hence $\pi_3(\widetilde{Y}, y) = \pi_3(S^2, x_2) = \mathbb{Z}$; for the case of the Hopf bundle above, the map φ_* is surjective, and so we get $\pi_3(\widetilde{Y}, y) = 0$. Clearly, this technique can be used to produce further examples that might be needed to test various hypothesis.

Taking the barycentric subdivision of X will yield a simplicial complex with the same property: It has an acyclic matching with one critical cell of dimensions 0, 2, and 4 each, but is of course homeomorphic to the regular CW version; in particular, it is not homotopy equivalent to $S^2 \vee S^4$. We leave finding such an acyclic matching to the reader.

Acknowledgments

I thank the University of Bremen and the Mathematical Sciences Research Institute at Berkeley for being able to conduct this work. I also thank Eric Clark, Richard Ehrenborg, and Serge Ochanine for their valuable comments.

References

[Clark and Ehrenborg ≥ 2011] E. Clark and R. Ehrenborg, "The Frobenius complex", preprint. To appear in *Ann. Comb.*

[Cohen 1973] M. M. Cohen, *A course in simple-homotopy theory*, Graduate Texts in Mathematics **10**, Springer, New York, 1973. MR 50 #14762 Zbl 0261.57009

[Forman 1998] R. Forman, "Morse theory for cell complexes", *Adv. Math.* **134**:1 (1998), 90–145. MR 99b:57050 Zbl 0896.57023

[Hatcher 2002] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. MR 2002k:55001 Zbl 1044.55001

[Kozlov 2008] D. Kozlov, *Combinatorial algebraic topology*, Algorithms and Computation in Mathematics **21**, Springer, Berlin, 2008. MR 2008j:55001 Zbl 1130.55001

Received November 12, 2009.

DMITRY N. KOZLOV
DEPARTMENT OF MATHEMATICS
BREMEN UNIVERSITY
BIBLIOTHEKSSTRASSE 1
D-28334 BREMEN
GERMANY
dfk@math.uni-bremen.de
http://www.math.uni-bremen.de/~dfk/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

REGULARITY OF CANONICAL AND DEFICIENCY MODULES FOR MONOMIAL IDEALS

MANOJ KUMMINI AND SATOSHI MURAI

We show that the Castelnuovo–Mumford regularity of the canonical or a deficiency module of the quotient of a polynomial ring by a monomial ideal is bounded by its dimension.

1. Introduction

Let $R = \Bbbk[x_1, \dots, x_n]$ be a standard graded polynomial ring over a field $\Bbbk$, and let $\mathfrak{m} = (x_1, \dots, x_n)$ be the homogeneous maximal ideal of R. We study the Castelnuovo–Mumford regularity of the modules $\operatorname{Ext}^i_R(R/I, \omega_R)$ when $I \subset R$ is a monomial ideal; here $\omega_R = R(-n)$ denotes the canonical module of R. The modules

$$\operatorname{Ext}^i_R(R/I, \omega_R) \quad \text{for } i > n - \dim R/I$$

are called the *deficiency modules* of R/I, while

$$\operatorname{Ext}^{n-\dim R/I}_R(R/I, \omega_R)$$

is called the *canonical module* of R/I.

For any homogeneous ideal $I \subseteq R$, the local cohomology modules $\mathrm{H}^i_{\mathfrak{m}}(R/I)$ are important in commutative algebra and algebraic geometry. One is often interested in the vanishing of homogeneous components of $\mathrm{H}^i_{\mathfrak{m}}(R/I)$. While one cannot expect the vanishing of $\mathrm{H}^i_{\mathfrak{m}}(R/I)$ in negative degrees (unless it has finite length), one can, using the local duality theorem of Grothendieck, obtain some information from $\operatorname{Ext}^{n-i}_R(R/I, \omega_R)$. For a finitely generated graded R-module M, its (*Castelnuovo–Mumford*) *regularity* $\operatorname{reg}(M)$ is an invariant that contains information about the stability of homogeneous components in sufficiently large degrees. In light of these, it is desirable to get bounds on $\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R))$. Such bounds were studied by L. T. Hoa and E. Hyry [2006] and by M. Chardin, D. T. Ha and Hoa [2009]; see also the references in those papers.

Unfortunately, canonical and deficiency modules can have large regularity. For a finitely generated graded R-module M, known bounds for $\operatorname{reg}(\operatorname{Ext}^i_R(M, \omega_R))$

MSC2000: 13D07, 13D45.
Keywords: canonical modules, Castelnuovo–Mumford regularity.

are large; see, for example, [Hoa and Hyry 2006, Theorems 9 and 14]. On the other hand, more optimal bounds for $\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R))$ are known to exist for certain classes of graded ideals I; see [Hoa and Hyry 2006, Section 4]. It is an interesting problem to find a class of graded ideals $I \subset R$ with optimal bounds for $\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R))$. In this paper, we focus on monomial ideals. It follows from the theory of square-free modules, introduced by K. Yanagawa [2000], that if I is a square-free monomial ideal, then $\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R)) \leq \dim \operatorname{Ext}^i_R(R/I, \omega_R)$. This bound is small, since $\dim \operatorname{Ext}^i_R(R/I, \omega_R) \leq n - i$; see [Bruns and Herzog 1993, Corollary 3.5.11].

While one cannot apply the theory of square-free modules to all monomial ideals, there are results that show that $\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R))$ is not large when I is a monomial ideal. For example, we see from [Takayama 2005, Proposition 1, page 333] that if $\operatorname{Ext}^i_R(R/I, \omega_R)$ has finite length, then its regularity is negative or equal to zero. Again, Hoa and Hyry [2006, Proposition 21] showed that if $\mathrm{H}^i_{\mathfrak{m}}(R/I)$ has finite length for $i = 0, 1, \dots, d-1$, where $d = \dim R/I$, then $\operatorname{reg}(\operatorname{Ext}^{n-d}_R(R/I, \omega_R)) \leq d$. We generalize these results in the next theorem:

Theorem 1.1. *Let $I \subseteq R$ be a monomial ideal. Then*

$$\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R)) \leq \dim \operatorname{Ext}^i_R(R/I, \omega_R) \quad \textit{for all } 0 \leq i \leq n.$$

Since $\dim \operatorname{Ext}^i_R(R/I, \omega_R) \leq n - i$, we immediately get this:

Corollary 1.2. *Let $I \subseteq R$ be a monomial ideal. Then*

$$\operatorname{reg}(\operatorname{Ext}^i_R(R/I, \omega_R)) \leq n - i \quad \textit{for all } 0 \leq i \leq n.$$

In general, this conclusion need not hold without the assumption that I is a monomial ideal; see [Chardin and D'Cruz 2003, Example 3.5].

Our approach to bounding the regularity of canonical and deficiency modules differs from that of Hoa and Hyry. We show that if I is a monomial ideal, then $\operatorname{Ext}^i_R(R/I, \omega_R)$ has a multigraded filtration, called the *Stanley filtration* and introduced by D. Maclagan and G. G. Smith [2005]; the bound on regularity follows from this filtration.

In the next section, we discuss some preliminaries on Stanley filtrations and local cohomology. In Section 3, we prove our main result.

2. Preliminaries

Hereafter we take R-modules to be graded by $\mathbb{Z}^n$, giving $\deg x_i = \boldsymbol{e}_i$, the i-th unit vector of $\mathbb{Z}^n$. We call this the *multigrading* of R and R-modules.

Notation 2.1. Let $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathbb{Z}^n$. Write

$$\boldsymbol{x}^{\boldsymbol{a}} = \prod_{i=1}^{n} x_i^{a_i} \in \Bbbk[x_1^{\pm 1}, \dots, x_n^{\pm 1}].$$

We say that $\boldsymbol{a}$ is the *degree* of $\boldsymbol{x}^{\boldsymbol{a}}$ and write $\deg \boldsymbol{x}^{\boldsymbol{a}} = \boldsymbol{a}$. Define $\operatorname{Supp}(\boldsymbol{a}) = \{i : a_i \neq 0\}$, and define $\boldsymbol{a}^+, \boldsymbol{a}^- \in \mathbb{N}^n$ by the conditions

$$\boldsymbol{a} = \boldsymbol{a}^+ - \boldsymbol{a}^- \quad \text{and} \quad \operatorname{Supp}(\boldsymbol{a}^+) \cap \operatorname{Supp}(\boldsymbol{a}^-) = \varnothing.$$

Write $\|\boldsymbol{a}\|$ for $\sum_{i=1}^{n} a_i$, the *total degree* of $\boldsymbol{a}$ (and of the monomial $\boldsymbol{x}^{\boldsymbol{a}}$). We will say that $\boldsymbol{a}$ (or equivalently $\boldsymbol{x}^{\boldsymbol{a}}$) is *square-free* if $a_i \in \{0, 1\}$ for all i. Let $[n] = \{1, \dots, n\}$. For $\Lambda \subseteq [n]$, we set $\boldsymbol{e}_\Lambda = \sum_{i \in \Lambda} \boldsymbol{e}_i$ and abbreviate the (square-free) monomial $\boldsymbol{x}^{\boldsymbol{e}_\Lambda}$ as x_Λ. The canonical module of R is $\omega_R = R(-\boldsymbol{e}_{[n]})$.

Let M be a finitely generated multigraded R-module. Let $m \in M$ be a homogeneous element, and let $G \subset \{x_1, \dots, x_n\}$ be a subset such that $um \neq 0$ for all monomials $u \in \Bbbk[G]$. The $\Bbbk$-subspace $\Bbbk[G]m$ of M generated by all the um, where u is a monomial in $\Bbbk[G]$, is called a *Stanley space*. A *Stanley decomposition* of M is a finite set $\mathcal{S}$ of pairs (m, G) of homogeneous elements $m \in M$ and $G \subseteq \{x_1, \dots, x_n\}$ such that $\Bbbk[G]m$ is a Stanley space for all $(m, G) \in \mathcal{S}$ and

$$M =_{\Bbbk} \bigoplus_{(m,G) \in \mathcal{S}} \Bbbk[G]m. \tag{1}$$

(We used "$=_{\Bbbk}$" to emphasize that the decomposition is only as vector spaces.) Properties of such decompositions have been widely studied; we follow the approach of [Maclagan and Smith 2005, Section 3], where Stanley decompositions were used to get bounds for multigraded regularity. Following [Maclagan and Smith 2005, Definition 3.7], we define a *Stanley filtration* to be a Stanley decomposition with an ordering of pairs $\{(m_i, G_i) : 1 \leq i \leq p\}$ such that

$$\Big(\sum_{i=1}^{j} R m_i\Big) \Big/ \Big(\sum_{i=1}^{j-1} R m_i\Big) = \Bbbk[G_j](-\deg m_j) \quad \text{for } j = 1, 2, \dots, p$$

as R-modules. Note, in this case, that

$$0 \subseteq R m_1 \subseteq \cdots \subseteq \sum_{i=1}^{j} R m_i \subseteq \cdots \subseteq \sum_{i=1}^{p} R m_i = M$$

is a prime filtration of M, as in [Eisenbud 1995, Proposition 3.7, page 93].

Proposition 2.2. *Let M be a multigraded R-module with a Stanley decomposition $\mathcal{S}$ such that* $(\deg m)^+$ *is square-free and* $G = \operatorname{Supp}((\deg m)^+)$ *for all* $(m, G) \in \mathcal{S}$. *Then, $\mathcal{S}$ gives a Stanley filtration. Moreover,* $\operatorname{reg} M \leq \max\{\|\deg m\| : (m, G) \in \mathcal{S}\}$.

Proof. We order $\mathcal{S} = \{(m_1, G_1), \dots, (m_p, G_p)\}$ so that $\|\deg m_1\| \geq \cdots \geq \|\deg m_p\|$. It follows from our hypothesis that

(2) $$\operatorname{span}_{\Bbbk}\{m_1, \dots, m_p\} = \operatorname{span}_{\Bbbk}\{m \in M : \operatorname{Supp}((\deg m)^+) \text{ is square-free}\},$$

where $\operatorname{span}_{\Bbbk}(V)$ denotes the $\Bbbk$-vector space spanned by the elements in V. We write $M^{(j)}$ for $\sum_{i=1}^{j} Rm_i$. We will now show, inductively on j, that

(a) $M^{(j-1)} :_R m_j = (x_k; x_k \notin G_j)$, and

(b) the set $\bigcup_{i=1}^{j}\{um_i : u \text{ is a monomial in } \Bbbk[G_i]\}$ is a $\Bbbk$-basis for $M^{(j)}$.

These imply that $\mathcal{S}$ is a Stanley filtration of M.

Let $j = 1$. We will show that $(0 :_R m_1) = (x_k; x_k \notin G_1)$. We have $um_1 \neq 0$ for all monomials $u \in \Bbbk[G_1]$ from the definition of the decomposition. Therefore we must show that $x_l m_1 = 0$ for any $x_l \notin G_1$. Let $x_l \notin G_1$. Then $(\deg x_l m_1)^+$ is square-free, and $x_l m_1 \in \operatorname{span}_{\Bbbk}\{m_1, \dots, m_p\}$ by (2). However, from the choice of m_1, we see that $x_l m_1 = 0$. Therefore $(0 :_R m_1) = (x_k; k \notin G_1)$, proving (a). Then (b) follows immediately.

Now, assume that $j > 1$ and that the assertion is known for all $i < j$. We first show (a). Let u be a monomial in $\Bbbk[G_j]$. By statement (b) for $j-1$, the set $\bigcup_{i=1}^{j-1}\{vm_i : v \text{ is a monomial in } \Bbbk[G_i]\}$ is a $\Bbbk$-basis for $M^{(j-1)}$. Since um_j is an element of the basis of M coming from the Stanley decomposition, um_j is not in the $\Bbbk$-linear span of $\bigcup_{i=1}^{j-1}\{vm_i : v \text{ is a monomial in } \Bbbk[G_i]\}$, that is, $um_j \notin M^{(j-1)}$. It remains to prove that $x_l m_j \in M^{(j-1)}$ for any $x_l \notin G_j$. Let $x_l \notin G_j$. Since $(\deg x_l m_j)^+$ is square-free, it follows from (2) and the ordering of the (m_i, G_i) that

$$x_l m_j \in \operatorname{span}_{\Bbbk}\{m_i : 1 \leq i \leq p,\ \deg m_i > \deg m_j\} \subseteq \operatorname{span}_{\Bbbk}\{m_1, \dots, m_{j-1}\}.$$

Therefore $x_l m_j \in M^{(j-1)}$, proving the statement (a) for j.

From (a), we see that the sequence

(3) $$0 \to M^{(j-1)} \to M^{(j)} \to \Bbbk[G_j]m_j \to 0$$

is exact. Now, statement (b) for j follows from the induction hypothesis.

Theorem 4.1 of [Maclagan and Smith 2005] essentially gives the assertion about regularity, but we give a quick proof here by showing that

$$\operatorname{reg} M^{(j)} \leq \max\{\|\deg m_i\| : 1 \leq i \leq j\} \quad \text{for all } 1 \leq j \leq p.$$

It holds for $j = 1$. For $j > 1$, it follows from [Eisenbud 1995, Corollary 20.19] and the exact sequence (3) that

$$\operatorname{reg} M^{(j)} \leq \max\{\operatorname{reg} M^{(j-1)}, \|\deg m_j\|\}.$$

Then induction completes the proof. ∎

Finally, we recall some basics of local cohomology. We follow [Bruns and Herzog 1993, Sections 3.5 and 3.6]. Let $\check{C}^\bullet$ be the Čech complex on $x_1, \dots, x_n$; the term at the i-th cohomological degree is

$$\check{C}^i = \bigoplus_{\Lambda \subseteq [n],\ |\Lambda| = i} R_{x_\Lambda},$$

where R_{x_Λ} denotes inverting the monomial x_Λ. Note that $\check{C}^\bullet$ is a complex of $\mathbb{Z}^n$-graded R-modules, with differentials of degree 0. For a finitely generated R-module M, we set $\check{C}^\bullet(M) = \check{C}^\bullet \otimes_R (M)$. Then $\mathrm{H}^i_\mathfrak{m}(M) = \mathrm{H}^i(\check{C}^\bullet(M))$.

Definition 2.3. Let $F \subseteq [n]$. We define $\check{C}^\bullet_F$ to be the subcomplex of $\check{C}^\bullet$ obtained by setting

$$\check{C}^i_F = \begin{cases} 0 & \text{if } i < |F|, \\ \bigoplus_{\substack{F \subseteq \Lambda \subseteq [n] \\ |\Lambda| = i}} R_{x_\Lambda} & \text{otherwise.} \end{cases}$$

Lemma 2.4. *Let I be a monomial ideal and $F \subseteq [n]$. If $\boldsymbol{a} \in \mathbb{Z}^n$ is such that* $\operatorname{Supp}(\boldsymbol{a}^-) = F$, *then* $\mathrm{H}^i_\mathfrak{m}(R/I)_{\boldsymbol{a}} = \mathrm{H}^i(\check{C}^\bullet_F \otimes_R (R/I))_{\boldsymbol{a}}$.

Proof. The proof of [Takayama 2005, Theorem 1] uses this argument implicitly. Since $\mathrm{H}^i_\mathfrak{m}(R/I)_{\boldsymbol{a}} = \mathrm{H}^i((\check{C}^\bullet(R/I))_{\boldsymbol{a}}$, it suffices to show that

$$(\check{C}^\bullet(R/I))_{\boldsymbol{a}} = (\check{C}^\bullet_F \otimes_R (R/I))_{\boldsymbol{a}}.$$

This, in turn, stems from the fact that $\check{C}^j_F \otimes_R (R/I)$ consists precisely of the direct summands of $\check{C}^j(R/I)$ that are nonzero in multidegree $\boldsymbol{a}$ for all $1 \le j \le n$. □

3. Proof of the main theorem

Lemma 3.1. *Let $I \subset R$ be a monomial ideal. Let $\boldsymbol{a} \in \mathbb{Z}^n$ and $j \in$* $\operatorname{Supp}(\boldsymbol{a}^+)$. *The multiplication map*

$$x_j : \operatorname{Ext}^i_R(R/I, \omega_R)_{\boldsymbol{a}} \to \operatorname{Ext}^i_R(R/I, \omega_R)_{\boldsymbol{a}+\boldsymbol{e}_j}$$

is bijective.

Proof. We first claim that the multiplication map

$$x_j : \mathrm{H}^{n-i}_\mathfrak{m}(R/I)_{-\boldsymbol{a}-\boldsymbol{e}_j} \to \mathrm{H}^{n-i}_\mathfrak{m}(R/I)_{-\boldsymbol{a}}$$

is bijective. By local duality [Bruns and Herzog 1993, Theorem 3.6.19], this map is the Matlis dual of the multiplication by x_j on $\operatorname{Ext}^i_R(R/I, \omega_R)_{\boldsymbol{a}}$; hence, it suffices to prove the claim above.

Set $F = \operatorname{Supp}(\boldsymbol{a}^+)$. Note that $\operatorname{Supp}(\boldsymbol{a}^+ + \boldsymbol{e}_j) = F$. For all i, the map x_j acts as a unit on $\check{C}^i_F$. Therefore the homomorphism of complexes

$$\check{C}^\bullet_F \otimes_R (R/I) \to \check{C}^\bullet_F \otimes_R (R/I)$$

induced by the multiplication map $x_j : \check{C}_F^i \otimes_R (R/I) \to \check{C}_F^i \otimes_R (R/I)$ is an isomorphism. The claim now follows from Lemma 2.4, which implies that

$$\begin{aligned} \mathrm{H}^i_{\mathfrak{m}}(R/I)_{-\boldsymbol{a}-\boldsymbol{e}_j} &= \mathrm{H}^i(\check{C}_F^\bullet \otimes_R (R/I))_{-\boldsymbol{a}-\boldsymbol{e}_j}, \\ \mathrm{H}^i_{\mathfrak{m}}(R/I)_{-\boldsymbol{a}} &= \mathrm{H}^i(\check{C}_F^\bullet \otimes_R (R/I))_{-\boldsymbol{a}}. \end{aligned} \qquad \square$$

The previous lemma says that, if I is a monomial ideal, then $\mathrm{Ext}^i_R(R/I, \omega_R)$ is a $(1, 1, \ldots, 1)$-determined module in the sense of [Miller 2000, Definition 2.1].

Proof of Theorem 1.1. For $F \subseteq [n]$, let $\mathcal{M}^i_F$ be a multigraded $\Bbbk$-basis for

$$\bigoplus_{\boldsymbol{a}\in\mathbb{N}^n,\ \mathrm{Supp}(\boldsymbol{a})\cap F=\varnothing} \mathrm{Ext}^i_R(R/I, \omega_R)_{\boldsymbol{e}_F-\boldsymbol{a}}.$$

Let $\mathcal{S}_i = \{(m, F) : F \subseteq [n] \text{ and } m \in \mathcal{M}^i_F\}$. It follows from Lemma 3.1 that $\mathcal{S}_i$ is a Stanley decomposition of $\mathrm{Ext}^i_R(R/I, \omega_R)$. In particular,

$$\dim \mathrm{Ext}^i(R/I, \omega_R) = \max\{|F| : \mathcal{M}^i_F \neq \varnothing\}.$$

By the construction of $\mathcal{M}^i_F$, this Stanley decomposition satisfies the assumption of Proposition 2.2. Therefore

$$\begin{aligned} \mathrm{reg}(\mathrm{Ext}^i_R(R/I, \omega_R)) &\leq \max_{F\subseteq[n]}\{\max\{\|\deg m\| : m \in \mathcal{M}^i_F\}\} \\ &\leq \max_{F\subseteq[n]}\{|F| : \mathcal{M}^i_F \neq \varnothing\} \\ &= \dim \mathrm{Ext}^i_R(R/I, \omega_R), \end{aligned}$$

as desired. (The second inequality follows since $\|\deg u\| = |F| - \|(\deg u)^-\|$ for any $u \in \mathcal{M}^i_F$.) $\square$

We remark that, by using [Takayama 2005, Theorem 1] and local duality, one can determine whether $\mathcal{M}^i_F \neq \varnothing$ from certain subcomplexes of the Stanley–Reisner complex of the radical $\sqrt{I}$ of I.

Acknowledgments.

The authors thank B. Ulrich for helpful comments. This paper was written when Murai was visiting Purdue University in September 2009. He would like to thank his host, G. Caviglia, for his hospitality.

References

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR 95h:13020 Zbl 0788.13005

[Chardin and D'Cruz 2003] M. Chardin and C. D'Cruz, "Castelnuovo–Mumford regularity: Examples of curves and surfaces", *J. Algebra* **270**:1 (2003), 347–360. MR 2004m:13036 Zbl 1056.14065

[Chardin et al. 2009] M. Chardin, D. T. Ha, and L. T. Hoa, "Castelnuovo–Mumford regularity of Ext modules and homological degree", preprint, 2009. arXiv 0903.4535

[Eisenbud 1995] D. Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR 97a:13001 Zbl 0819.13001

[Hoa and Hyry 2006] L. T. Hoa and E. Hyry, "Castelnuovo–Mumford regularity of canonical and deficiency modules", *J. Algebra* **305**:2 (2006), 877–900. MR 2007g:13023 Zbl 1108.13017

[Maclagan and Smith 2005] D. Maclagan and G. G. Smith, "Uniform bounds on multigraded regularity", *J. Algebraic Geom.* **14**:1 (2005), 137–164. MR 2005g:14098 Zbl 1070.14006

[Miller 2000] E. Miller, "The Alexander duality functors and local duality with monomial support", *J. Algebra* **231**:1 (2000), 180–234. MR 2001k:13028 Zbl 0968.13009

[Takayama 2005] Y. Takayama, "Combinatorial characterizations of generalized Cohen–Macaulay monomial ideals", *Bull. Math. Soc. Sci. Math. Roumanie* (*N.S.*) **48**:3 (2005), 327–344. MR 2006e:13017 Zbl 1092.13020

[Yanagawa 2000] K. Yanagawa, "Alexander duality for Stanley–Reisner rings and squarefree $\mathbb{N}^n$-graded modules", *J. Algebra* **225**:2 (2000), 630–645. MR 2000m:13036 Zbl 0981.13011

Received October 11, 2009. Revised June 7, 2010.

MANOJ KUMMINI
PURDUE UNIVERSITY
DEPARTMENT OF MATHEMATICS
150 N. UNIVERSITY ST.
WEST LAFAYETTE, IN 47907
UNITED STATES

nkummini@math.purdue.edu

SATOSHI MURAI
YAMAGUCHI UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCE
YAMAGUCHI 753-8512
JAPAN

murai@yamaguchi-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

$\mathrm{SL}_2(\mathbb{C})$-CHARACTER VARIETY OF A HYPERBOLIC LINK AND REGULATOR

WEIPING LI AND QINGXUE WANG

We analyze a special smooth projective variety Y^h arising from some one-dimensional irreducible slices on the $\mathrm{SL}_2(\mathbb{C})$-character variety of a hyperbolic link in S^3. We prove that a natural symbol obtained from these one-dimensional slices is a torsion in $K_2(\mathbb{C}(Y^h))$. By using the regulator map from K_2 to the corresponding Deligne cohomology, we get some variation formulas on some Zariski open subset of Y^h. From this we discuss a possible parametrized volume conjecture for both hyperbolic links and knots.

1. Introduction

This is the sequel to [Li and Wang 2008] on the generalized volume conjecture for a hyperbolic knot in S^3. In this paper, we shall study a hyperbolic link in S^3 and extend several results from the knot case. The main idea is to apply the regulator map in K-theory to the $\mathrm{SL}_2(\mathbb{C})$-character varieties of hyperbolic links.

For a link L in S^3, Kashaev [1995] introduced a sequence of complex numbers $\{K_N \mid N \text{ is an odd integer} > 1\}$, which were derived from a matrix version of the quantum dilogarithms. Kashaev's volume conjecture therein predicts that for any hyperbolic link L in S^3, the asymptotic behavior of his invariants $\{K_N\}$ regains the hyperbolic volume of $S^3 \setminus L$. Kashaev verified this for the figure eight knot. The volume conjecture provides an intriguing relationship between the quantum invariants and the hyperbolic volume, but we still do not fully understand it.

For the knot case, Murakami and Murakami [2001] showed that the Kashaev invariants $\{K_N\}$ can be identified with the values of the normalized colored Jones polynomial at the primitive N-th roots of unity. From this, they formulated a new version of volume conjecture, stating that the asymptotic behavior of the colored Jones invariants of any knot equals the Gromov simplicial volume of its complement in S^3. This version of the volume conjecture bridges the quantum invariants of the knot with its classical geometry and topology. However, this formulation does

---

Wang is supported by NSFC, grant number 10801034.
MSC2000: primary 57M25, 57M27; secondary 19F15, 14H50.
Keywords: Chern–Simons invariant, character variety, algebraic K-theory, hyperbolic links, volume conjecture, regulator map.

not fit well for links since it does not hold for many split links; see [Murakami et al. 2002]. Hence it is a very interesting question to see what is really behind the volume conjecture for links.

Following Witten's SU(2) topological quantum field theory, Gukov [2005] proposed a complex version of Chern–Simons theory and generalized the volume conjecture to a $\mathbb{C}^*$-parametrized version with parameter lying on the zero locus of the A-polynomial of the knot. In [Li and Wang 2008], we constructed a natural torsion element in K_2 of the function field of the curve defined by the A-polynomial. We then showed that the part from the A-polynomial in Gukov's generalized volume conjecture can be interpreted in terms of the regulator map on this torsion element. In particular, this implied the Bohr–Sommerfeld quantization condition posed by Gukov [2005, page 597].

It is natural to ask if there exists a parametrized volume conjecture for links in S^3, as Gukov showed for the knot case. This is the motivation of this paper. Now we have to deal with two problems for links with more than one component. First, its $\mathrm{SL}_2(\mathbb{C})$-character variety has dimension greater than one, and it is not clear how to define an A-polynomial for such a link that will contain geometric information like volume and Chern–Simons as in the knot case. Second, it is not clear how to relate the colored Jones polynomial to its $\mathrm{SL}_2(\mathbb{C})$-character variety. In this paper, we shall focus on the first problem for hyperbolic links. We introduce n curves on the geometric component of the character variety. From these curves, we obtain an n-dimensional smooth projective variety Y^h, where n is the number of the components of the link. We construct a natural torsion element in K_2 of the function field of Y^h. By applying the regulator map on this torsion element, we get the variation formulas (Theorem 3.13) on some Zariski open subset of Y^h. When the link has one component, we recover the results for hyperbolic knots. This suggests that there may exist a parametrized volume conjecture for hyperbolic links and the Y^h may provide a replacement for the zero locus of the A-polynomial of a knot. We do not know how to deal with the second problem, and only give some speculations at the end of Section 4.

On the other hand, Dupont [1987] used the dilogarithm to construct explicitly the Cheeger–Chern–Simons class associated to the second Chern polynomial. This result applied to a closed hyperbolic 3-manifold M gives a number in $\mathbb{C}/\mathbb{Z}$. Dupont also showed that the imaginary part of this number is the hyperbolic volume of M, while the real part is the Chern–Simons invariant of M. In general, for an odd-dimensional hyperbolic manifold of finite volume, Goncharov [1999] constructed an element in Quillen's algebraic K-group of $\mathbb{C}$ and proved that after applying the Borel regulator, we get the volume of the manifold. Here, we use the regulator map for the function field of Y^h; it can be regarded as an analogue of a family version of Dupont and Goncharov's for the $\mathrm{SL}_2(\mathbb{C})$-character variety of a hyperbolic link.

The paper is organized as follows. In Section 2, we review the basics of the $\mathrm{SL}_2(\mathbb{C})$-character variety of a hyperbolic link. We then study the properties of a smooth projective variety Y^h coming from the one-dimensional slices of the character variety. In Section 3, we recall the definitions and basic properties of K_2 of a commutative ring. We then state and prove our main results. In Section 4, we discuss a parametrized volume conjecture for hyperbolic links.

2. Character variety of a hyperbolic link

2a. Let L be a hyperbolic link in S^3 with n components $K_1, \ldots, K_n$. This means that the complement $S^3 \setminus L$ carries a complete hyperbolic structure of finite volume. Let $N(L)$ be an open tubular neighborhood of L in S^3. Then $M_L = S^3 \setminus N(L)$ is a compact 3-manifold with boundary ∂M_L a disjoint union of n tori $T_1, \ldots, T_n$, and is called the link exterior. Note that $\pi_1(S^3 \setminus L)$ and $\pi_1(M_L)$ are isomorphic. In the following, we shall identify them.

Let $R(M_L) = \mathrm{Hom}(\pi_1(M_L), \mathrm{SL}_2(\mathbb{C}))$ and $R(T_i) = \mathrm{Hom}(\pi_1(T_i), \mathrm{SL}_2(\mathbb{C}))$ for $i = 1, \ldots, n$ be the $\mathrm{SL}_2(\mathbb{C})$-representation spaces. We have the natural action of $\mathrm{SL}_2(\mathbb{C})$ on them by conjugation. According to [Culler and Shalen 1983], they are affine algebraic sets and so are the corresponding character varieties $X(M_L)$ and $X(T_i)$, which are the algebro-geometric quotients of $R(M_L)$ and $R(T_i)$ by $\mathrm{SL}_2(\mathbb{C})$. We then have the canonical surjective morphisms $t : R(M_L) \to X(M_L)$ and $t_i : R(T_i) \to X(T_i)$ that map a representation to its character. The inclusions of $\pi_1(T_i)$ into $\pi_1(M_L)$ induce the restriction map

$$r : X(M_L) \to X(T_1) \times \cdots \times X(T_n).$$

For details on character varieties, see [Culler and Shalen 1983; Culler et al. 1987; Cooper et al. 1994; Shalen 2002].

2b. Let $\rho_0 : \pi_1(M_L) \to \mathrm{SL}_2(\mathbb{C})$ be a representation associated to the complete hyperbolic structure on $S^3 \setminus L$. This representation is irreducible. Denote by χ_0 its character. Fix an irreducible component R_0 of $R(M_L)$ containing ρ_0. Then $X_0 = t(R_0)$ is an affine variety of dimension n [Culler and Shalen 1983; Shalen 2002]. We call X_0 a geometric component of the character variety. We define $Y_0 := \overline{r(X_0)}$, where the bar means the Zariski closure of the image $r(X_0)$ in $X(T_1) \times \cdots \times X(T_n)$.

For $g \in \pi_1(M_L)$, there is a regular function $I_g : X_0 \to \mathbb{C}$ defined by $I_g(\chi) = \chi(g)$ for all $\chi \in X_0$.

Proposition 2.1 [Culler and Shalen 1984, Proposition 2, page 539]. *Let γ_i be a noncontractible simple closed curve in the boundary torus T_i for $1 \le i \le n$. Let $g_i \in \pi_1(M_L)$ be an element whose conjugacy class corresponds to the free homotopy class of γ_i. Let k be an integer with $0 \le k \le n$, and let V be the algebraic subset of X_0 defined by the equations $I_{g_i}^2(\chi) = 4$, with $k < i \le n$. Let V_0 denote an* *irreducible component of V containing χ_{ρ_0}. If χ is a point of V_0, i is an integer with $k < i \le n$, and g is an element of the subgroup* $\mathrm{Im}(\pi_1(T_i) \to \pi_1(M_L))$ *(defined up to conjugacy), then we have $I_g(\chi) = \pm 2$. If also $k = 0$, then $V_0 = \{\chi_{\rho_0}\}$.*

The following generalizes the knot case; see [Culler and Shalen 1983; 1984].

Proposition 2.2. *Y_0 is an n-dimensional affine variety.*

Proof. It is clear that Y_0 is an affine variety. We need to show that $\dim Y_0 = n$. Since $\dim X_0 = n$, we have $\dim Y_0 \le n$. Assume that $\dim Y_0 = m < n$. Then for $y \in r(X_0)$, every component of the fiber $r^{-1}(y)$ has dimension $\ge n - m \ge 1$. Take $y = r(\chi_0)$; then there is an irreducible component C of the fiber $r^{-1}(y)$ containing χ_0 and $\dim C \ge 1$. For each boundary torus T_i and a nontrivial $g_i \in \mathrm{Im}(\pi_1(T_i) \to \pi_1(M_L))$, consider the regular function $I_{g_i} : X_0 \to \mathbb{C}$. For all $\chi \in C$, we have $I_{g_i}(\chi) = I_{g_i}(\chi_0)$. Since χ_0 is the character of the complete hyperbolic structure on M_L, we have $I^2_{g_i}(\chi) - 4 = I^2_{g_i}(\chi_0) - 4 = 0$ for all $\chi \in C$ and all $g_i \in \mathrm{Im}(\pi_1(T_i) \to \pi_1(M_L))$ with $1 \le i \le n$. Now we fix n nontrivial $g_i \in \mathrm{Im}(\pi_1(T_i) \to \pi_1(M_L))$ for $1 \le i \le n$. Consider the algebraic subset V of X_0 defined by the equations $I^2_{g_i} - 4 = 0$ for $1 \le i \le n$. By its construction, C is contained in an irreducible component V_0 of V containing χ_0. Hence $\dim V_0 \ge 1$. On the other hand, $V_0 = \{\chi_0\}$ by Proposition 2.1, a contradiction. Therefore, $\dim Y_0 = n$. □

For every boundary torus T_i, we fix a meridian-longitude basis $\{\mu_i, \lambda_i\}$ for $\pi_1(T_i) = H_1(T_i; \mathbb{Z})$. Given $1 \le i \le n$, we define X^i_0 as the subvariety of X_0 defined by the equations $I^2_{\mu_j} - 4 = 0$ for $j \ne i$ and $1 \le j \le n$. Let V_i be an irreducible component of X^i_0 containing χ_0.

Proposition 2.3. *V_i has dimension one for each $i = 1, \dots, n$.*

Proof. Since X^i_0 is defined by $n-1$ equations and $\dim X_0 = n$, every component of X^i_0 has dimension at least 1. Now assume that $\dim V_i \ge 2$. Let U be the subvariety of V_i defined by the equation $I^2_{\mu_i} - 4 = 0$, and let U_0 be the irreducible component of U containing χ_0. Then $\dim V_i \ge 2$ implies that $\dim U_0 \ge 1$. But this contradicts the last part of Proposition 2.1. Hence, $\dim V_i = 1$. □

Lemma 2.4. *Fix a nontrivial $g_i \in \mathrm{Im}(\pi_1(T_i) \to \pi_1(M_L))$, with $1 \le i \le n$.*

(1) *$I_{g_i} = \pm 2$ is a constant on every V_j with $j \ne i$.*

(2) *I_{g_i} is not a constant on V_i; hence it is not a constant on X_0 either.*

Proof. (1) follows from the definition of V_j and Proposition 2.1.

For (2), suppose I_{g_i} were a constant on V_i. Then $I_{g_i} = I_{g_i}(\chi_0) = \pm 2$. Consider the algebraic subset V of X_0 defined by the n equations $I^2_{\mu_j} = 4$ with $j \ne i$, and $I^2_{g_i} = 4$. Then V_i is contained in some irreducible component V_0 of V that contains χ_{ρ_0}. Hence $\dim V_0 \ge 1$, contradicting Proposition 2.1. □

For each $i = 1, \dots, n$, let p_i be the projection map from $X(T_1) \times \cdots \times X(T_n)$ to the i-th factor $X(T_i)$. Denote by $r_i : X_0 \to X(T_i)$ the composition of r and p_i.

Proposition 2.5. *For every $i = 1, \dots, n$, the Zariski closure W_i of the image $r_i(V_i)$ in $X(T_i)$ has dimension* 1.

Proof. It suffices to consider the case $i = 1$. Since $\dim V_1 = 1$ and r_1 is regular, $\dim W_1 \le 1$. Assume that $\dim W_1 = 0$. This means that $r_1(V_1)$ consists of a single point. Therefore, I_{g_1} is a constant on V_1 for any $g_1 \in \mathrm{Im}(\pi_1(T_1) \to \pi_1(M_L))$. This contradicts Lemma 2.4(2). □

2c. For $1 \le i \le n$, denote by $R_D(T_i)$ the subvariety of $R(T_i)$ that consists of the diagonal representations. For such a representation ρ, it is clear by taking the eigenvalues of $\rho(\mu_i)$ and $\rho(\lambda_i)$ that $R_D(T_i)$ is isomorphic to $\mathbb{C}^* \times \mathbb{C}^*$. We denote the coordinates by (l_i, m_i). Let $t_{i|D}$ be the restriction of t_i to $R_D(T_i) = \mathbb{C}^* \times \mathbb{C}^*$. Set $D_i = t_{i|D}^{-1}(W_i)$. By the proof of [Li and Wang 2006, Proposition 3.3], D_i is either irreducible or has two isomorphic irreducible components. Let $y^i \in D_i$ be the point corresponding to the character of the representation of the hyperbolic structure on $S^3 \setminus L$. Let Y_i be an irreducible component of D_i containing y^i. Then Y_i is an algebraic curve. Denote by $\overline{Y}_i$ the smooth projective model of Y_i. Denote by $\mathbb{C}(\overline{Y}_i)$ the function field of $\overline{Y}_i$ that is isomorphic to the function field $\mathbb{C}(Y_i)$ of Y_i. Note that when L is a hyperbolic knot ($n = 1$), Y_1 is the locus of the factor of the A-polynomial corresponding to the geometric component.

We define $Y^h = \overline{Y}_1 \times \overline{Y}_2 \times \cdots \times \overline{Y}_n$. Note that Y^h is an n-dimensional smooth projective variety. Let $\mathbb{C}(Y^h)$ be the function field of Y^h. For each i, we have the injective morphism $j_i : \mathbb{C}(Y_i) = \mathbb{C}(\overline{Y}_i) \to \mathbb{C}(Y^h)$ that is induced by the i-th projection from Y^h to $\overline{Y}_i$. In this way we take the $\mathbb{C}(Y_i)$ as subfields of $\mathbb{C}(Y^h)$. This also induces the map j on the K-groups:

$$j : \bigoplus_{i=1}^{n} K_2(\mathbb{C}(Y_i)) \to K_2(\mathbb{C}(Y^h)).$$

For $f_i, g_i \in \mathbb{C}(Y_i)$ with $i = 1, \dots, n$, we have $j(\sum_{i=1}^{n}\{f_i, g_i\}) = \prod_{i=1}^{n}\{f_i, g_i\}$, where we identify f_i and g_i as rational functions on Y^h via the injection j_i. Note that in this paper we use the multiplication in K_2 instead of addition.

Proposition 2.6. *There exists a finite field extension F of $\mathbb{C}(Y^h)$ with the property that for every $i = 1, \dots, n$, there is a representation $P_i : \pi_1(M_L) \to \mathrm{SL}_2(F)$ such that for $1 \le j \le n$, if $j \ne i$, the traces of $P_i(\lambda_j)$ and $P_i(\mu_j)$ are either 2 or -2. If $j = i$, then*

$$P_i(\lambda_i) = \begin{bmatrix} l_i & 0 \\ 0 & l_i^{-1} \end{bmatrix} \quad \textit{and} \quad P_i(\mu_i) = \begin{bmatrix} m_i & 0 \\ 0 & m_i^{-1} \end{bmatrix}.$$

Proof. By definition, W_i for each i is the Zariski closure of $r_i(V_i)$ in $X(T_i)$ and Y_i is mapped dominatingly to W_i. The canonical morphism $t : R_0 \to X_0$ is surjective, so we can choose a curve $E_i \subset R_0$ such that $t(E_i)$ is dense in V_i. Hence the composition $r_i \circ t : E_i \to W_i$ is dominating. Then the function fields $\mathbb{C}(E_i)$ and $\mathbb{C}(Y_i)$ are finite extensions of $\mathbb{C}(W_i)$. By [Culler and Shalen 1983, page 115], there is a tautological representation $p_i : \pi_1(M_L) \to \mathrm{SL}_2(\mathbb{C}(E_i))$, and the trace of $p_i(g)$ equals I_g for any $g \in \pi_1(M_L)$. The composite field F_i of $\mathbb{C}(E_i)$ and $\mathbb{C}(Y_i)$ is finite over both $\mathbb{C}(E_i)$ and $\mathbb{C}(Y_i)$. We shall view p_i as a representation in $\mathrm{SL}_2(F_i)$. Since $t(E_i)$ is dense in V_i, by Lemma 2.4 we have that the traces of $p_i(\lambda_j)$ and $p_i(\mu_j)$ are ± 2 if $j \neq i$, and the traces of $p_i(\lambda_i)$ and $p_i(\mu_i)$ are nonconstant functions on E_i if $j = i$. Since $p_i(\lambda_i)$ and $p_i(\mu_i)$ are commuting and their eigenvalues l_i and m_i are in F_i, the representation p_i is conjugate in $\mathrm{GL}_2(F_i)$ to a representation

$$P_i : \pi_1(M_L) \to \mathrm{SL}_2(F_i)$$

such that if $j \neq i$, the traces of $P_i(\lambda_j)$ and $P_i(\mu_j)$ are either 2 or -2. If $j = i$, then

$$P_i(\lambda_i) = \begin{bmatrix} l_i & 0 \\ 0 & l_i^{-1} \end{bmatrix} \quad \text{and} \quad P_i(\mu_i) = \begin{bmatrix} m_i & 0 \\ 0 & m_i^{-1} \end{bmatrix}.$$

Fix an algebraic closure $\overline{\mathbb{C}(Y^h)}$ of $\mathbb{C}(Y^h)$. As above, by viewing $\mathbb{C}(Y_i)$ as a subfield of $\overline{\mathbb{C}(Y^h)}$, we can identify the finite field extension F_i as a subfield of $\overline{\mathbb{C}(Y^h)}$. In $\overline{\mathbb{C}(Y^h)}$, take the composition K_i of F_i and $\mathbb{C}(Y^h)$ over $\mathbb{C}(Y_i)$. Then $F_i \subset K_i$, and K_i is a finite extension of $\mathbb{C}(Y^h)$ because the extension $F_i/\mathbb{C}(Y_i)$ is finite. Now let F be the composition of the fields $K_1, \ldots, K_n$ in $\overline{\mathbb{C}(Y^h)}$. Then F is a finite extension of $\mathbb{C}(Y^h)$ since each K_i is. Now compose each P_i with the embedding $\mathrm{SL}_2(F_i) \hookrightarrow \mathrm{SL}_2(F)$; the proposition follows. □

3. K-theory and Deligne cohomology

First we recall the definitions of K_2 of a commutative ring A; see [Milnor 1971]. Let $\mathrm{GL}(A)$ be the direct limit of the groups $\mathrm{GL}_n(A)$, and let $E(A)$ be the direct limit of the groups $E_n(A)$ generated by all $n \times n$ elementary matrices.

Definition 3.1. For $n \geq 3$, the *Steinberg group* $\mathrm{St}(n, A)$ is the group defined by generators x_{ij}^{λ} for $1 \leq i \neq j \leq n$, with $\lambda \in A$, subject to the relations

(i) $x_{ij}^{\lambda} \cdot x_{ij}^{\mu} = x_{ij}^{\lambda+\mu}$,

(ii) $[x_{ij}^{\lambda}, x_{jl}^{\mu}] = x_{il}^{\lambda\mu}$ for $i \neq l$, and

(iii) $[x_{ij}^{\lambda}, x_{kl}^{\mu}] = 1$ for $j \neq k$ and $i \neq l$.

We have the canonical homomorphism $\phi_n : \mathrm{St}(n, A) \to \mathrm{GL}_n(A)$ by $\phi(x_{ij}^{\lambda}) = e_{ij}^{\lambda}$, where $e_{ij}^{\lambda} \in \mathrm{GL}_n(A)$ is the elementary matrix with entry λ in the (i, j) place. Taking the direct limit as $n \to \infty$, we get $\phi : \mathrm{St}(A) \to \mathrm{GL}(A)$. Its image $\phi(\mathrm{St}(A))$ is equal to $E(A)$, the commutator subgroup of $\mathrm{GL}(A)$.

Definition 3.2. $K_2(A) = \mathrm{Ker}\,\phi$.

It is well known that $K_2(A)$ is the center of the Steinberg group $\mathrm{St}(A)$ and there is a canonical isomorphism $\alpha : H_2(E(A); \mathbb{Z}) \to K_2(A)$; see [Milnor 1971, Theorems 5.1 and 5.10], respectively.

3a. ***The symbol.*** Let U and V be two commuting elements of $E(A)$. Choose $u, v \in \mathrm{St}(A)$ such that $U = \phi(u)$ and $V = \phi(v)$. Then the commutator $[u, v] = uvu^{-1}v^{-1}$ is in the kernel of ϕ. Hence $[u, v] \in K_2(A)$. We can check that $[u, v]$ is independent of the choices of u and v, and we denote it by $U \star V$.

Lemma 3.3. (1) *The product is skew-symmetric*: $U \star V = (V \star U)^{-1}$.

(2) *It is bimultiplicative*: $(U_1 \cdot U_2) \star V = (U_1 \star V) \cdot (U_2 \star V)$.

(3) *It is conjugation invariant*: $(PUP^{-1}) \star (PVP^{-1}) = U \star V$ *for* $P \in \mathrm{GL}(A)$.

Proof. This is [Milnor 1971, Lemma 8.1]. For (3), we remark that since $E(A)$ is a normal subgroup of $\mathrm{GL}(A)$, the left side of the formula makes sense. If P, U and V are in $\mathrm{GL}(n, A)$, then choose $p \in \mathrm{St}(A)$ such that

$$\phi(p) = \begin{bmatrix} P & 0 \\ 0 & P^{-1} \end{bmatrix} \in E(A).$$

Now we have $\phi(pup^{-1}) = PUP^{-1}$ and $\phi(pvp^{-1}) = PVP^{-1}$. Hence

$$[pup^{-1}, pvp^{-1}] = p[u, v]p^{-1} = [u, v].$$ □

Given two units f and g of A, consider the matrices

$$D_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D'_g = \begin{bmatrix} g & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & g^{-1} \end{bmatrix}.$$

They are in $E(A)$ and commute. Define the symbol $\{f, g\} := D_f \star D'_g$.

Lemma 3.4 [Milnor 1971, Lemmas 8.2 and 8.3]. (1) *The symbol* $\{f, g\}$ *is skew-symmetric*: $\{f, g\} = \{g, f\}^{-1}$.

(2) *It is bimultiplicative*: $\{f_1 f_2, g\} = \{f_1, g\}\{f_2, g\}$.

(3) *Denote by* $\mathrm{diag}(f_1, \dots, f_n)$ *a diagonal matrix with diagonal entries the* f_i. *If* $f_1 \cdots f_n = g_1 \cdots g_n = 1$, *then*

$$\mathrm{diag}(f_1, \dots, f_n) \star \mathrm{diag}(g_1, \dots, g_n) = \{f_1, g_1\}\{f_2, g_2\} \cdots \{f_n, g_n\}.$$

where the right side means the product of the symbols $\{f_i, g_i\}$ *for* $1 \le i \le n$.

Let F be a field. Let $\mathrm{SL}(F)$ be the direct limit of the groups $\mathrm{SL}_n(F)$. We know that $\mathrm{SL}(F) = E(F)$ and any element of $\mathrm{SL}_n(F)$ is also naturally an element of $E(F)$.

Lemma 3.5. *Let $u, t \in F$.*

(1) $\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \star \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix} = 1.$

(2) $\begin{bmatrix} -1 & t \\ 0 & -1 \end{bmatrix} \star \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \star \begin{bmatrix} -1 & u \\ 0 & -1 \end{bmatrix}$ *and* $\begin{bmatrix} -1 & t \\ 0 & -1 \end{bmatrix} \star \begin{bmatrix} -1 & u \\ 0 & -1 \end{bmatrix}$ *are 2-torsion in $K_2(F)$.*

(3) *If U and V are two commuting matrices in $\mathrm{SL}_2(F)$ and their traces are 2 or -2, then $U \star V$ is 2-torsion in $K_2(F)$. In particular, if both have trace 2, then $U \star V = 1$.*

Proof. For $s \in F$, let

$$M(1, s) = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad M(-1, s) = \begin{bmatrix} -1 & s \\ 0 & -1 \end{bmatrix}.$$

In particular, $M(1, 0)$ is the 2×2 identity matrix and $M(-1, 0)$ is the 2×2 diagonal matrix with diagonal entries -1.

For (1), $M(1, t) \star M(1, u) = [x_{12}^t, x_{12}^u] = 1$ by the definition of $\mathrm{St}(A)$.

For (2), notice that by the definition, $M(1, 0) \star A = 1$ and $A \star A = 1$ for any $A \in E(F)$. By Lemma 3.3,

$$1 = (M(-1, 0) \cdot M(-1, 0)) \star M(1, s) = (M(-1, 0) \star M(1, s))^2,$$

so $M(-1, 0) \star M(1, s)$ is a 2-torsion in $K_2(F)$. Since

$$M(-1, t) = M(-1, 0) \cdot M(1, -t) \quad \text{and} \quad M(-1, u) = M(-1, 0) \cdot M(1, -u),$$

by Lemma 3.3 and the first part, we have

$$\begin{aligned} M(-1, t) \star M(1, u) &= (M(-1, 0) \star M(1, u))(M(1, -t) \star M(1, u)) \\ &= M(-1, 0) \star M(1, u), \\ M(-1, t) \star M(-1, u) &= (M(-1, 0) \star M(1, -u))(M(1, -t) \star M(-1, 0)); \end{aligned}$$

hence they are 2-torsion.

For (3), we can find $P \in \mathrm{GL}_2(F)$ such that

$$PUP^{-1} = \begin{bmatrix} \pm 1 & t \\ 0 & \pm 1 \end{bmatrix} \quad \text{and} \quad PVP^{-1} = \begin{bmatrix} \pm 1 & u \\ 0 & \pm 1 \end{bmatrix}.$$

Then it follows from the first two parts and Lemma 3.3(3). □

The following proposition slightly generalizes [Cooper et al. 1994, Lemma 4.1]. The proof is the same.

Proposition 3.6. *Let π be a free abelian group of rank two with $\{e_1, e_2\}$ its basis. Let $f : \pi \to E(A)$ be a group homomorphism defined by $f(e_1) = U$ and $f(e_2) = V$. Then there is a generator t of $H_2(\pi; \mathbb{Z})$ such that $\alpha(f_*(t)) = U \star V$. Here $\alpha : H_2(E(A); \mathbb{Z}) \to K_2(A)$ is the canonical isomorphism and $f_* : H_2(\pi; \mathbb{Z}) \to H_2(E(A); \mathbb{Z})$ is the homomorphism induced by f.*

Proof. Since π is abelian, U and V commute. $U \star V$ is well-defined. Let F be the free group on $\{e_1, e_2\}$. The homomorphism f gives rise to a commutative diagram of short exact sequences of groups:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & [F, F] & \longrightarrow & F & \longrightarrow & \pi & \longrightarrow & 0 \\
\downarrow & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f} & & \downarrow \\
0 & \longrightarrow & K_2(A) & \longrightarrow & \mathrm{St}(A) & \xrightarrow{\phi} & E(A) & \longrightarrow & 0,
\end{array}$$

where $f_2([e_1, e_2]) = U \star V$. Applying the homology spectral sequence to this diagram, we obtain the diagram

$$\begin{array}{ccc}
H_2(\pi; \mathbb{Z}) & \longrightarrow & H_0(\pi; H_1([F, F]; \mathbb{Z})) \\
\downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle g} \\
H_2(E(A); \mathbb{Z}) & \xrightarrow{\alpha} & K_2(A).
\end{array}$$

The top horizontal arrow is an isomorphism. The class of $[e_1, e_2]$ is the generator of $H_0(\pi; H_1([F, F]; \mathbb{Z}))$. It is mapped to $U \star V$ by g, which is induced by f_2. Let t be the generator of $H_2(\pi; \mathbb{Z})$ mapped to the class of $[e_1, e_2]$. Then we have $\alpha(f_*(t)) = U \star V$ by the commutative diagram. □

Corollary 3.7. (1) *If $U = \mathrm{diag}(u, u^{-1})$ and $V = \mathrm{diag}(v, v^{-1})$, where u, v are units of A, then there is a generator t of $H_2(\pi; \mathbb{Z})$ such that $\alpha(f_*(t)) = \{u, v\}^2$.*

(2) *Suppose A is a field. If U and V are two commuting matrices in $SL_2(A)$ and their traces are 2 or -2, then the image of any generator of $H_2(\pi; \mathbb{Z})$ is 2-torsion in $K_2(A)$.*

Proof. For (1), we have $U \star V = \{u, v\}\{u^{-1}, v^{-1}\} = \{u, v\}^2$ by Lemma 3.4.

For (2), $U \star V$ is 2-torsion in $K_2(F)$ by Lemma 3.5(3). □

Theorem 3.8. *For each $i = 1, \ldots, n$, there is an integer $\epsilon(i) = 1$ or -1 such that the symbol $\prod_{i=1}^{n} \{l_i, m_i\}^{\epsilon(i)}$ is a torsion element in $K_2(\mathbb{C}(Y^h))$.*

Proof. First, by Proposition 2.6, for each $i = 1, \ldots, n$ there exist a finite extension F of $\mathbb{C}(Y^h)$ and a representation $P_i : \pi_1(M_L) \to SL_2(F)$ such that for $1 \le j \le n$, the traces of $P_i(\lambda_j)$ and $P_i(\mu_j)$ are either 2 or -2 if $j \neq i$ and, if $j = i$,

$$P_i(\lambda_i) = \begin{bmatrix} l_i & 0 \\ 0 & l_i^{-1} \end{bmatrix} \quad \text{and} \quad P_i(\mu_i) = \begin{bmatrix} m_i & 0 \\ 0 & m_i^{-1} \end{bmatrix}.$$

The inclusions of $\pi_1(T_i)$ into $\pi_1(M_L)$ induce homomorphisms $\pi_1(T_i) \to E(F)$ by composition with P_i. This gives rise to homomorphisms

$$\bigoplus_{i=1}^{n} H_2(\pi_1(T_i); \mathbb{Z}) \xrightarrow{\alpha} H_2(\pi_1(M_L); \mathbb{Z}) \xrightarrow{\beta} H_2(E(F); \mathbb{Z}) = K_2(F) \tag{3-1}$$

in group homology, where $\alpha = j_{1*} + \cdots + j_{n*}$, $\beta = P_{1*} + \cdots + P_{n*}$, the j_{i*} are the morphisms on the group homology induced by the inclusions $j_i : \pi_1(T_i) \hookrightarrow \pi_1(M_L)$, and the P_{i*} are those induced by the P_i.

The orientation of M_L induces an orientation on each boundary torus T_i. Let $[T_i]$ be the orientation class of $H_2(T_i; \mathbb{Z}) = \mathbb{Z}$. By Corollary 3.7(1), for each i there is a generator ξ_i of $H_2(\pi_1(T_i))$ such that $P_{i*}(j_{i*}(\xi_i)) = \{l_i, m_i\}^2$. Since T_i is a $K(\pi_1(T_i), 1)$ space, $H_2(\pi_1(T_i); \mathbb{Z}) = H_2(T_i; \mathbb{Z})$. If $\xi_i = [T_i]$, define $\epsilon(i) = 1$; if $\xi_i = -[T_i]$, then define $\epsilon(i) = -1$.

Since L is a hyperbolic link, M_L is a $K(\pi_1(M_L), 1)$ space. Hence we have $H_2(\pi_1(M_L); \mathbb{Z}) = H_2(M_L; \mathbb{Z})$. Under this identification, we have

$$\alpha(\epsilon(1)\xi_1, \dots, \epsilon(n)\xi_n) = \sum_{i=1}^{n} [T_i] = [\partial M_L] = 0 \quad \text{in } H_2(M_L; \mathbb{Z}).$$

Therefore,

$$\beta(\alpha(\epsilon(1)\xi_1, \dots, \epsilon(n)\xi_n)) = 1 \quad \text{in } K_2(F). \tag{3-2}$$

On the other hand, we have

$$\begin{aligned} \beta(\alpha(\epsilon(1)\xi_1, \dots, \epsilon(n)\xi_n)) &= \beta\Big(\sum_{i=1}^{n} j_{i*}(\epsilon(i)\xi_i)\Big) \\ &= \sum_{k=1}^{n} P_{k*}\Big(\sum_{i=1}^{n} j_{i*}(\epsilon(i)\xi_i)\Big) \\ &= \sum_{i=1}^{n} P_{i*}(j_{i*}(\epsilon(i)\xi_i)) + \sum_{1 \le i \neq k \le n} P_{k*}(j_{i*}(\epsilon(i)\xi_i)) \\ &= \prod_{i=1}^{n} \{l_i, m_i\}^{2\epsilon(i)} \cdot \prod_{1 \le i \neq k \le n} P_k(\mu_i) \star P_k(\lambda_i), \end{aligned}$$

where the last step follows from Proposition 3.6 and Corollary 3.7. Note also that we use multiplication in $K_2(F)$.

Now $\prod_{1\le i\ne k\le n} P_k(\mu_i) \star P_k(\lambda_i)$ is 2-torsion by Corollary 3.7(2). Comparing with (3-2), we see that $\prod_{i=1}^n \{l_i, m_i\}^{2\epsilon(i)}$ is 2-torsion in $K_2(F)$. By the argument of [Li and Wang 2008, Proposition 3.2], $\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)}$ is torsion in $K_2(\mathbb{C}(Y^h))$. □

Remark 3.1. This theorem is a natural generalization of [Li and Wang 2008, Proposition 3.2], which concerned the hyperbolic knot case.

Remark 3.2. The proof of Theorem 3.8 uses the condition that the geometric component contains the character χ_0 of the complete hyperbolic structure. For a nongeometric component of the character variety, it is not clear whether we can still have the analogous torsion property on it.

3b. ***Deligne cohomology.*** Here we recall the definition of Deligne cohomology, give the construction of the regulator map, and apply it to our situation.

Let X be a nonsingular variety over $\mathbb{C}$. First recall the definition of the (holomorphic) Deligne cohomology groups of X. For more details, see [Beĭlinson 1984; Brylinski 2008; Esnault and Viehweg 1988]. We define the complex $\mathbb{Z}(p)_{\mathcal{D}}$ of sheaves on X by

$$\mathbb{Z}(p)_{\mathcal{D}} : \mathbb{Z}(p) \longrightarrow \mathcal{O}_X \xrightarrow{d} \Omega_X^1 \xrightarrow{d} \cdots \xrightarrow{d} \Omega_X^{p-1}, \tag{3-3}$$

where $\mathbb{Z}(p)$ is the constant sheaf $(2\pi\sqrt{-1})^p\mathbb{Z}$ and sits in degree zero, $\mathcal{O}_X$ is the sheaf of holomorphic functions on X, and Ω_X^i is the sheaf of holomorphic i-forms on X. The first map in (3-3) is the inclusion and d is the exterior differential. The Deligne cohomology groups of X are defined as the hypercohomology of the complex $\mathbb{Z}(p)_{\mathcal{D}}$:

$$H^q_{\mathcal{D}}(X; \mathbb{Z}(p)) := \mathbb{H}^q(X; \mathbb{Z}(p)_{\mathcal{D}}).$$

For example, the exponential exact sequence of sheaves on X

$$0 \to \mathbb{Z}(1) \to \mathcal{O}_X \to \mathcal{O}_X^* \to 0$$

gives rise to a quasiisomorphism between $\mathbb{Z}(1)_{\mathcal{D}}$ and $\mathcal{O}_X^*[-1]$, where $\mathcal{O}_X^*$ is the sheaf of nonvanishing holomorphic functions on X. Moreover there is a quasi-isomorphism between $\mathbb{Z}(2)_{\mathcal{D}}$ and the complex [Esnault and Viehweg 1988, page 46]

$$(\mathcal{O}_X^* \xrightarrow{d\log} \Omega_X^1)[-1].$$

Therefore, we have for any integer q

$$H^q_{\mathcal{D}}(X; \mathbb{Z}(1)) = H^{q-1}(X; \mathcal{O}_X^*) \quad \text{and} \quad H^q_{\mathcal{D}}(X; \mathbb{Z}(2)) = \mathbb{H}^{q-1}(X; \mathcal{O}_X^* \to \Omega_X^1).$$

On the other hand, Deligne [1991] interprets $\mathbb{H}^1(X; \mathcal{O}_X^* \to \Omega_X^1) = H^2_{\mathcal{D}}(X; \mathbb{Z}(2))$ as the group of holomorphic line bundles with (holomorphic) connections over X. For details, see [Brylinski 2008, Theorem 2.2.20].

Let $\mathbb{C}(X)$ be the function field of X. Given two functions $f, g \in \mathbb{C}(X)$, let $D(f, g)$ be the divisors of the zeros and poles of f and g, and let $|D(f, g)|$ denote its support. Then we have the morphism

$$(f, g) : X - |D(f, g)| \to \mathbb{C}^* \times \mathbb{C}^*,$$

given by $(f, g)(x) = (f(x), g(x))$.

Let $\mathcal{H}$ be the Heisenberg line bundle with connection on $\mathbb{C}^* \times \mathbb{C}^*$. For its construction, see [Bloch 1981] and [Ramakrishnan 1989, Section 4]. Pull back $\mathcal{H}$ along (f, g) to obtain a line bundle $r(f, g)$ with connection on $X - |D(f, g)|$. Hence $r(f, g) \in \mathbb{H}^1(V; \mathcal{O}_V^* \to \Omega_V^1) = H^2_{\mathcal{D}}(V; \mathbb{Z}(2))$, where $V = X - |D(f, g)|$. Moreover we can represent $r(f, g)$ in terms of Čech cocycles for $\mathbb{H}^1(V; \mathcal{O}_V^* \to \Omega_V^1)$. Indeed, choose an open covering $(U_i)_{i\in I}$ of V such that the logarithm $\log_i f$ of f is well-defined on every U_i. Then $r(f, g)$ is represented by the cocyle (c_{ij}, ω_i), with

$$c_{ij} = g^{(\log_j f - \log_i f)/(2\pi\sqrt{-1})} \qquad \text{on } U_i \cap U_j, \tag{3-4}$$

$$\omega_i = \frac{1}{2\pi\sqrt{-1}} \log_i f \ \frac{dg}{g} \qquad \text{on } U_i. \tag{3-5}$$

Its curvature is

$$R = \frac{1}{2\pi\sqrt{-1}} \frac{df}{f} \wedge \frac{dg}{g}. \tag{3-6}$$

Remark 3.3. There is a cup product on the Deligne cohomology groups [Beĭlinson 1984; Esnault and Viehweg 1988]. For $f, g \in H^0(X; \mathcal{O}_X^*) = H^1_{\mathcal{D}}(X; \mathbb{Z}(1))$ as above, the cup product $f \cup g$ is exactly the line bundle $r(f, g) \in H^2_{\mathcal{D}}(X; \mathbb{Z}(2))$.

Furthermore, we have the following properties of $r(f, g)$:

Proposition 3.9. *$r(f_1 f_2, g) = r(f_1, g) \otimes r(f_2, g)$, $r(f, g) = r(g, f)^{-1}$, and the Steinberg relation $r(f, 1 - f) = 1$ holds if $f \neq 0$ and $f \neq 1$.*

Proof. See [Bloch 1981; Esnault and Viehweg 1988] and [Ramakrishnan 1989, Section 4]. The proofs there assume that X is a curve. But they are valid for arbitrary X without change. To prove the Steinberg relation, we need the ubiquitous dilogarithm function. □

Corollary 3.10. *We have the regulator map*

$$r : K_2(\mathbb{C}(X)) \to \varinjlim_{U\subset X:\text{Zariski open}} H^2_{\mathcal{D}}(U; \mathbb{Z}(2)),$$

which maps the symbol $\{f, g\}$ to the line bundle $r(f, g)$.

This follows from the definition of K_2 and Proposition 3.9.

When $\dim X = 1$, the line bundle $r(f, g)$ is always flat, but $r(f, g)$ is not necessarily flat if $\dim X > 1$. Nevertheless:

Proposition 3.11. *If $x \in K_2(\mathbb{C}(X))$ is torsion, the corresponding line bundle $r(x)$ is flat.*

Proof. Let U be the Zariski open subset over which the line bundle $r(x)$ is defined. Since x is torsion in $K_2(\mathbb{C}(X))$, $r(x)$ is torsion in $\mathbb{H}^1(U; \mathbb{O}_U^* \to \Omega_U^1)$. Choose a suitable open covering $(U_i)_{i\in I}$ of U such that $r(x)$ is represented by a Čech cocyle (c_{ij}, ω_i) with $c_{ij} \in \mathbb{O}^*(U_i \cap U_j)$ and $\omega_i \in \Omega^1(U_i)$. Then there exists an integer $n > 0$ such that the class represented by the cocycle $((c_{ij})^n, n\omega_i)$ is zero. Hence, there exists $t_i \in \mathbb{O}_X^*(U_i)$ (or by a refinement covering of $\{U_i\}$) , such that

$$c_{ij}^n = \frac{t_j}{t_i} \quad \text{and} \quad \omega_i = \frac{1}{n}\frac{dt_i}{t_i}.$$

Therefore, $d\omega_i = 0$ for all i and the curvature is 0. □

Let $|D|$ be the support of the divisors of zeros and poles of the rational functions m_i and l_i on Y^h for $1 \le i \le n$. Define $Y_0^h = Y^h - |D|$. The line bundle $r(\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)})$ is well-defined over Y_0^h.

Corollary 3.12. *The line bundle $r(\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)})$ over Y_0^h is flat; therefore it is an element of $H^1(Y_0^h; \mathbb{C}^*)$.*

Proof. This follows from Theorem 3.8 and Proposition 3.11. □

Using the Čech cocycle for $r(f, g)$ given in (3-4) and (3-5), we can represent $r(\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)})$ as follows. Choose an open covering $\{U_\alpha\}_{\alpha\in\Lambda}$ of Y_0^h such that on every U_α, the logarithms of l_i are well-defined and denoted by $\log_\alpha l_i$. Then $r(\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)})$ is represented by the cocyle $(c_{\alpha\beta}, \omega_\alpha)$:

$$c_{\alpha\beta} = \prod_{i=1}^n m_i^{\epsilon(i)((\log_\beta l_i - \log_\alpha l_i))/(2\pi\sqrt{-1})} \quad \text{on } U_\alpha \cap U_\beta, \tag{3-7}$$

$$\omega_\alpha = \sum_{i=1}^n \frac{\epsilon(i)}{2\pi\sqrt{-1}}(\log_\alpha l_i)\frac{dm_i}{m_i} \quad \text{on } U_\alpha. \tag{3-8}$$

Let $t_0 = (l_1^0, m_1^0, \dots, l_n^0, m_n^0) \in Y_0^h$ be a point corresponding to the hyperbolic structure of the link complement $S^3\backslash L$. Then the monodromy of the flat line bundle $r(\prod_{i=1}^n \{m_i, l_i\}^{\epsilon(i)})$ give rises to the representation $M : \pi_1(Y_0^h, t_0) \to \mathbb{C}^*$. With its explicit descriptions (3-7) and (3-8), we have the following formula for M. Let γ be a loop based at t_0. Let $\log l_i$ be a branch of logarithm of l_i over $\gamma - \{t_0\}$, then by a direct calculation we have

$$M(\gamma) = \exp\bigg(\sum_{i=1}^n \Big(-\frac{\epsilon(i)}{2\pi\sqrt{-1}}\Big)\Big(\int_\gamma \log l_i \frac{dm_i}{m_i} - \log m_i(t_0)\int_\gamma \frac{dl_i}{l_i}\Big)\bigg); \tag{3-9}$$

see [Deligne 1991, (2.7.2)].

Now we have the main theorem:

Theorem 3.13. (i) *The real* 1*-form*

$$\eta = \sum_{i=1}^{n} \epsilon(i)(\log|l_i|\, d\arg m_i - \log|m_i|\, d\arg l_i)$$

is exact on Y_0^h*. Hence there exists a smooth function* $V : Y_0^h \to \mathbb{R}$ *such that*

$$dV = \sum_{i=1}^{n} \epsilon(i)(\log|l_i|\, d\arg m_i - \log|m_i|\, d\arg l_i).$$

(ii) *Suppose* $m_i^0 = 1$ *for* $1 \le i \le n$*. For a loop* γ *with initial point* t_0 *in* Y_0^h

$$\frac{1}{4\pi^2} \sum_{i=1}^{n} \epsilon(i) \int_\gamma (\log|m_i| d\log|l_i| + \arg l_i \; d\arg m_i) = \frac{p}{q},$$

where q *is the order of the symbol* $\prod_{i=1}^{n}\{l_i, m_i\}^{\epsilon(i)}$ *in* $K_2(\mathbb{C}(Y^h))$*, and* p *is some integer depending on the loop* $\gamma \in \pi_1(Y_0^h, t_0)$ *and the branches of* $\arg l_i$ *for* $1 \le i \le n$*.*

Proof. First, by (3-8), the curvature of the flat line bundle is

$$R = \sum_{i=1}^{n} \frac{\epsilon(i)}{2\pi\sqrt{-1}} \Big(\frac{dl_i}{l_i} \wedge \frac{dm_i}{m_i}\Big) = 0.$$

On the other hand, we have $d\eta = \mathrm{Im}(\sum_{i=1}^{n} \epsilon(i)(dl_i/l_i \wedge dm_i/m_i))$; hence η is a real closed 1-form.

Since the symbol $\prod_{i=1}^{n}\{l_i, m_i\}^{\epsilon(i)}$ has order q in $K_2(\mathbb{C}(Y^h))$, by (3-9) we have for a loop $\gamma \in \pi_1(Y_0^h, t_0)$ that

$$1 = M(\gamma)^q = \bigg(\exp\Big(\sum_{i=1}^{n}\Big(-\frac{\epsilon(i)}{2\pi\sqrt{-1}}\Big)\Big(\int_\gamma \log l_i \,\frac{dm_i}{m_i} - \log m_i(t_0) \int_\gamma \frac{dl_i}{l_i}\Big)\Big)\bigg)^q.$$

Decompose part of this into real and imaginary parts as

$$\sum_{i=1}^{n} \epsilon(i)\Big(\int_\gamma \log l_i \,\frac{dm_i}{m_i} - \log m_i(t_0) \int_\gamma \frac{dl_i}{l_i}\Big) = \mathrm{Re} + i\,\mathrm{Im},$$

Then we have $\exp(q \cdot \mathrm{Im}/(2\pi) + q \cdot \mathrm{Re}/(2\pi\sqrt{-1})) = 1$. Therefore, $\mathrm{Im} = 0$ and $q \cdot \mathrm{Re}/(2\pi\sqrt{-1}) = 2\pi\sqrt{-1}p$ for some integer p. A straightforward calculation or [Li and Wang 2008, Lemma 3.4] shows that

$$\begin{aligned} &\mathrm{Im} = \int_\gamma \eta, \\ (3\text{-}10) \quad &\mathrm{Re} = -\sum_{i=1}^{n} \epsilon(i) \int_\gamma (\log|m_i|\, d\log|l_i| + \arg l_i \;\, d\arg m_i) = \int_\gamma \xi. \end{aligned}$$

These immediately imply both parts of the theorem. □

Remark 3.4. When $n = 1$, our V is (up to sign) the volume function of the representation of the knot complement [Dunfield 1999]. For $n \geq 2$, up to some constant and signs related to the orientations on each boundary component of the hyperbolic link exterior, the function V should be closely related to the volume function given in [Hodgson 1986, Theorem 5.5].

Remark 3.5. From the proof of Theorem 3.8, the signs $\epsilon(i)$ for $1 \leq i \leq n$ are determined by the orientation of M_L on its n boundary tori. For knots, the sign can be neglected since there is only one term in the 1-form η. For links (where $n \geq 2$), if they are not the same, they could have quite contributions different from those in the knot case. On the other hand, it is not clear what are the exact geometric meanings of these signs for the link L.

Remark 3.6. If there exists any representation $\rho : \pi_1(Y^h) \to \mathrm{GL}_n(\mathbb{C})$ with $n \geq 2$, then Reznikov [1995, Theorem 1.1] proved that for all $i \geq 2$, the Chern classes $c_i \in H^{2i}_{\mathcal{D}}(Y^h; \mathbb{Z}(i))$ in the Deligne cohomology groups are torsion.

3c. ***On the Bohr–Sommerfeld quantization condition for hyperbolic links.*** We now discuss the Theorem 3.13(ii) from a symplectic point of view. When $n = 1$, this is the Bohr–Sommerfeld quantization condition proposed by Gukov for knots in [Gukov 2005, page 597], and is proved in [Li and Wang 2008, Theorem 3.3(2)].

Let Σ be a closed surface with fundamental group π. Its $\mathrm{SL}_2(\mathbb{C})$-character variety is the space of equivalence classes of representations from π into $\mathrm{SL}_2(\mathbb{C})$. This variety carries a natural complex-symplectic structure, where a complex-symplectic structure is a nondegenerate closed holomorphic exterior 2-form; see [Goldman 1984; 2004].

A homomorphism $\rho : \pi \to \mathrm{SL}_2(\mathbb{C})$ is irreducible if it has no proper linear invariant subspace of $\mathbb{C}^2$, and irreducible representations are stable points, denoted by $\mathrm{Hom}(\pi, \mathrm{SL}_2(\mathbb{C}))^s$. Now $\mathrm{SL}_2(\mathbb{C})$ acts freely and properly on $\mathrm{Hom}(\pi, \mathrm{SL}_2(\mathbb{C}))^s$, and the quotient $X^s(\Sigma) = \mathrm{Hom}(\pi, \mathrm{SL}_2(\mathbb{C}))^s / \mathrm{SL}_2(\mathbb{C})$ is an embedding onto an open subset in the geometric quotient $\mathrm{Hom}(\pi, \mathrm{SL}_2(\mathbb{C})) /\!/ \mathrm{SL}_2(\mathbb{C})$. Thus $X^s(\Sigma)$ is a smooth irreducible complex quasiaffine variety that is dense in the geometric quotient [Goldman 2004, Section 1]. Note that ρ is a nonsingular point if and only if $\dim Z(\rho)/Z(\mathrm{SL}_2(\mathbb{C})) = 0$, and this corresponds to the top stratum $X^s(\Sigma)$, where

$Z(u)$ is the centralizer of u in $\mathrm{SL}_2(\mathbb{C})$. If $\rho \in \mathrm{Hom}(\pi, \mathrm{SL}_2(\mathbb{C}))$ is a singular point (that is, $\dim Z(\rho)/Z(\mathrm{SL}_2(\mathbb{C})) > 0$), then all points of $\sigma \in \mathrm{Hom}(\pi, Z(Z(\rho)))^s$ with $\mathrm{stab}(\sigma) = Z(\sigma) = Z(\rho)$ have the same orbit type and form a stratification of the $\mathrm{SL}_2(\mathbb{C})$-character variety [Goldman 1984, Section 1].

We have the $\mathrm{SL}_2(\mathbb{C})$-character variety $X(T^2)$ of the torus T^2 as a surface in $\mathbb{C}^3$ given by

$$x^2 + y^2 + z^2 - xyz - 4 = 0.$$

See [Li and Wang 2006, Proposition 3.2]. There is a natural symplectic structure on the smooth top stratum $X^s(T^2)$ of $X(T^2)$, and there exists a symplectic structure ω on the character variety $X^s(\partial M_L) = \prod_{i=1}^n X^s(T_i^2)$ such that $X(M_L) \cap X^s(\partial M_L)$ (a subset of $X(M_L)$) is a Lagrangian subvariety of $X^s(\partial M_L)$, where $X^s(\partial M_L)$ is a smooth irreducible variety that is open and dense in $X(\partial M_L)$.

The inclusion $\partial M_L \to M_L$ indeed induces a degree one map on the irreducible components. Thus $r(X_0)^s$ (the smooth part of the image $r(X_0)$) is a Lagrangian submanifold of the symplectic manifold $X^s(\partial M_L)$. Note that the pullback of the symplectic 2-form on the double covering of $X^s(T_i^2)$ is again skew-symmetric and nondegenerate. The symplectic form $\tilde{\omega}_i$ induced by the map $t_i : r(X_0) \to X(T_i^2)$ gives the Lagrangian property for the corresponding pullback of the Lagrangian part $r(X_0^i)^s$. Hence we have the product Lagrangian smooth part of the pullback of $\prod_{i=1}^n r(X_0^i)^s$. Then we need to see that the smooth projective model preserves the Lagrangian and symplectic property.

Let $\tilde{X}(T_i^2)$ be the symplectic blowup of the double covering of $X(T_i^2)$ as in [McDuff and Salamon 1998]. The blowup in the complex category carries a natural symplectic structure on $\tilde{X}(T_i^2)$; see [McDuff and Salamon 1998, Section 7.1]. On the other hand, the corresponding part $\overline{Y}_i$ of Y_i (the irreducible component of D_i containing y_i) lies in the symplectic manifold $\tilde{X}(T_i^2)$.

Define a compatible Lagrangian blowup with respect to the complex blowup as following. Define a real submanifold $\tilde{\mathbb{R}}^n$ of $\mathbb{R}^n \times \mathbb{R}P^{n-1}$ (a subset of $\mathbb{C}^n \times \mathbb{C}P^{n-1}$) as a subspace of pairs (x, l) with $x = \mathrm{Re}(z) \in l$, where $l \in \mathbb{R}P^{n-1}$ is a real line in $\mathbb{R}^n$. If $I_{\mathbb{C}}$ is complex conjugation on $\mathbb{C}^n$ and $J_{\mathbb{C}P^{n-1}}$ is the complex involution on $\mathbb{C}P^{n-1}$ given by complex conjugation on each component, then

$$\begin{aligned}\tilde{\mathbb{R}}^n &= \mathrm{Fix}(I_{\mathbb{C}} \times J_{\mathbb{C}P^{n-1}}|_{\tilde{\mathbb{C}}^n}) \subset \tilde{\mathbb{C}}^n \\ &= \{(z_1, \dots, z_n; [w_1 : \cdots : w_n]) \mid w_j z_k = w_k z_j,\ 1 \le j, k \le n\}.\end{aligned}$$

It is clear that $\tilde{\mathbb{R}}^n$ is Lagrangian in $\tilde{\mathbb{C}}^n$. Hence the real Lagrangian blowup $\tilde{\overline{Y}}_i$ is Lagrangian in $\tilde{X}(T_i^2)$, and the Lagrangian submanifold $\tilde{Y}^h$ is Lagrangian in the symplectic manifold $\prod_{i=1}^n \tilde{X}(T_i^2)$. In this way, the symplectic and Lagrangian properties are preserved under the blowup, and we can treat the Lagrangian blowup in a real blowup by looking at the complex one.

Now we have a Lagrangian submanifold $\tilde{Y}_0^h$ in a symplectic manifold. Suppose $m_i^0 = 2$ for $1 \le i \le n$. For a loop γ with initial point t_0 in $\tilde{Y}_0^h$, Theorem 3.13(ii) gives

$$\frac{1}{4\pi^2}\sum_{i=1}^{n}\epsilon(i)\int_\gamma (\log|m_i|\,d\log|l_i| + \arg l_i\,d\arg m_i) = \frac{p}{q},$$

where p is some integer and q is the order of the symbol $\prod_{i=1}^n \{l_i, m_i\}^{\epsilon(i)}$ in $K_2(\mathbb{C}(Y^h))$. We shall call this result the Bohr–Sommerfeld quantization condition for hyperbolic links. It would be interesting to give an interpretation from mathematical physics, as what Gukov did for hyperbolic knots.

4. On a possible unified volume conjecture for both knots and links

By Corollary 3.12, the class $r(\prod_{i=1}^n \{l_i, m_i)^{\varepsilon_i})$ corresponds to a flat line bundle over Y_0^h; therefore the curvature of the holomorphic connection is zero. Formally this can be expressed as $d(\xi + \sqrt{-1}\eta) = 0$, where ξ and η are defined in (3-10). Hence, $(\xi + \sqrt{-1}\eta)/(2\pi\sqrt{-1})$ can be viewed as the Chern–Simons 1-*form* of the line bundle $r(\prod_{i=1}^n \{l_i, m_i)^{\varepsilon_i})$.

Given a point $p \in Y_0^h$, choose a path $\gamma : [0, 1] \to Y_0^h$ with $\gamma(1) = p$ and $\gamma(0) = t_0$ a point corresponding to the complete hyperbolic structure. Write

$$\gamma(t) = (l(t), m(t)) = (l_1(t), m_1(t), \dots, l_n(t), m_n(t)).$$

Recall that q is the order of the symbol $\prod_{i=1}^n \{l_i, m_i\}^{\varepsilon_i}$ in $K_2(\mathbb{C}(Y^h))$. Let $\mathrm{Vol}(L)$ and $\mathrm{CS}(L)$ be the volume and usual Chern–Simons invariant of the complete hyperbolic structure on $S^3 \setminus L$, respectively. Now we define

$$V(p) = \mathrm{Vol}(L) + 2\cdot\sum_{i=1}^{n}\epsilon(i)\int_\gamma (\log|l_i|\,d\arg m_i - \log|m_i|\,d\arg l_i). \tag{4-1}$$

$$U(p) = 4\pi^2\,\mathrm{CS}(L) + q\cdot\sum_{i=1}^{n}\epsilon(i)\int_\gamma (\log|m_i|\,d\log|l_i| + \arg l_i\,d\arg m_i). \tag{4-2}$$

According to Theorem 3.13, $R(p) = (2\pi)^{-1}(V(p) + \sqrt{-1}(2\pi)^{-1}U(p))$ is independent of the choices of the path γ and takes values in $\mathbb{C}/\mathbb{Z}$. We call

$$\frac{1}{4\pi^2}U(p)$$

the *special Chern–Simons invariant* of the hyperbolic link L at p. When $p = t_0$, it equals $\mathrm{CS}(L)$.

Remark 4.1. For $p \ne t_0$, $U(p)/(4\pi^2)$ is different from the usual Chern–Simons invariant for a 3-dimensional manifold. The latter comes from the transgressive 3-form of the second Chern class of the 3-dimensional manifold.

In order to formulate a parametrized conjecture parallel to the knot case as in [Li and Wang 2008, Conjecture 3.9], we have to find a way to relate the quantum invariants to the n-dimensional variety Y_0^h that comes from the $\mathrm{SL}_2(\mathbb{C})$ character variety. By the work of Kashaev [1995] and Baseilhac and Benedetti [2004], there exists an $\mathrm{SL}_2(\mathbb{C})$ quantum hyperbolic invariant for a hyperbolic link in S^3, which is conjectured to give the information of the volume and Chern–Simons at the point for the complete hyperbolic structure.

Here is a conjectural description. Given a point $p \in Y_0^h$ corresponding to an $\mathrm{SL}_2(\mathbb{C})$ representation of $\pi_1(M_L)$, let's assume that we can define certain quantum invariants $K_N(L, p)$. Then we formulate the following:

Conjecture 4.1 (a possibly unified parametrized volume conjecture).

$$\lim_{N\to\infty} \frac{\log K_N(L, p)}{N} = \frac{1}{2\pi}\Big(V(p) + \frac{\sqrt{-1}}{2\pi} U(p)\Big).$$

Remark 4.2. When L is a hyperbolic knot (that is, $n = 1$), Y^h is the smooth projective model of an irreducible component of the locus of the A-polynomial that contains the complete hyperbolic structure. Fix a number a. For $p = (l, m) \in Y_0^h$ with $m = -\exp(\sqrt{-1}\pi a)$, we take $K_N(L, p) = J_N(L, e^{2\pi\sqrt{-1}a/N})$, the values of the colored Jones polynomial of L evaluated at $e^{2\pi\sqrt{-1}\,a/N}$. Then Conjecture 4.1 reduces to the reformulated generalized volume conjecture (3.9) of [Li and Wang 2008] for hyperbolic knots. When γ is the constant path at t_0, or equivalently $p = t_0$, it reduces to the complexification of Kashaev's conjecture for hyperbolic knots; see [Murakami et al. 2002, Conjeture 1.2].

Remark 4.3. When $n \geq 2$, we can take $K_N(L, t_0)$ to be the Kashaev and Baseilhac–Benedetti invariant that is based on the triangulations of the manifold and is conjectured to give the information of the volume and Chern–Simons at the complete hyperbolic structure t_0. See [Baseilhac and Benedetti 2004, Section 5]. For a general $p \in Y_0^h$, we do not have a rigorous definition, although we expect that there is a way of deforming $K_N(L, t_0)$ to get $K_N(L, p)$.

Remark 4.4. If the point corresponding to the hyperbolic structure in Y_i is not smooth, then the point t_0 in the definition of (4-1) and (4-2) is not unique. If we make different choices of t_0, then $V(p)$ and $U(p)$ will differ by a constant, corresponding to choice made in the integrals in (4-1) and (4-2). We can modify the left side of the Conjecture 4.1 by this constant accordingly. So the choice of t_0 is not essential, and it seems that there is no canonical choice.

Remark 4.5. From the regulator point of view developed in this paper, we expect there exists a parametrized version of the volume conjecture for both hyperbolic links and knots.

Acknowledgments

Q. Wang is grateful for the support and hospitality of the Marie Curie Research Programme at DPMMS, University of Cambridge and the program ANR "Galois" at the Université Pierre et Marie Curie (Paris 6) and École Normale Supérieure, Paris. He thanks Professors Y. André and A. Scholl for their helpful discussions. We thank the referee for helpful comments and suggestions.

References

[Baseilhac and Benedetti 2004] S. Baseilhac and R. Benedetti, "Quantum hyperbolic invariants of 3-manifolds with $PSL(2, \mathbb{C})$-characters", *Topology* **43**:6 (2004), 1373–1423. MR 2005d:57015 Zbl 1065.57008

[Beĭlinson 1984] A. A. Beĭlinson, "Higher regulators and values of L-functions", pp. 181–238 Современные проблемы математики **24**, Akad. Nauk SSSR Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1984. In Russian; translated in *J. Sov. Math* **30**, 2036–2070. MR 86h:11103 Zbl 0588.14013

[Bloch 1981] S. Bloch, "The dilogarithm and extensions of Lie algebras", pp. 1–23 in *Algebraic K-theory* (Evanston, IL, 1980), edited by E. M. Friedlander and M. R. Stein, Lecture Notes in Math. **854**, Springer, Berlin, 1981. MR 83b:17010 Zbl 0469.14009

[Brylinski 2008] J.-L. Brylinski, *Loop spaces, characteristic classes and geometric quantization*, Birkhäuser, Boston, MA, 2008. MR 2008h:53155 Zbl 1136.55001

[Cooper et al. 1994] D. Cooper, M. Culler, H. Gillet, D. D. Long, and P. B. Shalen, "Plane curves associated to character varieties of 3-manifolds", *Invent. Math.* **118**:1 (1994), 47–84. MR 95g:57029 Zbl 0842.57013

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

[Culler and Shalen 1984] M. Culler and P. B. Shalen, "Bounded, separating, incompressible surfaces in knot manifolds", *Invent. Math.* **75**:3 (1984), 537–545. MR 85k:57010 Zbl 0542.57011

[Culler et al. 1987] M. Culler, C. M. Gordon, J. Luecke, and P. B. Shalen, "Dehn surgery on knots", *Ann. of Math.* (2) **125**:2 (1987), 237–300. MR 88a:57026 Zbl 0633.57006

[Deligne 1991] P. Deligne, "Le symbole modéré", *Inst. Hautes Études Sci. Publ. Math.* 73 (1991), 147–181. MR 93i:14030 Zbl 0749.14011

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

[Dupont 1987] J. L. Dupont, "The dilogarithm as a characteristic class for flat bundles", *J. Pure Appl. Algebra* **44**:1-3 (1987), 137–164. MR 88k:57032 Zbl 0624.57024

[Esnault and Viehweg 1988] H. Esnault and E. Viehweg, "Deligne–Beĭlinson cohomology", pp. 43–91 in *Beĭlinson's conjectures on special values of L-functions*, edited by M. Rapoport et al., Perspect. Math. **4**, Academic Press, Boston, MA, 1988. MR 89k:14008 Zbl 0656.14012

[Goldman 1984] W. M. Goldman, "The symplectic nature of fundamental groups of surfaces", *Adv. in Math.* **54**:2 (1984), 200–225. MR 86i:32042 Zbl 0574.32032

[Goldman 2004] W. M. Goldman, "The complex-symplectic geometry of $SL(2, \mathbb{C})$-characters over surfaces", pp. 375–407 in *Algebraic groups and arithmetic*, edited by S. G. Dani and G. Prasad, Tata Inst. Fund. Res., Mumbai, 2004. MR 2005i:53110 Zbl 1089.53060

[Goncharov 1999] A. Goncharov, "Volumes of hyperbolic manifolds and mixed Tate motives", *J. Amer. Math. Soc.* **12**:2 (1999), 569–618. MR 99i:19004 Zbl 0919.11080

[Gukov 2005] S. Gukov, "Three-dimensional quantum gravity, Chern–Simons theory, and the A-polynomial", *Comm. Math. Phys.* **255**:3 (2005), 577–627. MR 2006f:58029 Zbl 1115.57009

[Hodgson 1986] C. Hodgson, *Degeneration and regeneration of geometric structures on three-manifolds*, thesis, Princeton University, 1986.

[Kashaev 1995] R. M. Kashaev, "A link invariant from quantum dilogarithm", *Modern Phys. Lett. A* **10**:19 (1995), 1409–1418. MR 96j:81060 Zbl 1022.81574

[Li and Wang 2006] W. Li and Q. Wang, "An $SL_2(\mathbb{C})$ algebro-geometric invariant of knots", preprint, 2006. To appear in *Internat. J. Math.* arXiv math.GT/0610752

[Li and Wang 2008] W. Li and Q. Wang, "On the generalized volume conjecture and regulator", *Commun. Contemp. Math.* **10**:suppl. 1 (2008), 1023–1032. MR 2010a:57027 Zbl 05521680

[McDuff and Salamon 1998] D. McDuff and D. Salamon, *Introduction to symplectic topology*, 2nd ed., Oxford University Press, New York, 1998. MR 97b:58062 Zbl 0844.58029

[Milnor 1971] J. Milnor, *Introduction to algebraic K-theory*, Annals of Mathematics Studies **72**, Princeton University Press, 1971. MR 50 #2304 Zbl 0237.18005

[Murakami and Murakami 2001] H. Murakami and J. Murakami, "The colored Jones polynomials and the simplicial volume of a knot", *Acta Math.* **186**:1 (2001), 85–104. MR 2002b:57005 Zbl 0983.57009

[Murakami et al. 2002] H. Murakami, J. Murakami, M. Okamoto, T. Takata, and Y. Yokota, "Kashaev's conjecture and the Chern–Simons invariants of knots and links", *Experiment. Math.* **11**:3 (2002), 427–435. MR 2004a:57016 Zbl 1117.57300

[Ramakrishnan 1989] D. Ramakrishnan, "Regulators, algebraic cycles, and values of L-functions", pp. 183–310 in *Algebraic K-theory and algebraic number theory* (Honolulu, HI, 1987), edited by M. R. Stein and R. K. Dennis, Contemp. Math. **83**, Amer. Math. Soc., Providence, RI, 1989. MR 90e:11094 Zbl 0694.14002

[Reznikov 1995] A. Reznikov, "All regulators of flat bundles are torsion", *Ann. of Math.* (2) **141**:2 (1995), 373–386. MR 96a:14012 Zbl 0865.14003

[Shalen 2002] P. B. Shalen, "Representations of 3-manifold groups", pp. 955–1044 in *Handbook of geometric topology*, edited by R. J. Daverman and R. B. Sher, North-Holland, Amsterdam, 2002. MR 2003d:57002 Zbl 1012.57003

Received December 9, 2009.

WEIPING LI
DEPARTMENT OF MATHEMATICS
OKLAHOMA STATE UNIVERSITY
STILLWATER, OK 74078
UNITED STATES
wli@math.okstate.edu

QINGXUE WANG
SCHOOL OF MATHEMATICAL SCIENCES
FUDAN UNIVERSITY
SHANGHAI 200433
CHINA
qxwang@fudan.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

HYPERGEOMETRIC EVALUATION IDENTITIES AND SUPERCONGRUENCES

LING LONG

We apply some hypergeometric evaluation identities, including a strange valuation of Gosper, to prove several supercongruences related to special valuations of truncated hypergeometric series. In particular, we prove a conjecture of van Hamme.

1. Introduction

In this article, we use p to denote an odd prime. Zudilin [2009] proved several Ramanujan-type supercongruences using the Wilf–Zeilberger (WZ) method. One of them, conjectured by van Hamme, says that

$$\sum_{k=0}^{(p-1)/2}(4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3(-1)^k \equiv (-1)^{(p-1)/2}p \mod p^3, \tag{1}$$

where $(a)_k = a(a+1)\cdots(a+k-1)$ is the rising factorial for $a \in \mathbb{C}$ and $k \in \mathbb{N}$.

The first proof of (1) was given by Mortenson [2008]. It is said to be of Ramanujan-type because it is a p-adic version of Ramanujan's formula

$$\sum_{k=0}^{\infty}(4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3(-1)^k = \frac{2}{\pi}.$$

See [Zudilin 2009] for more Ramanujan-type supercongruences.

In this short note, we will present a new proof of (1), which summarizes our strategy in proving similar types of supercongruences.

McCarthy and Osburn [2008] proved van Hamme's conjecture [1997] that

$$\sum_{k=0}^{(p-1)/2}(4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^5 \equiv \begin{cases} -\dfrac{p}{\Gamma_p(3/4)^4} & \mod p^3 \quad \text{if } p \equiv 1 \mod 4, \\ 0 & \mod p^3 \quad \text{if } p \equiv 3 \mod 4, \end{cases}$$

where $\Gamma_p(\cdot)$ denotes the p-adic Gamma function.

Supported by NSA grant H98230-08-1-0076.
MSC2000: 33C20.
Keywords: Ramanujan supercongruences, hypergeometric identities.

Similarly, van Hamme has conjectured that for any prime $p > 3$,

$$\sum_{k=0}^{(p-1)/2} (6k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3 4^{-k} \equiv (-1)^{(p-1)/2} p \mod p^4. \tag{2}$$

This formula is supported by numerical evidence, but as van Hamme said, "we have no real explanation for our observations". In our exploration, it will become clear that such supercongruences are a result of extra symmetries, which we are able to interpret using hypergeometric evaluation identities. Of course, they can also be seen from other perspectives, such as the WZ method.

Meanwhile, it is known that some of the truncated hypergeometric series are related to the number of rational points on certain algebraic varieties over finite fields and further to coefficients of modular forms. For instance, based on the result of Ahlgren and Ono [2000], Kilbourn [2006] proved that

$$\sum_{k=0}^{(p-1)/2} \left(\frac{(\frac{1}{2})_k}{k!}\right)^4 \equiv a_p \mod p^3, \tag{3}$$

where a_p is the p-th coefficient of a weight 4 modular form

$$\eta(2z)^4\eta(4z)^4 := q\prod_{n\geq 1}(1-q^{2n})^4(1-q^{4n})^4, \quad \text{where } q = e^{2\pi i z}. \tag{4}$$

This is one instance of the supercongruences conjectured by Rodriguez-Villegas [2003], which relate special truncated hypergeometric series values and coefficients of Heck eigenforms. McCarthy [2009] proved another supercongruence of this type and his approach provides a general combinatorial framework for all these congruences.

We will establish a few supercongruences mainly via hypergeometric evaluation identities and combinatorics. Since there exist many amazing hypergeometric evaluation identities in the literature, we expect that our approach can be used to prove other interesting congruences.

Here is a summary of our results.

Theorem 1.1. *Let $p > 3$ be a prime and r be a positive integer. Then*

$$\sum_{k=0}^{(p^r-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^4 \equiv p^r \mod p^{3+r}.$$

Theorem 1.2. *Let $p > 3$ be a prime. Then*

$$\sum_{k=0}^{(p-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^6 \equiv p \cdot a_p \mod p^4.$$

Conjecture 1.3. *Let $p > 3$ be a prime and r be a positive integer. Then*

$$\sum_{k=0}^{(p^r-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^6 \equiv p^r \cdot a_{p^r} \mod p^{3+r},$$

where a_{p^r} is the p^r-th coefficient of (4).

Theorem 1.4. *Van Hamme's conjecture* (2) *is true.*

Theorem 1.5. *Let $p > 3$ be a prime. Then*

$$\sum_{k=0}^{(p-1)/2} (6k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3 \frac{(-1)^k}{8^k} \equiv (-1)^{(p^2-1)/8+(p-1)/2} p \mod p^2. \tag{5}$$

2. Preliminaries

Hypergeometric series. For any positive integer r,

$${}_{r+1}F_r\left[\begin{matrix} a_1, a_2, \ldots, a_{r+1}; z \\ b_1, \ldots, b_r \end{matrix}\right] = \sum_{k\ge 0} \frac{(a_1)_k \cdots (a_{r+1})_k}{k!(b_1)_k \cdots (b_r)_k} z^k,$$

where $(a)_k$ is the rising factorial and $z \in \mathbb{C}$. A hypergeometric series terminates if it is well-defined and at least one of the a_i is a negative integer. We will make use of this fact to produce various truncated hypergeometric series.

By the definition of the rising factorial,

$$\frac{(\frac{1}{2})_k}{k!} = 2^{-2k}\binom{2k}{k}. \tag{6}$$

Gamma function. Let $\Gamma(x)$ denote the usual Gamma function, which is defined for all $x \in \mathbb{C}$ except for the nonpositive integers. It satisfies some well known properties, such as $\Gamma(x+1) = x\Gamma(x)$. Thus, $(a)_k = \Gamma(a+k)/\Gamma(a)$ when $\Gamma(a) \neq 0$ and $\Gamma(a+k)$ are defined.

Another formula we need is Euler's reflection formula

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}.$$

Some combinatorics. We gather here some results in combinatorics to be used later. It is the author's pleasure to acknowledge that the approaches used in (7)–(10) are due to Zudilin. Here is a key idea of Zudilin for rising factorials; see also [Chan et al. 2010, Lemma 1]:

$$\begin{aligned}\left(\tfrac{1}{2}+\varepsilon\right)_k &= (\tfrac{1}{2}+\varepsilon)(\tfrac{1}{2}+\varepsilon+1)\cdots(\tfrac{1}{2}+\varepsilon+k-1) \\ &= \left(\tfrac{1}{2}\right)_k\Big(1 + 2\varepsilon\sum_{j=1}^{k}\frac{1}{2j-1} + 4\varepsilon^2 \sum_{1\le i<j\le k}\frac{1}{(2i-1)(2j-1)} + O(\varepsilon^3)\Big).\end{aligned} \tag{7}$$

Hence, $\left(\frac{1}{2}+\varepsilon\right)_k\left(\frac{1}{2}-\varepsilon\right)_k$ can be expanded as a power series of ε^2 as

$$\left(\tfrac{1}{2}+\varepsilon\right)_k\left(\tfrac{1}{2}-\varepsilon\right)_k=\left(\tfrac{1}{2}\right)_k^2\Bigl(1-4\varepsilon^2\sum_{j=1}^{k}\frac{1}{(2j-1)^2}+O(\varepsilon^4)\Bigr). \tag{8}$$

Similarly,

$$(1+\varepsilon)_k(1-\varepsilon)_k=(1)_k^2\Bigl(1-\varepsilon^2\sum_{j=1}^{k}\frac{1}{j^2}+O(\varepsilon^4)\Bigr). \tag{9}$$

Letting $\varepsilon=-p^r/2$ and $\varepsilon=p^r/2$ respectively in (7) and taking k to be an integer between 1 and $(p^r-1)/2$, we obtain

$$(-1)^k\binom{(p^r-1)/2}{k}\equiv\frac{(\frac{1}{2})_k}{k!}\mod p\quad\text{and}\quad\binom{(p^r-1)/2+k}{k}\equiv\frac{(\frac{1}{2})_k}{k!}\mod p.$$

Similarly, letting $\varepsilon=p^r/2$ in (8) and k be an integer between 1 and $(p^r-1)/2$, we have

$$(-1)^k\binom{(p^r-1)/2}{k}\binom{(p^r-1)/2+k}{k}\equiv\left(\frac{(\frac{1}{2})_k}{k!}\right)^2\mod p^2.$$

Lemma 2.1. *For any positive integer $n>1$,*

$$(2n+1)\sum_{k=0}^{n}\frac{1}{2k+1}\binom{n}{k}\binom{n+k}{k}(-1)^k=1. \tag{10}$$

Proof. We use the partial fraction decomposition

$$\frac{(t-1)(t-2)\cdots(t-n)}{t(t+1)\cdots(t+n)}=\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}\binom{n+k}{k}\frac{1}{t+k}.$$

Letting $t=1/2$, this becomes

$$(-1)^n\frac{2}{2n+1}=2\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}\binom{n+k}{k}\frac{1}{1+2k},$$

which is equivalent to the claim of the lemma. □

Lemma 2.2. *Let n be an odd positive integer. Then*

$$\frac{(\frac{3}{2}-\frac{1}{4}n)_{(n-1)/2}(1-\frac{1}{2}n)_{(n-1)/2}}{(2-\frac{1}{2}n)_{(n-1)/2}(1-\frac{1}{4}n)_{(n-1)/2}}=(-1)^{(n-1)/2}n.$$

Proof. Using $(a)_k = \Gamma(a+k)/\Gamma(a)$, we have

$$\begin{aligned}\frac{(\frac{3}{2}-\frac{1}{4}n)_{(n-1)/2}(1-\frac{1}{2}n)_{(n-1)/2}}{(2-\frac{1}{2}n)_{(n-1)/2}(1-\frac{1}{4}n)_{(n-1)/2}} &= \frac{\Gamma(\frac{3}{2}-\frac{1}{4}n+\frac{1}{2}(n-1))\Gamma(\frac{1}{2})\Gamma(2-\frac{1}{2}n)\Gamma(1-\frac{1}{4}n)}{\Gamma(\frac{3}{2}-\frac{1}{4}n)\Gamma(1-\frac{1}{2}n)\Gamma(\frac{3}{2})\Gamma(1-\frac{1}{4}n+\frac{1}{2}(n-1))} \\ &= \frac{(1-\frac{1}{2}n)}{\frac{1}{2}}\frac{\frac{1}{4}n\cdot\Gamma(\frac{1}{4}n)\Gamma(1-\frac{1}{4}n)}{(\frac{1}{2}-\frac{1}{4}n)\cdot\Gamma(\frac{1}{2}+\frac{n}{4})\Gamma(\frac{1}{2}-\frac{1}{4}n)} \\ &= n\cdot\frac{\sin(\pi/2-\pi n/4)}{\sin(\pi n/4)} = n\cdot\cot(\pi n/4) = (-1)^{(n-1)/2}n. \qquad \square\end{aligned}$$

Lemma 2.3. *Let n be an odd integer. Then*

$$\frac{(\frac{3}{2}-\frac{1}{4}n)_{(n-1)/2}}{(2-\frac{1}{2}n)_{(n-1)/2}}2^{(n-1)/2} = (-1)^{(n^2-1)/8+(n-1)/2}n.$$

Proof. We have

$$\frac{(\frac{3}{2}-\frac{1}{4}n)_{(n-1)/2}}{(2-\frac{1}{2}n)_{(n-1)/2}}2^{(n-1)/2} = \frac{(3-\frac{1}{2}n)(5-\frac{1}{2}n)\cdots\frac{n}{2}}{(2-\frac{1}{2}n)(3-\frac{1}{2}n)\cdots\frac{1}{2}} = \mathrm{sgn}\cdot n,$$

where $\mathrm{sgn} = (-1)^{\#}$ and # is the number of negative terms appearing in the fraction above. It is easy to see that

$$\# = \lfloor\tfrac{1}{2}(\tfrac{1}{2}n+1)\rfloor + \lfloor\tfrac{1}{2}n\rfloor - 2 \equiv \tfrac{1}{8}(n^2-1)+\tfrac{1}{2}(n-1) \mod 2. \qquad \square$$

Lemma 2.4 [Cai 2002]. *For any prime $p > 3$ and positive integer r,*

$$(11) \qquad (-1)^{(p^r-1)/2}\binom{p^r-1}{\frac{1}{2}(p^r-1)} \equiv \left(\frac{(\frac{1}{2})_{(p^r-1)/2}}{(\frac{1}{2}(p^r-1))!}\right)^2 \mod p^3.$$

Using (6), the congruence (11) is equivalent to

$$\binom{p^r-1}{\frac{1}{2}(p^r-1)} \equiv (-1)^{(p^r-1)/2}2^{2(p^r-1)} \mod p^3.$$

When $r = 1$, this was proved in [Morley 1895].

A generalized harmonic sum. Let $H_k^{(2)} := \sum_{j=1}^{k}\frac{1}{j^2}$.

Lemma 2.5 [Morley 1895]. *Let $p > 3$ be a prime. We have*

$$H_{(p-1)/2}^{(2)} \equiv 0 \mod p \quad \textit{and} \quad \sum_{j=1}^{(p-1)/2}\frac{1}{(2j-1)^2} \equiv 0 \mod p.$$

Using arguments in [Morley 1895] or elementary congruence, it is easy to see the following lemma holds.

Lemma 2.6. *Let $p>3$ be a prime. Then for every integer k between* 1 *and* $p-2$,

$$H_k^{(2)}+H_{p-1-k}^{(2)}\equiv 0 \mod p.$$

Lemma 2.7. *Let $p>3$ be a prime and s be a positive integer. Then*

$$\sum_{k=0}^{(p-1)/2}\left(\frac{(\frac{1}{2})_k}{k!}\right)^{2s}\cdot H_{2k}^{(2)}\equiv 0 \mod p.$$

Proof. Using the fact that

$$(-1)^k\binom{\frac{1}{2}(p-1)}{k}\equiv\frac{(\frac{1}{2})_k}{k!} \mod p,$$

we have

$$\begin{aligned}\sum_{k=0}^{(p-1)/2}\left(\frac{(\frac{1}{2})_k}{k!}\right)^{2s}H_{2k}^{(2)}&\equiv\sum_{k=0}^{(p-1)/2}\binom{\frac{1}{2}(p-1)}{k}^{2s}H_{2k}^{(2)} \mod p\\&=\frac{1}{2}\left(\sum_{k=0}^{(p-1)/2}\binom{\frac{1}{2}(p-1)}{k}^{2s}H_{2k}^{(2)}+\sum_{k=0}^{(p-1)/2}\binom{\frac{1}{2}(p-1)}{\frac{1}{2}(p-1)-k}^{2s}H_{p-1-2k}^{(2)}\right)\\&=\frac{1}{2}\left(\sum_{k=0}^{(p-1)/2}\binom{\frac{1}{2}(p-1)}{k}^{2s}(H_{2k}^{(2)}+H_{p-1-2k}^{(2)})\right)\\&\equiv 0 \mod p.\end{aligned}$$

□

2.1. ***An elementary p-adic analysis.*** Let $F(x_1,\dots,x_t;z)$ be a $(t+1)$-variable formal power series. For instance, it could be a scalar multiple of a terminating hypergeometric series as follows:

$$C\cdot{}_{r+1}F_r\left[\begin{matrix}a_1,a_2,\dots, & a_r, & -n;z\\ b_1,\dots,b_{r-1}, & b_r\end{matrix}\right].$$

Assume that by specifying values $x_i=a_i$ for $i=1,\dots,t$ and $z=z_0$, we have

$$F(a_1,\dots,a_t;z_0)\in\mathbb{Z}_p.$$

Now we fix z_0 and deform the parameters a_i into polynomials $a_i(x)\in\mathbb{Z}_p[x]$ such that $a_i(0)=a_i$ for all $1\le i\le t$, and assume that the resulting function $F(a_1(x),\dots,a_t(x);z_0)$ is a formal power series in x^2 with coefficients in $\mathbb{Z}_p$, that is, $F(a_1(x),\dots,a_t(x);z_0)=A_0+A_2x^2+A_4x^4+\cdots$ for $A_i\in\mathbb{Z}_p$, where $A_0=F(a_1,\dots,a_t;z_0)$.

Lemma 2.8. *Under the setting above, if* $p^s \mid A_2$ *for* $s = 1, 2$, *then*

$$F(a_1(p), \ldots, a_t(p); z_0) \equiv A_0 \mod p^{2+s}.$$

3. A new proof of (1)

We briefly outline our method for proving the next few supercongruences; we are motivated by [McCarthy and Osburn 2008] and [Mortenson 2008]. To each congruence, we first identify a corresponding hypergeometric evaluation identity, which with specified parameters is congruent to a target truncated hypergeometric series evaluation up to some power of p. Usually the power of p so obtained is weaker than the conjectural exponent. In our cases, we reduce the optimal congruences to some congruence combinatorial identities, which are established using additional hypergeometric evaluation identities or combinatorics.

Our strategy can be best implemented in the following new proof of (1). An identity of Whipple [1926, (5.1)] says

$${}_4F_3\left[\begin{matrix} a, 1+a/2, & c, & d; & -1 \\ & a/2, & 1+a-c, 1+a-d & \end{matrix}\right] = \frac{\Gamma(1+a-c)\Gamma(1+a-d)}{\Gamma(1+a)\Gamma(1+a-c-d)}.$$

Letting $a = \frac{1}{2}$, $c = \frac{1}{2} + \frac{1}{2}p$ and $d = \frac{1}{2} - \frac{1}{2}p$, we conclude immediately that

$$\sum_{k=0}^{(p-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3 (-1)^k \equiv \frac{\Gamma(1-\frac{1}{2}p)\Gamma(1+\frac{1}{2}p)}{\Gamma(\frac{1}{2})\Gamma(\frac{3}{2})} = (-1)^{(p-1)/2} p \mod p^2.$$

To achieve the congruence modulo p^3, we consider the expansion of the terminating hypergeometric series (it terminates since $(1-p)/2$ is a negative integer)

$$(12)\quad {}_4F_3\left[\begin{matrix} \frac{1}{2}(1-p), \frac{5}{4}, \frac{1}{2}(1-x), \frac{1}{2}(1+x); -1 \\ \frac{1}{4},\ 1+\frac{1}{2}x,\ \ 1-\frac{1}{2}x \end{matrix}\right]$$
$$= \sum_{k=0}^{(p-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3 (-1)^k + A_2 x^2 + \cdots \quad \text{for some } A_2 \in \mathbb{Z}_p.$$

By Lemma 2.8, if $p \mid A_2$, we are done. Now we follow Mortenson [2008] by using another hypergeometric evaluation identity, which is a specialization of Whipple's ${}_7F_6$ formula (see [Bailey 1935, page 28]):

$${}_6F_5\left[\begin{matrix} a, 1+\frac{1}{2}a, & b, & c, & d, & e; & -1 \\ \frac{1}{2}a, & 1+a-b, & 1+a-c, & 1+a-d, & 1+a-e & \end{matrix}\right]$$
$$= \frac{\Gamma(1+a-d)\Gamma(1+a-e)}{\Gamma(1+a)\Gamma(1+a-d-e)}\, {}_3F_2\left[\begin{matrix} 1+a-b-c, & d, & e; & 1 \\ & 1+a-b, & 1+a-c & \end{matrix}\right].$$

Letting $a=\frac{1}{2},\ b=\frac{1-x}{2},\ c=\frac{1}{2}(1+x),\ e=\frac{1}{2}(1-p)$ and $d=1$, we have

$$(13)\quad {}_6F_5\left[\begin{matrix}\frac{1}{2}, \frac{5}{4}, \frac{1}{2}(1-x), \frac{1}{2}(1+x), \frac{1}{2}(1-p), & 1; & -1\\ \frac{1}{4},\ 1+\frac{1}{2}x,\ \ 1-\frac{1}{2}x,\ \ \ \frac{1}{2}, & 1+\frac{1}{2}p & \end{matrix}\right]$$
$$=\frac{\Gamma(\frac{1}{2})\Gamma(1+\frac{1}{2}p)}{\Gamma(\frac{3}{2})\Gamma(\frac{1}{2}p)}\, {}_3F_2\left[\begin{matrix}\frac{1}{2}, & 1, & \frac{1}{2}-\frac{1}{2}p;1\\ & 1+\frac{1}{2}x, & 1-\frac{1}{2}x\end{matrix}\right].$$

Since $\Gamma(\frac{1}{2})\Gamma(1+\frac{1}{2}p)/((\Gamma(\frac{3}{2})\Gamma(\frac{1}{2}p))=p$, every x-coefficient above is in $p\mathbb{Z}_p$. Moreover, modulo p the left side of (12) is congruent to that of (13). So when we expand the left side of (12) in terms of x, the coefficients are all in $p\mathbb{Z}_p$. In particular, $p\,|\,A_2$ and this concludes the proof of (1).

4. Proofs of Theorems 1.1, 1.2, 1.4, and 1.5

Whipple [1926, (7.7)] proved that

$$(14)\quad {}_7F_6\left[\begin{matrix}a, 1+\frac{1}{2}a, & c, & d, & e, & f, & g; & 1\\ \frac{1}{2}a, & 1+a-c, & 1+a-d, & 1+a-e, & 1+a-f & 1+a-g; & \end{matrix}\right]$$
$$=\frac{\Gamma(1+a-e)\Gamma(1+a-f)\Gamma(1+a-g)\Gamma(1+a-e-f-g)}{\Gamma(1+a)\Gamma(1+a-f-g)\Gamma(1+a-e-f)\Gamma(1+a-e-f)}$$
$$\times\, {}_4F_3\left[\begin{matrix}1+a-c-d, & e, & f, & g; & 1\\ & e+f+g-a, & 1+a-c, & 1+a-d & \end{matrix}\right],$$

provided the ${}_4F_3$ is a terminating series.

Proof of Theorem 1.1. Let r be a positive integer and $p>3$ a prime. In (14), we let

$$a=\tfrac{1}{2},\quad c=\tfrac{1}{2}+i\tfrac{1}{2}p^r,\quad d=\tfrac{1}{2}-i\tfrac{1}{2}p^r,\quad e=\tfrac{1}{2}+\tfrac{1}{2}p^r,\quad f=\tfrac{1}{2}-\tfrac{1}{2}p^r,\quad g=1,$$

where $i=\sqrt{-1}$. Then following McCarthy and Osburn's argument, we know the left side of (14) is congruent to

$$\sum_{k=0}^{(p^r-1)/2}(4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^4 \mod p^{4r}$$

and the right side of (14) equals

$$\frac{\Gamma(1-\frac{1}{2}p^r)\Gamma(1+\frac{1}{2}p^r)\Gamma(-\frac{1}{2})}{\Gamma(\frac{3}{2})\Gamma(-\frac{1}{2}p^r)\Gamma(\frac{1}{2}p^r)}\, {}_4F_3\left[\begin{matrix}\frac{1}{2}, \frac{1}{2}+\frac{1}{2}p^r, & \frac{1}{2}-\frac{1}{2}p^r, & 1; & 1\\ \frac{3}{2}, & 1-i\frac{1}{2}p^r, & 1+i\frac{1}{2}p^r & \end{matrix}\right].$$

Since

$$\frac{\Gamma(1-\frac{1}{2}p^r)\Gamma(1+\frac{1}{2}p^r)\Gamma(-\frac{1}{2})}{\Gamma(\frac{3}{2})\Gamma(-\frac{1}{2}p^r)\Gamma(\frac{p^r}{2})}=p^{2r},$$

it suffices to prove

$$p^r \cdot \sum_{k=0}^{(p^r-1)/2} \frac{1}{2k+1}\left(\frac{(\frac{1}{2})_k}{k!}\right)^2 \equiv 1 \mod p^3 \quad \text{for } p > 3.$$

Recall that Lemma 2.1 says for any odd integer $n > 1$,

$$(2n+1)\sum_{k=0}^{n} \frac{(-1)^k}{2k+1}\binom{n}{k}\binom{n+k}{k} = 1.$$

Therefore, combining this identity, congruence (8), and Lemma 2.4, we have

$$\begin{aligned} p^r \cdot \sum_{k=0}^{(p^r-1)/2} \frac{1}{2k+1}\left(\frac{(\frac{1}{2})_k}{k!}\right)^2 &= p^r \cdot \sum_{k=0}^{(p^r-1)/2-1} \frac{1}{2k+1}\left(\frac{(\frac{1}{2})_k}{k!}\right)^2 + \left(\frac{(\frac{1}{2})_{(p^r-1)/2}}{(\frac{1}{2}(p^r-1))!}\right)^2 \\ &\equiv p^r \cdot \sum_{k=0}^{(p^r-1)/2-1} \frac{(-1)^k}{2k+1}\binom{\frac{1}{2}(p^r-1)}{k}\binom{\frac{1}{2}(p^r-1)+k}{k} \\ &\qquad + (-1)^{(p^r-1)/2}\binom{p^r-1}{\frac{1}{2}(p^r-1)} \mod p^3 \\ &\equiv 1 \mod p^3. \end{aligned}$$

□

Proof of Theorem 1.2. In (14), take

$$a = \tfrac{1}{2}, \quad c = \tfrac{1}{2} + i\tfrac{1}{2}p, \quad d = \tfrac{1}{2} - i\tfrac{1}{2}p, \quad e = \tfrac{1}{2} - \tfrac{1}{2}p, \quad f = \tfrac{1}{2} + \tfrac{1}{2}p, \quad g = \tfrac{1}{2} - p^4.$$

Then the left side of (14) is congruent to

$$\sum_{k=0}^{(p-1)/2} (4k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^6 \mod p^4.$$

Meanwhile, the right side of (14) is congruent to

$$\begin{aligned} &\frac{\Gamma(1-\frac{1}{2}p)\Gamma(1+\frac{1}{2}p)}{\Gamma(\frac{1}{2})\Gamma(\frac{3}{2})} \frac{\Gamma(1+p^4)\Gamma(p^4)}{\Gamma(\frac{1}{2}+\frac{1}{2}p+p^4)\Gamma(\frac{1}{2}-\frac{1}{2}p+p^4)} \\ &\qquad \times \sum_{k=0}^{(p-1)/2} \frac{(\frac{1}{2})_k^2(\frac{1}{2}+\frac{1}{2}p)_k(\frac{1}{2}-\frac{1}{2}p)_k}{k!^2(1-i\frac{1}{2}p)_k(1+i\frac{1}{2}p)_k} \mod p^4, \end{aligned}$$

where

$$\frac{\Gamma(1-\frac{1}{2}p)\Gamma(1+\frac{1}{2}p)}{\Gamma(\frac{1}{2})\Gamma(\frac{3}{2})} = (-1)^{(p-1)/2} p$$

and

$$\frac{\Gamma(1+p^4)\Gamma(p^4)}{\Gamma(\frac{1}{2}+\frac{1}{2}p+p^4)\Gamma(\frac{1}{2}-\frac{1}{2}p+p^4)} = \frac{(p^4-\frac{1}{2}(p-1))_{(p-1)/2}}{(1+p^4)_{(p-1)/2}}$$
$$\equiv \frac{(-\frac{1}{2}(p-1))(-\frac{1}{2}(p-1)+1)\cdots(-1)}{1\cdot 2\cdots(\frac{1}{2}(p-1))} \mod p = (-1)^{(p-1)/2}.$$

Therefore, Theorem 1.2 follows from the result of Kilbourn (see (3)) and the next lemma. □

Lemma 4.1. *Let $p>3$ be a prime, then*

$$\sum_{k=0}^{(p-1)/2} \frac{(\frac{1}{2})_k^2(\frac{1}{2}+\frac{1}{2}p)_k(\frac{1}{2}-\frac{1}{2}p)_k}{k!^2(1-i\frac{1}{2}p)_k(1+i\frac{1}{2}p)_k} \equiv \sum_{k=0}^{(p-1)/2}\left(\frac{(\frac{1}{2})_k}{k!}\right)^4 \mod p^3.$$

Proof. Expand

$$\sum_{k=0}^{(p-1)/2} \frac{(\frac{1}{2})_k^2(\frac{1}{2}+\frac{1}{2}x)_k(\frac{1}{2}-\frac{1}{2}x)_k}{k!^2(1-i\frac{1}{2}x)_k(1+i\frac{1}{2}x)_k} = \sum_{k=0}^{(p-1)/2}\left(\frac{(\frac{1}{2})_k}{k!}\right)^4(1+b_{2,k}x^2+b_{4,k}x^4+\cdots).$$

Using (8) and (9), we have

$$b_{2,k} = -\sum_{j=1}^{k}\frac{1}{(2j-1)^2} - \frac{1}{4}\sum_{j=1}^{k}\frac{1}{j^2} = -\sum_{j=1}^{2k}\frac{1}{j^2}.$$

The claim is verified by using Lemma 2.8 and taking $s=2$ in Lemma 2.7. □

Proof of Theorem 1.4. We start with the following combinatorial identity.

Lemma 4.2.
$$\sum_{k=0}^{(p-1)/2}(6k+1)\frac{(\frac{1}{2})_k(\frac{1}{2}-\frac{1}{2}p)_k(\frac{1}{2}+\frac{1}{2}p)_k}{(1)_k(1+\frac{1}{4}p)_k(1-\frac{1}{4}p)_k}\frac{1}{4^k} = (-1)^{(p-1)/2}p.$$

Proof. Recall that [Gessel 1995, (31.1)] says

$${}_5F_4\left[\begin{matrix}\frac{1}{2}+a-c, & -n, & n+1, & 2-2c+n, & \frac{5}{3}-\frac{2}{3}c+\frac{1}{3}n; \frac{1}{4}\\ & 2-c+n, & \frac{2}{3}-\frac{2}{3}c+\frac{1}{3}n, & n-2a+2, & \frac{3}{2}-c\end{matrix}\right]$$
$$= \frac{(2-c)_n(2-2a)_n}{(3-2c)_n(\frac{3}{2}-a)_n}.$$

Letting $a=\frac{1}{2}+\frac{1}{4}p$, $c=\frac{1}{2}+\frac{1}{4}p$, and $n=\frac{1}{2}(p-1)$ and using Lemma 2.2, we have

$${}_5F_4\left[\begin{matrix}\frac{1}{2}, \frac{1}{2}, \frac{7}{6}, \frac{1}{2}-\frac{1}{2}p, \frac{1}{2}+\frac{1}{2}p; \frac{1}{4}\\ \frac{1}{2}, \frac{1}{6}, 1-\frac{1}{4}p, \ 1+\frac{1}{4}p\end{matrix}\right] = \frac{(\frac{3}{2}-\frac{1}{4}p)_{(p-1)/2}(1-\frac{1}{2}p)_{(p-1)/2}}{(2-\frac{1}{2}p)_{(p-1)/2}(1-\frac{1}{4}p)_{(p-1)/2}}$$
$$= (-1)^{(p-1)/2}p.$$ □

Lemma 4.3. *The function*

$$\left(\sum_{k=0}^{(p-1)/2}(6k+1)\frac{(\frac{1}{2})_k(\frac{1}{2}-\frac{1}{2}x)_k(\frac{1}{2}+\frac{1}{2}x)_k}{(1)_k(1+\frac{1}{4}x)_k(1-\frac{1}{4}x)_k}\frac{1}{4^k}\right)\Big/\left(\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\left(\frac{(\frac{1}{2})_k}{k!}\right)^3\right)$$

is a formal power series in x^2 with coefficients in $\mathbb{Z}_p$. Its x^2 coefficient is zero modulo p.

Proof. We use the *strange* valuation of Gosper:

$${}_5F_4\left[\begin{matrix}2a, & 2b, & 1-2b, & 1+\frac{2}{3}a, & -n; & \frac{1}{4}\\ & a+b-1, & a+b+\frac{1}{2}, & \frac{2}{3}a, & 1+2a+2n & \end{matrix}\right]=\frac{(a+\frac{1}{2})_n(a+1)_n}{(a+b+\frac{1}{2})_n(a-b+1)_n}.$$

See [Gessel and Stanton 1982, (1.2)]. Let $a=\frac{1}{4}$, $b=\frac{1}{4}-\frac{1}{4}x$ and $n=\frac{1}{2}(p-1)$. Then the left side of the above equals

(15)
$${}_5F_4\left[\begin{matrix}\frac{1}{2}, \frac{1}{2}-\frac{1}{2}x, \frac{1}{2}+\frac{1}{2}x, & \frac{7}{6}, & \frac{1}{2}-\frac{1}{2}p; \frac{1}{4}\\ \frac{1}{2}+p, & \frac{1}{6}, & 1-\frac{1}{4}x, \; 1+\frac{1}{4}x\end{matrix}\right]=\frac{(\frac{3}{4})_{(p-1)/2}(\frac{5}{4})_{(p-1)/2}}{(1-\frac{1}{4}x)_{(p-1)/2}(1+\frac{1}{4}x)_{(p-1)/2}}.$$

We remark that

(16)
$${}_5F_4\left[\begin{matrix}\frac{1}{2}, \frac{1}{2}-\frac{1}{2}x, \frac{1}{2}+\frac{1}{2}x, & \frac{7}{6}, & \frac{1}{2}-\frac{1}{2}p; \frac{1}{4}\\ \frac{1}{2}+p, & \frac{1}{6}, & 1-\frac{1}{4}x, \; 1+\frac{1}{4}x\end{matrix}\right]$$
$$\equiv\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\frac{(\frac{1}{2})_k(\frac{1}{2}-\frac{1}{2}x)_k(\frac{1}{2}+\frac{1}{2}x)_k}{(1)_k(1+\frac{1}{4}x)_k(1-\frac{1}{4}x)_k} \mod p.$$

When $x=0$, the right hand side of (15) equals $(\frac{3}{4})_{(p-1)/2}(\frac{5}{4})_{(p-1)/2}/(1)^2_{(p-1)/2}$, which is in $p\mathbb{Z}_p$. In fact, if $p\equiv 1 \bmod 4$ then $\frac{5}{4}+\frac{1}{4}(p-1)-1=\frac{1}{4}p$, and if $p\equiv 3 \bmod 4$, then $\frac{3}{4}+\frac{1}{4}(p-3)=\frac{1}{4}p$, while $(1)_{(p-1)/2}$ is a p-adic unit. It is not difficult to see that p divides $((3)/4)_{(p-1)/2}(\frac{5}{4})_{(p-1)/2}/(1)^2_{(p-1)/2}$ exactly. Consequently, if we expand

$${}_5F_4\left[\begin{matrix}\frac{1}{2}, \frac{1}{2}-\frac{1}{2}x, \frac{1}{2}+\frac{1}{2}x, & \frac{7}{6}, & \frac{1}{2}-\frac{1}{2}p; \frac{1}{4}\\ \frac{1}{2}+p, & \frac{1}{6}, & 1-\frac{1}{4}x, \; 1+\frac{1}{4}x\end{matrix}\right]$$

in terms of formal power series of x (in fact, x^2), each coefficient is in $p\mathbb{Z}_p$. Thus the coefficients of the right side of (16), including the coefficient of x^2, are all divisible by p. By Lemmas 2.8 and 4.2,

$$\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\left(\frac{(\frac{1}{2})_k}{k!}\right)^3\equiv(-1)^{(p-1)/2}p \mod p^3.$$

Namely,

$$\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\left(\frac{(\frac12)_k}{k!}\right)^3=(-1)^{(p-1)/2}p+ap^3 \quad \text{for some } a\in\mathbb{Z}_p.$$

The statement of Theorem 1.4 is equivalent to $a\in p\mathbb{Z}_p$.

The quotient

$$\left(\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\frac{(\frac12)_k(\frac12-\frac12x)_k(\frac12+\frac12x)_k}{(1)_k(1+\frac14x)_k(1-\frac14x)_k}\right)\Bigg/\left(\sum_{k=0}^{(p-1)/2}\frac{6k+1}{4^k}\frac{(\frac12)_k}{(1)_k}\right) \tag{17}$$

is a formal power series in x^2 with p-integral coefficients, since the denominators are divisible by p exactly. The same conclusion applies to

$$\begin{aligned}&{}_5F_4\left[\begin{matrix}\frac12, \frac12-\frac12x, \frac12+\frac12x, & \frac76, & \frac12-\frac12p;\frac14\\ \frac12+p, & \frac16, & 1-\frac14x,\ 1+\frac14x\end{matrix}\right]\Bigg/{}_5F_4\left[\begin{matrix}\frac12, & \frac12, & \frac12, & \frac76, & \frac12-\frac12p;\frac14\\ \frac12+p, & \frac16, & 1-\frac14x, & 1+\frac14x\end{matrix}\right]\\
&={}_5F_4\left[\begin{matrix}\frac12, \frac12-\frac12x, \frac12+\frac12x, & \frac76, & \frac12-\frac12p;\frac14\\ \frac12+p, & \frac16, & 1-\frac14x,\ 1+\frac14x\end{matrix}\right]\Bigg/\left(\frac{(\frac34)_{(p-1)/2}(\frac54)_{(p-1)/2}}{(1)^2_{(p-1)/2}}\right)\\
&=\frac{(1)^2_{(p-1)/2}}{(1-\frac14x)_{(p-1)/2}(1+\frac14x)_{(p-1)/2}}.\end{aligned}$$

On the other hand, by (9), the x^2 coefficient of

$$\frac{(1)^2_{(p-1)/2}}{(1-\frac14x)_{(p-1)/2}(1+\frac14x)_{(p-1)/2}}$$

is a scalar multiple of $H^{(2)}_{(p-1)/2}$, which is in $p\mathbb{Z}_p$ by Lemma 2.5; so is the x^2 coefficient of (17). □

By Lemma 2.8 and the analysis above,

$$\frac{(-1)^{(p-1)/2}p}{(-1)^{(p-1)/2}p+ap^3}=\frac{(-1)^{(p-1)/2}}{(-1)^{(p-1)/2}+ap^2}\equiv 1 \mod p^3;$$

hence $a\in p\mathbb{Z}_p$, which concludes the proof of Theorem 1.4. □

Lemma 4.4.

$$\sum_{k=0}^{(p-1)/2}(6k+1)\frac{(\frac12)_k(\frac12-\frac12p)_k(\frac12+\frac12p)_k}{(1)_k(1+\frac14p)_k(1-\frac14p)_k}\frac{(-1)^k}{8^k}=(-1)^{(p^2-1)/8+(p-1)/2}p.$$

Proof. This time, we use [Gessel 1995, last identity of page 544]

$$ {}_4F_3\left[\begin{matrix} 2a+n+1, & n+1, & \frac{2}{3}a+\frac{1}{3}n+\frac{4}{3}, & -n; -\frac{1}{8} \\ & a+\frac{3}{2}+n, & \frac{2}{3}a+\frac{1}{3}n+\frac{1}{3}, & 1+a \end{matrix}\right] = \frac{(a+\frac{3}{2})_n}{(2a+2)_n}2^n. $$

Letting $a=-\frac{1}{4}p$ and $n=\frac{1}{2}(p-1)$ and using Lemma 2.3, we have

$$ \begin{aligned} {}_4F_3\left[\begin{matrix} \frac{1}{2}, \frac{7}{6}, \frac{1}{2}+\frac{1}{2}p, \frac{1}{2}-\frac{1}{2}p; -\frac{1}{8} \\ \frac{1}{6}, 1-\frac{1}{4}p, \ 1+\frac{1}{4}p \end{matrix}\right] &= \frac{(\frac{3}{2}-\frac{1}{4}p)_{(p-1)/2}}{(2-\frac{1}{2}p)_{(p-1)/2}}2^{(p-1)/2} \\ &= (-1)^{(p^2-1)/8+(p-1)/2}p. \end{aligned} $$ □

Proof of Theorem 1.5. Equation (5) is a consequence of Lemma 4.4. □

Remark 1. Van Hamme's conjecture that

$$ \sum_{k=0}^{(p-1)/2}(6k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3\frac{(-1)^k}{8^k} \equiv (-1)^{(p^2-1)/8+(p-1)/2}p \mod p^3 $$

holds if

$$ \sum_{k=0}^{(p-1)/2}(6k+1)\left(\frac{(\frac{1}{2})_k}{k!}\right)^3\left(\sum_{j=1}^{k}\frac{1}{(2j-1)^2}-\frac{1}{16}\sum_{j=1}^{k}\frac{1}{j^2}\right)\frac{(-1)^k}{8^k} \equiv 0 \mod p. $$

The proof of the latter is left to the interested reader.

Remark 2. In [2009], Zudilin proved the congruence (2) modulo p^2 and the congruence (5) modulo p.

Acknowledgments

I would like to thank Heng Huat Chan and Wadim Zudilin for their encouragement, enlightening discussions and valuable comments. In particular, Zudilin pointed me to a few useful ideas and the reference [Gessel and Stanton 1982]. I further thank the anonymous referees for their detailed comments, including pointing out the reference [Cai 2002], on an earlier version of this article.

References

[Ahlgren and Ono 2000] S. Ahlgren and K. Ono, "A Gaussian hypergeometric series evaluation and Apéry number congruences", *J. Reine Angew. Math.* **518** (2000), 187–212. MR 2001c:11057 Zbl 0940.33002

[Bailey 1935] W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Mathematics and Mathematical Physics **32**, Cambridge University Press, London, 1935. MR 32 #2625 Zbl 0011.02303

[Cai 2002] T. Cai, "A congruence involving the quotients of Euler and its applications, I", *Acta Arith.* **103**:4 (2002), 313–320. MR 2003d:11007 Zbl 1008.11001

[Chan et al. 2010] H. H. Chan, L. Long, and W. Zudilin, "A supercongruence motivated by the Legendre family of elliptic curves", *Mat. Zametki* **88**:4 (2010), 621–625. In Russian; translated in *Mathematical Notes* **88**:4 (2010), 599–602.

[Gessel 1995] I. M. Gessel, "Finding identities with the WZ method", *J. Symbolic Comput.* **20**:5-6 (1995), 537–566. MR 97j:05010 Zbl 0908.33004

[Gessel and Stanton 1982] I. Gessel and D. Stanton, "Strange evaluations of hypergeometric series", *SIAM J. Math. Anal.* **13**:2 (1982), 295–308. MR 83c:33002 Zbl 0486.33003

[van Hamme 1997] L. van Hamme, "Some conjectures concerning partial sums of generalized hypergeometric series", pp. 223–236 in *p-adic functional analysis* (Nijmegen, 1996), edited by W. H. Schikhof et al., Lecture Notes in Pure and Appl. Math. **192**, Dekker, New York, 1997. MR 98k:33011 Zbl 0895.11051

[Kilbourn 2006] T. Kilbourn, "An extension of the Apéry number supercongruence", *Acta Arith.* **123**:4 (2006), 335–348. MR 2007e:11049 Zbl 1170.11008

[McCarthy 2009] D. McCarthy, "Supercongruence conjectures of Rodriguez–Villegas", preprint, 2009. arXiv 0907.5089

[McCarthy and Osburn 2008] D. McCarthy and R. Osburn, "A p-adic analogue of a formula of Ramanujan", *Arch. Math.* (*Basel*) **91**:6 (2008), 492–504. MR 2009k:11191 Zbl 1175.33004

[Morley 1895] F. Morley, "Note on the congruence $2^{4n} \equiv (-1)^n (2n)!/(n!)^2$, where $2n+1$ is prime", *Annals of Math.* **9** (1895), 168–170.

[Mortenson 2008] E. Mortenson, "A p-adic supercongruence conjecture of van Hamme", *Proc. Amer. Math. Soc.* **136**:12 (2008), 4321–4328. MR 2010g:11203 Zbl 1171.11061

[Rodriguez-Villegas 2003] F. Rodriguez-Villegas, "Hypergeometric families of Calabi–Yau manifolds", pp. 223–231 in *Calabi–Yau varieties and mirror symmetry* (Toronto, ON, 2001), edited by N. Yui and J. D. Lewis, Fields Inst. Commun. **38**, Amer. Math. Soc., Providence, RI, 2003. MR 2005b:11086 Zbl 1062.11038

[Whipple 1926] F. J. W. Whipple, "On well-posed series, generalised hypergeometric series having parameters in pairs, each pair with the same sum", *Proc. London Math. Soc.* **24**:1 (1926), 247–263.

[Zudilin 2009] W. Zudilin, "Ramanujan-type supercongruences", *J. Number Theory* **129**:8 (2009), 1848–1857. MR 2522708 Zbl 05569026

Received December 1, 2009. Revised September 27, 2010.

LING LONG
DEPARTMENT OF MATHEMATICS
IOWA STATE UNIVERSITY
396 CARVER HALL
AMES, IA 50011
UNITED STATES

linglong@iastate.edu
http://www.math.iastate.edu/linglong

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

NECESSARY AND SUFFICIENT CONDITIONS FOR UNIT GRAPHS TO BE HAMILTONIAN

H. R. MAIMANI, M. R. POURNAKI AND S. YASSEMI

The unit graph corresponding to an associative ring R is the graph obtained by setting all the elements of R to be the vertices and defining distinct vertices x and y to be adjacent if and only if $x + y$ is a unit of R. By a constructive method, we derive necessary and sufficient conditions for unit graphs to be Hamiltonian.

1. Introduction

A graph is *Hamiltonian* if it has a cycle that visits every vertex exactly once; such a cycle is called a *Hamiltonian cycle*. In general, the problem of finding a Hamiltonian cycle in a given graph is an *NP*-complete problem and a special case of the traveling salesman problem. It is a problem in combinatorial optimization studied in operations research and theoretical computer science; see [Garey and Johnson 1979]. The only known way to determine whether a given graph has a Hamiltonian cycle is to undertake an exhaustive search, and until now no theorem giving a necessary and sufficient condition for a graph to be Hamiltonian was known. The study of Hamiltonian graphs has long been an important topic. See [Gould 2003] for a survey, updating earlier surveys in this area.

Let n be a positive integer, and let $\mathbb{Z}_n$ be the ring of integers modulo n. Grimaldi [1990] defined a graph $G(\mathbb{Z}_n)$ based on the elements and units of $\mathbb{Z}_n$. The vertices of $G(\mathbb{Z}_n)$ are the elements of $\mathbb{Z}_n$, and distinct vertices x and y are defined to be adjacent if and only if $x + y$ is a unit of $\mathbb{Z}_n$. For a positive integer m, it follows that $G(\mathbb{Z}_{2m})$ is a $\varphi(2m)$-regular graph, where φ is the Euler phi function. In case $m \geq 2$, the graph $G(\mathbb{Z}_{2m})$ can be expressed as the union of $\varphi(2m)/2$ Hamiltonian cycles. The odd case is not quite so easy, but the structure is clear and the results are similar to the even case. We recall that a *cone* over a graph is obtained by taking

The research of H. R. Maimani and S. Yassemi was in part supported by a grant from IPM (numbers 89050211 and 89130213). The research of M. R. Pournaki was in part supported by a grant from the Academy of Sciences for the Developing World (TWAS–UNESCO Associateship — Ref. FR3240126591).
MSC2000: primary 05C45; secondary 13M05.
Keywords: Hamiltonian cycle, Hamiltonian graph, finite ring.

the categorical product of the graph and a path with a loop at one end, and then identifying all the vertices whose second coordinate is the other end of the path. When p is an odd prime, $G(\mathbb{Z}_p)$ can be expressed as a cone over a complete partite graph with $(p-1)/2$ partitions of size two. This leads to an explicit formula for the chromatic polynomial of $G(\mathbb{Z}_p)$. Grimaldi [1990] also concludes with some properties of the graphs $G(\mathbb{Z}_{p^m})$, where p is a prime number and $m \geq 2$. Recently, the authors of this paper generalized $G(\mathbb{Z}_n)$ to $G(R)$, the unit graph of R, where R is an arbitrary associative ring with nonzero identity and studied the properties of this graph; see [Ashrafi et al. 2010; Maimani et al. 2010].

By a constructive method, we derive necessary and sufficient conditions for unit graphs to be Hamiltonian.

2. Preliminaries and the main result

Throughout the paper, by a graph we mean a finite undirected graph without loops or multiple edges. Also all rings are finite commutative with nonzero identity. For undefined terms and concepts, see [West 1996; Atiyah and Macdonald 1969].

We first start with recalling some notions from graph theory. For a graph G and for any two vertices x and y of G, we recall that a *walk* between x and y is a sequence $x = v_0, e_1, v_1, \ldots, e_k, v_k = y$ of vertices and edges of G, denoted by

$$x = v_0 \to v_1 \to \cdots \to v_k = y,$$

such that for every i with $1 \leq i \leq k$, the edge e_i has endpoints v_{i-1} and v_i. Also a *path* between x and y is a walk between x and y without repeated vertices. A *cycle* of a graph is a path such that the start and end vertices are the same. Two cycles are considered the same if they consist of the same vertices and edges. The number of edges (counting repeats) in a walk, path or a cycle, is called its *length*. A *Hamiltonian path* (*cycle*) in G is a path (cycle) in G that visits every vertex exactly once. A graph is called *Hamiltonian* if it contains a Hamiltonian cycle. Also a graph G is called *connected* if for any vertices x and y of G there is a path between x and y.

We now define the unit graph corresponding to a ring. Let R be a ring and $U(R)$ be the set of unit elements of R. The *unit graph* of R, denoted by $G(R)$, is the graph obtained by setting all the elements of R to be the vertices and defining distinct vertices x and y to be adjacent if and only if $x + y \in U(R)$. The graphs in Figure 1 are the unit graphs of the rings indicated. It is easy to see that, for given rings R and S, if $R \cong S$ as rings, then $G(R) \cong G(S)$ as graphs. This point is illustrated in Figure 2.

We continue this section by collecting some notions from ring theory. First of all, for a given ring R, the *Jacobson radical* $J(R)$ of R is defined to the intersection

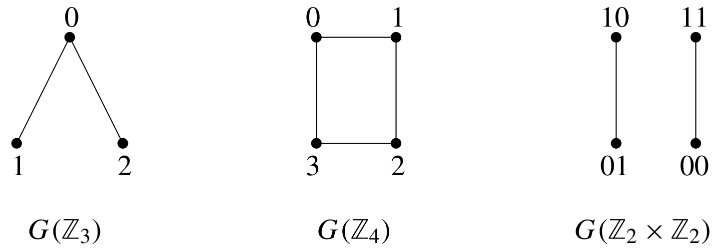


Figure 1. Unit graphs of some specific rings.

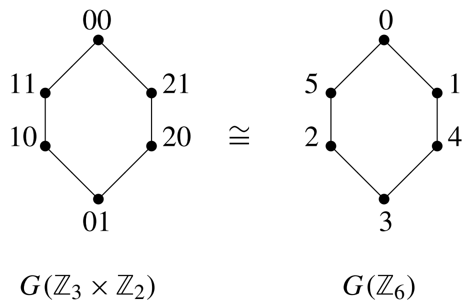


Figure 2. Unit graphs of two isomorphic rings.

of all maximal ideals of R. Let R be a ring and let k be a positive integer. An element $r \in R$ is said to be *k-good* if we may write $r = u_1 + \cdots + u_k$, where $u_1, \ldots, u_k \in U(R)$. The ring R is said to be *k-good* if every element of R is k-good. Following [Goldsmith et al. 1998], we now define an invariant of a ring, called the unit sum number, which expresses in a fairly precise way how the units generate the ring. The *unit sum number* $\boldsymbol{u}(R)$ of R is given by

- $\min\{k \mid R \text{ is } k\text{-good}\}$ if R is k-good for some $k \geq 1$,
- ω if R is not k-good for every k, but every element of R is k-good for some k (that is, when at least $U(R)$ generates R additively), and
- ∞ otherwise (that is, when $U(R)$ does not generate R additively).

For example, let D be a division ring. If $|D| \geq 3$, then $\boldsymbol{u}(D) = 2$; whereas if $|D| = 2$, that is, $D = \mathbb{Z}_2$, the field of two elements, then $\boldsymbol{u}(\mathbb{Z}_2) = \omega$. We have also $\boldsymbol{u}(\mathbb{Z}_2 \times \mathbb{Z}_2) = \infty$ — see [Ashrafi and Vámos 2005] for unit sum numbers of some other rings. The topic of unit sum numbers seems to have arisen with a paper by Zelinsky [1954], in which he shows that if V is any finite- or infinite-dimensional vector space over a division ring D, then every linear transformation is the sum of two automorphisms unless $\dim V = 1$ and D is the field of two elements. Interest in this topic increased recently after Goldsmith, Pabst and Scott [1998] defined the unit sum number. For additional historical background, see [Vámos 2005], which also contains references to recent work in this area.

We are now ready to state the main result of this paper. The proof is given in Section 3 by a sequence of lemmas and propositions.

Theorem 2.1. *Let R be a ring such that $R \not\cong \mathbb{Z}_2$ and $R \not\cong \mathbb{Z}_3$. Then the following statements are equivalent*:

(a) *The unit graph $G(R)$ is Hamiltonian.*

(b) *The ring R cannot have $\mathbb{Z}_2 \times \mathbb{Z}_2$ as a quotient.*

(c) *The ring R is generated by its units.*

(d) *The unit sum number of R is less than or equal to ω.*

(e) *The unit graph $G(R)$ is connected.*

3. The proofs

In this section we state and prove some lemmas that will be used in the proof of Theorem 2.1. For the convenience of the reader we state without proof a few known results in the form of propositions that will be used in the proofs. We also recall some definitions and notations for later use.

A *bipartite* graph is one whose vertex-set is partitioned into two (not necessarily nonempty) disjoint subsets so that the two end vertices for each edge lie in distinct partitions. Among bipartite graphs, a *complete bipartite* graph is one in which each vertex is joined to every vertex that is not in the same partition. The complete bipartite graph with two partitions of size m and n is denoted by $K_{m,n}$.

The following result characterizes the complete bipartite unit graphs of rings.

Proposition 3.1 [Ashrafi et al. 2010, Theorem 3.5]. *Let R be a ring and $\mathfrak{m}$ be a maximal ideal of R such that $|R/\mathfrak{m}| = 2$. Then $G(R)$ is a bipartite graph. The unit graph $G(R)$ is a complete bipartite graph if and only if R is a local ring.*

The degrees of all vertices of a unit graph is given by the following result. For a graph G and for a vertex x of G, the *degree* $\deg(x)$ of x is the number of edges of G incident with x.

Proposition 3.2 [Ashrafi et al. 2010, Proposition 2.4]. *Let R be a ring. Then the following statements hold for the unit graph of R*:

(1) *If* $2 \notin U(R)$, *then* $\deg(x) = |U(R)|$ *for every* $x \in R$.

(2) *If* $2 \in U(R)$, *then* $\deg(x) = |U(R)| - 1$ *for every* $x \in U(R)$ *and* $\deg(x) = |U(R)|$ *for every* $x \in R \setminus U(R)$.

We also need the following well known result due to Dirac, which initiated the study of Hamiltonian graphs. This work was continued by Ore [1960].

Proposition 3.3 [Dirac 1952, Theorem 3]. *If G is a graph with n vertices, $n \geq 3$, and every vertex has degree at least $n/2$, then G is Hamiltonian.*

Lemma 3.4. *Let R be a local ring with $|R| \geq 4$. Then the unit graph $G(R)$ is Hamiltonian.*

Proof. Suppose $\mathfrak{m}$ is the unique maximal ideal of R. There are two possibilities: either $|R/\mathfrak{m}| = 2$ or $|R/\mathfrak{m}| > 2$.

First, suppose that $|R/\mathfrak{m}| = 2$. In this case, Proposition 3.1 implies that the unit graph $G(R)$ is a complete bipartite graph. Moreover, its proof shows that $\mathfrak{m}$ and $R \setminus \mathfrak{m}$ are the partite sets of $G(R)$. Since $|R/\mathfrak{m}| = 2$, we conclude that $|\mathfrak{m}| = |R \setminus \mathfrak{m}|$ and so $G(R) \cong K_{|\mathfrak{m}|,|\mathfrak{m}|}$. The assumptions $|R| \geq 4$ and $|R/\mathfrak{m}| = 2$ imply that $|\mathfrak{m}| \geq 2$ and thus $G(R)$ is Hamiltonian.

Second, suppose that $|R/\mathfrak{m}| > 2$. In this case, Proposition 3.2 implies that $\deg(x) \geq |U(R)| - 1$ for all $x \in R$. We claim that $|U(R)| - 1 \geq |R|/2$. To show this, note that R is a local ring with $|R| \geq 4$. If $|R| = 4$, then the assumption $|R/\mathfrak{m}| > 2$ implies that $|\mathfrak{m}| < 2$ and so $\mathfrak{m} = 0$. Therefore R is a field and so $|U(R)| = 3$. Thus $|U(R)| - 1 = 2 = |R|/2$. If $|R| = 5$, then R is again a field and so $|U(R)| = 4$. Thus $|U(R)| - 1 = 3 > 2.5 = |R|/2$. If $|R| \geq 6$, then since R is local with $|R/\mathfrak{m}| > 2$, we conclude that $|U(R)| \geq 2|R|/3$. Therefore $|U(R)| - 1 \geq (2|R|/3) - 1 \geq |R|/2$. Thus the claim holds and so $\deg(x) \geq |R|/2$ for every $x \in R$. Therefore Proposition 3.3 implies that $G(R)$ is Hamiltonian. □

The following result gives us information about the existence of a Hamiltonian cycle in unit graphs of the direct product of a ring and a field.

Lemma 3.5. *Let T be a ring with Hamiltonian unit graph and let F be a field. If $F \ncong \mathbb{Z}_2$, then the unit graph $G(T \times F)$ is Hamiltonian.*

Proof. Since the unit graph $G(T)$ is Hamiltonian, there is a Hamiltonian cycle with length $n = |T|$ in $G(T)$, say

$$0 = a_1 \to a_2 \to \cdots \to a_{n-1} \to a_n \to a_{n+1} = 0.$$

Either the characteristic of F is equal to 2 or it is not.

First, suppose the latter. In this case we may assume that

$$F = \{0, x_1, \ldots, x_{(|F|-1)/2}, -x_1, \ldots, -x_{(|F|-1)/2}\}.$$

If n is even and $|F| \geq 5$, then $x_2 \neq -x_1$ and so $x_1 + x_2$ is a unit element of F. Now consider the following paths in the unit graph $G(T \times F)$:

$$\begin{aligned}
P_0 &: (0,0) \to (a_2, x_1) \to (a_3, 0) \to (a_4, x_1) \to \cdots \to (a_n, x_1),\\
P_1 &: (0, x_2) \to (a_2, 0) \to (a_3, x_2) \to \cdots \to (a_{n-1}, x_2) \to (a_n, 0),\\
P_2 &: (0, x_1) \to (a_2, x_2) \to (a_3, x_1) \to \cdots \to (a_n, x_2).
\end{aligned}$$

Also for every i with $3 \leq i \leq (|F|-1)/2$, consider the path

$$P_i : (0, x_i) \to (a_2, x_i) \to (a_3, x_i) \to \cdots \to (a_n, x_i),$$

and for every i with $1 \leq i \leq (|F|-1)/2$, consider the path

$$P'_i : (0, -x_i) \to (a_2, -x_i) \to \cdots \to (a_n, -x_i).$$

It is easy to see that P_{i-1} is adjacent to P_i for every i with $1 \leq i \leq (|F|-1)/2$ and P'_{i-1} is adjacent to P'_i for every i with $2 \leq i \leq (|F|-1)/2$, and $P_{(|F|-1)/2}$ is adjacent to P'_1. Therefore $P_0 P_1 P_2 P_3 \cdots P_{(|F|-1)/2} P'_1 \cdots P'_{(|F|-1)/2}(0, 0)$ is a Hamiltonian cycle in the unit graph $G(T \times F)$, which shows that it is Hamiltonian. If n is even and $|F| = 3$, then $F \cong \mathbb{Z}_3$ and thus the cycle

$$\begin{aligned}(a_1, 1) &\to (a_2, 0) \to (a_3, 2) \to (a_4, 2) \to (a_3, 0)\\ &\to (a_2, 1) \to (a_3, 1) \to \cdots \to (a_{n-2}, 1) \to (a_{n-1}, 1)\\ &\to (a_1, 2) \to (a_2, 2) \to (a_1, 0) \to (a_n, 1) \to (a_1, 1),\end{aligned}$$

is a Hamiltonian cycle in the unit graph $G(T \times F)$, and thus it is Hamiltonian.

If n is odd and $|F| \geq 5$, consider the path

$$P_0 : (a_1, 0) \to (a_2, x_1) \to \cdots \to (a_n, 0) \to (a_1, x_1) \to (a_2, 0) \to \cdots \to (a_n, x_1),$$

and for $1 \leq i \leq (|F|-1)/2$ consider the paths

$$\begin{aligned}P_i &: (a_1, x_i) \to (a_2, x_i) \to \cdots \to (a_n, x_i),\\ P'_i &: (a_1, -x_i) \to (a_2, -x_i) \to \cdots \to (a_n, -x_i).\end{aligned}$$

It is easy to see that $P_0 P_1 \cdots P_{(|F|-1)/2} P'_1 \cdots P'_{(|F|-1)/2}(a_1, 0)$ is a Hamiltonian cycle in the unit graph $G(T \times F)$ and thus it is Hamiltonian. If n is odd and $|F| = 3$, we may obtain a Hamiltonian cycle in the unit graph $G(T \times F)$ by replacing the eleven end-vertices in the cycle above with

$$\begin{aligned}(a_{n-3}, 1) &\to (a_{n-2}, 1) \to (a_{n-1}, 0) \to (a_n, 2) \to (a_{n-1}, 1)\\ &\to (a_n, 1) \to (a_1, 0) \to (a_2, 2) \to (a_1, 2) \to (a_n, 0) \to (a_1, 1).\end{aligned}$$

This shows that the unit graph $G(T \times F)$ is Hamiltonian.

Second, suppose that characteristic of F is equal to 2. Therefore we have $|F| \geq 4$. In this case we may assume that

$$F = \{x_1, \ldots, x_{2^m}\} = \{x_{2i-1}, x_{2i} \mid 1 \leq i \leq 2^{m-1}\}.$$

If n is even, then for every i with $1 \leq i \leq 2^{m-1}$, consider the following paths in the unit graph $G(T \times F)$:

$$\begin{aligned}P_i &: (a_1, x_{2i-1}) \to (a_2, x_{2i}) \to \cdots \to (a_n, x_{2i}),\\ P'_i &: (a_1, x_{2i}) \to (a_2, x_{2i-1}) \to \cdots \to (a_n, x_{2i-1}).\end{aligned}$$

Since $|F| \geq 4$, it is clear that $P_1 P'_{2^{m-1}} P_2 P'_{2^{m-1}-1} \cdots P_{2^{m-1}} P'_1 (0, x_1)$ is a Hamiltonian cycle in the unit graph $G(T \times F)$ and thus it is Hamiltonian.

If n is odd, then consider the path

$$P_i : (a_1, x_{2i-1}) \to (a_2, x_{2i}) \to \cdots \to (a_{n-1}, x_{2i}) \to (a_n, x_{2i-1}) \to \cdots \to (a_n, x_{2i}).$$

Therefore $P_1 P_2 \cdots P_{2^{m-1}} (a_1, x_1)$ is a Hamiltonian cycle in the unit graph $G(T \times F)$ and thus it is Hamiltonian. □

In the sequel we need Lemmas 3.7, 3.8 and 3.10. But first, we state the following proposition, which is useful in the proof of Lemma 3.7. Recall that a *clique* of a graph G is a complete subgraph of G. Also a *coclique* (also called an *independent set of vertices*) in a graph G is a set of pairwise nonadjacent vertices.

Proposition 3.6 [Ashrafi et al. 2010, Lemma 2.7]. *Let R be a ring and suppose that $J(R)$ denotes the Jacobson radical of R. Suppose $x, y \in R$.*

(a) *If $x + J(R)$ and $y + J(R)$ are adjacent in the unit graph $G(R/J(R))$, then every element of $x + J(R)$ is adjacent to every element of $y + J(R)$ in the unit graph $G(R)$.*

(b) *If $2x \in U(R)$, then $x + J(R)$ is a clique in the unit graph $G(R)$.*

(c) *If $2x \notin U(R)$, then $x + J(R)$ is a coclique in the unit graph $G(R)$.*

Lemma 3.7. *Let T be a ring and let R be a local ring with unique maximal ideal $\mathfrak{m}$. If the unit graph $G(T \times R/\mathfrak{m})$ is Hamiltonian, then the unit graph $G(T \times R)$ is Hamiltonian.*

Proof. Since the unit graph $G(T \times R/\mathfrak{m})$ is Hamiltonian, there is a Hamiltonian cycle in $G(T \times R/\mathfrak{m})$, say

$$(a_1, y_1 + \mathfrak{m}) \to \cdots \to (a_n, y_n + \mathfrak{m}) \to (a_1, y_1 + \mathfrak{m}),$$

where $n = |T \times R/\mathfrak{m}|$. Let $\mathfrak{m} = \{x_1, \ldots, x_t\}$. Therefore for every i with $1 \leq i \leq t$, we have $y_i + \mathfrak{m} = \{y_i + x_1, \ldots, y_i + x_t\}$ and so $T \times R = \bigcup_{i=1}^{n} M_i$, where $M_i = \{(a_i, y_i + x_j) \mid 1 \leq j \leq t\}$. It is easy to see that for every r with $1 \leq r \leq n-1$, every element of M_r is adjacent to every element of M_{r+1}. Also every element of M_n is adjacent to every element of M_1. Let S_r for $1 \leq r \leq n-1$ be a subgraph of the unit graph $G(T \times R)$ with vertex-set $M_r \cup M_{r+1}$ and edge-set $\{(a_r, y_r + x_j) \to (a_{r+1}, y_{r+1} + x_\ell) \mid 1 \leq j, \ell \leq t\}$. Also let S_n be a subgraph of the unit graph $G(T \times R)$ with vertex-set $M_n \cup M_1$ and edge-set $\{(a_n, y_n + x_j) \to (a_1, y_1 + x_\ell) \mid 1 \leq j, \ell \leq t\}$. It is easy to see that S_r for $1 \leq r \leq n$ is a Hamiltonian complete bipartite subgraph of the unit graph $G(T \times R)$. For every r with $1 \leq r \leq n-1$, let P_r be a Hamiltonian path of S_r with initial vertex $(a_r, y_r + x_1)$ and end point $(a_{r+1}, y_{r+1} + x_1)$. Also let P_n be a Hamiltonian path of S_n with initial vertex $(a_n, y_n + x_1)$ and end point $(a_1, y_1 + x_1)$. Now we consider the following two cases:

Case 1: n is even. In this case, the cycle

$$P_1 \to P_3 \to \cdots \to P_{n-1} \to (a_1, y_1 + x_1)$$

is a Hamiltonian cycle in the unit graph $G(T \times R)$ and thus it is Hamiltonian.

Case 2: n is odd. In this case, since $|T \times R/\mathfrak{m}|$ is odd, $|R/\mathfrak{m}|$ is odd. This implies that $|R|$ is odd and so $2 \in U(R)$. We may assume that $y_1 + \mathfrak{m} = \mathfrak{m}$. Therefore $y_n + \mathfrak{m} \neq \mathfrak{m}$. Now Proposition 3.6 implies that the subgraph induced by M_n is a clique. Therefore the cycle

$$P_1 \to P_3 \to \cdots \to P_{n-2} \to (a_n, y_n + x_1) \to \cdots \to (a_n, y_n + x_t) \to (a_1, y_1 + x_1)$$

is a Hamiltonian cycle in the unit graph $G(T \times R)$ and thus it is Hamiltonian. □

Lemma 3.8. *Let $R \cong R_1 \times \cdots \times R_n$, where every R_i is a local ring with maximal ideal $\mathfrak{m}_i$. Suppose that $R \not\cong \mathbb{Z}_3$ and for every i with $1 \le i \le n$, we have $R_i/\mathfrak{m}_i \not\cong \mathbb{Z}_2$. Then the unit graph $G(R)$ is Hamiltonian.*

Proof. We prove the lemma by induction on n. If $n = 1$, then R is local and assumptions imply that $|R| \ge 4$. Therefore by using Lemma 3.4 we conclude that the unit graph $G(R)$ is Hamiltonian. Now suppose that the lemma holds true for $n-1$. Consider $T = R_1 \times \cdots \times R_{n-1}$ and $F = R_n/\mathfrak{m}_n$. There are two possibilities: either $T \cong \mathbb{Z}_3$ or $T \not\cong \mathbb{Z}_3$.

First, suppose that $T \cong \mathbb{Z}_3$. If $|R_n| \ge 4$, then by Lemma 3.5 the unit graph $G(R) \cong G(\mathbb{Z}_3 \times R_n)$ is Hamiltonian. If $|R_n| = 3$, then $R_n \cong \mathbb{Z}_3$ and so $R \cong \mathbb{Z}_3 \times \mathbb{Z}_3$. Therefore the cycle

$$(0,0) \to (1,1) \to (0,1) \to (2,1) \to (2,0) \to (2,2) \to (0,2) \to (1,0) \to (1,2) \to (0,0),$$

is a Hamiltonian cycle in the unit graph $G(R) \cong G(\mathbb{Z}_3 \times \mathbb{Z}_3)$ and thus it is Hamiltonian.

Second, suppose that $T \not\cong \mathbb{Z}_3$. In this case the induction hypothesis implies that the unit graph $G(T)$ is Hamiltonian. On the other hand, $F \cong R_n/\mathfrak{m}_n$ is a field with $|F| \ge 3$. Therefore Lemma 3.5 implies that the unit graph $G(T \times F)$ is Hamiltonian. Therefore by applying Lemma 3.7, we conclude that the unit graph $G(R)$ is Hamiltonian. □

We need the following result to give a proof of Lemma 3.10.

Proposition 3.9 [Chartrand and Oellermann 1993, Theorem 8.6]. *Let G be a bipartite graph with partite sets X and Y such that $|X| = |Y| = n \ge 2$. If $\deg(x) > n/2$ for every vertex x of G, then G is Hamiltonian.*

Lemma 3.10. *Let $R \cong R_1 \times \cdots \times R_n \times \mathbb{Z}_2$, where every R_i is a local ring with maximal ideal $\mathfrak{m}_i$. If $R_i/\mathfrak{m}_i \not\cong \mathbb{Z}_2$ for every i with $1 \le i \le n$, then the unit graph $G(R)$ is Hamiltonian.*

Proof. We prove the lemma by induction on n. If $n = 1$, then $R \cong R_1 \times \mathbb{Z}_2$. In this case, it is easy to see that the unit graph $G(R)$ is a bipartite graph with partite sets $X = R_1 \times \{0\}$ and $Y = R_1 \times \{1\}$. On the other hand, by Proposition 3.2(1), we have $\deg(x) = |U(R)| = |U(R_1)| > |U(R_1)|/2 \geq |R|/4$ for every vertex x in $G(R)$. Therefore, by Proposition 3.9, the unit graph $G(R)$ is Hamiltonian.

Now suppose that the lemma holds for $n-1$. The induction hypothesis implies that the unit graph $G(R_1 \times \cdots \times R_{n-1} \times \mathbb{Z}_2)$ is Hamiltonian. On the other hand, $F \cong R_n/\mathfrak{m}_n$ is a field with $|F| \geq 3$. Therefore Lemma 3.5 implies that the unit graph $G(R_1 \times \cdots \times R_{n-1} \times \mathbb{Z}_2 \times F)$ is Hamiltonian and so by applying Lemma 3.7 we conclude that the unit graph $G(R)$ is Hamiltonian. □

A *cycle graph* is a graph that consists of a single cycle. The following result characterizes the unit graphs of rings that are cycle graphs.

Proposition 3.11 [Ashrafi et al. 2010, Theorem 3.2]. *Let R be a ring. Then the unit graph $G(R)$ is a cycle graph if and only if R is isomorphic to either*

(a) $\mathbb{Z}_4$,

(b) $\mathbb{Z}_6$, *or*

(c) $\left\{\left[\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right] \mid a, b \in \mathbb{Z}_2\right\}$.

The next result gives a sufficient condition for a unit graph to be Hamiltonian.

Lemma 3.12. *Let R be a ring such that $R \ncong \mathbb{Z}_2$ and $R \ncong \mathbb{Z}_3$. If R cannot have $\mathbb{Z}_2 \times \mathbb{Z}_2$ as a quotient, then the unit graph $G(R)$ is Hamiltonian.*

Proof. Every ring is isomorphic to a direct product of local rings; see [McDonald 1974, page 95]. Therefore we may write $R \cong R_1 \times \cdots \times R_n$, where every R_i is a local ring with maximal ideal $\mathfrak{m}_i$. We claim that $|U(R)| \geq 2$. To show this, suppose to the contrary that $|U(R)| = 1$. This implies that $|J(R)| = 1$, where $J(R)$ denotes the Jacobson radical of R. Therefore $|\mathfrak{m}_1 \times \cdots \times \mathfrak{m}_n| = 1$ and so $|\mathfrak{m}_i| = 1$ for every i with $1 \leq i \leq n$. Therefore R_i for $1 \leq i \leq n$ is a field and thus $R \cong \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2$, where $\mathbb{Z}_2$ occurs n times in the product. Now the assumption implies that $R \cong \mathbb{Z}_2$, a contradiction. Thus the claim holds and we have $|U(R)| \geq 2$.

First, suppose $|U(R)| = 2$. In this case, by Proposition 3.2, the unit graph $G(R)$ is a 2-regular connected graph and so is a cycle graph. Hence by Proposition 3.11, R is isomorphic to either $\mathbb{Z}_4$, $\mathbb{Z}_6$, or $\left\{\left[\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right] \mid a, b \in \mathbb{Z}_2\right\}$. It is easy to see that the unit graph of each of them is Hamiltonian and therefore so is the unit graph $G(R)$.

Second, suppose that $|U(R)| \geq 3$. By the assumption, $R_i/\mathfrak{m}_i \ncong \mathbb{Z}_2$ for every i, except for possibly at most one i. If $R_i/\mathfrak{m}_i \ncong \mathbb{Z}_2$ for every i, then by Lemma 3.8 the unit graph $G(R)$ is Hamiltonian. If for one i, say n, we have $R_n/\mathfrak{m}_n \cong \mathbb{Z}_2$, then by Lemma 3.10 the unit graph $G(R_1 \times \cdots \times R_n \times \mathbb{Z}_2)$ is Hamiltonian. Now by applying Lemma 3.7 we conclude that the unit graph $G(R)$ is Hamiltonian. □

Proof of Theorem 2.1. (a) implies (b): By assumption, the unit graph $G(R)$ is Hamiltonian and so it is obviously connected. Therefore, by [Ashrafi et al. 2010, Theorem 4.3], we have $\boldsymbol{u}(R) \leq \omega$. This means that the ring R is generated by its units and thus by [Raphael 1974, Corollary 7] it cannot have $\mathbb{Z}_2 \times \mathbb{Z}_2$ as a quotient.

(b) implies (a): This holds by Lemma 3.12.

(b) is equivalent to (c): This holds by [Raphael 1974, Corollary 7].

(c) is equivalent to (d): This is true by definition.

(d) is equivalent to (e): This holds by [Ashrafi et al. 2010, Theorem 4.3]. □

Acknowledgments

The authors would like to thank the referee for carefully reading the paper and for his/her suggestions. Part of this work was done while M. R. Pournaki visited the Delhi Center of the Indian Statistical Institute (ISID). He would like to thank the Academy of Sciences for the Developing World (TWAS) and ISID for sponsoring his visits to New Delhi in July and August 2007 and January 2010. Especially he thanks Professor Rajendra Bhatia for the hospitality enjoyed at ISID.

References

[Ashrafi and Vámos 2005] N. Ashrafi and P. Vámos, "On the unit sum number of some rings", *Q. J. Math.* **56**:1 (2005), 1–12. MR 2005k:11220 Zbl 1100.11036

[Ashrafi et al. 2010] N. Ashrafi, H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Unit graphs associated with rings", *Comm. Algebra* **38**:8 (2010), 2851–2871. MR 2730284 Zbl 05803773

[Atiyah and Macdonald 1969] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley, Reading, MA, 1969. MR 39 #4129 Zbl 0175.03601

[Chartrand and Oellermann 1993] G. Chartrand and O. R. Oellermann, *Applied and algorithmic graph theory*, McGraw-Hill, New York, 1993. MR 1211413

[Dirac 1952] G. A. Dirac, "Some theorems on abstract graphs", *Proc. London Math. Soc.* (3) **2** (1952), 69–81. MR 13,856e Zbl 0047.17001

[Garey and Johnson 1979] M. R. Garey and D. S. Johnson, *Computers and intractability, A guide to the theory of NP-completeness*, W. H. Freeman, San Francisco, CA, 1979. MR 80g:68056 Zbl 0411.68039

[Goldsmith et al. 1998] B. Goldsmith, S. Pabst, and A. Scott, "Unit sum numbers of rings and modules", *Quart. J. Math. Oxford Ser.* (2) **49**:195 (1998), 331–344. MR 99i:16060 Zbl 0933.16035

[Gould 2003] R. J. Gould, "Advances on the Hamiltonian problem — a survey", *Graphs Combin.* **19**:1 (2003), 7–52. MR 2004a:05092 Zbl 1024.05057

[Grimaldi 1990] R. P. Grimaldi, "Graphs from rings", pp. 95–103 in *Proceedings of the Twentieth Southeastern Conference on Combinatorics, Graph Theory, and Computing* (Boca Raton, FL, 1989), vol. 71, edited by F. Hoffman et al., 1990. MR 90m:05122 Zbl 0747.05091

[Maimani et al. 2010] H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Weakly perfect graphs arising from rings", *Glasg. Math. J.* **52**:3 (2010), 417–425. MR 2679902 Zbl 05799531

[McDonald 1974] B. R. McDonald, *Finite rings with identity*, Pure and Applied Mathematics **28**, Marcel Dekker, New York, 1974. MR 50 #7245 Zbl 0294.16012

[Ore 1960] O. Ore, "Note on Hamilton circuits", *Amer. Math. Monthly* **67** (1960), 55. MR 22 #9454 Zbl 0089.39505

[Raphael 1974] R. Raphael, "Rings which are generated by their units", *J. Algebra* **28** (1974), 199–205. MR 49 #7300 Zbl 0271.16013

[Vámos 2005] P. Vámos, "2-good rings", *Q. J. Math.* **56**:3 (2005), 417–430. MR 2006e:16055 Zbl 1156.16303

[West 1996] D. B. West, *Introduction to graph theory*, Prentice Hall, Upper Saddle River, NJ, 1996. MR 96i:05001 Zbl 0845.05001

[Zelinsky 1954] D. Zelinsky, "Every linear transformation is a sum of nonsingular ones", *Proc. Amer. Math. Soc.* **5** (1954), 627–630. MR 16,8c Zbl 0056.11002

Received December 28, 2009. Revised July 14, 2010.

H. R. MAIMANI
MATHEMATICS SECTION, DEPARTMENT OF BASIC SCIENCES
SHAHID RAJAEE TEACHER TRAINING UNIVERSITY
P.O. BOX 16785-163
TEHRAN
IRAN

and

SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM)
P.O. BOX 19395-5746
TEHRAN
IRAN
maimani@ipm.ir

M. R. POURNAKI
DEPARTMENT OF MATHEMATICAL SCIENCES
SHARIF UNIVERSITY OF TECHNOLOGY
P.O. BOX 11155-9415
TEHRAN
IRAN
pournaki@ipm.ir
http://math.ipm.ac.ir/pournaki/

S. YASSEMI
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
COLLEGE OF SCIENCE
UNIVERSITY OF TEHRAN
TEHRAN
IRAN

and

SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM)
P.O. BOX 19395-5746
TEHRAN
IRAN
yassemi@ipm.ir
http://math.ipm.ac.ir/yassemi/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

INSTABILITY OF THE GEODESIC FLOW FOR THE ENERGY FUNCTIONAL

DOMENICO PERRONE

Let $(S^n(r), g_0)$ be the canonical sphere of radius r. Denote by $\widetilde{G}_s$ the Sasaki metric on the unit tangent bundle $T_1S^n(r)$ induced from g_0 and by $\widetilde{\widetilde{G}}_s$ the Sasaki metric on $T_1T_1S^n(r)$ induced from $\widetilde{G}_s$. We resolve here, for $n \geq 7$, a question raised by Boeckx, González–Dávila, and Vanhecke: namely, we prove that the geodesic flow

$$\xi : (T_1S^n(r), \widetilde{G}_s) \to (T_1T_1S^n(r), \widetilde{\widetilde{G}}_s)$$

is an unstable harmonic vector field for any $r > 0$ and $n \geq 7$. In particular, in the case $r = 1$, ξ is an unstable harmonic map. We show that these results are invariant under a four-parameter deformation of the Sasaki metric $\widetilde{\widetilde{G}}_s$.

1. Introduction

Let (M, g) be a compact Riemannian manifold and $\mathfrak{X}^1(M)$ the set of all smooth unit vector fields on (M, g), which we suppose to be nonempty, equivalently, the Euler–Poincaré characteristic of M vanishes. Let $(T_1M, \widetilde{G}_s)$ be the unit tangent sphere bundle equipped with the Sasaki metric $\widetilde{G}_s$. A unit vector field $U \in \mathfrak{X}^1(M)$ determines a map between (M, g) and $(T_1M, \widetilde{G}_s)$ and the energy $E_{\widetilde{G}_s}(U)$ is defined as the energy of the corresponding map

$$U : (M, g) \to (T_1M, \widetilde{G}_s).$$

A unit vector field U is said to be a *harmonic vector field* if it is a critical point for the energy functional $E_{\widetilde{G}_s}$ restricted to $\mathfrak{X}^1(M)$ [Wiegmink 1995; Wood 1997]. Harmonic unit vector fields aren't harmonic maps unless an additional curvature condition is satisfied [Han and Yim 1998; Abbassi et al. 2009a].

For the unit sphere S^{2m+1}, $m > 1$, the Hopf vector fields are unstable harmonic unit vector fields [Wood 1997]. The unit vector fields of minimum energy on the unit sphere S^3 are precisely the Hopf vector fields, equivalently, the unit Killing

The author was supported by funds of the MIUR (PRIN 07) and of the Università del Salento.
MSC2000: 53C43, 53D25.
Keywords: geodesic flow, canonical sphere, stability, energy functional, harmonic maps, natural Riemannian metrics.

vector fields, and no others [Brito 2000]. Contact metric manifolds which Reeb vector field is harmonic are called H-contact manifolds [Perrone 2004]. In [Perrone 2009a] we studied the stability of the Reeb vector field of a compact H-contact three-manifold. If the unit tangent bundle itself is taken as the source manifold of unit vector fields, then a distinguished unit vector field, namely, the *geodesic flow vector field* ξ, appears in a natural way (it is collinear, with a constant factor, to the Reeb vector field of the standard contact metric structure on T_1M).

Let (M, g) be a Riemannian manifold locally isometric to a two-point homogeneous space, that is, locally flat or locally isometric to a rank-one symmetric space. Boeckx and Vanhecke [2000] proved that $\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \widetilde{\widetilde{G}}_s)$ is a harmonic vector field (and a harmonic map), where $\widetilde{\widetilde{G}}_s$ is the corresponding Sasaki metric on T_1T_1M.

Concerning the stability of the geodesic flow ξ we have few results. Boeckx et al. [2002] studied the stability of ξ as harmonic vector field when such a M is in addition compact (note that, by [Borel 1963], compact quotients always exist) and satisfies some other conditions. More precisely, the authors proved that if $n \geq 3$ and M is of nonpositive curvature with nonzero first Betti number, then the geodesic flow $\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \widetilde{\widetilde{G}}_s)$ is an unstable harmonic vector field. In the positive curvature case they considered a space of constant curvature and proved a similar yet weaker result. Indeed, in such case, they proved that the existence of nonzero Killing vector fields implies the instability of ξ for the energy functional $E_{\widetilde{\widetilde{G}}}$, in certain ranges of the dimension and the curvature. With these results, the question of stability of ξ remains open, particularly in the case of a compact quotient of a two-point homogeneous space of positive curvature. The most intriguing one, according to Boeckx et al. [2002], concerns the unit spheres $S^n(1)$ for $n > 2$. Their method does not give any answers in this case.

Recently, the papers [Abbassi et al. 2009a; 2009b; 2010a; Perrone 2009b; 2010] examined the question of when a vector field $V : (M, g) \to (TM, G)$ and a unit vector field $U : (M, g) \to (T_1 M, \widetilde{G})$ are harmonic vector fields and define harmonic maps, where G is a natural Riemannian metric on TM and $\widetilde{G}$ is its restriction to the unit tangent sphere bundle T_1M. (Natural Riemannian metrics form a very large family, which includes the Sasaki metric, the Cheeger–Gromoll metric, metrics of Cheeger–Gromoll type [Benyounes et al. 2007] and the Kaluza–Klein metrics [Wood 1990].) The restrictions $\widetilde{G}$ of such metrics to T_1M possess a simpler form and globally depend on four real parameters a, b, c, d satisfying some inequalities (the parameters $a = 1,\ b = c = d = 0$ define the Sasaki metric $\widetilde{G}_s$). Suppose that (M, g) is a Riemannian manifold locally isometric to a two-point homogeneous space and T_1M, $T_\rho T_1M$ are equipped with arbitrary natural Riemannian metrics $\widetilde{G}$ and $\widetilde{\widetilde{G}}$ respectively. Then, Abbassi et al. [2010b] proved that the geodesic flow $\xi : (T_1M, \widetilde{G}_s) \to (T_\rho T_1M, \widetilde{\widetilde{G}})$ is always a harmonic vector field, and it also defines a harmonic map under some conditions on the coefficients determining the natural Riemannian metrics.

The main purpose of this paper is to study the stability of the geodesic flow

$$\xi : (T_1 S^n(r), \widetilde{G}_s) \to (T_1 T_1 S^n(r), \widetilde{\widetilde{G}}),$$

where $S^n(r)$ is the canonical sphere of radius r and $\widetilde{\widetilde{G}}$ is an arbitrary natural Riemannian metric on $T_1 T_1 S^n(r)$ induced from the Sasaki metric $\widetilde{G}_s$ on $T_1 S^n(r)$ (see Theorem 4.2 and Theorem 5.3). In particular, we get that the geodesic flow

$$\xi : (T_1 S^n(r), \widetilde{G}_s) \to (T_1 T_1 S^n(r), \widetilde{\widetilde{G}})$$

is an unstable harmonic vector field (and an unstable harmonic map) for any $r > 0$, $n \geq 7$, and for any natural Riemannian metric $\widetilde{\widetilde{G}}$ on $T_1 T_1 S^n(r)$ induced from the Sasaki metric $\widetilde{G}_s$. When $\widetilde{\widetilde{G}} = \widetilde{\widetilde{G}}_s$, we resolve the question of posed in [Boeckx et al. 2002, page 202] for any $n \geq 7$. In order to get all these results, we use the Hessian form of the energy functional

$$E_{\widetilde{G}} : \mathfrak{X}^1(M) \to \mathbb{R}, U \mapsto E_{\widetilde{G}}(U) = E\big(U : (M, g) \to (T_1 M, \widetilde{G})\big),$$

for an arbitrary natural Riemannian metric $\widetilde{G}$ (see Theorem 3.2). It should be noted that the instability of the Hopf vector fields on S^{2m+1}, $m > 1$, and the stability (instability) results given in [Perrone 2009a] are invariant under a four-parameter deformation of the Sasaki metric $\widetilde{G}_s$ on $T_1 M$ (see Corollary 3.4).

2. Natural Riemannian metrics on $T_1 M$

Let (M, g) be an n-dimensional Riemannian manifold and ∇ its Levi-Civita connection. We denote by R the Riemannian curvature tensor of (M, g) with the sign convention $R(X, Y)Z = -\nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z + \nabla_{[X,Y]} Z$. Moreover, we denote by Ric the Ricci tensor of type $(0, 2)$, by Q the corresponding endomorphism field and by τ the scalar curvature.

At any point (x, u) of the *tangent bundle* TM, the tangent space of TM splits into the horizontal and vertical subspaces with respect to ∇:

$$(TM)_{(x,u)} = \mathcal{H}_{(x,u)} \oplus \mathcal{V}_{(x,u)}.$$

For any vector $X \in M_x$, there exists a unique vector $X^h \in \mathcal{H}_{(x,u)}$ (the *horizontal lift* of X to $(x, u) \in TM$), such that $p_* X^h = X$, where $p : TM \to M$ is the natural projection. The *vertical lift* of a vector $X \in M_x$ to $(x, u) \in TM$ is a vector $X^v \in \mathcal{V}_{(x,u)}$ such that $X^v(df) = Xf$, for all smooth functions f on M. Here we consider 1-forms df on M as smooth functions on TM. The map $X \to X^h$ is an isomorphism between the vector spaces M_x and $\mathcal{H}_{(x,u)}$. Similarly, the map $X \to X^v$ is an isomorphism between M_x and $\mathcal{V}_{(x,u)}$. Each tangent vector

$\widetilde{Z} \in (TM)_{(x,u)}$ can be written in the form $\widetilde{Z} = X^h + Y^v$, where $X, Y \in M_x$ are uniquely determined vectors. The *geodesic flow* ξ on TM is a vector field given, in terms of local coordinates, by

$$\xi_{(x,u)} = u^h_{(x,u)} = \sum_i u^i (\partial/\partial x^i)^h_{(x,u)}, \quad \text{where} \quad u = \sum_i u^i (\partial/\partial x^i)_x \in M_x.$$

The *natural Riemannian metrics* form a wide family of Riemannian metrics on TM. These metrics depend on several smooth functions from $\mathbb{R}^+ = [0, +\infty)$ to $\mathbb{R}$ and as their name suggests, they arise from a very "natural" construction starting from a Riemannian metric g over M (see [Abbassi and Sarih 2005; Abbassi et al. 2010a] and the references in [Abbassi 2008]). Given an arbitrary g-natural metric G on the tangent bundle TM of a Riemannian manifold (M, g), there are six smooth functions $\alpha_i, \beta_i : \mathbb{R}^+ \to \mathbb{R}$, $i = 1, 2, 3$, such that for every u, X, $Y \in M_x$, we have

$$\begin{aligned}
&G_{(x,u)}(X^h, Y^h) = (\alpha_1+\alpha_3)(r^2) g_x(X, Y) + (\beta_1+\beta_3)(r^2) g_x(X, u) g_x(Y, u),\\
&G_{(x,u)}(X^h, Y^v) = \alpha_2(r^2) g_x(X, Y) + \beta_2(r^2) g_x(X, u) g_x(Y, u),\\
&G_{(x,u)}(X^v, Y^h) = G_{(x,u)}(X^h, Y^v),\\
&G_{(x,u)}(X^v, Y^v) = \alpha_1(r^2) g_x(X, Y) + \beta_1(r^2) g_x(X, u) g_x(Y, u),
\end{aligned} \tag{2-1}$$

where $r^2 = g_x(u, u)$. Put

$$\begin{aligned}
\phi_i(t) &= \alpha_i(t) + t\beta_i(t),\\
\alpha(t) &= \alpha_1(t)(\alpha_1 + \alpha_3)(t) - \alpha_2^2(t),\\
\phi(t) &= \phi_1(t)(\phi_1 + \phi_3)(t) - \phi_2^2(t),
\end{aligned}$$

for all $t \in \mathbb{R}^+$. Then, a g-natural metric G on TM is Riemannian if and only if

$$\alpha_1(t) > 0, \;\; \phi_1(t) > 0, \;\; \alpha(t) > 0, \;\; \phi(t) > 0 \quad \text{for all } t \in \mathbb{R}^+. \tag{2-2}$$

The Sasaki metric G_s, the Cheeger–Gromoll metric, metrics of Cheeger–Gromoll type [Benyounes et al. 2007] and the Kaluza–Klein metrics, as commonly defined on principal bundle [Wood 1990], belong to the subclass of g-natural Riemannian metrics on TM for which horizontal and vertical distribution are mutually orthogonal (i.e., $\alpha_2 = \beta_2 = 0$). More generally, g-natural Riemannian metrics on TM for which horizontal and vertical distribution are mutually orthogonal are called *metrics of Kaluza–Klein type* [Perrone 2010].

Next, the *tangent sphere bundle of radius r* over a Riemannian manifold (M, g), is the hypersurface $T_r M = \{(x, u) \in TM : g_x(u, u) = r^2\}$. The tangent space of

$T_r M$ at a point $(x,u) \in T_r M$ is given by

$$(T_r M)_{(x,u)} = \{X^h + Y^v : X \in M_x, Y \in \{u\}^\perp \subset M_x\}. \tag{2-3}$$

We call *g-natural metrics on* $T_r M$ the restrictions of g-natural metrics of TM to its hypersurface $T_r M$. These metrics possess a simpler form. Precisely, taking in account of (2-1) and (2-3), every natural Riemannian metric $\widetilde{G}$ on $T_r M$ is necessarily induced by a natural Riemannian metric G on TM of the special form (see also [Abbassi 2008; Abbassi et al. 2009a]):

$$\begin{aligned} G_{(x,u)}(X^h, Y^h) &= (a+c)\, g_x(X,Y) + \beta\, g_x(X,u) g_x(Y,u), \\ G_{(x,u)}(X^h, Y^v) &= G_{(x,u)}(X^v, Y^h) = b\, g_x(X,Y), \\ G_{(x,u)}(X^v, Y^v) &= a\, g_x(X,Y), \end{aligned} \tag{2-4}$$

for three real constants a, b, c and a smooth function $\beta : [0,\infty) \to \mathbb{R}$. It is easily seen that G is obtained by the general expression (2-1) when

$$\alpha_1 = a, \quad \alpha_2 = b, \quad \alpha_3 = c, \quad \beta_1 = \beta_2 = 0, \quad \beta_3 = \beta, \tag{2-5}$$

Such a metric $\widetilde{G}$ on $T_r M$ only depends on the value $d = \beta(r^2)$ of β at r^2. From (2-2) and (2-5) it follows that $\widetilde{G}$ is Riemannian if and only if

$$a > 0, \quad \alpha := a(a+c) - b^2 > 0 \quad \text{and} \quad \phi = a(a+c+r^2 d) - b^2 > 0. \tag{2-6}$$

By (2-4), horizontal and vertical lifts are orthogonal with respect to $\widetilde{G}$ if and only if $b = 0$. Moreover, metrics satisfying $b = 0$ are all and the ones induced by natural Riemannian metrics of Kaluza–Klein type. For this reason, a natural Riemannian metric $\widetilde{G}$ on $T_r M$ will be said to be of *Kaluza–Klein type* if and only if horizontal and vertical lifts are $\widetilde{G}$-orthogonal, that is, $b = 0$ in (2-4). Notice that the Sasaki metric, the Cheeger–Gromoll metric, metrics of Cheeger–Gromoll type and the Kaluza–Klein metrics belong to the subclass of natural Riemannian metrics on $T_1 M$ of Kaluza–Klein type. Moreover, an arbitrary natural Riemannian metric $\widetilde{G}$ on $T_r M$ can be considered as a deformation on four parameters (a,b,c,d) of the Sasaki metric $\widetilde{G}_s$ (which is defined by $a = 1$, $b = c = d = 0$).

When $r = 1$, $T_1 M$ is called *unit tangent sphere bundle*. Now, if $\widetilde{G}$ is an arbitrary g-natural Riemannian metric on $T_1 M$, then by (2-4) it follows that the geodesic flow vector field ξ on $T_1 M$ has constant length $\|\xi\|_{\widetilde{G}} = \sqrt{a+c+d}$ (not necessarily equal to 1). Note that $a+c+d > 0$, since $a > 0$ and $\phi = a(a+c+d) - b^2 > 0$. Hence, ξ defines a map $\xi : T_1 M \to T_\rho T_1 M$ where $\rho := \sqrt{a+c+d}$; if $\widetilde{G} = \widetilde{G}_s$, then $\rho = 1$.

3. The Hessian form for the energy $E_{\widetilde{G}}$

Let (M,g) be a compact Riemannian manifold of dimension n. Every unit vector field U on M defines a map between (M,g) and $(T_1M,\widetilde{G}_s)$ and we can define $E_{\widetilde{G}_s}(U)$, *the energy of* U, as the energy of the corresponding map:

$$E_{\widetilde{G}_s}(U)=\frac{1}{2}\int_M \|dU\|^2\,v_g=\frac{n}{2}\,\mathrm{vol}(M,g)+\frac{1}{2}\int_M \|\nabla U\|^2\,dv_g.$$

$E(U)$ is equal, up to constants, to $B(U)=\int_M \|\nabla U\|^2\,dv_g$ which is known as the total bending of U [Wiegmink 1995]. Here dv_g denotes the canonical measure on (M,g). U is called a *harmonic vector field* if it is critical for the energy functional

$$E_{\widetilde{G}_s}:\mathfrak{X}^1(M)\to\mathbb{R},\ U\mapsto E_{\widetilde{G}_s}(U)=E\big(U:(M,g)\to(T_1M,\widetilde{G}_s)\big).$$

The corresponding critical point condition "$\bar{\Delta}V$ is collinear to V" has been determined in [Wiegmink 1995] (see also [Wood 1997]), where $\bar{\Delta}U=-\mathrm{tr}\nabla^2U$ is the *rough Laplacian* at U. This critical point condition has a tensorial character and may also be considered on non compact manifolds.

Now, consider on T_1M an arbitrary g-natural Riemannian metric $\widetilde{G}$. Then a unit vector field U defines a mapping from (M,g) to $(T_1M,\widetilde{G})$ and we can consider the energy functional

$$E_{\widetilde{G}}:\mathfrak{X}^1(M)\to\mathbb{R},\ U\mapsto E_{\widetilde{G}}(U)=E\big(U:(M,g)\to(T_1M,\widetilde{G})\big)=\int_M e(U)\,dv_g,$$

where $e(U)$ is the energy density of $U:(M,g)\to(T_1M,\widetilde{G})$ and is given by [Abbassi et al. 2009a]

$$2e(U)=n(a+c)+d+a\,\|\nabla U\|^2+2b\,\mathrm{div}\,U, \tag{3-1}$$

and so, integrating over M we get

$$E_{\widetilde{G}}(U)=\frac{1}{2}[n(a+c)+d]\,\mathrm{vol}(M,g)+\frac{a}{2}\int_M \|\nabla U\|^2 dv_g. \tag{3-2}$$

In [Abbassi et al. 2009a] we proved that the critical point condition for the energy $E_{\widetilde{G}_s}$ is invariant under a four-parameter deformation of the Sasaki metric $\widetilde{G}_s$. More precisely:

Theorem 3.1 [Abbassi et al. 2009a]. *Let (M,g) be a compact Riemannian manifold of dimension n. Then, a unit vector field $U\in\mathfrak{X}^1(M)$ is a harmonic vector field for the energy $E_{\widetilde{G}}$ if and only if U is a harmonic vector field for the energy $E_{\widetilde{G}_s}$, that is, $\Delta U=\|\nabla U\|^2U$. Moreover, $U:(M,g)\to(T_1M,\widetilde{G})$ is a harmonic map if and only if U is a harmonic vector field and*

$$b\,QU+a\ \mathrm{tr}[R(\nabla.U,U)\cdot]=(b\,\|\nabla V\|^2-d\,\mathrm{div}\,U)U+d\nabla_UU. \tag{3-3}$$

In the case of the Sasaki metric $\widetilde{G}_s$, (3-3) gives a result of [Han and Yim 1998].

Wiegmink [1995] obtained the second variation formula for the energy $E_{\widetilde{G}_s}$. The second variation formula for the energy $E_{\widetilde{G}}$ could be deduced directly from (3-1) by using Theorem 3.1. In the sequel, we include the proof for completeness. Let U be a harmonic vector field for the energy $E_{\widetilde{G}}$, and $U(t)$ a variation of U in $\mathfrak{X}^1(M)$. Then, by (3-1) we have

$$2e(t) := 2e(U(t)) = n(a+c) + d + a\,\|\nabla U(t)\|^2 + 2b \operatorname{div} U(t),$$

and integrating over M, we find

$$E_{\widetilde{G}}(t) := E_{\widetilde{G}}(U(t)) = \frac{n(a+c)+d}{2}\operatorname{vol}(M,g) + \frac{a}{2}\int_M \|\nabla U(t)\|^2 dv_g. \tag{3-4}$$

Differentiating (3-4) we obtain

$$E'_{\widetilde{G}}(t) = a\int_M g\big(\nabla U(t), \nabla U'(t)\big)\,dv_g,$$

and hence

$$E''_{\widetilde{G}}(t) = a\int_M g\big(\nabla U'(t), \nabla U'(t)\big)\,dv_g + a\int_M g\big(\nabla U(t), \nabla U''(t)\big)\,dv_g.$$

Therefore

$$E''_{\widetilde{G}}(0) = a\int_M \|\nabla W\|^2 dv_g + a\int_M g(\nabla U, \nabla A)\,dv_g,$$

where $W = U'(0)$ is orthogonal to U and $A = U''(0)$. On the other hand, for any $X, Y \in \mathfrak{X}(M)$, by a direct calculation, one gets the Bochner-type formula (see [Poor 1981, page 158] for $X = Y$):

$$\Delta g(X,Y) = g(\bar{\Delta}X, Y) + g(X, \bar{\Delta}Y) - 2g(\nabla X, \nabla Y), \tag{3-5}$$

where Δ is the Laplacian acting on functions. This formula implies

$$\int_M g(\bar{\Delta}U, A)\,dv_g = \int_M g(\nabla U, \nabla A) dv_g,$$

where, using Theorem 3.1, $\bar{\Delta}U = \|\nabla U\|^2 U$. Then

$$E''_{\widetilde{G}}(0) = a\int_M (\|\nabla W\|^2 + \|\nabla U\|^2 g(U,A))\,dv_g.$$

Moreover, $\|\nabla U\|^2 = 1$ implies

$$\|W\|^2 = g\big(U'(0), U'(0)\big) = -g\big(U(0), U''(0)\big) = -g(U,A).$$

Thus, we get:

Theorem 3.2. *Let (M, g) be a compact Riemannian manifold. If $U \in \mathfrak{X}^1(M)$ is a critical point of the energy functional $E_{\tilde{G}}$. Then*

$$(\text{Hess} E_{\tilde{G}})_U(W) = a \int_M \big(\|\nabla W\|^2 - \|\nabla U\|^2 \|W\|^2\big)\, dv_g \tag{3-6}$$

for any $W \in U^{\perp}$.

When T_1M is equipped with the Sasaki metric $\widetilde{G}_s$, we get the Hessian form given in [Wiegmink 1995].

Corollary 3.3. *Let (M, g) be a compact Riemannian manifold and U a unit vector field on M. Then the property of $U : (M, g) \to (T_1M, \widetilde{G}_s)$ being a stable* (*or unstable*) *harmonic vector field is invariant under a four-parameter deformation of the Sasaki metric $\widetilde{G}_s$ on T_1M.*

Wood [1997] showed that for the unit sphere S^{2m+1}, $m > 1$, the Hopf vector fields are unstable for the energy $E_{\widetilde{G}_s}$. Contact metric manifolds which Reeb vector field is harmonic are called H-contact manifolds [Perrone 2004]. Recently, in [Perrone 2009a] we studied the stability of the Reeb vector field of a compact H-contact three manifold for the energy $E_{\widetilde{G}_s}$. From Corollary 3.3 we get:

Corollary 3.4. *The instability of the Hopf vector fields on S^{2m+1}, $m > 1$, and the stability* (*or instability*) *results given in* [Perrone 2009a] *are invariant under a four-parameter deformation of the Sasaki metric $\widetilde{G}_s$ on T_1M.*

4. Instability of the geodesic flow

Let (M, g) be a Riemannian manifold locally isometric to a two-point homogeneous space, that is, locally flat or locally isometric to a rank-one symmetric space. We denote by $\widetilde{G}_s$ the Sasaki metric on T_1M, by $\tilde{\tilde{G}}_s$ the corresponding Sasaki metric on T_1T_1M and by $\tilde{\tilde{G}}$ an arbitrary natural Riemannian metric on T_1T_1M constructed from $\widetilde{G}_s$. Boeckx and Vanhecke [2000] proved that $\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \tilde{\tilde{G}}_s)$ is a harmonic map, in particular ξ is a harmonic vector field for the energy $E_{\tilde{\tilde{G}}}$. About the stability of ξ, we have:

Theorem 4.1 [Boeckx et al. 2002]. *Let (M, g) be a compact quotient of a two-point homogeneous space of nonpositive curvature and with first Betti number $b_1(M) \neq 0$, $\dim M = n \geq 3$. Then the geodesic flow ξ on T_1M is unstable for the energy $E_{\tilde{\tilde{G}}}$.*

In the positive curvature case they proved a similar yet weaker result. Indeed, in such case, the existence of nonzero Killing vector fields implies the instability of ξ for the energy functional $E_{\widetilde{G}_s}$, in certain ranges of the dimension n and of curvature. With these results, the question of stability of ξ remains open. The most intriguing one (according to [Boeckx et al. 2002, page 202]) concerns the unit spheres $S^n(1)$ for $n > 2$. Their method does not give any answers in this case.

Now, we consider on T_1M the Sasaki metric $\widetilde{G}_s$ while on T_1T_1M consider an arbitrary natural Riemannian metric $\widetilde{\widetilde{G}}$ constructed from $\widetilde{G}_s$, where (M, g) is a compact quotient of a two-point homogeneous space of dimension n. Abbassi et al. [2010b, Theorem 5] proved that $\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \widetilde{\widetilde{G}})$ is a harmonic vector field for the energy $E_{\tilde{\tilde{G}}}$. From Theorem 3.2 we have that the geodesic flow ξ is stable (or unstable) with respect to $E_{\tilde{\tilde{G}}}$ if and only if it has the same property with respect to $E_{\tilde{\tilde{G}}_s}$, that is, when $\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \widetilde{\widetilde{G}}_s)$. So we consider Hess $E_{\tilde{\tilde{G}}_s}$; from the general expression (3-6), we have

$$(\text{Hess } E_{\tilde{\tilde{G}}_s})_\xi(W) = \int_{T_1M} \big(\|\tilde{\nabla} W\|^2 - \|\tilde{\nabla}\xi\|^2 \|W\|^2\big)\, dv_{\tilde{G}_s} \tag{4-1}$$

for any vector field W on T_1M such that $\widetilde{G}_s(\xi, W) = 0$, where $\tilde{\nabla}$ is the Levi-Civita connection of $(T_1M, \widetilde{G}_s)$. If X is an arbitrary vector field on M, the tangential lift $X^t_z = X^v_z - g_x(X_x, u)u^v$, $z = (x, u)$, is a vector field on T_1M orthogonal to ξ, but the horizontal lift X^h in general is not. For that reason, we define the modified horizontal $\bar{X}^h_z = X^h_z - g(X_p, u)\xi_z$, $z = (p, u)$. This vector field on T_1M is orthogonal to ξ and tangent to T_1M. Moreover, we have, from [Boeckx et al. 2002, Lemma 1, page 206],

$$\int_{T_1M} \big(\|\tilde{\nabla} X^t\|^2 - \|\tilde{\nabla}\xi\|^2 \|X^t\|^2\big)\, dv_{\tilde{G}_s} = a_{n-1} \int_M \big(\|\nabla X\|^2 + A_t \|X\|^2\big)\, dv_g, \tag{4-2}$$

$$\int_{T_1M} \big(\|\tilde{\nabla} \bar{X}^h\|^2 - \|\tilde{\nabla}\xi\|^2 \|\bar{X}^h\|^2\big)\, dv_{\tilde{G}_s} = a_{n-1} \int_M \big(\|\nabla X\|^2 + A_h \|X\|^2\big)\, dv_g, \tag{4-3}$$

where $\dfrac{n}{n-1}\, a_{n-1}$ is the volume of the unit sphere S^{n-1}, and

$$A_t = \frac{5-2n}{4n(n-1)(n+2)}\|R\|^2 - \frac{\tau^2}{2n^2(n+2)} + \frac{\tau}{n} - n + 2,$$

$$A_h = \frac{4-n}{4n(n-1)(n+2)}\|R\|^2 - \frac{\tau^2}{2n(n-1)(n+2)} + \frac{(n-2)\tau}{n(n-1)} - n + 3.$$

Denote by Δ_1 the Laplacian acting on 1-forms. Recall that Δ_1 also acts on vector fields via duality and it is related to the rough Laplacian $\bar{\Delta}$ and the Ricci operator Q by the well-known Weitzenböck formula [Poor 1981, page 168]:

$$\Delta_1 = \bar{\Delta} + Q. \tag{4-4}$$

Moreover, for any $X \in \mathfrak{X}(M)$, from (3-5) we have

$$-\tfrac{1}{2}\Delta\|X\|^2 = \|\nabla X\|^2 - g(\bar{\Delta}X, X). \tag{4-5}$$

Then (4-4) and (4-5) imply that

$$-\tfrac{1}{2}\Delta\|X\|^2 = \|\nabla X\|^2 - g(\Delta_1 X, X) + \mathrm{Ric}(X, X).$$

As M is locally isometric to a two-point homogeneous space, it is Einstein, that is, $\mathrm{Ric} = (\tau/n)g$, the above equation gives

$$\int_M \|\nabla X\|^2 dv_g = \int_M (g(\Delta_1 X, X) - \frac{\tau}{n}\|X\|^2)\, dv_g. \tag{4-6}$$

Then, (4-1), (4-2), and (4-6) imply

$$(\mathrm{Hess}\, E_{\tilde{\tilde{G}}_s})_\xi(X^t) = a_{n-1}\int_M \Big(g(\Delta_1 X, X) + \Big(A_t - \frac{\tau}{n}\Big)\|X\|^2\Big)\, dv_g, \tag{4-7}$$

$$(\mathrm{Hess}\, E_{\tilde{\tilde{G}}_s})_\xi(\bar{X}^h) = a_{n-1}\int_M \Big(g(\Delta_1 X, X) + \Big(A_h - \frac{\tau}{n}\Big)\|X\|^2\Big)\, dv_g. \tag{4-8}$$

Let λ_1 the first eigenvalue of the Laplacian Δ acting on functions. Consider an eigenfunction f related to the eigenvalue λ_1. Set $\omega = df$, so that

$$\Delta_1\omega = (d\delta + \delta d)\, df = d\delta df = d\Delta f = \lambda_1 df = \lambda_1\omega.$$

Hence, if X_0 is the vector field defined by $g(X_0, \cdot) = \omega$, we obtain

$$\Delta_1 X_0 = \lambda_1 X_0.$$

Consequently, $(\mathrm{Hess}\, E_{\tilde{\tilde{G}}_s})_\xi(X_0^t) < 0$ if and only if λ_1 satisfies

$$\lambda_1 < \frac{\tau}{n} - A_t = \frac{2n-5}{4n(n-1)(n+2)}\|R\|^2 + \frac{\tau^2}{2n^2(n+2)} + n - 2, \tag{4-9}$$

and $(\mathrm{Hess}\, E_{\tilde{\tilde{G}}_s})_\xi(\bar{X}_0^h) < 0$ if and only if λ_1 satisfies

$$\begin{aligned} \lambda_1 &< \frac{\tau}{n} - A_h \\ &= \frac{n-4}{4n(n-1)(n+2)}\|R\|^2 + \frac{\tau^2}{2n(n-1)(n+2)} + \frac{\tau}{n(n-1)} + n - 3. \end{aligned} \tag{4-10}$$

Now, suppose that (M, g) is a space of constant curvature $\kappa > 0$. Then,

$$\tau = n(n-1)\kappa, \quad \|R\|^2 = 2n(n-1)\kappa^2 = \frac{2\tau^2}{n(n-1)} \quad \text{and}$$

$$A_t - \frac{\tau}{n} = \frac{(5-2n)2n(n-1)\kappa^2}{4n(n-1)(n+2)} - \frac{n^2(n-1)^2\kappa^2}{2n^2(n+2)} - (n-2),$$

that is,

$$\frac{\tau}{n} - A_t = (n-2)\Big(\frac{\kappa^2}{2} + 1\Big) > 0 \quad \text{for any } n > 2.$$

Moreover,

$$A_h - \frac{\tau}{n} = \frac{(4-n)2n(n-1)\kappa^2}{4n(n-1)(n+2)} - \frac{n^2(n-1)^2\kappa^2}{2n(n-1)(n+2)} + \frac{(n-2)n(n-1)\kappa}{n(n-1)} - (n-3) - \frac{\tau}{n}$$
$$= \frac{(2-n)}{2}\kappa^2 - \kappa - (n-3),$$

that is,

$$\frac{\tau}{n} - A_h = \frac{(n-2)}{2}\kappa^2 + \kappa + n - 3.$$

Therefore, by (4-9), (Hess $E_{\tilde{\tilde{G}}_s})_\xi(X_0^t) < 0$ if and only if λ_1 satisfies

$$\lambda_1 < \frac{\tau}{n} - A_t = (n-2)\Big(\frac{\kappa^2}{2} + 1\Big) \tag{4-11}$$

and, by (4-10), (Hess $E_{\tilde{\tilde{G}}_s})_\xi(\bar{X}_0^h) < 0$ if and only if λ_1 satisfies

$$\lambda_1 < \frac{\tau}{n} - A_h = \frac{(n-2)}{2}\kappa^2 + \kappa + n - 3. \tag{4-12}$$

Now, for a space of constant sectional curvature $\kappa > 0$, a result of Lichnerowicz and Obata [Berger et al. 1971, pages 179–180] states that the eigenvalue λ_1 satisfies $\lambda_1 \geq n\kappa$, where the equality holds if and only if M is isometric to the canonical sphere of radius $r = \sqrt{1/\kappa}$. So, for the sphere $S^n(r)$ of radius $r > 0$, that is of constant sectional curvature $\kappa = 1/r^2$, the conditions (4-11), (4-12) become

$$(\kappa^2 - 2\kappa + 2)\left(n - \frac{2(\kappa^2+2)}{\kappa^2 - 2\kappa + 2}\right) > 0, \tag{4-13}$$

$$(\kappa^2 - 2\kappa + 2)\left(n - \frac{2\kappa^2 - 2\kappa + 6}{\kappa^2 - 2\kappa + 2}\right) > 0. \tag{4-14}$$

Examining these expressions, we conclude:

If n and κ satisfy one of the following conditions, then (4-11) is satisfied:

- $\kappa > 0$ and $n \geq 7$,
- $\kappa \in]0, 1[\,\cup\,]2, +\infty[$ and $n \geq 6$,
- $\kappa \in]0, \frac{1}{3}(5-\sqrt{7})[\,\cup\,]\frac{1}{3}(5+\sqrt{7}), +\infty[$ and $n \geq 5$,
- $\kappa \in]0, 2-\sqrt{2}[\,\cup\,]2+\sqrt{2}, +\infty[$ and $n \geq 4$,
- $\kappa \in]0, 3-\sqrt{7}[\,\cup\,]3+\sqrt{7}, +\infty[$ and $n \geq 3$.

If n and κ satisfy one of the following conditions, then (4-12) is satisfied:

- $\kappa > 0$ and $n \geq 7$,
- $\kappa \in]0, 1[\,\cup\,]\frac{3}{2}, +\infty[$ and $n \geq 6$,
- $\kappa \in]0, \frac{2}{3}[\,\cup\,]2, +\infty[$ and $n \geq 5$,
- $\kappa \in]0, 3-2\sqrt{2}[\,\cup\,]3+2\sqrt{2}, +\infty[$ and $n \geq 4$,
- $\kappa \in]4, +\infty[$ and $n \geq 3$.

Summarizing:

Theorem 4.2. *Let $S^n(r)$ be the canonical sphere of radius r, and let $\kappa = 1/r^2$. If one of the following conditions holds, then the geodesic flow ξ on $T_1 S^n(r)$ is unstable for the energy $E_{\tilde{\tilde{G}}}$:*

- $\kappa > 0$ *and* $n \geq 7$,
- $\kappa \in\,]0,1[\, \cup\,]\frac{3}{2},+\infty[$ *and* $n \geq 6$,
- $\kappa \in\,]0,\frac{2}{3}[\, \cup\,]2,+\infty[$ *and* $n \geq 5$,
- $\kappa \in\,]0,2-\sqrt{2}[\, \cup\,]2+\sqrt{2},+\infty[$ *and* $n \geq 4$,
- $\kappa \in\,]0,3-\sqrt{7}[\, \cup\,]4,+\infty[$ *and* $n \geq 3$.

Corollary 4.3. *The geodesic flow ξ on $T_1 S^n(1)$ is unstable for the energy $E_{\tilde{\tilde{G}}}$, for $n \geq 7$.*

The two-dimensional case. Let (M,g) be a compact Riemannian surface of constant curvature $\kappa > 0$. If $\kappa < 1$, Theorem 7 of [Boeckx et al. 2002] gives that the geodesic flow ξ on T_1M is an unstable harmonic vector field for the energy $E_{\tilde{\tilde{G}}_s}$. If $\kappa = 1$, (T_1M, G_s) is a compact Riemannian three-manifold of constant curvature $c = \frac{1}{4}$ and ξ is a unit Killing vector field. Brito [2000] proved that the unit vector fields of minimum energy on the unit sphere S^3 are precisely the unit Killing vector fields, and no others. Recently, we proved an analogue of Brito's theorem for a compact Sasakian three-manifold [Perrone 2008, page 20]. A consequence of its proof gives: *the unit vector fields of minimum energy on a compact Riemannian three-manifold of constant sectional curvature $c \geq 0$ are precisely the unit Killing vector fields, and no others.*

Other positively curved two-point homogeneous spaces. There are known analogues of Theorem 4.2 for other compact positively curved two-point homogeneous spaces, though with different conditions. We mention:

– For the real projective space $\mathbb{RP}^n$ of constant sectional curvature $\kappa > 0$, we know from [Gallot 1980, page 38] that $\lambda_1 = 2(n+1)\kappa$. The conditions (4-11) and (4-12) become

$$n(\kappa^2 - 4\kappa + 2) - 2(\kappa^2 + 2\kappa + 2) > 0, \quad n(\kappa^2 - 4\kappa + 2) - 2(\kappa^2 + \kappa + 3) > 0.$$

Examining this inequality we find that if $n \geq 3$ and $\kappa \in\,]0, 8-\sqrt{62}[\, \cup\,]14,+\infty[$, the geodesic flow ξ on $T_1\mathbb{RP}^n$ is unstable for the energy $E_{\tilde{\tilde{G}}}$.

– For the complex projective space $\mathbb{CP}^m$, $n = 2m$, of constant holomorphic sectional curvature $\mu > 0$, we have, from [Gray and Vanhecke 1979, page 177] and [Gallot 1980, page 38],

$$\tau = m(m+1)\mu, \quad \|R\|^2 = 2m(m+1)\mu^2, \quad \lambda_1 = (m+1)\mu. \tag{4-15}$$

Using this, we obtain conditions, like Theorem 4.2, which imply the instability of the geodesic flow on the unit tangent sphere bundle of the corresponding space. For $m > 1$, the condition $\lambda_1 + A_t - \tau/n < 0$ becomes

$$(m-1)(2m+11)\mu^2 - 16(m+1)(2m-1)\mu + 32(m-1)(2m-1) > 0.$$

The other condition, $\lambda_1 + A_h - \tau/n < 0$, becomes

$$(m-1)(m+4)\mu^2 - 4(m+1)(4m-3)\mu + 8(2m-3)(2m-1) > 0.$$

A similar remark applies to the next two examples. The references are also the same.

– For the quaternionic projective space, $n = 4m$, of constant quaternionic sectional curvature $\nu > 0$, we have

$$\tau = 4m(m+2)\nu, \|R\|^2 = 4m(5m+1)\nu^2, \lambda_1 = 2(m+1)\nu. \tag{4-16}$$

– For the Cayley projective plane, $n = 16$, of maximum sectional curvature $\zeta > 0$,

$$\tau = 144\zeta, \|R\|^2 = 576\zeta^2, \lambda_1 = 48\zeta. \tag{4-17}$$

5. Instability of harmonic maps defined by the geodesic flow

In the theory of harmonic maps, a fundamental question concerns the existence of harmonic maps between two given Riemannian manifolds (M, g) and (M', g'). If (M, g) is compact and (M', g') is of nonpositive sectional curvature, there exists a harmonic map $f : (M, g) \to (M', g')$ in each homotopy class [Eells and Sampson 1964]. However, there is no general existence result when (M', g') does not satisfy this condition. This fact makes it interesting to find examples of harmonic maps having such a target manifold. Since the standard existence theory for harmonic maps does not apply, examples have to be constructed ad hoc.

Now, let $\widetilde{G}$ be an arbitrary Riemannian g-natural metric on T_1M. By (2-4), the geodesic flow vector field ξ on T_1M has constant length $\|\xi\|_{\widetilde{G}} = \rho = \sqrt{a+c+d}$ (not necessarily equal to 1). Hence, we can study the harmonicity of the geodesic flow as a map $\xi : T_1M \to T_\rho T_1M$. We equip $T_\rho T_1M$ with an arbitrary g-natural Riemannian metric $\widetilde{\widetilde{G}}$ coming from $\widetilde{G}$. By (2-6), $\widetilde{\widetilde{G}}$ will depend on four constants a', b', c', d', satisfying

$$a' > 0, \quad a'(a'+c') - (b')^2 > 0, \quad a'(a'+c'+\rho^2 d') - (b')^2 > 0.$$

The following result shows that in many cases, the geodesic flow also defines a harmonic map.

Theorem 5.1 [Abbassi et al. 2010b]. *Let (M, g) be a two-point homogeneous space. The map $\xi : (T_1M, \widetilde{G}) \to (T_\rho T_1M, \widetilde{\widetilde{G}})$ is a harmonic map if and only if*

$$na\alpha b' \sum_{i=1}^{n-1} \lambda_i^2 = \big[a'b^3 d + 2b'\alpha(\alpha - b^2)\big]\tau - n(n-1)b'\alpha(a+c)^2, \tag{5-1}$$

where $\alpha = a(a+c) - b^2$ and the λ_i are the eigenvalues of the Jacobi operator $R_u = R(\cdot\,, u)u$.

In particular, if $\widetilde{G} = \widetilde{G}_s$ (i.e., $a = 1, b = c = d = 0$) and M has constant sectional curvature κ, then $\lambda_i = \kappa$, $\tau = n(n-1)\kappa$ and (5-1) becomes $n(n-1)b'(\kappa-1)^2 = 0$. Thus we get:

Theorem 5.2. *Let (M, g) be a space of constant sectional curvature κ.*

(i) *If $\kappa = 1$, the geodesic flow determines a harmonic map*

$$\xi : (T_1M, \widetilde{G}_s) \to (T_1T_1M, \widetilde{\widetilde{G}})$$

for any natural Riemannian metric $\widetilde{\widetilde{G}}$ on T_1T_1M induced from $\widetilde{G}_s$.

(ii) *If $\kappa \neq 1$, the geodesic flow determines a harmonic map*

$$\xi : (T_1M, \widetilde{G}_s) \to (T_\rho T_1M, \widetilde{\widetilde{G}})$$

if and only if $\widetilde{\widetilde{G}}$ is of Kaluza–Klein type, that is, $b' = 0$.

Since instability for the energy restricted to $\mathfrak{X}^1(T_1M)$ clearly implies instability in the large sense, combining Theorem 4.2 and Theorem 5.2 we get:

Theorem 5.3. (i) *The geodesic flow vector field on $T_1S^n(1)$, $n > 6$, determines an unstable harmonic map $\xi : (T_1S^n(1), \widetilde{G}_s) \to (T_1T_1S^n(1), \widetilde{\widetilde{G}})$ for any natural Riemannian metric $\widetilde{\widetilde{G}}$ on $T_1T_1S^n(1)$ induced from $\widetilde{G}_s$.*

(ii) *Let $S^n(\kappa)$ be the canonical sphere of constant curvature κ, where*

$$\kappa \in \,]0, 3 - \sqrt{7}[\, \cup \,]4, +\infty[,$$

and let $n \geq 3$. Then the geodesic flow on $T_1S^n(\kappa)$ determines an unstable harmonic map

$$\xi : (T_1S^n(\kappa), \widetilde{G}_s) \to (T_1T_1S^n(\kappa), \widetilde{\widetilde{G}})$$

for any metric of Kaluza–Klein type $\widetilde{\widetilde{G}}$ on $T_1T_1S^n(\kappa)$ induced from $\widetilde{G}_s$.

References

[Abbassi 2008] M. T. K. Abbassi, "g-natural metrics: new horizons in the geometry of tangent bundles of Riemannian manifolds", *Note Mat.* **28**:Suppl. 1 (2008), 6–35. MR 2011a:53039 Zbl 05579511

[Abbassi and Sarih 2005] M. T. K. Abbassi and M. Sarih, "On some hereditary properties of Riemannian g-natural metrics on tangent bundles of Riemannian manifolds", *Differential Geom. Appl.* **22**:1 (2005), 19–47. MR 2005k:53051 Zbl 1068.53016

[Abbassi et al. 2009a] M. T. K. Abbassi, G. Calvaruso, and D. Perrone, "Harmonicity of unit vector fields with respect to Riemannian g-natural metrics", *Differential Geom. Appl.* **27**:1 (2009), 157–169. MR 2009m:53169 Zbl 1185.53070

[Abbassi et al. 2009b] M. T. K. Abbassi, G. Calvaruso, and D. Perrone, "Some examples of harmonic maps for g-natural metrics", *Ann. Math. Blaise Pascal* **16**:2 (2009), 305–320. MR 2010m:53097 Zbl 1183.58008

[Abbassi et al. 2010a] M. Abbassi, G. Calvaruso, and D. Perrone, "Harmonic sections of tangent bundles equipped with Riemannian g-natural metrics", *Quart. J. Math.* (2010), 1–30.

[Abbassi et al. 2010b] M. T. K. Abbassi, G. Calvaruso, and D. Perrone, "Harmonic maps defined by the geodesic flow", *Houston J. Math.* **36**:1 (2010), 69–90. MR 2610782

[Benyounes et al. 2007] M. Benyounes, E. Loubeau, and C. M. Wood, "Harmonic sections of Riemannian vector bundles, and metrics of Cheeger–Gromoll type", *Differential Geom. Appl.* **25**:3 (2007), 322–334. MR 2008e:53118 Zbl 1128.53037

[Berger et al. 1971] M. Berger, P. Gauduchon, and E. Mazet, *Le spectre d'une variété riemannienne*, Lecture Notes in Math. **194**, Springer, Berlin, 1971. MR 43 #8025 Zbl 0223.53034

[Boeckx and Vanhecke 2000] E. Boeckx and L. Vanhecke, "Harmonic and minimal vector fields on tangent and unit tangent bundles", *Differential Geom. Appl.* **13**:1 (2000), 77–93. MR 2001f:53138 Zbl 0973.53053

[Boeckx et al. 2002] E. Boeckx, J. C. González-Dávila, and L. Vanhecke, "Stability of the geodesic flow for the energy", *Comment. Math. Univ. Carolin.* **43**:2 (2002), 201–213. MR 2003g:53147 Zbl 1090.53035

[Borel 1963] A. Borel, "Compact Clifford-Klein forms of symmetric spaces", *Topology* **2** (1963), 111–122. MR 26 #3823 Zbl 0116.38603

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

[Eells and Sampson 1964] J. Eells, Jr. and J. H. Sampson, "Harmonic mappings of Riemannian manifolds", *Amer. J. Math.* **86** (1964), 109–160. MR 29 #1603 Zbl 0122.40102

[Gallot 1980] S. Gallot, "Variétés dont le spectre ressemble à celui de la sphère", pp. 33–52 in *Analysis on manifolds* (Metz, 1979), Astérisque **80**, Soc. Math. France, Paris, 1980. MR 82k:58092

[Gray and Vanhecke 1979] A. Gray and L. Vanhecke, "Riemannian geometry as determined by the volumes of small geodesic balls", *Acta Math.* **142**:3-4 (1979), 157–198. MR 81i:53038 Zbl 0428.53017

[Han and Yim 1998] D.-S. Han and J.-W. Yim, "Unit vector fields on spheres, which are harmonic maps", *Math. Z.* **227**:1 (1998), 83–92. MR 99c:58044 Zbl 0891.53024

[Perrone 2004] D. Perrone, "Contact metric manifolds whose characteristic vector field is a harmonic vector field", *Differential Geom. Appl.* **20**:3 (2004), 367–378. MR 2005a:53134 Zbl 1061.53028

[Perrone 2008] D. Perrone, "On the volume of unit vector fields on Riemannian three-manifolds", *C. R. Math. Acad. Sci. Soc. R. Can.* **30**:1 (2008), 11–21. MR 2009i:53028 Zbl 1168.53019

[Perrone 2009a] D. Perrone, "Stability of the Reeb vector field of H-contact manifolds", *Math. Z.* **263**:1 (2009), 125–147. MR 2010g:53147 Zbl 1173.53015

[Perrone 2009b] D. Perrone, "Unit vector fields on real space forms which are harmonic maps", *Pacific J. Math.* **239**:1 (2009), 89–104. MR 2009j:53082 Zbl 1151.53059

[Perrone 2010] D. Perrone, "Minimality, harmonicity and CR geometry for Reeb vector fields", *Int. J. of Math.* **9** (2010), 1189–1218.

[Poor 1981] W. A. Poor, *Differential geometric structures*, McGraw-Hill, New York, 1981. MR 83k:53002 Zbl 0493.53027

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

[Wood 1990] C. M. Wood, "An existence theorem for harmonic sections", *Manuscripta Math.* **68**:1 (1990), 69–75. MR 91d:58055 Zbl 0713.58010

[Wood 1997] C. M. Wood, "On the energy of a unit vector field", *Geom. Dedicata* **64**:3 (1997), 319–330. MR 98e:58064 Zbl 0878.58017

Received February 9, 2010.

DOMENICO PERRONE
DIPARTIMENTO DI MATEMATICA "E. DE GIORGI"
UNIVERSITÀ DEL SALENTO
I-73100 LECCE
ITALY

domenico.perrone@unisalento.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

STRING STRUCTURES AND CANONICAL 3-FORMS

CORBETT REDDEN

Using basic homotopy constructions, we show that isomorphism classes of string structures on spin bundles are naturally given by certain degree 3 cohomology classes, which we call string classes, on the total space of the bundle. Using a Hodge isomorphism, we then show that the harmonic representative of a string class gives rise to a canonical 3-form on the base space, refining the associated differential character. We explicitly calculate this 3-form for homogeneous metrics on 3-spheres, and we discuss how the cohomology theory tmf could potentially encode obstructions to positive Ricci curvature metrics.

1. Introduction

Degree four characteristic classes arise as obstructions in several ways in math and theoretical physics. This is analogous to the way the Stiefel–Whitney classes w_1 and w_2 encode obstructions to orientations and spin structures on a manifold M. One usually encounters the degree four classes when considering structures analogous to the spin structure, but on mapping spaces $\mathrm{Map}(\Sigma, M)$, where Σ is a 1- or 2-dimensional manifold. It is common to say that $\frac{1}{2}p_1(M) = 0 \in H^4(M; \mathbb{Z})$ is the obstruction to forming a *string structure* on a manifold M.

In this paper, we only deal with a homotopy-theoretic version of string structures. While geometric notions, such as [Coquereaux and Pilch 1989; Stolz and Teichner 2004; Waldorf 2009], are necessary for applications, we show we can recover some of this geometric information from the topological data for free. When dealing with these degree 4 classes, one usually must also deal with the associated differential characters. We naturally obtain globally defined forms representing these characters. We also speculate on the possibility that the string orientation of tmf may encode obstructions to positive Ricci curvature metrics. This would be analogous to the obstructions for positive scalar curvature metrics encoded in the spin orientation of KO.

This work was partially supported by the NSF RTG grant DMS-0739208.
MSC2000: primary 57R15, 58J28; secondary 58A14, 55N34, 53C05.
Keywords: string structure, differential characters, positive Ricci curvature, elliptic cohomology.

Throughout, all manifolds will be compact, connected, oriented, smooth, and without boundary. We now set up notation. Let G be a compact, simply connected, simple Lie group and $\lambda \in H^4(BG;\mathbb{Z})$ a universal characteristic class. Let $P \xrightarrow{\pi} M$ be a principal G-bundle with connection Θ, and let

$$\lambda(P) \in H^4(M;\mathbb{Z}), \quad \lambda(\Theta) \in \Omega^4(M), \quad \check{\lambda}(\Theta) \in \check{H}^4(M),$$

respectively be the naturally induced characteristic class, Chern–Weil form, and Cheeger–Simons differential character. The differential character is closely related to the Chern–Simons form $CS_\lambda(\Theta) \in \Omega^3(P)$. Finally, g will be a Riemannian metric on M. (In Section 2, we also consider more general G and λ.)

To the characteristic class

$$\lambda \in H^4(BG;\mathbb{Z}) \cong H^3(G;\mathbb{Z}) \cong \pi_3(G) \cong \mathbb{Z},$$

one can associate a topological group $\widetilde{G}_\lambda$ and homomorphism $\widetilde{G}_\lambda \to G$ killing off the corresponding element in $\pi_3(G)$ and inducing isomorphisms in all higher homotopy groups [Stolz 1996; Stolz and Teichner 2004; Baez et al. 2007; Henriques 2008; Schommer-Pries 2009]. When $G = \mathrm{Spin}(k)$ and $\lambda = \frac{1}{2}p_1$, the resulting group is commonly known as String(k). While the actual groups $\widetilde{G}_\lambda$ are not easy to describe, their homotopy type is clearly fixed. Therefore, we base our constructions only on the homotopy type. While more concrete models of $\widetilde{G}_\lambda$ lead to more geometric definitions of $\widetilde{G}_\lambda$-structures, we only consider the problem of lifting the classifying map from BG to $\widetilde{BG_\lambda}$. We call a specific choice of lift a trivialization of the cohomology class λ.

In Section 2, we show that, up to homotopy, such trivializations of λ are naturally equivalent to cohomology classes $\mathcal{S} \in H^3(P;\mathbb{Z})$ that restrict to $\Omega\lambda \in H^3(G;\mathbb{Z})$ on the fibers. Here, $\Omega\lambda$ is the class that universally transgresses to λ. These classes $\mathcal{S}$ are referred to as λ-trivialization classes. In the case where $G = \mathrm{Spin}(k)$ and $\lambda = \frac{1}{2}p_1$, we see that the homotopy class of a string structure is equivalent to its string class $\mathcal{S}$. An important consequence is that one can describe an element in string bordism by a spin manifold M and string class $\mathcal{S} \in H^3(\mathrm{Spin}(TM);\mathbb{Z})$.

In fact, these statements hold in greater generality, and Section 2 considers the more general case where G is a topological group and $\lambda \in H^n(BG;H)$. Homotopy classes of lifts to $\widetilde{BG_\lambda}$ still induce canonical classes in $H^{n-1}(P;H)$, and there is an equivalence when $\widetilde{H}^i(BG;H) = 0$ for $i < n$.

In Section 3, we analyze the harmonic representative of a λ-trivialization class $\mathcal{S}$ on $P \xrightarrow{\pi} M$. The metric on P is naturally induced by a connection Θ on P and a Riemannian metric g on M. The harmonic 3-forms on P, in an adiabatic limit, were previously analyzed in [Redden 2008]. In Theorem 3.7, we see that the induced Hodge isomorphism $H^3(P;\mathbb{R}) \xrightarrow{\cong} \Omega^3(P)$ sends the string class $\mathcal{S}$ to the 3-form $CS_\lambda(\Theta) - \pi^* H_{\mathcal{S},g,\Theta}$. We call the form $H_{\mathcal{S},g,\Theta} \in \Omega^3(M)$ the canonical 3-form associated to the λ-trivialization class, metric, and connection. Proposition 3.12 states

$$d^* H_{\mathcal{S},g,\Theta} = 0 \in \Omega^2(M) \quad \text{and} \quad \check{H}_{\mathcal{S},g,\Theta} = \check{\lambda}(\Theta) \in \check{H}^4(M),$$

where $\check{H}_{\mathcal{S},g,\Theta}$ is the induced differential character. Thus, the form $H_{\mathcal{S},g,\Theta}$ lifts $\check{\lambda}(\Theta)$ to take values in $\mathbb{R}$ instead of $\mathbb{R}/\mathbb{Z}$. This lift is independent of the metric on M; the metric picks out the forms with smallest norm lifting $\check{\lambda}(\Theta)$.

We note that any time one encounters $\check{\lambda}(\Theta)$ and $\lambda(P) = 0$, the form $H_{\mathcal{S},g,\Theta}$ is relevant because it gives a purely local version of $\check{\lambda}(\Theta)$. This situation arises in theoretical physics under the guise of anomaly cancellation. It also arises when constructing the loop group extension bundle $\widehat{LP} \to LP$ restricting to $\widehat{LG} \to LG$.

In Sections 4–6, we deal exclusively with string structures on the frame bundle $\mathrm{Spin}(TM) \to M$ of a manifold with spin structure. Given a string class $\mathcal{S}$ and Riemannian metric g, we use the Levi-Civita connection to produce the canonical 3-form $H_{\mathcal{S},g}$.

Section 4 is largely motivational and provides background information on how string structures arise and why they are important. In particular, we discuss the string orientation

$$M\mathrm{String} \xrightarrow{\sigma} \mathrm{tmf}$$

of the cohomology theory of topological modular forms [Hopkins 2002] and its tentative relationship to index theory on loop spaces. This is analogous to the well-understood relationship between KO-theory and index theory. A theorem of Hitchin shows that the spin orientation of KO encodes obstructions to positive scalar curvature metrics. In the hope of an analogous theorem, Question 4.3 asks, If $(M, \mathcal{S}, g)$ is a closed Riemannian n-manifold with string class $\mathcal{S}$ satisfying both $\mathrm{Ric}(g) > 0$ and $H_{\mathcal{S},g} = 0$, does this imply that $\sigma[M, \mathcal{S}] = 0 \in \mathrm{tmf}^{-n}(pt)$?

In Section 5, we give an equivalent reformulation of Question 4.3. One can use the canonical 3-form $H_{\mathcal{S},g}$ to modify the Levi-Civita connection, inducing a metric connection $\nabla^{\mathcal{S},g}$ with torsion. Since the metric is used to "raise an index" of $H_{\mathcal{S},g}$, the global rescaling of M determines a canonical 1-parameter family of connections associated to $H_{\mathcal{S},g}$. This converges to the Levi-Civita connection in the large volume limit. Proposition 5.4 states that the simultaneous condition $(\mathrm{Ric}(g) > 0,\ H_{\mathcal{S},g} = 0)$ is equivalent to the modified connection having positive Ricci curvature in a small volume limit. An interesting side note is the alternate description of the Levi-Civita connection. For a fixed metric g, the Levi-Civita connection is the unique metric connection maximizing the Ricci curvature.

In Section 6, we examine Question 4.3 in the case where $M = S^3$ with a homogeneous metric (under the left or right action of $S^3 \cong \mathrm{SU}(2)$). In this case, the answer to Question 4.3 is yes, but not if either of the conditions $(\mathrm{Ric}(g) > 0,\ H_{\mathcal{S},g} = 0)$ are weakened in an obvious way. The only $(\mathcal{S}, g)$ satisfying both conditions is the string class and round metric induced from D^4. In this case, the string bordism class is obviously 0. However, there is a 1-parameter family of left-invariant metrics g satisfying $\mathrm{Ric}(g) \geq 0$ and $H_{\mathcal{R},g} = 0$, where $\mathcal{R}$ is induced by the right-invariant framing. Since $\sigma[S^3, \mathcal{R}] = 1/24 \in \mathrm{tmf}^{-3}(pt)$, we see that our question would have a negative answer if one were to weaken the curvature condition. Also, one can find Ricci positive metrics g such that $H_{\mathcal{R},g}$ is arbitrarily small, so one cannot easily weaken the condition $H_{\mathcal{S},g} = 0$ either.

We close by noting that Section 3 is part of a more general story. The results of [Redden 2008] and Section 2 imply that the adiabatic-harmonic representative of a spin^c class gives a canonical 2-form refining the flat differential character $\check{W}_3(\Theta)$. Similarly, the harmonic representative of an SU-class on a $U(n)$-bundle canonically gives a 1-form refining the character $\check{c}_1(\Theta)$. It appears there is a very general relationship between certain cohomology classes on a bundle P, their harmonic representatives, and the associated differential characters. The author is currently attempting to prove and properly understand these relations.

2. Trivializations of characteristic classes

In this section we make some observations on the general theory of trivializing a characteristic class, and we apply it to the Pontrjagin class $\frac{1}{2}p_1 \in H^4(B\mathrm{Spin}; \mathbb{Z})$ to obtain results in subsequent sections. In the case of spin structures, or trivializations of w_2, the results in this section are quite standard. In fact, this section is essentially a rewriting of [Lawson and Michelsohn 1989, Chapter 2.1] so that it applies in greater generality.

Since we will frequently use the notions of homotopy fibers and Eilenberg–Mac Lane spaces, we recall a couple of key facts. If H is an abelian group,[1] then an Eilenberg–Mac Lane space of type $K(H, n)$ is a space with the only nontrivial homotopy group being $\pi_n K(H, n) \cong H$. The space $K(H, n)$ is unique up to homotopy and is the classifying space for ordinary cohomology; that is, for a CW-complex X,

$$H^n(X; H) \cong [X, K(H, n)],$$

where the right side is homotopy classes of based maps $X \to K(H, n)$. Furthermore, the loopspace functor Ω induces a homotopy equivalence

$$\Omega K(H, n) \simeq K(H, n-1) \quad \text{for } n > 1.$$

The homotopy fiber of a map is defined as the pullback of the pathspace fibration. Given a space Y with basepoint y_1, we obtain the pathspace

$$PY = \{\gamma : [0, 1] \to Y \mid \gamma(1) = y_1\}.$$

[1] H is unrelated to the canonical forms in subsequent sections.

The natural map $PY \to Y$ given by $\gamma(0)$ is a fibration whose fiber is homotopic to ΩY. In fact, ΩY acts on the total space of this fibration. The homotopy fiber $\widetilde{X}_f$ of a map $X \xrightarrow{f} Y$ is then the actual pullback of PY. If the homotopy fiber construction is repeated, one obtains a sequence of fibrations homotopic to

$$\cdots \Omega\widetilde{X}_f \to \Omega X \xrightarrow{\Omega f} \Omega Y \to \widetilde{X}_f \to X \xrightarrow{f} Y.$$

Now, let G be a connected topological group (of CW type so that standard classifying space constructions apply). Then, BG is the classifying space for G-bundles, and $H^*(BG; H)$ is the cohomology of BG with coefficients in H. The examples we will be concerned with are when G is a classical Lie group such as SO(n) or Spin(n), and $H = \mathbb{Z}$ or $\mathbb{Z}/2$. Consider a universal characteristic class $\lambda \in H^n(BG; H)$, equivalent to a homotopy class of maps

$$BG \xrightarrow{\lambda} K(H, n).$$

We fix a specific map λ and will not distinguish notationally between the map and the cohomology class.

Let $\widetilde{BG_\lambda}$ be the homotopy fiber of $BG \xrightarrow{\lambda} K(H, n)$. This gives rise to the sequence

$$\cdots \to G \xrightarrow{\Omega\lambda} K(H, n-1) \to \widetilde{BG_\lambda} \to BG \xrightarrow{\lambda} K(H, n)$$

of fibrations up to homotopy. Let $P \xrightarrow{\pi} M$ be a principal G-bundle over the space M; that is, P has a free continuous (right) G-action with quotient map $\pi : P \to P/G \cong M$. Any such bundle P can be obtained as the pullback of the universal bundle

$$\begin{array}{ccc} P & \xrightarrow{f^*} & EG \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ M & \xrightarrow{f} & BG. \end{array}$$

Consequently, any G-bundle has a natural characteristic class

$$\lambda(P) := f^*\lambda \in H^n(M; H).$$

Definition 2.1. A trivialization of the characteristic class λ on P is a lift of the classifying map to $\widetilde{BG_\lambda}$, that is, a lift $\tilde{f}$

$$\begin{array}{ccc} & & \widetilde{BG_\lambda} \\ & \overset{\tilde{f}}{\nearrow} & \downarrow \\ M & \xrightarrow{f} & BG. \end{array}$$

We say two trivializations $\tilde{f}_0$ and $\tilde{f}_1$ are homotopic if they are homotopic through the space of lifts, that is, if there exists a homotopy $\widetilde{F} : [0,1] \times M \to \widetilde{BG_\lambda}$ such that $\widetilde{F}|_0 = \tilde{f}_0$, $\widetilde{F}|_1 = \tilde{f}_1$, and $\widetilde{F}|_t$ is a lift of f for all $t \in [0,1]$.

Proposition 2.2. *Let $P \xrightarrow{\pi} M$ be a G-bundle classified by the map $f : M \to BG$.*

(1) *There exists a trivialization of λ on P if and only if $\lambda(P) = 0 \in H^n(M; H)$.*

(2) *If $\lambda(P) = 0$, the set of trivializations of λ up to homotopy has a free and transitive action of $H^{n-1}(M; H)$; that is, it is an $H^{n-1}(M; H)$-torsor.*

Proof. Part (1) follows from the definition of the homotopy fiber. A lift $\tilde{f}$ is precisely the choice of a nullhomotopy of $\lambda \circ f : M \to K(H, n)$, and $\lambda \circ f$ is nullhomotopic precisely when the cohomology class $\lambda(P) = 0$.

For part (2), assume an initial trivialization $\tilde{f}_0$. This is equivalent to a global section $\tilde{f}_0 : M \to f^*\widetilde{BG_\lambda}$, and $\widetilde{BG_\lambda} \to BG$ is a fibration with fibers of type $\Omega K(H, n) \simeq K(H, n-1)$. In fact the H-space $\Omega K(H, n)$ acts fiberwise on $\widetilde{BG_\lambda}$, so a global section $\tilde{f}_0$ induces a fiber homotopy equivalence

$$\begin{array}{ccc} M \times \Omega K(H,n) & \xrightarrow{\simeq} & f^*\widetilde{BG_\lambda} \\ & \searrow \quad \swarrow & \\ & M. & \end{array}$$

Therefore, the homotopy class of any other section $\tilde{f}_1 : M \to f^*\widetilde{BG_\lambda}$ is equivalent to the homotopy class of a function $M \to \Omega K(H, n) \simeq K(H, n-1)$. □

Note that the connectedness of G implies that BG is simply connected, so we don't have to use local coefficients when dealing with the cohomology of fibers. The cohomology of any fiber is canonically isomorphic to $H^*(G; H)$, and we have a well-defined "restriction to fibers" map in cohomology, given by

$$i^* : H^*(P; H) \to H^*(G; H).$$

Proposition 2.3. (1) *A trivialization $\tilde{f}$ of $\lambda(P)$ gives a canonical cohomology class in $H^{n-1}(P; H)$ that restricts on fibers to the class $\Omega\lambda \in H^{n-1}(G; H)$.*

(2) *The cohomology class in* (1) *only depends on the homotopy class of $\tilde{f}$.*

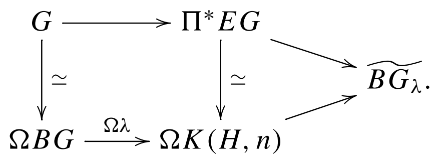
(3) *Furthermore, $H^{n-1}(M; H)$ acts equivariantly on the homotopy classes of λ-trivializations and $H^{n-1}(P; H)$ via π^*.*

Proof. For part (1), consider the universal pullback bundle

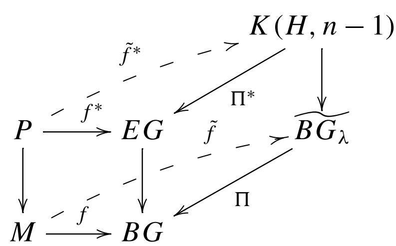
$$\begin{array}{ccc} EG & \xleftarrow{\Pi^*} & \Pi^* EG \\ \downarrow & & \downarrow \\ BG & \xleftarrow{\Pi} & \widetilde{BG_\lambda} \end{array}$$

Then, a lift $\tilde{f} : M \to \widetilde{BG_\lambda}$ such that $\Pi \circ \tilde{f} = f$ is equivalent to a G-equivariant map $\tilde{f}^* : P \to \Pi^* EG$ such that $\Pi^* \circ \tilde{f}^* = f^*$.

Since EG is contractible, $\Pi^* EG$ is a $K(H, n-1)$ space, as evidenced by the natural homotopy equivalence of fibrations given by



Therefore, any lift $\tilde{f}$ is equivalent to $\tilde{f}^* : P \to \Pi^* EG \simeq K(H, n-1)$. When restricted to a fiber, $\tilde{f}^* : G \to \Pi^* EG$ is equivalent to $\Omega\lambda : G \to K(H, n-1)$. This is shown in the following commutative diagram:



For part (2), a homotopy $\widetilde{F}$ between any two trivializations $\tilde{f}_0$ and $\tilde{f}_1$ naturally lifts to an equivariant homotopy $\widetilde{F}^*$ between the bundle maps $\tilde{f}_0^*$ and $\tilde{f}_1^*$. Therefore, the cohomology class $\tilde{f}^* \in H^{n-1}(P; H)$ of a trivialization only depends on the homotopy class of $\tilde{f}$.

For part (3), the fiberwise action of $\Omega K(H, n)$ on $\widetilde{BG_\lambda} \xrightarrow{\Pi} BG$ naturally pulls back via π^* to an action on $\Pi^* EG$. If $\tilde{f}_1 = \phi \cdot \tilde{f}_0$, where $\phi : M \to \Omega K(H, n)$, then

$$\tilde{f}_1^* = \pi^* \phi \cdot \tilde{f}_0^*.$$

Therefore, if two homotopy classes trivializations $[\tilde{f}_0]$ and $[\tilde{f}_1]$ differ by $[\phi] \in H^{n-1}(M, H)$, their natural cohomology classes $[\tilde{f}_0^*], [\tilde{f}_1^*] \in H^{n-1}(P, H)$ differ by $\pi^*[\phi]$. □

The previous proposition gives a map

$$\{\lambda\text{-trivializations}\}/\sim \;\longrightarrow\; \{\mathcal{S} \in H^{n-1}(P; H) \mid i^*\mathcal{S} = \Omega\lambda \in H^{n-1}(G; H)\} \tag{2.4}$$

that is equivariant under the natural $H^{n-1}(M; H)$ action. Here, $\sim$ denotes equivalence up to homotopy. In general, this map is neither injective nor surjective. We will refer to such a cohomology class $\mathcal{S}$ as λ-trivialization class.

Proposition 2.5. *Suppose $\widetilde{H}^i(G; H) = 0$ for $i < n-1$. Then (2.4) is a bijection.*

Proof. The connectedness of G implies the E_2 term in the Leray–Serre cohomology spectral sequence for $EG \to BG$ is

$$E_2^{r,s} \cong H^r(BG; H^s(G; H)),$$

and the contractibility of EG implies that $E_\infty^{r,s} = 0$ for $(r, s) \neq (0, 0)$. This, combined with the vanishing of $H^i(G; H)$ for $i < n-1$, implies that the transgression

$$d_n : E_n^{0,n-1} \cong H^{n-1}(G; H) \to E_n^{n,0} \cong H^n(BG; H)$$

is an isomorphism. In fact, Lemma 2.6 says that $d_n(\Omega\lambda) = \lambda$.

The Leray–Serre cohomology spectral sequence for $P \xrightarrow{\pi} M$ is pulled back from the sequence for the universal bundle. This results in the exact sequence

$$\begin{array}{l} 0 \to H^{n-1}(M; H) \xrightarrow{\pi^*} H^{n-1}(P; H) \xrightarrow{i^*} H^{n-1}(G; H) \xrightarrow{d_n} H^n(M; H) \\ \hspace{22em} \Omega\lambda \mapsto \lambda(P). \end{array}$$

If $\lambda(P) = 0$, then the action of $H^{n-1}(M; H)$ is free and transitive on classes in $H^{n-1}(P; H)$ restricting to $\Omega\lambda$. Since (2.4) is an equivariant map, and both sides are torsors for $H^{n-1}(M; H)$, it must be a bijection. □

Lemma 2.6. *Suppose that* $\widetilde{H}^i(X) = 0$ *for* $i < n$. *Then, the cohomology transgression for the pathspace fibration* $\Omega X \hookrightarrow PX \to X$ *is the inverse of the loop functor; that is,* $d_n^{-1} = \Omega$ *in*

$$H^{n-1}(\Omega X; H) \underset{\Omega}{\overset{d_n}{\rightleftarrows}} H^n(X; H). \quad (\cong)$$

Proof. For the fibration $\Omega X \hookrightarrow PX \to X$, the transgression and loop functor are related by

$$\begin{array}{ccc} H^n(X; H) & \xrightarrow{\Omega} & H^{n-1}(\Omega X; H) \\ \downarrow & & \uparrow \\ E_n^{n,0} & \xleftarrow{d_n} & E_n^{0,n-1}. \end{array}$$

This follows from the general relationship between the transgression and cohomology loop suspension [Serre 1951]. If $\widetilde{H}^i(X; H) = 0$ for $i < n$, then there is no room for any nontrivial differentials in the Serre spectral sequence until d_n. Therefore, $E_n^{n,0} \cong H^n(X; H)$, $E_n^{0,n-1} \cong H^{n-1}(\Omega X; H)$ and d_n is an isomorphism with inverse Ω. □

Finally, we wish to make a general note about *stable* cohomology classes. The usual examples are Chern classes, Pontryagin classes, and Stiefel–Whitney classes, and they correspond to the stable cohomology of classifying spaces for the groups

$U(k)$ and $O(k)$. In general, assume one has a sequence of groups $\{G(k)\}$ and natural inclusions $G(k) \hookrightarrow G(k+1)$ inducing maps

$$\cdots \to BG(k) \to BG(k+1) \to BG(k+2) \to \cdots$$

such that the cohomology stabilizes. We then refer to the cohomology of $BG = \lim_{k\to\infty} BG(k)$. Any cohomology class $\lambda \in H^n(BG; H)$ is stable and defines a sequence of cohomology classes $\lambda_k \in H^n(BG(k); H)$ for all k:

$$H^n(BG; H) \to H^n(BG(k); H), \quad \lambda \mapsto \lambda_k,$$

though the k-subscript is usually unnecessary and dropped. Given a $G(k)$-bundle $P(k)$ classified by $f : M \to BG(k)$, one can stably extend to a $G(k+l)$-bundle $P(k+l)$ by $M \xrightarrow{f} BG(k) \to BG(k+l)$. It is obvious that the characteristic class is stable in that $\lambda_{k+l}(P(k+l)) = \lambda_k(P(k)) \in H^n(M; H)$.

Proposition 2.7. *Consider $\lambda \in H^n(BG; H)$. A trivialization of λ_k on any $G(k)$-bundle naturally induces a trivialization of λ on any stable extension of P.*

Proof. This follows from the naturality of homotopy fibers. If we consider the inclusion map $\iota : BG(k) \to BG(k+1)$, then

$$\widetilde{BG(k)}_\lambda = \lambda_k^* PK(H,n) = (\lambda_{k+1} \circ \iota)^* PK(H,n) = \iota^* \widetilde{BG(k+1)}_\lambda.$$

Drawing this bundle map, we have

$$\begin{array}{ccccc} & & \widetilde{BG(k)}_\lambda & \longrightarrow & \widetilde{BG(k+1)}_\lambda \\ & \overset{\tilde f}{\nearrow} & \downarrow & & \downarrow \\ M & \xrightarrow{f} & BG(k) & \longrightarrow & BG(k+1). \end{array}$$

Any trivialization of $\widetilde{BG(k)}_\lambda$ naturally extends to a trivialization of $\widetilde{BG(k+1)}_\lambda$ by composition, and this process can be continued indefinitely. □

To the $G(k_1)$-bundle $P_1 \xrightarrow{\pi_1} M$ and $G(k_2)$-bundle $P_2 \xrightarrow{\pi_2} M$ we can associate the $G(k_1) \times G(k_2)$-bundle $P_1 \times_M P_2 \to M$. If there are inclusions

$$BG(k_1) \times BG(k_2) \xrightarrow{\iota_1 \times \iota_2} BG(k_1 + k_2),$$

the bundle $P_1 \times_M P_2$ is also naturally a $G(k_1 + k_2)$-bundle.

Suppose that $H^i(BG; H) = 0$ for $i < n$. The Kunneth formula then implies the additivity of $\lambda \in H^n(BG; H)$:

$$\lambda(P_1 \times_M P_2) = \lambda(P_1) + \lambda(P_2) \in H^n(M; H). \tag{2.8}$$

The bottom square of the diagram below then commutes up to homotopy, implying the existence of the dotted arrow map.

$$\begin{array}{ccc}
\widetilde{BG(k_1)}_\lambda \times \widetilde{BG(k_2)}_\lambda & \dashrightarrow & \widetilde{BG(k_1+k_2)}_\lambda \\
\downarrow & & \downarrow \\
BG(k_1) \times BG(k_2) & \longrightarrow & BG(k_1+k_2) \\
\downarrow{\scriptstyle \lambda\times\lambda} & & \downarrow{\scriptstyle \lambda} \\
K(H,n) \times K(H,n) & \longrightarrow & K(H,n)
\end{array}$$

Therefore, a trivialization of λ on the bundles P_1 and P_2 induces a trivialization of λ on $P_1 \times_M P_2$ when viewed as a $G(k_1+k_2)$-bundle (at least up to homotopy). This can also be seen explicitly in terms of cohomology classes.

Proposition 2.9. *For $l = 1, 2$, let $P_l \xrightarrow{\pi_l} M$ be a $G(k_l)$-bundle. Let $P_1 \times_M P_2 \to M$ be the $G(k_1) \times G(k_2)$-bundle and $P \to M$ the induced $G(k_1+k_2)$-bundle. Assume $H^i(BG(k); H) = 0$ for $i < n$ (here $k = k_1, k_2, k_1+k_2$) and $\lambda \in H^n(BG; H)$. Then up to homotopy, a λ-trivialization on any two of $\{P, P_1, P_2\}$ induces a λ-trivialization on the third.*

Proof. Equation (2.8) implies the existence of a λ-trivialization on the third bundle if the other two admit λ-trivializations. Proposition 2.5 states the choice of a trivialization, up to homotopy, is equivalent to a λ-trivialization class $\mathcal{S}_i \in H^{n-1}(P_i; H)$ restricting to $\Omega\lambda$ on the fibers. We now show that the choice of λ-trivialization class on any two bundles determines one on the third bundle.

Note that there are natural bundle maps

$$\begin{array}{ccccc}
 & & P_1 \times_M P_2 & \xrightarrow{\iota_1\times\iota_2} & P \\
 & \swarrow{\scriptstyle \pi_1} & & {\scriptstyle \pi_2}\searrow & \\
P_1 & & & & P_2.
\end{array}$$

We seek solutions to the equation

$$(\iota_1 \times \iota_2)^* \mathcal{S} = \pi^* \mathcal{S}_1 + \pi^* \mathcal{S}_2.$$

Just as in Proposition 2.5, the following commutative diagram is obtained from the Serre spectral sequences for the bundles P and $P_1 \times_M P_2$:

$$\begin{array}{ccccccc}
0 \to H^{n-1}(M) & \longrightarrow & H^{n-1}(P) & \longrightarrow & H^{n-1}(G(k_1+k_2)) & \longrightarrow & H^n(M) \\
\| & & \downarrow{\scriptstyle (\iota_1\times\iota_2)^*} & & \downarrow{\scriptstyle (\iota_1\times\iota_2)^*} & & \| \\
0 \to H^{n-1}(M) & \to & H^{n-1}(P_1 \times_M P_2) & \to & H^{n-1}(G(k_1)) \oplus H^{n-1}(G(k_2)) & \to & H^n(M)
\end{array}$$

The cohomology coefficients are all H but suppressed for spacing purposes.

We know that $(\iota_1 \times \iota_2)^* \lambda_{k_1+k_2} = \lambda_{k_1} \oplus \lambda_{k_2}$. For any three classes $\mathcal{S}, \mathcal{S}_1, \mathcal{S}_2$ in the respective bundles, the exact sequence implies

$$(\iota_1 \times \iota_2)^* \mathcal{S} - \pi_1^* \mathcal{S}_1 - \pi_2^* \mathcal{S}_2 = \pi^* \phi$$

for a unique $\phi \in H^{n-1}(M; H)$. If we fix two of the classes $\mathcal{S}, \mathcal{S}_1, \mathcal{S}_2$, modifying the third by ϕ gives us a solution to our desired equation. □

The previous proposition is useful when dealing with cobordism theories. In the Pontryagin–Thom construction, the relevant extra structure takes place on the stable normal bundle. Suppose the m-manifold M already has a G-structure on the stable normal bundle $\nu(M)$. A lift of the classifying map to $\widetilde{BG_\lambda}$ induces maps on the Thom spaces, which in turn give an element in the $\widetilde{G}_\lambda$-bordism group $M\widetilde{G}_\lambda^{-m}(pt)$.

However, it is often easier or more desirable to describe structures on the tangent bundle. For any manifold M, $TM \oplus \nu(M)$ is canonically isomorphic to the trivial bundle, so Proposition 2.9 often allows us to construct cobordism classes while only dealing with TM, or $G(TM)$.

Corollary 2.10. *Let $\lambda \in H^n(BG; H)$ be a stable class and suppose $H^i(BG; H)$ vanishes for $i < n$. Then, an m-manifold M with G-structure and λ-trivialization class $\mathcal{S} \in H^{n-1}(G(TM); H)$ canonically determines a $\widetilde{G}_\lambda$-bordism class $[M, \mathcal{S}]$ in $M\widetilde{G}_\lambda^{-m}(pt)$.*

We now apply Propositions 2.2, 2.3, and 2.5 and Corollary 2.10 to recover standard information on spin and spin^c structures as well as a convenient description of string structures.

2a. *Spin structures.* For $k > 2$, $\pi_1(\mathrm{SO}(k)) \cong \mathbb{Z}/2$, and the nontrivial double cover is known as Spin(k). The Hurewicz image of the generator of $\pi_1(\mathrm{SO}(k))$ is the generator of $H^1(\mathrm{SO}(k); \mathbb{Z}/2)$, which transgresses to $w_2 \in H^2(B\mathrm{SO}(k); \mathbb{Z}/2)$. It is then clear that

$$\widetilde{B\mathrm{SO}(k)}_{w_2} \simeq B\mathrm{Spin}(k).$$

Moreover, there is a spin orientation of KO-theory $\alpha : M\mathrm{Spin} \to \mathrm{KO}$. Propositions 2.2, 2.3, 2.5, and Corollary 2.10 imply the following.

Proposition 2.11. *Let $P \xrightarrow{\pi} M$ be a principal* SO(k)*-bundle.*

- *P admits a spin structure if and only if $w_2(P) = 0 \in H^2(M; \mathbb{Z}/2)$.*
- *The set of spin structures up to isomorphism is naturally equivalent to the set of spin classes $\mathcal{S} \in H^1(P; \mathbb{Z}/2)$ that restrict to the nontrivial class in $H^1(\mathrm{SO}(k); \mathbb{Z}/2)$.*
- *The set of spin structures up to isomorphism is a torsor for $H^1(M; \mathbb{Z}/2)$.*

- *An oriented m-manifold M with spin class $\mathcal{S} \in H^1(\mathrm{SO}(TM); \mathbb{Z}/2)$ gives rise to the bordism class $[M, \mathcal{S}] \in M\mathrm{Spin}^{-m}(pt)$ and the* KO*-theory class* $\alpha[M, \mathcal{S}] \in \mathrm{KO}^{-m}(pt)$.

Geometrically, the statements can be understood by interpreting $H^1(\,\cdot\,; \mathbb{Z}/2)$ in terms of double covers; see [Lawson and Michelsohn 1989, Chapter 2.1]. Then, a spin structure on P is an equivariant double cover of P restricting fiberwise to the nontrivial double cover of $\mathrm{SO}(k)$.

2b. Spin^c *structures.* For $k > 2$,

$$H^1(\mathrm{SO}(k); \mathbb{Z}) = 0 \quad \text{and} \quad H^2(\mathrm{SO}(k); \mathbb{Z}) \cong H_1(\mathrm{SO}(k); \mathbb{Z}) \cong \pi_1(\mathrm{SO}(k)) \cong \mathbb{Z}/2.$$

The group $\mathrm{Spin}^c(k) = \mathrm{Spin}(k) \times_{\mathbb{Z}/2} S^1$ is a nontrivial S^1-bundle over $\mathrm{SO}(k)$ and hence classified by the generator of $H^2(\mathrm{SO}(k); \mathbb{Z})$; this generator transgresses to $W_3 \in H^3(B\mathrm{SO}(k); \mathbb{Z}) \cong \mathbb{Z}/2$. Therefore,

$$\widetilde{B\mathrm{SO}(k)}_{W_3} \simeq B\mathrm{Spin}^c(k).$$

Furthermore, there is a spin^c orientation $M\mathrm{Spin}^c \to K$ of K-theory. Propositions 2.2, 2.3, 2.5, and Corollary 2.10 imply the following.

Proposition 2.12. *Let $P \xrightarrow{\pi} M$ be a principal* $\mathrm{SO}(k)$*-bundle.*

- *P admits a spin^c structure if and only if $W_3(P) = 0 \in H^3(M; \mathbb{Z})$.*
- *The set of spin^c structures up to homotopy is naturally equivalent to the set of classes $\mathcal{S} \in H^2(P; \mathbb{Z})$ that restrict to the nontrivial class in* $H^2(\mathrm{SO}(k); \mathbb{Z})$.
- *The set of spin^c structures up to homotopy is a torsor for $H^2(M; \mathbb{Z})$.*
- *An oriented m-manifold M with spin^c class $\mathcal{S} \in H^2(\mathrm{SO}(TM); \mathbb{Z})$ gives rise to the bordism class $[M, \mathcal{S}] \in M\mathrm{Spin}^{c\ -m}(pt)$ and K-theory class $\in K^{-m}(pt)$.*

Again, the statements above all have direct geometric interpretations based on $K(\mathbb{Z}, 2) \simeq BS^1$. A spin^c structure is an equivariant S^1-extension of P restricting to the nontrivial extension on fibers. One can always tensor an S^1-bundle over P with the pullback of an S^1-bundle on M.

2c. *String structures.* Let G be any compact simple simply connected Lie group. Then, $\pi_2(G) = 0$ and $\pi_3(G) \cong H^3(G; \mathbb{Z}) \cong \mathbb{Z}$. What happens when you kill $\pi_3(G)$? The 3-connected cover $G\langle 4\rangle \to G$ cannot be a finite-dimensional Lie group, since any connected nonabelian Lie group has nontrivial π_3. However, there do exist topological groups $\widetilde{G} \to G$ that are 3-connected coverings. Various constructions can be found in [Stolz 1996; Stolz and Teichner 2004; Baez et al. 2007; Henriques 2008; Schommer-Pries 2009]. The results of all these imply the following (and usually one only needs G to be semisimple):

Choose a "level" $\lambda \in H^4(BG;\mathbb{Z}) \cong H^3(G;\mathbb{Z})$. Then, there exists a topological group and continuous homomorphism $G\langle\lambda\rangle \to G$ such that $G\langle\lambda\rangle$ has the homotopy type of the fiber of $G \xrightarrow{\Omega\lambda} K(\mathbb{Z},3)$. Applying the classifying space functor gives

$$BG\langle\lambda\rangle \to BG \xrightarrow{\lambda} K(\mathbb{Z},4).$$

When this construction is applied to $G = \mathrm{Spin}(k)$ with $\lambda = \frac{1}{2}p_1 \in H^4(B\mathrm{Spin}(k);\mathbb{Z})$, the resulting topological group is known as $\mathrm{String}(k)$. Trivializations of $\frac{1}{2}p_1$ are commonly referred to as string structures. Applying the classifying space functor gives us $B\mathrm{String}(k)$, and it is clear that

$$B\mathrm{String}(k) \simeq \widetilde{B\mathrm{Spin}(k)}_{\frac{1}{2}p_1}.$$

Remark 2.13. While multiple models for $G\langle\lambda\rangle$ exist, there is no "easy" model like the one Clifford algebras provide for the Spin groups. One must deal with some combination of higher categories, von Neumann algebras, or gerbes, each of which have particular subtleties. In this paper, we avoid these subtleties by only considering the homotopy type of $G\langle\lambda\rangle$. While we lose some information, we can characterize lifts of structure groups purely in terms of ordinary cohomology classes.

Remark 2.14. One should be careful when talking about spin, spin^c, and string structures up to homotopy. In addition to ignoring geometric considerations, these structures are naturally categories and have automorphisms; we only deal with isomorphism classes. The automorphisms play an important role, especially if one wishes to talk about structures locally or glue together manifolds with structures. See [Stolz and Teichner 2004; Waldorf 2009] for more concrete and categorical models of string structures.

For $k \geq 3$, $\mathrm{Spin}(k)$ is simply connected and compact; thus $\widetilde{H}^i(\mathrm{Spin}(k);\mathbb{Z}) = 0$ for $i < 3$. Also, there is a generalized cohomology theory tmf that has a string orientation $M\mathrm{String} \xrightarrow{\sigma} \mathrm{tmf}$, as discussed more in Section 4. Propositions 2.2, 2.3 and 2.5 and Corollary 2.10 then imply the following statements, which can obviously be rewritten for arbitrary $\lambda \in H^4(BG;\mathbb{Z})$ (except for the string orientation).

Definition 2.15. Let $P \xrightarrow{\pi} M$ be a principal $\mathrm{Spin}(k)$-bundle for $k \geq 3$.

- A string structure on a principal $\mathrm{Spin}(k)$-bundle $P \to M$ is a lift of the classifying map to $B\mathrm{String}(k)$, that is, a lift $\tilde{f}$

$$\begin{array}{ccc} & & B\mathrm{String}(k) \\ & \overset{\tilde{f}}{\nearrow} & \downarrow \\ M & \xrightarrow{f} & B\mathrm{Spin}(k). \end{array}$$

- A string class $\mathcal{S} \in H^3(P;\mathbb{Z})$ is a cohomology class that restricts fiberwise to the stable generator of $H^3(\mathrm{Spin}(k);\mathbb{Z})$.

Proposition 2.16. *Let $P \xrightarrow{\pi} M$ be a principal* Spin(k)*-bundle for $k \geq 3$.*

- *P admits a string structure if and only if $\frac{1}{2}p_1(P) = 0 \in H^4(M;\mathbb{Z})$.*
- *Up to homotopy, the choice of a string structure is equivalent to the choice of a string class $\mathcal{S} \in H^3(P;\mathbb{Z})$.*
- *If $\mathcal{S}$ is a string class, then so is $\mathcal{S} + \pi^*\phi$ for $\phi \in H^3(M;\mathbb{Z})$. This natural action of $H^3(M;\mathbb{Z})$ on string classes is free and transitive; that is, the set of string classes is a torsor for $H^3(M;\mathbb{Z})$.*
- *A spin m-manifold M with string class $\mathcal{S} \in H^3(\mathrm{Spin}(TM);\mathbb{Z})$ determines canonical classes $[M,\mathcal{S}] \in M\mathrm{String}^{-m}(pt)$ and $\sigma[M,\mathcal{S}] \in \mathrm{tmf}^{-m}(pt)$.*

Remark 2.17. The cohomology $H^3(\mathrm{Spin}(k);\mathbb{Z})$ does not stabilize until $k = 5$, so we briefly describe the stable generator in dimensions 3 and 4. Under the low-dimensional isomorphisms with the symplectic groups, the groups Spin(k) for $k = 3, 4, 5$ are related through the diagram

$$\begin{array}{ccccc}
\mathrm{Spin}(3) & \hookrightarrow & \mathrm{Spin}(4) & \hookrightarrow & \mathrm{Spin}(5) \\
\downarrow\cong & & \downarrow\cong & & \downarrow\cong \\
\mathrm{Sp}(1) & \xrightarrow{\mathrm{Id}\times\mathrm{Id}} & \mathrm{Sp}(1)\times\mathrm{Sp}(1) & \hookrightarrow & \mathrm{Sp}(2) \\
\downarrow & & \downarrow & & \downarrow \\
\mathbb{H} & \xrightarrow{\mathrm{Id}\times\mathrm{Id}} & \mathbb{H}\oplus\mathbb{H} & \hookrightarrow & \mathrm{GL}(\mathbb{H},2).
\end{array}$$

The Spin(4) decomposition is induced by left and right multiplications of the unit quaternions. The second inclusion $\mathrm{Spin}(4) \hookrightarrow \mathrm{Spin}(5)$ is isomorphic to the matrix inclusion $\mathrm{Sp}(1)\times\mathrm{Sp}(1) \hookrightarrow \mathrm{Sp}(2)$ along the diagonal. Since $H^3(\mathrm{Sp}(k);\mathbb{Z})$ stabilizes at $k = 1$, we denote by 1 a generator of $H^3(\mathrm{Sp}(1);\mathbb{Z}) \cong H^3(\mathrm{SU}(2);\mathbb{Z})$. Then

$$\begin{array}{ccccc}
H^3(\mathrm{Spin}(5);\mathbb{Z}) & \longrightarrow & H^3(\mathrm{Spin}(4);\mathbb{Z}) & \longrightarrow & H^3(\mathrm{Spin}(3);\mathbb{Z}) \\
1 & \longmapsto & (1,1) & \longmapsto & 2.
\end{array}$$

We originally defined $p_1 = -c_2$. Therefore, we see that $\Omega\frac{1}{2}p_1 \in H^3(\mathrm{Spin}(3);\mathbb{Z})$ is twice a generator, and in fact $\Omega\frac{1}{2}p_1 = -2\Omega c_2 \in H^3(S^3;\mathbb{Z})$, where Ωc_2 is the usual generator.

3. Harmonic representative of a string class

In Section 2 we showed that, up to homotopy, a string structure on a principal Spin(k)-bundle $P \to M$ is equivalent to a string class $\mathcal{S} \in H^3(P;\mathbb{Z})$. In this section we consider the harmonic representative of a string class. This will depend on the choice of a metric on a closed manifold M, a connection on P, and it involves taking an adiabatic limit. We also must pass from $\mathbb{Z}$ coefficients to $\mathbb{R}$ coefficients and lose torsion information. The harmonic representative of $\mathcal{S}$ is the Chern–Simons 3-form associated to the connection, minus a 3-form on M representing the differential cohomology class $\frac{1}{2}\check{p}_1(\Theta)$. The corresponding result holds for arbitrary $\lambda \in H^4(BG;\mathbb{Z})$, where G is a compact, simple, simply connected Lie group.

3a. ***Background: Differential characters.*** The canonical 3-form of Theorem 3.7 is best understood in the language of differential characters, originally developed in [Cheeger and Simons 1985]. See also [Freed 2002]. Let $C_i(M)$ and $Z_i(M)$ denote the group of smooth i-chains and cycles on M, respectively. Let $\Omega^i_{\mathbb{Z}}(M)$ denote the closed differential i-forms with integral periods; that is, their image in $H^i(M;\mathbb{R})$ lies in the image of $H^i(M;\mathbb{Z}) \to H^i(M;\mathbb{R})$. The group of differential characters $\check{H}^i(M)$ is defined as certain homomorphisms satisfying a transgression property:

$$\check{H}^i(M) := \{\chi : Z_{i-1}(M) \to \mathbb{R}/\mathbb{Z} \mid \text{there exists } \omega \in \Omega^i(M) \text{ satisfying} \\ \textstyle\int_\Sigma c^*\omega = \chi(\partial c) \mod \mathbb{Z} \text{ for all } c : \Sigma \to M \in C_i(M)\}$$

The form ω associated to a character χ must be unique, and in fact $\omega \in \Omega^i_{\mathbb{Z}}(M)$. The character σ also determines a cohomology class in $H^i(M;\mathbb{Z})$ whose image in $H^i(M;\mathbb{R})$ is the same as $[\omega]$. These two maps induce the short exact sequences

$$0 \to \frac{\Omega^{i-1}(M)}{\Omega^{i-1}_{\mathbb{Z}}(M)} \to \check{H}^i(M) \to H^i(M;\mathbb{Z}) \to 0, \tag{3.1}$$

$$0 \to H^{i-1}(M;\mathbb{R}/\mathbb{Z}) \to \check{H}^i(M) \to \Omega^i_{\mathbb{Z}}(M) \to 0, \tag{3.2}$$

$$0 \to \frac{H^{i-1}(M;\mathbb{R})}{H^{i-1}(M;\mathbb{Z})} \to \check{H}^i(M) \to H^i(M;\mathbb{Z}) \times_{H^i(M;\mathbb{R})} \Omega^i_{\mathbb{Z}}(M) \to 0. \tag{3.3}$$

In fact, these exact sequences uniquely characterize the groups $\check{H}^i(M)$ [Simons and Sullivan 2008]; one can refer to $\check{H}^*(M)$ as the differential cohomology of M without specifying the exact model being used, just as one refers to ordinary cohomology without specifying the model.

The importance of differential cohomology is due to the natural factoring of the Chern–Weil homomorphism through $\check{H}^*(M)$. Any compact Lie group G and universal class $\lambda \in H^{2i}(BG;\mathbb{Z})$ determine the following for any G-bundle $P \to M$ with connection Θ:

Characteristic class	$\lambda(P) \in H^{2i}(M;\mathbb{Z})$
Chern–Weil form	$\lambda(\Theta) \in \Omega^{2i}(M)$
Chern–Simons form	$\mathrm{CS}_\lambda(\Theta) \in \Omega^{2i-1}(P)$
Differential character	$\check{\lambda}(\Theta) \in \check{H}^{2i}(M)$

The integral class and form associated to $\check{\lambda}(\Theta)$ are $\lambda(P)$ and $\lambda(\Theta)$, respectively.

Suppose G is compact, semisimple and simply connected. Let $\lambda \in H^4(BG; \mathbb{Z})$. Then, as discussed in [Freed 1995], the associated Chern–Weil form is

$$\lambda(\Theta) = \langle \Omega \wedge \Omega \rangle \in \Omega^4(M),$$

where Ω is the curvature of Θ, and $\langle \cdot , \cdot \rangle$ is a suitably normalized Ad-invariant inner product on $\mathfrak{g}$. In this case, the Chern–Simons form is

$$\mathrm{CS}_\lambda(\Theta) = \langle \Theta \wedge \Omega \rangle - \tfrac{1}{6}\langle \Theta \wedge [\Theta \wedge \Theta] \rangle \in \Omega^3(P),$$

and

$$[i^* \mathrm{CS}_\lambda(\Theta)] = \Omega\lambda \in H^3(G; \mathbb{R}). \tag{3.4}$$

Suppose that $c : X \to M$ is a 3-cycle. The assumptions on G imply that $c^*P \to X$ admits a global section p. Then

$$\check{\lambda}(\Theta)(c) = \int_X p^*(c^* \mathrm{CS}_\lambda(\Theta)) \mod \mathbb{Z}.$$

Hence, the information contained in $\check{\lambda}(\Theta) \in \check{H}^4(M)$ is simply the $\mathbb{R}/\mathbb{Z}$-periods of the Chern–Simons 3-form. One is forced to only consider the $\mathbb{R}/\mathbb{Z}$-periods because different global sections will give different $\mathbb{R}$-periods. Note that when M is a connected oriented 3-manifold, (3.3) implies

$$\check{H}^4(M) \cong H^3(M; \mathbb{R})/H^3(M; \mathbb{Z}) \cong \mathbb{R}/\mathbb{Z},$$

and the isomorphism is given by evaluating on the fundamental cycle $[M]$. On a 3-manifold, the element $\check{\lambda}(\Theta) \in \check{H}^4(M) \cong \mathbb{R}/\mathbb{Z}$ is often called the Chern–Simons [1974] invariant or number of the connection Θ. This invariant motivated the theory of differential characters.

3b. *Hodge isomorphism on P.* A Riemannian metric g on an n-manifold M induces the Hodge star $* : \Lambda^i TM \to \Lambda^{n-i} TM$, creating the codifferential

$$d^* := (-1)^{n(i+1)+1} * d * : \Omega^i(M) \to \Omega^{i-1}(M).$$

The Hodge Laplacian is the operator

$$\Delta_g = dd^* + d^*d = (d + d^*)^2 : \Omega^i(M) \to \Omega^i(M).$$

When M is closed (compact with no boundary), classical Hodge theory states that there is a canonical isomorphism

$$H^i(M; \mathbb{R}) \xrightarrow[\simeq]{\Pi_{\mathrm{Ker}\,\Delta_g}} \mathrm{Ker}\,\Delta_g \subset \Omega^i(M).$$

We will later denote $\operatorname{Ker}\Delta_g$ by $\mathcal{H}^i(M)$, though the forms in $\mathcal{H}^3(P)$ will only be harmonic in a limit.

Let (M, g) be a closed Riemannian manifold, and let $P \xrightarrow{\pi} M$ be a principal G-bundle with connection Θ (G a compact, simple, simply connected Lie group). This naturally gives rise to a one-parameter family of right-invariant Riemannian metrics on P:

$$g_\delta := \delta^{-2}\pi^* g \oplus g_G \quad \text{for} \quad \delta > 0,$$

where g_G is any biinvariant metric on G. (The metric g_G exists since G is compact, and it is unique up to a scaling constant because G is simple.) Conceptually, g_δ is given by using the connection to decompose TP into horizontal and vertical spaces; the metrics on M and G determine metrics on the horizontal and vertical components, respectively.

For any $\delta > 0$, we have the harmonic forms $\operatorname{Ker}\Delta^3_{g_\delta} \subset \Omega^3(P)$. In general this finite-dimensional subspace varies with δ, and we will not be concerned with $\operatorname{Ker}\Delta_{g_\delta}$ for any particular δ. Instead, we analyze the *adiabatic limit*, the limit as $\delta \to 0$. Note that we had to choose the metric g_G. For this reason, it seems natural to introduce the scaling factor δ and take a limit, thus removing the dependence on the initial choice of g_G. Indeed, this is supported by concrete calculations, where the adiabatic limit appears to be of most interest.

Theorem 3.5 [Mazzeo and Melrose 1990; Dai 1991; Forman 1995]. *The* 1*-parameter space* $\operatorname{Ker}\Delta^i_{g_\delta} \subset \Omega^i(P)$ *smoothly extends to* $\delta = 0$. *Furthermore, there is a spectral sequence computing* $\lim_{\delta\to 0}\operatorname{Ker}\Delta_{g_\delta}$ *that is isomorphic to the Serre spectral sequence.*

This theorem holds in greater generality, and the context of each cited paper applies to the principal G-bundles with metric that we are considering. The spectral sequence mentioned is a Hodge-theoretic sequence, the details of which are given in [Forman 1995] and also summarized in [Redden 2008]. The fact that $\operatorname{Ker}\Delta_{g_\delta}$ extends continuously to $\delta = 0$ (as a path in Grassmannian space) implies that there is still a Hodge isomorphism

$$H^i(P;\mathbb{R}) \xrightarrow[\simeq]{\Pi_{\operatorname{Ker}\Delta_0}} \lim_{\delta\to 0}\operatorname{Ker}\Delta_{g_\delta} \subset \Omega^i(P).$$

We now introduce the notation

$$\begin{aligned}
\mathcal{H}^i(M) &:= \operatorname{Ker}\Delta_g \subset \Omega^i(M),\\
\mathcal{H}^i(P) &:= \lim_{\delta\to 0}\operatorname{Ker}\Delta_{g_\delta} \subset \Omega^i(P),\\
\mathcal{H}^i(G) &:= \operatorname{Ker}\Delta_{g_G} \subset \Omega^i(G).
\end{aligned}$$

In [Redden 2008], the spectral sequence interpretation of $\mathcal{H}^3(P)$ was used to give the following description of harmonic 3-forms on P in the adiabatic limit.

Theorem 3.6 [Redden 2008, Proposition 4.5 and Theorem 4.6]. *Consider the set* $(P \xrightarrow{\pi} M, g, \Theta)$, *where* G *is a compact simple Lie group. If* $\lambda(P) = 0 \in H^4(M; \mathbb{R})$, *then*

$$\mathcal{H}^3(P) = \mathbb{R}[\mathrm{CS}_\lambda(\Theta) - \pi^* h] \oplus \pi^* \mathcal{H}^3(M),$$

where $h \in \Omega^3(M)$ *is the unique coexact form satisfying* $dh = \lambda(\Theta)$.

When G is also simply connected, the Serre spectral sequence gives the following exact sequence, as seen in Proposition 2.5:

$$\begin{array}{ccccccccc} 0 \longrightarrow & H^3(M;\mathbb{Z}) & \xrightarrow{\pi^*} & H^3(P;\mathbb{Z}) & \xrightarrow{i^*} & H^3(G;\mathbb{Z}) & \xrightarrow{d_4} & H^4(M;\mathbb{Z}) \\ & & & \mathcal{S} & \overset{?}{\dashrightarrow} & \Omega\lambda & \longmapsto & \lambda(P) \end{array}$$

Theorem 3.7. *Consider* $(P \xrightarrow{\pi} M, g, \Theta)$ *where* G *is a simply connected compact simple Lie group. Suppose that* $\lambda(P) = 0 \in H^4(M; \mathbb{Z})$ *and that* $\mathcal{S} \in H^3(P; \mathbb{Z})$ *is a* λ*-trivialization class, that is,* $i^*\mathcal{S} = \Omega\lambda \in H^3(G; \mathbb{Z})$. *Then, the image of* $\mathcal{S}$ *under the Hodge isomorphism is of the form*

$$\begin{array}{ccc} H^3(P;\mathbb{Z}) \to H^3(P;\mathbb{R}) & \xrightarrow[\simeq]{\Pi_{\mathrm{Ker}\,\Delta_0}} & \mathcal{H}^3(P) \subset \Omega^3(P) \\ \mathcal{S} \mapsto & & \mathrm{CS}_\lambda(\Theta) - \pi^* H_{\mathcal{S},g,\Theta}, \end{array}$$

where $H_{\mathcal{S},g,\Theta} \in \Omega^3(M)$. *Alternatively,* $\Pi_{\mathrm{Ker}\,\Delta_0}\mathcal{S} - \mathrm{CS}_\lambda(\Theta) \in \pi^*\Omega^3(M)$.

Proof. The orthogonal decomposition of $\mathcal{H}^3(P)$ in Theorem 3.6 corresponds to a splitting $H^3(P; \mathbb{R}) \cong H^3(G; \mathbb{R}) \oplus H^3(M; \mathbb{R})$. We know $\pi^* H^3(M; \mathbb{R})$ restricts to $0 \subset H^3(G; \mathbb{R})$. As mentioned in (3.4), both $\mathrm{CS}_\lambda(\Theta) - \pi^* h$ and $\mathcal{S}$ cohomologically restrict to $\Omega\lambda \in H^3(G; \mathbb{R})$, so

$$\Pi_{\mathrm{Ker}\,\Delta_0}\mathcal{S} - (\mathrm{CS}_\lambda(\Theta) - \pi^* h) \in \pi^* \mathcal{H}^3(M).$$

Therefore, the harmonic representative of $\mathcal{S}$ must be of the form

$$\mathrm{CS}_\lambda(\Theta) - \pi^* h - \pi^* h',$$

with $h' \in \mathcal{H}^3(M)$, and we define $H_{\mathcal{S},g,\Theta} := h + h' \in \Omega^3(M)$. □

Remark 3.8. The theorem does not hold without taking an adiabatic limit. For a general $\delta > 0$,

$$\Pi_{\mathrm{Ker}\,\Delta_{g_\delta}}\mathcal{S} - \mathrm{CS}_\lambda(\Theta) \notin \pi^*\Omega^3(M),$$

but instead will contain forms with bidegree (2,1) and (1,2) in the (horizontal, vertical) decomposition of $\Omega^3(P)$ given by the connection.

Remark 3.9. When restricted to the fibers, the Chern–Simons form is the standard harmonic (biinvariant) form representing $\Omega\lambda$; that is, $i^* \mathrm{CS}_\lambda(\Theta) \in \mathcal{H}^3(G)$. Just as $\mathcal{S}$ is a cohomological extension of $\Omega\lambda$ to all of P, we see that $\Pi_{\operatorname{Ker}\Delta_0}\mathcal{S} = \mathrm{CS}_\lambda(\Theta) - \pi^* H_{\mathcal{S},g,\Theta}$ is a harmonic extension of $\Omega\lambda$ to all of P.

3c. ***Properties of canonical 3-form.*** Theorem 3.7 gives a canonical construction

$$\{\lambda\text{-triv classes}\} \times \mathrm{Met}(M) \times \mathcal{A}(P) \to \Omega^3(M), \quad \mathcal{S}, g, \Theta \mapsto H_{\mathcal{S},g,\Theta}. \tag{3.10}$$

We call $H_{\mathcal{S},g,\Theta}$ the canonical 3-form associated to $(\mathcal{S}, g, \Theta)$. While Theorem 3.7 only uses information about $\mathcal{S}$ as a class in $H^3(P;\mathbb{R})$, the integrality becomes necessary when understanding $H_{\mathcal{S},g,\Theta}$ in terms of differential characters. The exact sequence (3.1) gives rise to

$$\begin{gathered} 0 \to \Omega^3_{\mathbb{Z}}(M) \to \Omega^3(M) \to \check{H}^4(M) \to H^4(M;\mathbb{Z}) \to 0 \\ H_{\mathcal{S},g,\Theta} \mapsto \check{H}_{\mathcal{S},g,\Theta}, \end{gathered} \tag{3.11}$$

where the character $\check{H}_{\mathcal{S},g,\Theta}$ obtained via $\Omega^3(M) \to \check{H}^4(M)$ is given by simply by integrating $H_{\mathcal{S},g,\Theta}$ on cycles and reducing mod $\mathbb{Z}$. Also, note that $H^3(M;\mathbb{Z})$ acts naturally on $\{\lambda\text{-triv classes}\}$, and it also acts on $\Omega^3(M) \times \mathrm{Met}(M)$ by adding a harmonic representative.

Proposition 3.12. *The construction* (3.10) *is equivariant with respect to the natural action of* $H^3(M;\mathbb{Z})$; *that is,* $H_{\mathcal{S}+\pi^*\phi,g,\Theta} = H_{\mathcal{S},g,\Theta} + \Pi_{\operatorname{Ker}\Delta_g}\phi$.
Furthermore, the forms $H_{\mathcal{S},g,\Theta}$ *satisfy the following*:

- $d^* H_{\mathcal{S},g,\Theta} = 0 \in \Omega^2(M)$,
- $dH_{\mathcal{S},g,\Theta} = \lambda(\Theta) \in \Omega^4(M)$,
- $\check{H}_{\mathcal{S},g,\Theta} = \check{\lambda}(\Theta) \in \check{H}^4(M)$.

Proof. The action of $H^3(M;\mathbb{Z})$ on λ-trivialization classes is given by addition under π^*, and the action on $\Omega^3(M)$ is given by adding the harmonic representative (with respect to a fixed metric M). Theorem 3.6 implies that for $\phi \in H^3(M;\mathbb{Z})$ with harmonic representative $\Pi_{\operatorname{Ker}\Delta_g}\phi \in \mathcal{H}^3(M)$,

$$\pi^*(\Pi_{\operatorname{Ker}\Delta_g}\phi) = \Pi_{\operatorname{Ker}\Delta_0}(\pi^*\phi) \in \Omega^3(P).$$

The property $d^* H_{\mathcal{S},g,\Theta} = 0$ also follows directly from Theorem 3.6. That $\Pi_{\operatorname{Ker}\Delta_0}\mathcal{S}$ is closed implies

$$\begin{aligned} d(\mathrm{CS}_\lambda(\Theta) - \pi^* H_{\mathcal{S},g,\Theta}) &= 0, \\ \pi^*\lambda(\Theta) - \pi^* dH_{\mathcal{S},g,\Theta} &= 0, \\ dH_{\mathcal{S},g,\Theta} &= \lambda(\Theta), \end{aligned}$$

with the last equality following from π^* being injective on forms.

Finally, suppose that $X \xrightarrow{c} M$ is a smooth 3-cycle on M. Then the value of $\check{\lambda}(\Theta)$ on (X, c) is

$$\check{\lambda}(\Theta)(c) = c^*\check{\lambda}(\Theta) \in \check{H}^4(X) \cong \mathbb{R}/\mathbb{Z}.$$

Now, standard obstruction theory implies that $c^*P \to X$ admits a global section $p : X \to c^*P$, and it is easy to see that

$$p^*c^*\check{\mathrm{CS}}_\lambda(\Theta) = p^*\check{\mathrm{CS}}_\lambda(c^*\Theta) = c^*\check{\lambda}(\Theta) \in \check{H}^4(X).$$

Because $\mathrm{CS}_\lambda(\Theta) - \pi^* H_{\mathcal{S},g,\Theta} \in \Omega^3_{\mathbb{Z}}(P)$, we have $\check{\mathrm{CS}}_\lambda(\Theta) = \pi^*\check{H}_{\mathcal{S},g,\Theta} \in \check{H}^3(P)$ and hence

$$\begin{aligned} c^*\check{\lambda}(\Theta) = p^*c^*\check{\mathrm{CS}}_\lambda(\Theta) &= p^*c^*\pi^*\check{H}_{\mathcal{S},g,\Theta} \\ &= p^*\pi^*c^*\check{H}_{\mathcal{S},g,\Theta} = c^*\check{H}_{\mathcal{S},g,\Theta} \in \check{H}^4(X). \end{aligned}$$

This implies that for all 3-cycles $X \xrightarrow{c} M$

$$\check{H}_{\mathcal{S},g,\Theta}(c) = \check{\lambda}(\Theta)(c),$$

and hence $\check{\lambda}(\Theta) = \check{H}_{\mathcal{S},g,\Theta} \in \check{H}^4(M)$. (This also implies $dH_{\mathcal{S},g,\Theta} = \lambda(\Theta)$.) □

Integrating the form $H_{\mathcal{S},g,\Theta}$ naturally gives values in $\mathbb{R}$, and Proposition 3.12 says that reducing mod $\mathbb{Z}$ gives the same values as $\check{\lambda}(P)$. In other words, the choice of a λ-trivialization class naturally gives a lift

(3.13)
$$\begin{array}{ccc} & & \mathbb{R} \\ & \overset{H_{\mathcal{S},g,\Theta}}{\nearrow} & \big\downarrow \\ Z_3(M) & \xrightarrow{\check{\lambda}(\Theta)} & \mathbb{R}/\mathbb{Z}, \end{array}$$

and the action of $H^3(M; \mathbb{Z})$ modifies the lift by the induced map $Z_3(M) \to \mathbb{Z}$. While the actual form $H_{\mathcal{S},g,\Theta}$ depends on the choice of a metric, this lift does not.

Proposition 3.14. *The lift in* (3.13) *is independent of the choice of metric g.*

Proof. If g_0 and g_1 are two different metrics, then (3.11) implies

$$H_{\mathcal{S},g_1,\Theta} - H_{\mathcal{S},g_0,\Theta} \in \Omega^3_{\mathbb{Z}}(M).$$

The space of Riemannian metrics is contractible, so

$$H_{\mathcal{S},g_1,\Theta} - H_{\mathcal{S},g_0,\Theta} \in d\Omega^2(M).$$ □

The role of the metric in (3.10) is to pick out the forms $H_{\mathcal{S},g,\Theta}$ with smallest norm still satisfying $\check{H}_{\mathcal{S},g,\Theta} = \check{\lambda}(\Theta)$. We denote the lift by $H_{\mathcal{S},\Theta} \in \check{H}^4_{\mathbb{R}}(M)$. Here we use the nonstandard notation of $\check{H}^4_{\mathbb{R}}(M)$ to denote characters $Z_3(M) \to \mathbb{R}$ satisfying the usual transgression assumption.

In summary, the construction (3.10) induces lifts of the standard differential character construction, which are encoded in the following diagram:

$$\begin{array}{ccc} \{\lambda\text{-triv classes}\} \times \mathrm{Met}(M) \times \mathscr{A}(P) & \xrightarrow{H_{\mathscr{S},g,\Theta}} & \Omega^3(M) \\ \downarrow & & \downarrow \\ \{\lambda\text{-triv classes}\} \times \mathscr{A}(P) & \xrightarrow{H_{\mathscr{S},\Theta}} & \check{H}^4_{\mathbb{R}}(M) \\ \downarrow & & \downarrow \\ \mathscr{A}(P) & \xrightarrow{\check{\lambda}(\Theta)} & \check{H}^4(M). \end{array}$$

Remark 3.15. Stolz and Teichner [2004] define a *geometric* trivialization of $\lambda(P)$ as a trivialization of the extended Chern–Simons field theory on $P \to M$. This includes defining a lift of the differential character $\check{\lambda}(\Theta)$ to take values in $\mathbb{R}$, and it aligns nicely with the construction above. In fact, if $H^3(M;\mathbb{Z})$ has no torsion, then the choice of a lift of $\check{\lambda}(\Theta)$ to $\check{H}^4_{\mathbb{R}}(M)$ is equivalent to the choice of a λ-trivialization class. Waldorf [2009] gives an explicit model for string structures in terms of trivializations of a Chern–Simons 2-gerbe, and shows that a string structure produces a 3-form on M. The 3-forms obtained in our construction are a proper subset of those he obtained, analogous to the relationship between forms representing a de Rham class and harmonic forms.

Note that one can also directly define the lift $H_{\mathscr{S},\Theta}$ without using the Hodge isomorphism. On a 3-cycle $c : X \to M$,

$$H_{\mathscr{S},\Theta}(c) = \int_X p^*(\mathrm{CS}_\lambda(c^*\Theta) - c^*S), \tag{3.16}$$

where p is any global section, and S is any de Rham representative of $\mathscr{S}$. This is a simple consequence of $S = \mathrm{CS}_\lambda(\Theta) - H_{\mathscr{S},g,\Theta} + d\beta$. It is also easy to verify that the integral on the right side is independent of p. In cases like Lemma 3.18, this allows us to calculate the form $H_{\mathscr{S},g,\Theta}$ without solving a differential equation.

Suppose the G-bundle $P \xrightarrow{\pi} M$ is topologically trivial. Then, the choice of a global section $p : M \to P$ is equivalent to a trivialization $P \cong M \times G$. The canonical λ-trivialization on $M \times G$ induces one on P, and the corresponding cohomology class is given by the Kunneth isomorphism

$$H^3(P;\mathbb{Z}) \cong H^3(M;\mathbb{Z}) \oplus H^3(G;\mathbb{Z}), \quad \mathscr{S} \leftrightarrow (0, \Omega\lambda). \tag{3.17}$$

Lemma 3.18. *Suppose $P \xrightarrow{\pi} M$ is a trivial bundle with λ-trivialization class $\mathscr{S}$ induced by the trivialization $p : M \to P$. Then,*

$$H_{\mathscr{S},g,\Theta} - p^*\,\mathrm{CS}_\lambda(\Theta) \in d\Omega^2(M).$$

In particular, $p^*\,\mathrm{CS}_\lambda(\Theta) = H_{\mathcal{S},\Theta}$ *as elements of* $\check{H}^4_{\mathbb{R}}(M)$. *If* $d^*p^*\,\mathrm{CS}_\lambda(\Theta) = 0$, *then* $p^*\,\mathrm{CS}_\lambda(\Theta) = H_{\mathcal{S},g,\Theta} \in \Omega^3(M)$.

Proof. As seen in (3.17), $p^*\mathcal{S} = 0 \in H^3(M;\mathbb{Z})$. Therefore, (3.16) simplifies to

$$\int_X c^* H_{\mathcal{S},g,\Theta} = \int_X c^* p^*\,\mathrm{CS}_\lambda(\Theta)$$

for all 3-cycles $c : X \to M$, so $[H_{\mathcal{S},g,\Theta} - p^*\,\mathrm{CS}_\lambda(\Theta)] = 0 \in H^3(M;\mathbb{R})$. □

One usually chooses $\lambda \in H^4(BG;\mathbb{Z}) \cong \mathbb{Z}$ to be the generator. This is because $\widetilde{BG_\lambda}$ is the universal extension. This universality is also reflected in the associated canonical 3-forms.

Proposition 3.19. *If* $\mathcal{S} \in H^3(P;\mathbb{Z})$ *is a* λ*-trivialization class and* $\ell \in \mathbb{Z}$, *then* $\ell\mathcal{S}$ *is an* $\ell\lambda$*-trivialization class, and* $H_{\ell\mathcal{S},g,\Theta} = \ell H_{\mathcal{S},g,\Theta} \in \Omega^3(P)$.

Proof. The first statement is obvious, and the second follows from the linearity of the Hodge isomorphism. □

We now apply the construction above to $G = \mathrm{Spin}(k)$ for $k \geq 3$ with $\lambda = \frac{1}{2}p_1 \in H^4(B\mathrm{Spin}(k);\mathbb{Z})$ to canonically produce 3-forms associated to string structures. Since $\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2)$ is not simple, we define the canonical 3-form when $k = 4$ to be the one obtained by stabilizing to Spin(5), a process that does not affect $\frac{1}{2}\check{p}_1(\Theta)$.[2]

Theorem 3.20. *Let* $P \xrightarrow{\pi} M$ *be a principal* $\mathrm{Spin}(k)$*-bundle* ($k \geq 3$) *with connection* Θ *over the Riemannian manifold* (M, g). *Under the Hodge isomorphism* (*in an adiabatic limit*), *a string class* $\mathcal{S} \in H^3(P;\mathbb{Z})$ *is represented by*

$$\Pi_{\mathrm{Ker}\,\Delta_0}\mathcal{S} = \mathrm{CS}_{\frac{1}{2}p_1}(\Theta) - \pi^* H_{\mathcal{S},g,\Theta} \in \Omega^3(P).$$

The canonical form $H_{\mathcal{S},g,\Theta} \in \Omega^3(M)$ *is such that*

- $d^* H_{\mathcal{S},g,\Theta} = 0 \in \Omega^2(M)$,
- $\check{H}_{\mathcal{S},g,\Theta} = \frac{1}{2}\check{p}_1(\Theta) \in \check{H}^4(M)$, *and*
- *the construction of* $H_{\mathcal{S},g,\Theta}$ *is equivariant with respect to the natural action of* $H^3(M;\mathbb{Z})$.

In particular, consider the case where (M, g) is a Riemannian manifold with spin structure satisfying $\frac{1}{2}p_1(M) = 0 \in H^4(M;\mathbb{Z})$. Then, we can let $P = \mathrm{Spin}(TM)$, and we call a string structure on $\mathrm{Spin}(TM)$ a string structure on M. Letting Θ be the Levi-Civita connection, this gives a map

$$\{\text{String classes on } M\} \times \mathrm{Met}(M) \to \Omega^3(M), \quad \mathcal{S}, g \mapsto H_{\mathcal{S},g}. \tag{3.21}$$

[2]The arguments in [Redden 2008] can be extended to semisimple groups, and Theorem 3.20 also holds for $k = 4$ without stabilizing.

4. Canonical 3-forms and the string orientation of tmf

We now review how string structures arise and give a possible new application of the canonical 3-forms $H_{\mathcal{G},g}$ from (3.21). First recall some classical results from index theory; an excellent source is [Lawson and Michelsohn 1989]. Suppose M is an oriented closed manifold. A priori, one cannot form a spinor bundle $\mathrm{SO}(M) \times_{\mathrm{SO}(n)} S^{\pm} \to M$, because the spinor representations $\mathrm{SO}(n) \to \mathrm{GL}(S^{\pm})$ are only projective. The choice of a spin structure, discussed in Section 2a, allows one to define the spinor bundle $S^{\pm}_M := \mathrm{Spin}(M) \times_{\mathrm{Spin}(n)} S^{\pm}$ and Dirac operator $\not{D}_M : \Gamma(S^{\pm}) \to \Gamma(S^{\mp})$.

While the Fredholm operator $\not{D}_M$ depends on the spin structure, the Atiyah–Singer index theorem states that its index does not, and in fact

$$\mathrm{index}(\not{D}_M) = \hat{A}(M) \in \mathbb{Z}.$$

Here, $\hat{A}(M)$ is a topological invariant determined by a manifold's Pontryagin classes and is defined for any oriented manifold. In general $\hat{A}(M) \in \mathbb{Q}$, but $\hat{A}(M) \in \mathbb{Z}$ when $w_2(M) = 0$. There is also a refinement of $\hat{A}(M)$ given by the spin orientation $\alpha : M\mathrm{Spin} \to \mathrm{KO}$. This refinement can be thought of as the Clifford-linear index, and it does depend on the spin structure.

(4.1)
$$\begin{array}{ccc} & & \mathrm{KO}^{-n}(pt) \\ & \overset{\alpha}{\nearrow} & \downarrow \\ M\mathrm{Spin}^{-n}(pt) & \xrightarrow{\hat{A}} & \mathbb{Z}. \end{array}$$

The KO-invariants usually appear in family index theorems, but they also contain interesting information for a single manifold due to the torsion in $\mathrm{KO}^{-*}(pt)$.

Index theory is now a central part of mathematics, and one of its powerful applications is to the problem of when a closed manifold admits positive scalar curvature metrics. The Lichnerowicz–Weitzenböck formula $\not{D}_M^2 = \nabla^*\nabla + \frac{1}{4}s$, which relates $\not{D}_M$ to a positive operator and the scalar curvature s, implies the following: If a closed spin manifold M admits a metric of positive scalar curvature, then $\mathrm{index}(\not{D}_M) = \hat{A}(M) = 0$ [Lichnerowicz 1962]. Furthermore, $\alpha[M] = 0 \in \mathrm{KO}^{-n}(pt)$ for all spin structures [Hitchin 1974]. In fact, for simply connected spin manifolds of dimension ≥ 5, all the α-invariants vanish if and only if M admits a metric of positive scalar curvature [Stolz 1992].

There is an analogous story, though not fully developed, involving the Witten genus, index theory on loop spaces, and elliptic cohomology. Witten [1988] used intuition from theoretical physics and defined a topological invariant $\varphi_W(M)$ known as the Witten genus. He claimed it should be the S^1-equivariant index of the Dirac operator on the free loop space LM; that is,

$$\text{“index}^{S^1} \not{D}_{LM}\text{”} = \varphi_W(M).$$

We place the left side in quotes because of analytic difficulties in defining a good theory of Fredholm operators on infinite-dimensional manifolds. However, the Witten genus (and other elliptic genera) are well defined, and one can make formal sense of index theory on LM by using localization formulas or the representation theory of loop groups. For a good overview on these ideas, see [Liu 1996]. While $\varphi_W(M) \in \mathbb{Q}[[q]][q^{-1}]$ for any oriented manifold, for a string manifold, $\varphi_W(M)$ is the q-expansion of a modular form (MF) with integer coefficients and weight $n/2$, and we say $\varphi_W(M) \in MF_n$. The intuitive reason is that when $\frac{1}{2}p_1(M) = 0$, one can define the spinor bundle on LM [Coquereaux and Pilch 1989]. We wish to form $L\,\mathrm{Spin}(M) \times_{L\,\mathrm{Spin}(n)} S \to LM$, where S is a positive energy representation of $L\,\mathrm{Spin}(n)$. However, these representations are all projective, so one must pass to an S^1-extension $L\,\widehat{\mathrm{Spin}(n)} \to L\,\mathrm{Spin}(n)$. Topologically, our string class $\mathcal{S} \in H^3(\mathrm{Spin}(M); \mathbb{Z})$ transgresses to a class in $H^2(L\,\mathrm{Spin}(M); \mathbb{Z})$ that defines an isomorphism class of S^1-extension $L\,\widehat{\mathrm{Spin}(M)} \to L\,\mathrm{Spin}(M) \to LM$. We say that a string structure on M transgresses to a spin structure on LM (though in this paper we have only discussed isomorphism classes of such structures). This led to the following conjecture.

Conjecture 4.2 (Höhn and Stolz [Stolz 1996]). *Let M be a closed oriented n-manifold admitting spin and string structures. If M admits a metric of positive Ricci curvature, then the Witten genus $\varphi_W(M)$ vanishes.*

Stolz's heuristic argument comes from the hope that there is some Weizenböck-type formula such that positive Ricci curvature on M implies positive scalar curvature on LM, which in turn implies $\mathrm{Ker}(\not{D}_{LM}) = \varphi_W(M) = 0$. Though this reasoning is far from rigorous, there are no known counterexamples, and the conjecture holds true for homogeneous spaces and complete intersections. To the author's knowledge, there are no known examples of simply connected closed manifolds admitting metrics of positive scalar curvature, but not metrics of positive Ricci curvature. If the conjecture is true, it would provide examples of such manifolds.

Just as KO-theory refines the $\hat{A}$-genus, there is a cohomology theory tmf, or topological modular forms, with string-orientation refining the Witten genus (see [Hopkins 2002]):

$$\begin{array}{ccc} & & \mathrm{tmf}^{-n}(pt) \\ & \overset{\sigma}{\nearrow} & \downarrow \\ M\mathrm{String}^{-n}(pt) & \xrightarrow{\ \varphi_W\ } & MF_n \end{array}$$

The map $\mathrm{tmf}^{-*}(pt) \to MF_*$ is a rational isomorphism, but it is not integrally surjective or injective. In particular, $\mathrm{tmf}^{-*}(pt)$ contains a great deal of torsion. While defining tmf is a subtle process, informally tmf is the universal elliptic cohomology theory, or the elliptic cohomology theory associated to the universal moduli stack of elliptic curves. Despite several attempts [Baas et al. 2004; Hu and Kriz 2004; Segal 1988; Stolz and Teichner 2004], there is still no geometric description of tmf. However, it is believed that tmf should provide a natural home for family index theorems on loop spaces.

One might hope that all the refined invariants in tmf also vanish for string manifolds admitting positive Ricci curvature metrics, giving an analogy of Hitchin's theorem. However, there exist a fair number of compact nonabelian Lie groups (thus admitting positive Ricci curvature metrics) that are sent to torsion elements in $\mathrm{tmf}^{-*}(pt)$ via their left-invariant framing [Hopkins 2002]. In Section 6, we investigate the case where $M = S^3$.

Conceptually, this is still compatible with the analogy to classical index theory. The group $\mathrm{Spin}(n)$ is a discrete cover of $\mathrm{SO}(n)$, so there are no local differences between the bundles $\mathrm{Spin}(M)$ and $\mathrm{SO}(M)$ and their connections. However, $\mathrm{Spin}^c(n) \to \mathrm{SO}(n)$ is an S^1-extension, and one must choose a connection on the S^1-bundle $\mathrm{Spin}^c(M) \to \mathrm{SO}(M)$. The curvature of this connection appears in the Weizenböck formula for the spin^c Dirac operator. Since $\mathrm{String}(M) \to \mathrm{Spin}(M)$ has $K(\mathbb{Z}, 2)$-fibers, string structures are more analogous to spin^c structures. When constructing the S^1-extension $L\,\widehat{\mathrm{Spin}(M)} \to L\,\mathrm{Spin}(M)$, one really needs an S^1-extension with connection [Coquereaux and Pilch 1989]. The form $\mathrm{CS}_{\frac{1}{2}p_1}(g) - \pi^* H_{\mathcal{S},g} \in \Omega^3(\mathrm{Spin}(M))$ representing $\mathcal{S}$ transgresses to the curvature (minus a canonically defined term) of this connection on $L\,\mathrm{Spin}(M)$. One would reasonably expect any Weizenböck-type formula for $\not{D}_{LM}$ to also involve the form $H_{\mathcal{S},g}$. We ask the following question in an attempt to formulate a connection between tmf and obstructions for certain types of curvature.

Question 4.3. Let M be a closed n-dimensional manifold with spin structure such that $\frac{1}{2}p_1(M) = 0 \in H^4(M; \mathbb{Z})$, and let $\mathcal{S}$ be a specified string class. Suppose there exists a metric g such that

$$\mathrm{Ric}(g) > 0 \quad \text{and} \quad H_{\mathcal{S},g} = 0 \in \Omega^3(M).$$

Does this imply that

$$\sigma[M, \mathcal{S}] = 0 \in \mathrm{tmf}^{-n}(pt)?$$

Remark 4.4. The condition $H_{\mathcal{S},g} = 0$ for some string class $\mathcal{S}$ is equivalent to $\frac{1}{2}\check{p}_1(g) = 0 \in \check{H}^4(M)$. This is a strong condition and is not usually satisfied for generic metrics. While a great deal of information about the characters $\frac{1}{2}\check{p}_1(g)$

is known for certain manifolds, the author is not aware of any general results guaranteeing the existence or nonexistence of such metrics.

Remark 4.5. The condition $H_{\mathcal{S},g} = 0$ is conformally invariant; if $H_{\mathcal{S},g} = 0$, then $H_{\mathcal{S},e^f g} = 0$ for any conformally related metric $e^f g$. This follows from the conformal invariance of $\frac{1}{2}\check{p}_1(g)$ and the fact that $0 \in \mathcal{H}^3(M)$ for all metrics.

We close this discussion by noting that $\not{D}_M$ and $\not{D}_{LM}$ can both be thought of as partition functions of certain 1- and 2-dimensional supersymmetric nonlinear sigma models [Witten 1999]. These sigma models require spin and string structures, respectively. In the 2-dimensional sigma models, the form $H_{\mathcal{S},g}$ is used to trivialize the natural connection on a certain determinant line bundle [Witten 1999; Alvarez and Singer 2002]. Sometimes, terms in the action of these sigma models are combined and written as the connection $\nabla^{\mathcal{S},g}$ discussed in Section 5.

Stolz and Teichner [2004] have shown that KO^{-n} is homotopy equivalent to the space of supersymmetric 1-dimensional Euclidean field theories of degree n, and the spin orientation is (up to homotopy) given by the previously mentioned sigma model. The hope is that the analogous statement should hold for 2-dimensional field theories with the string orientation σ given by these sigma models. In this context, Question 4.3 is essentially asking, If one does not have to add in the terms $H_{\mathcal{S},g}$, does positivity of the Ricci curvature imply that the corresponding sigma model is qualitatively trivial?

5. Metric connections with torsion

Question 4.3 can be reformulated in terms of the Ricci curvature of a metric connection with torsion. Given a string class and metric $(\mathcal{S}, g)$, we define the torsion tensor $T^{\mathcal{S},g}$ by

$$T^{\mathcal{S},g} := g^{-1} H_{\mathcal{S},g} \in \Omega^1(M; \mathfrak{gl}(TM)),$$

where $H_{\mathcal{S},g}$ is the canonical 3-form from (3.21). This is simply a case of "raising indices" and is equivalent to saying $g(T_X^{\mathcal{S},g} Y, Z) = H_{\mathcal{S},g}(X, Y, Z)$, or in coordinates $T_{ij}^k = g^{rk} H_{ijr}$. Then

$$\nabla^{\mathcal{S},g} := \nabla^g + \tfrac{1}{2} T^{\mathcal{S},g}$$

is a metric connection with torsion $T^{\mathcal{S},g}$, where ∇^g is the Levi-Civita connection.

In general, torsion tensor T of a connection is called *totally skew-symmetric* if $gT \in \Omega^3(M)$, that is, $g(T(\cdot\,,\cdot),\cdot)$ is skew-symmetric in all three variables. By construction, $\nabla^{\mathcal{S},g}$ is a metric connection with totally skew-symmetric torsion. We also note that the connection $\nabla^{\mathcal{S},g}$ still preserves the geodesics of the Levi-Civita connection. In general for a fixed metric g, we have the following equalities of subsets of connections on TM:

$$\{\text{Metric connections}\} = \left\{\begin{matrix}\text{Metric connections}\\ \text{with } \nabla^g\text{-geodesics}\end{matrix}\right\} = \left\{\begin{matrix}\text{Metric connections with}\\ \text{totally skew-symmetric}\\ \text{torsion}\end{matrix}\right\}$$

One can easily prove this by writing any connection ∇ as $\nabla^g + A$ and plugging into the geodesic equation $\nabla_X X = 0$ and metric equation $g(\nabla_X Y, Z) = -g(Y, \nabla_X Z)$.

For a torsion connection $\nabla^T = \nabla^g + \frac{1}{2}T$, we can still define the curvature tensor

$$R^T_{X,Y} Z := (\nabla^T_X \nabla^T_Y - \nabla^T_Y \nabla^T_X - \nabla^T_{[X,Y]})Z,$$

and Ricci tensor

$$\operatorname{Ric}^T(X, Y) := \sum_i g(R^T_{e_i,X} Y, e_i),$$

where $\{e_i\}$ is any orthonormal basis. We let Ric^g denote the Ricci tensor of the Levi-Civita connection.

Lemma 5.1. *Suppose that $\nabla^T = \nabla^g + \frac{1}{2}T$ is a metric connection with totally skew-symmetric torsion satisfying $gT = H \in \Omega^3(M)$. Then the Ricci tensor satisfies*

$$\operatorname{Ric}^T(X, Y) = \operatorname{Ric}^g(X, Y) - \tfrac{1}{4}\textstyle\sum_i g(T_{e_i} X, T_{e_i} Y) - \tfrac{1}{2} d^* H(X, Y).$$

Proof. Let $\langle \cdot , \cdot \rangle$ denote $g(\cdot , \cdot)$. Simply expanding using $\nabla^T = \nabla^g + \frac{1}{2}T$, we get

$$\begin{aligned}\langle R^T_{e_i,X} Y, e_i\rangle &= \langle \nabla^T_{e_i}\nabla^T_X Y - \nabla^T_X \nabla^T_{e_i} Y - \nabla^T_{[e_i,X]} Y, e_i\rangle \\ &= \langle R^g_{e_i,X} Y, e_i\rangle - \tfrac{1}{4}\langle T_X T_{e_i} Y, e_i\rangle + \tfrac{1}{2}\langle \nabla^g_{e_i} T_X Y - T_X \nabla^g_{e_i} Y - T_{\nabla^g_{e_i} X} Y, e_i\rangle \\ &\qquad + \tfrac{1}{2}\langle T_{\nabla^g_{e_i} X - \nabla^g_X e_i - [e_i,X]} Y, e_i\rangle \\ &= \langle R^g_{e_i,X} Y, e_i\rangle - \tfrac{1}{4}\langle T_{e_i} X, T_{e_i} Y\rangle + \tfrac{1}{2}\langle (\nabla^g_{e_i} T)(X, Y), e_i\rangle.\end{aligned}$$

The last term is easily seen to be a tensor. Using a normal orthonormal frame $\{e_i\}$ at a point (that is, $\nabla_{e_i} e_j = 0$), one easily calculates that

$$\textstyle\sum_i \langle (\nabla^g_{e_i} T)(e_j, e_k), e_i\rangle = \sum_i \partial_i T^i_{jk} = \sum_i \partial_i H_{ijk} = -d^* H(e_j, e_k). \qquad \square$$

The usual Ricci tensor Ric^g is symmetric, and Lemma 5.1 shows that the skew-symmetric part of Ric^T is $-\frac{1}{2}d^*H$. Because the canonical form $H_{\mathcal{S},g}$ satisfies $d^* H_{\mathcal{S},g} = 0$, it gives rise to a metric connection $\nabla^{\mathcal{S},g}$ with symmetric Ricci tensor. For an arbitrary metric connection ∇^T, we refer to the Ricci curvature $\operatorname{Ric}^T(X) := \operatorname{Ric}^T(X, X)$ as the symmetric component of Ric^T, which satisfies

$$\operatorname{Ric}^g(X) - \operatorname{Ric}^T(X) = \tfrac{1}{4}\textstyle\sum_i \|T_{e_i} X\|^2 \geq 0,$$

with equality for all X precisely when $T = 0$. This gives an alternative description of the Levi-Civita connection.

Corollary 5.2. *For a fixed Riemannian metric g, the Levi-Civita connection is the unique metric connection that maximizes the Ricci curvature.*

One convenient property of both the Levi-Civita connection and the usual Ricci tensor is the invariance under a global scaling. A quick check shows that for $\epsilon > 0$,

$$\mathrm{Ric}^{\epsilon g}(X) = \sum_i \epsilon g(R^{\epsilon g}_{\epsilon^{-1/2}e_i, X} X, \epsilon^{-1/2} e_i) = \sum_i g(R^g_{e_i, X} X, e_i) = \mathrm{Ric}^g(X).$$

The form $H_{\mathcal{S},g}$ was constructed using a Hodge isomorphism in an adiabatic limit, giving us the scale invariance $H_{\mathcal{S},\epsilon g} = H_{\mathcal{S},g}$. However, we use the metric to change $H_{\mathcal{S},g}$ into a torsion tensor. Therefore,

$$T^{\mathcal{S},\epsilon g} = (\epsilon g)^{-1} H_{\mathcal{S},\epsilon g} = \epsilon^{-1} T^{\mathcal{S},g}.$$

It is more natural then to consider the 1-parameter family of connections $\nabla^{\mathcal{S},\epsilon g}$ than any fixed $\nabla^{\mathcal{S},g}$. In the large volume limit, as $\epsilon \to \infty$, this connection converges to the Levi-Civita connection ∇^g. In the small volume limit, as $\epsilon \to 0$, the terms $T^{\mathcal{S},\epsilon g}$ blow up and $\nabla^{\mathcal{S},\epsilon g}$ does not converge to a connection unless $H_{\mathcal{S},g} = 0$.

Question 5.3. Let $(M, g, \mathcal{S})$ be an n-dimensional Riemannian manifold with string class. Suppose that the Ricci tensor of the modified connection $\nabla^{\mathcal{S},g}$ is strictly positive in the small volume scaling limit; that is

$$\lim_{\epsilon \to 0} \mathrm{Ric}(\nabla^{\mathcal{S},\epsilon g}) > 0.$$

Does this imply $\sigma[M, \mathcal{S}] = 0 \in \mathrm{tmf}^{-n}(pt)$?

Proposition 5.4. *Question 4.3 is equivalent to Question 5.3.*

Proof. This follows directly from the description of the Ricci tensor in Lemma 5.1, which implies

$$\mathrm{Ric}^{\epsilon g, \mathcal{S}}(X) = \mathrm{Ric}^g(X) - \tfrac{1}{4}\epsilon \textstyle\sum_i \|T_{e_i} X\|^2.$$

Consequently, if $H_{\mathcal{S},g} \neq 0$, then $\mathrm{Ric}^{\epsilon g,\mathcal{S}}(X) \xrightarrow{\epsilon \to 0} -\infty$ for some X. The simultaneous conditions $\mathrm{Ric}(g) > 0$ and $H_{\mathcal{S},g} = 0$ are equivalent to $\mathrm{Ric}(\nabla^{\mathcal{S},\epsilon g}) > 0$ for arbitrarily small ϵ. □

6. Homogeneous metrics on S^3

We now examine the canonical 3-forms obtained when $M = S^3$ with a homogeneous metric, and we compare the results with Question 4.3. We see that it has an affirmative answer in this special situation, but it would not if the conditions were weakened. In particular, there exists a 1-dimensional family of left-invariant metrics g with nonnegative Ricci curvature such that the right-invariant framing $\mathcal{R}$ produces $H_{\mathcal{R},g} = 0$ and $\sigma[S^3, \mathcal{R}] \neq 0 \in \mathrm{tmf}^{-3}(pt)$. The previous sentence is also true with left and right swapped.

6a. ***String structures on S^3.*** Using the isomorphism $S^3 \cong \mathrm{SU}(2) \cong \mathrm{Sp}(1)$, the left- and right-invariant framings induce two string classes, which we denote $\mathscr{L}$ and $\mathscr{R}$. The disc D^4 inherits a standard framing from its inclusion $D^4 \subset \mathbb{R}^4$, and this restricts to a framing of the stable tangent bundle for $\partial D^4 = S^3$. We denote the induced string class by ∂D^4 and note that, by construction, the string bordism class $[S^3, \partial D^4] = 0 \in M\mathrm{String}^{-3}(pt)$.

The set of string classes is a torsor for $H^3(S^3; \mathbb{Z}) \cong \mathbb{Z}$, an affine copy of $\mathbb{Z}$. In other words, the difference between any two string classes is naturally an integer. We now determine where the three previously defined string classes live on this affine line, and we use $\Omega c_2 \in H^3(S^3; \mathbb{Z})$ as our standard generator. The left and right framings are related by

$$S^3 \times \mathrm{Spin}(3) \xrightarrow{L} \mathrm{Spin}(S^3) \xleftarrow{R} S^3 \times \mathrm{Spin}(3),$$

and the composition $R^{-1} \circ L$ is the Adjoint representation lifted to Spin:

$$S^3 \cong \mathrm{SU}(2) \xrightarrow{\mathrm{Ad}} \mathrm{Spin}(\mathfrak{su}(2)) \cong \mathrm{Spin}(3).$$

The difference $\mathscr{L} - \mathscr{R}$ is equal to $\pi^*(\mathrm{Ad}^* \, \Omega \tfrac{1}{2} p_1)$. The Adjoint representation here is an isomorphism of Lie groups and hence an isomorphism on cohomology. As mentioned in Remark 2.17, there is a factor of 2 and minus sign at work: The class $\Omega \tfrac{1}{2} p_1$ is twice a generator of $H^3(S^3; \mathbb{Z})$, and stably $p_1 = -c_2$. Hence $\Omega \tfrac{1}{2} p_1$ is mapped to $-2\Omega c_2$, or $-2 \in \mathbb{Z} \cong H^3(S^3; \mathbb{Z})$, and we use the shorthand $\mathscr{L} + 2 = \mathscr{R}$.

Similarly, we examine the difference between the left-framing and the bounding string structure, and in doing so reference Remark 2.17. The string structure induced from D^4 is a framing of the stable tangent bundle. The normal bundle $\nu \to S^3$ is trivial, and we have the standard isomorphisms of bundles over S^3:

$$\mathrm{Spin}(TS^3 \oplus \mathbb{R}) \cong \mathrm{Spin}(TS^3 \oplus \nu) \cong \mathrm{Spin}(D^4) \cong \mathrm{Spin}(4).$$

The difference in framing of the two stable bundles differs by the left-multiplication map $S^3 \to \mathrm{Spin}(4)$ given by considering S^3 as the unit quaternions. Under the standard isomorphisms $S^3 \cong \mathrm{SU}(2)$ and $\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2)$, this left-multiplication map is the inclusion into the first factor:

$$\mathrm{SU}(2) \xrightarrow{\mathrm{Id} \times \{1\}} \mathrm{SU}(2) \times \mathrm{SU}(2) \cong \mathrm{Spin}(4).$$

The induced map on cohomology sends $\Omega \tfrac{1}{2} p_1$ to $-\Omega c_2$, or $-1 \in \mathbb{Z} \cong H^3(S^3; \mathbb{Z})$. Therefore,

$$\mathscr{L} + 1 = \partial D^4 \quad \text{and} \quad \mathscr{L} + 2 = \partial D^4 + 1 = \mathscr{R}.$$

The Adams e-invariant gives an isomorphism $\pi_3^s \xrightarrow{\cong} \mathbb{Z}/24$ and sends the left and right framings to the two generators [Atiyah and Smith 1974]. Our calculations also verify this explicitly. On a framed $(4k-1)$-dimensional manifold M, the e-invariant can be computed as follows. Choose a spin manifold W such that $\partial W = M$ as spin manifolds; such a manifold exists because $M\mathrm{Spin}^{4k-1}(pt) = 0$. Using the framing of TM, define the Pontryagin classes $p_i(W, M)$ as relative classes in $H^*(W, M)$. We then obtain $\hat{A}(W, M)$ by evaluating $\hat{A}(TW, TM)$ on the fundamental class of W, where $\hat{A}(TW, TM)$ is the $\hat{A}$-polynomial with relative Pontryagin classes. Then,

$$e[M] = \begin{cases} \hat{A}(W, M) \mod \mathbb{Z} & \text{for } k \text{ even,} \\ \frac{1}{2}\hat{A}(W, M) \mod \mathbb{Z} & \text{for } k \text{ odd.} \end{cases}$$

The e-invariant is well-defined as an element of $\mathbb{Q}/\mathbb{Z}$, since choosing a different W' will give $\hat{A}(W', M) - \hat{A}(W, M) = \hat{A}(W' \cup_M (-W))$, which is an integer (or even integer) by the Atiyah–Singer index theorem.

If we include metrics so that $(W, \tilde{g})$ is a Riemannian spin manifold with boundary (M, g), then we naturally have the Pontryagin forms $p_i(\tilde{g}) \in \Omega^{4k}(W)$.

Proposition 6.1. *If $(M, g, \mathcal{S})$ is a Riemannian spin 3-manifold with string class, then*

$$e(M, \mathcal{S}) = -\frac{1}{48}\int_W p_1(\tilde{g}) + \frac{1}{24}\int_M H_{\mathcal{S},g} \mod \mathbb{Z}.$$

Proof.

$$e[M, \mathcal{S}] = \frac{1}{2}\int_W \hat{A}(W, M) = \frac{1}{2}\int_W (1 - \tfrac{1}{24}p_1(W, M) + \cdots) = -\frac{1}{48}\int_W p_1(W, M).$$

We now construct a de Rham representative of $p_1(W, M)$. If $\partial W = M$, then consider the bordism $W \cup_M ([0, 1] \times M)$ obtained by gluing ∂W to $\{0\} \times M$. The string class $\mathcal{S}$ gives a stable trivialization p of $\mathrm{Spin}(TM)$ up to homotopy, and we let Θ_p denote the induced flat connection. Denoting the Levi-Civita connection on $\mathrm{Spin}(TM)$ by Θ_g, we have the connection $\Theta(t)$ on $[0, 1] \times M$, where

$$\Theta(t) = t\Theta_p + (1 - t)\Theta_g.$$

Finally, define $\tilde{\Theta}$ to be the connection on $\mathrm{Spin}(W \cup_M ([0, 1] \times M))$ induced by $\Theta_{\tilde{g}}$ and $\Theta(t)$. The form $p_1(\tilde{\Theta})$ is a de Rham representative of $p_1(W, M)$, and

$$\begin{aligned} \int_{\widetilde{W}} p_1(\tilde{\Theta}) &= \int_W p_1(\tilde{g}) + \int_{M^3}\int_{[0,1]} p_1(\Theta(t)) \\ &= \int_W p_1(\tilde{g}) + \int_M \mathrm{CS}_{p_1}(\Theta_p, \Theta_g) = \int_W p_1(\tilde{g}) - 2\int_M \mathrm{CS}_{\frac{1}{2}p_1}(\Theta_g, \Theta_p), \end{aligned}$$

where $\mathrm{CS}_\lambda(\Theta_g, \Theta_p)$ is the general Chern–Simons transgression between two connections. Lemmas 6.3 and 3.18 together imply

$$\int_M \mathrm{CS}_{\frac{1}{2}p_1}(\Theta_g, \Theta_p) = \int_M p^* CS(\Theta_g) = \int_M H_{\mathcal{S},g}.$$

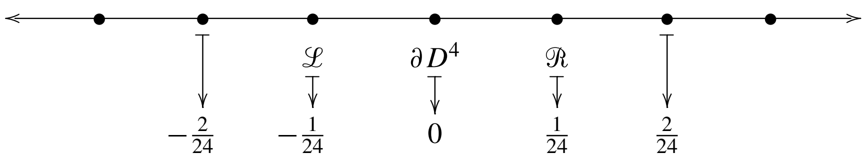
Therefore,

$$-\frac{1}{48}\int_{\widetilde{W}} p_1(\tilde{\Theta}) = -\frac{1}{48}\int_W p_1(\tilde{g}) + \frac{1}{24}\int_M \mathrm{CS}_{\frac{1}{2}p_1}(\Theta_g, \Theta_p)$$
$$= -\frac{1}{48}\int_W p_1(\tilde{g}) + \frac{1}{24}\int_M H_{\mathcal{S},g}. \qquad \square$$

Corollary 6.2. *When $M = S^3$ and g is the standard round metric,*

$$e(S^3, \mathcal{S}) = \frac{1}{24}\int_{S^3} H_{\mathcal{S},g} \mod \mathbb{Z}.$$

In the next subsection, we calculate $H_{\mathcal{S},g}$ for all left-invariant metrics on S^3. Equation (6.8) and the corollary above imply that $e[S^3, \mathcal{L}] = -\frac{1}{24}$, $e[S^3, \partial D^4] = 0$, and $e[S^3, \mathcal{R}] = \frac{1}{24}$. Below is a pictorial description of the space of string classes on S^3 and their corresponding string bordism class under $e : M\mathrm{String}^{-3} \xrightarrow{\cong} \mathbb{Z}/24$.



Lemma 6.3. *If $p : M \to P$ is a global section and Θ_p the induced flat connection, then*

$$\mathrm{CS}_\lambda(\Theta, \Theta_p) = p^* \mathrm{CS}_\lambda(\Theta) \in \Omega^3(M).$$

Proof. This lemma is essentially a tautology. Using the notation of [Freed 2002], in general $\mathrm{CS}_\lambda(\Theta_1, \Theta_0) := \int_{[0,1]} \lambda(\Theta_t) \in \Omega^{2i-1}(M)$, where $\Theta_t := t\Theta_1 + (1-t)\Theta_0$ is a connection on $[0, 1] \times P \to [0, 1] \times M$. Then,

$$\mathrm{CS}_\lambda(\Theta) := \mathrm{CS}_\lambda(\pi^*\Theta, \Theta_{\mathrm{taut}}) \in \Omega^{2i-1}(P),$$

where Θ_{taut} is the trivial connection induced by the canonical section of $\pi^* P$. Since one can compute these transgression forms via local frames, and by definition $p^*\Theta_p = 0$, we easily see

$$\mathrm{CS}_\lambda(\Theta, \Theta_p) = \int_{[0,1]} \lambda\big(p^*(t\Theta + (1-t)\Theta_p)\big) = \int_{[0,1]} \lambda\big(p^* t\Theta\big)$$
$$= p^* \int_{[0,1]} \lambda(t\Theta) = p^* \mathrm{CS}_\lambda(\Theta). \qquad \square$$

6b. ***Calculation of canonical 3-forms.*** We now investigate Question 4.3 by considering left-invariant metrics on $S^3 \cong \mathrm{SU}(2)$; that is, metrics g on $\mathrm{SU}(2)$ such that left multiplication is an isometry. As noted in Proposition 6.11, the calculations for right-invariant metrics only differ from those for left-invariant metrics by a sign. Any such left-invariant metric is determined by its behavior on the tangent space at the identity, so we are considering metrics on the Lie algebra $\mathfrak{su}(2)$ of left-invariant vector fields. A global rescaling of g leaves the Ricci tensor and canonical form $H_{\mathcal{L},g}$ invariant; hence it does not affect the outcome of Question 4.3. The space of left-invariant metrics, up to change of oriented basis and global rescaling, is the 2-dimensional space $\mathrm{Sym}^2_{>0}(\mathbb{R}^3)/(\mathrm{SO}(\mathbb{R}^3)\times\mathbb{R}_+)$, where $\mathrm{Sym}^2_{>0}(\mathbb{R}^3)$ denotes the 6-dimensional space of positive-definite 3×3-matrices.

We now give a more computationally explicit description of this space. Let $\{e_1, e_2, e_3\}$ be the standard basis for $\mathfrak{su}(2)$ satisfying

$$[e_1, e_2] = 2e_3, [e_2, e_3] = 2e_1, [e_3, e_1] = 2e_2.$$

When $\{e_1, e_2, e_3\}$ is an orthonormal basis, the metric is biinvariant and equal to the standard round metric on $S^3 \subset D^4$. For any $\alpha_1, \alpha_2 \in \mathbb{R}_{>0}$, define the left-invariant metric g_{α_1,α_2} by declaring $\{\alpha_1 e_1, \alpha_2 e_2, e_3\}$ to be an orthonormal basis. In the case where $\alpha_2 = 1$, we recover the 1-parameter family of Berger metrics on S^3. Based on knowledge from [Milnor 1976], it suffices to consider the 2-parameter family of metrics $\{g_{\alpha_1,\alpha_2}\}$.

Lemma 6.4. *If g is a left-invariant metric on* SU(2), *then there exists $\alpha_1, \alpha_2 \in \mathbb{R}_{>0}$ such that g_{α_1,α_2} is isometric to a constant multiple of g.*

Proof. Lemma 4.1 [Milnor 1976] implies that there exists an orthonormal basis $\{E_1, E_2, E_3\}$ for g such that

$$[E_1, E_2] = \lambda_3 E_3, \quad [E_2, E_3] = \lambda_1 E_1, \quad [E_3, E_1] = \lambda_2 E_2,$$

where $\lambda_i \in \mathbb{R}_{>0}$. (Milnor's e_i correspond to our E_i.) For any $(\lambda_1, \lambda_2, \lambda_3)$, it is clear that the orthonormal basis

$$\left\{\tfrac{1}{2}\sqrt{\lambda_2\lambda_3}e_1, \tfrac{1}{2}\sqrt{\lambda_3\lambda_1}e_2, \tfrac{1}{2}\sqrt{\lambda_1\lambda_2}e_3\right\}$$

defines a left-invariant metric isometric to the original g. Finally, we normalize so that the coefficient of e_3 is 1. Hence, there is a surjective map

$$\mathbb{R}^2_{>0} \to \{\text{Left-invariant metrics}\}/\{\text{Isom} \times \text{Scale}\}, \quad \alpha_1, \alpha_2 \mapsto g_{\alpha_1,\alpha_2}. \qquad \square$$

We first calculate the Ricci curvature for g_{α_1,α_2}. A straightforward computation gives the covariant derivative of the Levi-Civita connection in our invariant frame. The nonzero components are

$$\begin{aligned}
\langle \nabla_{\alpha_1 e_1} \alpha_2 e_2, e_3\rangle &= \alpha_1\alpha_2 + \alpha_1/\alpha_2 - \alpha_2/\alpha_1,\\
\langle \nabla_{\alpha_2 e_2} \alpha_3 e_3, \alpha_1 e_1\rangle &= \alpha_1\alpha_2 - \alpha_1/\alpha_2 + \alpha_2/\alpha_1,\\
\langle \nabla_{\alpha_3 e_3} \alpha_1 e_1, \alpha_1 e_2\rangle &= -\alpha_1\alpha_2 + \alpha_1/\alpha_2 + \alpha_2/\alpha_1.
\end{aligned}$$

The Ricci tensor is then diagonalized with eigenvalues

$$\begin{aligned}\operatorname{Ric}(\alpha_1 e_1) &= 2(\ \ \alpha_1\alpha_2 - \alpha_1/\alpha_2 + \alpha_2/\alpha_1)(-\alpha_1\alpha_2 + \alpha_1/\alpha_2 + \alpha_2/\alpha_1),\\ \operatorname{Ric}(\alpha_1 e_2) &= 2(-\alpha_1\alpha_2 + \alpha_1/\alpha_2 + \alpha_2/\alpha_1)(\ \ \alpha_1\alpha_2 + \alpha_1/\alpha_2 - \alpha_2/\alpha_1),\\ \operatorname{Ric}(e_3) &= 2(\ \ \alpha_1\alpha_2 + \alpha_1/\alpha_2 - \alpha_2/\alpha_1)(\ \ \alpha_1\alpha_2 - \alpha_1/\alpha_2 + \alpha_2/\alpha_1).\end{aligned}$$

Solving inequalities tells us that the Ricci curvature is strictly positive if and only if (α_1, α_2) is in the interior of the region bounded by the three curves

$$\alpha_2 = \sqrt{\frac{\alpha_1^2}{1+\alpha_1^2}}, \quad \alpha_2 = \sqrt{\frac{\alpha_1^2}{-1+\alpha_1^2}}, \quad \alpha_2 = \sqrt{\frac{-\alpha_1^2}{-1+\alpha_1^2}}. \tag{6.5}$$

This region is shown in Figure 1, left. The Ricci curvature is nonnegative with one zero eigenvalue on the three boundary curves.

Now we calculate the canonical 3-form $H_{\mathscr{L},g_{\alpha_1,\alpha_2}} \in \Omega^3(S^3)$. For dimensional reasons, $H_{\mathscr{L},g_{\alpha_1,\alpha_2}}$ is harmonic and therefore

$$H_{\mathscr{L},g_{\alpha_1,\alpha_2}} \in \mathscr{H}^3(S^3) \cong H^3(S^3;\mathbb{R}) \cong \mathbb{R}$$

with its value in $\mathbb{R}$ determined by integrating over S^3. Lemma 3.18 says that we can calculate $H_{\mathscr{L},g_{\alpha_1,\alpha_2}}$ by simply calculating the Chern–Simons 3-form $\mathrm{CS}_{\frac{1}{2}p_1}(g_{\alpha_1,\alpha_2})$ on the global frame $\{\alpha_1 e_1, \alpha_2 e_2, e_3\}$. This is a straightforward, though lengthy, calculation.

For the class $\frac{1}{2}p_1$, the Chern–Simons form is

$$\mathrm{CS}_{\frac{1}{2}p_1}(\Theta) = -\frac{1}{16\pi^2}\operatorname{Tr}(\Omega\wedge\Theta - \tfrac{1}{6}\Theta\wedge[\Theta\wedge\Theta]),$$

with Tr being the ordinary matrix trace. The normalization constant can be seen from $\frac{1}{2}p_1(\Theta) = -\frac{1}{2}c_2(\Theta) = -\frac{1}{2}\frac{1}{8\pi^2}\operatorname{Tr}(\Omega\wedge\Omega)$. The frame $\{e_i\}$ gives rise to the dual frame $\{e^i\}$ on $\mathfrak{su}(2)^*$. In our global frame, the Chern–Simons form is a constant multiple of $e^1\wedge e^2\wedge e^3$, the standard volume form for $\mathrm{SU}(2)\cong S^3 \subset D^4$. Using a direct calculation along with $\int_{S^3} e^1\wedge e^2\wedge e^3 = 2\pi^2$, we obtain

$$\begin{aligned}\int_{S^3} H_{\mathscr{L},g_{\alpha_1,\alpha_2}} &= -\frac{1}{16\pi^2}\int_{S^3}\operatorname{Tr}(\Theta\wedge\Omega - \tfrac{1}{6}\Theta\wedge[\Theta\wedge\Theta])\\ &= -\frac{\alpha_1^6\alpha_2^6 - \alpha_1^6\alpha_2^4 - \alpha_1^4\alpha_2^6 - \alpha_1^6\alpha_2^2 - \alpha_1^2\alpha_2^6 - \alpha_1^4\alpha_2^2 - \alpha_1^2\alpha_2^4 + 4\alpha_1^4\alpha_2^4 + \alpha_1^6 + \alpha_2^6}{\alpha_1^4\alpha_2^4}.\end{aligned} \tag{6.6}$$

See Figure 1, right, for a graph of this function. If we set $\alpha_2 = 1$ and only consider the usual Berger metrics, we obtain

$$\int_{S^3} H_{\mathscr{L},g_{\alpha_1,1}} = -2 + \frac{2\alpha_1^2 - 1}{\alpha_1^4}. \tag{6.7}$$

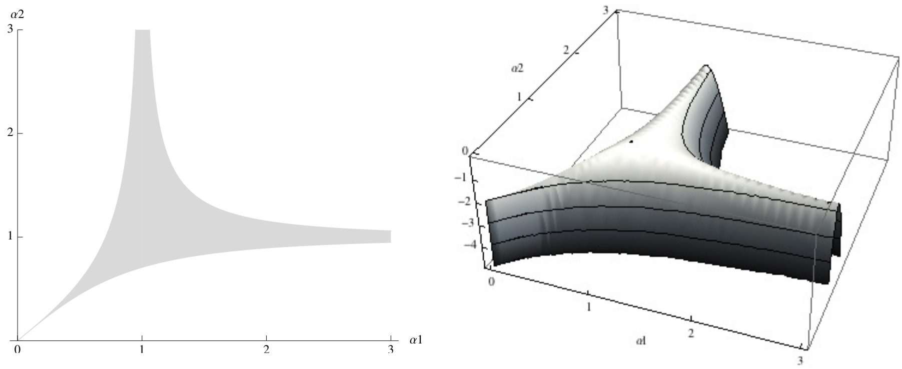


Figure 1. At left: Region with positive Ricci curvature. At right: Values of $\int_{S^3} H_{\mathscr{L},g_{\alpha_1,\alpha_2}}$.

These values are graphed in Figure 2, left. Note that when reduced mod $\mathbb{Z}$, (6.7) coincides with the calculation performed in the original [Chern and Simons 1974]. If we set $\alpha_1 = \alpha_2 = 1$, we obtain the standard biinvariant metric and see that

$$\int_{S^3} H_{\mathscr{L},g_{1,1}} = -1, \quad \int_{S^3} H_{\partial D^4,g_{1,1}} = 0, \quad \int_{S^3} H_{\mathscr{R},g_{1,1}} = 1. \tag{6.8}$$

We now analyze (6.6) on the region Ric ≥ 0. The only critical point occurs at $\alpha_1 = \alpha_2 = 1$, where $\int_{S^3} H_{\mathscr{L},g_{\alpha_1,\alpha_2}} = -1$ is a maximal value. Furthermore, $\int_{S^3} H_{\mathscr{L},g_{\alpha_1,\alpha_2}} = -2$ identically on the three curves bounding the region of positive Ricci curvature. So, we have the range of values

$$\{\textstyle\int_{S^3} H_{\mathscr{L},g} \mid \mathrm{Ric}(g) > 0,\ g \text{ left-invariant}\} = (-2, -1].$$

Figure 1, right, demonstrates this with the help of Mathematica; the level curves for -2 are precisely the three functions from (6.5).

Due to the equivariance of the canonical 3-form under change of string class (see Proposition 3.12), our calculation using $\mathscr{L}$ gives us $H_{\mathscr{S},g_{\alpha_1,\alpha_2}}$ for any other string class $\mathscr{S}$ by

$$\int_{S^3} H_{\mathscr{L}+j,g_{\alpha_1,\alpha_2}} = j + \int_{S^3} H_{\mathscr{L},g_{\alpha_1,\alpha_2}} \quad \text{for any } j \in \mathbb{Z} \cong H^3(S^3;\mathbb{Z}).$$

Therefore,

$$\{\textstyle\int_{S^3} H_{\mathscr{L}+j,g} \mid \mathrm{Ric}(g) > 0,\ g \text{ left-invariant}\} = (-2+j, -1+j]. \tag{6.9}$$

To graphically demonstrate this, Figure 2, left, shows the canonical 3-forms for various string classes on the 1-parameter family of left-invariant Berger metrics.

The entire previous discussion was based on left-invariant Riemannian metrics. What if we had decided to use right-invariant metrics? Given an inner product g_e

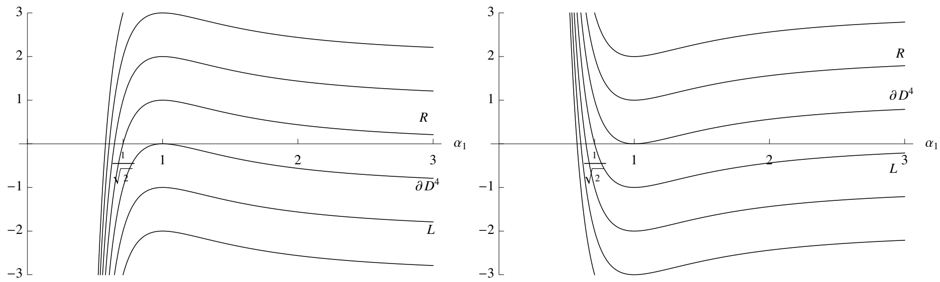


Figure 2. $\int_{S^3} H_{\mathcal{S},g_{\alpha_1,\alpha_2}}$ on Berger metrics, for left- and right-invariant metrics, respectively.

on $T_e\,\mathrm{SU}(2)$, we can form a left-invariant metric g^L and a right-invariant metric g^R by left or right multiplying g_e. The canonical 3-forms are related by the following easy lemma, whose proof is at the end of this section.

Lemma 6.10. $H_{\mathcal{L},g^L} = -H_{\mathcal{R},g^R}$.

This fact is graphically demonstrated in Figure 2, right. In the case of the Berger metrics, note that the Ricci curvature is positive for all $\alpha_1 > 1/\sqrt{2}$, and the Ricci curvature is nonnegative with a 0 eigenvalue at $\alpha_1 = 1/\sqrt{2}$.

Proposition 6.11. *Suppose the string class and (left or right)-invariant Riemannian metric $(\mathcal{S}, g)$ on S^3 satisfy*

$$\mathrm{Ric}(g) > 0 \quad \textit{and} \quad H_{\mathcal{S},g} = 0.$$

Then $\mathcal{S} = \partial D^4$ and g is the biinvariant round metric. Consequently,

$$\sigma[S^3, \mathcal{S}] = 0 \in \mathrm{tmf}^{-3}(pt) \cong \mathbb{Z}/24.$$

Proof. If g is a left-invariant metric with positive Ricci curvature and $H_{\mathcal{S},g} = 0$, then (6.9) implies that $\mathcal{S} = \mathcal{L} + 1 = \partial D^4$ with g the biinvariant metric $g_{1,1}$.

If g is a right-invariant metric, Lemma 6.10 and (6.9) imply that

$$\{\int_{S^3} H_{\mathcal{R}+j,g} \mid \mathrm{Ric}(g) > 0,\ g \text{ right-invariant}\} = [1 + j, 2 + j). \tag{6.12}$$

If $H_{\mathcal{S},g} = 0$, then $\mathcal{S} = \mathcal{R} - 1 = \partial D^4$ and $g = g_{1,1}$. Finally, $[S^3, \partial D^4] = 0 \in M\mathrm{String}^{-3}$, so $\sigma[S^3, \partial D^4] = 0 \in \mathrm{tmf}^{-3}$. □

We conclude that in this case, Question 4.3 has a very nontrivial affirmative answer. In particular, there are 1-dimensional families of left- and right-invariant metrics that are Ricci nonnegative and satisfy $H_{\mathcal{R},g} = 0$ and $H_{\mathcal{L},g} = 0$, respectively. Furthermore, as evidenced by Figure 2, one can find Ricci positive metrics with

$H_{\mathscr{L},g}$ arbitrarily small but nonzero. Finally, we point out that for any string class $\mathscr{S}$, the lift of the Chern–Simons invariant

$$\mathrm{Met}(S^3) \xrightarrow{\int H_{\mathscr{S},g}} \mathbb{R}$$

is surjective. The 1-parameter families of left- and right-invariant Berger metrics in Figure 2 show this.

Proof of Lemma 6.10. In a left- or right-invariant frame, the connection is computed purely in terms of the Lie bracket on vector fields. On a Lie group G, one can define two Lie algebra structures $[\,\cdot\,,\cdot\,]_L$ and $[\,\cdot\,,\cdot\,]_R$ corresponding to the usual Lie bracket on left- or right-invariant vector fields. For $X, Y \in T_eG$, these are related by

$$[X, Y]_L = -[X, Y]_R.$$

If Θ_L, Θ_R denote the connections in the two frames, we have $\Theta_L = -\Theta_R$ and $\Omega_L = \Omega_R$, so

$$\mathrm{Tr}(\Theta_L \wedge \Omega_L - \tfrac{1}{6}\Theta_L \wedge [\Theta_L \wedge \Theta_L]) = -\,\mathrm{Tr}(\Theta_R \wedge \Omega_R - \tfrac{1}{6}\Theta_R \wedge [\Theta_R \wedge \Theta_R]). \quad \square$$

Acknowledgments

Much of this paper is based on work done in my PhD thesis [Redden 2006], and the paper itself was prepared during my time at both Stony Brook University and Michigan State University. I would like to especially thank Stephan Stolz for encouragement, insight, and numerous suggestions. I also thank Peter Teichner, Blaine Lawson, Liviu Nicolaescu, Hisham Sati, and Florin Dumitrescu for useful discussions, questions, and comments.

References

[Alvarez and Singer 2002] O. Alvarez and I. M. Singer, "Beyond the elliptic genus", *Nuclear Phys. B* **633**:3 (2002), 309–344. MR 2003h:58047 Zbl 0995.58016

[Atiyah and Smith 1974] M. F. Atiyah and L. Smith, "Compact Lie groups and the stable homotopy of spheres", *Topology* **13** (1974), 135–142. MR 49 #8013 Zbl 0282.55008

[Baas et al. 2004] N. A. Baas, B. I. Dundas, and J. Rognes, "Two-vector bundles and forms of elliptic cohomology", pp. 18–45 in *Topology, geometry and quantum field theory*, edited by U. Tillmann, London Math. Soc. Lecture Note Ser. **308**, Cambridge Univ. Press, 2004. MR 2005e:55007 Zbl 1106.55004

[Baez et al. 2007] J. C. Baez, D. Stevenson, A. S. Crans, and U. Schreiber, "From loop groups to 2-groups", *Homology, Homotopy Appl.* **9**:2 (2007), 101–135. MR 2009c:22022 Zbl 1122.22003

[Cheeger and Simons 1985] J. Cheeger and J. Simons, "Differential characters and geometric invariants", pp. 50–80 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 87g:53059 Zbl 0621.57010

[Chern and Simons 1974] S. S. Chern and J. Simons, "Characteristic forms and geometric invariants", *Ann. of Math.* (2) **99** (1974), 48–69. MR 50 #5811 Zbl 0283.53036

[Coquereaux and Pilch 1989] R. Coquereaux and K. Pilch, "String structures on loop bundles", *Comm. Math. Phys.* **120**:3 (1989), 353–378. MR 90k:58012 Zbl 0672.55008

[Dai 1991] X. Dai, "Adiabatic limits, nonmultiplicativity of signature, and Leray spectral sequence", *J. Amer. Math. Soc.* **4**:2 (1991), 265–321. MR 92f:58169 Zbl 0736.58039

[Forman 1995] R. Forman, "Spectral sequences and adiabatic limits", *Comm. Math. Phys.* **168**:1 (1995), 57–116. MR 96g:58176 Zbl 0827.58001

[Freed 1995] D. S. Freed, "Classical Chern–Simons theory, I", *Adv. Math.* **113**:2 (1995), 237–303. MR 96h:58019 Zbl 0844.58039

[Freed 2002] D. S. Freed, "Classical Chern–Simons theory, II", *Houston J. Math.* **28**:2 (2002), 293–310. MR 2003f:55022 Zbl 1030.53033

[Henriques 2008] A. Henriques, "Integrating L_∞-algebras", *Compos. Math.* **144**:4 (2008), 1017–1045. MR 2441255 Zbl 1152.17010

[Hitchin 1974] N. Hitchin, "Harmonic spinors", *Advances in Math.* **14** (1974), 1–55. MR 50 #11332 Zbl 0284.58016

[Hopkins 2002] M. J. Hopkins, "Algebraic topology and modular forms", pp. 291–317 in *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), vol. 1, edited by T. Li, Higher Ed. Press, Beijing, 2002. MR 2004g:11032 Zbl 1031.55007

[Hu and Kriz 2004] P. Hu and I. Kriz, "Conformal field theory and elliptic cohomology", *Adv. Math.* **189**:2 (2004), 325–412. MR 2005m:55008 Zbl 1071.55004

[Lawson and Michelsohn 1989] H. B. Lawson, Jr. and M.-L. Michelsohn, *Spin geometry*, Princeton Mathematical Series **38**, Princeton University Press, 1989. MR 91g:53001 Zbl 0688.57001

[Lichnerowicz 1962] A. Lichnerowicz, "Laplacien sur une variété riemannienne et spineurs", *Atti Accadem. Nazionale Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (8) **33** (1962), 187–191. MR 27 #5203 Zbl 0118.37601

[Liu 1996] K. Liu, "Modular forms and topology", pp. 237–262 in *Moonshine, the Monster, and related topics* (South Hadley, MA, 1994), edited by C. Dong and G. Mason, Contemp. Math. **193**, Amer. Math. Soc., Providence, RI, 1996. MR 97f:58124 Zbl 0974.58020

[Mazzeo and Melrose 1990] R. R. Mazzeo and R. B. Melrose, "The adiabatic limit, Hodge cohomology and Leray's spectral sequence for a fibration", *J. Differential Geom.* **31**:1 (1990), 185–213. MR 90m:58004 Zbl 0702.58007

[Milnor 1976] J. Milnor, "Curvatures of left invariant metrics on Lie groups", *Advances in Math.* **21**:3 (1976), 293–329. MR 54 #12970 Zbl 0341.53030

[Redden 2006] C. Redden, *Canonical metric connections associated to string structures*, thesis, University of Notre Dame, 2006, available at http://tinyurl.com/6czpzwh. MR 2709440

[Redden 2008] C. Redden, "Harmonic forms on principal bundles", preprint, 2008. arXiv 0810.4578

[Schommer-Pries 2009] C. Schommer-Pries, "A finite-dimensional string 2-group", preprint, 2009. arXiv 0911.2483

[Segal 1988] G. Segal, "Elliptic cohomology (after Landweber-Stong, Ochanine, Witten, and others)", pp. 187–201 in *Séminaire Bourbaki, Vol. 1987/88, Exp. No. 695–699*, Astérisque **161-162**, Société Mathématique de France, Paris, 1988. MR 91b:55005 Zbl 0686.55003

[Serre 1951] J.-P. Serre, "Homologie singulière des espaces fibrés. Applications", *Ann. of Math.* (2) **54** (1951), 425–505. MR 13,574g Zbl 0045.26003

[Simons and Sullivan 2008] J. Simons and D. Sullivan, "Axiomatic characterization of ordinary differential cohomology", *J. Topol.* **1**:1 (2008), 45–56. MR 2009e:58035 Zbl 1163.57020

[Stolz 1992] S. Stolz, "Simply connected manifolds of positive scalar curvature", *Ann. of Math.* (2) **136**:3 (1992), 511–540. MR 93i:57033 Zbl 0784.53029

[Stolz 1996] S. Stolz, "A conjecture concerning positive Ricci curvature and the Witten genus", *Math. Ann.* **304**:4 (1996), 785–800. MR 96k:58209 Zbl 0856.53033

[Stolz and Teichner 2004] S. Stolz and P. Teichner, "What is an elliptic object?", pp. 247–343 in *Topology, geometry and quantum field theory*, edited by U. Tillmann, London Math. Soc. Lecture Note Ser. **308**, Cambridge Univ. Press, 2004. MR 2005m:58048 Zbl 1107.55004

[Waldorf 2009] K. Waldorf, "String connections and Chern-Simons theory", preprint, 2009. arXiv 0906.0117

[Witten 1988] E. Witten, "The index of the Dirac operator in loop space", pp. 161–181 in *Elliptic curves and modular forms in algebraic topology* (Princeton, NJ, 1986), edited by P. S. Landweber, Lecture Notes in Math. **1326**, Springer, Berlin, 1988. MR 970288 Zbl 0679.58045

[Witten 1999] E. Witten, "Index of Dirac operators", pp. 475–511 in *Quantum fields and strings: A course for mathematicians* (Princeton, NJ, 1996/1997), vol. 1, edited by P. Deligne et al., Amer. Math. Soc., Providence, RI, 1999. MR 2000i:58037 Zbl 1170.58307

Received December 11, 2009.

CORBETT REDDEN
DEPARTMENT OF MATHEMATICS
MICHIGAN STATE UNIVERSITY
A-318 WELLS HALL
EAST LANSING, MI 48824
UNITED STATES

redden@math.msu.edu
http://www.math.msu.edu/~redden

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

DUAL PAIRS AND CONTRAGREDIENTS OF IRREDUCIBLE REPRESENTATIONS

BINYONG SUN

Let G be one of the classical groups $\mathrm{GL}(n)$, $\mathrm{U}(n)$, $\mathrm{O}(n)$ or $\mathrm{Sp}(2n)$, over a nonarchimedean local field of characteristic zero. It is well known that the contragredient of an irreducible admissible smooth representation of G is isomorphic to a twist of it by an automorphism of G. We prove that similar results hold for double covers of G that occur in the study of local theta correspondences.

1. Introduction and the results

Fix a nonarchimedean local field $\Bbbk$ of characteristic zero. We introduce the notation in order to treat the four classes of classical groups $\mathrm{GL}(n)$, $\mathrm{U}(n)$, $\mathrm{O}(n)$ and $\mathrm{Sp}(2n)$ simultaneously. Let A be a $\Bbbk$-algebra and τ be a $\Bbbk$-algebra involution of A such that

$$(A,\tau)=\begin{cases}(\Bbbk\times\Bbbk,\ \text{the nontrivial automorphism}),\\ (\text{a quadratic field extension of }\Bbbk,\ \text{the nontrivial automorphism}), \ \text{or}\\ (\Bbbk,\ \text{the trivial automorphism}).\end{cases}$$

Let $\epsilon=\pm 1$ and let E be an ϵ-Hermitian A-module; namely, E is a free A-module of finite rank equipped with a nondegenerate $\Bbbk$-bilinear map

$$\langle\,\cdot\,,\,\cdot\,\rangle_E : E\times E\to A$$

satisfying $\langle u,v\rangle_E=\epsilon\langle v,u\rangle_E^{\tau}$ and $\langle au,v\rangle_E=a\langle u,v\rangle_E$ for $a\in A$ and $u,v\in E$. Denote by $\mathrm{U}(E)$ the group of all A-module automorphisms of E that preserve the form $\langle\,\cdot\,,\,\cdot\,\rangle_E$. Depending on the choice of A and ϵ, it is either a general linear group, a unitary group, an orthogonal group or a symplectic group.

Following Mœglin, Vigneras and Waldspurger [1987, Proposition 4.I.2], we extend $\mathrm{U}(E)$ to a larger group $\breve{\mathrm{U}}(E)$ consisting of pairs $(g,\delta)\in\mathrm{GL}_{\Bbbk}(E)\times\{\pm 1\}$ such that either

$$\delta=1 \quad\text{and}\quad g\in\mathrm{U}(E),$$

Supported by NSFC grants number 10801126 and 10931006.
MSC2000: 22E35, 22E46.
Keywords: contragredient representation, dual pair, irreducible representation.

or

$$\delta = -1,$$
$$g(au) = a^\tau g(u) \quad \text{for } a \in A,\ u \in E, \qquad \text{and}$$
$$\langle gu, gv\rangle_E = \langle v, u\rangle_E \quad \text{for } u, v \in E.$$

Clearly $\breve{\mathrm{U}}(E)$ contains $\mathrm{U}(E)$ as a subgroup of index two.

In general, if π is a representation of a group H and g is an element of a group that acts on H as automorphisms, we define the twist π^g to be the representation of H that has the same underlying space as that of π, and whose action is given by $\pi^g(h) := \pi(gh)$ for $h \in H$. If $\breve{H}$ is a group containing H as a subgroup of index two, we always let it act on H by conjugation:

$$\mathrm{Ad} : \breve{H} \times H \to H, \quad (\breve{g}, x) \mapsto \mathrm{Ad}_{\breve{g}}(x) := \breve{g}x\breve{g}^{-1}.$$

It is a classical result in linear algebra that

(1) $$\breve{g}x\breve{g}^{-1} \text{ is conjugate to } x^{-1} \text{ inside } \mathrm{U}(E)$$

for all $\breve{g} \in \breve{\mathrm{U}}(E) \setminus \mathrm{U}(E)$ and all $x \in \mathrm{U}(E)$. For example, when $\mathrm{U}(E)$ is a general linear group, this amounts to saying that every square matrix is conjugate to its transpose. For orthogonal groups, this says that every element of an orthogonal group is conjugate to its inverse. The following considerations (which lead to Theorem 1.1 below) appear in [Mœglin et al. 1987]. By the localization principle of Bernšteĭn and Zelevinskiĭ [1976, Theorem 6.9 and Theorem 6.15.A], result (1) implies that

(2) $$f(\breve{g}x\breve{g}^{-1}) = f(x^{-1}) \quad \text{(as generalized functions on } \mathrm{U}(E))$$

for all Ad-invariant generalized functions f on $\mathrm{U}(E)$ and all $\breve{g} \in \breve{\mathrm{U}}(E) \setminus \mathrm{U}(E)$. For the usual notion of generalized functions, see [Sun 2009, Section 2]. We get the following well known result by (2) and by considering characters of irreducible admissible smooth representations (which are conjugation invariant generalized functions).

Theorem 1.1 [Mœglin et al. 1987, Theorem 4.II.1]. *Let $\breve{g} \in \breve{\mathrm{U}}(E) \setminus \mathrm{U}(E)$, and let π be an irreducible admissible smooth representation of* $\mathrm{U}(E)$. *Then $\pi^\vee$ is isomorphic to $\pi^{\breve{g}}$.*

Here and as usual, we use "$^\vee$" to indicate the contragredient of an admissible smooth representation of a totally disconnected locally compact group.

If E is a symplectic space, that is, if $\epsilon = -1$ and $A = \Bbbk$, then $\breve{\mathrm{Sp}}(E) := \breve{\mathrm{U}}(E)$ is equal to the subgroup of $\mathrm{GSp}(E)$ with similitudes ± 1. Denote by

(3) $$1 \to \{\pm 1\} \to \widetilde{\mathrm{Sp}}(E) \to \mathrm{Sp}(E) \to 1$$

the metaplectic cover of the symplectic group $\mathrm{Sp}(E)$. It is shown in [Mœglin et al. 1987, page 36] that there is a unique continuous action

$$\widetilde{\mathrm{Ad}} : \breve{\mathrm{Sp}}(E) \times \widetilde{\mathrm{Sp}}(E) \to \widetilde{\mathrm{Sp}}(E) \tag{4}$$

of $\breve{\mathrm{Sp}}(E)$ on $\widetilde{\mathrm{Sp}}(E)$ as group automorphisms that lifts the adjoint action

$$\mathrm{Ad} : \breve{\mathrm{Sp}}(E) \times \mathrm{Sp}(E) \to \mathrm{Sp}(E)$$

and leaves the central element $-1 \in \widetilde{\mathrm{Sp}}(E)$ fixed.

We first extend Theorem 1.1 to the case of metaplectic groups:

Theorem 1.2. *Assume that E is a symplectic space. Let $\breve{g} \in \breve{\mathrm{Sp}}(E) \setminus \mathrm{Sp}(E)$, and let π be a genuine irreducible admissible smooth representation of $\widetilde{\mathrm{Sp}}(E)$. Then $\pi^\vee$ is isomorphic to $\pi^{\breve{g}}$.*

Here and henceforth, "genuine" means that the central element $-1 \in \widetilde{\mathrm{Sp}}(E)$ acts via the scalar multiplication by -1.

Remark. In the case that the character of π is a locally integrable function, Theorem 1.2 is proved in [Mœglin et al. 1987, Theorem 4.II.2].

Harish-Chandra [1999] proved locally integrability of irreducible characters for p-adic linear reductive groups, but he did not treat metaplectic groups.

The proofs of Theorem 1.1 in [Mœglin et al. 1987] and Theorem 1.2 in Section 2 do not depend on locally integrability of irreducible characters.

Now we consider dual pairs. Write $\epsilon' := -\epsilon$, and let $(E', \langle \cdot, \cdot \rangle_{E'})$ be an ϵ'-Hermitian A-module. Then $\boldsymbol{E} := E \otimes_A E'$ is a skew-Hermitian A-module under the form $\langle u \otimes u', v \otimes v' \rangle_{\boldsymbol{E}} := \langle u, v \rangle_E \langle u', v' \rangle_{E'}$. Write $\boldsymbol{E}_{\Bbbk} := \boldsymbol{E}$, viewed as a $\Bbbk$-symplectic space under the form $\langle u, v \rangle_{\boldsymbol{E}_{\Bbbk}} := \mathrm{tr}_{A/\Bbbk}(\langle u, v \rangle_{\boldsymbol{E}})$. Put

$$G := \mathrm{U}(E), \quad \breve{G} := \breve{\mathrm{U}}(E), \quad G' := \mathrm{U}(E'), \quad \breve{G}' := \breve{\mathrm{U}}(E').$$

The group G obviously maps to the symplectic group $\mathrm{Sp}(\boldsymbol{E}_{\Bbbk})$. Define the fiber product $\widetilde{G} := \widetilde{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}) \times_{\mathrm{Sp}(\boldsymbol{E}_{\Bbbk})} G$. This is a double cover of G that depends on both E and E'.

In what follows, we define an action

$$\widetilde{\mathrm{Ad}} : \breve{G} \times \widetilde{G} \to \widetilde{G} \tag{5}$$

that lifts the adjoint action $\mathrm{Ad} : \breve{G} \times G \to G$ and fixes the central element $-1 \in \widetilde{G}$. Let $\breve{g} = (g, \delta) \in \breve{G}$. Choose an arbitrary element $(g', \delta) \in \breve{G}'$. Then

$$\breve{\boldsymbol{g}} := (g \otimes g', \delta) \in \breve{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}),$$

and the automorphism

$$\widetilde{\mathrm{Ad}}_{\breve{\boldsymbol{g}}} \times \mathrm{Ad}_{\breve{g}} : \widetilde{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}) \times G \to \widetilde{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}) \times G \tag{6}$$

leaves the subgroup $\widetilde{G}$ stable. It restricts to an automorphism

$$\widetilde{\mathrm{Ad}}_{\breve{g}} : \widetilde{G} \to \widetilde{G} \tag{7}$$

that is independent of the choice of g'. We obtain (5) by gluing (7) for all $\breve{g} \in \breve{G}$.

The following is a generalization of Theorem 1.2 in the setting of dual pairs.

Theorem 1.3. *Let $\breve{g} \in \breve{G} \setminus G$, and let π be a genuine irreducible admissible smooth representation of $\widetilde{G}$. Then $\pi^\vee$ is isomorphic to $\pi^{\breve{g}}$.*

Remark. When $E' = A = \Bbbk$ and $\epsilon = -1$, Theorem 1.3 specializes to Theorem 1.2. The statement for the general case reduces essentially to those of Theorem 1.1 and Theorem 1.2. Theorem 1.3 is proved in Section 3.

Theorem 1.3 has the following consequence, which is known to experts (up to a proof of Theorem 1.2). As far as the author knows, no proof of it in full generality is found in the literature.

Theorem 1.4. *Denote by ω_ψ the smooth oscillator representation of $\widetilde{\mathrm{Sp}}(\boldsymbol{E}_\Bbbk)$ corresponding to a nontrivial character ψ of $\Bbbk$. Then for all genuine irreducible admissible smooth representation π of $\widetilde{G}$ and π' of $\widetilde{G}'$, we have*

$$\dim \mathrm{Hom}_{G\times G'}(\omega_\psi \otimes \pi \otimes \pi', \mathbb{C}) = \dim \mathrm{Hom}_{G\times G'}(\omega_\psi^\vee \otimes \pi^\vee \otimes \pi'^\vee, \mathbb{C}).$$

Here $\widetilde{G}' := \widetilde{\mathrm{Sp}}(\boldsymbol{E}_\Bbbk) \times_{\mathrm{Sp}(\boldsymbol{E}_\Bbbk)} G'$ is a double cover of G'. Note that both $\omega_\psi \otimes \pi \otimes \pi'$ and $\omega_\psi^\vee \otimes \pi^\vee \otimes \pi'^\vee$, which are originally representations of $\widetilde{G} \times \widetilde{G}'$, descend to representations of $G \times G'$.

Remark. In a follow-up paper [Li et al. 2009], Theorem 1.4 is used to prove multiplicity preservations in theta correspondences (for all residue characteristics), that is, the dimension in Theorem 1.4 is at most one. This is the main reason for providing a detailed proof of Theorem 1.4 here.

In the archimedean case, the analog of Theorem 1.4 is proved by T. Przebinda [1988, Theorem 5.5], while the analog of Theorem 1.3 is a consequence of [1988, Theorem 2.6]. His method is different from ours in that he uses the Langlands classification.

As shown in [Przebinda 1988], Theorem 1.4 together with the Howe duality conjecture implies that theta lifting maps Hermitian representations to Hermitian representations.

2. Theorem 1.2 and its analog

Throughout this section, we assume that $\epsilon = -1$.

2.1. ***Skew Hermitian modules and Jacobi groups.*** As in the last section, E is an ϵ-Hermitian A-module, and $E_{\Bbbk} := E$ is a symplectic space under the form

$$\langle u, v\rangle_{E_{\Bbbk}} := \mathrm{tr}_{A/\Bbbk}(\langle u, v\rangle_E).$$

Denote by $\mathrm{H}(E) := E_{\Bbbk} \times \Bbbk$ the Heisenberg group associated to $E_{\Bbbk}$, whose multiplication is given by $(u, t)(u', t') := (u + u', t + t' + \langle u, u'\rangle_{E_{\Bbbk}})$. The group $\breve{\mathrm{U}}(E)$ acts on $\mathrm{H}(E)$ as group automorphisms by

$$(g, \delta)(u, t) := (gu, \delta t). \tag{8}$$

It defines a semidirect product $\breve{\mathrm{J}}(E) := \breve{\mathrm{U}}(E) \ltimes \mathrm{H}(E)$, which contains $\mathrm{J}(E) := \mathrm{U}(E) \ltimes \mathrm{H}(E)$ as a subgroup of index two.

The results of this note depend heavily on the following.

Lemma 2.1 [Sun 2009, Theorem D]. *Let f be a generalized function on* $\mathrm{J}(E)$. *If f is invariant under conjugations by* $\mathrm{U}(E)$, *that is,*

$$f(gxg^{-1}) = f(x) \quad \textit{for all } g \in \mathrm{U}(E),$$

then

$$f(\breve{g}x\breve{g}^{-1}) = f(x^{-1}) \quad \textit{for all } \breve{g} \in \breve{\mathrm{U}}(E) \setminus \mathrm{U}(E).$$

Actually, we only need the following lemma, which is much weaker.

Lemma 2.2. *Let f be a conjugation-invariant generalized function on* $\mathrm{J}(E)$. *Then*

$$f(\breve{g}x\breve{g}^{-1}) = f(x^{-1}) \quad \textit{for all } \breve{g} \in \breve{\mathrm{J}}(E) \setminus \mathrm{J}(E).$$

A consequence of Lemma 2.2 is this:

Proposition 2.3. *Let $\breve{g} \in \breve{\mathrm{J}}(E) \setminus \mathrm{J}(E)$, and let π be an irreducible admissible smooth representation of* $\mathrm{J}(E)$. *Then $\pi^\vee$ is isomorphic to $\pi^{\breve{g}}$.*

Proof. Denote by f the character of π, which is thus a conjugation-invariant generalized function on $\mathrm{J}(E)$. Therefore

$$f(\breve{g}x\breve{g}^{-1}) = f(x^{-1}) \tag{9}$$

by Lemma 2.2. The left side of (9) is the character of $\pi^{\breve{g}}$, and the right side is the character of $\pi^\vee$. Therefore $\pi^{\breve{g}}$ and $\pi^\vee$ have the same character, and they are thus isomorphic to each other. □

2.2. ***Proof of Theorem 1.2 and its analog.*** We reuse the notation of Section 2.1. Denote by

$$\widetilde{\mathrm{U}}(E) := \widetilde{\mathrm{Sp}}(E_{\Bbbk}) \times_{\mathrm{Sp}(E_{\Bbbk})} \mathrm{U}(E)$$

the double cover of $\mathrm{U}(E)$ induced by the metaplectic cover

$$1 \to \{\pm 1\} \to \widetilde{\mathrm{Sp}}(E_{\Bbbk}) \to \mathrm{Sp}(E_{\Bbbk}) \to 1. \tag{10}$$

As in (5), we have an action

$$\widetilde{\mathrm{Ad}} : \breve{\mathrm{U}}(E) \times \widetilde{\mathrm{U}}(E) \to \widetilde{\mathrm{U}}(E). \tag{11}$$

The following theorem reduces to Theorem 1.2 when $A = \Bbbk$.

Theorem 2.4. *Assume that $\epsilon = -1$. Let $\breve{g} \in \breve{\mathrm{U}}(E) \setminus \mathrm{U}(E)$, and let π be a genuine irreducible admissible smooth representation of $\widetilde{\mathrm{U}}(E)$. Then $\pi^\vee$ is isomorphic to $\pi^{\breve{g}}$.*

Proof. Denote by ω_ψ the smooth oscillator representation of $\widetilde{\mathrm{Sp}}(E_\Bbbk) \ltimes \mathrm{H}(E)$ that corresponds to a nontrivial character ψ of $\Bbbk$. Up to isomorphism, this is the only genuine smooth representation that, as a representation of $\mathrm{H}(E)$, is irreducible and has central character ψ.

Both ω_ψ and π are viewed as smooth representations of $\tilde{\mathrm{J}}(E) := \widetilde{\mathrm{U}}(E) \ltimes \mathrm{H}(E)$, via the restriction and the inflation, respectively. The tensor product $\omega_\psi \otimes \pi$ descends to an irreducible admissible smooth representation of $\mathrm{J}(E)$ [Sun 2009, Lemma 5.3].

The actions of $\breve{\mathrm{U}}(E)$ on $\widetilde{\mathrm{U}}(E)$, $\mathrm{U}(E)$ and $\mathrm{H}(E)$ induce its actions on the semi-direct products $\tilde{\mathrm{J}}(E)$ and $\mathrm{J}(E)$. By Proposition 2.3,

$$(\omega_\psi \otimes \pi)^{\breve{g}} \cong (\omega_\psi \otimes \pi)^\vee$$

as irreducible admissible smooth representations of $\mathrm{J}(E)$, or equivalently

$$\omega_\psi^{\breve{g}} \otimes \pi^{\breve{g}} \cong \omega_\psi^\vee \otimes \pi^\vee.$$

Note that $\omega_\psi^{\breve{g}} \cong \omega_\psi^\vee$ as smooth representations of $\tilde{\mathrm{J}}(E)$. (This is a special case of Lemma 3.3.) Therefore

$$\omega_\psi^\vee \otimes \pi^{\breve{g}} \cong \omega_\psi^\vee \otimes \pi^\vee. \tag{12}$$

As in the proof of [Sun 2009, Lemma 5.3], we have

$$\pi^{\breve{g}} \cong \mathrm{Hom}_{\mathrm{H}(E)}(\omega_\psi^\vee, \omega_\psi^\vee \otimes \pi^{\breve{g}}). \tag{13}$$

Here the right side carries the action of $\widetilde{\mathrm{U}}(E)$ given by $(\tilde{g}\phi)(v) := g(\phi(\tilde{g}^{-1}v))$, where

$$\tilde{g} \in \widetilde{\mathrm{U}}(E), \quad \phi \in \mathrm{Hom}_{\mathrm{H}}(\omega_\psi^\vee, \omega_\psi^\vee \otimes \pi^{\breve{g}}), \quad v \in \omega_\psi^\vee,$$

and g is the image of $\tilde{g}$ under the covering map $\widetilde{\mathrm{U}}(E) \to \mathrm{U}(E)$. Similarly,

$$\pi^\vee \cong \mathrm{Hom}_{\mathrm{H}}(\omega_\psi^\vee, \omega_\psi^\vee \otimes \pi^\vee). \tag{14}$$

We finish the proof by combining (12), (13) and (14). □

3. Proofs of Theorem 1.3 and Theorem 1.4

3.1. ***Proof of Theorem 1.3 for symplectic groups.*** Now we return to the notation of Section 1. First assume that $A = \Bbbk$ and $\epsilon = -1$. Then G is a symplectic group and is thus perfect, that is, G equals its own commutator group. Consequently, there is only one action of $\breve{G}$ on $\widetilde{G}$ that lifts the adjoint action and fixes the central element $-1 \in \widetilde{G}$. There are two cases.

Case 1. The covering map $\widetilde{G} \to G$ splits. Then $\widetilde{G} = G \times \{\pm 1\}$, and Theorem 1.3 is one case of Theorem 1.1.

Case 2. The covering map $\widetilde{G} \to G$ does not split. Then $\widetilde{G} = \widetilde{\mathrm{Sp}}(E)$ [Moore 1968, Theorem 10.4], and Theorem 1.3 is one case of Theorem 1.2.

3.2. ***Proof of Theorem 1.3 when $A \neq \Bbbk$.*** Assume that $A \neq \Bbbk$. Then $\mathrm{U}(\boldsymbol{E})$ is a general linear group or a unitary group.

Lemma 3.1. *There exists a genuine character on* $\widetilde{\mathrm{U}}(\boldsymbol{E})$.

Proof. It is well known that the exact sequence

$$1 \to \mathbb{C}^\times \to (\widetilde{\mathrm{Sp}}(\boldsymbol{E}_\Bbbk) \times \mathbb{C}^\times)/\,\mathrm{diag}(\{\pm 1\}) \to \mathrm{Sp}(\boldsymbol{E}_\Bbbk) \to 1$$

splits continuously over $\mathrm{U}(\boldsymbol{E})$ (this is trivial for general linear groups, and for unitary groups, see [Kudla 1994, Proposition 4.1] or [Harris et al. 1996, Section 1]). Write ι for such a splitting and write $p : \widetilde{\mathrm{U}}(\boldsymbol{E}) \to \mathrm{U}(\boldsymbol{E})$ for the covering map. Then $x \in \widetilde{\mathrm{U}}(\boldsymbol{E}) \mapsto x^{-1}\, \iota(p(x)) \in \mathbb{C}^\times$ is a genuine character. □

Lemma 3.2. *There exists a genuine character χ of $\widetilde{G}$ such that $\chi^{\breve{g}} = \chi^{-1}$ for all $\breve{g} \in \breve{G} \setminus G$.*

Proof. As in Section 1, let $\breve{g} = (g, -1) \in \breve{G} \setminus G$ and $(g', -1) \in \breve{G}' \setminus G'$, and write $\breve{\boldsymbol{g}} := (g \otimes g', -1) \in \breve{\mathrm{U}}(\boldsymbol{E}) \setminus \mathrm{U}(\boldsymbol{E})$. It is obvious that the diagram

$$\begin{array}{ccc} \widetilde{\mathrm{U}}(\boldsymbol{E}) & \xrightarrow{\widetilde{\mathrm{Ad}}_{\breve{\boldsymbol{g}}}} & \widetilde{\mathrm{U}}(\boldsymbol{E}) \\ \uparrow & & \uparrow \\ \widetilde{G} & \xrightarrow{\widetilde{\mathrm{Ad}}_{\breve{g}}} & \widetilde{G} \end{array} \tag{15}$$

commutes.

Take a character $\chi_{\boldsymbol{E}}$ as in Lemma 3.1, and denote by χ its restriction to $\widetilde{G}$. Then

$$\begin{aligned} \chi^{\breve{g}} &= (\chi_{\boldsymbol{E}}|_{\widetilde{G}})^{\breve{g}} \\ &= (\chi_{\boldsymbol{E}}^{\breve{\boldsymbol{g}}})|_{\widetilde{G}} && \text{by commutativity of (15)} \\ &= (\chi_{\boldsymbol{E}}^{-1})|_{\widetilde{G}} && \text{by Theorem 2.4} \\ &= \chi^{-1}. \end{aligned}$$

□

Fix χ as in Lemma 3.2. Let $\breve{g} \in \breve{G} \setminus G$, and let π be a genuine irreducible admissible smooth representation of $\widetilde{G}$. Then $\pi \otimes \chi$ descends to an irreducible admissible smooth representation of G. By Theorem 1.1, $(\pi \otimes \chi)^{\breve{g}} \cong (\pi \otimes \chi)^{\vee}$, or equivalently, $\pi^{\breve{g}} \otimes \chi^{\breve{g}} \cong \pi^{\vee} \otimes \chi^{-1}$. Therefore, $\pi^{\breve{g}} \cong \pi^{\vee}$ since $\chi^{\breve{g}} = \chi^{-1}$. This proves Theorem 1.3 when $A \neq \Bbbk$.

3.3. ***Proof of Theorem 1.3 for orthogonal groups.*** Assume that $A = \Bbbk$ and $\epsilon = 1$, that is, G is an orthogonal group. In what follows, we show that Lemma 3.2 still holds in this case. Fix a complete polarization $E' = E'_+ \oplus E'_-$ of the symplectic space E'. Then $\boldsymbol{E} = \boldsymbol{E}_+ \oplus \boldsymbol{E}_-$ is a complete polarization of the symplectic space $\boldsymbol{E}$, where $\boldsymbol{E}_\pm := E \otimes E'_\pm$. Depending on this polarization, we define a skew-Hermitian $\Bbbk \times \Bbbk$-module $\boldsymbol{E}'$ as follows. As an abelian group, $\boldsymbol{E}' = \boldsymbol{E}$. The scalar multiplication is given by

$$(ae_1 + be_2)(u + v) := au + bv \quad \text{for } a, b \in \Bbbk,\ u \in \boldsymbol{E}_+, v \in \boldsymbol{E}_-,$$

where $e_1 := (1, 0)$ and $e_2 := (0, 1)$ are the two idempotent elements of $\Bbbk \times \Bbbk$. The skew-Hermitian form is given by

$$\langle u_+ + u_-, v_+ + v_- \rangle_{\boldsymbol{E}'} := \langle u_+, v_- \rangle_{\boldsymbol{E}}\, e_1 + \langle u_-, v_+ \rangle_{\boldsymbol{E}}\, e_2,$$

where $u_+, v_+ \in \boldsymbol{E}_+$, $u_-, v_- \in \boldsymbol{E}_-$.

Let $\breve{g} = (g, -1) \in \breve{G} \setminus G$. Choose an element $(g', -1) \in \breve{G}' \setminus G'$ such that $g'(E'_+) = E'_-$ and $g'(E'_-) = E'_+$. Then

$$\breve{\boldsymbol{g}} := (g \otimes g', -1) \in \breve{\mathrm{U}}(\boldsymbol{E}') \setminus \mathrm{U}(\boldsymbol{E}'),$$

and we have a commutative diagram

$$\begin{array}{ccc} \widetilde{\mathrm{U}}(\boldsymbol{E}') & \xrightarrow{\widetilde{\mathrm{Ad}}_{\breve{\boldsymbol{g}}}} & \widetilde{\mathrm{U}}(\boldsymbol{E}') \\ \uparrow & & \uparrow \\ \widetilde{G} & \xrightarrow{\widetilde{\mathrm{Ad}}_{\breve{g}}} & \widetilde{G}. \end{array}$$

Take a genuine character $\chi_{\boldsymbol{E}'}$ of $\widetilde{\mathrm{U}}(\boldsymbol{E}')$ as in Lemma 3.1, and denote by χ its restriction to $\widetilde{G}$. Then as in the proof of Lemma 3.2, we show that χ fulfills the requirement of Lemma 3.2. Now we argue as in the end of the last subsection, and prove Theorem 1.3 for orthogonal groups.

3.4. ***Proof of Theorem 1.4.*** The group

$$\breve{\boldsymbol{G}} := \breve{G} \times_{\{\pm 1\}} \breve{G}' = \{(g, g', \delta) \mid (g, \delta) \in \breve{G},\ (g', \delta) \in \breve{G}'\}$$

contains $\boldsymbol{G} := G \times G'$ as a subgroup of index two. Define a homomorphism

$$\xi : \breve{\boldsymbol{G}} \to \breve{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}), \quad (g, g', \delta) \mapsto (g \otimes g', \delta).$$

By using the covering map $\widetilde{G} \times \widetilde{G}' \to \boldsymbol{G} = G \times G'$ and the map $\xi|_{\boldsymbol{G}} : \boldsymbol{G} \to \mathrm{Sp}(\boldsymbol{E}_{\Bbbk})$, we form the semidirect product $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$ as in Section 2.1. Let $\breve{\boldsymbol{G}}$ act on $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$ as group automorphisms by

$$\breve{\boldsymbol{g}}(x, y, z) := (\widetilde{\mathrm{Ad}}_{\breve{g}}(x),\ \widetilde{\mathrm{Ad}}_{\breve{g}'}(y),\ \xi(\breve{\boldsymbol{g}})z), \tag{16}$$

where

$$\breve{\boldsymbol{g}} = (g, g', \delta), \quad \breve{g} = (g, \delta), \quad \breve{g}' = (g', \delta),$$

and the last term of the right hand side of (16) is defined as in (8).

Let ω_ψ, π and π' be as in Theorem 1.4.

Lemma 3.3. *View ω_ψ as an admissible smooth representation of $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$ (via the restriction). Then for every $\breve{\boldsymbol{g}} \in \breve{\boldsymbol{G}} \setminus \boldsymbol{G}$, we have*

$$\omega_\psi^\vee \cong \omega_\psi^{\breve{\boldsymbol{g}}}.$$

Proof. Recall that the group $\breve{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk})$ acts on $\widetilde{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}) \ltimes \mathrm{H}(\boldsymbol{E})$ diagonally through its action on the two factors. We have

$$\omega_\psi^\vee \cong \omega_\psi^{\xi(\breve{\boldsymbol{g}})} \tag{17}$$

as smooth oscillator representations of $\widetilde{\mathrm{Sp}}(\boldsymbol{E}_{\Bbbk}) \ltimes \mathrm{H}(\boldsymbol{E})$, since both correspond to the character ψ^{-1}. We prove the lemma by restricting both sides of (17) to the group $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$. □

Lemma 3.4. *Via the inflations, view π and π' as admissible smooth representations of $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$. Then for every $\breve{\boldsymbol{g}} \in \breve{\boldsymbol{G}} \setminus \boldsymbol{G}$, we have*

$$\pi^\vee \cong \pi^{\breve{\boldsymbol{g}}} \quad \textit{and} \quad \pi'^\vee \cong \pi'^{\breve{\boldsymbol{g}}}. \tag{18}$$

Proof. Write $\breve{\boldsymbol{g}} = (g, g', -1)$ and $\breve{g} = (g, -1)$. By Theorem 1.3, we have $\pi^\vee \cong \pi^{\breve{g}}$ as irreducible admissible smooth representations of $\widetilde{G}$. By pulling back this isomorphism to the group $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$, we obtain the first isomorphism of (18). The second isomorphism follows similarly. □

Lemma 3.5. *For every $\breve{\boldsymbol{g}} \in \breve{\boldsymbol{G}} \setminus \boldsymbol{G}$, we have*

$$\omega_\psi^\vee \otimes \pi^\vee \otimes \pi'^\vee \cong (\omega_\psi \otimes \pi \otimes \pi')^{\breve{\boldsymbol{g}}} \tag{19}$$

as smooth representations of $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$.

Proof. This is a combination of Lemma 3.3 and Lemma 3.4. □

Fix an element $\breve{\boldsymbol{g}} \in \breve{\boldsymbol{G}} \setminus \boldsymbol{G}$. Since the action of $\breve{\boldsymbol{g}}$ stabilizes the subgroup $\widetilde{G} \times \widetilde{G}'$ of $(\widetilde{G} \times \widetilde{G}') \ltimes \mathrm{H}(\boldsymbol{E})$, we have

$$\mathrm{Hom}_{\widetilde{G} \times \widetilde{G}'}(\omega_\psi \otimes \pi \otimes \pi', \mathbb{C}) = \mathrm{Hom}_{\widetilde{G} \times \widetilde{G}'}((\omega_\psi \otimes \pi \otimes \pi')^{\breve{\boldsymbol{g}}}, \mathbb{C}). \tag{20}$$

Now Theorem 1.4 is a consequence of (19) and (20).

Acknowledgments

I thank the referee for many detailed suggestions.

References

[Bernšteĭn and Zelevinskiĭ 1976] I. N. Bernšteĭn and A. V. Zelevinskiĭ, "Representations of the group GL(n, F), where F is a local non-Archimedean field", *Uspehi Mat. Nauk* **31**:3 (1976), 5–70. In Russian; translated in *Russian Mathematical Surveys* **31**:3 (1976), 1–68. MR 54 #12988

[Harish-Chandra 1999] Harish-Chandra, *Admissible invariant distributions on reductive p-adic groups*, University Lecture Series **16**, American Mathematical Society, Providence, RI, 1999. MR 2001b:22015 Zbl 0928.22017

[Harris et al. 1996] M. Harris, S. S. Kudla, and W. J. Sweet, "Theta dichotomy for unitary groups", *J. Amer. Math. Soc.* **9**:4 (1996), 941–1004. MR 96m:11041 Zbl 0870.11026

[Kudla 1994] S. S. Kudla, "Splitting metaplectic covers of dual reductive pairs", *Israel J. Math.* **87**:1-3 (1994), 361–401. MR 95h:22019 Zbl 0840.22029

[Li et al. 2009] J.-S. Li, B. Sun, and Y. Tian, "The multiplicity one conjecture for local theta correspondences", preprint, 2009. To appear in *Invent. Math.* arXiv 0903.1419

[Mœglin et al. 1987] C. Mœglin, M.-F. Vignéras, and J.-L. Waldspurger, *Correspondances de Howe sur un corps p-adique*, Lecture Notes in Mathematics **1291**, Springer, Berlin, 1987. MR 91f:11040 Zbl 0642.22002

[Moore 1968] C. C. Moore, "Group extensions of p-adic and adelic linear groups", *Inst. Hautes Études Sci. Publ. Math.* 35 (1968), 157–222. MR 39 #5575 Zbl 0159.03203

[Przebinda 1988] T. Przebinda, "On Howe's duality theorem", *J. Funct. Anal.* **81**:1 (1988), 160–183. MR 89j:22031 Zbl 0678.22007

[Sun 2009] B. Sun, "Multiplicity one theorems for Fourier–Jacobi models", preprint, 2009. arXiv 0903.1417

Received December 7, 2009. Revised April 21, 2010.

BINYONG SUN
ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
BEIJING 100190
CHINA

sun@math.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

ON THE NUMBER OF PAIRS OF POSITIVE INTEGERS $x_1, x_2 \le H$ SUCH THAT x_1x_2 IS A k-TH POWER

DOYCHIN I. TOLEV

We find an asymptotic formula for the number of pairs of positive integers $x_1, x_2 \le H$ such that the product x_1x_2 is a k-th power.

1. Notation

Let H be a sufficiently large positive number and $k \ge 2$ be a fixed integer. By the letters $j, l, m, n, u, v, x, y, z$ we denote positive integers. The letter p is reserved for primes, and $\prod_p$ denotes a product over all primes. By the letters s and w, we denote complex numbers, and $i = \sqrt{-1}$. By ε we denote an arbitrary small positive number. The constants in the Vinogradov and Landau symbols are absolute or depend on ε and k. As usual, $\zeta(s)$ is the Riemann zeta function. By V_k we denote the set of k-free numbers (that is, positive integers not divided by a k-th power of a prime), and N_k is the set of k-th powers of natural numbers. We denote by $\mu(n)$ the Möbius function and by $\tau(n)$ the number of positive divisors of n. Further, we define $\eta(n) = \prod_{p|n} p$. We write (u, v) for the greatest common divisor of u and v. We assume that $\min(1, 0^{-1}) = 1$.

2. Introduction and statement of the result

Let $S_k(H)$ be the number of pairs of positive integers $x_1, x_2 \le H$ whose product x_1x_2 is in N_k. We will establish an asymptotic formula for $S_k(H)$. This problem is related to a result of Heath-Brown and Moroz [1999]. They considered the diophantine equation $x_1x_2x_3 = x_0^3$ and found an asymptotic formula for the number of primitive solutions such that $1 \le x_1, x_2, x_3 \le H$.

It is easy to find an asymptotic formula for the quantity

$$S_k^*(H) = \#\{x_1, x_2 \mid x_1, x_2 \le H,\ (x_1, x_2) = 1,\ x_1x_2 \in N_k\}.$$

Supported by Sofia University grant 028/2009.
MSC2000: 11D45.
Keywords: counting solutions of Diophantine equations.

Indeed, if $(x_1, x_2) = 1$, then $x_1 x_2 \in N_k$ exactly when $x_1 \in N_k$ and $x_2 \in N_k$. Hence

$$\begin{aligned} S_k^*(H) &= \#\{x_1, x_2 \mid x_1, x_2 \le H,\ (x_1, x_2) = 1,\ x_1 \in N_k,\ x_2 \in N_k\} \\ &= \sum_{\substack{z_1, z_2 \le H^{1/k}, \\ (z_1, z_2) = 1}} 1, \end{aligned}$$

and using the well-known property of the Möbius function we get

$$S_k^*(H) = \sum_{z_1, z_2 \le H^{1/k}} \sum_{d \mid (z_1, z_2)} \mu(d) = \sum_{d \le H^{1/k}} \mu(d) \Big(\frac{H^{1/k}}{d} + O(1) \Big)^2.$$

Therefore

$$\begin{aligned} S_k^*(H) &= H^{2/k} \sum_{d \le H^{1/k}} \frac{\mu(d)}{d^2} + O(H^{1/k} \log H) \\ &= \zeta(2)^{-1} H^{2/k} + O(H^{1/k} \log H). \end{aligned} \tag{1}$$

It is also easy to evaluate $S_2(H)$. Indeed, we have

$$S_2(H) = \sum_{d \le H} \sum_{\substack{x_1, x_2 \le H, \\ (x_1, x_2) = d, \\ x_1 x_2 \in N_2}} 1 = \sum_{d \le H} \sum_{\substack{y_1, y_2 \le H/d, \\ (y_1, y_2) = 1, \\ y_1 y_2 d^2 \in N_2}} 1 = \sum_{d \le H} S_2^*(H/d).$$

Now we apply (1) and after calculations that we leave to the reader, we find

$$S_2(H) = \zeta(2)^{-1} H \log H + O\,(H)\,.$$

However it is not clear how to apply (1) in order to evaluate $S_k(H)$ for $k \ge 3$.

Another quantity related to $S_k(H)$ is

$$T_k(H) = \#\{x_1, x_2 \mid x_1 x_2 \le H^2,\ x_1 x_2 \in N_k\} = \sum_{n \le H^{2/k}} \tau(n^k).$$

Using well-known analytic methods, based on Perron's formula and the simplest properties of $\zeta(s)$, we are able to prove the asymptotic formula

$$T_k(H) \sim \gamma_k\, H^{2/k}\, (\log H)^k,$$

where $\gamma_k > 0$ depends only on k. In this paper we show that using the same analytic tools, as well as an idea of Heath-Brown and Moroz [1999], we may find an asymptotic formula for $S_k(H)$ for any $k \ge 2$:

Theorem. *For any integer $k \ge 2$, we have*

$$S_k(H) = c_k H^{2/k} (\log H)^{k-1} + O(H^{2/k} (\log H)^{k-2}), \tag{2}$$

where

$$c_k = \frac{\mathcal{P}_k}{((k-1)!)^2}\Bigl(1+\frac{1}{k^{k-2}}\sum_{k/2<m\le k-1}\frac{(-1)^{k-m}(2m-k)^{k-1}\binom{k-1}{m}}{k-m}\Bigr), \tag{3}$$

$$\mathcal{P}_k = \prod_p \Bigl(1-\frac{1}{p}\Bigr)^{k-1}\Bigl(1+\frac{k-1}{p}\Bigr). \tag{4}$$

3. Some lemmas

Lemma 1. (i) *Every positive integer x can be represented uniquely in the form $x = yz$, where $y \in V_k$ and $z \in N_k$.*

(ii) *Every integer $y \in V_k$ can be written uniquely in the form $y = u_1u_2^2u_3^3\cdots u_{k-1}^{k-1}$, where $u_j \in V_2$ for $1 \le j \le k-1$ and $(u_i, u_j) = 1$ for $1 \le i, j \le k-1,\ i \ne j$.*

(iii) *If $y_1, y_2 \in V_k$ and $y_1y_2 \in N_k$, then $\eta(y_1) = \eta(y_2) = (y_1y_2)^{1/k}$.*

Proof. The proofs of (i) and (ii) can by obtained easily from the fundamental theorem of arithmetic and we leave this to the reader. Let us prove (iii). By our assumption, any prime in the factorization of y_1y_2 occurs with exponent at most $2k-2$, and hence with exponent exactly k. Since the exponent of each prime in y_1 and y_2 is $\le k-1$, the integers y_1 and y_2 have the same prime factors. □

The next lemma is a version of the Perron formula. Denote

$$E(\gamma) = \begin{cases} 1 & \text{if } \gamma \ge 1, \\ 0 & \text{if } 0 < \gamma < 1. \end{cases} \tag{5}$$

Lemma 2. *If $\gamma > 0,\ 0 < c < c_0$ and $T > 1$, then*

$$E(\gamma) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{\gamma^s}{s}\,ds + O(\gamma^c \min(1, T^{-1}|\log\gamma|^{-1})).$$

The constant in the Landau symbol depends only on c_0.

Proof. This is a slightly simplified version of a lemma from [Davenport 2000, Section 17]. □

Some of the basic properties of Riemann's zeta function are presented in the next lemma.

Lemma 3. (i) *$\zeta(s)$ is meromorphic in the complex plane and has a pole only at $s = 1$. It is simple and with a residue equal to* 1.

(ii) *If* $\mathrm{Re}(s) > 1$, *then* $\zeta(s) = \prod_p (1-p^{-s})^{-1}$.

(iii) *If* $\mathrm{Re}(s) \ge \sigma > 1$, *then* $\zeta(s) \ll (\sigma-1)^{-1} + 1$.

(iv) *If* $1/2 \le \sigma_0 \le 1$, $\sigma \ge \sigma_0$ *and* $|t| \ge 2$, *then* $\zeta(\sigma+it) \ll |t|^{(1-\sigma_0)/2+\varepsilon}$.

(v) *There exist* $\lambda_0 > 0$ *such that if* $X \geq 2$, $|t| \leq X$ *and* $\sigma \geq 1 - \lambda_0/\log X$, *then* $\zeta(\sigma + it) \neq 0$.

Proof. See [Titchmarsh 1986, Chapters 1–3 and 5]. □

4. Proof of the theorem

4.1. We already considered the case $k = 2$, so we may assume that $k \geq 3$.

Working as in [Heath-Brown and Moroz 1999] we apply Lemma 1(i) and find that $S_k(H)$ is equal to the number of quadruples y_1, y_2, z_1, z_2 such that

$$y_1, y_2 \in V_k, \quad z_1, z_2 \in N_k, \quad y_1 z_1 \leq H, \quad y_2 z_2 \leq H, \quad y_1 z_1 y_2 z_2 \in N_k.$$

Obviously the last of the above conditions is equivalent to $y_1 y_2 \in N_k$ because z_1 and z_2 are k-th powers. Hence

$$S_k(H) = \sum_{\substack{y_1, y_2 \leq H, \\ y_1, y_2 \in V_k, \\ y_1 y_2 \in N_k}} \sum_{\substack{m_j \leq (H/y_j)^{1/k}, \\ j=1,2}} 1 = \sum_{\substack{y_1, y_2 \leq H, \\ y_1, y_2 \in V_k, \\ y_1 y_2 \in N_k}} ((H/y_1)^{1/k} + O(1))((H/y_2)^{1/k} + O(1)).$$

Expanding brackets, we get

$$S_k(H) = H^{2/k} U_k(H) + O(H^{1/k} W_k(H)), \tag{6}$$

where

$$U_k(H) = \sum_{\substack{y_1, y_2 \leq H, \\ y_1, y_2 \in V_k, \\ y_1 y_2 \in N_k}} (y_1 y_2)^{-1/k} \quad \text{and} \quad W_k(H) = \sum_{\substack{y_1, y_2 \leq H, \\ y_1, y_2 \in V_k, \\ y_1 y_2 \in N_k}} y_1^{-1/k}.$$

Using Lemma 1(iii), we see that for a given y_1 the integer y_2 is determined uniquely. Therefore we have

$$U_k(H) = \sum_{\substack{y \leq H, \\ y \in V_k, \\ \eta(y)^k \leq Hy}} \eta(y)^{-1} \quad \text{and} \quad W_k(H) = \sum_{\substack{y \leq H, \\ y \in V_k, \\ \eta(y)^k \leq Hy}} y^{1/k} \eta(y)^{-1}. \tag{7}$$

To prove the theorem we have to find an asymptotic formula for $U_k(H)$ and to estimate $W_k(H)$.

4.2. Consider first $W_k(H)$. Applying Lemma 1(ii), we get

$$\begin{aligned} W_k(H) &\leq \sum_{u_1 u_2^2 \cdots u_{k-1}^{k-1} \leq H} \frac{(u_1 u_2^2 \cdots u_{k-1}^{k-1})^{1/k}}{u_1 u_2 \cdots u_{k-1}} \\ &= \sum_{u_1 u_2^2 \cdots u_{k-2}^{k-2} \leq H} u_1^{-1+1/k} u_2^{-1+2/k} \cdots u_{k-2}^{-1+(k-2)/k} \sum_{u_{k-1} \leq \left(\frac{H}{u_1 u_2^2 \cdots u_{k-2}^{k-2}}\right)^{1/(k-1)}} u_{k-1}^{-1/k}. \end{aligned}$$

The inner sum is $\ll H^{1/k}(u_1u_2^2\dots u_{k-2}^{k-2})^{-1/k}$; hence

$$W_k(H) \ll H^{1/k} \sum_{u_1u_2^2\cdots u_{k-2}^{k-2}\le H} (u_1u_2\dots u_{k-2})^{-1} \ll H^{1/k}(\log H)^{k-2}. \tag{8}$$

It remains to show that

$$U_k(H) = c_k(\log H)^{k-1} + O\left((\log H)^{k-2}\right). \tag{9}$$

Formula (2) is a consequence of (6), (8) and (9).

4.3. Using (5) and (7), we write $U_k(H)$ in the form

$$U_k(H) = \sum_{\substack{y\le H,\\ y\in V_k}} \eta(y)^{-1}E(Hy\eta(y)^{-k}).$$

We put

$$c = (\log H)^{-1} \quad \text{and} \quad T = (\log H)^{100k^3} \tag{10}$$

and applying Lemma 2 we find that

$$U_k(H) = U^{(1)} + O(\Delta), \tag{11}$$

where

$$U^{(1)} = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{H^s}{s}\Phi(s)ds, \quad \text{and} \quad \Phi(s) = \sum_{\substack{y\le H,\\ y\in V_k}} y^s\eta(y)^{-ks-1} \tag{12}$$

and $\Delta = \sum_{\substack{y\le H,\\ y\in V_k}} \eta(y)^{-1}\min(1, T^{-1}|\log(Hy\eta(y)^{-k})|^{-1})$.

4.4. Consider first the sum Δ. We put

$$\varkappa = T^{-1/2} \tag{13}$$

and write

$$\Delta = \Delta_1 + \Delta_2, \tag{14}$$

where in Δ_1 the summation is taken over y satisfying $|\log(H\,y\,\eta(y)^{-k})| \ge \varkappa$ and in Δ_2 over the other y. To estimate Δ_1 we apply Lemma 1(iii), (10) and (13) to find

$$\begin{aligned}\Delta_1 &\ll T^{-1/2}\sum_{\substack{y\le H,\\ y\in V_k}} \eta(y)^{-1} \ll T^{-1/2} \sum_{u_1u_2^2\cdots u_{k-1}^{k-1}\le H} (u_1u_2\cdots u_{k-1})^{-1}\\ &\ll \frac{(\log H)^{k-1}}{T^{1/2}} \ll 1.\end{aligned} \tag{15}$$

Consider Δ_2. Using its definition and Lemma 1(iii), we find

$$\begin{aligned}\Delta_2 &\ll \sum_{\substack{u_1,u_2,\dots,u_{k-1}:\\ |\log(H/(u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2 u_{k-1}))|<\varkappa}} (u_1u_2\cdots u_{k-1})^{-1} \\ &\ll \sum_{He^{-\varkappa}<u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2u_{k-1}<He^{\varkappa}} (u_1u_2\cdots u_{k-1})^{-1} \\ &\ll \sum_{u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2<2H} (u_1u_2\cdots u_{k-2})^{-1} \sum_{\frac{He^{-\varkappa}}{u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2}<u_{k-1}<\frac{He^{\varkappa}}{u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2}} u_{k-1}^{-1}.\end{aligned}$$

To estimate the inner sum we apply the obvious inequality

$$\sum_{a<n\le b} n^{-1} \le a^{-1} + \log(b/a) \quad \text{for } 0<a<b \tag{16}$$

and find that

$$\Delta_2 \ll \sum_{u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2<2H} \frac{H^{-1}u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2+\varkappa}{u_1u_2\cdots u_{k-2}} \ll H^{-1}\Delta_3+\varkappa(\log H)^{k-2}, \tag{17}$$

where

$$\Delta_3 = \sum_{u_1^{k-1}u_2^{k-2}\cdots u_{k-2}^2<2H} u_1^{k-2}u_2^{k-3}\cdots u_{k-2}. \tag{18}$$

If $k>3$, then

$$\begin{aligned}\Delta_3 &\ll \sum_{u_1^{k-1}u_2^{k-2}\cdots u_{k-3}^3<2H} u_1^{k-2}u_2^{k-3}\cdots u_{k-3}^2 \sum_{u_{k-2}<(2H/(u_1^{k-1}u_2^{k-2}\cdots u_{k-3}^3))^{1/2}} u_{k-2} \\ &\ll H \sum_{u_1^{k-1}u_2^{k-2}\cdots u_{k-3}^3<2H} (u_1u_2\cdots u_{k-3})^{-1} \ll H(\log H)^{k-3}.\end{aligned} \tag{19}$$

The last estimate for Δ_3 is obviously true also for $k=3$. From (10), (13)–(15), (17) and (19), we get

$$\Delta \ll (\log H)^{k-3}. \tag{20}$$

4.5. Consider the expression $\Phi(s)$ defined by (12). Let c and T be specified by (10) and

$$T_1 = 2kT. \tag{21}$$

We apply Lemma 2 again and show that if $\mathrm{Re}(s) = c$, then

$$\Phi(s) = \frac{1}{2\pi i}\int_{c-iT_1}^{c+iT_1} \frac{H^w}{w}\mathcal{M}(s, w)\, dw + O(\Delta^*), \tag{22}$$

where

$$\mathcal{M}(s, w) = \sum_{y=1, y\in V_k}^{\infty} y^{s-w}\,\eta(y)^{-ks-1}, \tag{23}$$

$$\Delta^* = \sum_{y=1, y\in V_k}^{\infty} \eta(y)^{-kc-1}\min(1, T_1^{-1}|\log(H/y)|^{-1}). \tag{24}$$

To justify (22) we note from Euler's identity, (10) and parts (ii) and (iii) of Lemma 3 it follows that

$$\sum_{\substack{y=1,\\ y\in V_k}}^{\infty} \eta(y)^{-kc-1} = \prod_p\Big(1+\frac{k-1}{p^{kc+1}}\Big) \ll \zeta^{k-1}(kc+1) \ll c^{-k+1} \ll (\log H)^{k-1}. \tag{25}$$

Hence $\mathcal{M}(s, w)$ is absolutely and uniformly convergent in $\mathrm{Re}(s) = \mathrm{Re}(w) = c$ because under this assumption we have $\mathcal{M}(s, w) \ll \sum_{y=1, y\in V_k}^{\infty} \eta(y)^{-kc-1}$. This completes the verification of (22).

4.6. Consider the expression Δ^* defined by (24). We write it in the form

$$\Delta^* = \Delta_1^* + \Delta_2^*, \tag{26}$$

where the summation in Δ_1^* is taken over y such that $|\log(H/y)| \ge \varkappa$ and in Δ_2^* over the other y. Using (10), (13), (21) and (25), we find

$$\Delta_1^* \ll T^{-1/2}\sum_{y=1, y\in V_k}^{\infty} \eta(y)^{-kc-1} \ll (\log H)^{k-1-50k^3} \ll 1. \tag{27}$$

To estimate Δ_2^* we apply Lemma 1(iii) and (10), (13), (16) to get

$$\begin{aligned}
\Delta_2^* &\ll \sum_{\substack{He^{-\varkappa}<y<He^{\varkappa},\\ y\in V_k}} \eta(y)^{-1} \ll \sum_{He^{-\varkappa}<u_1u_2^2\cdots u_{k-1}^{k-1}<He^{\varkappa}} (u_1u_2\cdots u_{k-1})^{-1}\\
&\ll \sum_{u_2^2u_3^3\cdots u_{k-1}^{k-1}<2H} (u_2u_3\cdots u_{k-1})^{-1} \sum_{\frac{He^{-\varkappa}}{u_2^2u_3^3\cdots u_{k-1}^{k-1}}<u_1<\frac{He^{\varkappa}}{u_2^2u_3^3\cdots u_{k-1}^{k-1}}} u_1^{-1}\\
&\ll \sum_{u_2^2u_3^3\cdots u_{k-1}^{k-1}<2H} \frac{H^{-1}u_2^2u_3^3\cdots u_{k-1}^{k-1}+\varkappa}{u_2u_3\cdots u_{k-1}}\\
&\ll H^{-1}\Delta_3 + 1, \qquad (28)
\end{aligned}$$

where Δ_3 is given by (18). Applying (19), (26)–(28) we find

$$\Delta^* \ll (\log H)^{k-3}. \tag{29}$$

We substitute in formula (12) the expression for $\Phi(s)$ given by (22) and find a new form of $U^{(1)}$. Using (10) and (29) we see that the contribution to $U^{(1)}$ coming from Δ^* is

$$\ll (\log H)^{k-3} \int_{-T}^{T} \frac{dt}{\sqrt{c^2+t^2}} \ll (\log H)^{k-2}.$$

Therefore, taking also into account (11) and (20), we find

$$U_k(H) = \frac{1}{(2\pi i)^2} \int_{c-iT}^{c+iT} \frac{H^s}{s} \int_{c-iT_1}^{c+iT_1} \frac{H^w}{w} \mathcal{M}(s, w)\, dw\, ds + O((\log H)^{k-2}). \tag{30}$$

4.7. For a fixed s satisfying $\mathrm{Re}(s) = c$ the infinite series $\mathcal{M}(s, w)$ defined by (23) is absolutely and uniformly convergent for $\mathrm{Re}(w) \geq c$ and represents a holomorphic function in $\mathrm{Re}(w) > c$. Applying Euler's identity we find

$$\begin{aligned} \mathcal{M}(s, w) &= \prod_p (1 + p^{-ks-1}(p^{s-w} + p^{2(s-w)} + \cdots + p^{(k-1)(s-w)})) \\ &= \prod_p \Big(1 + \sum_{j=1}^{k-1} p^{-(k-j)s-jw-1}\Big). \end{aligned}$$

Using Lemma 3(ii), we conclude that for $\mathrm{Re}(s) = c$ and $\mathrm{Re}(w) \geq c$, we have

$$\mathcal{M}(s, w) = \mathcal{K}(s, w) \prod_{j=1}^{k-1} \zeta((k-j)s + jw + 1), \tag{31}$$

where

$$\mathcal{K}(s, w) = \prod_p \Big(\Big(1 + \sum_{j=1}^{k-1} p^{-(k-j)s-jw-1}\Big) \prod_{j=1}^{k-1} (1 - p^{-(k-j)s-jw-1})\Big).$$

It is clear that there exists $\delta = \delta(k) \in (0, 1/100)$ such that in the region

$$\mathrm{Re}(s) > -\delta \quad \text{and} \quad \mathrm{Re}(w) > -\delta \tag{32}$$

the function $\mathcal{K}(s, w)$ is holomorphic with respect to s as well as to w and satisfies

$$0 < |\mathcal{K}(s, w)| \ll 1. \tag{33}$$

We have also

$$\mathcal{K}(0, 0) = \mathcal{P}_k, \tag{34}$$

where $\mathcal{P}_k$ is given by (4).

Suppose that we have a fixed $s = c + it$ with $-T \le t \le T$. From (31), (33) and Lemma 3(i), we conclude that the function $H^w w^{-1}\mathcal{M}(s, w)$ has a meromorphic continuation to $\operatorname{Re}(w) > -\delta$ and that poles may occur only at the points

$$w = 0 \quad \text{and} \quad w = (1 - k/m)s \quad \text{for } 1 \le m \le k - 1. \tag{35}$$

All these points are actually simple poles. Indeed, for $w = 0$ this follows immediately from (33) and parts (i) and (v) of Lemma 3. In the case $1 \le m \le k - 1$, the point $w = (1 - k/m)s$ is a simple pole of $\zeta((k - m)s + mw + 1)$ and, due to Lemma 3(v) and (10), it cannot be a pole or zero of $\zeta((k - j)s + jw + 1)$ for $1 \le j \le k - 1$ with $j \ne m$.

For $1 \le m \le k - 1$, we denote by $\mathcal{R}_m(s)$ the residue of $H^w w^{-1}\mathcal{M}(s, w)$ at $w = (1-k/m)s$ and let $\mathcal{R}_0(s)$ be the residue at $w = 0$. A straightforward calculation, based on the arguments above, (33) and Lemma 3(i), leads to

$$\begin{aligned} \mathcal{R}_0(s) &= \mathcal{K}(s, 0) \prod_{j=1}^{k-1} \zeta(js + 1), \\ \mathcal{R}_m(s) &= \frac{H^{(1-k/m)s}}{(m-k)s} \mathcal{K}\Big(s, \Big(1 - \frac{k}{m}\Big)s\Big) \prod_{\substack{j=1,\\ j\ne m}}^{k-1} \zeta\Big(k\Big(1 - \frac{j}{m}\Big)s + 1\Big) \quad \text{for } 1 \le m \le k - 1. \end{aligned} \tag{36}$$

4.8. Let us define

$$\theta = \frac{\delta}{2k^3}. \tag{37}$$

By (10) and (21) and since $s = c + it$, where $-T \le t \le T$, we see that all points (35) are inside the rectangle with vertices $c - iT_1$, $-\theta - iT_1$, $-\theta + iT_1$, $c + iT_1$. Applying the residue theorem we find that

$$\int_{c-iT_1}^{c+iT_1} \frac{H^w}{w}\mathcal{M}(s, w)\, dw = 2\pi i \sum_{m=0}^{k-1} \mathcal{R}_m(s) + I_1 + I_2 + I_3,$$

where

$$\begin{aligned} I_1 &= \int_{c-iT_1}^{-\theta-iT_1} \frac{H^w}{w}\mathcal{M}(s, w)\, dw, \qquad I_2 = \int_{-\theta-iT_1}^{-\theta+iT_1} \frac{H^w}{w}\mathcal{M}(s, w)\, dw, \\ I_3 &= \int_{-\theta+iT_1}^{c+iT_1} \frac{H^w}{w}\mathcal{M}(s, w)\, dw. \end{aligned}$$

From the formula above and (30) we get

$$U_k(H) = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{H^s}{s} \sum_{m=0}^{k-1} \mathcal{R}_m(s)\, ds \; + J_1 + J_2 + J_3 + O((\log H)^{k-2}). \tag{38}$$

Here J_μ are the contributions coming from I_μ for $\mu = 1, 2, 3$ and we will see that we may neglect them.

To estimate J_μ we will first show that if $s = c + it$, where $|t| \le T$, and if w belongs to some of the sets of integration of I_1, I_2 or I_3, then

$$\mathcal{M}(s, w) \ll T^{k^2\theta}. \tag{39}$$

Having in mind (31) and (33), we see that in order to verify this it is enough to establish that for s and w satisfying the conditions above, we have

$$\zeta(\lambda) \ll T^{k\theta}, \quad \text{where} \quad \lambda = (k-j)s + jw + 1 \quad \text{for } 1 \le j \le k-1. \tag{40}$$

If $w = \beta + iT_1$ (or $w = \beta - iT_1$), where $-\theta \le \beta \le c$, then from (10), (21), (37) it follows that for the number λ given by (40), we have $\mathrm{Re}(\lambda) \ge 1 - k\theta$ and $T \ll |\mathrm{Im}(\lambda)| \ll T$. Hence the estimate (40) is a consequence of Lemma 3(iv). Suppose now that $w = -\theta + it_1$, where $|t_1| \le T_1$. From (10), (21) and (37), we get $\mathrm{Re}(\lambda) \ge 1 - k\theta$ and $|\mathrm{Im}(\lambda)| \ll T$. If $|\mathrm{Im}(\lambda)| \ge 2$, then the estimate (40) follows again from Lemma 3(iv). In the case $|\mathrm{Im}(\lambda)| < 2$ we use also the inequality $\mathrm{Re}(\lambda) \le 1 - \theta/2$ to conclude that $\zeta(\lambda) \ll 1$, so the estimate (40) is true again.

From the definitions of J_μ and (10), (21), (37) and (39), we find

$$J_1, J_3 \ll \int_{-T}^{T} \frac{1}{\sqrt{c^2+t^2}} \int_{-\theta}^{c} \frac{T^{k^2\theta}}{\sqrt{\beta^2+T_1^2}}\, d\beta\, dt \ll c^{-1} + \log T \ll \log H,$$
$$J_2 \ll \int_{-T}^{T} \frac{1}{\sqrt{c^2+t^2}} \int_{-T_1}^{T_1} \frac{H^{-\theta}T^{k^2\theta}}{\sqrt{\theta^2+t_1^2}}\, dt_1\, dt \ll H^{-\theta}(c^{-1} + \log T)T^{k^2\theta} \log T \ll 1.$$

This means that the terms J_μ in formula (38) can be omitted. Then using (36), we get

$$U_k(H) = \frac{1}{2\pi i}\Big(\mathfrak{N}_0 + \sum_{m=1}^{k-1} \frac{1}{m-k}\mathfrak{N}_m\Big) + O((\log H)^{k-2}), \tag{41}$$

where

$$\mathfrak{N}_m = \int_{c-iT}^{c+iT} \Xi_m(s)\, ds \tag{42}$$

and

$$\Xi_0(s) = s^{-1} H^s \mathcal{K}(s, 0) \prod_{j=1}^{k-1} \zeta(js+1), \tag{43}$$

$$\Xi_m(s) = s^{-2} H^{(2-k/m)s} \mathcal{K}\Big(s, \Big(1 - \frac{k}{m}\Big)s\Big) \prod_{\substack{j=1,\\ j\ne m}}^{k-1} \zeta\Big(k\Big(1 - \frac{j}{m}\Big)s + 1\Big) \quad \text{for } 1 \le m \le k-1. \tag{44}$$

4.9. Consider first $\mathfrak{N}_m$ for $1 \le m \le k/2$. Since $\Xi_m(s)$ is a holomorphic function in the rectangle with vertices $c - iT$, $\theta - iT$, $\theta + iT$ and $c + iT$, we have

$$\begin{aligned}\mathfrak{N}_m &= \int_{c-iT}^{\theta-iT} \Xi_m(s)\,ds + \int_{\theta-iT}^{\theta+iT} \Xi_m(s)\,ds + \int_{\theta+iT}^{c+iT} \Xi_m(s)\,ds \\ &= \mathfrak{N}_m^{(1)} + \mathfrak{N}_m^{(2)} + \mathfrak{N}_m^{(3)},\end{aligned} \tag{45}$$

say. If s belongs to the sets of integration of $\mathfrak{N}_m^{(1)}$ or $\mathfrak{N}_m^{(3)}$ and if $1 \le j \le k-1$, $j \ne m$, then from Lemma 3(iv), it follows that

$$\zeta(k(1 - j/m)s + 1) \ll T^{k^2\theta}.$$

Hence, using (33), (37) and our assumption $1 \le m \le k/2$, we find

$$\mathfrak{N}_m^{(1)}, \mathfrak{N}_m^{(3)} \ll \int_c^\theta \frac{H^{(2-k/m)\beta}}{\beta^2 + T^2} T^{k^3\theta}\,d\beta \ll T^{k^3\theta - 2} \ll 1. \tag{46}$$

Suppose now that s belongs to the set of integration of $\mathfrak{N}_m^{(2)}$ (that is, $s = \theta + it$ for $|t| \le T$) and consider the number $\tilde{\lambda} = k(1 - j/m)s + 1$. It is easy to see that for each j that occurs in (44), we have

$$\operatorname{Re}(\tilde{\lambda}) \ge 1 - k^2\theta, \quad |\operatorname{Re}(\tilde{\lambda}) - 1| \ge \theta, \quad |\operatorname{Im}(\tilde{\lambda})| \le k^2|t|.$$

Hence an application of Lemma 3(iv) gives

$$\zeta(\tilde{\lambda}) \ll (1 + |t|)^{k^2\theta}.$$

Therefore

$$\mathfrak{N}_m^{(2)} \ll \int_{-T}^{T} \frac{H^{(2-k/m)\theta}}{\theta^2 + t^2}(1 + |t|)^{k^3\theta}\,dt \ll 1. \tag{47}$$

From (45)–(47), we get $\mathfrak{N}_m \ll 1$ for $1 \le m \le k/2$ and using (41) we find

$$U_k(H) = \frac{1}{2\pi i}\Bigl(\mathfrak{N}_0 + \sum_{k/2 < m \le k-1} \frac{1}{m-k}\mathfrak{N}_m\Bigr) + O((\log H)^{k-2}). \tag{48}$$

4.10. Consider now $\mathfrak{N}_m$ for $k/2 < m \le k-1$. The function $\Xi_m(s)$ has a pole only at $s = 0$ and it is not difficult to compute that the corresponding residue is equal to

$$\mathcal{L}_m(\log H)^{k-1} + O((\log H)^{k-2}),$$

where

$$\mathcal{L}_m = \frac{(2m-k)^{k-1}(-1)^{k-m-1}\binom{k-1}{m}\mathcal{P}_k}{((k-1)!)^2 k^{k-2}}. \tag{49}$$

We leave the standard verification to the reader. From (42) and the residue theorem we get

$$\mathfrak{N}_m = 2\pi i \mathscr{L}_m (\log H)^{k-1} + \mathfrak{N}'_m + \mathfrak{N}''_m + \mathfrak{N}'''_m + O((\log H)^{k-2}), \tag{50}$$

where

$$\mathfrak{N}'_m = \int_{c-iT}^{-\theta-iT} \Xi_m(s)\, ds, \quad \mathfrak{N}''_m = \int_{-\theta-iT}^{-\theta+iT} \Xi_m(s)\, ds, \quad \mathfrak{N}'''_m = \int_{-\theta+iT}^{c+iT} \Xi_m(s)\, ds.$$

Using Lemma 3(iv), we find that if s belongs to the set of integration of some of the integrals above, then the product of the values of the zeta-function in the definition (44) is $\ll T^{k^3\theta}$. Hence from (10), (33), (37) and our assumption $k/2 < m \le k-1$, it follows that

$$\begin{aligned} \mathfrak{N}'_m, \mathfrak{N}'''_m &\ll \int_{-\theta}^{c} \frac{T^{k^3\theta}}{\beta^2+T^2}\, d\beta \ll 1 \\ \mathfrak{N}''_m &\ll \int_{-T}^{T} \frac{H^{-(2-k/m)\theta}}{\theta^2+t^2} T^{k^3\theta}\, dt \ll H^{-(2-k/m)\theta}\, T^{k^3\theta} \ll 1. \end{aligned} \tag{51}$$

From (50) and (51), we find

$$\mathfrak{N}_m = 2\pi i \mathscr{L}_m (\log H)^{k-1} + O((\log H)^{k-2}) \quad \text{for } k/2 < m \le k-1. \tag{52}$$

4.11. It remains to consider $\mathfrak{N}_0$. It is not difficult to see that the function $\Xi_0(s)$ specified by (43) has a pole only at $s = 0$, with residue equal to

$$\mathscr{L}_0 (\log H)^{k-1} + O((\log H)^{k-2}),$$

where

$$\mathscr{L}_0 = \frac{\mathscr{P}_k}{((k-1)!)^2}. \tag{53}$$

From (42) and the residue theorem we find

$$\mathfrak{N}_0 = 2\pi i \mathscr{L}_0 (\log H)^{k-1} + \mathfrak{N}'_0 + \mathfrak{N}''_0 + \mathfrak{N}'''_0 + O((\log H)^{k-2}),$$

where

$$\mathfrak{N}'_0 = \int_{c-iT}^{-\theta-iT} \Xi_0(s)\, ds, \quad \mathfrak{N}''_0 = \int_{-\theta-iT}^{-\theta+iT} \Xi_0(s)\, ds, \quad \mathfrak{N}'''_0 = \int_{-\theta+iT}^{c+iT} \Xi_0(s)\, ds.$$

Arguing as above, we conclude that $\mathfrak{N}'_0$, $\mathfrak{N}''_0$, $\mathfrak{N}'''_0 \ll 1$ (we leave the verification to the reader). Hence

$$\mathfrak{N}_0 = 2\pi i \mathscr{L}_0 (\log H)^{k-1} + ((\log H)^{k-2}). \tag{54}$$

From (3), (34), (48), (49), and (52)–(54), we obtain (9), and the proof of the theorem is complete. □

References

[Davenport 2000] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics **74**, Springer, New York, 2000. MR 2001f:11001 Zbl 1002.11001

[Heath-Brown and Moroz 1999] D. R. Heath-Brown and B. Z. Moroz, "The density of rational points on the cubic surface $X_0^3 = X_1X_2X_3$", *Math. Proc. Cambridge Philos. Soc.* **125**:3 (1999), 385–395. MR 2000f:11080 Zbl 0938.11016

[Titchmarsh 1986] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd ed., Oxford University Press, 1986. MR 88c:11049 Zbl 0601.10026

Received September 17, 2009.

DOYCHIN I. TOLEV
FACULTY OF MATHEMATICS AND INFORMATICS
SOFIA UNIVERSITY "ST. KL. OHRIDSKY"
5 J.BOURCHIER
1164 SOFIA
BULGARIA

dtolev@fmi.uni-sofia.bg
http://www.tolev.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 249, No. 2, 2011

CORRECTION TO THE ARTICLE A FLOER HOMOLOGY FOR EXACT CONTACT EMBEDDINGS

KAI CIELIEBAK AND URS ADRIAN FRAUENFELDER

Volume **239**:2 (2009), 251–316

The paper in question included an appendix, titled "A Wasserman-type theorem for the Rabinowitz action functional", where we showed that the Rabinowitz action functional is generically Morse–Bott and the Morse–Bott manifold is the disjoint union of the energy hypersurface itself, representing the constant Reeb orbits, and a circle for each Reeb orbit. The treatment of multiple covered Reeb orbits contained a gap, which is filled in this note.

Appendix B of s devoted to showing that the Rabinowitz action functional is generically Morse–Bott and the corresponding Morse–Bott manifold is the disjoint union of the energy hypersurface itself, representing the constant Reeb orbits, and a circle for each Reeb orbit. Here we fix a gap in the proof, pointed out to us by Will Merry and Gabriel Paternain.

In the Claim in Step 2 of the proof of Theorem B.1 we asserted that $\bar{D}S(H, w)$ is surjective for every $(H, w) \in S^{-1}(0)$ whenever w is not a fixed point of the S^1-action. This assertion is incorrect as stated; it is only true if the underlying Reeb orbit v is simple. The trouble is inequality (70), which a priori only holds in a neighborhood of t_0, and might fail to hold globally on the circle if the Reeb orbit is multiply covered and hence comes back to $v(t_0)$. Therefore the proof of Theorem B.1 as it stands only proves that the Rabinowitz action functional is generically Morse–Bott on the constant and simple Reeb orbits.

To prove the full assertion of Theorem B.1 we need to show in addition that generically no root of unity arises as an eigenvalue of the linearized Reeb flow at a simple periodic orbit. But this fact follows from a classical theorem of C. Robinson [1970, Lemma 19].

Here is how this works. For $T > 0$ and $k \in \mathbb{N}$, denote by $\mathcal{U}(T, k) \subset C_c^\infty(V)$ the subset of Hamiltonians H with the following properties. If $k = 1$, then $\mathcal{U}(T, 1)$ consists of all Hamiltonians such that the Rabinowitz action functional $\mathcal{A}^H$ is Morse–Bott at the constant Reeb orbits and all simple Reeb orbits of period less

MSC2000: 53D10, 53D40.
Keywords: contact manifolds, Floer homology, Rabinowitz action functional.

than or equal to T (since the Reeb orbit is allowed to traverse backwards we here actually mean the absolute value of the period). If $k \geq 2$ then $\mathcal{U}(T, k) \subset \mathcal{U}(T, 1)$ consists of all $H \in \mathcal{U}(T, 1)$ with the additional property that the linearized Reeb flow at each simple Reeb orbit of period less than or equal to T has no eigenvalues equal to roots of unity of order less than or equal to k. As it follows from our arguments in the proof of Theorem B.1, for each $T > 0$ the subset $\mathcal{U}(T, 1)$ is open and dense in $C_c^\infty(V)$. If $H \in \mathcal{U}(T, 1)$, we deduce from the Arzelà–Ascoli Theorem that there are only finitely many simple Reeb orbits of period at most T. Hence by Robinson's result for each $k \in \mathbb{N}$ the subset $\mathcal{U}(T, k)$ is dense in $\mathcal{U}(T, 1)$. Again by Arzelà–Ascoli $\mathcal{U}(T, k)$ is also open in $\mathcal{U}(T, 1)$. Hence we conclude that for each $T > 0$ and for each $k \in \mathbb{N}$ the set $U(T, k)$ is open and dense in $C_c^\infty(V)$. Now set

$$\mathcal{U} = \bigcap_{\substack{N \in \mathbb{N} \\ k \in \mathbb{N}}} \mathcal{U}(N, k).$$

The subset $\mathcal{U}$ is obviously of second category in $C_c^\infty(V)$ and if $H \in \mathcal{U}$ then the Rabinowitz action functional $\mathcal{A}^H$ is Morse–Bott at the constants and at all simple Reeb orbits. Moreover, the linearized Reeb flow at each simple Reeb orbit has no root of unity as eigenvalue. Hence $\mathcal{A}^H$ is Morse–Bott at all Reeb orbits and its critical manifold consists of the disjoint union of a copy of the hypersurface and circles for each nontrivial Reeb orbit. This fills up the gap in Appendix B.

References

[Robinson 1970] R. C. Robinson, "Generic properties of conservative systems", *Amer. J. Math.* **92** (1970), 562–603. MR 42 #8517

Received September 28, 2010.

KAI CIELIEBAK
DEPARTMENT OF MATHEMATICS
LUDWIG-MAXIMILIAN UNIVERSITY
THERESIENSTRASSE 39
D-80333 MUNICH
GERMANY
kai@mathematik.uni-muenchen.de

URS ADRIAN FRAUENFELDER
DEPARTMENT OF MATHEMATICS AND RESEARCH INSTITUTE OF MATHEMATICS
SEOUL NATIONAL UNIVERSITY
SEOUL 151-747
SOUTH KOREA
frauenf@snu.ac.kr

CONTENTS

Volume 249, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 249 No. 2 February 2011

0030-8730(201102)249:2;1-B