Pacific Journal of Mathematics

Volume 269 No. 2 June 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

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

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

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.257

TOTARO'S QUESTION FOR SIMPLY CONNECTED GROUPS OF LOW RANK

JODI BLACK AND RAMAN PARIMALA

Let k be a field and let G be a connected linear algebraic group over k. In a 2004 paper, Totaro asked whether a torsor X under G and over k which admits a zero cycle of degree d also admits a closed étale point of degree dividing d. We consider this question in the setting where G is a simply connected, semisimple group of rank at most 2 and k is of characteristic different from 2.

Introduction

Serre [1995, p. 233] raised the following question:

Serre's question: Let k be a field and let G be a connected linear algebraic group defined over k. Let X be a G-torsor over k. Suppose X admits a zero cycle of degree 1. Does X have a k-rational point?

An affirmative answer to Serre's question is known in a number of special cases. See, for example, [Sansuc 1981; Bayer-Fluckiger and Lenstra 1990; Black 2011a; 2011b]. Burt Totaro [2004] posed the following generalization of Serre's question:

Totaro's question: Let k be a field and let G be a connected linear algebraic group defined over k. Let X be a G-torsor over k. Suppose X admits a zero cycle of degree d. Does X have a closed étale point of degree dividing d?

An affirmative answer to Totaro's question when $G = \mathrm{PGL}_n$ is a classical result in the theory of central simple algebras. Tits [1992] associated to any absolutely simple, linear algebraic k-group G, an integer $n(G)$. The values of $n(G)$ are shown in Table 1 below, where ν denotes the 2-adic valuation. One can show that for any G-torsor X, there is a separable field extension L/k such that X has a rational point over L and $[L:k]$ divides $n(G)^2$ [Serre 1995, Section 2.3]. Thus, Tits' construction gives an affirmative answer to Totaro's question provided $n(G)^2$ divides d. Garibaldi and Hoffmann [2006] give an affirmative answer to Totaro's question for semisimple groups which are of type G_2, of reduced type F_4 or simply

Parimala is partially supported by National Science Foundation grant DMS-1001872.
MSC2010: primary 11E72; secondary 11E57, 11E81.
Keywords: zero cycles, Galois cohomology, classical groups,

Type of group	$n(G)$
A_n	$2(n+1)$
B_n	2^n
C_n	$2^{\nu(n)+1}$
D_n $(n \neq 4)$	$2^{n+\nu(n)}$

Table 1. Values of $n(G)$ for classical groups.

connected of type ${}^1E^0_{6,6}$ or ${}^1E^{28}_{6,2}$. Their work extended previous results of Totaro [2004] which gave an affirmative answer for split, simply connected groups of type G_2, F_4 and E_6. Results in [Black 2011b] give an affirmative answer to Totaro's question in the case where G is a simply connected or adjoint, semisimple, classical group and d is prime to $n(G)$.

In this paper we show the following:

Theorem 0.1. *The answer to Totaro's question is yes if k is of characteristic different from* 2 *and G is a semisimple*, *simply connected*, *classical group such that rank $G_{\bar{k}} \leq 2$.*

1. Galois cohomology

Let k be a field, let k_s be a separable closure of k and let $\Gamma_k = \mathrm{Gal}(k_s/k)$ be the absolute Galois group of k. We write $H^1(k, G)$ for the first Galois cohomology set $H^1(\Gamma_k, G(k_s))$. Given any finite field extension L/k there is a canonical restriction map $H^1(k, G) \to H^1(L, G)$. If $\lambda \in H^1(k, G)$ is any element, we write λ_L for the image of λ under the restriction map $H^1(k, G) \to H^1(L, G)$.

For our convenience, we will consider the formulation of Totaro's question in Galois cohomology:

Totaro's question: Let k be a field and let G be a connected linear algebraic group defined over k. Let $\{L_i\}_{1\leq i\leq m}$ be a set of finite field extensions of k and let $d = \gcd\{[L_i : k]_{1\leq i\leq m}\}$. If $\lambda_{L_i} = 1$ for all i, is there a finite, separable field extension F of k such that $\lambda_F = 1$ and $[F : k]$ divides d?

2. Results

In this section, we consider Totaro's question for various groups G.

The case $G = \mathrm{SL}_1(A)$.

Theorem 2.1. *The answer to Totaro's question is yes if $G = \mathrm{SL}_1(A)$ for A a central simple algebra over k of prime index.*

Proof. Let $\{L_i\}_{1\leq i\leq m}$ be a set of finite field extensions of k and suppose $\lambda \in H^1(k, \mathrm{SL}_1(A))$ is an element such that $\lambda_{L_i} = 1$ for all i. Let $d = \gcd\{[L_i : k]_{1\leq i\leq m}\}$.

We will find F/k separable such that $\lambda_F = 1$ and $[F:k]$ divides d.

Since by [Knus et al. 1998, Theorem 29.2], $H^1(k, \mathrm{GL}_1(A)) = 1$, the short exact sequence

$$1 \longrightarrow \mathrm{SL}_1(A) \longrightarrow \mathrm{GL}_1(A) \xrightarrow{\mathrm{Nrd}} G_m \longrightarrow 1$$

induces the long exact sequence

$$A^* \xrightarrow{\mathrm{Nrd}} k^* \longrightarrow H^1(k, \mathrm{SL}_1(A)) \longrightarrow 1 \tag{2.1.1}$$

in Galois cohomology, where Nrd is the reduced norm. By (2.1.1) above,

$$H^1(k, \mathrm{SL}_1(A)) \cong k^*/\operatorname{Nrd}(A^*),$$

and we can identify λ with the class of an element of k^* which is in $\operatorname{Nrd}(A_{L_i})$ for all i. For simplicity, we will also refer to this element as λ. Let the index of A be s and choose L contained in A a separable field extension of k of degree s which splits A [Gille and Szamuely 2006, Propositions 4.5.3 and 4.5.4]. Then $\operatorname{Nrd}(A_L) = L^*$ and λ is in $\operatorname{Nrd}(A_L)$. So if s divides d we may take $F = L$. Recall that s is prime. So if s does not divide d then $\gcd(s, d) = 1$. It is well known that $N_{L/k}(\operatorname{Nrd}(A_L)) \subseteq \operatorname{Nrd}(A)$. In particular, $\lambda^s = N_{L/k}(\lambda)$ is in $\operatorname{Nrd}(A)$. Since $\operatorname{Nrd}(A)$ is a group and $N_{L_i/k}(\lambda) \in \operatorname{Nrd}(A)$ for all i, we find that λ^d is in $\operatorname{Nrd}(A)$. In turn, λ is in $\operatorname{Nrd}(A)$ and we can take $F = k$. □

The case $G = \mathrm{SU}(A, \sigma)$.

Theorem 2.2. *The answer to Totaro's question is yes if k is of characteristic different from 2 and $G = \mathrm{SU}(A, \sigma)$ for a central simple algebra A of degree 3 over K, $k = K^\sigma$ and $[K:k] = 2$.*

Proof. Let $\{L_i\}_{1 \le i \le m}$ be a set of finite field extensions of k and suppose $\lambda \in H^1(k, \mathrm{SU}(A, \sigma))$ is an element such that $\lambda_{L_i} = 1$ for all i. Let $d = \gcd\{[L_i : k]\}$. We will find F/k separable such that $\lambda_F = 1$ and $[F:k]$ divides d.

The case where d is coprime to 2 and 3 was covered in [Black 2011b, Theorem 3.4]. If $6 \mid d$, we take L to be a separable extension of K of degree dividing 3 which splits A. Since K/k is Galois, L/k is separable of degree dividing 6. Since $H^1(K, \mathrm{SU}(A, \sigma)) = H^1(K, \mathrm{SL}_1(A))$ and L splits A, $H^1(L, \mathrm{SU}(A, \sigma)) = \{1\}$ by Hilbert's Theorem 90. Therefore, for any $\lambda \in H^1(k, \mathrm{SU}(A, \sigma))$, $\lambda_L = 1$ and we can take $F = L$. Now suppose $2 \mid d$ and $3 \nmid d$. Fix an index i such that $[L_i : k]$ is prime to 3 and $\lambda_{L_i} = 1$. Consider $L_i K$, the compositum of L_i and K. Since, by assumption, 3 is prime to $[L_i : k]$, and $[K:k] = 2$, we know that 3 is prime to $[L_i K : k]$. Therefore, 3 is prime to $[L_i K : K]$. Let L be a separable splitting field of A such that $[L:K]$ is equal to the index of A. Since $\deg_K(A) = 3$, either $[L:K] = 1$ or $[L:K] = 3$. In either case, L, $L_i K$ is a pair of field extensions of K such that $\lambda_L = 1 = \lambda_{L_i K}$ and gcd$\{[L:K],[L_iK:K]\}$ is 1. Since $H^1(K,\mathrm{SU}(A,\sigma)) = H^1(K,\mathrm{SL}_1(A))$ we have $\lambda_K = 1$ by Theorem 2.1, and we can take $F = K$. The final setting to consider is the case where $3 \mid d$ and $2 \nmid d$. Since d is odd, we can fix an index i such that $[L_i : k]$ is odd and $\lambda_{L_i} = 1$. Let $R_{K/k}G_m$ be the Weil transfer of G_m and let $R^1_{K/k}G_m$ be defined as the kernel of the norm map $N_{K/k} : R_{K/k}G_m \to G_m$. The short exact sequence

$$1 \to \mathrm{SU}(A,\sigma) \to U(A,\sigma) \to R^1_{K/k}G_m \to 1$$

induces the commutative diagram

$$\begin{tikzcd}
K^{*1} \arrow[r, "\delta"] \arrow[d] & H^1(k,\mathrm{SU}(A,\sigma)) \arrow[r, "j"] \arrow[d] & H^1(k,U(A,\sigma)) \arrow[d] \\
(K\otimes L_i)^{*1} \arrow[r] & H^1(L_i,\mathrm{SU}(A,\sigma)) \arrow[r] & H^1(L_i,U(A,\sigma))
\end{tikzcd}$$

where K^{*1} and $(K\otimes L_i)^{*1}$ denote the norm-one elements in K^* and $(K\otimes L_i)^*$ respectively. By a result of Bayer-Fluckiger and Lenstra [1990, Theorem 2.1], $j(\lambda) = 1$. In particular, we can choose $\alpha \in K^{*1}$ such that $\delta(\alpha) = \lambda$. In the case where A is split, $H^1(K,\mathrm{SU}(A,\sigma)) = H^1(K,\mathrm{SL}_1(A)) = \{1\}$. Then, since K and L_i are field extensions of coprime degree with $\lambda_K = \lambda_{L_i} = 1$, the desired result holds by [Black 2011b, Theorem 4.4]. Since $\deg(A) = 3$, if A is not split, then A is a division algebra and by [Albert 1963] (see also [Knus et al. 1998, Theorem 19.14]), there is a k-subalgebra L of A such that L/k is étale of degree three. Since A is division, L is a field. Consider the diagram

$$\begin{tikzcd}
U(A,\sigma)(k) \arrow[r] \arrow[d] & K^{*1} \arrow[r, "\delta"] \arrow[d] & H^1(k,\mathrm{SU}(A,\sigma)) \arrow[r, "j"] \arrow[d] & H^1(k,U(A,\sigma)) \arrow[d] \\
U(A,\sigma)(L) \arrow[r] & (K\otimes L)^{*1} \arrow[r] & H^1(L,\mathrm{SU}(A,\sigma)) \arrow[r] & H^1(L,U(A,\sigma))
\end{tikzcd}$$

For $x \in (K\otimes L)^{*1}$, write $x = y^{-1}\bar{y}$ for $y \in (K\otimes L)^*$ where $\bar{\ }$ denotes the nontrivial automorphism of K/k. Since $A\otimes L$ is split, y is a reduced norm from $A\otimes L$. In view of [Merkurjev 1995, Proposition 6.1], the image of $\mathrm{Nrd}(U(A,\sigma) \to (K\otimes L)^{*1})$ contains x. Thus $\lambda_L = 1$ and we may take $F = L$. $\square$

***The case* $\mathbf{G = Spin}(q)$.** The following result will be useful:

Proposition 2.3. *Let k be a field of characteristic different from 2 and let q be a quadratic form over k of dimension ≤ 5. Let $\lambda \in H^1(k,\mathrm{Spin}(q))$ be any element. Then there exists a (separable) field extension F of k such that $[F:k]$ divides 2 and $\lambda_F = 1$.*

Proof. Consider the short exact sequence

$$1 \longrightarrow \mu_2 \xrightarrow{i} \mathrm{Spin}(q) \xrightarrow{\pi} O^+(q) \longrightarrow 1,$$

which induces the exact sequence in Galois cohomology

$$H^1(k, \mu_2) \xrightarrow{i} H^1(k, \mathrm{Spin}(q)) \xrightarrow{\pi} H^1(k, O^+(q)). \tag{2.3.1}$$

The pointed set $H^1(k, O^+(q))$ classifies quadratic forms over k of the same dimension and discriminant as q. Let $q' = \pi(\lambda)$. Then $q \perp -q'$ has even dimension, trivial discriminant and trivial Clifford invariant since q' is in the image of π. Thus $q \perp -q' \in I^3(k)$.

First consider the case where $\dim(q) < 4$. Then, $\dim(q \perp -q') < 8$ and by the Arason–Pfister Hauptsatz [Lam 1980, Chapter X, Hauptsatz 5.1], $q \perp -q'$ is hyperbolic. Equivalently, $q \cong q'$ and $q' = 1$ in $H^1(k, O^+(q))$. Using the exactness of (2.3.1), choose η in $H^1(k, \mu_2)$ such that $i(\eta) = \lambda$. Since $H^1(k, \mu_2) \cong k^*/k^{*2}$ we can choose F/k a field extension of degree at most 2 such that $\eta_F = 1 \in H^1(F, \mu_2)$. By commutativity of (2.3.2) below, $\lambda_F = 1$ in $H^1(F, \mathrm{Spin}(q))$.

$$\begin{tikzcd} H^1(k, \mu_2) \arrow[r] \arrow[d] & H^1(k, \mathrm{Spin}(q)) \arrow[d] \\ H^1(F, \mu_2) \arrow[r] & H^1(F, \mathrm{Spin}(q)) \end{tikzcd} \tag{2.3.2}$$

Suppose instead that $\dim(q) = 4$. Let $d = \mathrm{disc}(q)$ and write $q = a\langle 1, b, c, bcd\rangle$. By [Lam 1980, Chapter XII, Proposition 2.4], there is an element $\alpha \in k^*$ such that $q' \cong \alpha q$ and we may write $q \perp -q' \cong \langle 1, -\alpha\rangle q = a\langle 1, -\alpha\rangle\langle 1, b, c, bcd\rangle$. Let e_2 be the map from $I^2(k) \to H^2(k, \mu_2)$ induced by the Clifford invariant. Since $q \perp -q' \in I^3(k)$, $e_2(q \perp -q') = (d) \cup (\alpha) = 0 \in H^2(k, \mu_2)$ [Elman et al. 2008, 16.2] and so $\langle 1, -\alpha, -d, \alpha d\rangle$ is hyperbolic. Equivalently, $\langle 1, -\alpha\rangle d \cong \langle 1, -\alpha\rangle$ and $q \perp -q' \cong a\langle 1, -\alpha\rangle\langle 1, b, c, bc\rangle = a\langle 1, -\alpha\rangle\langle 1, b\rangle\langle 1, c\rangle$. Let $F = k(\sqrt{-b})$. Then $[F : k] \leq 2$, $(q \perp -q')_F$ is hyperbolic and $q'_F = 1 \in H^1(F, O^+(q))$. Consider the diagram

$$\begin{tikzcd} O^+(q)(k) \arrow[r] \arrow[d] & H^1(k, \mu_2) \arrow[r] \arrow[d] & H^1(k, \mathrm{Spin}(q)) \arrow[r] \arrow[d] & H^1(k, O^+(q)) \arrow[d] \\ O^+(q)(F) \arrow[r, "\mathrm{sn}"] & H^1(F, \mu_2) \arrow[r, "i"] & H^1(F, \mathrm{Spin}(q)) \arrow[r, "\pi"] & H^1(F, O^+(q)) \end{tikzcd} \tag{2.3.3}$$

By commutativity of the right rectangle, $\pi(\lambda_F) = 1$ and by the exactness of the bottom row, $\lambda_F \in \mathrm{im}(i)$. But since $q \cong a\langle 1, b, c, bcd\rangle$, q_F is isotropic. Thus, the spinor norm $\mathrm{sn}: O^+(q)(F) \to H^1(F, \mu_2)$ is onto [Baeza 1978, p. 78] and therefore, since $\lambda_F \in \mathrm{im}(i)$, $\lambda_F = 1$.

Now suppose $\dim(q) = 5$. Since $q \perp -q'$ is a rank 10 form in $I^3(k)$, it is isotropic [Lam 1980, Chapter XII, Proposition 2.8]. Therefore q and q' have a common slot and we can write $q = \langle a \rangle \perp q_1$ and $q' = \langle a \rangle \perp q_2$. Since $q_1 \perp -q_2 \in I^3 k$ is rank 8, we can proceed as in the rank 4 case and find a field extension F of k of degree at most 2 such that $(q_1 \perp -q_2)_F$ is hyperbolic and $(q_1)_F$ is isotropic. By the Arason–Pfister Hauptsatz, $(q \perp -q')_F$ is hyperbolic and thus $q_F \cong q'_F$ and $\pi(\lambda_F) = q'_F = 1 \in H^1(F, O^+(q))$. Thus λ_F is in the image of $i : H^1(F, \mu_2) \to H^1(F, \mathrm{Spin}(q))$. However, $(q_1)_F$ being isotropic, q_F is isotropic and $\mathrm{sn} : O^+(q)(F) \to H^1(F, \mu_2)$ is onto. Therefore, i is the zero map and $\lambda_F = 1$. □

Theorem 2.4. *The answer to Totaro's question is yes if k is of characteristic different from* 2 *and* $G = \mathrm{Spin}(q)$ *for q a quadratic form of dimension* ≤ 5.

Proof. Let $\{L_i\}_{1 \leq i \leq m}$ be a set of finite field extensions of k and suppose $\lambda \in H^1(k, \mathrm{Spin}(q))$ is an element such that $\lambda_{L_i} = 1$ for all i. Let $d = \gcd\{[L_i : k]\}$. We want to find F/k separable such that $\lambda_F = 1$ and $[F : k]$ divides d. If d is odd we are done by [Black 2011b, Theorem 3.7] and can take $F = k$. If d is even, by Proposition 2.3, there is a separable extension F/k of degree at most 2 such that $\lambda_F = 1$. □

Theorem 2.5. *The answer to Totaro's question is yes if k is of characteristic different from* 2 *and* $G = \mathrm{Sp}(A, \sigma)$ *where A is a central simple algebra with symplectic involution and* $\deg(A)$ *is* 2 *or* 4.

Proof. Let q be a quadratic form of dimension 3 (resp. 5) with trivial discriminant. Then the even Clifford algebra $A = C_0(V, q)$ is a central simple algebra of degree 2 (resp. 4) and the canonical involution on the Clifford algebra is symplectic and $\mathrm{Spin}(q) \cong \mathrm{Sp}(A, \sigma)$ [Knus et al. 1998, Section 15.C]. Moreover, every algebra A of degree 2 or 4 with a symplectic involution arises in this way. Thus, a positive answer to Totaro's question for $\mathrm{Sp}(A, \sigma)$ follows from Proposition 2.3. □

The case $G = \mathrm{Spin}(A, \sigma)$.

Theorem 2.6. *The answer to Totaro's question is yes if k is of characteristic different from* 2 *and* $G = \mathrm{Spin}(A, \sigma)$, *where A is a central simple algebra of degree* 4 *over k and σ is an orthogonal involution on A.*

Proof. Let $\{L_i\}_{1 \leq i \leq m}$ be a set of finite field extensions of k and suppose $\lambda \in H^1(k, \mathrm{Spin}(A, \sigma))$ is an element such that $\lambda_{L_i} = 1$ for all i. Let $d = \gcd\{[L_i : k]\}$. We will find F/k separable such that $\lambda_F = 1$ and $[F : k]$ divides d.

By [Black 2011b, Theorem 3.7], when d is odd we may take $F = k$. So we may suppose that d is even. Suppose (A, σ) has trivial discriminant. Then

$(A,\sigma) \cong (Q_1 \otimes Q_2, \tau_1 \otimes \tau_2)$ [Knus et al. 1998, Corollary 15.12], where Q_1 and Q_2 are quaternion algebras with the symplectic involution given by conjugation. In turn $\mathrm{Spin}(A,\sigma) \cong \mathrm{SL}_1(Q_1) \times \mathrm{SL}_1(Q_2)$ [Knus et al. 1998, Corollary 15.13]. There exist $\lambda_1, \lambda_2 \in k^*$ such that $\lambda = (\bar{\lambda}_1, \bar{\lambda}_2)$ with $\bar{\lambda}_i \in k^*/\operatorname{Nrd}(Q_i) \cong H^1(k, \mathrm{SL}_1(Q_i))$ for $i = 1, 2$. In the case $4 \mid d$, let F_1, F_2 be extensions of k of degree at most 2 which split Q_1 and Q_2 respectively. Then $\lambda_{F_1F_2} = 1$ and $[F_1F_2 : k]$ divides 4. In the case $2 \mid d$ and $4 \nmid d$, we can fix an L_j/k such that $[L_j : k] = 2m$, where m is odd and $\lambda_{L_j} = 1$. Following arguments as in [Garibaldi and Hoffmann 2006, Lemma 1.5] we suppose without loss of generality that $k \subseteq L \subseteq L_j$ with $[L : k]$ odd, $[L_j : L] = 2$ and $\lambda_{L_j} = 1$. Let N_{Q_1}, N_{Q_2} be the norm forms for the quaternion algebras Q_1, Q_2 respectively and let $\phi_1 = \langle 1, -\lambda_1\rangle N_{Q_1}$ and $\phi_2 = \langle 1, -\lambda_2\rangle N_{Q_2}$. The fact that $\lambda_{L_j} = 1$ implies that ϕ_1, ϕ_2 are hyperbolic over L_j. Then by [Garibaldi and Hoffmann 2006, Lemma 1.4] there exists $\mu \in k^*$ such that $\phi_1 \cong \langle 1, \mu\rangle\tilde{\phi}_1$ and $\phi_2 \cong \langle 1, \mu\rangle\tilde{\phi}_2$, where $\tilde{\phi}_1, \tilde{\phi}_2$ are 2-fold Pfister forms. Let $F = k(\sqrt{-\mu})$. Then ϕ_1, ϕ_2 are hyperbolic over F and thus $\lambda_1 \in \operatorname{Nrd}(Q_{1_F})$ and $\lambda_2 \in \operatorname{Nrd}(Q_{2_F})$. That is, $\lambda_F = 1$. Also, F/k is separable and degree at most 2 by construction.

Suppose instead that (A,σ) has nontrivial discriminant. One can associate to (A,σ) its Clifford algebra Q, which is a quaternion algebra with center $K = k(\sqrt{\delta})$, where $\delta = \operatorname{disc}(A,\sigma)$ [Knus et al. 1998, Theorem 15.7]. Then $\mathrm{Spin}(A,\sigma) = R_{K/k}\,\mathrm{SL}_1(Q)$ [Knus et al. 1998, Proposition 15.10] and $H^1(k, \mathrm{Spin}(A,\sigma)) = H^1(K, \mathrm{SL}_1(Q))$. If Q is split, $\lambda = 1$ and we take $F = k$. So suppose Q is not split. If $4 \mid d$ we can take F a splitting field of Q such that F/K is a separable extension of degree 2. Since

$$H^1(F, \mathrm{Spin}(A,\sigma)) = H^1(K \otimes F, \mathrm{SL}_1(Q)) \cong H^1(F \times F, \mathrm{SL}_1(Q)) = \{1\},$$

we obtain $\lambda_F = 1$. Further $[F : k] = 4$, and since F/K and K/k are separable, F/k is separable. We are left to consider the case where (A,σ) has nontrivial discriminant and $4 \nmid d$ and $2 \mid d$.

Consider the short exact sequence

$$1 \to R_{K/k}\,\mathrm{SL}_1(Q) \to R_{K/k}\,\mathrm{GL}_1(Q) \to R_{K/k}G_m \to 1,$$

which induces

$$\mathrm{GL}_1(Q)(K) \xrightarrow{\mathrm{Nrd}} K^* \longrightarrow H^1(K, \mathrm{SL}_1(Q)) \longrightarrow 1.$$

Choose $\lambda \in H^1(K, \mathrm{SL}_1(Q))$ such that $\lambda_{L_i} = 1$ for all i and let $\beta \in K^*$ satisfy $\delta(\beta) = \lambda$. Following [Garibaldi and Hoffmann 2006, Lemma 1.5], we may suppose that $\lambda_{L_j} = 1$ where $k \subseteq L \subseteq L_j$ and $[L_j : L] = 2$.

$$
\begin{tikzcd}
\mathrm{GL}_1(Q)(K) \arrow[r, "\mathrm{Nrd}"] \arrow[d] & K^* \arrow[r] \arrow[d] & H^1(K, \mathrm{SL}_1(Q)) \arrow[r] \arrow[d] & 1 \\
\mathrm{GL}_1(Q)(K \otimes_k L_j) \arrow[r, "\mathrm{Nrd}"] & (K \otimes_k L_j)^* \arrow[r] & H^1(K \otimes_k L_j, \mathrm{SL}_1(Q)) \arrow[r] & 1
\end{tikzcd} \tag{2.6.1}
$$

Write $L_j = L(\sqrt{a})$ for $a \in L^*/L^{*2}$. Let f be the norm form on Q and let f^0 denote the norm form restricted to the traceless elements of Q, which we denote by Q^0. Since $\lambda_{L_j} = 0$, choose $x_0, y_0 \in Q \otimes L$ such that

$$\beta = f(x_0 + y_0\sqrt{a}). \tag{2.6.2}$$

If $y_0 = 0$ we have $\beta \in \mathrm{Nrd}(Q \otimes L)$, and, L/K being of odd degree, this implies $\beta \in \mathrm{Nrd}(Q)$. We take $F = k$. So suppose $y_0 \neq 0$. Since Q is a division algebra, $f(y_0) \neq 0$ and

$$\beta = f(x_0) + af(y_0). \tag{2.6.3}$$

If we let b_f denote the adjoint bilinear form, we have

$$b_f(x_0, y_0) = 0 \tag{2.6.4}$$

and

$$\beta f(y_0^{-1}) = f(x_0 y_0^{-1}) + a, \tag{2.6.5}$$

where the reduced trace $\mathrm{trd}(x_0 y_0^{-1})$ vanishes by (2.6.4). Therefore,

$$\beta f(y_0^{-1}) = f^0(x_0 y_0^{-1}) + a. \tag{2.6.6}$$

Let $f = f_1 + \sqrt{\delta} f_2$ with f_1 and f_2 quadratic forms on Q with values in k. Further let $f^0 = f_1^0 + \sqrt{\delta} f_2^0$ where f_1^0, f_2^0 are quadratic forms on Q^0 with values in k. Setting $z_0 = y_0^{-1}$ and $w_0 = x_0 y_0^{-1}$, we have

$$a = \beta_1 f_1(z_0) + \beta_2 \delta f_2(z_0) - f_1^0(w_0), \tag{2.6.7}$$

$$0 = \beta_1 f_2(z_0) + \beta_2 f_1(z_0) - f_2^0(w_0), \tag{2.6.8}$$

with $z_0 \in Q \otimes L$ and $w_0 \in Q^0 \otimes L$. Define k-quadratic forms $q_1 : Q \oplus Q^0 \to k$ and $q_2 : Q \oplus Q^0 \to k$ by

$$q_1(z, w) = \beta_1 f_1(z) + \beta_2 \delta f_2(z) - f_1^0(w), \tag{2.6.9}$$

$$q_2(z, w) = \beta_1 f_2(z) + \beta_2 f_1(z) - f_2^0(w), \tag{2.6.10}$$

for $z \in Q$ and $w \in Q_0$. Since $y_0 \neq 0$, $z_0 = y_0^{-1} \neq 0$ and (z_0, w_0) is a nontrivial zero of q_2 over L. Then by Springer's theorem [1952], q_2 has a nontrivial zero (z_1, w_1)

over k. By a general position argument, we may assume that $z_1 \neq 0$. Let

$$\alpha = \beta_1 f_1(z_1) + \beta_2 \delta f_2(z_1) - f_1^0(w_1). \tag{2.6.11}$$

We have

$$0 = \beta_1 f_2(z_0) + \beta_2 f_1(z_0) - f_2^0(w_1). \tag{2.6.12}$$

Adding these two equations, we find

$$\alpha = \beta f(z_1) - f^0(w_1), \tag{2.6.13}$$

or, equivalently,

$$\beta f(z_1) = \alpha + f^0(w_1). \tag{2.6.14}$$

Let $F = k(\sqrt{\alpha})$. Then $[F:k] \le 2$, $(\sqrt{\alpha} + w_1)z_1^{-1} \in Q_F$ and $\beta = \mathrm{Nrd}((\sqrt{\alpha} + w_1)z_1^{-1})$. Thus, $\lambda_F = 1$. □

Theorem 2.7. *The answer to Totaro's question is yes if k is of characteristic different from* 2 *and $G = \mathrm{SU}(A, \sigma)$ where A is a quaternion algebra with unitary involution σ.*

Proof. The norm algebra $N_{K/k}(A, \sigma)$ equals (B, τ) for B a central simple algebra of degree 4 and τ an orthogonal involution on B. Since $\mathrm{Spin}(B, \tau) \cong \mathrm{SU}(A, \sigma)$, that Totaro's question has an affirmative answer in this case is a consequence of Theorem 2.6. □

3. Conclusion

Theorem 3.1. *The answer to Totaro's question is yes for k a field of characteristic different from* 2 *and G a simply connected*, *semisimple*, *classical group of rank* ≤ 2.

Proof. We suppose in all cases that G is simply connected and semisimple and that the rank of $G_{\bar{k}} \le 2$. If G is of type ${}^1A_1$ or ${}^1A_2$ then G is of the form $\mathrm{SL}_1(A)$ for A a central simple algebra of degree 2 or 3 [Knus et al. 1998, Theorem 26.9]. A positive answer to Totaro's question for a group of this form was shown in Theorem 2.1. If G is of type ${}^2A_1$ then $G = \mathrm{SU}(A, \sigma)$ for A a central simple algebra of degree 2 with unitary involution σ. The proof for this case was given in Theorem 2.7. If G is of type ${}^2A_2$ then G is of the form $\mathrm{SU}(A, \sigma)$, where A is a central simple algebra of degree 3 with unitary involution σ [Knus et al. 1998, Theorem 26.9]. Thus an affirmative answer to Totaro's question for a group of type ${}^2A_2$ follows from Theorem 2.2 above. If G is of type B_1 or B_2, then $G = \mathrm{Spin}(q)$ for q a quadratic form of dimension 3 or 5 [Knus et al. 1998, Theorem 26.12] and the desired result was proven in Theorem 2.4. If G is of type C_1 or C_2, then $G = \mathrm{Sp}(A, \sigma)$, where A is a central simple algebra of degree 2 or 4 and σ is a symplectic involution on A. The proof of our result in this case was covered in Theorem 2.5. If G is of type D_2 then either $G = \mathrm{Spin}(q)$ for q a quadratic form of dimension 2 or 4 or G is of the form $\mathrm{Spin}(A, \sigma)$ for A a central simple algebra over k of degree 4 and σ an orthogonal involution on A [Knus et al. 1998, Theorem 26.15]. In the first case the desired results follows from Theorem 2.4 and in the latter it follows from Theorem 2.6. □

Remark 3.2. Since Garibaldi and Hoffman [2006] have given a proof in the case G is of type G_2, Totaro's question has a positive answer for any simply connected, semisimple group of rank ≤ 2.

References

[Albert 1963] A. A. Albert, "On involutorial associative division algebras", *Scripta Math.* **26** (1963), 309–316. MR 31 #3451 Zbl 0147.28702

[Baeza 1978] R. Baeza, *Quadratic forms over semilocal rings*, Lecture Notes in Mathematics **655**, Springer, Berlin, 1978. MR 58 #10972 Zbl 0382.10014

[Bayer-Fluckiger and Lenstra 1990] E. Bayer-Fluckiger and H. W. Lenstra, Jr., "Forms in odd degree extensions and self-dual normal bases", *Amer. J. Math.* **112**:3 (1990), 359–373. MR 91h:11030 Zbl 0729.12006

[Black 2011a] J. Black, "Implications of the Hasse principle for zero cycles of degree one on principal homogeneous spaces", *Proc. Amer. Math. Soc.* **139**:12 (2011), 4163–4171. MR 2012g:11071 Zbl 1257.11038

[Black 2011b] J. Black, "Zero cycles of degree one on principal homogeneous spaces", *J. Algebra* **334** (2011), 232–246. MR 2012g:12008 Zbl 05990153

[Elman et al. 2008] R. Elman, N. Karpenko, and A. S. Merkurjev, *The algebraic and geometric theory of quadratic forms*, American Mathematical Society Colloquium Publications **56**, American Mathematical Society, Providence, RI, 2008. MR 2009d:11062 Zbl 1165.11042

[Garibaldi and Hoffmann 2006] S. Garibaldi and D. W. Hoffmann, "Totaro's question on zero-cycles on G_2, F_4 and E_6 torsors", *J. London Math. Soc.* (2) **73**:2 (2006), 325–338. MR 2007g:11049 Zbl 1092.11021

[Gille and Szamuely 2006] P. Gille and T. Szamuely, *Central simple algebras and Galois cohomology*, Cambridge Studies in Advanced Mathematics **101**, Cambridge University Press, Cambridge, 2006. MR 2007k:16033 Zbl 1137.12001

[Knus et al. 1998] M.-A. Knus, A. S. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications **44**, American Mathematical Society, Providence, RI, 1998. MR 2000a:16031 Zbl 0955.16001

[Lam 1980] T. Y. Lam, *The algebraic theory of quadratic forms*, Benjamin/Cummings, Reading, MA, 1980. MR 83d:10022 Zbl 0437.10006

[Merkurjev 1995] A. S. Merkurjev, "Норменный принцип для алгебраических групп", *Algebra i Analiz* **7**:2 (1995), 77–105. Translated as "The norm principle for algebraic groups", *St. Petersburg Mathematical Journal* **7**:2 (1996), 243–264. MR 96k:20088 Zbl 0859.20039

[Sansuc 1981] J.-J. Sansuc, "Groupe de Brauer et arithmétique des groupes algébriques linéaires sur un corps de nombres", *J. Reine Angew. Math.* **327** (1981), 12–80. MR 83d:12010 Zbl 0468.14007

[Serre 1995] J.-P. Serre, *Cohomologie Galoisienne: progrès et problèmes*, pp. 229–257, Astérisque **227**, Société Mathématique de France, Paris, 1995, Available at http://eudml.org/doc/110186. Séminaire Bourbaki, Vol. 1993/94, Exp. No. **783**:4. MR 97d:11063 Zbl 0837.12003

[Springer 1952] T. A. Springer, "Sur les formes quadratiques d'indice zéro", *C. R. Acad. Sci. Paris* **234** (1952), 1517–1519. MR 13,815j Zbl 0046.24303

[Tits 1992] J. Tits, "Sur les degrés des extensions de corps déployant les groupes algébriques simples", *C. R. Acad. Sci. Paris Sér. I Math.* **315**:11 (1992), 1131–1138. MR 93m:20059 Zbl 0823.20042

[Totaro 2004] B. Totaro, "Splitting fields for E_8-torsors", *Duke Math. J.* **121**:3 (2004), 425–455. MR 2005h:11081 Zbl 1048.11031

Received November 28, 2012. Revised February 19, 2013.

JODI BLACK
DEPARTMENT OF MATHEMATICS
BUCKNELL UNIVERSITY
LEWISBURG, PA 17837
UNITED STATES
jodi.black@bucknell.edu

RAMAN PARIMALA
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY
400 DOWMAN DRIVE W401
ATLANTA, GA 30322
UNITED STATES
parimala@mathcs.emory.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.269

UNIFORM HYPERBOLICITY OF THE CURVE GRAPHS

BRIAN H. BOWDITCH

We show that there is a universal constant, k, such that the curve graph associated to any compact orientable surface is k-hyperbolic. Independent proofs of this have been given by Aougab, by Hensel, Przytycki and Webb, and by Clay, Rafi and Schleimer.

1. Introduction

Let Σ be a closed orientable surface of genus g, together with a (possibly empty) finite set $\Pi \subseteq \Sigma$. Set $p = |\Pi|$. We assume that $3g + p \geq 5$. Let $\mathcal{G} = \mathcal{G}(g, p)$ be the curve graph associated to (Σ, Π); that is, the 1-skeleton of the curve complex as originally defined in [Harvey 1981]. Its vertex set, $V(\mathcal{G})$, is the set of free homotopy classes of nontrivial nonperipheral closed curves in $\Sigma \setminus \Pi$; and two such curves are deemed to be adjacent in $\mathcal{G}$ if they can be realised disjointly in $\Sigma \setminus \Pi$. These, and related, complexes are now central tools in geometric group theory and hyperbolic geometry.

In [Masur and Minsky 1999], it was shown that, for all g, p, $\mathcal{G}(g, p)$ is hyperbolic in the sense of [Gromov 1987]. In [Bowditch 2006], henceforth abbreviated [B], it was shown that the hyperbolicity constant, k, is bounded above by a function that is logarithmic in $g + p$. In fact, we show here that k can be chosen independently of g and p:

Theorem 1.1. *There is a universal constant, $k \in \mathbb{N}$, such that $\mathcal{G}(g, p)$ is k-hyperbolic for all g, p with $3g + p \geq 5$.*

We will give some estimates for k (though certainly not optimal) in Section 4.

Independent proofs of this result have been found by Aougab [2013], by Hensel, Przytycki and Webb [Hensel, Przytycki and Webb 2013], and by Clay, Rafi and Schleimer [Clay, Rafi and Schleimer 2013]. The proofs in these last two papers are combinatorial in nature, while Aougab's proof is based on broadly similar principles to those described here, though the specifics are different. Both this paper and [Aougab 2013] make use of riemannian geometry. The argument of [Hensel, Przytycki and Webb 2013] seems to give the best constants.

MSC2010: 20F65.
Keywords: uniform hyperbolicity, curve graph, surface.

Given Theorem 1.1, one can also obtain uniform bounds for the bounded geodesic image theorem of [Masur and Minsky 2000]. For this, one can combine the description of quasigeodesic lines in [B] with an unpublished argument of Leininger. In fact, a more direct approach, just using hyperbolicity, has recently been found by Webb [2013].

We remark that Theorem 1.1 does not imply uniform hyperbolicity of the curve complexes (with simplices realised as regular euclidean simplices) since their 1-skeleta are not uniformly quasi-isometrically embedded — there is an arbitrarily large contraction of distances as the complexity increases.

The proof of Theorem 1.1 consists primarily of going through the arguments of [B] with more careful bookkeeping of constants. This is accomplished in Section 2 here. In Sections 3 and 4 here, we show that much of this can be bypassed. In fact, we only really need a few results from [B], notably Lemmas 1.3, 4.4 and 4.5, together with the construction of singular euclidean structures described in Section 5 thereof.

We were motivated to look again at that paper after reading some estimates in [Tang 2013] which relate distances to intersection number.

2. Proofs

In this section, we will prove Proposition 2.6, which, together with Proposition 3.1 of [B], implies Theorem 1.1.

We will use the following different measures of the "complexity" of Σ, Π, tailored to different parts of the argument: $\xi_0 = 2g + p - 4$, $\xi_1 = 2g + p - 1$, $\xi_2 = 2g + p + 6$. For $\alpha, \beta \in V(\mathcal{G})$, we write $\iota(\alpha, \beta)$ for the intersection number, and $d(\alpha, \beta)$ for the combinatorial distance in the curve graph.

Lemma 2.1. *If $\gamma, \delta \in V(\mathcal{G})$, with $\iota(\gamma, \delta) \le \xi_0 + 1$, then $d(\gamma, \delta) \le 2$.*

Proof. We realise γ, δ in $\Sigma \setminus \Pi$ so that $|\gamma \cap \delta| = \iota(\gamma, \delta) = n$, say. Now, $\gamma \cup \delta$ is a graph with n vertices and $2n$ edges, and hence Euler characteristic $-n$. If $d(\gamma, \delta) > 2$, then $\gamma \cup \delta$ fills $\Sigma \setminus \Pi$ and so this Euler characteristic must be at most that of $\Sigma \setminus \Pi$, namely, $2 - 2g - p$. Thus $n \ge 2g + p - 2$. Taking the contrapositive, if $n \le \xi_0 + 1 = 2g + p - 3$, then $d(\gamma, \delta) \le 2$. ☐

Now, Lemma 1.3 of [B] shows that if $\alpha, \beta \in V(\mathcal{G})$ with $2\iota(\alpha, \beta) \le ab$ for $a, b \in \mathbb{N}$, then there is some $\gamma \in V(\mathcal{G})$ with $\iota(\alpha, \gamma) \le a$ and $\iota(\beta, \gamma) \le b$. Applying this q times, together with Lemma 2.1, we get:

Corollary 2.2. *If $q \in \mathbb{N}$ and $\alpha, \beta \in V(\mathcal{G})$ with $2^q \iota(\alpha, \beta) \le \xi_0^{q+1}$, then $d(\alpha, \beta) \le 2(q+1)$.*

Definition. By a *region* in Σ, we mean a subsurface, $H \subseteq \Sigma$, with $\partial H \cap \Pi = \varnothing$. A region is *trivial* if it is a topological disc containing at most one point of Π. An *annulus* in Σ is a region $A \subseteq \Sigma \setminus \Pi$ homeomorphic to $S^1 \times [0, 1]$ such that no component of $\Sigma \setminus A$ is trivial.

The core curve of an annulus therefore determines an element of $V(\mathcal{G})$.

Suppose that ρ is a riemannian metric on Σ. We allow for a finite number of cone singularities (which need bear no relation to Π). We define the *width* of an annulus $A \subseteq \Sigma$ to be the length of a shortest path in A connecting its two boundary components.

The following lemma is a slight variation of Lemma 5.1 of [B]. We follow a similar argument, but taking more care with constants.

The proof will make use of the following notion. Let α be an essential nonperipheral closed curve in $\Sigma \setminus \Pi$.

Definition. A *bridge* (across α) is an arc, $\delta \subseteq \Sigma \setminus \Pi$, with $\partial\delta = \delta \cap \alpha$ such that no component of $\Sigma \setminus (\alpha \cup \delta)$ is a disc not meeting Π.

In other words, $\alpha \cup \delta$ is an embedded π_1-injective theta-curve in $\Sigma \setminus \Pi$, i.e., it is the union of three arcs which meet precisely in their endpoints and are pairwise nonhomotopic relative to their endpoints.

Lemma 2.3. *Let ρ be a (singular) riemannian metric on Σ with* $\mathrm{area}(\Sigma) = 1$. *Suppose that $3g + p \geq 5$. Suppose that there is a constant $h > 0$ such that for any trivial region $\Delta \subseteq \Sigma$ we have* $\mathrm{area}(\Delta) \leq h(\mathrm{length}(\partial\Delta))^2$. *Then Σ contains an annulus of width at least $\eta = \frac{1}{4}\xi_1\xi_2\sqrt{h}$.*

Proof. To avoid technical details obscuring the exposition, we will relax inequalities so that they are assumed to hold up to an arbitrarily small additive constant $\epsilon > 0$. Thus, for example, a "shortest" curve will be assumed to be shortest to within ϵ. This will allow us, for instance, to adjust paths so that they can be assumed to avoid Π. Finally, we can allow $\epsilon \to 0$. In what follows any "curve" in $\Sigma \setminus \Pi$ will be assumed to be essential and nonperipheral, i.e., it does not bound a trivial region in Σ.

Let $\eta_0 = 1/4\xi_2\sqrt{h}$. We claim that there are curves, $\alpha, \beta \subseteq \Sigma \setminus \Pi$ with $\rho(\alpha, \beta) \geq \eta_0$. Given this, we let $\phi : \Sigma \to [0, \eta_0] = [0, \xi_1\eta]$ be a 1-lipschitz map with $\alpha \subseteq \phi^{-1}(0)$ and $\beta \subseteq \phi^{-1}(\xi_1\eta)$. Given any $i \in \{1, \dots, \xi_1 - 1\}$, we can find a multicurve, $\gamma_i \subseteq \phi^{-1}(i\eta)$, which separates Σ into exactly two components, S_i^α, S_i^β, containing α and β respectively. We can assume $\gamma_i \cap \Pi = \emptyset$, and that $S_i^\alpha \subseteq S_{i+1}^\alpha$ for all i. These multicurves cut Σ into ξ_1 regions $M_i = S_i^\alpha \cap S_{i-1}^\beta$ (where $M_0 = S_1^\alpha$ and $M_{\xi_1} = S_{\xi_1-1}^\beta$). At least one of these must have a component which is an annulus (otherwise each $M_i \setminus \Pi$ would have negative Euler characteristic, giving the contradiction that the Euler characteristic of $\Sigma \setminus \Pi$ is at most $-\xi_1 < 2 - 2g - p$). This annulus must have width at least η as required.

To find α and β, we take α to be a shortest curve in $\Sigma \setminus \Pi$. We suppose, for contradiction, that if $\beta \subseteq \Sigma \setminus \Pi$ is any curve, then $\rho(\alpha, \beta) < \eta_0$. Let $\lambda = 2\eta_0$.

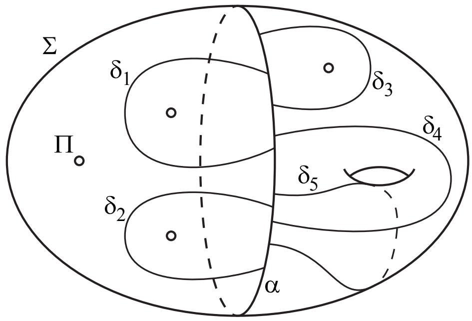


Figure 1. Example of a curve with bridges, $(g, p) = (1, 4)$.

We first claim that there is a collection of disjoint bridges, $\delta_1, \ldots, \delta_n$, across α with length$(\delta_i) < \lambda$ for all i and with each component of $\Sigma \setminus (\alpha \cup \delta_1 \cup \cdots \cup \delta_n)$ trivial. (An example is shown in Figure 1.)

To prove this claim, let $N(\alpha, t)$ be the metric t-neighbourhood of α in Σ. Let $G(t)$ be the image of $\pi_1(N(\alpha, t) \setminus \Pi)$ in $\pi_1(\Sigma \setminus \Pi)$. Note that $G(0)$ is infinite cyclic, and $G(\eta_0) = \pi_1(\Sigma \setminus \Pi)$. As t increases from 0 to η_0, $G(t)$ gets bigger at certain critical times, $t_1, \ldots, t_n$. At these times, we can suppose we have added another generator, which we can represent as a bridge, δ_i, of length at most $2t_i < 2\eta_0 = \lambda$. Thus, inductively, $G(t_i)$ is supported on $\alpha \cup \delta_1 \cup \cdots \cup \delta_i$. It follows that $\alpha \cup \delta_1 \cup \cdots \cup \delta_n$ must fill $\Sigma \setminus \Pi$ (that is, carries all of $\pi_1(\Sigma \setminus \Pi)$), otherwise we could find a curve, β, with $\rho(\alpha, \beta) \geq \eta_0$. This gives us our collection of bridges as claimed.

Let $l = \text{length}(\alpha)$. We now claim that $l \leq 6\lambda$. So, suppose, to the contrary, that $l > 6\lambda$.

Given any i, write $\alpha = \alpha_i \cup \alpha'_i$, where α_i and α'_i are respectively the shorter and longer arcs with endpoints at $\partial \delta_i$. Thus

$$\text{length}(\alpha_i) \leq l/2 \quad \text{and} \quad \text{length}(\alpha_i \cup \delta_i) \leq l/2 + \lambda < l.$$

By minimality of α, $\alpha_i \cup \delta_i$ must be trivial or peripheral, i.e., it bounds a trivial region in Σ. This region must be a disc containing exactly one point of Π. Since this is true of all bridges δ_i, we already get a contradiction if $g > 0$ (and we can deduce that $l \leq 3\lambda$ in this case). So we can assume that $g = 0$, and so α cuts Σ into two discs, H_0 and H_1. We have $|\Pi \cap H_i| \geq 2$, and we can assume that $|\Pi \cap H_0| \geq 3$.

Note also that, if $\alpha'_i \cup \delta_i$ is nontrivial, then $\text{length}(\alpha'_i \cup \delta_i) \geq \text{length}(\alpha)$ and so $\text{length}(\alpha_i) \leq \text{length}(\delta_i) < \lambda$.

Now H_0 must contain at least two bridges from our collection. We can assume these are δ_1 and δ_2. Recall that $\delta_1 \cap \delta_2 = \varnothing$. From the above, it follows that $\text{length}(\alpha_1) < \lambda$ and $\text{length}(\alpha_2) < \lambda$. Since δ_1 and δ_2 cannot cross, we must have $\alpha_1 \cap \alpha_2 = \varnothing$.

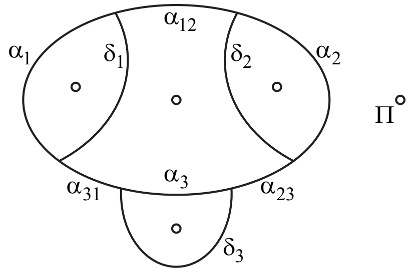


Figure 2. Picture of three bridges, $(g, p) = (0, 5)$.

Now let δ_3 be a bridge in H_1. As before, $\text{length}(\alpha_3) \le l/2$, and so for $i = 1, 2$, $\text{length}(\alpha_i \cup \alpha_3 \cup \delta_i \cup \delta_3) \le 3\lambda + l/2$. Now $\alpha_1 \cap \alpha_3 = \varnothing$ (otherwise $\alpha_1 \cup \alpha_3 \cup \delta_1 \cup \delta_3$ would contain a curve of length at most $3\lambda + l/2 < l$). Similarly, $\alpha_2 \cap \alpha_3 = \varnothing$. Now, given $i, j \in \{1, 2, 3\}$, let α_{ij} be the component of $\alpha \setminus (\alpha_1 \cup \alpha_2 \cup \alpha_3)$ between α_i and α_j (see Figure 2). Let θ_{ij} be the curve in Σ with image $\alpha_{ij} \cup \alpha_i \cup \alpha_j \cup \delta_i \cup \delta_j$, which passes through α_{ij} exactly twice. Together, the curves θ_{12}, θ_{23} and θ_{31} pass twice through each edge of $\alpha \cup \delta_1 \cup \delta_2 \cup \delta_3$, and so their lengths sum to at most $2l + 6\lambda$. We arrive at the contradiction that the length of at least one of the θ_{ij} is at most $\frac{1}{3}(2l + 6\lambda) < l$.

This shows that $l \le 6\lambda$ as claimed.

After removing some of the bridges if necessary, we can assume that at most two of the complementary components are discs not meeting Π, and so $n \le 2g + p$. Let $\sigma = \alpha \cup \delta_1 \cup \cdots \cup \delta_n$. Thus $\text{length}(\sigma) < 6\lambda + n\lambda = (n+6)\lambda \le (2g + p + 6)\lambda = \xi_2\lambda$.

Since each component of $\Sigma \setminus \sigma$ is trivial, we must have $\text{area}(\Sigma) \le h(2\,\text{length}\,\sigma)^2$ (the worst case being when $\Sigma \setminus \sigma$ is connected). But we have assumed that $\text{area}(\Sigma) = 1$ and so $1 < h(2\xi_2\lambda)^2$. Now, $\lambda = 2\eta_0 = 2(1/4\xi_2\sqrt{h}) = 1/2\xi_2\sqrt{h}$, so we arrive at the contradiction that $1 < 1$.

This shows that there must be a curve, β, in $\Sigma \setminus \Pi$ with $\rho(\alpha, \beta) \ge \eta_0$ as claimed. □

In fact, the argument also applies if $(g, p) = (1, 1)$. If $(g, p) = (0, 4)$, we will only need to consider a special case, namely, the quotient of a euclidean torus by an involution with four fixed points. In that case, we can set $\eta = 1/2$.

We will now set $h = 1/2\pi$. This gives $\eta = 1/4\xi_1\xi_2\sqrt{1/2\pi} = \sqrt{2\pi}/4\xi_1\xi_2$. As in Section 5 of [B], we define $R = \sqrt{2}/\eta$. In this case therefore, $R = (4/\sqrt{\pi})\xi_1\xi_2$.

Now suppose that α, β are weighted multicurves in the sense defined in [B]. (In other words, each is a measured lamination whose support is a disjoint union of curves.)

Definition. The *weighted intersection number*, $\iota(\alpha, \beta)$, of α and β is the sum $\sum_{i,j} \lambda_i \lambda_j \iota(\alpha_i, \beta_j)$, where α_i and β_j vary over the components of the support of α and β, where λ_i and λ_j are the respective weighting on them, and where $\iota(\alpha_i, \beta_j) \in \mathbb{N}$ is the usual geometric intersection number.

We write $d(\alpha, \beta) = \min_{i,j}\{d(\alpha_i, \beta_j)\}$, again where α_i and β_j vary over the components of α, β.

Given $\gamma \in V(\mathcal{G})$ we set $l(\gamma) = l_{\alpha\beta}(\gamma) = \max\{\iota(\alpha, \gamma), \iota(\beta, \gamma)\}$ (interpreting γ as a one-component multicurve of unit weight). One can think of $l(\gamma)$ as describing a "length" in a singular euclidean structure arising from α and β (see Section 5 of [B]).

Lemma 2.4. *Suppose that α, β are weighted multicurves with $\iota(\alpha, \beta) = 1$ and $d(\alpha, \beta) \geq 2$. Then there is some $\delta \in V(\mathcal{G})$ with $l(\delta) \leq R$ such that $\iota(\gamma, \delta) \leq Rl(\gamma)$ for all $\gamma \in V(\mathcal{G})$ (where R is defined as above).*

Proof. This is just a restating of Lemma 4.1 of [B] for this particular definition of R. The proof is the same. Suppose first that $\alpha \cup \beta$ fills $\Sigma \setminus \Pi$. As in Section 5 of that paper, we construct a singular euclidean surface, tiled by rectangles, dual to $\alpha \cup \beta$. The cone angles are all multiples of π, and all cone singularities of angle π lie in Π. Thus, any trivial region, $\Delta \subseteq \Pi$, contains at most one cone point of angle less than 2π. Passing to a branched double cover over this cone point (if it exists) we are reduced to considering the case where all cone angles are at least 2π. But then the worst case is a round circle in the euclidean plane [Weil 1926] which would give $\text{area}(\Delta) = \text{length}(\partial\Delta)^2/4\pi$. We can therefore set $h = 2(1/4\pi) = 1/2\pi$. Now apply Lemma 2.3, and set δ to be a core curve of that annulus. The statement then follows exactly as in [B] (at the end of Section 5 thereof). (In [B], h was given inaccurately as $\pi/2$.)

If $\alpha \cup \beta$ does not fill $\Sigma \setminus \Pi$, we get instead a singular euclidean structure on a "smaller" surface, namely a region of Σ with each boundary component collapsed to a point. However, this process can only decrease ξ_1 and ξ_2, so we again get an annulus of width at least η. (This case is the reason we needed a version of Lemma 2.3 when $3g + p = 4$. In the case where $(g, p) = (0, 4)$, note that $1/2$ is certainly greater than the required $\sqrt{2\pi}/120$.) ☐

Given $r \geq 0$, set $L(\alpha, \beta, r) = \{\gamma \in V(\mathcal{G}) \mid l(\gamma) \leq r\}$. Note that the curve δ given by Lemma 2.4 lies in $L(\alpha, \beta, R)$.

Lemma 2.5. *Suppose that $2g + p \geq 195$. Suppose that α, β are weighted multicurves with $\iota(\alpha, \beta) = 1$ and $d(\alpha, \beta) \geq 2$. Then, the diameter of $L(\alpha, \beta, 2R)$ in $\mathcal{G}$ is at most 20.*

Proof. Let δ be as given by Lemma 2.4. If $\gamma \in L(\alpha, \beta, 2R)$, then $l(\gamma) \leq 2R$, so $\iota(\gamma, \delta) \leq 2R^2$. If we knew that $16\,\iota(\gamma, \delta) \leq \xi_0^5$, then Corollary 2.2 with $q = 4$ would give $d(\gamma, \delta) \leq 10$ and the result would follow.

It is therefore sufficient that $16(2R^2) \leq \xi_0^5$. Recall that $R = (4/\sqrt{\pi})\xi_1\xi_2$, so this reduces to $32(4/\sqrt{\pi})^2\xi_1^2\xi_2^2 \leq \xi_0^5$, that is, $512\xi_1^2\xi_2^2 \leq \pi\xi_0^5$. In other words, we want

$$(*) \qquad 512(2g + p - 1)^2(2g + p + 6)^2 \leq \pi(2g + p - 4)^5,$$

which holds whenever $2g + p \geq 195$. □

We now assume that $2g + p \geq 195$.

Recall that Lemma 4.3 of [B] states that $L(\alpha, \beta, R)$ has diameter bounded by some constant D (which there, depended on R). Since $L(\alpha, \beta, R) \subseteq L(\alpha, \beta, 2R)$, we have now verified Lemma 4.3 of [B] with $D = 20$. Recall that Lemma 4.2 of [B], more generally, placed a bound on the diameter of $L(\alpha, \beta, r)$ depending on r and R (specifically, $\operatorname{diam} L(\alpha, \beta, r) \leq 2Rr + 2$). This was used in the proof of Lemma 4.12 [B]. We can now use Lemma 2.5 above, in place of Lemma 4.2 of [B], to give a proof of Lemma 4.12 of [B] with the constant $4D$ now replaced by 40. We can now proceed as in [B] to prove Lemma 4.13 and Proposition 4.11 of that paper. In fact, the improvement on Lemma 4.12 allows us, respectively, to replace the constants $14D$ by $10D$ and $18D$ by $14D$, where $D = 20$. Thus, the original diameter bound of $18D$ of Proposition 4.11 of [B] now becomes 280.

Recall that Proposition 3.1 of [B] gives a criterion for hyperbolicity depending on a constant, K, in the hypotheses. The three clauses (1), (2), and (3) of those hypotheses were verified respectively by Lemma 4.10, Proposition 4.11 and Lemma 4.9. These respectively gave K bounded by $4D$, $18D$, and $2D$, which we can now replace by 80, 280 and 40. In particular, we have shown:

Proposition 2.6. *If $2g + p \geq 195$, then the curve graph $\mathscr{G}(g, p)$ satisfies the hypotheses of Proposition 3.1 of [B] with $K = 280$.*

For $2g + p \geq 195$, one can now explicitly estimate k from the proof of Proposition 3.1 of [B]. In fact, one can do better.

3. A criterion for hyperbolicity

We give a self-contained account of a criterion for hyperbolicity which is related to, but simpler than, that used in [B]. In particular, it does not require the condition on moving centres (clause (2) of Proposition 3.1 of [B]) which complicated the argument there. Essentially the same statement can be found in Section 3.13 of [Masur and Schleimer 2013], though without a specific estimate for the hyperbolicity constant arising (or the final clause about Hausdorff distance). Our proof uses an idea to be found in [Gilman 2002], but bypasses use of the isoperimetric inequality. Since this criterion has many applications, this may be of some independent interest. For definiteness, we say that a space is *k-hyperbolic* if, in every geodesic triangle, each side lies in a k-neighbourhood of the union of the other two.

Proposition 3.1. *Given $h \geq 0$, there is some $k \geq 0$ with the following property. Suppose that G is a connected graph, and that for each $x, y \in V(G)$, we have associated a connected subgraph, $\mathscr{L}(x, y) \subseteq G$, with $x, y \in \mathscr{L}(x, y)$. Suppose that:*

(1) *For all* $x, y, z \in V(G)$,

$$\mathcal{L}(x, y) \subseteq N(\mathcal{L}(x, z) \cup \mathcal{L}(z, y), h).$$

(2) *For any* $x, y \in V(G)$ *with* $d(x, y) \leq 1$, *the diameter of* $\mathcal{L}(x, y)$ *in* G *is at most* h.

Then G *is* k*-hyperbolic. In fact, we can take any* $k \geq (3m - 10h)/2$, *where* m *is any positive real number satisfying* $2h(6 + \log_2(m + 2)) \leq m$. *Moreover, for all* $x, y \in V(G)$, *the Hausdorff distance between* $\mathcal{L}(x, y)$ *and any geodesic from* x *to* y *is bounded above by* $m - 4h$.

Here, d is the combinatorial metric on G, and $N(\,\cdot\,, h)$ denotes h-neighbourhood. Note that we can assume that $\mathcal{L}(x, y) = \mathcal{L}(y, x)$ (on replacing $\mathcal{L}(x, y)$ with $\mathcal{L}(x, y) \cup \mathcal{L}(y, x)$). Note that the condition on m is monotonic: if it holds for m, it holds strictly for any $m' > m$.

Proof. Given any $x, y \in V(G)$, let $\mathcal{I}(x, y)$ be the set of all geodesics from x to y. Given any $n \in \mathbb{N}$, write

$$f(n) = \max\{d(w, \alpha) \mid (\exists x, y \in V(G))\, d(x, y) \leq n,\ \alpha \in \mathcal{I}(x, y),\ w \in \mathcal{L}(x, y)\}.$$

In other words, $f(n)$ is the minimal $f \geq 0$ such that $\mathcal{L}(x, y) \subseteq N(\alpha, f)$ for any geodesic, α, connecting any two vertices x, y at most n apart.

We first claim that $f(n) \leq (2 + [\log_2 n])h$ (compare [Gilman 2002]). To see this, write $l = d(x, y) \leq n$ and $p = [\log_2 l] + 2$. Let $z \in V(G)$ be a "near midpoint" of α; that is, it cuts α into two subpaths, α^- and α^+, whose lengths differ by at most 1. By (1), $\mathcal{L}(x, y) \subseteq N(\mathcal{L}(x, z) \cup \mathcal{L}(z, y), h)$. We now choose near midpoints of each of the paths α^+ and α^- and then continue inductively. After at most $p - 1$ steps, we see that $\mathcal{L}(x, y) \subseteq N\big(\bigcup_{i=0}^{l-1} \mathcal{L}(x_i, x_{i+1}), (p-1)h\big)$ where $x = x_0, x_1, \ldots, x_l = y$ is the sequence of vertices along α. Applying (2) now gives $\mathcal{L}(x, y) \subseteq N(\alpha, ph)$, and so $f(n) \leq ph$ as claimed.

In fact, we aim to show that $f(n)$ is bounded purely in terms of h. We proceed as follows.

Let $t = f(n) + 2h + 1$. Choose any $w \in \mathcal{L}(x, y)$. Let $l_0 = \max\{0, d(w, x) - t\}$ and $l_1 = \max\{0, d(w, y) - t\}$. Since $l = d(x, y)$, we have $l \leq l_0 + l_1 + 2t$, and so we can find vertices x', y' in α cutting it into subpaths $\alpha = \alpha_0 \cup \delta \cup \alpha_1$, where $d(x, x') \leq l_0$, $d(x', y') \leq 2t$, and $d(y', y) \leq l_1$. If $x = x'$ we leave out α_0, and/or if $y = y'$ we leave out α_1. (We can always assume that $x' \neq y'$.)

Note that $d(w, \alpha_0) \geq d(w, x) - d(x, x') \geq d(w, x) - l_0$. Therefore, if $x \neq x'$, then $l_0 = d(w, x) - t$, and so $d(w, \alpha_0) \geq t$. But $d(x, x') \leq d(x, y) \leq n$ and so $\mathcal{L}(x, x') \subseteq N(\alpha_0, f(n))$. It follows that $d(w, \mathcal{L}(x, x')) \geq t - f(n) = 2h + 1$. In other words, if $x \neq x'$, then $d(w, \mathcal{L}(x, x')) \geq 2h + 1$. Similarly, if $y \neq y'$, then $d(w, \mathcal{L}(y', y)) \geq 2h + 1$. But

$$w \in \mathcal{L}(x, y) \subseteq N(\mathcal{L}(x, x') \cup \mathcal{L}(x', y') \cup \mathcal{L}(y', y), 2h)$$

and so $d(w, \mathscr{L}(x', y')) \leq 2h$. Now $d(x', y') \leq 2t$ and so $\mathscr{L}(x', y') \subseteq N(\delta, f(2t))$. Thus, $w \in N(\delta, f(2t)+2h) \subseteq N(\alpha, f(2t)+2h)$. Since w was an arbitrary point of $\mathscr{L}(x, y)$, it follows that

$$f(n) \leq f(2t)+2h = f(2f(n)+4h+2)+2h.$$

Writing $F(n) = 2f(n)+4h+2$, we have shown that $F(n) \leq F(F(n))+4h$ for all n.

Now, from the earlier claim,

$$F(n) \leq 2((2+\log_2 n)h)+4h+2 = 2h(4+\log_2 n)+2.$$

Suppose m is as in the statement of the theorem. Writing $r = m+2$, we have $2h(6+\log r)+2 \leq r$, and so $F(n)+4h \leq 2h(6+\log_2 n)+2 < n$ for any $n > r$.

In summary, we have shown that

$$F(n) \leq F(F(n))+4h$$

for all n, and that

$$F(n)+4h < n$$

for all $n > r$. It follows that $F(n) \leq r$ for all n (otherwise, we have the contradiction $F(n) \leq F(F(n))+4h < F(n)$). It now follows that $f(n) \leq s$, where $s = (r/2)-2h-1 = (m/2)-2h$.

We have shown that for all $x, y \in V(G)$ and $\alpha \in \mathscr{I}(x, y)$, we have $\mathscr{L}(x, y) \subseteq N(\alpha, s)$. It now follows also that $\alpha \subseteq N(\mathscr{L}(x, y), 2s)$. Since if $w \in \alpha$, then w cuts α into two subpaths, α^- and α^+. Since $\mathscr{L}(x, y)$ is connected and contains x, y, we can find some $v \in \mathscr{L}(x, y)$ and $v^\pm \in \alpha^\pm$ with $d(v, v^\pm) \leq s$. Now $d(w, \{v^-, v^+\}) \leq s$, so $d(v, w) \leq 2s$. We deduce that $d(w, \mathscr{L}(x, y)) \leq 2s$ as required.

Now suppose that $x, y, z \in V(G)$ and that $\alpha \in \mathscr{I}(x, y)$, $\beta \in \mathscr{I}(x, z)$, and $\gamma \in \mathscr{I}(y, z)$. We have

$$\alpha \subseteq N(\mathscr{L}(x, y), 2s) \subseteq N(\mathscr{L}(x, z) \cup \mathscr{L}(z, y), 2s+h) \subseteq N(\beta \cup \gamma, k),$$

where

$$k = 3s+h \leq 3((r/2)-2h-1)+h = (3m-10h)/2.$$

Thus, G is k-hyperbolic. □

4. Estimation of constants

Given Proposition 3.1 of this paper, we can extract information more efficiently from [B], and bypass much of the proof of Theorem 1.1. Given $\alpha, \beta \in V(\mathscr{G}(g, p))$ with $d(\alpha, \beta) \geq 2$ and $t \in \mathbb{R}$, let $\Lambda_{\alpha\beta}(t) = L((e^t/\iota)\alpha, (e^{-t}/\iota)\beta, R)$, where $\iota = \iota(\alpha, \beta) > 0$.

Now, $\iota((e^t/\iota)\alpha, (e^{-t}/\iota)\beta) = 1$. Therefore if $2g+p \geq 195$, then by Lemma 2.4, $\Lambda_{\alpha\beta}(t) \neq \varnothing$. Let $\mathscr{L}(\alpha, \beta)(t)$ be the full subgraph of $\mathscr{G}$ with vertex set $\Lambda_{\alpha\beta}(t)$. It is not hard to see that $\mathscr{L}(\alpha, \beta)(t)$ is connected. (For example, the standard argument, going back to work of Lickorish, for showing that $\mathscr{G}$ itself is connected effectively does this. This involves interpolating between two curves by a series of surgery operations, see Lemma 1.3 of [B] for example. These can only decrease the intersection number with any fixed curve.) It follows easily that $\mathscr{L}(\alpha, \beta) = \bigcup_{t\in\mathbb{R}} \mathscr{L}(\alpha, \beta)(t)$ is connected. Note that the vertex set of $\mathscr{L}(\alpha, \beta)$ is the "line" $\Lambda_{\alpha\beta} = \bigcup_{t\in\mathbb{R}} \Lambda_{\alpha\beta}(t)$ as defined in [B]. Note also that $\alpha, \beta \in \Lambda_{\alpha\beta}$. If $d(\alpha, \beta) \leq 1$, we set $\Lambda_{\alpha\beta} = \{\alpha, \beta\}$, so that $\mathscr{L}(\alpha, \beta)$ is a single vertex or edge.

We can now verify that the collection $(\mathscr{L}(\alpha, \beta))_{\alpha,\beta\in V(\mathscr{G})}$ satisfies the hypotheses of Proposition 3.1 here with $h = 40$. Condition (2) is immediate. For condition (1), let $\alpha, \beta, \gamma \in V(G)$. If these three curves all pairwise intersect, then we set $\tau = \frac{1}{2}\log_e(\iota(\alpha, \beta)\iota(\alpha, \gamma)/\iota(\beta, \gamma))$. As in Lemma 4.5 of [B], we see that if $t \leq \tau$, the diameter of $\mathscr{L}(\alpha, \beta)(t) \cup \mathscr{L}(\alpha, \gamma)(t)$ is at most 40 (since we can set $D = 20$). Similarly, if $t \geq \tau$ then $\mathscr{L}(\alpha, \beta)(t) \cup \mathscr{L}(\beta, \gamma)(t)$ has diameter at most 40. Thus, $\mathscr{L}(\alpha, \beta) \subseteq N(\mathscr{L}(\alpha, \gamma) \cup \mathscr{L}(\gamma, \beta), h)$ with $h = 40$. The cases where at least two of the curves α, β, γ are disjoint follow from a slight modification of this argument, as in [B]. This now gives $m \leq 1320$ and $k \leq 1780$. This shows that if $2g + p \geq 195$, then $\mathscr{G}(p, q)$ is 1780-hyperbolic.

In fact, since we are now only using Lemma 4.3 of [B], we can replace $2R$ by R in Lemma 2.5 here, so that the requirement $16(2R^2) \leq \xi_0^5$ becomes $16R^2 \leq \xi_0^5$, and so we can replace the resulting factor of 512 in $(*)$ by 256. It is therefore sufficient that $2g + p \geq 107$. We have shown that if $2g + p \geq 107$, then $\mathscr{G}(g, p)$ is 1780-hyperbolic.

We can deal with lower complexity surfaces using larger values of q from Corollary 2.2. In general, we require that

$$2^{q+4}(2g + p - 1)^2(2g + p + 6)^2 \leq \pi(2g + p - 4)^{q+1}.$$

For example, with $q = 5$, this is satisfied for $2g + p \geq 26$. This gives

$$D = 4(q + 1) = 24, \quad h = 2D = 48, \quad m \leq 1584, \quad k \leq 2136.$$

In other words, if $2g + p \geq 26$, then $\mathscr{G}(g, p)$ is 2064-hyperbolic. Similarly (with $q = 6$), if $2g + p \geq 14$, then $\mathscr{G}(g, p)$ is 2492-hyperbolic, and so on.

For the cases where $2g + p \leq 6$, we need to revert to previous arguments. The estimates and methods in [Tang 2013] might give improvements for some of the lower complexities.

There is scope for other improvements in various directions. For the bounds on complexity for example, suppose $p = 0$. In the proof of Lemma 2.3 we don't have to worry about trivial regions, so we can easily obtain $l \leq 2\lambda$, allowing us to reset $\xi_2 = 2g + 2$. We can also reset $\xi_1 = 2g$. For Corollary 2.2, we could set $h = 1/4\pi$, further decreasing R by a factor of $\sqrt{2}$. In Lemma 1.3 of [B], we can eliminate the factor of 2 in the hypotheses, and thereby weaken those of Corollary 2.2 here to saying that $\iota(\alpha, \beta) \le x_0^q$. The fact that we have replaced $2R$ by R also gives us another factor of 2, so that our requirement, when $q = 4$, now becomes $R^2 \le \xi_0^5$. Together these now give $8(2g)^2(2g+2)^2 \le \pi(2g-4)^5$, that is, $4g^2(g+1)^2 \le \pi(g-2)^5$, which holds for $g \ge 8$. In other words, $\mathcal{G}(g, 0)$ is 1780-hyperbolic for $g \ge 8$.

We remark that in [Hensel, Przytycki and Webb 2013], it is shown that every curve graph is "17-hyperbolic" in the sense that, for every geodesic triangle, there is a vertex at a distance of no more than 17 from each of its sides. From this, one can easily derive a uniform hyperbolicity constant in the sense we have defined it.

References

[Aougab 2013] T. Aougab, "Uniform hyperbolicity of the graphs of curves", *Geom. Topol.* **17**:5 (2013), 2855–2875. MR 3190300 Zbl 1273.05050

[Bowditch 2006] B. H. Bowditch, "Intersection numbers and the hyperbolicity of the curve complex", *J. Reine Angew. Math.* **598** (2006), 105–129. MR 2009b:57034 Zbl 1119.32006

[Clay, Rafi and Schleimer 2013] M. Clay, K. Rafi, and S. Schleimer, "Uniform hyperbolicity of the curve graph via surgery sequences", preprint, 2013. arXiv 1302.5519

[Gilman 2002] R. H. Gilman, "On the definition of word hyperbolic groups", *Math. Z.* **242**:3 (2002), 529–541. MR 2004b:20062 Zbl 1047.20033

[Gromov 1987] M. Gromov, "Hyperbolic groups", pp. 75–263 in *Essays in group theory* (Berkeley, CA, 1985), edited by S. M. Gersten, Math. Sci. Res. Inst. Publ. **8**, Springer, New York, 1987. MR 89e:20070 Zbl 0634.20015

[Harvey 1981] W. J. Harvey, "Boundary structure of the modular group", pp. 245–251 in *Riemann surfaces and related topics* (Stony Brook, NY, 1978), edited by I. Kra and B. Maskit, Ann. of Math. Stud. **97**, Princeton University Press, 1981. MR 83d:32022 Zbl 0461.30036

[Hensel, Przytycki and Webb 2013] S. Hensel, P. Przytycki, and R. C. H. Webb, "1-slim triangles and uniform hyperbolicity for arc graphs and curve graphs", preprint, 2013. To appear in *J. Eur. Math. Soc.* arXiv 1301.5577

[Masur and Minsky 1999] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves, I: Hyperbolicity", *Invent. Math.* **138**:1 (1999), 103–149. MR 2000i:57027 Zbl 0941.32012

[Masur and Minsky 2000] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves, II: Hierarchical structure", *Geom. Funct. Anal.* **10**:4 (2000), 902–974. MR 2001k:57020 Zbl 0972.32011

[Masur and Schleimer 2013] H. A. Masur and S. Schleimer, "The geometry of the disk complex", *J. Amer. Math. Soc.* **26**:1 (2013), 1–62. MR 2983005 Zbl 1272.57015

[Tang 2013] R. Tang, *Covering maps and hulls in the curve complex*, thesis, University of Warwick, Coventry, 2013.

[Webb 2013] R. C. H. Webb, "Uniform bounds for bounded geodesic image theorems", preprint, 2013. To appear in *J. Reine Angew. Math.* arXiv 1301.6187

[Weil 1926] A. Weil, "Sur les surfaces à courbure négative", *C. R. Acad. Sci.* (*Paris*) **182** (1926), 1069–1071. JFM 52.0712.05

Received March 8, 2013. Revised June 24, 2013.

BRIAN H. BOWDITCH
MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.281

CONSTANT GAUSSIAN CURVATURE SURFACES IN THE 3-SPHERE VIA LOOP GROUPS

DAVID BRANDER, JUN-ICHI INOGUCHI AND SHIMPEI KOBAYASHI

In this paper we study constant positive Gauss curvature K surfaces in the 3-sphere $\mathbb{S}^3$ with $0 < K < 1$, as well as constant negative curvature surfaces. We show that the so-called *normal* Gauss map for a surface in $\mathbb{S}^3$ with Gauss curvature $K < 1$ is Lorentz harmonic with respect to the metric induced by the second fundamental form if and only if K is constant. We give a uniform loop group formulation for all such surfaces with $K \neq 0$, and use the generalized d'Alembert method to construct examples. This representation gives a natural correspondence between such surfaces with $K < 0$ and those with $0 < K < 1$.

Introduction

The study of isometric immersions from space forms into space forms is a classical and important problem of differential geometry. This subject has its origin in realizability of the hyperbolic plane geometry in Euclidean 3-space $\mathbb{E}^3$. As is well known, Hilbert [1901] proved the nonexistence of isometric immersions of the hyperbolic plane into $\mathbb{E}^3$. Analogous results hold for surfaces in the 3-sphere $\mathbb{S}^3$ and hyperbolic 3-space $\mathbb{H}^3$ as follows:

Theorem [Spivak 1979]. *There is no complete surface in $\mathbb{S}^3$ with constant curvature $K < 1$, $K \neq 0$. Moreover, the only complete immersion of constant positive curvature in $\mathbb{S}^3$ is a totally umbilic round sphere. There is no complete surface in $\mathbb{H}^3$ with constant curvature $K < -1$.*

Due to the complicated structure (nonlinearity) of the integrability condition (Gauss–Codazzi–Ricci equations) of isometric immersions between space forms, in the past decades many results on nonexistence, rather than explicit constructions of

Brander is partially supported by FNU grant *Symmetry Techniques in Differential Geometry*. Inoguchi is partially supported by Kakenhi 21546067, 24540063. Kobayashi is partially supported by Kakenhi 23740042.
MSC2010: primary 53A10; secondary 53C42, 53C43.
Keywords: constant curvature, 3-sphere, generalized Weierstrass representation, nonlinear d'Alembert formula.

examples, have been obtained. For this direction we refer the reader to the survey article [Borisenko 2001].

Another reason for the focus on nonexistence may be the presence of singularities. Surfaces in $\mathbb{S}^3$ with constant Gauss curvature $K < 1$ always have singularities, excepting the flat case $K = 0$ (in fact there exist infinitely many flat tori in $\mathbb{S}^3$ [Kitagawa 1988]). Recently, however, there has been some movement to broaden the class of surfaces to include those with singularities, and a number of interesting studies of the geometry of these; see, for example, [Saji et al. 2009].

On the other hand, one can see that under the asymptotic Chebyshev net parametrization, the Gauss–Codazzi equations of surfaces in $\mathbb{S}^3$ with constant curvature $K < 1$ ($K \neq 0$) are reduced to the sine-Gordon equation. The sine-Gordon equation also arises as the Gauss–Codazzi equation of pseudospherical surfaces in $\mathbb{E}^3$ (surfaces of constant negative curvature) and is associated to harmonic maps from a Lorentz surface into the 2-sphere.

By virtue of loop group techniques, an infinite-dimensional d'Alembert-type representation for solutions is available for surfaces associated to Lorentz harmonic maps. This d'Alembert-type representation is a special case of the generalized DPW method (described in [Brander and Dorfmeister 2009]): for the case of Riemannian harmonic maps the method reduces to an analogue of the Weierstrass representation; for the case of Lorentz harmonic maps, one obtains an analogue of the d'Alembert solution for the wave equation. More precisely, all solutions are given in terms of two functions, each of one variable only. This type of construction method can be traced back to [Kričever 1980]. An example of an application of this method is the solution in [Brander and Svensson 2013] of the geometric Cauchy problem for pseudospherical surfaces in $\mathbb{E}^3$, as well as for timelike constant mean curvature surfaces in Lorentz–Minkowski 3-space $\mathbb{L}^3$. The key ingredient is the generalized d'Alembert representation for Lorentz harmonic maps of Lorentz surfaces into semi-Riemannian symmetric spaces. See also [Dorfmeister 2008] and the references therein for more examples. One can expect that the approach can be adapted to other classes of isometric immersion problems.

These observations motivate us to establish a loop group method (generalized d'Alembert formula) for surfaces in $\mathbb{S}^3$ of constant curvature $K < 1$. We shall in fact give such a solution that covers all such surfaces with $K \neq 0$. The key point is to discover *which Gauss map* (there are several definitions for surfaces in $\mathbb{S}^3$) is the right one to make the connection with harmonic maps.

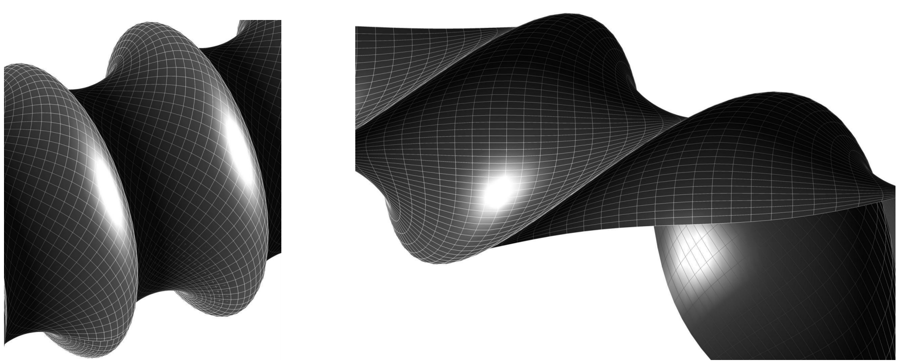
Outline of this article. This paper is organized as follows: After prerequisite knowledge in Sections 1 and 2, we will give a loop group formulation for surfaces in $\mathbb{S}^3$ of constant curvature $K < 1$ ($K \neq 0$) in Section 3. In particular, we will show that the Lorentz harmonicity (with respect to the conformal structure determined by the second fundamental form) of the *normal Gauss map* of a surface with curvature $K < 1$ is equivalent to the constancy of K. The normal Gauss map is the left translation, to the Lie algebra $\mathfrak{su}(2)$, of the unit normal n to the immersion f into $\mathbb{S}^3 = \mathrm{SU}(2)$ — in symbols, $\nu = f^{-1}n$.

The harmonicity of the normal Gauss map enables us to construct constant curvature surfaces in terms of Lorentz harmonic maps. We establish a loop group theoretic d'Alembert representation for surfaces in $\mathbb{S}^3$ with constant curvature $K < 1$ ($K \neq 0$). In Section 4 we give a relation between the surfaces in $\mathbb{S}^3$ and pseudospherical surfaces in $\mathbb{E}^3$, and show how the well-known Sym formula for the latter surfaces arises naturally from our construction. Finally, we give a detailed analysis of the limiting procedure $K \to 0$.

The paper ends with some explicit examples constructed by our method. All of the images shown here were produced using a numerical implementation of the method in Matlab. The code, *ksphere*, can be found, at the time of writing, at http://davidbrander.org/software.html.

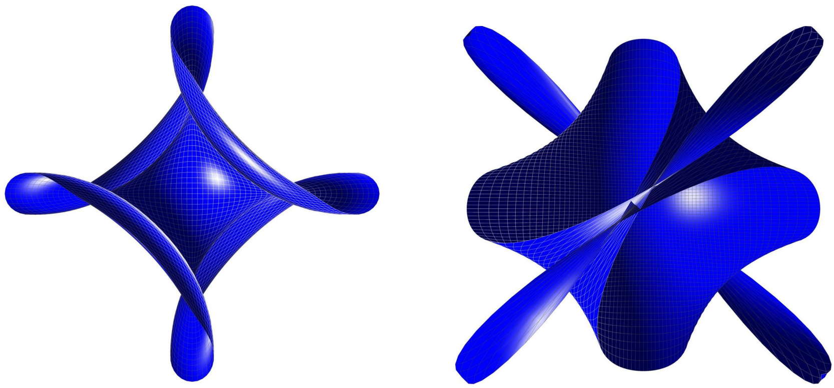
Examples. Figure 1 shows the well-known pseudospherical surface of revolution, together with a corresponding constant negative curvature surface in $\mathbb{S}^3$ obtained by a different projection from the same loop group frame. The surface in $\mathbb{S}^3$ is mapped diffeomorphically to $\mathbb{R}^3$ by the stereographic projection for rendering. See Example 4.1 below.

Figure 2 shows Amsler's pseudospherical surface in $\mathbb{E}^3$, which contains two intersecting straight lines, together with a corresponding surface of constant curvature $K = 16/25$ in $\mathbb{S}^3$ also obtained from the same loop group frame. The two straight lines correspond to two great circles. The great circles appear as straight lines in



$\mu = 1,\ K = -1$, target $\mathbb{E}^3$ $\qquad$ $\mu = 4,\ K = -\frac{16}{25}$, target $\mathbb{S}^3$

Figure 1. Left: a pseudospherical surface of revolution in $\mathbb{E}^3$. Right: a constant negative curvature analogue in $\mathbb{S}^3$. See also Figure 3.



$\mu = 1,\ K = -1$, target $\mathbb{E}^3$ $\mu = -4,\ K = 16/25$, target $\mathbb{S}^3$

Figure 2. Left: Amsler's surface in $\mathbb{E}^3$. Right: a constant positive curvature analogue in $\mathbb{S}^3$. See also Figure 4.

the image obtained by stereographic projection to $\mathbb{R}^3$. This example shows that although the singular sets in the coordinate domain are the same for every surface in the family, the type of singularity can change. The surface obtained at $\mu = -4$ apparently has a swallowtail singularity at a point where the surfaces obtained at $\mu = 1$ and $\mu = 4$ (see Example 4.2 below) each have a cuspidal edge. This suggests that the singularities of constant curvature surfaces in $\mathbb{S}^3$ are also worth investigating.

Comparison with other methods. It should be noted that Ferus and Pedit [1996] gave a very nice loop group representation for isometric immersions of space forms $M^n_c \to \widetilde{M}^{n+k}_{\tilde{c}}$ with flat normal bundle for any $c \neq \tilde{c}$ with $c \neq 0 \neq \tilde{c}$. Finite-type solutions can be generated using the modified AKS theory described in [ibid.], and all solutions can, in principle, be constructed from curved flats using the generalized DPW method described in [Brander and Dorfmeister 2009]. For the case of surfaces, as in the present article, the construction of Ferus and Pedit is quite different from the Lorentzian harmonic map approach used here. For surfaces, the Lorentzian harmonic map representation is probably more useful, since one obtains, via the generalized d'Alembert method, all solutions from essentially arbitrary pairs of functions of one variable only; this is the key, for example, to the solution of the geometric Cauchy problem in [Brander and Svensson 2013]. If one were to use the setup in [Ferus and Pedit 1996], and the generalized DPW method of [Brander and Dorfmeister 2009], which is the analogue of generalized d'Alembert method, one instead obtains a curved flat in the Grassmannian $\mathrm{SO}(4)/(\mathrm{SO}(2) \times \mathrm{SO}(2))$ as the basic data, which is not as simple. In contrast, our basic data are essentially arbitrary functions of one variable.

Another interesting difference between the two approaches is the following: we will show below that the loop group frame corresponding to a surface of constant curvature $K < 0$ in $\mathbb{S}^3$ also corresponds to a surface with $0 < K < 1$ in $\mathbb{S}^3$, giving some kind of Lawson correspondence between two surfaces, one of which has negative curvature and the other positive. This correspondence is obtained by evaluating at a different value of the loop parameter λ. On the other hand, in [Brander 2007], the loop group maps of Ferus and Pedit are also found to produce Lawson-type correspondences between various isometric immersions of space forms by evaluating in different ranges of λ. In this case, however, one does not obtain such a correspondence between surfaces with positive and negative curvature.

Finally, we should observe that Xia [2007] has also studied isometric immersions of constant curvature surfaces in space forms via loop group methods. In that work, for surfaces in $\mathbb{S}^3$, the group SO(4) is used (as opposed to $\mathrm{SU}(2) \times \mathrm{SU}(2)$, used here) and a loop group representation for the surfaces is given. However, the generalized d'Alembert method to construct solutions is not given, and neither is the equivalence of this problem with Lorentz harmonic maps via the normal Gauss map. It turns out to be difficult to find a suitable loop group decomposition in the SO(4) setup used in [ibid.], which is really the setup for Lorentz harmonic maps into the Grassmannian $\mathrm{SO}(4)/(\mathrm{SO}(2) \times \mathrm{SO}(2))$. The essential problem is that the surfaces in question are not associated to arbitrary harmonic maps in the Grassmannian, but very special ones. In contrast, our use of the group $\mathrm{SU}(2) \times \mathrm{SU}(2)$ leads naturally to the normal Gauss map, the harmonicity of which is a basic characterization of these surfaces; this leads to a straightforward solution in terms of the known method for Lorentz harmonic maps.

1. Preliminaries

The symmetric space $\mathbb{S}^3$. Let $\mathbb{E}^4$ be the *Euclidean* 4-*space* with standard inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_0 y_0 + x_1 y_1 + x_2 y_2 + x_3 y_3.$$

We denote by $e_0 = (1, 0, 0, 0)$, $e_1 = (0, 1, 0, 0)$, $e_2 = (0, 0, 1, 0)$, $e_3 = (0, 0, 0, 1)$ the natural basis of $\mathbb{E}^4$.

The *orthogonal group* O(4) is defined by

$$\mathrm{O}(4) = \{A \in \mathrm{GL}(4, \mathbb{R}) \mid A^T A = I\}.$$

Here I is the identity matrix. We denote by SO(4) the identity component of O(4) (called the *rotation group*).

Let us denote by $\mathbb{S}^3$ the unit 3-sphere in $\mathbb{E}^4$ centered at the origin. The unit 3-sphere is a simply connected Riemannian space form of constant curvature 1.

The rotation group SO(4) acts isometrically and transitively on $\mathbb{S}^3$, and the isotropy subgroup at e_0 is SO(3). Hence $\mathbb{S}^3 = \mathrm{SO}(4)/\mathrm{SO}(3)$. This representation is a Riemannian symmetric space representation of $\mathbb{S}^3$ with involution $\mathrm{Ad}_{\mathrm{diag}(-1,1,1,1)}$.

The unit tangent sphere bundle. Let us denote by $\mathrm{U}\mathbb{S}^3$ the *unit tangent sphere bundle* of $\mathbb{S}^3$. Namely, $\mathrm{U}\mathbb{S}^3$ is the manifold of all unit tangent vectors of $\mathbb{S}^3$, and is identified with the submanifold

$$\{(\boldsymbol{x}, \boldsymbol{v}) \mid \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{v}, \boldsymbol{v} \rangle = 1,\ \langle \boldsymbol{x}, \boldsymbol{v} \rangle = 0\}$$

of $\mathbb{E}^4 \times \mathbb{E}^4$. The tangent space $T_{(\boldsymbol{x},\boldsymbol{v})}\mathrm{U}\mathbb{S}^3$ at a point $(\boldsymbol{x}, \boldsymbol{v})$ is expressed as

$$T_{(\boldsymbol{x},\boldsymbol{v})}\mathrm{U}\mathbb{S}^3 = \{(X, V) \in \mathbb{E}^4 \times \mathbb{E}^4 \mid \langle \boldsymbol{x}, X \rangle = \langle \boldsymbol{v}, V \rangle = 0,\ \langle \boldsymbol{x}, V \rangle + \langle \boldsymbol{v}, X \rangle = 0\}.$$

Define a 1-form ω on $\mathrm{U}\mathbb{S}^3$ by

$$\omega_{(\boldsymbol{x},\boldsymbol{v})}(X, V) = \langle X, \boldsymbol{v} \rangle = -\langle \boldsymbol{x}, V \rangle.$$

Then one can see that ω is a *contact form* on $\mathrm{U}\mathbb{S}^3$; that is, $(\mathrm{d}\omega)^2 \wedge \omega \neq 0$. The distribution

$$\mathcal{D}_{(\boldsymbol{x},\boldsymbol{v})} := \{(X, V) \in T_{(\boldsymbol{x},\boldsymbol{v})}\mathrm{U}\mathbb{S}^3 \mid \omega_{(\boldsymbol{x},\boldsymbol{v})}(X, V) = 0\}$$

is called the *canonical contact structure* of $\mathrm{U}\mathbb{S}^3$.

The rotation group SO(4) acts on $\mathrm{U}\mathbb{S}^3$ via the action $A \cdot (\boldsymbol{x}, \boldsymbol{v}) = (A\boldsymbol{x}, A\boldsymbol{v})$. It is easy to see that under this action the unit tangent sphere bundle $\mathrm{U}\mathbb{S}^3$ is a homogeneous space of SO(4). The isotropy subgroup at $(\boldsymbol{e}_0, \boldsymbol{e}_1)$ is

$$\left\{ \begin{pmatrix} I_2 & 0 \\ 0 & b \end{pmatrix} \,\middle|\, b \in \mathrm{SO}(2) \right\}.$$

Here I_2 is the identity matrix of rank 2. Hence, $\mathrm{U}\mathbb{S}^3 = \mathrm{SO}(4)/\mathrm{SO}(2)$. The invariant Riemannian metric induced on $\mathrm{U}\mathbb{S}^3 = \mathrm{SO}(4)/\mathrm{SO}(2)$ is a normal homogeneous metric (and hence naturally reductive), but not Riemannian symmetric. Note that $\mathrm{U}\mathbb{S}^3$ coincides with the *Stiefel manifold* of oriented 2-frames in $\mathbb{E}^4$.

The space of geodesics. Next, we consider $\mathrm{Geo}(\mathbb{S}^3)$, the space of all oriented geodesics in $\mathbb{S}^3$. Take a geodesic $\gamma \in \mathrm{Geo}(\mathbb{S}^3)$; then γ is given by the intersection of $\mathbb{S}^3$ with an oriented 2-dimensional linear subspace W in $\mathbb{E}^4$. By identifying γ with W, the space $\mathrm{Geo}(\mathbb{S}^3)$ is identified with the Grassmann manifold $\mathrm{Gr}_2(\mathbb{E}^4)$ of oriented 2-planes in Euclidean 4-space. The natural projection $\pi_1 : \mathrm{U}\mathbb{S}^3 \to \mathrm{Geo}(\mathbb{S}^3)$ is regarded as the map

$$\pi_1(\boldsymbol{x}, \boldsymbol{v}) = \text{the geodesic } \gamma \text{ satisfying the conditions } \gamma(0) = \boldsymbol{x},\ \gamma'(0) = \boldsymbol{v}.$$

The rotation group SO(4) acts isometrically and transitively on $\mathrm{Geo}(\mathbb{S}^3)$. The isotropy subgroup at $e_0 \wedge e_1$ is $\mathrm{SO}(2) \times \mathrm{SO}(2)$.

Therefore, the tangent space $T_{e_0\wedge e_1}\,\mathrm{Geo}(\mathbb{S}^3)$ is identified with the linear subspace

$$\left\{\begin{pmatrix} 0 & 0 & -x_2 & -x_3 \\ 0 & 0 & -x_{21} & -x_{31} \\ x_2 & x_{21} & 0 & 0 \\ x_3 & x_{31} & 0 & 0 \end{pmatrix}\right\}$$

of $\mathfrak{so}(4)$. The standard invariant complex structure J on $\mathrm{Geo}(\mathbb{S}^3) = \mathrm{Gr}_2(\mathbb{E}^4)$ is given explicitly by

$$J\begin{pmatrix} 0 & 0 & -x_2 & -x_3 \\ 0 & 0 & -x_{21} & -x_{31} \\ x_2 & x_{21} & 0 & 0 \\ x_3 & x_{31} & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & x_{21} & x_{31} \\ 0 & 0 & -x_2 & -x_3 \\ -x_{21} & x_2 & 0 & 0 \\ -x_{31} & x_3 & 0 & 0 \end{pmatrix}.$$

One can see that $\mathrm{Gr}_2(\mathbb{E}^4)$ is a Hermitian symmetric space with Ricci tensor $2\langle\cdot\,,\cdot\rangle$. The Kähler form Ω is related to the contact form ω by $\pi_1^*\Omega = d\omega$.

2. Surface theory in $\mathbb{S}^3$

The Lagrangian and Legendrian Gauss maps. Let $f : M \to \mathbb{S}^3 \subset \mathbb{E}^4$ be a conformal immersion of a Riemann surface with unit normal vector field n. Then we define the (*Lagrangian*) *Gauss map* L of f by

$$L := f \wedge n : M \to \mathrm{Gr}_2(\mathbb{E}^4).$$

One can see that L is an immersion and, in addition, that it is *Lagrangian* with respect to the canonical symplectic form Ω of $\mathrm{Gr}_2(\mathbb{E}^4)$; that is, $L^*\Omega = 0$. Under the identification $\mathrm{Gr}_2(\mathbb{E}^4) = \mathrm{Geo}(\mathbb{S}^3)$, the Lagrangian Gauss map is referred as the oriented normal geodesic of f (and called the *spherical Gauss map*).

On the other hand, we have a map $\mathscr{L} := (f, n) : M \to \mathrm{U}\mathbb{S}^3$. This map is *Legendrian* with respect to the canonical contact structure of $\mathrm{U}\mathbb{S}^3$; that is, $\mathscr{L}^*\omega = 0$. This map $\mathscr{L}$ is called the *Legendrian Gauss map* of f.

Parallel surfaces. An oriented *geodesic congruence* in $\mathbb{S}^3$ is an immersion of a 2-manifold M into the space $\mathrm{Geo}(\mathbb{S}^3)$ of geodesics. Now, let $f : M \to \mathbb{S}^3$ be a surface with unit normal n. Then a *normal geodesic congruence* through f at a distance r is the map $f^r : M \to \mathbb{S}^3$ defined by

$$f^r := \cos r\, f + \sin r\, n.$$

If f satisfies the condition $\cos(2r) - \sin(2r)H + \sin^2(r)K \neq 0$, then f^r is an immersion. Here, H and K are the mean and Gauss curvatures of f, respectively. If f^r is an immersion, then it is called the *parallel surface* of f at the distance r. The correspondence $f \mapsto f^r$ is called the *parallel transformation*.

Legendrian lifts, frontals and fronts. The Gauss map L of an oriented surface $f : M \to \mathbb{S}^3$ with unit normal n is a Lagrangian immersion into $\mathrm{Gr}_2(\mathbb{E}^4)$. Conversely, we have:

Proposition 2.1 [Palmer 1994]. *Let $L : M \to \mathrm{Gr}_2(\mathbb{E}^4)$ be a Lagrangian immersion. Then locally L is a projection of a Legendrian immersion $\mathscr{L} : M \to \mathrm{U}\mathbb{S}^3$. The Legendrian immersion is unique up to parallel transformations.*

The Legendrian immersion $\mathscr{L}$ is called a *Lie surface* by Palmer. If $f : M \to \mathbb{S}^3$ is an immersion with unit normal n, then $\mathscr{L} := (f, n)$ is a Legendrian immersion into $\mathrm{U}\mathbb{S}^3$. However, even if $\mathscr{L}$ is a Legendrian immersion, $f := \pi_2 \circ \mathscr{L}$ need not be an immersion, although it possesses a unit normal n. Here, $\pi_2 : \mathrm{U}\mathbb{S}^3 \to \mathbb{S}^3$ is the natural projection.

Remark 2.2. A smooth map $f : M \to \mathbb{S}^3$ is called a *frontal* if for any point $p \in M$ there exists a neighborhood $\mathscr{U}$ of p and a unit vector field n along f defined on $\mathscr{U}$ such that $\langle \mathrm{d}f, n\rangle = 0$. A frontal is said to be *coorientable* if there exists a unit vector field n along f such that $\langle \mathrm{d}f, n\rangle = 0$. Namely, a coorientable frontal is a smooth map $f : M \to \mathbb{S}^3$ that has a lift $\mathscr{L} = (f, n)$ to $\mathrm{U}\mathbb{S}^3$ satisfying the *Legendrian condition* $\mathscr{L}^*\omega = \langle \mathrm{d}f, n\rangle = 0$. A coorientable frontal is called a *front* if its Legendrian lift is an immersion.

Our main interest is surfaces of constant curvature $K < 1$ in $\mathbb{S}^3$. Except in the case $K = 0$, any surface of constant Gauss curvature $K < 1$ has singularities. A theory of the singularities of fronts can be found in [Arnold 1990]. Geometric concepts such as curvature and completeness for surfaces with singularities have been defined by Saji, Umehara and Yamada in [Saji et al. 2009].

Asymptotic coordinates. Hereafter, assume that the Gaussian curvature K is less than 1. This implies that the second fundamental form II derived from n is a possibly singular Lorentzian metric on M.

Represent K as $K = 1 - \rho^2$ for some positive function ρ, and take a local *asymptotic coordinate system* (u, v) defined on a simply connected domain $\mathbb{D} \subset M$. Then the first and second fundamental forms I and II are given by (see, e.g., [Moore 1972])

$$\mathrm{I} = A^2\,\mathrm{d}u^2 + 2AB\cos\phi\,\mathrm{d}u\,\mathrm{d}v + B^2\,\mathrm{d}v^2, \quad \mathrm{II} = 2\rho AB\sin\phi\,\mathrm{d}u\,\mathrm{d}v. \tag{2-1}$$

Note that asymptotic coordinates (u, v) are *conformal* with respect to the second fundamental form. We may regard M as a (singular) *Lorentz surface* [Weinstein 1996] with respect to the conformal structure determined by II (called the *second conformal structure* [Klotz 1963; Milnor 1983]). Thus, one can see that

$$C = A^2\,\mathrm{d}u^2 + B^2\,\mathrm{d}v^2$$

is well defined on M.

The *Gauss equation* is given by

$$\phi_{uv} - \left(\frac{\rho_v}{2\rho}\frac{B}{A}\sin\phi\right)_v - \left(\frac{\rho_u}{2\rho}\frac{A}{B}\sin\phi\right)_u + (1-\rho^2)AB\sin\phi = 0.$$

Now, we introduce functions a and b by $a = A\rho$ and $b = B\rho$. The *Codazzi equations* are

$$a_v - \frac{\rho_v}{2\rho}a + \frac{\rho_u}{2\rho}b\cos\phi = 0, \quad b_u - \frac{\rho_u}{2\rho}b + \frac{\rho_v}{2\rho}a\cos\phi = 0. \tag{2-2}$$

The Codazzi equations imply that if K is constant, then we have $a_v = b_u = 0$. In addition, the Gauss–Codazzi equations are invariant under the deformation

$$a \mapsto \lambda a, \; b \mapsto \lambda^{-1} b, \quad \lambda \in \mathbb{R}^* := \mathbb{R}\setminus\{0\}.$$

Thus, there exists a one-parameter deformation $\{f_\lambda\}_{\lambda\in\mathbb{R}^*}$ of f preserving the second fundamental form and the Gauss curvature. The resulting family is called the *associated family* of f. The existence of the associated family motivates us to study constant Gauss curvature surfaces in $\mathbb{S}^3$ by loop group methods.

3. The loop group formulation

The* SU(2) × SU(2) *frame. Let us now identify $\mathbb{S}^3$ with SU(2), via

$$(z, w) \in \mathbb{S}^3 \subset \mathbb{R}^4 = \mathbb{C}^2 \quad \longleftrightarrow \quad \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} \in \mathrm{SU}(2).$$

The standard metric g on $\mathbb{S}^3$ is then given by left translating $V, W \in T_x\mathbb{S}^3$ to the tangent space at the identity, $T_e\,\mathrm{SU}(2) = \mathfrak{su}(2)$; that is,

$$g(V, W) := \langle x^{-1}V, x^{-1}W\rangle,$$

where the inner product on $\mathfrak{su}(2)$ is given by $\langle X, Y\rangle = -\operatorname{Tr}(XY)/2$. The natural basis $\{e_0, e_1, e_2, e_3\}$ of $\mathbb{E}^4$ is identified with

$$e_0 = e = I_2, \quad e_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Note that $\{e_1, e_2, e_3\}$ is an orthonormal basis of $\mathfrak{su}(2)$. We have the commutators $[e_1, e_2] = 2e_3$, $[e_2, e_3] = 2e_1$ and $[e_3, e_1] = 2e_2$, so the cross product on $\mathbb{E}^3$ is given by $A \times B = \frac{1}{2}[A, B]$. Note that $\mathbb{S}^3$ is represented by $(\mathrm{SU}(2)\times\mathrm{SU}(2))/\,\mathrm{SU}(2)$ as a Riemannian symmetric space. The natural projection is given by $(G, F) \mapsto GF^{-1}$.

Let M be a simply connected 2-manifold, and suppose given an immersion $f : M \to \mathbb{S}^3$, with global asymptotic coordinates (u, v) and first and second fundamental forms as above at (2-1). Set $\theta = \phi/2$ and

$$\xi_1 = \cos\theta\, e_1 - \sin\theta\, e_2 = \begin{pmatrix} 0 & e^{-i\theta} \\ -e^{i\theta} & 0 \end{pmatrix},$$

$$\xi_2 = \cos\theta\, e_1 + \sin\theta\, e_2 = \begin{pmatrix} 0 & e^{i\theta} \\ -e^{-i\theta} & 0 \end{pmatrix}.$$

Then $\langle \xi_1, \xi_2 \rangle = \cos\phi$, and so we can define a map $F : M \to \mathrm{SU}(2)$ by the equations

$$f^{-1} f_u = A \,\mathrm{Ad}_F\, \xi_1, \quad f^{-1} f_v = B \,\mathrm{Ad}_F\, \xi_2, \quad f^{-1} n = \mathrm{Ad}_F\, e_3, \tag{3-1}$$

where n is the unit normal given by $n = (AB)^{-1} f(f^{-1} f_u \times f^{-1} f_v)$.

Setting $G = fF$, the map $\mathcal{F} = (F, G) : M \to \mathrm{SU}(2) \times \mathrm{SU}(2)$ is a lift of f, and the projection to SU(2) is given by

$$f = GF^{-1}.$$

We call $\mathcal{F}$ the *coordinate frame* for f. We now want to get expressions for the Maurer–Cartan forms of F and G. Differentiating $G = fF$ and substituting in the expressions at (3-1) for $f^{-1} f_u$ and $f^{-1} f_v$, we obtain

$$\begin{aligned} G^{-1}G_u - F^{-1}F_u &= A\xi_1, \\ G^{-1}G_v - F^{-1}F_v &= B\xi_2. \end{aligned} \tag{3-2}$$

Now, write $F^{-1}F_u = a_1 e_1 + a_2 e_2 + a_3 e_3$. Differentiating the expression $f^{-1} f_u = A \,\mathrm{Ad}_F\, \xi_1$, we obtain

$$\begin{aligned} f^{-1} f_{uu} &= A^2 \,\mathrm{Ad}_F\, \xi_1^2 + \frac{\partial A}{\partial u} \mathrm{Ad}_F\, \xi_1 + A \,\mathrm{Ad}_F [F^{-1}F_u, \xi_1] + A \,\mathrm{Ad}_F\, \frac{\partial \xi_1}{\partial u} \\ &= A \,\mathrm{Ad}_F \bigg(-Ae_0 + A^{-1} \frac{\partial A}{\partial u} \xi_1 \\ &\qquad\qquad + [a_1 e_1 + a_2 e_2 + a_3 e_3, \cos\theta\, e_1 - \sin\theta\, e_2] + \frac{\partial \xi_1}{\partial u} \bigg) \\ &= A(-2a_1 \sin\theta - 2a_2 \cos\theta) \,\mathrm{Ad}_F\, e_3 + \mathrm{Ad}_F (d_0 e_0 + d_1 e_1 + d_2 e_2), \end{aligned}$$

where we are only interested in the coefficient of $\mathrm{Ad}_F\, e_3$, that is, of $f^{-1}n$. Since the second fundamental form is assumed to be $\mathrm{II} = 2\rho AB \sin\phi \,\mathrm{d}u\,\mathrm{d}v$, we know that $\langle f^{-1}n, f^{-1} f_{uu} \rangle = 0$. Hence the coefficient of n in the above equation is zero: $0 = A(-2a_1 \sin\theta - 2a_2 \cos\theta)$, or again

$$a_2 = -a_1 \tan\theta.$$

Next, differentiating $f^{-1} f_v = B \,\mathrm{Ad}_F\, \xi_2$ with respect to u, we deduce

$$f^{-1} f_{uv} = AB \,\mathrm{Ad}_F (\xi_1 \xi_2) + \mathrm{Ad}_F \left(\frac{\partial B}{\partial u} \xi_2 + B[F^{-1}F_u, \xi_2] + B \frac{\partial \xi_2}{\partial u} \right),$$

and the coefficient of $\mathrm{Ad}_F\, e_3$ on the right-hand side is

$$AB\sin 2\theta + B(2a_1\sin\theta - 2a_2\cos\theta).$$

Substituting in $a_2 = -a_1\tan\theta$, the equation $\langle f^{-1}n, f^{-1}f_{uv}\rangle = \rho AB\sin\phi$ then becomes

$$\rho AB\sin 2\theta = AB\sin 2\theta + B\,4a_1\sin\theta.$$

Hence, $a_1 = A(\rho-1)\cos(\theta)/2$, and $a_2 = -A(\rho-1)\sin(\theta)/2$. Writing $U_0 := a_3e_3$, we have

$$F^{-1}F_u = U_0 + \frac{\rho-1}{2}A\xi_1.$$

From the equations (3-2) we also have

$$G^{-1}G_u = U_0 + \frac{\rho+1}{2}A\xi_1.$$

Similarly, one obtains the expressions

$$F^{-1}F_v = V_0 - \frac{\rho+1}{2}B\xi_2 \quad \text{and} \quad G^{-1}G_v = V_0 - \frac{\rho-1}{2}B\xi_2,$$

where V_0 is a scalar times e_3.

The Maurer–Cartan form $\alpha = F^{-1}\mathrm{d}F$ of F thus has the expression

$$\begin{aligned} &\alpha = \alpha_0 + \alpha_1 + \alpha_{-1},\\ &\alpha_0 = U_0\,\mathrm{d}u + V_0\,\mathrm{d}v, \quad \alpha_1 = \frac{\rho-1}{2}A\xi_1\,\mathrm{d}u, \quad \alpha_{-1} = -\frac{\rho+1}{2}B\xi_2\,\mathrm{d}v, \end{aligned} \tag{3-3}$$

and similarly, $G^{-1}\,\mathrm{d}G = \beta = \beta_0 + \beta_1 + \beta_{-1}$, with

$$\beta_0 = U_0\,\mathrm{d}u + V_0\mathrm{d}v, \quad \beta_1 = \frac{\rho+1}{2}A\xi_1\,\mathrm{d}u, \quad \beta_{-1} = -\frac{\rho-1}{2}B\xi_2\,\mathrm{d}v. \tag{3-4}$$

One can check that U_0 and V_0 are of the form $U_0 = -\theta_u e_3/2$ and $V_0 = \theta_v e_3/2$.

Ruh–Vilms property. Now we investigate Lorentz harmonicity, with respect to the second conformal structure, of the *normal Gauss map* $\nu = f^{-1}n$ of f. By definition, ν takes values in the unit 2-sphere $\mathbb{S}^2 = \mathrm{Ad}_{\mathrm{SU}(2)}\, e_3$ in the Lie algebra $\mathfrak{su}(2)$. Since f and ν are given by $f = GF^{-1}$, $\nu = \mathrm{Ad}_F\, e_3$, we have

$$\nu_u = \mathrm{Ad}_F[U, e_3], \quad \nu_v = \mathrm{Ad}_F[V, e_3],$$

where $U = F^{-1}F_u$ and $V = F^{-1}F_v$. From these we have

$$\begin{aligned} \frac{\partial}{\partial v}\nu_u &= \mathrm{Ad}_F\left(\left[\frac{\partial U}{\partial v}, e_3\right] + [V,[U,e_3]]\right),\\ \frac{\partial}{\partial u}\nu_v &= \mathrm{Ad}_F\left(\left[\frac{\partial V}{\partial u}, e_3\right] + [U,[V,e_3]]\right). \end{aligned}$$

Next, we have

$$\begin{aligned}
\left[\frac{\partial U}{\partial v}, e_3\right] &= \{A(\rho-1)\sin\theta\}_v e_1 + \{A(\rho-1)\cos\theta\}_v e_2,\\
\left[\frac{\partial V}{\partial v}, e_3\right] &= -\{B(\rho+1)\sin\theta\}_u e_1 + \{B(\rho+1)\cos\theta\}_u e_2,\\
[V,[U,e_3]] &= -A(\rho-1)\theta_v(\cos(\theta)e_1 - \sin(\theta)e_2) + \tfrac{1}{2}AB(\rho^2-1)\cos(2\theta)e_3,\\
[U,[V,e_3]] &= B(\rho+1)\theta_u(\cos(\theta)e_1 + \sin(\theta)e_2) + \tfrac{1}{2}AB(\rho^2-1)\cos(2\theta)e_3.
\end{aligned}$$

Here we recall that a smooth map $\nu : M \to \mathbb{S}^2 \subset \mathbb{E}^3$ of a Lorentz surface M into the 2-sphere is said to be a *Lorentz harmonic map* (or *wave map*) if its tension field with respect to any Lorentzian metric in the conformal class vanishes. This is equivalent to the existence of a function k such that

$$\nu_{uv} = k\nu$$

for any conformal coordinates (u, v).

First, $(\nu_u)_v$ is parallel to ν if and only if

$$A_v(\rho-1) + A\rho_v = 0.$$

Inserting the Codazzi equation (2-2) into this, we get

$$a(\rho+1)\rho_v - b(\rho-1)(\cos\phi)\rho_u = 0. \tag{3-5}$$

Analogously, $(\nu_v)_u$ is parallel to ν if and only if

$$B_u(1+\rho) + B\rho_u = 0.$$

Inserting the Codazzi equation again, we get

$$b(\rho-1)\rho_v - a(\rho+1)(\cos\phi)\rho_v = 0. \tag{3-6}$$

Thus ν is Lorentz harmonic if and only if (3-5) and (3-6) hold. The system (3-5)–(3-6) can be written in matrix form as

$$\begin{pmatrix} b(\rho-1) & -a(\rho+1)\cos\phi \\ -b(\rho-1)\cos\phi & a(\rho+1) \end{pmatrix} \begin{pmatrix} \rho_u \\ \rho_v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The determinant of the coefficient matrix is $ab(\rho^2-1)\sin^2\phi$. Thus, under the condition $\rho \neq 1$, i.e., $K \neq 0$, we have ν is Lorentz harmonic if and only if K is constant.

In case $\rho = 1$, we have $U = -\theta_u e_3/2$ and so $\nu_u = \mathrm{Ad}_F[U, e_3] = 0$. Hence $\nu_{uv} = 0$. Thus g is Lorentz harmonic.
goodbreak

Theorem 3.1. *Let $f : M \to \mathbb{S}^3$ be an isometric immersion of Gauss curvature $K < 1$. Then the normal Gauss map ν is Lorentz harmonic with respect to the conformal structure determined by the second fundamental form if and only if K is constant.*

This characterization is referred as the *Ruh–Vilms property* for constant curvature surfaces in $\mathbb{S}^3$ with $K < 1$.

Remark 3.2. Under the identification $\mathbb{S}^3 = (\mathrm{SU}(2)\times\mathrm{SU}(2))/\,\mathrm{SU}(2)$, the space $\mathrm{Geo}(\mathbb{S}^3)$ is identified with the Riemannian product

$$\mathbb{S}^2\times\mathbb{S}^2 = (\mathrm{SU}(2)\times\mathrm{SU}(2))/(\mathrm{U}(1)\times\mathrm{U}(1)).$$

The Lagrangian Gauss map L corresponds to the map

$$L \longleftrightarrow (nf^{-1}, f^{-1}n) = (\mathrm{Ad}_G\, e_3, \mathrm{Ad}_F\, e_3)$$

(see [Aiyama and Akutagawa 2000; Kitagawa 1995]). Thus the Ruh–Vilms property can be rephrased as follows:

Corollary 3.3. *Let $f : M \to \mathbb{S}^3$ be an isometric immersion of Gauss curvature $K < 1$. Then the Lagrangian Gauss map L is Lorentz harmonic with respect to the conformal structure determined by the second fundamental form if and only if K is constant.*

The Legendrian Gauss map has the formula $\mathscr{L} = (f, n) = (GF^{-1}, Ge_3F^{-1})$.

Remark 3.4. Consider an oriented minimal surface $f : M \to \mathbb{S}^3$ with unit normal n. Then its Lagrangian Gauss map $L = f \wedge n$ is a *harmonic map* with respect to the conformal structure determined by the *first fundamental form*. Hence L is a minimal Lagrangian surface in the Grassmannian [Palmer 1994, Proposition 3.1]; see also [Castro and Urbano 2007].

The loop group formulation for constant curvature surfaces. Let α and β be as defined above at (3-3) and (3-4). Let us now define the family of 1-forms

$$\alpha^\lambda = \alpha_0 + \lambda\alpha_1 + \lambda^{-1}\alpha_{-1},$$

where λ is a (nonzero) complex parameter. The integrability conditions for the 1-forms α and β are $\mathrm{d}\alpha + \alpha\wedge\alpha = 0$ and $\mathrm{d}\beta + \beta\wedge\beta = 0$. Using these two equations, which must already be satisfied, it is straightforward to deduce that α^λ is integrable for all λ if and only if ρ is *constant*; in other words, if and only if the immersion f has *constant* curvature $1-\rho^2$. In this case we have, of course, $\alpha = \alpha^1$, but also

$$\beta = \alpha^\mu, \quad \text{where } \mu = \frac{\rho+1}{\rho-1}. \tag{3-7}$$

From now on, we assume that ρ is constant, so $\mathrm{d}\alpha^\lambda + \alpha^\lambda\wedge\alpha^\lambda = 0$ for all nonzero complex values of λ. Let us choose coordinates for the ambient space such that

$$F(u_0, v_0) = f(u_0, v_0) = G(u_0, v_0) = I \tag{3-8}$$

at some base point (u_0, v_0). We further assume that M is simply connected. Then we can integrate the equations

$$\widehat{F}^{-1}\,\mathrm{d}\widehat{F} = \alpha^\lambda, \quad \widehat{F}(u_0, v_0) = I,$$

to obtain a map $\widehat{F} : M \to \mathcal{G} = \Lambda\mathrm{SL}(2, \mathbb{C})_{\sigma\tau}$. Here the *twisted loop group* $\Lambda\mathrm{SL}(2, \mathbb{C})_{\sigma\tau}$ is the fixed point subgroup of the free loop group $\Lambda\mathrm{SL}(2, \mathbb{C})$ by the involutions σ and τ, which are defined as

$$\sigma x(\lambda) := \mathrm{Ad}_{\mathrm{diag}(1,-1)}(x(-\lambda)), \quad \tau x(\lambda) := \overline{x((\bar{\lambda}))^T}^{\,-1}.$$

Elements of $\mathcal{G}$ take values in SU(2) for real values of λ.

By definition, we have $F = \widehat{F}|_{\lambda=1}$, and, moreover, from (3-7) and the initial condition (3-8) we also have $G = \widehat{F}|_{\lambda=\mu}$. Thus, $\widehat{F}$ can be thought of as a lift of the coordinate frame $\mathcal{F} = (F, G)$, with the projections $(F, G) = (\widehat{F}|_{\lambda=1}, \widehat{F}|_{\lambda=\mu})$ and

$$f = \widehat{F}\big|_{\lambda=\mu}\widehat{F}^{-1}\big|_{\lambda=1}. \tag{3-9}$$

Thus we may call the map $\widehat{F}$ the *extended coordinate frame* for f.

Let us now consider a general map into the twisted loop group $\mathcal{G}$ that has a similar Maurer–Cartan form to α^λ; first, let K be the diagonal subgroup of SU(2) and $\mathfrak{su}(2) = \mathfrak{k} + \mathfrak{p}$ be the symmetric space decomposition induced by $\mathbb{S}^2 = G/K = \mathrm{SU}(2)/\mathrm{U}(1)$ of the Lie algebra; that is,

$$\mathfrak{k} = \mathrm{span}(e_3) \quad \text{and} \quad \mathfrak{p} = \mathrm{span}(e_1, e_2).$$

Definition 3.5. Let M be a simply connected subset of $\mathbb{R}^2$ with coordinates (u, v). An *admissible frame* is a smooth map $\widehat{F} : M \to \mathcal{G}$, the Maurer–Cartan form of which has the Fourier expansion

$$\widehat{F}^{-1}\,\mathrm{d}\widehat{F} = \alpha_0 + \lambda B_1\,\mathrm{d}u + \lambda^{-1}B_{-1}\,\mathrm{d}v, \qquad \alpha_0 \in \mathfrak{k} \otimes \Omega^1(M),\ B_{\pm 1}(u, v) \in \mathfrak{p}.$$

The admissible frame is *regular at a point p* if $B_1(p)$ and $B_{-1}(p)$ are linearly independent, and $\widehat{F}$ is called *regular* if it is regular at every point.

Note that the extended coordinate frame for a constant curvature $1-\rho^2$ immersion, defined above, is a regular admissible frame on M. Conversely:

Lemma 3.6. *Let $\widehat{F} : M \to \mathcal{G}$ be a regular admissible frame. Let μ be any real number not equal to* 1 *or* 0. *Then the map $f : M \to \mathbb{S}^3 = \mathrm{SU}(2)$ defined by the projection* (3-9) *is an immersion of constant curvature*

$$K_\mu = 1 - \rho^2, \quad \textit{where } \rho := \frac{\mu+1}{\mu-1}.$$

The first and second fundamental form are given by

$$\mathrm{I} = A^2\,\mathrm{d}u^2 + 2AB\cos\phi\,\mathrm{d}u\,\mathrm{d}v + B^2\,\mathrm{d}v^2, \quad \mathrm{II} = 2\rho AB\sin\phi\,\mathrm{d}u\,\mathrm{d}v, \tag{3-10}$$

where $A = (\mu-1)|B_1|$, $B = (\mu^{-1}-1)|B_{-1}|$, *and* ϕ *is the angle between* B_1 *and* B_{-1}.

Proof. Set

$$F := \widehat{F}\big|_{\lambda=1}, \quad G := \widehat{F}\big|_{\lambda=\mu},$$

so that $f = GF^{-1}$. Differentiating this formula and using the expressions $F^{-1}\mathrm{d}F = \alpha_0 + B_1\,\mathrm{d}u + B_{-1}\,\mathrm{d}v$ and $G^{-1}\mathrm{d}G = \alpha_0 + \mu B_1\,\mathrm{d}u + \mu^{-1}B_{-1}\,\mathrm{d}v$, we obtain

$$f^{-1}f_u = (\mu-1)\,\mathrm{Ad}_F\,B_1, \quad f^{-1}f_v = (\mu^{-1}-1)\,\mathrm{Ad}_F\,B_{-1}. \tag{3-11}$$

Thus, since $B_{\pm 1}$ are linearly independent for a regular admissible frame, the map f is an immersion and the first fundamental form is given by

$$(\mu-1)^2|B_1|^2\,\mathrm{d}u^2 + 2(\mu-1)\Big(\frac{1}{\mu}-1\Big)\cos\phi\,|B_1|\,|B_{-1}|\,\mathrm{d}u\,\mathrm{d}v + \Big(\frac{1}{\mu}-1\Big)^2|B_{-1}|^2\,\mathrm{d}v^2,$$

where ϕ is the angle between B_1 and B_{-1}. This gives formula $(3\text{-}10)_1$ for the first fundamental form.

It remains to show formula $(3\text{-}10)_2$ for the *second* fundamental form, from which it will follow that the intrinsic curvature is $1-\rho^2$. Since $B_{\pm 1}$ take values in $\mathfrak{p}$, and e_3 is perpendicular to $\mathfrak{p}$, it follows from equations (3-11) that a choice of unit normal is given by

$$n = f\,\mathrm{Ad}_F\,e_3.$$

Differentiating equations (3-11) then leads to

$$\begin{aligned}
\langle f^{-1}f_{uu}, f^{-1}n\rangle &= \langle f^{-1}f_{vv}, f^{-1}n\rangle = 0,\\
\langle f^{-1}f_{uv}, f^{-1}n\rangle &= (1-\mu)(1+\mu^{-1})|B_1|\,|B_{-1}|\sin\phi\\
&= \rho\,(\mu-1)(\mu^{-1}-1)|B_1|\,|B_{-1}|\sin\phi,
\end{aligned}$$

which gives the formula at (3-11) for II. □

Note that

$$\begin{aligned}
&K_\mu \in (0,1] \quad \text{and} \quad K_{-1} = 1 &&\text{for } \mu < 0,\\
&K_\mu < 0 \quad \text{and} \quad \lim_{\mu\to 1} K_\mu = -\infty &&\text{for } \mu > 0.
\end{aligned}$$

The Legendrian Gauss map and Lagrangian Gauss map of $f = \widehat{F}_{\lambda=\mu}\widehat{F}^{-1}_{\lambda=1}$ are given respectively by

$$\mathcal{L} = (\widehat{F}_{\lambda=\mu}\widehat{F}^{-1}_{\lambda=1}, \widehat{F}_{\lambda=\mu}e_3\widehat{F}^{-1}_{\lambda=1}), \quad L = (\mathrm{Ad}_{\widehat{F}_{\lambda=\mu}}e_3, \mathrm{Ad}_{\widehat{F}_{\lambda=1}}e_3).$$

The generalized d'Alembert representation. As we have shown, the problem of finding a nonflat constant curvature immersion $f : M \to \mathbb{S}^3$ with $K < 1$ is equivalent to finding an admissible frame. As a matter of fact, Definition 3.5 of an admissible frame is identical to the extended SU(2) frame for a pseudospherical surface in the Euclidean space $\mathbb{E}^3$ (see, for example, [Brander and Svensson 2013; Dorfmeister and Sterling 2002; Toda 2005]). The surfaces in $\mathbb{E}^3$ are obtained from the same frame, not by the projection (3-9) but by the so-called *Sym formula*. We will explain the connection between these problems in the next section, but the point we are making here is that the problem of constructing these admissible frames by the generalized d'Alembert representation has already been solved in [Toda 2005].

A presentation of the method, using similar definitions to those found here, can be found in [Brander and Svensson 2013]. The building blocks of any admissible frame are these:

Definition 3.7 [Brander and Svensson 2013, Definition 5.1]. Let I_u and I_v be two real intervals, with coordinates u and v respectively. A potential pair (η_+, η_-) is a pair of smooth $\Lambda\mathfrak{sl}(2,\mathbb{C})_{\sigma\tau}$-valued 1-forms on I_u and I_v respectively with Fourier expansions in λ

$$\eta_+ = \sum_{j=-\infty}^{1} (\eta_+)_j \lambda^j \, \mathrm{d}u, \quad \eta_- = \sum_{j=-1}^{\infty} (\eta_-)_j \lambda^j \, \mathrm{d}v.$$

The potential pair is called *regular* if $[(\eta_+)_1]_{12} \neq 0$ and $[(\eta_-)_{-1}]_{12} \neq 0$.

The admissible frame $\widehat{F}$ is then obtained by solving $F_\pm^{-1}\,\mathrm{d}F_\pm = \eta_\pm$ with initial conditions $F_\pm(0) = I$, and thereafter performing at each (u, v) a *Birkhoff decomposition* [Pressley and Segal 1986]

$$F_+^{-1}(u)F_-(v) = H_-(u,v)H_+(u,v), \quad \text{with } H_\pm(u,v) \in \Lambda^\pm\mathrm{SL}(2,\mathbb{C}),$$

and then setting $\widehat{F}(u,v) = F_+(u)H_-(u,v)$.

Example solutions, using a numerical implementation of this method, are computed below.

4. Limiting cases: pseudospherical surfaces in Euclidean space and flat surfaces in the 3-sphere

In this section we discuss the interpretation of admissible frames at degenerate values of the loop parameter μ, namely, the case $\mu = 1$, which was excluded from the above construction, and the limit $\mu \to 0$ or $\mu \to \infty$.

Relation to pseudospherical surfaces in Euclidean space $\mathbb{E}^3$***.*** As alluded to above, in addition to the constant Gauss curvature $K = 1 - \rho^2$ surfaces in $\mathbb{S}^3$ of Lemma 3.6, one also obtains from a regular admissible frame $\widehat{F}$ a constant negative curvature -1 surface in $\mathbb{E}^3$ by the *Sym formula*

$$\check{f} = 2\frac{\partial \widehat{F}}{\partial \lambda}\widehat{F}^{-1}\big|_{\lambda=1}. \tag{4-1}$$

Here we explain how this formula arises naturally from the construction of surfaces in $\mathbb{S}^3 = \mathrm{SU}(2)$.

Obviously the projection formula

$$f_\mu = \widehat{F}\big|_{\lambda=\mu}\widehat{F}^{-1}\big|_{\lambda=1}$$

for the surface in $\mathbb{S}^3$ degenerates to a constant map for $\mu = 1$. On the other hand, we can see that $K_\mu = 1-(\mu+1)^2/(\mu-1)^2$ approaches $-\infty$ when μ approaches 1. This suggests that we multiply our projection formula by some factor, allowing the size of the sphere to vary, so that K approaches some finite limit instead, in order to have an interpretation for the map at $\mu = 1$. Set

$$\check{f}_\mu = \frac{2}{1-\mu}(f_\mu - e_0).$$

Note that $e_0 = (1,0,0,0)$ under our identification $\mathbb{E}^4 = \mathfrak{su}(2) + \mathrm{span}(e_0)$. Now, for $\mu \neq 1$ the function f_μ is a constant curvature K_μ surface in $\mathbb{S}^3$, and $\check{f}_\mu$ is obtained by a constant dilation of $\mathbb{E}^4$ by the factor $2(1-\mu)^{-1}$, plus a constant translation which has no geometric significance. It follows that $\check{f}_\mu$ is a surface in a (translated) sphere of radius $2(1-\mu)^{-1}$, and that $\check{f}_\mu$ has constant curvature

$$\check{K}_\mu = (1/4)(1-\mu)^2 K_\mu. \tag{4-2}$$

Now, consider the function $g : M \times (1-\varepsilon, 1+\varepsilon) \to \mathbb{E}^4$, for some small positive real number ε, given by

$$g(u,v,\lambda) = 2\big(\widehat{F}(u,v)\big|_{\lambda=\lambda}\widehat{F}(u,v)^{-1}\big|_{\lambda=1} - e_0\big).$$

This function is differentiable in all arguments, and

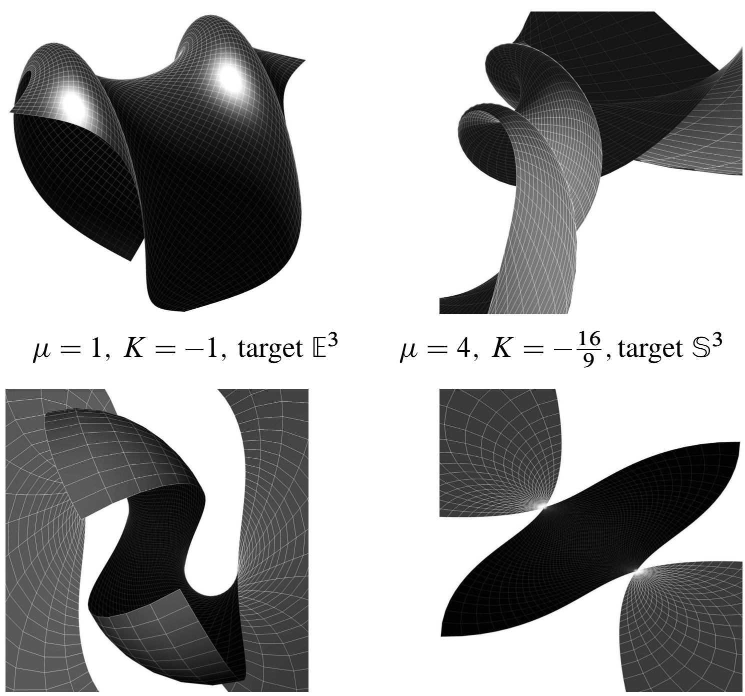
$$\frac{\partial g}{\partial \lambda}\bigg|_{\lambda=1} = 2\lim_{\mu\to 1}\frac{F\big|_{\lambda=\mu}F^{-1}\big|_{\lambda=1} - e_0}{1-\mu} = \lim_{\mu\to 1}\check{f}_\mu.$$

Hence the limit on the right-hand side exists and is a smooth function $M \to \mathbb{E}^4$. On the other hand, differentiating the definition of g, we obtain the right-hand side of the Sym formula (4-1). Note that since F is SU(2)-valued, this expression takes values in the Lie algebra $\mathfrak{su}(2) = \mathrm{span}(e_1, e_2, e_2)$, which in our representation of $\mathbb{E}^4$ is the hyperplane $x_0 = 0$. In other words, $\lim_{\mu\to 1}\check{f}_\mu$ takes values in $\mathbb{E}^3 \subset \mathbb{E}^4$. Assuming that our surface in $\mathbb{S}^3$ is regular, then one can verify that the regularity assumption on the frame F implies that this map is an immersion, and it is clear from expression (4-2) that this surface has constant curvature -1.

Example 4.1. In Figure 3, various projections of the same admissible frame are plotted. These are computed using the generalized d'Alembert method (see [Toda 2005]), using the potential pair

$$\eta_+ = A\,\mathrm{d}u, \quad \eta_- = A\,\mathrm{d}v, \quad A = \begin{pmatrix} 0 & -\lambda^{-1} + i\lambda \\ \lambda^{-1} + i\lambda & 0 \end{pmatrix}.$$

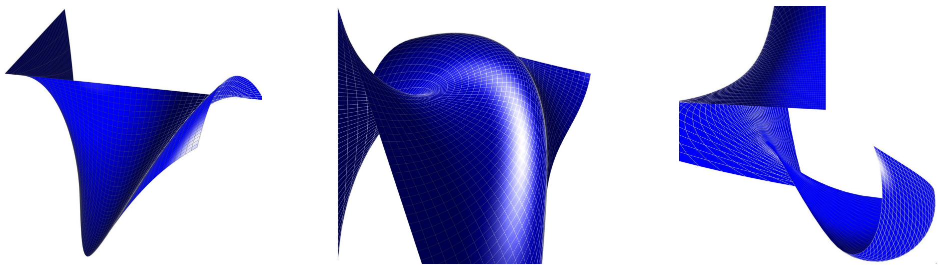
The first image, the surface in $\mathbb{E}^3$ obtained via the Sym formula (4-1), is part of a hyperbolic surface of revolution (a plot of a larger region is shown in Figure 1). The two cuspidal edges that can be seen in this image also appear in the other surfaces at the same places in the coordinate domain, because the condition on the admissible frame for the surface to be regular is independent of μ. The surfaces in $\mathbb{S}^3$ are of course distorted by the stereographic projection, which is taken from the south



$\mu = 1,\ K = -1$, target $\mathbb{E}^3$ $\quad$ $\mu = 4,\ K = -\frac{16}{9}$, target $\mathbb{S}^3$

$\mu = -4,\ K = \frac{16}{25}$, target $\mathbb{S}^3$ $\quad$ $\mu = -1,\ K = 1$, target $\mathbb{S}^3$

Figure 3. Surfaces obtained from one admissible frame evaluated at different values of μ. All images are of the same coordinate patch. The first image is obtained via the Sym formula, and the others are in $\mathbb{S}^3$, stereographically projected to $\mathbb{R}^3$ for plotting.



$\mu=1,\ K=-1$, target $\mathbb{E}^3$ $\quad\mu=4,\ K=-\frac{16}{9}$, target $\mathbb{S}^3$ $\quad\mu=-4,\ K=\frac{16}{25}$, target $\mathbb{S}^3$

Figure 4. Amsler's surface and generalizations in the 3-sphere. The surfaces are obtained from one admissible frame evaluated at different values of μ. All images are of the same coordinate patch.

pole $(-1,0,0,0) \in \mathbb{E}^4$; the north pole $(1,0,0,0)$ is at the center of the coordinate domain plotted. The last image is in fact planar, the projection of a part of a totally geodesic hypersphere $\mathbb{S}^2 \subset \mathbb{S}^3$. In this case, each of the two singular curves in the coordinate domain maps to a single point in the surface.

Example 4.2. Amsler's surface in $\mathbb{E}^3$ can be computed by the generalized d'Alembert method, using the potential pair

$$\eta_+ = \begin{pmatrix} 0 & i\lambda \\ i\lambda & 0 \end{pmatrix} \mathrm{d}u, \quad \eta_- = \begin{pmatrix} 0 & -\lambda^{-1} \\ \lambda^{-1} & 0 \end{pmatrix} \mathrm{d}v.$$

The image of a rectangle $[0,a] \times [0,b]$ in the positive quadrant of the uv-plane is plotted in Figure 4, evaluated at 3 different values of μ. The coordinate axes correspond to straight lines for the surface in $\mathbb{E}^3$, and to great circles for the surfaces in $\mathbb{S}^3$, which project to straight lines under the stereographic projection from the south pole. The north pole $(1,0,0,0)$ corresponds to $(u,v) = (0,0)$.

The singular set in the coordinate patch corresponds to a cuspidal edge in each of the first two images, but contains a swallowtail singularity in the third. See also Figure 2.

Relation to flat surfaces in the 3-sphere. We have considered above the surfaces f_μ, obtained by the projection

$$\widehat{F}|_{\lambda=\mu}\, \widehat{F}^{-1}|_{\lambda=1}, \tag{4-3}$$

for all nonzero real values of μ. We now consider the limit as μ approaches 0 or ∞. From the formula $K_\mu = 1-(\mu+1)^2/(\mu-1)^2$, it is clear that the limiting surface, if it exists, will be flat. We discuss the case $\mu \to 0$ here.

Observe that the admissible frame $\widehat{F}$ has a pole at $\lambda = 0$, so we cannot evaluate (4-3) at $\mu = 0$. However, in the Maurer–Cartan form of $\widehat{F}$, the factor λ^{-1} appears only as a coefficient of $\mathrm{d}v$. Hence a change of coordinates could remove the pole in λ. For $\mu > 0$, we set $\tilde{u} = u$ and $\tilde{v} = v/\mu$, so that

$$f_\mu(u, v) = f_\mu(\tilde{u}, \mu\tilde{v}) =: g_\mu(\tilde{u}, \tilde{v}).$$

For simplicity, let us assume that M is a rectangle $(a, b) \times (c, d) \subset \mathbb{R}^2$, containing the origin $(0, 0)$ and with coordinates (u, v). We denote by M_μ the same rectangle in the coordinates $(\tilde{u}, \tilde{v})$, that is, $M_\mu = (a, b) \times (c/\mu, d/\mu)$, and we define $M_0 := (a, b) \times (-\infty, \infty)$.

We have already seen that for $\mu > 0$, the map $g_\mu : M_\mu \to \mathbb{S}^3$ is an immersion of constant curvature $K_\mu = 1 - (\mu + 1)^2/(\mu - 1)^2$, since this is just the same map as f_μ in different coordinates. For fixed $\mu_0 \in (0, 1)$, if $0 < \mu < \mu_0$ then $M_\mu \supset M_{\mu_0}$, and so we can restrict g_μ to M_{μ_0} and talk about a family of maps $g_\mu : M_{\mu_0} \to \mathbb{S}^3$ with a fixed domain.

Lemma 4.3. *For any fixed $\mu_0 \in (0, 1)$, the family of maps $g_\mu : M_{\mu_0} \to \mathbb{S}^3$ extends real analytically in μ to $\mu = 0$. Moreover, the map $g_0 : M_{\mu_0} \to \mathbb{S}^3$ extends to the whole of $M_0 = (a, b) \times (-\infty, \infty)$, and is an immersion of zero Gaussian curvature.*

Proof. Write $\widehat{G}_\mu(\tilde{u}, \tilde{v}) = \widehat{F}(\tilde{u}, \mu\tilde{v}) = \widehat{F}(u, v)$, so $\widehat{G}_\mu : M_\mu \to \mathcal{G}$. Then

$$g_\mu(\tilde{u}, \tilde{v}) = H_\mu(\tilde{u}, \tilde{v}) K_\mu^{-1}(\tilde{u}, \tilde{v}), \quad \text{where } H_\mu := \widehat{G}_\mu\big|_{\lambda=\mu},\ K_\mu := \widehat{G}_\mu\big|_{\lambda=1}.$$

Since $\widehat{F}$ is an admissible frame, we can write

$$\begin{aligned} \widehat{F}^{-1}\,\mathrm{d}\widehat{F} &= (U_0 + \lambda U_1)\,\mathrm{d}u + (V_0 + \lambda^{-1} V_1)\,\mathrm{d}v \\ &= \big(U_0(\tilde{u}, \mu\tilde{v}) + \lambda U_1(\tilde{u}, \mu\tilde{v})\big)\,\mathrm{d}\tilde{u} + \big(\mu V_0(\tilde{u}, \mu\tilde{v}) + \mu\lambda^{-1} V_1(\tilde{u}, \mu\tilde{v})\big)\,\mathrm{d}\tilde{v}, \end{aligned}$$

and thus

$$H_\mu^{-1}\,\mathrm{d}H_\mu = (U_0 + \mu U_1)\,\mathrm{d}\tilde{u} + (\mu V_0 + V_1)\,\mathrm{d}\tilde{v},$$

so

$$H_0^{-1}\,\mathrm{d}H_0 = U_0(\tilde{u}, 0)\,\mathrm{d}\tilde{u} + V_1(\tilde{u}, 0)\,\mathrm{d}\tilde{v},$$

and

$$K_\mu^{-1}\,\mathrm{d}K_\mu = (U_0 + U_1)\,\mathrm{d}\tilde{u} + (\mu V_0 + \mu V_1)\,\mathrm{d}\tilde{v},$$

so

$$K_0^{-1}\,\mathrm{d}K_0 = (U_0(\tilde{u}, 0) + U_1(\tilde{u}, 0))\,\mathrm{d}\tilde{u}.$$

Since H_μ and K_μ are both obviously real analytic in μ in a neighborhood of $\mu = 0$, so also is g_μ. Finally, the 1-forms $\gamma = H_0^{-1}\,\mathrm{d}H_0$ and $\delta = K_0^{-1}\,\mathrm{d}K_0$ are both integrable on $M_{\mu_0} = (a, b) \times (c/\mu_0, d/\mu_0)$ for any fixed μ_0. But, since the coefficients of the 1-forms are constant in $\tilde{v}$, this means that they are in fact integrable on the whole of $(a, b) \times (-\infty, \infty)$. This implies the claim. $\square$

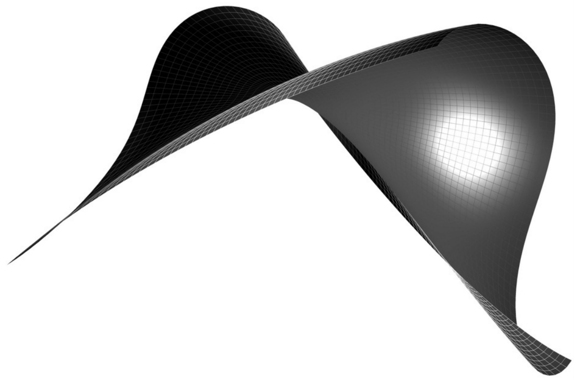


Figure 5. The surface g_μ for $\mu = 10^{-9}$, obtained from the same admissible frame used in Figure 3.

Using the expressions γ and δ above, we obtain the formula

$$g_0^{-1}\,\mathrm{d}g_0 = \mathrm{Ad}_{K_0}(-U_1(\tilde{u},0)\,\mathrm{d}\tilde{u} + V_1(\tilde{u},0)\,\mathrm{d}\tilde{v}),$$

from which we have the following expression for the first fundamental form of g_0:

$$I(\tilde{u},\tilde{v}) = (|B_1|^2\,\mathrm{d}\tilde{u}^2 - 2\cos\phi\,|B_1||B_{-1}|\,\mathrm{d}\tilde{u}\,\mathrm{d}\tilde{v} + |B_{-1}|^2\,\mathrm{d}\tilde{v}^2)\big|_{(\tilde{u},0)}.$$

Letting $\mu \to 0$ in expression (3-10), we conclude that the second fundamental form of g_0 is

$$\mathrm{II} = 2|B_1(\tilde{u},0)|\,|B_{-1}(\tilde{u},0)|\sin(\phi(\tilde{u},0))\,\mathrm{d}\tilde{u}\,\mathrm{d}\tilde{v}.$$

Example 4.4. In Figure 5 is shown the surface g_μ for $\mu = 10^{-9}$, obtained from the same admissible frame $\widehat{F}_\mu$ used in Example 4.1. A square region in the $(\tilde{u},\tilde{v})$-plane is plotted, approximately equal to the region $(a,b)\times(c/\mu, d/\mu)$ in the uv-plane, where $(a,b)\times(c,d)$ is the region plotted in Example 4.1. The region plotted here is actually slightly larger, in order to make the singular set visible. As μ approaches zero, the cuspidal edges, which, in the nonflat surface, were approximated by $v = \pm u + \text{constant}$, are now approaching curves of the form $\tilde{v} = \text{constant}$.

References

[Aiyama and Akutagawa 2000] R. Aiyama and K. Akutagawa, "Kenmotsu type representation formula for surfaces with prescribed mean curvature in the 3-sphere", *Tohoku Math. J.* (2) **52**:1 (2000), 95–105. MR 2001k:53009 Zbl 1008.53012

[Arnold 1990] V. I. Arnold, *Singularities of caustics and wave fronts*, Mathematics and its Applications (Soviet Series) **62**, Kluwer, Dordrecht, 1990. MR 93b:58019 Zbl 0734.53001

[Borisenko 2001] A. A. Borisenko, "Isometric immersions of space forms in Riemannian and pseudo–Riemannian spaces of constant curvature", *Uspekhi Mat. Nauk* **56**:3 (2001), 3–78. In Russian; translated in *Russ. Math. Surv.* **56**:3 (2001), 425–497. MR 2003a:53086 Zbl 1038.53062

[Brander 2007] D. Brander, "Curved flats, pluriharmonic maps and constant curvature immersions into pseudo–Riemannian space forms", *Ann. Global Anal. Geom.* **32**:3 (2007), 253–275. MR 2008i:53075 Zbl 1127.37048

[Brander and Dorfmeister 2009] D. Brander and J. Dorfmeister, "Generalized DPW method and an application to isometric immersions of space forms", *Math. Z.* **262** (2009), 143–172. MR 2009m:37186 Zbl 1167.37031

[Brander and Svensson 2013] D. Brander and M. Svensson, "The geometric Cauchy problem for surfaces with Lorentzian harmonic Gauss maps", *J. Differential Geom.* **93**:1 (2013), 37–66. MR 3019511 Zbl 06201312

[Castro and Urbano 2007] I. Castro and F. Urbano, "Minimal Lagrangian surfaces in $\mathbb{S}^2 \times \mathbb{S}^2$", *Comm. Anal. Geom.* **15**:2 (2007), 217–248. MR 2008j:53107 Zbl 1185.53063

[Dorfmeister 2008] J. Dorfmeister, "Generalized Weierstraß representations of surfaces", pp. 55–111 in *Surveys on geometry and integrable systems*, edited by M. Guest et al., Adv. Stud. Pure Math. **51**, Math. Soc. Japan, Tokyo, 2008. MR 2010a:58020 Zbl 1168.53031

[Dorfmeister and Sterling 2002] J. Dorfmeister and I. Sterling, "Finite type Lorentz harmonic maps and the method of Symes", *Differential Geom. Appl.* **17**:1 (2002), 43–53. MR 2003f:58032 Zbl 1027.58012

[Ferus and Pedit 1996] D. Ferus and F. Pedit, "Isometric immersions of space forms and soliton theory", *Math. Ann.* **305**:2 (1996), 329–342. MR 97d:53061 Zbl 0866.53046

[Hilbert 1901] D. Hilbert, "Ueber Flächen von constanter Gaussscher Krümmung", *Trans. Amer. Math. Soc.* **2**:1 (1901), 87–99. MR 1500557 JFM 32.0608.01

[Kitagawa 1988] Y. Kitagawa, "Periodicity of the asymptotic curves on flat tori in S^3", *J. Math. Soc. Japan* **40**:3 (1988), 457–476. MR 89g:53080 Zbl 0642.53059

[Kitagawa 1995] Y. Kitagawa, "Embedded flat tori in the unit 3-sphere", *J. Math. Soc. Japan* **47**:2 (1995), 275–296. MR 96e:53093 Zbl 0836.53035

[Klotz 1963] T. Klotz, "Some uses of the second conformal structure on strictly convex surfaces", *Proc. Amer. Math. Soc.* **14** (1963), 793–799. MR 27 #2917 Zbl 0126.17302

[Kričever 1980] I. M. Kričever, "An analogue of the d'Alembert formula for the equations of a principal chiral field and the sine–Gordon equation", *Dokl. Akad. Nauk SSSR* **253**:2 (1980), 288–292. In Russian; translated in *Sov. Math. Dokl.* **22** (1980), 79–84. MR 82k:35095 Zbl 0496.35073

[Milnor 1983] T. K. Milnor, "Harmonic maps and classical surface theory in Minkowski 3-space", *Trans. Amer. Math. Soc.* **280**:1 (1983), 161–185. MR 85e:58037 Zbl 0532.53047

[Moore 1972] J. D. Moore, "Isometric immersions of space forms in space forms", *Pacific J. Math.* **40** (1972), 157–166. MR 46 #4442 Zbl 0238.53033

[Palmer 1994] B. Palmer, "Buckling eigenvalues, Gauss maps and Lagrangian submanifolds", *Differential Geom. Appl.* **4**:4 (1994), 391–403. MR 95j:53107 Zbl 0819.53028

[Pressley and Segal 1986] A. Pressley and G. Segal, *Loop groups*, Clarendon Press, New York, 1986. MR 88i:22049 Zbl 0618.22011

[Saji et al. 2009] K. Saji, M. Umehara, and K. Yamada, "The geometry of fronts", *Ann. of Math.* (2) **169**:2 (2009), 491–529. MR 2010e:58042 Zbl 1177.53014

[Spivak 1979] M. Spivak, *A comprehensive introduction to differential geometry, IV*, 2nd ed., Publish or Perish, Wilmington, DE, 1979. MR 82g:53003d Zbl 0439.53004

[Toda 2005] M. Toda, "Initial value problems of the sine–Gordon equation and geometric solutions", *Ann. Global Anal. Geom.* **27**:3 (2005), 257–271. MR 2007f:35205 Zbl 1077.53010

[Weinstein 1996] T. Weinstein, *An introduction to Lorentz surfaces*, de Gruyter Expositions in Mathematics **22**, de Gruyter, Berlin, 1996. MR 98a:53104 Zbl 0881.53001

[Xia 2007] Q. Xia, "Generalized Weierstrass representations of surfaces with the constant Gauss curvature in pseudo–Riemannian three-dimensional space forms", *J. Math. Phys.* **48**:4 (2007), 042301, 18. MR 2008d:53012 Zbl 1137.53305

Received January 25, 2013.

DAVID BRANDER
INSTITUT FOR MATEMATIK OG COMPUTER SCIENCE
MATEMATIKTORVET
BYGNING 303B
TECHNICAL UNIVERSITY OF DENMARK
DK-2800
KONGENS LYNGBY
DENMARK
D.Brander@mat.dtu.dk

JUN-ICHI INOGUCHI
DEPARTMENT OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
YAMAGATA UNIVERSITY
YAMAGATA 990-8560
JAPAN
inoguchi@sci.kj.yamagata-u.ac.jp

SHIMPEI KOBAYASHI
DEPARTMENT OF MATHEMATICS
HOKKAIDO UNIVERSITY
SAPPORO 060-0810
JAPAN
shimpei@math.sci.hokudai.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.305

ON EMBEDDINGS INTO COMPACTLY GENERATED GROUPS

PIERRE-EMMANUEL CAPRACE AND YVES CORNULIER

We prove that there is a second-countable locally compact group that does not embed as a closed subgroup in any compactly generated locally compact group, and discuss various related embedding and nonembedding results.

1. Introduction

The Higman–Neumann–Neumann (HNN) theorem [Higman et al. 1949] ensures that every countable group embeds as a subgroup of a finitely generated group, indeed 2-generated (relying on a fundamental construction referred to since then as *HNN-extension*). This was a major breakthrough, providing some of the first evidence that finitely generated groups are not structurally simpler than countable groups and thus are far from tame or classifiable. B. H. and H. Neumann [Neumann and Neumann 1959] gave an alternative construction, showing for instance that every countable k-solvable group (that is, solvable of derived length at most k) embeds as a subgroup of a finitely generated $(k+2)$-solvable group. Further refinements by P. Hall [1954] and P. Schupp [1976] in a slightly different direction showed that every countable group embeds in a 2-generated *simple* group.

In the present paper, we address similar questions in the context of *locally compact topological groups*, which will be abbreviated henceforth by the term *l.c. groups*. Recall that locally compact groups are a natural generalization of discrete groups, the counterpart of *countability* (resp. *finite generation*), being *σ-compactness* (resp. *compact generation*). A prototypical example of an embedding of a noncompactly generated l.c. group into a compactly generated one is the embedding of the p-adic additive group $\mathbf{Q}_p$ into the affine group $\mathbf{Q}_p \rtimes \mathbf{Q}_p^\times$ (or its discrete cousin, the embedding of the additive underlying group of the ring $\mathbf{Z}[1/p]$ into the Baumslag–Solitar group $\mathbf{Z}[1/p] \rtimes_p \mathbf{Z}$).

It is natural to ask whether analogues of the HNN theorem hold in the context of l.c. groups. In that context, an *embedding* $\varphi : H \to G$ of an l.c. group H to an

Caprace was partly supported by FNRS grant F.4520.11 and the European Research Council (grant #278469). Cornulier was partly supported by ANR grant GSG..
MSC2010: primary 22D05; secondary 05C63, 20E18, 57S30.
Keywords: locally compact groups, compactly generated groups, topologically simple groups, unrestricted wreath products.

l.c. group G is defined as a continuous injective homomorphism (with potentially nonclosed image). In the nondiscrete setting, several natural variants of the question can be considered:

- Given a σ-compact l.c. group H, is there any embedding $\varphi : H \to G$ into a compactly generated l.c. group G?
- Is there one with closed image?
- Is there one with open image?

It turns out that whenever the topology on H is nondiscrete, the answers to these questions are not always positive, and depend heavily on the algebraic structure of H. The results of this note are intended to illustrate that matter of fact. We start with a positive result in the case where the algebraic structure of H is the simplest possible, namely, when H is abelian:

Theorem 1.1. *Every σ-compact abelian l.c. group A embeds as an open subgroup of a compactly generated group G, which can be chosen to be* 3*-solvable. If moreover A is totally disconnected*, *second countable*, *or both*, *then G can also be chosen to enjoy the same additional property.*

In particular, the additive group of adeles, defined as a restricted product of all $\mathbf{Q}_p$ (see Section 4 for the definition) is isomorphic to an open subgroup of a compactly generated locally compact group. In contrast, the adeles are used to prove the following result, which shows in particular that Theorem 1.1 cannot be generalized to solvable groups:

Theorem 1.2. *There exists a σ-compact* metabelian *l.c. group M not isomorphic to any closed subgroup of any compactly generated l.c. group.*

Moreover M can be chosen to be second countable and totally disconnected.

The proof of Theorem 1.2 is based on the now classical observation, due to H. Abels [1974, Beispiel 5.2], that every compactly generated l.c. group admits, in a somewhat natural way, a continuous proper action on a connected graph of bounded degree (see Proposition 2.1 below). Using similar ideas, we obtain the following result, which shows that the HNN theorem fails in the nondiscrete setting, even if one allows embeddings with potentially nonclosed images:

Theorem 1.3. *There exists a second-countable* (*hence σ-compact*) *topologically simple totally disconnected l.c. group S such that every continuous* (*or even abstract*) *homomorphism of S to any compactly generated l.c. group is trivial.*

This stands in sharp contrast with the discrete case. We remark also that local compactness is absolutely essential to this result, since it is known from [Pestov 1986] that every σ-compact topological Hausdorff group is isomorphic to a closed subgroup of some compactly generated topological Hausdorff group.

We finally present a result illustrating the difference between embeddings with closed and open images:

Theorem 1.4. *There exists a second-countable* (*hence* σ*-compact*) *l.c. group* H *that is isomorphic to a* closed *subgroup of a compactly generated l.c. group*, *but not to any* open *subgroup of any compactly generated l.c. group.*

Moreover, H *can be chosen to be of the form* $K \rtimes \Gamma$, *with* Γ *discrete abelian and* K *compact abelian and either connected or profinite. It can also be chosen to be a Lie group.*

One part of the implication in Theorem 1.4 is the following general fact, which is based on a wreath product construction:

Proposition 1.5. *Any compact-by-{countable discrete} l.c. group embeds as a closed subgroup in a compact-by-{finitely generated discrete} l.c. group.*

Similarly to Theorem 1.1, this proposition illustrates that embedding theorems can hold in the nondiscrete case when the algebraic or topological structure of the group H is not too complicated.

We finish by mentioning some related natural questions which we have not been able to answer.

Question 1.6. Is every second-countable (real) Lie group isomorphic to a closed subgroup of a compactly generated locally compact group? Of a compactly generated Lie group? Same questions for p-adic Lie groups.

The answer to Question 1.6 with "closed subgroup" replaced by "open subgroup" is negative for both real and p-adic Lie groups; see the examples in Section 6.

2. Locally compact groups, Lie groups, and locally finite graphs

We shall use the following general result about l.c. groups; the first part follows from the solution to Hilbert's fifth problem, the second is an elementary but crucial observation due to H. Abels:

Proposition 2.1. *Let* G *be an l.c. group and* V *be any identity neighborhood.*

(i) (Yamabe) *If* G *is connected-by-compact* (*i.e.*, *if* G/G° *is compact*), *then* V *contains a compact normal subgroup* K *of* G *such that* G/K *is a connected Lie group.*

(ii) (Abels) *If* G *is totally disconnected and compactly generated*, *then* V *contains a compact normal subgroup* W *of* G *such that* G/W *admits a faithful continuous proper vertex-transitive action on some connected locally finite graph.*

Proof. For (i), see [Montgomery and Zippin 1955, Theorem IV.4.6]. For (ii), originally observed in [Abels 1974, Beispiel 5.2], refer to [Monod 2001, §11.3]. □

We deduce the following useful criterion for the nonexistence of embeddings into compactly generated l.c. groups:

Proposition 2.2. *Let H be an l.c. group. The following are equivalent*:

(1) *Every continuous homomorphism of H to a compactly generated l.c. group is trivial.*

(2) *The following two conditions are satisfied*:

   (a) *Every continuous homomorphism of H to a compactly generated totally disconnected l.c. group is trivial.*

   (b) *For any n, every continuous linear representation $H \to \mathrm{GL}_n(\mathbf{C})$ is trivial.*

Moreover, *a sufficient condition for* (a) *is that H has no nontrivial continuous action on any connected graph of bounded degree.*

Proof. That (1) implies (2) is immediate. Conversely, assume that (2) holds. Let G be a compactly generated l.c. group and $f : H \to G$ a continuous homomorphism. Considering the composite map $H \to G \to G/G^\circ$ in view of (a), we see that $f(H) \subset G^\circ$. If f is not the trivial map, some identity neighborhood V in G does not contain the image of f. By Proposition 2.1(i), there is a compact normal subgroup K of G° contained in V such that $L = G^\circ/K$ is a (connected) Lie group. So the composite map $H \to L$ is nontrivial. Using the adjoint representation of L and (b), we see that it maps H into the center of L. On the other hand, it follows from Pontryagin duality and (b) that H admits no nontrivial continuous homomorphism to any abelian l.c. group. So we get a contradiction, and thus f is the trivial homomorphism.

Let us now assume that H has no nontrivial continuous action on any connected graph of bounded degree and let us check that (a) holds. Let $f : H \to G$ be a continuous homomorphism, with G a compactly generated totally disconnected l.c. group. If f is nontrivial, some identity neighborhood V in G does not contain the image of f. By Proposition 2.1(ii), there is a compact normal subgroup K of G contained in V such that G/K acts continuously, faithfully, and vertex-transitively on a connected locally finite graph. The hypothesis made on H implies that the restriction of this action to H is trivial. Thus $f(H)$ is contained in K, hence in V, which is a contradiction. □

Remark 2.3. We do not know if, conversely, (a) implies that H has no nontrivial continuous action on any connected graph of bounded degree. In other words, does the existence of a continuous nontrivial action on a connected graph of bounded degree imply the existence of such an action on a vertex-transitive graph? The same question, replacing "nontrivial" by "proper," can also naturally be addressed.

Finally, we record an elementary fact, allowing us in suitable situations to exclude actions on some connected locally finite graphs:

Lemma 2.4. *Let G be an l.c. group acting continuously by automorphisms on a connected graph all of whose vertices have degree $\leq d$. Then every vertex stabilizer O is open in G and, for any prime $p > d$, every closed pro-p subgroup of O acts trivially on the graph. In particular, if G admits an open pro-p-group, then the action has an open kernel.*

Proof. Let O be a vertex stabilizer, which is open in G since the action on the graph is assumed continuous. Given any closed subgroup H of O which acts nontrivially on the graph, there is a vertex v fixed by H and adjacent to some vertex that is not fixed by H. In particular, H admits some nontrivial continuous permutation action on the set of neighbors of v, which is a set of at most d elements. It follows that H cannot be pro-p for any $p > d$. □

3. Proof of Theorem 1.1

Recall that any totally disconnected l.c. group contains compact open subgroups. Moreover, every abelian l.c. group A has a (noncanonical) decomposition as a topological direct product $\mathbf{R}^n \times W$, where W is compact-by-discrete, and the discrete quotient is countable as soon as A is σ-compact. Those facts could be used to deduce (a part of) Theorem 1.1 from Proposition 1.5. This is however not what we shall do here, and we present rather a simpler direct argument.

We begin with an easy classical result:

Lemma 3.1. *There exists a finitely generated group Γ whose center contains a free abelian group Z of countable rank; Γ can be chosen to be 3-solvable.*

Proof. If t is an indeterminate, the reader can check that the three matrices

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

generate a group containing the set of all matrices of the form

$$\begin{pmatrix}1&0&P(t)\\0&1&0\\0&0&1\end{pmatrix},\qquad P(t)\in\mathbf{Z}[t,1/t],$$

as a central, infinitely generated subgroup. (This construction is due to Hall [1954, Theorem 7].) □

Lemma 3.2. *If G is a σ-compact l.c. group, then it has a cocompact closed separable subgroup.*

Proof. By the Kakutani–Kodaira theorem, there is a compact normal subgroup K such that G/K is second countable. So G/K admits a dense countable subset. Lift this subset to G, and let D be the abstract (countable) group it generates. So $G = \overline{KD} = K\overline{D}$ since K is compact. Thus $\overline{D}$ is cocompact; moreover, it is separable by construction. □

Proof of Theorem 1.1. By Lemma 3.2, there is a cocompact closed separable subgroup in A. In other words, there is a homomorphism $f : Z \to A$ whose image has cocompact closure, where $Z = \mathbf{Z}^{(\omega)}$ is the restricted product of countably many copies of the infinite cyclic group. In view of Lemma 3.1, the group Z can be embedded as a central subgroup of a finitely generated group Γ (which can be chosen to be 3-solvable). The graph F of f is a closed discrete central subgroup of $\Gamma \times A$. Since f is injective, it follows that the mapping of A into $G = (\Gamma \times A)/F$ is injective. Moreover A has open image (because the quotient map is open). So A lies as an open (and central) subgroup of G. The latter group is compactly generated: indeed, the closure of the subgroup generated by a finite generating subset of Γ is cocompact. By construction, if A is second countable or totally disconnected, then so is G. □

4. Proof of Theorem 1.2

Consider $B_p = \mathbf{Q}_p \rtimes_p \mathbf{Z}$, where the notation $\rtimes_p$ means that the $\mathbf{Z}$-action is through multiplication by powers of p. Let A be the group of adeles; that is, the set of elements in $\prod_p \mathbf{Q}_p$ (p ranging over all primes) whose projection in $\prod \mathbf{Q}_p/\mathbf{Z}_p$ is finitely supported, endowed with the ring topology for which $\prod \mathbf{Z}_p$ is a compact open subring. In the product $\prod_p B_p = \left(\prod_p \mathbf{Q}_p\right) \rtimes \prod_p \mathbf{Z}$, consider the subgroup $Z = \bigoplus_p \mathbf{Z}$ and endow it with the discrete topology. Finally, define $M = A \rtimes Z \subset \prod_p B_p$. The group M is metabelian and admits a unique Hausdorff group topology for which $\prod_p \mathbf{Z}_p$ is a compact open subgroup. In particular, M is locally compact.

Theorem 1.2 is a consequence of the following:

Proposition 4.1. *There is no embedding of M as a closed subgroup of any compactly generated l.c. group.*

More precisely, given any continuous homomorphism $f : M \to G$ to a compactly generated l.c. group G, there exists p_0 such that $\overline{f(\mathbf{Q}_p)}$ is a compact connected group for all $p \geq p_0$.

We begin with two lemmas on homomorphisms from $\mathbf{Q}_p$ into locally compact groups:

Lemma 4.2. *For every continuous homomorphism $f : \mathbf{Q}_p \to G$ of $\mathbf{Q}_p$ to a connected-by-compact l.c. group G, the closure of the image $\overline{f(\mathbf{Q}_p)}$ is compact and connected.*

Proof. Assume first that G is a virtually connected Lie group. Since $\mathbf{Q}_p$ is divisible, it has no nontrivial finite quotient. Thus $\overline{f(\mathbf{Q}_p)}$ is a closed abelian subgroup of a connected Lie group, so is isomorphic to a product $\Gamma \times \mathbf{R}^k \times T$ for some finitely generated abelian group Γ and torus T. Invoking again that $\mathbf{Q}_p$ has no nontrivial finite quotient, we find $\Gamma = \{0\}$. Since $\mathbf{R}^k \times T$ has no small subgroup, the kernel of f must be open in $\mathbf{Q}_p$. In particular, $f(\mathbf{Q}_p)$ is a torsion group, from which we infer that $k = 0$. Therefore $\overline{f(\mathbf{Q}_p)}$ is a torus; in particular, it is compact.

Coming back to the general case, we now let W be the maximal compact normal subgroup of G, which exists by Proposition 2.1(i). Proposition 4.1 ensures that G/W is a virtually connected Lie group. By the special case above, we deduce that, denoting $K = \overline{f(\mathbf{Q}_p)}$, the group KW/W is compact. Hence K is compact as well. Since K/K° is profinite and $\mathbf{Q}_p$ has no nontrivial finite quotient, $K = K^\circ$; that is, K is connected. □

Remark 4.3. It follows from Pontryagin duality that $\mathbf{Q}_p$ has a continuous homomorphism with dense image into the circle, and also has an injective continuous homomorphism with dense image into the Pontryagin dual $\widehat{\mathbf{Q}}$ of $\mathbf{Q}$, which is a connected compact group.

Lemma 4.4. *Every nontrivial continuous homomorphism $f : \mathbf{Q}_p \to G$ of $\mathbf{Q}_p$ to a totally disconnected l.c. group G is proper, and either has a compact open kernel or is an isomorphism to its (closed) image.*

Proof. We can suppose that f has dense image, so G is abelian. Let U be a compact open subgroup in G. Then $f^{-1}(U)$ is an open subgroup of $\mathbf{Q}_p$. If it is all of $\mathbf{Q}_p$, then $U = G$, and since U is profinite and $\mathbf{Q}_p$ has no nontrivial finite quotient, it follows that $U = \{1\}$, contradicting that f is nontrivial. Otherwise, $f^{-1}(U)$ is a compact open subgroup, so f is proper and, in particular, has closed image and is the quotient map by some compact subgroup, giving the two possibilities. □

Lemma 4.5. *Every continuous homomorphism $f : B_p \to G$ of the group $B_p = \mathbf{Q}_p \rtimes_p \mathbf{Z}$ to a totally disconnected l.c. group G satisfies one of two alternatives: either f is a topological isomorphism to its closed image, or $f(\mathbf{Q}_p)$ is trivial.*

Proof. If $f(\mathbf{Q}_p)$ is nontrivial, then f is proper in restriction to $\mathbf{Q}_p$ by Lemma 4.4. Since the only compact subgroup of $\mathbf{Q}_p$ that is normal in B_p is the trivial group, it follows from Lemma 4.4 that the restriction of f to $\mathbf{Q}_p$ is an isomorphism to its closed image.

Let Ω be the normalizer of $f(\mathbf{Q}_p)$ in G; this is a closed subgroup and there is a unique continuous homomorphism $\rho : \Omega \to \mathbf{Z}$ such that conjugation by $g \in \Omega$ on $f(\mathbf{Q}_p)$ multiplies the Haar measure by $p^{\rho(g)}$. In restriction to $\mathbf{Z}$, we see that $\rho \circ f$ is the identity. It follows that $f(\mathbf{Q}_p)$ is open in $f(B_p)$ and that f is proper. □

Proof of Proposition 4.1. Let $f : M \to G$ be an arbitrary continuous homomorphism to a compactly generated l.c. group G. Note that G/G° is a compactly generated totally disconnected l.c. group. Therefore by Proposition 2.1(ii) it has a continuous proper action on a connected graph of degree d, for some d. By Lemma 2.4, for every $p > d$, the restriction to $\mathbf{Q}_p$ of the G-action on this graph has an open kernel. Hence, by Lemma 4.5, for all $p > d$ the action of $\mathbf{Q}_p$ on this graph is trivial.

Let W/G° be the (compact) kernel of the G-action on the graph. Thus W is connected-by-compact and contains $f(\mathbf{Q}_p)$ for all $p > d$. In view of Lemma 4.2, this implies that for all $p > d$, the group $\overline{f(\mathbf{Q}_p)}$ is compact and connected. □

Remark 4.6. Proposition 4.1 shows that there is no injective continuous homomorphism from M to any *totally disconnected* compactly generated l.c. group. On the other hand, M admits an injective continuous homomorphism (not proper!) to a compactly generated l.c. group, which can be obtained as follows: start from the dense embedding $\mathbf{Q} \subset \mathbf{Q}_p$; it induces a dense embedding $\mathbf{Q}_p \subset \widehat{\mathbf{Q}}$, where $\widehat{\mathbf{Q}}$ is the Pontryagin dual of $\mathbf{Q}$ (this is a compact connected group). Multiplication by p is an automorphism of $\mathbf{Q}$ and thus induces a topological automorphism of $\widehat{\mathbf{Q}}$, also given by multiplication by p. So we obtain a continuous injective homomorphism $M \to \prod_p \widehat{\mathbf{Q}} \rtimes Z$, where the p-th component of Z acts on the p-th component of the compact group $\prod_p \widehat{\mathbf{Q}}$ by multiplication by p. By Proposition 1.5, the latter group $\prod_p \widehat{\mathbf{Q}} \rtimes Z$ embeds as a closed subgroup of a compactly generated l.c. group.

5. Some groups of permutations

5A. *A nonembedding criterion.*

Proposition 5.1. *Let H be a topologically simple totally disconnected locally compact group. Assume that H has a compact open subgroup K such that for every k, the group K possesses for some prime $p > k$ a closed subgroup topologically isomorphic to a nontrivial pro-p-group* (*for example*, *K has some element of order p*). *Then H admits no nontrivial continuous homomorphism into any compactly generated locally compact group.*

Proof. We use the criteria from Proposition 2.2 applied to H, in which we can replace "nontrivial" by "faithful" since H is topologically simple. Thus, we only have to show

(1) H has no faithful continuous action on any connected graph of bounded degree;

(2) H has no faithful continuous representation into $\mathrm{GL}_n(\mathbf{C})$ for any n.

Condition (2) is immediate as H has small nontrivial subgroups whereas $\mathrm{GL}_n(\mathbf{C})$ has none.

Now consider a continuous action of H on a connected graph of bounded degree, say $\leq d$. Fix a vertex x_0. Then the stabilizer K_{x_0} of x_0 in K is open, hence of finite index in K. Therefore, the hypothesis implies that K_{x_0}, and hence also the full stabilizer H_{x_0}, contains a nontrivial pro-p-subgroup L for some prime $p > d$. But Lemma 2.4 implies that L acts trivially on the graph, so (1) holds. □

5B. ***Proof of Theorem 1.3.*** Here we prove the continuous case of Theorem 1.3. The case of abstract homomorphisms is postponed to Section 5C.

There exist various sources of topologically simple groups satisfying the criterion of Proposition 5.1 and hence the conclusions of Theorem 1.3. We shall content ourselves with describing one of them, following a construction of Akin, Glasner and Weiss [Akin et al. 2008, §4]; we point out that those examples were independently obtained as part of a more general construction by Willis [2007, §3].

The construction goes as follows: Fix an infinite index set J (we can have $J = \mathbf{N}$ in mind). Fix a family $u = (u_k)_{k\in J}$ of integers greater than 2. Define the graph $\mathcal{G} = \mathcal{G}(u)$ (nonoriented, without self-loops) as a disjoint union of complete graphs $\mathcal{G}_k$ on u_k elements; we denote the vertex set by $\mathcal{G}(u)$ as well. Let us call the *height function* h the function $\mathcal{G} \to J$ mapping any $v \in \mathcal{G}_k$ to k. Note that h completely characterizes the graph structure.

Given a self-map $f : \mathcal{G} \to \mathcal{G}$, we call a vertex $v \in \mathcal{G}_u$ *singular* if $h(f(v)) \neq u$. We call the self-map f *almost regular* if only finitely many vertices are singular. If f is a permutation, we say that f is an *almost automorphism* of the graph with height function $(\mathcal{G}, h)$ if both f and f^{-1} are almost regular. The group of almost automorphisms of $(\mathcal{G}, h)$ is denoted by S (or $S(u)$ if we need specify it). Its subgroup of automorphisms of $(\mathcal{G}, h)$ (consisting of those f that preserve the height and the graph structure) is denoted by K (or $K(u)$).

Note that K is naturally isomorphic to the product $\prod_{k\in J} \mathrm{Sym}(u_k)$, which makes it a compact group. The group S is endowed with the unique left-invariant topology making K a compact open subgroup; this topology is obviously locally compact and is a group topology, as checked in [Akin et al. 2008, §4]. It is the union of an increasing filtering family of compact subgroups (K_F), where F ranges over finite subsets of J and K_F is defined as those elements in S all of whose singularities and pairs of singularities lie in $\bigcup_{i\in F} \mathcal{G}_i$; note that $K_\varnothing = K$ and that K_F is topologically isomorphic to

$$\mathrm{Sym}\Big(\sum_{i\in F} u_i\Big) \times \prod_{k\in J\setminus F} \mathrm{Sym}(u_k).$$

Define K_F^+ as its closed subgroup

$$\mathrm{Alt}\Big(\sum_{i\in F} u_i\Big) \times \prod_{k\in J\setminus F} \mathrm{Alt}(u_k).$$

The filtering family (K_F^+) is increasing; we define S^+ as an abstract group as the union $\bigcup_{n\in F} K_F^+$. Endow it with the left-invariant topology making $K_\varnothing^+$ a compact open subgroup. For the same reason as for S, this is a group topology.

Finally, we define $A < S$ and $A^+ < S^+$ as the subgroups consisting of the *finitary* permutations, i.e., the permutations with finite support. Clearly A is the group of all finitary permutations on the vertices of $\mathscr{G}$, while A^+ is the index-two subgroup of A consisting of the alternating finitary permutations.

Remark 5.2. It is easily seen that A^+ and A are dense as subgroups of S^+ and S, respectively. Moreover, A^+ is also dense in S: indeed, since A is dense, it is enough to show that any transposition $(x\ y)$ in S can be approximated by a sequence of elements of A^+. This is indeed the case, using a sequence of double transpositions $(x\ y)(x_n\ y_n)$ with x_n, y_n distinct elements of the same height k_n and $n \mapsto k_n$ injective.

This implies in particular that the embedding of S^+ into S, which is continuous, is not closed: indeed, its image is a proper subgroup which is dense since it contains A^+.

Remark 5.3. In [Akin et al. 2008], it is shown that S has a dense conjugacy class under the assumption that the mapping $k \mapsto u_k$ has finite fibers (which implies that J is countable). The precise statement of [ibid., Theorem 4.4] actually shows that such a conjugacy class can then be found inside S^+, and also shows that S^+ itself admits a dense conjugacy class.

Let us show the following related but independent result. For the moment, the family (u_k) of integers greater than 2 is arbitrary:

Proposition 5.4. *Every nontrivial normal subgroup of S^+ or S contains A^+, and is thus dense. In particular, S^+ and S are both topologically simple.*

Note that S^+ and S are not abstractly simple, since A^+ is a proper dense normal subgroup in both.

Proof. Let s be a nontrivial element in S^+ or S and $t \in A^+$. Let N be the normal subgroup generated by s; we show that $t \in N$. For some finite subset F of J such that $\sum_{j\in F} u_j \geq 5$, the element s belongs to K_F and t has support in the finite set $X = \bigcup_{i\in F} \mathscr{G}_i$ (which has at least 5 elements). The commutator s' of s and a suitable element of $\mathrm{Alt}(X)$ is a nontrivial element of $N \cap \mathrm{Alt}(X)$. By simplicity of $\mathrm{Alt}(X)$, it follows that $t \in N$. $\square$

We deduce the following corollary, which implies Theorem 1.3:

Corollary 5.5. *If $(u_k)_{k\in J}$ is unbounded, the groups $S(u)^+$ and $S(u)$ admit no nontrivial continuous homomorphism into any compactly generated l.c. group.*

Proof. We have to check that the hypotheses of Proposition 5.1 are fulfilled. The topological simplicity is ensured by Proposition 5.4. The local condition also holds: since (u_k) is unbounded, every neighborhood of the identity contains finite symmetric groups of all orders and thus contains elements of all possible finite orders. □

5C. *Abstract homomorphisms of S and S^+.* We start with the following converse to Corollary 5.5:

Proposition 5.6. *If J is countable and $(u_k)_{k\in J}$ is bounded, then $S(u)^+$ and $S(u)$ are both embeddable as open subgroups in compactly generated l.c. groups, which can be taken as topologically finitely generated.*

Proof. The $\mathscr{G}_k$ form a countable partition of $\mathscr{G}$ by subsets of bounded cardinality, so there exists a permutation σ of $\mathscr{G}$ globally preserving this partition and having finitely many orbits on $\mathscr{G}$. Then σ normalizes S and S^+ as well as K and K^+. Therefore the semidirect products $S \rtimes \langle\sigma\rangle$ and $S^+ \rtimes \langle\sigma\rangle$ are well defined. They are totally disconnected locally compact groups containing S and S^+ respectively as open subgroups. Moreover they act naturally by permutations of $\mathscr{G}$. The subgroup generated by A^+ and σ is finitely generated (when σ is transitive, this group was introduced by B. H. Neumann [1937, p. 127]). Since A^+ is dense in S and S^+, it follows that $S \rtimes \langle\sigma\rangle$ and $S^+ \rtimes \langle\sigma\rangle$ are topologically finitely generated, hence compactly generated. □

Using two theorems of S. Thomas, it is possible to improve Corollary 5.5 in the case where the sequence (u_k) tends to infinity.

Theorem 5.7. *Assume that $k \mapsto u_k$ has finite fibers. Then $S(u)^+$ admits no nontrivial* abstract *homomorphism into any compactly generated l.c. group.*

Proof. Since J is countable, we suppose for convenience $J = \mathbf{N}$. We invoke the criterion from Proposition 2.2, applied to the group $S^+ = S(u)^+$ endowed with the discrete topology. Thus, it is enough to show that

(1) S^+ has no nontrivial action on any connected graph of bounded degree;

(2) S^+ has no nontrivial representation into $\mathrm{GL}_n(\mathbf{C})$.

Both conditions can be checked with the help of the following result: Consider the subgroup $L_k = \prod_{j\geq k} \mathrm{Alt}(u_j)$ of S^+. Observe that S^+ is generated by the alternating finitary group A^+ and L_k (because A^+ is dense), so it follows from Proposition 5.4 that S^+ is normally generated by L_k.

Next, we use a result of Thomas [1999, Theorem 1.10] that every (abstract) subgroup of at most countable index in L_k is open. This immediately shows that every action of S^+ on a graph of at most countable valency is continuous, so (1) follows from the proof of Corollary 5.5 (which, through the proof of Proposition 5.1, shows that S^+ admits no nontrivial continuous action on any connected graph of bounded valency).

Suppose S^+ has a nontrivial linear representation ρ into some $\mathrm{GL}_d(\mathbf{C})$ over a field. Let m_i be the dimension of the smallest nontrivial representation of the alternating group $\mathrm{Alt}(i)$; then m_i tends to infinity (it can be shown that $m_i = i-1$ for $i \geq 7$, but a nice argument based on commutation [Abért 2006] gives a completely elementary lower bound $\simeq \sqrt{i}$). Fix k so that $m_{u_j} > d$ for all $j \geq k$. Since L_k normally generates S^+, the representation ρ is nontrivial in restriction to L_k. By another result of Thomas [1999, Theorem 2.1], any nontrivial subgroup of $\mathrm{GL}_d(\mathbf{C})$ admits a subgroup of at most countable index. Apply this to $\rho(L_k)$ and let H be its inverse image in L_k. By the choice of k, the kernel of ρ contains the direct sum $\bigoplus_{j \geq k} \mathrm{Alt}(u_j)$, which is dense. So H is dense; on the other hand, the first-mentioned result of Thomas implies that H is open. We thus reach a contradiction. ☐

We have seen in Corollary 5.5 that unboundedness of the sequence (u_k) was sufficient to guarantee the absence of nontrivial homomorphisms of $S(u)$ or $S(u)^+$ to a compactly generated locally compact group. In contrast to this, the next result shows that the hypothesis that (u_k) tends to infinity in Theorem 5.7 cannot be weakened to the unboundedness of the sequence:

Proposition 5.8. *The quotient of the group S (resp. S^+) by its subgroup of finitary permutations can be identified with*

$$\prod_{j \in J} \mathrm{Sym}(u_j) \Big/ \bigoplus_{j \in J} \mathrm{Sym}(u_j) \quad \left(\textit{resp.} \prod_{j \in J} \mathrm{Alt}(u_j) \Big/ \bigoplus_{j \in J} \mathrm{Alt}(u_j) \right).$$

In particular:

1. *S has an uncountable abstract abelianization and has proper subgroups of finite index (such subgroups are necessarily dense).*
2. *S^+ has proper subgroups of finite index if and only if $k \mapsto u_k$ has an infinite fiber. It has a nontrivial abelianization if and only if $u^{-1}(\{3,4\})$ is infinite; and in this case the abelianization is uncountable.*

Proof. The first statement follows from the fact that $S = AK$, so $S/A = AK/A = K/(A \cap K)$; the argument for S^+ is similar.

Denoting by C_p the cyclic group of order p, the signature map induces a canonical surjection

$$\prod_{j \in J} \mathrm{Sym}(u_j) \Big/ \bigoplus_{j \in J} \mathrm{Sym}(u_j) \to C_2^J / C_2^{(J)},$$

proving that S has an uncountable abelianization, and, by taking a suitable quotient, admits subgroups of index 2, proving (1). (Observe that S^+ has index 2 in the kernel of the surjection $S \to C_2^J / C_2^{(J)}$.)

Recall that $u_k \geq 3$ for all k. Concerning S^+, if u has finite fibers, then, by Theorem 5.7, S^+ has no nontrivial linear representation, and therefore has no proper subgroup of finite index. Also observe that if $F = u^{-1}(\{3,4\})$ is finite, then S^+ is generated by the perfect groups A^+ and $\prod_{j \notin F} \mathrm{Alt}(u_j)$ and thus is perfect.

Conversely, assume u has a finite fiber $u^{-1}(\{m\})$. We obtain a surjective homomorphism

$$S^+ \to \mathrm{Alt}(m)^J / \mathrm{Alt}(m)^{(J)}.$$

Taking the limit with respect to a nonprincipal ultrafilter yields a nontrivial finite quotient. Also, if $m \in \{3,4\}$, then S^+ admits either $\mathrm{Alt}(3)^J / \mathrm{Alt}(3)^{(J)}$ or $\mathrm{Alt}(4)^J / \mathrm{Alt}(4)^{(J)}$ as a quotient, and thus admits either $C_2^J / C_2^{(J)}$ or $C_3^J / C_3^{(J)}$ as an uncountable abelian quotient. □

6. Proof of Theorem 1.4

Our first example is the following: let $\widehat{\mathbf{Q}}$ be the Pontryagin dual of the discrete additive group $\mathbf{Q}$. So $\widehat{\mathbf{Q}}$ is a connected torsion-free compact group, and by Pontryagin duality, its automorphism group can be identified with the group of automorphisms of the group $\mathbf{Q}$, namely, the multiplicative group $\mathbf{Q}^\times$. The first example is then

$$H_1 = \widehat{\mathbf{Q}} \rtimes \Lambda,$$

where Λ is an arbitrary infinitely generated subgroup of $\mathbf{Q}^\times$ endowed with the discrete topology (recall that $\mathbf{Q}^\times$ is isomorphic to the product of its subgroup of order 2 and a free abelian group of countable rank).

Our second example is very similar in construction. Fix a prime p. Recall that the group $\mathbf{Z}_p^\times$ is uncountable (it is known to be isomorphic to the product of a finite abelian group with $\mathbf{Z}_p$). Let Λ be a countable infinitely generated subgroup of $\mathbf{Z}_p^\times$, and endow Λ with the discrete topology. Our second example is

$$H_2 = \mathbf{Z}_p \rtimes \Lambda.$$

A third example is

$$H_3 = \mathbf{R} \rtimes \Lambda,$$

where Λ is a countable infinitely generated subgroup of $\mathbf{R}^\times$; this is a Lie group.

Proposition 6.1. *If an l.c. group G admits an isomorphic copy of H_i $(i = 1,2,3)$ as an open subgroup, then it admits a discrete quotient that is an infinitely generated abelian group. In particular, G is not compactly generated.*

Proof. Since the identity component $H^\circ = \widehat{\mathbf{Q}}$ is open in H, it is open in G and thus $H^\circ = G^\circ$ is open and normal in G. Thus the action by conjugation of G on $G^\circ = \widehat{\mathbf{Q}}$ defines a continuous homomorphism $\phi : G \to \mathbf{Q}^\times$ that is the identity on Λ and trivial on G°. So $\mathrm{Ker}(\phi)$ is open and the image of ϕ contains Λ and is thus an infinitely generated abelian group. This concludes the proof of the proposition for H_1. The proof for H_3 is similar.

Let us now deal with H_2; since $\mathbf{Z}_p$ is not connected, the previous argument does not work. The subgroup $\mathbf{Z}_p$ being compact and open in G, it is commensurated by G; its abstract topological commensurator is the group $\mathbf{Q}_p^\times$, so G naturally admits a continuous homomorphism to $\mathbf{Q}_p^\times$ whose kernel contains $\mathbf{Z}_p$. Let us check this directly: observe that if $g \in G$, then there exists n such that $g(p^n\mathbf{Z}_p)g^{-1} \subset \mathbf{Z}_p$, and then there exists a unique $\lambda(g)$, not depending on n, such that the conjugation by g, in restriction to $p^n\mathbf{Z}_p$, coincides with the multiplication by $\lambda(g)$. An immediate verification shows that λ is a homomorphism. In restriction to Λ, the map λ is the identity, and $\mathrm{Ker}(\lambda)$ is open in G since it contains $\mathbf{Z}_p$. Thus $G/\mathrm{Ker}(\lambda)$ is a discrete abelian group containing Λ and therefore fails to be finitely generated. □

In order to conclude the proof of Theorem 1.4, it remains to show that those examples admit embeddings as closed subgroups into some compactly generated l.c. groups.

For H_2, such an embedding can be obtained as follows. First embed Λ into a finitely generated group Γ (this is possible by Lemma 3.1). Thus Λ can be diagonally embedded as a discrete subgroup into $\mathbf{Z}_p^\times \times \Gamma$. This embedding extends to a continuous embedding of $\mathbf{Z}_p \rtimes \Lambda$ into $(\mathbf{Z}_p \rtimes \mathbf{Z}_p^\times) \times \Gamma$. This second embedding is continuous and injective; moreover, it is proper, since it is a discrete embedding in restriction to the cocompact subgroup Λ.

An obvious similar construction works for the third example $\mathbf{R} \rtimes \Lambda$. However, both embeddings rely on the fact that Λ is contained in a compactly generated l.c. group of automorphisms of the normal subgroup (either $\mathbf{Z}_p$ or $\mathbf{R}$). For H_1, we use the following topological version of a classical theorem of Krasner and Kaloujnine [1951]:

Recall that given two groups K and Q, the *unrestricted wreath product* $K \bar{\wr} Q$ is the semidirect product $K^Q \rtimes Q$, where Q acts on K^Q by shifting on the left, namely, $q \cdot f(r) = f(q^{-1}r)$. Assume now that K is a topological group and Q is a discrete group. Then the product topology on $K^Q \times Q$ makes $K \bar{\wr} Q$ a topological group.

Theorem 6.2. *For every l.c. group H that is an extension of a compact normal subgroup K by a discrete quotient Q, there is an embedding of H as a closed subgroup of the unrestricted wreath product $K \bar{\wr} Q = K^Q \rtimes Q$.*

Proof of Proposition 1.5. Let Γ be a finitely generated group containing Q. By Theorem 6.2, there is a closed embedding $H \leqslant K \bar{\wr} Q$. By the definition of the unrestricted wreath product, the embedding $Q \leqslant \Gamma$ extends to a closed embedding $K \bar{\wr} Q \leqslant K \bar{\wr} \Gamma$. □

Proof of Theorem 6.2. We begin by a general construction, not relying on the group topologies. Let $\pi : H \to Q$ be a surjective group homomorphism with kernel K. We will define, in a canonical way, a set $X = X(\pi)$ with commuting actions of $K \bar{\wr} Q$ and H such that the $(K \bar{\wr} Q)$-action is simply transitive and the H-action is free. Given a choice of $x \in X$, this yields a unique injective homomorphism $F_x : H \to K \bar{\wr} Q$ mapping $h \in H$ to the unique element $s = F_x(h) \in K \bar{\wr} Q$ such that $hx = s^{-1}x$. The latter homomorphism depends on the choice of x but is canonically defined up to postcomposition by inner automorphisms of $K \bar{\wr} Q$.

The set X is defined to be the set of functions $f : Q \to H$ such that $\pi \circ f$ is a left translation of Q by some element $\theta(f)$. Note that $X \neq \varnothing$; indeed, it contains the set of set-theoretic sections $Q \to H$ of π, which are the elements f in X such that $\theta(f) = 1$.

Let K^Q act on X as follows. If $u \in K^Q$, define

$$u \cdot f(q) = f(q)u(q)^{-1}.$$

If $f \in X$, then $u \cdot f \in X$ and $\theta(u \cdot f) = \theta(f)$, because

$$\pi \circ (u \cdot f)(q) = \pi(f(q)u(q)^{-1}) = \pi(f(q)) = \theta(f)q.$$

This is clearly an action.

Further, let Q act on X as follows. If $r \in Q$, define

$$r \cdot f(q) = f(r^{-1}q).$$

Note that $\pi(r \cdot f(q)) = \pi(f(r^{-1}q)) = \theta(f)r^{-1}q$, so $r \cdot f \in X$ and $\theta(r \cdot f) = \theta(f)r^{-1}$.

We next claim that these actions define an action of the semidirect product $K \bar{\wr} Q$ on X. To verify the claim, we need to show that for all $f \in X$, $u \in K^Q$ and $r \in Q$, we have

$$v \cdot f = r \cdot (u \cdot (r^{-1} \cdot f)),$$

where $v \in K^Q$ is defined as $v : q \mapsto u(r^{-1}q)$. In other words, we have $v = rur^{-1}$ in the wreath product $K \bar{\wr} Q$. Given $q \in Q$, we have

$$v \cdot f(q) = f(q)v(q)^{-1} = f(q)u(r^{-1}q)^{-1}.$$

On the other hand, we have

$$\begin{aligned} r \cdot (u \cdot (r^{-1} \cdot f))(q) &= (u \cdot (r^{-1} \cdot f))(r^{-1}q) \\ &= (r^{-1} \cdot f)(r^{-1}q)u(r^{-1}q)^{-1} \\ &= f(rr^{-1}q)u(r^{-1}q)^{-1} \\ &= f(q)u(r^{-1}q)^{-1}, \end{aligned}$$

so $v \cdot f(q) = r \cdot (u \cdot (r^{-1} \cdot f))(q)$ for all $q \in Q$, as desired.

A straightforward verification shows that the action of $K \bar{\wr} Q$ on X that has just been defined is simply transitive.

Finally, the H-action on X is defined as follows: if $g \in H$ and f is a function $Q \to H$, define

$$g \cdot f(q) = gf(q).$$

If $f \in X$ and $g \in H$ and $q \in Q$, we have

$$(\pi \circ (g \cdot f))(q) = \pi((g \cdot f)(q)) = \pi(gf(q)) = \pi(g)\pi(f(q)) = \pi(g)\theta(f)q,$$

so $g \cdot f \in X$ and $\theta(g \cdot f) = \pi(g)\theta(f)$.

We immediately see that the action of H, which is free, commutes with both the action of K^Q and the action of Q, and thus commutes with the action of $K \bar{\wr} Q$. So, we have for $x \in X$ an injective homomorphism $F_x : H \to K \bar{\wr} Q$ as defined above.

Assume now that K is a topological group, while Q is still assumed to be discrete, so $K \bar{\wr} Q$ is a topological group. Endow H^Q with the product topology, and endow $X \subset H^Q$ with the topology induced by inclusion, namely, the pointwise convergence topology. It is straightforward that the actions of $K \bar{\wr} Q$ and H on X are continuous and that orbital maps $K \bar{\wr} Q \to X$ are homeomorphisms. It follows that the homomorphism F_x is continuous.

Let us now assume that K is compact, so $K \bar{\wr} Q$ and X are both locally compact. (As soon as Q is infinite, the converse holds; namely, $K \bar{\wr} Q$ is locally compact if and only if K is compact.) We claim that the homomorphism F_x is then proper. Checking this amounts to verifying that the H-action on X is proper: Let U_1, U_2 be nonempty compact subsets of X and let us check that $I = \{g \in H : gU_1 \subset U_2\}$ has compact closure. By compactness, $\theta(U_2)$ is finite, and therefore we deduce that $\pi(I)$ is finite. Since π is proper, it follows that I has compact closure. $\square$

Acknowledgment

We thank Pierre de la Harpe for discussions and in particular for pointing out the reference [Pestov 1986].

References

[Abels 1974] H. Abels, "Specker–Kompaktifizierungen von lokal kompakten topologischen Gruppen", *Math. Z.* **135**:4 (1974), 325–361. MR 49 #9114 Zbl 0275.22011

[Abért 2006] M. Abért, "Representing graphs by the non-commuting relation", *Publ. Math. Debrecen* **69**:3 (2006), 261–269. MR 2008j:05162 Zbl 1121.05056

[Akin et al. 2008] E. Akin, E. Glasner, and B. Weiss, "Generically there is but one self homeomorphism of the Cantor set", *Trans. Amer. Math. Soc.* **360**:7 (2008), 3613–3630. MR 2008m:22009 Zbl 1144.22007

[Hall 1954] P. Hall, “Finiteness conditions for soluble groups”, *Proc. London Math. Soc.* (3) **4** (1954), 419–436. MR 17,344c Zbl 0056.25603

[Higman et al. 1949] G. Higman, B. H. Neumann, and H. Neumann, “Embedding theorems for groups”, *J. London Math. Soc.* **24** (1949), 247–254. MR 11,322d Zbl 0034.30101

[Krasner and Kaloujnine 1951] M. Krasner and L. Kaloujnine, “Produit complet des groupes de permutations et problème d’extension de groupes, III”, *Acta Sci. Math. Szeged* **14** (1951), 69–82. MR 14,242d Zbl 0045.30301

[Monod 2001] N. Monod, *Continuous bounded cohomology of locally compact groups*, Lecture Notes in Mathematics **1758**, Springer, Berlin, 2001. MR 2002h:46121 Zbl 0967.22006

[Montgomery and Zippin 1955] D. Montgomery and L. Zippin, *Topological transformation groups*, Interscience, New York, 1955. MR 17,383b Zbl 0068.01904

[Neumann 1937] B. H. Neumann, “Some remarks on infinite groups”, *J. London Math. Soc.* **1**:2 (1937), 120–127. Zbl 0016.29501

[Neumann and Neumann 1959] B. H. Neumann and H. Neumann, “Embedding theorems for groups”, *J. London Math. Soc.* **34** (1959), 465–479. MR 29 #1267 Zbl 0563.33002

[Pestov 1986] V. G. Pestov, “On compactly generated topological groups”, *Mat. Zametki* **40**:5 (1986), 671–676, 699. In Russian; translated in *Math. Notes* **40**:5 (1986), 880–882. MR 88j:22002 Zbl 0617.22001

[Schupp 1976] P. E. Schupp, “Embeddings into simple groups”, *J. London Math. Soc.* (2) **13**:1 (1976), 90–94. MR 53 #5758 Zbl 0363.20026

[Thomas 1999] S. Thomas, “Infinite products of finite simple groups, II”, *J. Group Theory* **2**:4 (1999), 401–434. MR 2000m:20043 Zbl 0938.20025

[Willis 2007] G. A. Willis, “Compact open subgroups in simple totally disconnected groups”, *J. Algebra* **312**:1 (2007), 405–417. MR 2008d:22005 Zbl 1119.22005

Received November 28, 2012. Revised May 18, 2013.

PIERRE-EMMANUEL CAPRACE
DEPARTEMENT DE MATHEMATIQUES
UNIVERSITE CATHOLIQUE DE LOUVAIN, IRMP
CHEMIN DU CYCLOTRON 2
1348 LOUVAIN-LA-NEUVE
BELGIUM

pe.caprace@uclouvain.be

YVES CORNULIER
LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ PARIS-SUD 11
BÂTIMENT 425
91405 ORSAY
FRANCE

yves.cornulier@math.u-psud.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.323

VARIATIONAL REPRESENTATIONS FOR N-CYCLICALLY MONOTONE VECTOR FIELDS

ALFRED GALICHON AND NASSIF GHOUSSOUB

Given a convex bounded domain Ω in $\mathbb{R}^d$ and an integer $N \geq 2$, we associate to any *jointly N-monotone* $(N-1)$-tuplet $(u_1, u_2, \dots, u_{N-1})$ of vector fields from Ω into $\mathbb{R}^d$ a Hamiltonian H on $\mathbb{R}^d \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d$ that is concave in the first variable, jointly convex in the last $N-1$ variables, and such that

$$(u_1(x), u_2(x), \dots, u_{N-1}(x)) = \nabla_{2,\dots,N} H(x, x, \dots, x)$$

for almost all $x \in \Omega$. Moreover, H is N-antisymmetric in a sense made precise later, and also N-sub-antisymmetric in the sense that for all $X \in \Omega^N$ the sum $\sum_{i=0}^{N-1} H(\sigma^i(X)) \leq 0$ is nonpositive, σ being the permutation that shifts the coordinates of X leftward one slot and places the first coordinate last. This result can be seen as an extension of a theorem of E. Krauss, which associates to any monotone operator a concave-convex antisymmetric saddle function. We also give various variational characterizations of vector fields that are almost everywhere N-monotone, showing that they are dual to the class of measure-preserving N-involutions on Ω.

1. Introduction

Given a domain Ω in $\mathbb{R}^d$, recall that a single-valued map u from Ω to $\mathbb{R}^d$ is said to be *N-cyclically monotone* if for every cycle $x_1, \dots, x_N, x_{N+1} = x_1$ of points in Ω, one has

$$\sum_{i=1}^{N} \langle u(x_i), x_i - x_{i+1} \rangle \geq 0. \tag{1}$$

A classical theorem of Rockafellar [Phelps 1993] states that a map u from Ω to $\mathbb{R}^d$

Galichon's research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ ERC grant agreement n° 313699. Support from FiME, Laboratoire de Finance des Marchés de l'Énergie (http://www.fime-lab.org) is gratefully acknowledged. Ghoussoub is partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
MSC2010: 49J40.
Keywords: N-cyclically monotone vector fields, Krauss theorem.

is *N-cyclically monotone for every $N \geq 2$* if and only if

$$u(x) \in \partial\phi(x) \quad \text{for all } x \in \Omega, \tag{2}$$

where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a convex function. On the other hand, a result of E. Krauss [1985] yields that u is a monotone map, i.e., a 2-cyclically monotone map, if and only if

$$u(x) \in \partial_2 H(x, x) \quad \text{for all } x \in \Omega, \tag{3}$$

where H is a concave-convex antisymmetric Hamiltonian on $\mathbb{R}^d \times \mathbb{R}^d$, and $\partial_2 H$ is the subdifferential of H as a convex function in the second variable.

In this paper, we extend the result of Krauss to the class of N-cyclically monotone vector fields, where $N \geq 3$. We shall give a representation for a family of $N-1$ vector fields, which may or may not be individually N-cyclically monotone. Here is the needed concept.

Definition 1. Let $u_1, \dots, u_{N-1}$ be bounded vector fields from a domain $\Omega \subset \mathbb{R}^d$ into $\mathbb{R}^d$. We shall say that the $(N-1)$-tuple $(u_1, u_2, \dots, u_{N-1})$ is *jointly N-monotone* if for every cycle $x_1, \dots, x_{2N-1}$ of points in Ω such that $x_{N+i} = x_i$ for $1 \leq i \leq N-1$, one has

$$\sum_{i=1}^{N} \sum_{\ell=1}^{N-1} \langle u_l(x_i), x_i - x_{l+i} \rangle \geq 0. \tag{4}$$

Examples of jointly N-monotone families of vector fields:

- It is clear that $(u, 0, 0, \dots, 0)$ is jointly N-monotone if and only if u is N-monotone.
- More generally, if each u_ℓ is N-monotone, then the family $(u_1, u_2, \dots, u_{N-1})$ is jointly N-monotone. Actually, one only needs that for $1 \leq \ell \leq N-1$, the vector field u_ℓ be *(N, ℓ)-monotone* in the following sense: for every cycle $x_1, \dots, x_{N+\ell}$ of points in Ω such that $x_{N+i} = x_i$ for $1 \leq i \leq \ell$, we have

$$\sum_{i=1}^{N} \langle u_\ell(x_i), x_i - x_{\ell+i} \rangle \geq 0. \tag{5}$$

  This notion is sometimes weaker than N-monotonicity since if ℓ divides N, then it suffices for u to be N/ℓ -monotone in order to be an (N, ℓ)-monotone vector field. For example, if u_1 and u_3 are 4-monotone operators and u_2 is 2-monotone, then the triplet (u_1, u_2, u_3) is jointly 4-monotone.
- Another example is if (u_1, u_2, u_3) are vector fields such that u_2 is 2-monotone and

$$\langle u_1(x) - u_3(y), x - y \rangle \geq 0 \quad \text{for every } x, y \in \mathbb{R}^d.$$

In this case, the triplet (u_1, u_2, u_3) is jointly 4-monotone. In particular, if u_1 and u_2 are both 2-monotone, then the triplet (u_1, u_2, u_1) is jointly 4-monotone.

- More generally, it is easy to show that $(u, u, \dots, u)$ is jointly N-monotone if and only if u is 2-cyclically monotone.

We shall always denote by σ the cyclic permutation on $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ defined by

$$\sigma(x_1, x_2, \dots, x_{N-1}, x_N) = (x_2, x_3, \dots, x_N, x_1).$$

We let

$$\mathcal{H}_N(\Omega) = \left\{ H \in C(\Omega^N) : \sum_{i=0}^{N-1} H(\sigma^i(x_1, \dots, x_N)) = 0 \right\} \tag{6}$$

be the family of continuous Hamiltonians on Ω^N that are *N-antisymmetric*, i.e., satisfy the condition to the right of the colon in (6). We say that H is *N-sub-antisymmetric on* Ω if

$$\sum_{i=0}^{N-1} H(\sigma^i(x_1, \dots, x_N)) \le 0 \quad \text{on } \Omega^N. \tag{7}$$

We shall also say that a function F of two variables is *N-cyclically sub-anti-symmetric on* Ω if

$$\begin{aligned} &F(x, x) = 0 \quad \text{and} \\ &\sum_{i=1}^{N} F(x_i, x_{i+1}) \le 0 \quad \text{for all cyclic families } x_1, \dots, x_N, x_{N+1} = x_1 \text{ in } \Omega. \end{aligned} \tag{8}$$

Note that if a function $H(x_1, \dots, x_N)$ N-sub-antisymmetric and if it only depends on the first two variables, then the function $F(x_1, x_2) := H(x_1, x_2, \dots, x_N)$ is N-cyclically sub-antisymmetric.

We associate to any function H on Ω^N the functional given by on $\Omega \times (\mathbb{R}^d)^{N-1}$

$$L_H(x, p_1, \dots, p_{N-1}) = \sup \left\{ \sum_{i=1}^{N-1} \langle p_i, y_i \rangle - H(x, y_1, \dots, y_{N-1}) : y_i \in \Omega \right\}. \tag{9}$$

Note that if Ω is convex and if H is convex in the last $N-1$ variables, then L_H is nothing but the Legendre transform of $\tilde{H}$ with respect to the last $N-1$ variables, where $\tilde{H}$ is the extension of H over $(\mathbb{R}^d)^N$, defined by $\tilde{H} = H$ on Ω^N and $\tilde{H} = +\infty$ outside Ω^N. Since $H(x, \dots, x) = 0$ for any $H \in \mathcal{H}_N(\Omega)$, we have, for any such H,

$$L_H(x, p_1, \dots, p_{N-1}) \ge \sum_{i=1}^{N-1} \langle x, p_i \rangle, \tag{10}$$

for $x \in \Omega$ and $p_1, \dots, p_{N-1} \in \mathbb{R}^d$. To formulate variational principles for such vector fields, we shall consider the class of σ-invariant probability measures on Ω^N, which are those $\pi \in \mathcal{P}(\Omega^N)$ such that for all $h \in L^1(\Omega^N, d\pi)$, we have

$$\int_{\Omega^N} h(x_1, \dots, x_N)\, d\pi = \int_{\Omega^N} h(\sigma(x_1, \dots, x_N))\, d\pi. \tag{11}$$

We set

$$\mathcal{P}_{\text{sym}}(\Omega^N) = \{\pi \in \mathcal{P}(\Omega^N) : \pi \ \sigma\text{-invariant probability on } \Omega^N\}. \tag{12}$$

For a given probability measure μ on Ω, we also consider the class

$$\mathcal{P}^{\mu}_{\text{sym}}(\Omega^N) = \{\pi \in \mathcal{P}_{\text{sym}}(\Omega^N) : \text{proj}_1 \pi = \mu\}, \tag{13}$$

i.e., the set of all $\pi \in \mathcal{P}_{\text{sym}}(\Omega^N)$ with a given first marginal μ, meaning that

$$\int_{\Omega^N} f(x_1)\, d\pi(x_1, \dots, x_N) = \int_{\Omega} f(x_1)\, d\mu(x_1) \quad \text{for every } f \in L^1(\Omega, \mu). \tag{14}$$

Now consider the set $\mathcal{S}(\Omega, \mu)$ of μ-measure-preserving transformations on Ω, which can be identified with a closed subset of the sphere of $L^2(\Omega, \mathbb{R}^d)$. We shall also consider the subset of $\mathcal{S}(\Omega, \mu)$ consisting of N-involutions, that is,

$$\mathcal{S}_N(\Omega, \mu) = \{S \in \mathcal{S}(\Omega, \mu) : S^N = I \ \mu\text{-a.e.}\}.$$

2. Monotone vector fields and N-antisymmetric Hamiltonians

In this section, we establish the following extension of a theorem of Krauss.

Theorem 2. *Let $N \geq 2$ be an integer, and let $u_1, \dots, u_{N-1}$ be bounded vector fields from a convex domain $\Omega \subset \mathbb{R}^d$ into $\mathbb{R}^d$.*

1) *If the $(N-1)$-tuple $(u_1, \dots, u_{N-1})$ is jointly N-monotone, then there exists an N-sub-antisymmetric Hamiltonian H that is zero on the diagonal of Ω^N, concave in the first variable, convex in the other $N-1$ variables, and such that*

$$(u_1(x), \dots, u_{N-1}(x)) = \nabla_{2,\dots,N} H(x, x, \dots, x) \quad \textit{for a.e. } x \in \Omega. \tag{15}$$

Moreover, H is N-antisymmetric in the sense that

$$H(x_1, x_2, \dots, x_N) + H_{2,\dots,N}(x_1, x_2, \dots, x_N) = 0, \tag{16}$$

where $H_{2,\dots,N}$ is the concavification of the function $K(\boldsymbol{x}) = \sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x}))$ with respect to the last $N-1$ variables.

Furthermore, there exists a continuous N-antisymmetric Hamiltonian $\overline{H}$ on Ω^N, such that

$$L_{\overline{H}}(x, u_1(x), u_2(x), \dots, u_{N-1}(x)) = \sum_{i=1}^{N-1} \langle u_i(x), x\rangle \quad \textit{for all } x \in \Omega. \tag{17}$$

2) *Conversely, if* $(u_1, \dots, u_{N-1})$ *satisfies* (15) *for some* N*-sub-antisymmetric Hamiltonian* H *that is zero on the diagonal of* Ω^N, *concave in the first variable, and convex in the other variables, then the* $(N-1)$*-tuple* $(u_1, \dots, u_{N-1})$ *is jointly* N*-monotone on* Ω.

Remark 3. In the case $N = 2$, $K(\boldsymbol{x}) = H(x_2, x_1)$ is concave with respect to x_2, hence $H_2(x_1, x_2) = H(x_2, x_1)$, and (16) becomes

$$H(x_1, x_2) + H(x_2, x_1) = 0;$$

thus H is antisymmetric, recovering well-known results [Krauss 1985; Ghoussoub 2009; Ghoussoub and Moameni 2013a; Millien 2011].

Lemma 4. *Assume the* $(N-1)$*-tuple of bounded vector fields* $(u_1, \dots, u_{N-1})$ *on* Ω *is jointly* N*-monotone. Define*

$$f(x_1, \dots, x_N) := \sum_{l=1}^{N-1} \langle u_l(x_1), x_1 - x_{l+1} \rangle$$

and let $\tilde{f}$ *be the convexification of* f *with respect to the first variable, given by*

$$\begin{aligned} &\tilde{f}(x_1, x_2, \dots, x_N) \\ &= \inf\Big\{\sum_{k=1}^{n} \lambda_k f(x_1^k, x_2, \dots, x_N) : n \in \mathbb{N},\ \lambda_k \geq 0,\ \sum_{k=1}^{n} \lambda_k = 1,\ \sum_{k=1}^{n} \lambda_k x_1^k = x_1\Big\}. \end{aligned} \tag{18}$$

1) *We have* $f \geq \tilde{f}$ *on* Ω^N.
2) $\tilde{f}$ *is convex in the first variable and concave with respect to the other variables.*
3) $\tilde{f}(x, x, \dots, x) = 0$ *for each* $x \in \Omega$.
4) $\tilde{f}$ *satisfies*

$$\sum_{i=0}^{N-1} \tilde{f}(\sigma^i(x_1, \dots, x_N)) \geq 0 \quad \textit{on } \Omega^N. \tag{19}$$

Proof. Since the $(N-1)$-tuple $(u_1, \dots, u_{N-1})$ is jointly N-monotone, it is easy to see that the function

$$f(x_1, \dots, x_N) := \sum_{l=1}^{N-1} \langle u_l(x_1), x_1 - x_{l+1} \rangle$$

is linear in the last $N-1$ variables, that $f(x, x, \dots, x) = 0$, and that

$$\sum_{i=0}^{N-1} f(\sigma^i(x_1, \dots, x_N)) \geq 0 \quad \text{on } \Omega^N. \tag{20}$$

It is also clear that $f \geq \tilde{f}$, that $\tilde{f}$ is convex with respect to the first variable x_1, and that it is concave with respect to the other variables $x_2, \dots, x_N$, since f itself is concave (actually linear) with respect to $x_2, \dots, x_N$. We now show that $\tilde{f}$ satisfies (19).

For that, we fix $x_1, x_2, \dots, x_N$ in Ω and consider $(x_1^k)_{k=1}^n$ in Ω, and $(\lambda_k)_k$ in $\mathbb{R}$ such that $\lambda_k \geq 0$ such that $\sum_{k=1}^n \lambda_k = 1$ and $\sum_{k=1}^n \lambda_k x_1^k = x_1$. For each k, we have

$$f(x_1^k, x_2, \dots, x_N) + f(x_2, \dots, x_N, x_1^k) + \cdots + f(x_N, x_1^k, x_2, \dots, x_{N-1}) \geq 0.$$

Multiplying by λ_k, summing over k, and using that f is linear in the last $N-1$ variables, we have

$$\sum_{k=1}^n \lambda_k f(x_1^k, x_2, \dots, x_N) + f(x_2, \dots, x_N, x_1) + \cdots + f(x_N, x_1, x_2, \dots, x_{N-1}) \geq 0.$$

By taking the infimum, we obtain

$$\tilde{f}(x_1, x_2, \dots, x_N) + \sum_{i=1}^{N-1} f(\sigma^i(x_1, x_2, \dots, x_N)) \geq 0.$$

Let now $n \in \mathbb{N}$, $\lambda_k \geq 0$, $x_N^k \in \Omega$ be such that $\sum_{k=1}^n \lambda_k = 1$ and $\sum_{k=1}^n \lambda_k x_2^k = x_2$. For every $1 \leq k \leq n$, we have

$$\tilde{f}(x_1, x_2^k, x_3, \dots, x_N) + f(x_2^k, x_3, \dots, x_1) + \cdots + f(x_N, x_1, x_2^k, x_3, \dots, x_{N-1}) \geq 0.$$

Multiplying by λ_k, summing over k and using that $\tilde{f}$ is convex in the first variable and f is linear in the last $N-1$ variables, we obtain

$$\begin{aligned}
&\tilde{f}(x_1, x_2, x_3, \dots, x_N) + \sum_{k=1}^n \lambda_k f(x_2^k, x_3, \dots, x_1) + \cdots + f(x_N, x_1, x_2, x_3, \dots, x_{N-1}) \\
&\quad \geq \sum_{k=1}^n \lambda_k \tilde{f}(x_1, x_2^k, x_3, \dots, x_N) + \sum_{k=1}^n \lambda_k f(x_2^k, x_3, \dots, x_1) \\
&\qquad\qquad + \cdots + \sum_{k=1}^n \lambda_k f(x_N, x_1, x_2^k, x_3, \dots, x_{N-1}) \\
&\quad \geq 0.
\end{aligned}$$

By taking the infimum over all possible such choices, we get

$$\tilde{f}(x_1, x_2, x_3, \dots, x_N) + \tilde{f}(x_2, x_3, \dots, x_1) + \cdots + f(x_N, x_1, x_2, x_3, \dots, x_{N-1}) \geq 0.$$

By repeating this procedure with $x_3, \dots, x_{N-1}$, we get

$$\sum_{i=0}^{N-2} \tilde{f}(\sigma^i(x_1, x_2, \dots, x_N)) + f(x_N, x_1, x_2, x_3, \dots, x_{N-1}) \geq 0.$$

Finally, since

$$f(x_N, x_1, x_2, x_3, \dots, x_{N-1}) \geq -\sum_{i=0}^{N-2} \tilde{f}(\sigma^i(x_1, x_2, \dots, x_N))$$

and since $\tilde{f}$ is concave in the last $N-1$ variables, the function

$$x_N \to -\sum_{i=0}^{N-2} \tilde{f}(\sigma^i(x_1, x_2, \dots, x_N))$$

for fixed $x_1, x_2, \dots, x_{N-1}$ is a convex minorant of $x_N \to f(x_N, x_1, x_2, \dots, x_{N-1})$. It follows that

$$\begin{aligned} f(x_N, x_1, x_2, x_3, \dots, x_{N-1}) &\geq \tilde{f}(x_N, x_1, x_2, x_3, \dots, x_{N-1}) \\ &\geq -\sum_{i=0}^{N-2} \tilde{f}(\sigma^i(x_1, x_2, \dots, x_N)), \end{aligned}$$

which yields $\sum_{i=0}^{N-1} \tilde{f}(\sigma^i(x_1, x_2, \dots, x_N)) \geq 0$. This implies that $\tilde{f}(x, x, \dots, x) \geq 0$ for $x \in \Omega$.

On the other hand, since $\tilde{f}(x, x, \dots, x) \leq f(x, x, \dots, x) = 0$, we get that $\tilde{f}(x, x, \dots, x) = 0$ for all $x \in \Omega$. □

Proof of Theorem 2. Assume the $(N-1)$-tuple of vector fields $(u_1, \dots, u_{N-1})$ is jointly N-monotone on Ω, and consider the function

$$f(x_1, \dots, x_N) := \sum_{l=1}^{N-1} \langle u_l(x_1), x_1 - x_{l+1} \rangle$$

as well as its convexification with respect to the first variable $\tilde{f}(x_1, \dots, x_N)$.

By Lemma 4, the function $\psi(x_1, \dots, x_N) := -\tilde{f}(x_1, \dots, x_N)$ satisfies the following properties:

(i) $x_1 \to \psi(x_1, \dots, x_N)$ is concave.

(ii) $(x_2, x_3, \dots, x_N) \to \psi(x_1, \dots, x_N)$ is convex.

(iii) $\psi(x_1, \dots, x_N) \geq -f(x_1, \dots, x_N) = \sum_{l=1}^{N-1} \langle u_l(x_1), x_{l+1} - x_1 \rangle$.

(iv) ψ is N-sub-antisymmetric.

Now consider the family $\overline{\mathcal{H}}$ of functions $H : \Omega^N \to \mathbb{R}$ such that

1) $H(x_1, x_2, \dots, x_N) \geq \sum_{l=1}^{N-1} \langle u_l(x_1), x_{l+1} - x_1 \rangle$ for every N-tuple $(x_1, \dots, x_N)$ in Ω^N,

2) H is concave in the first variable,

3) H is jointly convex in the last $N-1$ variables,

4) H is N-sub-antisymmetric,

5) H is zero on the diagonal of Ω^N.

Note that $\overline{\mathcal{H}} \neq \varnothing$ since ψ belongs to $\overline{\mathcal{H}}$. Note that any H satisfying conditions 1 and 4 automatically satisfies 5. Indeed, by N-sub-antisymmetry, for all $\boldsymbol{x} = (x_1, \ldots, x_N) \in \Omega^N$ we have

$$H(\boldsymbol{x}) \leq -\sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x})) \leq -\sum_{i=1}^{N-1} \psi(\sigma^i(\boldsymbol{x})). \tag{21}$$

This also yields that

$$\sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_{\ell+1} - x_1 \rangle \leq H(\boldsymbol{x}) \leq -\sum_{i=2}^{N} \sum_{\ell=1}^{N-1} \langle u_\ell(x_i), x_i - x_{i+\ell} \rangle, \tag{22}$$

where we denote $x_{i+N} := x_i$ for $i = 1, \ldots, \ell$. This yields that $H(x, x, \ldots, x) = 0$ for any $x \in \Omega$.

It is also easy to see that every directed family $(H_i)_i$ in $\overline{\mathcal{H}}$ has a supremum $H_\infty \in \overline{\mathcal{H}}$, meaning that $\overline{\mathcal{H}}$ is a Zorn family, and therefore has a maximal element H.

Now consider the function

$$\overline{H}(\boldsymbol{x}) = \frac{1}{N}\left((N-1)H(\boldsymbol{x}) - \sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x}))\right).$$

(i) $\overline{H}$ is N-antisymmetric, since $\overline{H}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N-1}[H(\boldsymbol{x}) - H(\sigma^i(\boldsymbol{x}))]$, and each summand is N-antisymmetric.

(ii) $\overline{H} \geq H$ on Ω^N, since $N[\overline{H}(\boldsymbol{x}) - H(\boldsymbol{x})] = -\sum_{i=0}^{N-1} H(\sigma^i(\boldsymbol{x})) \geq 0$ (because H itself is N-sub-antisymmetric).

The maximality of H would have implied that $H = \overline{H}$ is N-antisymmetric if only $\overline{H}$ was jointly convex in the last $N-1$ variables, but since this is not necessarily the case, we consider for $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ the function

$$K(x_1, x_2, \ldots, x_N) = K(\boldsymbol{x}) := -\sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x})),$$

which is already concave in the first variable x_1. Its convexification in the last $N-1$ variables, that is,

$$K^{2,\ldots,N}(\boldsymbol{x}) = \inf\left\{\sum_{i=1}^{n} \lambda_i K(x_1, x_2^i, \ldots, x_N^i) : \lambda_i \geq 0,\ \sum_{i=1}^{n} \lambda_i (x_2^i, \ldots, x_N^i, 1) = (x_2, \ldots, x_N, 1)\right\},$$

is still concave in the first variable, but is now convex in the last $N-1$ variables. Moreover,

$$H \le K^{2,\dots,N} \le K = -\sum_{i=1}^{N-1} H \circ \sigma^i. \tag{23}$$

Indeed, $K^{2,\dots,N} \le K$ from the definition of $K^{2,\dots,N}$, while $H \le K^{2,\dots,N}$ because $H \le K$ and H is already convex in the last $N-1$ variables. It follows that

$$H \le \frac{(N-1)H + K^{2,\dots,N}}{N} \le \frac{(N-1)H + K}{N} = \frac{1}{N}\Big((N-1)H - \sum_{i=1}^{N-1} H \circ \sigma^i\Big) = \bar{H}.$$

The function $H' = ((N-1)H + K^{2,\dots,N})/N$ belongs to the family $\overline{\mathcal{H}}$ and therefore $H = H'$ by the maximality of H.

This finally yields that H is N-sub-antisymmetric, that $H(x, \dots, x) = 0$ for all $x \in \Omega$ and that

$$H(\boldsymbol{x}) + H_{2,\dots,N}(\boldsymbol{x}) = 0 \quad \text{for every } \boldsymbol{x} \in \Omega^N,$$

where $H_{2,\dots,N} = -K^{2,\dots,N}$, which for a fixed x_1 is nothing but the concavification of $(x_2, \dots, x_N) \to \sum_{i=1}^{N-1} H(\sigma^i(x_1, x_2, \dots, x_N))$.

Note now that since for any $x_1, \dots, x_N$ in Ω

$$H(x_1, x_2, \dots x_N) \ge \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_{\ell+1} - x_1 \rangle, \tag{24}$$

and

$$H(x_1, x_1, \dots, x_1) = 0, \tag{25}$$

we have

$$H(x_1, x_2, \dots, x_N) - H(x_1, \dots, x_1) \ge \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_{\ell+1} - x_1 \rangle. \tag{26}$$

Since H is convex in the last $N-1$ variables, this means that for all $x \in \Omega$, we have

$$(u_1(x), u_2(x), \dots, u_{N-1}(x)) \in \partial_{2,\dots,N} H(x, x, \dots, x), \tag{27}$$

as claimed in (15). This also yields

$$L_H(x, u_1(x), \dots, u_{N-1}(x)) + H(x, x, \dots, x) = \sum_{\ell=1}^{N-1} \langle u_\ell(x), x \rangle \quad \text{for all } x \in \Omega.$$

In other words, $L_H(x, u_1(x), \dots, u_{N-1}(x)) = \sum_{\ell=1}^{N-1} \langle u_\ell(x), x \rangle$ for all $x \in \Omega$. As above, consider

$$\overline{H}(\boldsymbol{x}) = \frac{1}{N}\bigg((N-1)H(\boldsymbol{x}) - \sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x}))\bigg).$$

We have $\overline{H} \in \overline{\mathcal{H}}_N(\Omega)$ and $\overline{H} \geq H$, and therefore $L_{\overline{H}} \leq L_H$. On the other hand, for all $x \in \Omega$ we have

$$\begin{aligned} L_{\overline{H}}(x, u_1(x), \dots, u_{N-1}(x)) &= L_{\overline{H}}(x, u_1(x), \dots, u_{N-1}(x)) + \overline{H}(x, x, \dots, x) \\ &\geq \sum_{\ell=1}^{N-1} \langle u_\ell(x), x\rangle. \end{aligned}$$

To prove (17), we use the appendix in [Ghoussoub and Moameni 2013b] to deduce that for $i = 2, \dots, N$, the gradients $\nabla_i H(x, x, \dots, x)$ actually exist for a.e. x in Ω.

The converse is straightforward since if (27) holds, then (26) does, and since we also have (25), then the property that $(u_1, \dots, u_{N-1})$ is jointly N-monotone follows from (24) and the sub-antisymmetry of H. □

In the case of a single N-monotone vector field, we can obviously apply the above theorem to the $(N-1)$-tuple $(u, 0, \dots, 0)$, which is then N-monotone, to find an N-sub-antisymmetric Hamiltonian H, which is concave in the first variable and convex in the last $N-1$ variables such that

$$(-u(x), u(x), 0, \dots, 0) = \nabla H(x, x, \dots, x) \quad \text{for a.e. } x \in \Omega. \tag{28}$$

However, in this case we can restrict ourselves to N-cyclically sub-antisymmetric functions of two variables and establish the following extension of the theorem of Krauss.

Theorem 5. *If u is N-cyclically monotone on Ω, then there exists a concave-convex function of two variables F that is N-cyclically sub-antisymmetric and zero on the diagonal, such that*

$$(-u(x), u(x)) \in \partial F(x, x) \quad \textit{for all } x \in \Omega, \tag{29}$$

where ∂H is the subdifferential of H as a concave-convex function [Rockafellar 1970]. *Moreover*,

$$u(x) = \nabla_2 F(x, x) \quad \textit{for a.e. } x \in \Omega. \tag{30}$$

Proof. Let $f(x, y) = \langle u(x), x - y\rangle$ and let $f^1(x, y)$ be its convexification in x for fixed y, that is,

$$f^1(x, y) = \inf\bigg\{\sum_{k=1}^{n} \lambda_k f(x_k, y) : \lambda_k \geq 0,\ \sum_{k=1}^{n} \lambda_k = 1,\ \sum_{k=1}^{n} \lambda_k x_k = x\bigg\}. \tag{31}$$

Since $f(x, x) = 0$, f is linear in y, and $\sum_{i=1}^{N} f(x_i, x_{i+1}) \geq 0$ for any cyclic family

$x_1, \dots, x_N, x_{N+1} = x_1$ in Ω, it is easy to show that $f \geq f^1$ on Ω, f^1 is convex in the first variable and concave with respect to the second, $f^1(x, x) = 0$ for each $x \in \Omega$, and that f^1 is N-cyclically supersymmetric in the sense that for any cyclic family $x_1, \dots, x_N, x_{N+1} = x_1$ in Ω, we have $\sum_{i=1}^N f^1(x_i, x_{i+1}) \geq 0$.

Now consider $F(x, y) = -f^1(x, y)$ and note that $x \to F(x, y)$ is concave, $y \to F(x, y)$ is convex, $F(x, y) \geq -f(x, y) = \langle u(x), y - x\rangle$ and F is N-cyclically sub-antisymmetric. By the antisymmetry, we have

$$\langle u(x_1), x_2 - x_1\rangle \leq F(x_1, x_2) \leq \langle u(x_2), x_2 - x_1\rangle, \tag{32}$$

which yields that $(-u(x), u(x)) \in \partial F(x, x)$ for all $x \in \Omega$.

Since F is antisymmetric and concave-convex, the possibly multivalued map $x \to \partial_2 F(x, x)$ is monotone on Ω, and therefore single-valued and differentiable almost everywhere [Phelps 1993]. This completes the proof. $\square$

Remark 6. We cannot expect to have a function F such that $\sum_{i=1}^N F(x_i, x_{i+1}) = 0$ for all cyclic families $x_1, \dots, x_N, x_{N+1} = x_1$ in Ω. Actually, we believe that the only function satisfying such an N-antisymmetry for $N \geq 3$ must be of the form $F(x, y) = f(x) - f(y)$. This is why one needs to consider functions of N variables in order to get N-antisymmetry. In other words, the function defined by

$$H(x_1, x_2, \dots, x_N) := \frac{1}{N}\left((N-1)F(x_1, x_2) - \sum_{i=2}^{N-1} F(x_i, x_{i+1})\right) \tag{33}$$

is N-antisymmetric in the sense of (6) and $H(x_1, x_2, \dots, x_N) \geq F(x_1, x_2)$ for all $(x_1, x_2, \dots, x_N)$ in Ω^N.

3. Variational characterization of monotone vector fields

In order to simplify the exposition, we shall always assume in the sequel that $d\mu$ is Lebesgue measure dx normalized to be a probability on Ω. We shall also assume that Ω is convex and that its boundary has measure zero.

Theorem 7. *Let $u_1, \dots, u_{N-1} : \Omega \to \mathbb{R}^d$ be bounded measurable vector fields. The following properties are then equivalent*:

1) *The $(N-1)$-tuple $(u_1, \dots, u_{N-1})$ is jointly N-monotone a.e.*, *that is*, *there exists a measure-zero set Ω_0 such that $(u_1, \dots, u_{N-1})$ is jointly N-monotone on $\Omega \setminus \Omega_0$.*
2) *The infimum of the Monge–Kantorovich problem*

$$\inf\left\{\int_{\Omega^N} \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_1 - x_{\ell+1}\rangle \, d\pi(x_1, x_2, \dots, x_N)) : \pi \in \mathcal{P}^{\mu}_{\mathrm{sym}}(\Omega^N)\right\} \tag{34}$$

is equal to zero, and is therefore attained by the push-forward of μ by the map $x \to (x, x, \dots, x)$.

3) $(u_1, \dots, u_{N-1})$ *is in the polar of* $\mathcal{S}_N(\Omega, \mu)$ *in the following sense*:

$$\inf\left\{\int_\Omega \sum_{\ell=1}^{N-1} \langle u_\ell(x), x - S^\ell x\rangle \, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\} = 0. \tag{35}$$

4) *The following holds*:

$$\inf\left\{\int_\Omega \sum_{\ell=1}^{N-1} |u_\ell(x) - S^\ell x|^2 \, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\} = \sum_{\ell=1}^{N-1} \int_\Omega |u_\ell(x) - x|^2 \, d\mu. \tag{36}$$

5) *There exists an N-sub-antisymmetric Hamiltonian H which is concave in the first variable, convex in the last $N-1$ variables, and vanishing on the diagonal such that*

$$(u_1(x), \dots, u_{N-1}(x)) = \nabla_{2,\dots,N} H(x, x, \dots, x) \quad \textit{for a.e. } x \in \Omega. \tag{37}$$

Moreover, H is N-symmetric in the sense of (16).

6) *The following duality holds*:

$$\inf\left\{\int_\Omega L_H(x, u_1(x), \dots, u_{N-1}(x)) \, d\mu : H \in \mathcal{H}_N(\Omega)\right\}$$
$$= \sup\left\{\int_\Omega \sum_{\ell=1}^{N-1} \langle u_\ell(x), S^\ell x\rangle \, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\}$$

and the latter is attained at the identity map.

We start with the following lemma, which identifies those probabilities in $\mathcal{P}^\mu_{\mathrm{sym}}(\Omega^N)$ that are carried by graphs of functions from Ω to Ω^N.

Lemma 8. *Let $S : \Omega \to \Omega$ be a μ-measurable map. The following properties are equivalent*:

1) *The image of μ by the map $x \to (x, Sx, \dots, S^{N-1}x)$ belongs to $\mathcal{P}^\mu_{\mathrm{sym}}(\Omega^N)$.*

2) *S is μ-measure-preserving and $S^N(x) = x$ μ-a.e.*

3) *For any bounded Borel measurable N-antisymmetric H on Ω^N, we have* $\int_\Omega H(x, Sx, \dots, S^{N-1}x) \, d\mu = 0$.

Proof. Clearly 1) implies 3), since $\int_{\Omega^N} H(\boldsymbol{x}) \, d\pi(\boldsymbol{x}) = 0$ for any N-antisymmetric Hamiltonian H and any $\pi \in \mathcal{P}^\mu_{\mathrm{sym}}(\Omega^N)$.

That 2) implies 1) is also straightforward since if π is the push-forward of μ by a map of the form $x \to (x, Sx, \dots, S^{N-1}x)$, where S is a μ-measure-preserving S

with $S^N x = x$ μ-a.e. on Ω, then for all $h \in L^1(\Omega^N, d\pi)$, we have

$$\begin{aligned}\int_{\Omega^N} h(x_1, \ldots, x_N)\, d\pi &= \int_\Omega h(x, Sx, \ldots, S^{N-1}x)\, d\mu(x)\\ &= \int_\Omega h(Sx, S^2x, \ldots, S^{N-1}x, S^N x)\, d\mu(x)\\ &= \int_\Omega h(Sx, S^2x, \ldots, S^{N-1}x, x)\, d\mu(x)\\ &= \int_{\Omega^N} h(\sigma(x_1, \ldots, x_N))\, d\pi.\end{aligned}$$

We now prove that 2) and 3) are equivalent. Assuming first that S is μ-measure-preserving such that $S^N = I$ μ-a.e., then for every Borel bounded N-antisymmetric H, we have

$$\begin{aligned}\int_\Omega H(x, Sx, S^2x, \ldots, S^{N-1}x)\, d\mu &= \int_\Omega H(Sx, S^2x, \ldots, S^{N-1}x, x)\, d\mu\\ &= \cdots = \int_\Omega H(S^{N-1}x, x, Sx, \ldots, S^{N-2}x)\, d\mu.\end{aligned}$$

Since H is N-antisymmetric, we can see that

$$\begin{aligned}H(x, Sx, \ldots, S^{N-1}x) + H(Sx, S^2x, \ldots, S^{N-1}x, x)&\\ + \cdots + H(S^{N-1}x, x, Sx, \ldots, S^{N-2}x) &= 0.\end{aligned}$$

It follows that $N\int_\Omega H(x, Sx, S^2x, \ldots, S^{N-1}x)\, d\mu = 0$.

For the reverse implication, assume $\int_\Omega H(x, Sx, S^2x, \ldots, S^{N-1}x)\, d\mu = 0$ for every N-antisymmetric Hamiltonian H. By testing this identity with the Hamiltonians

$$H(x_1, x_2, \ldots, x_N) = f(x_1) - f(x_i),$$

where f is any continuous function on Ω, one gets that S is μ-measure-preserving. Now take the Hamiltonian

$$H(x_1, x_2, \ldots, x_N) = |x_1 - Sx_N| - |Sx_1 - x_2| - |x_2 - Sx_1| + |Sx_2 - x_3|.$$

Note that $H \in \mathcal{H}_N(\Omega)$ since it is of the form

$$H(x_1, \ldots, x_N) = f(x_1, x_2, x_N) - f(x_2, x_3, x_1).$$

Now test the above identity with such an H to obtain

$$0 = \int_\Omega H(x, Sx, S^2x, \ldots, S^{N-1}x)\, d\mu = \int_\Omega |x - SS^{N-1}x|\, d\mu.$$

It follows that $S^N = I$ μ-a.e. on ω, and we are done. $\square$

Proof of Theorem 7. To show that 1) implies 2), it suffices to notice that if π is a σ-invariant probability measure on Ω^N such that $\mathrm{proj}_1\pi = \mu$, then

$$\begin{aligned}\int_{\Omega^N} \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_1 - x_{\ell+1}\rangle \, d\pi(x_1, \dots, x_N) \\ &= \frac{1}{N}\sum_{i=1}^{N}\int_{\Omega^N}\sum_{\ell=1}^{N-1}\langle u_\ell(x_i), x_i - x_{i+\ell}\rangle \, d\pi(x_1,\dots,x_N) \\ &= \frac{1}{N}\int_{\Omega^N}\Bigl(\sum_{i=1}^{N}\sum_{\ell=1}^{N-1}\langle u_\ell(x_i), x_i - x_{i+\ell}\rangle\Bigr)\, d\pi(x_1,\dots,x_N) \\ &\geq 0,\end{aligned}$$

since $(u_1, \dots, u_{N-1})$ is jointly N-monotone. On the other hand, if π is the σ-invariant measure obtained by taking the image of $\mu := dx$ by $x \to (x, \dots, x)$, then

$$\int_{\Omega^N} \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_1 - x_{\ell+1}\rangle \, d\pi(x_1, \dots, x_N) = 0.$$

To show that 2) implies 3), let S be a μ-measure-preserving transformation on Ω such that $S^N = I$ μ-a.e. on Ω. Then the image π_S of μ by the map

$$x \to (x, Sx, S^2x, \dots, S^{N-1}x)$$

is σ-invariant, hence

$$\int_{\Omega^N} \sum_{\ell=1}^{N-1} \langle u_\ell(x_1), x_1 - x_{\ell+1}\rangle \, d\pi_S(x_1, \dots, x_N) = \int_{\Omega} \sum_{\ell=1}^{N-1} \langle u_\ell(x), x - S^\ell x\rangle \, d\mu \geq 0.$$

By taking $S = I$, we get that the infimum is necessarily zero.

The equivalence of 3) and 4) follows immediately from developing the square.

We now show that 3) implies 1). Take N points $x_1, x_2, \dots, x_N$ in Ω, and let $R > 0$ be such that $B(x_i, R) \subset \Omega$. Consider the transformation

$$S_R(x) = \begin{cases} x - x_1 + x_2 & \text{for } x \in B(x_1, R), \\ x - x_2 + x_3 & \text{for } x \in B(x_2, R), \\ & \vdots \\ x - x_N + x_1 & \text{for } x \in B(x_N, R), \\ x & \text{otherwise.}\end{cases}$$

It is easy to see that S_R is a measure-preserving transformation and that $S_R^N = \mathrm{Id}$.

We then have

$$0 \leq \int_{\Omega} \sum_{\ell=1}^{N-1} \langle u_\ell(x), x - S_R^\ell x \rangle \, d\mu \leq \sum_{i=1}^{N} \int_{B(x_i, R)} \sum_{\ell=1}^{N-1} \langle u_\ell(x), x_i - x_{\ell+i} \rangle \, d\mu.$$

Letting $R \to 0$, we get from Lebesgue's density theorem that

$$\frac{1}{|B(x_i, R)|} \int_{B(x_i, R)} \langle u_\ell(x), x_i - x_{\ell+i} \rangle d\mu \to \langle u_\ell(x_i), x_i - x_{\ell+i} \rangle,$$

from which follows that $(u_1, \dots, u_{N-1})$ are jointly N-monotone a.e. on Ω. The fact that 1) is equivalent to 5) follows immediately from Theorem 2.

To prove that 5) implies 6), note that for all $p_i \in \mathbb{R}^d$, $x \in \Omega$, $y_i \in \Omega$, $i = 1, \dots, N-1$,

$$L_H(x, p_1, \dots, p_{N-1}) + H(x, y_1, \dots, y_{N-1}) \geq \sum_{i=1}^{N-1} \langle p_i, y_i \rangle,$$

which yields that for any $S \in \mathcal{S}_N(\Omega, \mu)$,

$$\int_{\Omega} [L_H(x, u_1(x), \dots, u_{N-1}(x)) \, d\mu + H(x, Sx, \dots, S^{N-1}x)] \, d\mu$$
$$\geq \int_{\Omega} \sum_{\ell=1}^{N-1} \langle u_\ell(x), S^\ell x \rangle \, d\mu.$$

If $H \in \mathcal{H}_N(\Omega)$ and $S \in \mathcal{S}_N(\Omega, \mu)$, we then have $\int_{\Omega} H(x, Sx, \dots, S^{N-1}x) \, d\mu = 0$, and therefore

$$\int_{\Omega} L_H(x, u_1(x), \dots, u_{N-1}(x)) \, d\mu \geq \int_{\Omega} \sum_{\ell=1}^{N-1} \langle u_\ell(x), S^\ell x \rangle \, d\mu.$$

If now H is the N-sub-antisymmetric Hamiltonian obtained by 5), which is concave in the first variable and convex in the last $N-1$ variables, then

$$L_H(x, u_1(x), \dots, u_{N-1}(x)) + H(x, x, \dots, x) = \sum_{\ell=1}^{N-1} \langle u_\ell(x), x \rangle \quad \text{for all } x \in \Omega \setminus \Omega_0,$$

and therefore $\int_{\Omega} L_H(x, u_1(x), \dots, u_{N-1}(x)) \, d\mu = \sum_{\ell=1}^{N-1} \int_{\Omega} \langle u_\ell(x), x \rangle \, d\mu$.

Now consider

$$\overline{H}(\boldsymbol{x}) = \frac{1}{N} \left((N-1) H(\boldsymbol{x}) - \sum_{i=1}^{N-1} H(\sigma^i(\boldsymbol{x})) \right).$$

As before, we have $\overline{H} \in \mathcal{H}_N(\Omega)$ and $\overline{H} \geq H$. Since $L_{\overline{H}} \leq L_H$, we have

$$\int_\Omega L_{\overline{H}}(x, u_1(x), \ldots, u_{N-1}(x))\, d\mu = \sum_{\ell=1}^{N-1} \int_\Omega \langle u_\ell(x), x\rangle\, d\mu$$

and 6) is proved.

Finally, note that 6) readily implies 3), which means that $(u_1, \ldots, u_{N-1})$ is then jointly N-monotone. □

We now consider again the case of a single N-cyclically monotone vector field.

Corollary 9. *Let $u : \Omega \to \mathbb{R}^d$ be a bounded measurable vector field. The following properties are then equivalent*:

1) *The vector field u is N-cyclically monotone a.e., that is, there exists a measure-zero set Ω_0 such that u is N-cyclically monotone on $\Omega \setminus \Omega_0$.*
2) *The infimum of the Monge–Kantorovich problem*

$$\inf\left\{\int_{\Omega^N} \langle u(x_1), x_1 - x_2\rangle\, d\pi(\boldsymbol{x}) : \pi \in \mathcal{P}^{\mu}_{\text{sym}}(\Omega^N)\right\} \tag{38}$$

is equal to zero, and is therefore attained by the push-forward of μ by the map $x \to (x, x, \ldots, x)$.

3) *The vector field u is in the polar of $\mathcal{S}_N(\Omega, \mu)$, that is,*

$$\inf\left\{\int_\Omega \langle u(x), x - Sx\rangle\, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\} = 0. \tag{39}$$

4) *The projection of u on $\mathcal{S}_N(\Omega, \mu)$ is the identity map, that is,*

$$\inf\left\{\int_\Omega |u(x) - Sx|^2\, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\} = \int_\Omega |u(x) - x|^2\, d\mu. \tag{40}$$

5) *There exists an N-cyclically sub-antisymmetric function H of two variables, which is concave in the first variable, convex in the second variable, vanishing on the diagonal and such that*

$$u(x) = \nabla_2 H(x, x) \quad \textit{for a.e. } x \in \Omega. \tag{41}$$

6) *The following duality holds*:

$$\inf\left\{\int_\Omega L_H(x, u(x), 0, \ldots, 0)\, d\mu : H \in \mathcal{H}_N(\Omega)\right\}$$
$$= \sup\left\{\int_\Omega \langle u(x), Sx\rangle\, d\mu : S \in \mathcal{S}_N(\Omega, \mu)\right\}$$

and the latter is attained at the identity map.

Proof. This is an immediate application of Theorem 7 applied to the $(N-1)$-tuplet vector fields $(u, 0, \dots, 0)$, which is clearly jointly N-monotone on $\Omega\setminus\Omega_0$, whenever u is N-monotone on $\Omega \setminus \Omega_0$. □

Remark 10. The sets of μ-measure-preserving N-involutions $(\mathcal{S}_N(\Omega, \mu))_N$ do not form a nested family, that is, $\mathcal{S}_N(\Omega, \mu)$ is not necessarily included in $\mathcal{S}_M(\Omega, \mu)$, whenever $N \le M$, unless of course M is a multiple of N. On the other hand, the above theorem shows that their polar sets, i.e.,

$$\mathcal{S}_N(\Omega, \mu)^0 = \left\{u \in L^2(\Omega, \mathbb{R}^d) : \int_\Omega \langle u(x), x - Sx\rangle \, d\mu \ge 0 \text{ for all } S \in \mathcal{S}_N(\Omega, \mu)\right\},$$

which coincide with the N-cyclically monotone maps, satisfy

$$\mathcal{S}_{N+1}(\Omega, \mu)^0 \subset \mathcal{S}_N(\Omega, \mu)^0,$$

for every $N \ge 1$. This can also be seen directly. Indeed, it is clear that a 2-involution is a 4-involution but not necessarily a 3-involution. On the other hand, assume that u is a 3-cyclically monotone operator. Then for any transformation $S : \Omega \to \Omega$, we have

$$\int_\Omega \langle u(x), x - Sx\rangle \, d\mu + \int_\Omega \langle u(Sx), Sx - S^2x\rangle \, d\mu + \int_\Omega \langle u(S^2x), S^2x - x\rangle \, d\mu \ge 0.$$

Now if S is measure-preserving, we have

$$\int_\Omega \langle u(x), x - Sx\rangle \, d\mu + \int_\Omega \langle u(x), x - Sx\rangle \, d\mu + \int_\Omega \langle u(S^2x), S^2x - x\rangle \, d\mu \ge 0,$$

and if $S^2 = I$, then $\int_\Omega \langle u(x), x - Sx\rangle \, d\mu \ge 0$, which means that $u \in \mathcal{S}_2(\Omega, \mu)^0$. Similarly, one can show that any $(N+1)$-cyclically monotone operator belongs to $\mathcal{S}_N(\Omega, \mu)^0$. In other words, $\mathcal{S}_{N+1}(\Omega, \mu)^0 \subset \mathcal{S}_N(\Omega, \mu)^0$ for all $N \ge 2$. Note that $\mathcal{S}_1(\Omega, \mu)^0 = \{I\}^0 = L^2(\Omega, \mathbb{R}^d)$, while

$$\begin{aligned}\mathcal{S}(\Omega, \mu)^0 &= \textstyle\bigcap_N \mathcal{S}_N(\Omega, \mu)^0 \\ &= \{u \in L^2(\Omega, \mathbb{R}^d), u = \nabla\phi \text{ for some convex function } \phi \text{ in } W^{1,2}(\mathbb{R}^d)\},\end{aligned}$$

in view of classical results of Rockafellar [1970] and Brenier [1991].

Remark 11. In [Ghoussoub and Moameni 2013b], the preceding result is extended to give a similar decomposition for any family of bounded measurable vector fields $u_1, u_2, \dots, u_{N-1}$ on Ω. It is shown there that there exists a measure-preserving N-involution S on Ω and an N-antisymmetric Hamiltonian H on Ω^N such that for $i = 1, \dots, N-1$, we have

$$u_i(x) = \nabla_{i+1} H(x, Sx, S^2x, \dots, S^{N-1}x) \quad \text{for a.e. } x \in \Omega.$$

Acknowledgement

We are grateful to the anonymous referee for a careful reading of this paper, which led to several improvements.

References

[Brenier 1991] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions", *Comm. Pure Appl. Math.* **44**:4 (1991), 375–417. MR 92d:46088 Zbl 0738.46011

[Ghoussoub 2009] N. Ghoussoub, *Self-dual partial differential systems and their variational principles*, Springer, New York, 2009. MR 2010c:35001 Zbl 05366497

[Ghoussoub and Moameni 2013a] N. Ghoussoub and A. Moameni, "A self-dual polar factorization for vector fields", *Comm. Pure Appl. Math.* **66**:6 (2013), 905–933. MR 3043385 Zbl 1264.49048

[Ghoussoub and Moameni 2013b] N. Ghoussoub and A. Moameni, "Symmetric Monge–Kantorovich problems and polar decompositions of vector fields", preprint, 2013. arXiv 1302.2886

[Krauss 1985] E. Krauss, "A representation of arbitrary maximal monotone operators via subgradients of skew-symmetric saddle functions", *Nonlinear Anal.* **9**:12 (1985), 1381–1399. MR 88a:47046 Zbl 0619.47042

[Millien 2011] P. Millien, "On a polar factorization theorem", Master's thesis, University of British Columbia, Vancouver, 2011, http://tinyurl.com/millienthesis.

[Phelps 1993] R. R. Phelps, *Convex functions, monotone operators and differentiability*, 2nd ed., Lecture Notes in Math. **1364**, Springer, Berlin, 1993. MR 94f:46055 Zbl 0921.46039

[Rockafellar 1970] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series **28**, Princeton University Press, 1970. MR 43 #445 Zbl 0193.18401

Received January 7, 2013. Revised October 2, 2013.

ALFRED GALICHON
ECONOMICS DEPARTMENT
SCIENCES PO PARIS
28 RUE DES SAINTS-PÈRES
75007 PARIS
FRANCE
alfred.galichon@science-po.fr

NASSIF GHOUSSOUB
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER BC V6T 1Z2
CANADA
nassif@math.ubc.ca

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.341

RESTRICTED SUCCESSIVE MINIMA

MARTIN HENK AND CARSTEN THIEL

We give bounds on the successive minima of an o-symmetric convex body under the restriction that the lattice points realizing the successive minima are not contained in a collection of forbidden sublattices. Our investigations extend former results to forbidden full-dimensional lattices, to all successive minima and complement former results in the lower-dimensional case.

1. Introduction

Let $\mathcal{K}_o^n$ be the set of all o-symmetric convex bodies in $\mathbb{R}^n$ with nonempty interior, i.e., $K \in \mathcal{K}_o^n$ is an n-dimensional compact convex set satisfying $K = -K$. The volume, i.e., the n-dimensional Lebesgue measure, of a subset $X \subset \mathbb{R}^n$ is denoted by $\operatorname{vol} X$. By a lattice $\Lambda \subset \mathbb{R}^n$ we understand a free $\mathbb{Z}$-module of rank $\operatorname{rg} \Lambda \leq n$. The set of all lattices is denoted by $\mathcal{L}^n$, and $\det \Lambda$ denotes the determinant of $\Lambda \in \mathcal{L}^n$, that is the $(\operatorname{rg} \Lambda)$-dimensional volume of a fundamental cell of Λ.

For $K \in \mathcal{K}_o^n$ and $\Lambda \in \mathcal{L}^n$, Minkowski introduced the i-th successive minimum $\lambda_i(K, \Lambda)$, $1 \leq i \leq \operatorname{rg} \Lambda$, as the smallest positive number λ such that λK contains at least i linearly independent lattice points of Λ, i.e.,

$$\lambda_i(K, \Lambda) = \min\{\lambda \in \mathbb{R}_{\geq 0} : \dim(\lambda K \cap \Lambda) \geq i\}, \quad 1 \leq i \leq \operatorname{rg} \Lambda.$$

Minkowski's first fundamental theorem (see, e.g., [Gruber 2007, Sections 22–23]) on successive minima establishes an upper bound on the first successive minimum in terms of the volume of a convex body. More precisely, for $K \in \mathcal{K}_o^n$ and $\Lambda \in \mathcal{L}^n$ with $\operatorname{rg} \Lambda = r$, it may be formulated as

$$\lambda_1(K, \Lambda)^r \operatorname{vol}_r(K \cap \operatorname{lin} \Lambda) \leq 2^r \det \Lambda, \tag{1-1}$$

where $\operatorname{vol}_r(\,\cdot\,)$ denotes the r-dimensional volume, here with respect to the subspace $\operatorname{lin} \Lambda$, the linear hull of Λ. In the case $r = n$ we just write $\operatorname{vol}(\,\cdot\,)$. One of the many successful applications of this inequality is related to "Siegel's lemma", a catch-all term for results bounding the norm of a nontrivial lattice point lying in a linear subspace given as $\ker A$ where $A \in \mathbb{Z}^{m \times n}$ is an integral matrix of rank m. For instance, with respect to the maximum norm $|\cdot|_\infty$, it was shown in [Bombieri

MSC2010: primary 11H06; secondary 52C07.
Keywords: successive minima, lattice, convex body, short lattice vectors, forbidden sublattices.

and Vaaler 1983] (see also [Ball and Pajor 1990]) that there exists a $z \in \ker A \backslash \{0\}$ such that

$$|z|_\infty \leq \sqrt{\det(AA^\intercal)}^{\frac{1}{n-m}}.$$

In fact, this follows by (1-1), where $K = [-1, 1]^n$ is the cube of edge length 2, $\Lambda = \ker A \cap \mathbb{Z}^n$ is an $(n-m)$-dimensional lattice of determinant at most $\sqrt{\det(AA^\intercal)}$, and Vaaler's result [1979] on the minimal volume of a slice of a cube, which here gives $\mathrm{vol}_{n-m}([-1, 1]^n \cap \mathrm{lin}\, \Lambda) \geq 2^{n-m}$. For generalizations of Siegel's lemma to number fields we refer to [Bombieri and Vaaler 1983; Fukshansky 2006a; 2006b; Gaudron 2009; Gaudron and Rémond 2012a; Vaaler 2003].

Motivated by questions in Diophantine approximation, Fukshansky [2006a] studied an inverse problem to that addressed in Siegel's lemma, namely to bound the norm of lattice points which are not contained in the union of proper sublattices. To describe his result we need a bit more notation.

For a collection of sublattices $\Lambda_i \subset \Lambda$, $1 \leq i \leq s$, with $\bigcup_{i=1}^{s} \Lambda_i \neq \Lambda$ we call

$$\lambda_i\Big(K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\Big) = \min\Big\{\lambda \in \mathbb{R}_{\geq 0} : \dim\Big(\lambda K \cap \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\Big) \geq i\Big\}, \quad 1 \leq i \leq \mathrm{rg}\, \Lambda,$$

the *i-th restricted successive minimum* of K with respect to $\Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i$. Observe that by the compactness of K and the discreteness of $\Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i$ these minima are well-defined. Furthermore, they behave nicely with respect to dilations, as for $\mu > 0$ we have

$$\lambda_i\Big(\mu K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\Big) = \lambda_i\Big(K, \frac{1}{\mu}\Big(\Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\Big)\Big) = \frac{1}{\mu}\lambda_i\Big(K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\Big). \tag{1-2}$$

Moreover, for a lattice $\Lambda \in \mathscr{L}^n$, $r = \mathrm{rg}\, \Lambda$, and a basis $(b_1, \dots, b_r)$, $b_j \in \mathbb{R}^n$, of Λ, let $v(\Lambda) \in \mathbb{R}^{\binom{n}{r}}$ be the vector with entries $\det B_j$, where B_j is an $r \times r$ submatrix of $(b_1, \dots, b_r)$. Observe that up to the order of the coordinates the vector is independent of the given basis, and on account of the Cauchy–Binet formula the Euclidean norm of $v(\Lambda)$ is the determinant of the lattice. With this notation, Fukshansky [2006a, Theorem 1.1] proved

$$\lambda_1\left([-1, 1]^n, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\right) \leq \left(\tfrac{3}{2}\right)^{r-1} r^r \left(\sum_{i=1}^{s} \frac{1}{|v(\Lambda_i)|_\infty} + \sqrt{s}\right) |v(\Lambda)|_\infty + 1, \tag{1-3}$$

for proper sublattices Λ_i, $1 \leq i \leq s$, where *proper* means $\mathrm{rg}\, \Lambda_i < \mathrm{rg}\, \Lambda = r$. This result was generalized and improved in various ways in [Gaudron 2009] and [Gaudron and Rémond 2012a]. In particular, (1-3) has been extended to all o-symmetric bodies as well as to the adelic setting (see also [Gaudron and Rémond 2012b, Lemma 3.2] for an application). For instance, the following is a simplified version of [Gaudron 2009, Theorem 6.1] when we assume that $\operatorname{rg}\Lambda_i = \operatorname{rg}\Lambda - 1 = r - 1$ (see also [Gaudron and Rémond 2012a, Theorem 2.2, Corollary 3.3]):

$$\lambda_1\Bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\Bigr) \le \nu \max_{1\le i\le s}\left\{1,\ \frac{\nu^{r-1}\operatorname{vol}(K\cap\operatorname{lin}\Lambda_i)}{\omega_r\det\Lambda_i},\ \left(\frac{\nu}{\lambda_1(K,\Lambda\cap\operatorname{lin}\Lambda_i)}\right)^{\frac{r-2}{2}}\right\}, \tag{1-4}$$

where $\nu = 7r(s\omega_r\det\Lambda/\operatorname{vol}K)^{1/r}$ and ω_r is the volume of the r-dimensional unit ball.

In our first theorem we want to complement these results on forbidden lower-dimensional lattices by a bound which takes care of the size or the structure of the individual forbidden sublattices such that the bound becomes essentially (1-1) if $\lambda_1(K,\Lambda_i)\to\infty$ for $1\le i\le s$. In this case the bounds in (1-3) and (1-4) still have a dependency on s of order $\sqrt{s}$ and $s^{1/r}$, respectively. Here we have the following result.

Theorem 1.1. *Let $K\in\mathcal{K}_o^n$, $\Lambda\in\mathcal{L}^n$, $\operatorname{rg}\Lambda = n\ge 2$, and let $\Lambda_i\subset\Lambda$, $1\le i\le s$, $\operatorname{rg}\Lambda_i\le n-1$, be sublattices. Then*

$$\lambda_1\Bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\Bigr) < 6^{n-1}\frac{\det\Lambda}{\lambda_1(K,\Lambda)^{n-2}\operatorname{vol}K}\left(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\right) + \sqrt[n]{2^n\frac{\det\Lambda}{\operatorname{vol}K}}.$$

Note that, if $s=0$ or all the $\lambda_1(K,\Lambda_i)$ are very large, we get essentially (1-1).

Our second main theorem deals with forbidden full-dimensional sublattices — those for which $\operatorname{rg}\Lambda_i = \operatorname{rg}\Lambda$, $1\le i\le s$.

Theorem 1.2. *Let $K\in\mathcal{K}_o^n$, $\Lambda\in\mathcal{L}^n$, $\operatorname{rg}\Lambda = n\ge 2$, and let $\Lambda_i\subset\Lambda$, $1\le i\le s$, $\operatorname{rg}\Lambda_i = n$, be sublattices such that $\bigcup_{i=1}^{s}\Lambda_i\ne\Lambda$. Then*

$$\lambda_1\Bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\Bigr) < \frac{2^n\det\Lambda}{\lambda_1(K,\overline{\Lambda})^{n-1}\operatorname{vol}K}\left(\sum_{i=1}^{s}\frac{\det\overline{\Lambda}}{\det\Lambda_i} - s + 1\right) + \lambda_1(K,\overline{\Lambda}),$$

where $\overline{\Lambda} = \bigcap_{i=1}^{s}\Lambda_i$.

In the special case $s=1$, since we may the assume $\lambda_1(K,\Lambda_1)=\lambda_1(K,\Lambda)$, we get the following immediate consequence:

Corollary 1.3. *Let $K\in\mathcal{K}_o^n$, $\Lambda\in\mathcal{L}^n$, $\operatorname{rg}\Lambda = n\ge 2$, and let $\Lambda_1\subsetneq\Lambda$, $\operatorname{rg}\Lambda_1 = n$, be a sublattice. Then*

$$\lambda_1(K,\Lambda\setminus\Lambda_1) \le \frac{2^n\det\Lambda}{\lambda_1(K,\Lambda_1)^{n-1}\operatorname{vol}K} + \lambda_1(K,\Lambda).$$

The following example shows that the bound in Theorem 1.2 as well as the one of the corollary above cannot be improved in general by a multiplicative factor.

Example 1.4. Let $K \in \mathcal{K}_o^2$ be the rectangle $K = [-1, 1] \times [-\alpha, \alpha]$ of edge-lengths 2 and 2α, $\alpha \leq 1$, and of volume 4α. Let $\Lambda = \mathbb{Z}^2$, and define the sublattices

$$\Lambda_1 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_2 \equiv 0 \bmod 2\}, \quad \Lambda_2 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_1 \equiv 0 \bmod p\},$$

where $p > 2$ is a prime. Then $\det \Lambda = 1$, $\det \Lambda_1 = 2$, $\det \Lambda_2 = p$, and

$$\overline{\Lambda} = \Lambda_1 \cap \Lambda_2 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_2 \equiv 0 \bmod 2,\ z_1 \equiv 0 \bmod p\}$$

with $\det \overline{\Lambda} = 2p$. For $\alpha \leq 2/p$ we therefore have $\lambda_1(K, \overline{\Lambda}) = p$. Regarding the set $\Lambda \backslash (\Lambda_1 \cup \Lambda_2)$, we observe that the lattice points on the axes are forbidden, but not $(1, 1)^\intercal$ and so $\lambda_1(K, \Lambda \backslash (\Lambda_1 \cup \Lambda_2)) = 1/\alpha$. Putting everything together, the bound in Theorem 1.2 evaluates for $\alpha \leq 2/p$ to

$$\frac{1}{\alpha} = \lambda_1(\Lambda \backslash (\Lambda_1 \cup \Lambda_2)) < \frac{4}{p\, 4\alpha}(p+1) + p = \frac{1}{\alpha} + \frac{1}{p\alpha} + p.$$

Hence for $\alpha = 2/p^2$ and $p \to \infty$ the bound cannot be improved by a multiplicative factor.

In the situation of Corollary 1.3, i.e., when we consider only the forbidden lattice Λ_1, the upper bound in the corollary evaluates to $1/\alpha + 1$, whereas, as before, $\lambda_1(K, \Lambda \backslash \Lambda_1) = 1/\alpha$.

Before beginning with the proofs of our results we would like to mention a closely related problem, namely to cover $K \cap \Lambda$, $K \in \mathcal{K}_o^n$, by a minimal number $\gamma(K)$ of lattice hyperplanes. Obviously, having a $\nu > 0$ with $\gamma(\nu K) \geq s + 1$ implies that

$$\lambda_1\left(K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\right) \leq \nu$$

in the case of lower-dimensional sublattices Λ_i. For bounds on $\gamma(K)$ in terms of the successive minima and other functionals from the geometry of numbers we refer to [Bárány et al. 2001; Bezdek and Hausel 1994; Bezdek and Litvak 2009].

Finally, we remark that restricted successive minima have also been investigated from an algorithmic point of view. Blömer and Naewe [2007] studied the complexity of computing $\lambda_1\big(K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\big)$ for $s = 1$ and when K is the unit ball of an l_p-norm. They show, among other things, that some of the well-known lattice problems, like the shortest or closest lattice vector problem, are polynomial reducible to computing/approximating $\lambda_1(K, \Lambda \backslash \Lambda_1)$. Moreover, as in the case of these lattice problems an LLL-reduced basis (see [Grötschel et al. 1993, Chapter 5]) can be used to find in polynomial time a lattice vector b which approximates $\lambda_1(B_n, \Lambda \backslash \Lambda_1)$ up to a factor of 2^{n-1} [Blömer and Naewe 2007, Theorem 3.6]. Here B_n is the unit ball of the Euclidean norm. Hence, Theorem 1.1 implies (see [Grötschel et al. 1993, Theorem 5.3.13a] for a similar result in the standard setting $s = 0$):

Corollary 1.5. *Let* $\Lambda \in \mathscr{L}^n$, $\operatorname{rg}\Lambda = n \geq 2$, *and let* $\Lambda_1 \subset \Lambda$, $\operatorname{rg}\Lambda_1 \leq n-1$, *be a sublattice. There exists a polynomial time algorithm for computing a vector* $b \in \Lambda \backslash \Lambda_1$ *of Euclidean length*

$$\|b\| < 2^{n-1}\left(6^{n-1}\frac{\det\Lambda}{\lambda_1(K,\Lambda)^{n-2}\operatorname{vol}K}\frac{1}{\lambda_1(K,\Lambda_1)} + \sqrt[n]{2^n\frac{\det\Lambda}{\operatorname{vol}K}}\right).$$

It seems to be a challenging problem to extend this result to more than one forbidden sublattice as well as to full-dimensional forbidden lattices.

The paper is organized as follows. The proof of Theorem 1.1 will be given in the next section and full-dimensional forbidden sublattices, i.e., Theorem 1.2, will be treated in Section 3. In each of the sections we also present some extensions of the results above to higher successive minima, i.e., to $\lambda_i\big(K, \Lambda\backslash\bigcup_{i=1}^s \Lambda_i\big)$, $i > 1$.

2. Avoiding lower-dimensional sublattices

In the course of the proof we have to estimate the number of lattice points in a centrally symmetric convex body, i.e., to bound $|K \cap \Lambda|$ from below and above. Assuming $K \in \mathscr{K}_o^n$ and $\operatorname{rg}\Lambda = n$, we will use as a lower bound a classical result of van der Corput (see, e.g., [Gruber and Lekkerkerker 1987, p. 51]):

$$|K \cap \Lambda| \geq 2\left\lfloor \frac{\operatorname{vol}K}{2^n\det\Lambda}\right\rfloor + 1 > \frac{\operatorname{vol}K}{2^{n-1}\det\Lambda} - 1. \tag{2-1}$$

As upper bound we will use a bound in terms of the first successive minima by Betke, Henk and Wills [Betke et al. 1993]:

$$|K \cap \Lambda| \leq \left(\frac{2}{\lambda_1(K,\Lambda)} + 1\right)^n. \tag{2-2}$$

Proof of Theorem 1.1. By scaling K with $\lambda_1(K,\Lambda)$ we may assume without loss of generality that $\lambda_1(K,\Lambda) = 1$, i.e., K contains no nontrivial lattice point in its interior (cf. (1-2)). Let $n_i = \operatorname{rg}\Lambda_i < n$. For $\gamma \geq 1$, since $\lambda_1(K,\Lambda_i) \geq \lambda_1(K,\Lambda) = 1$ we get, from (2-2),

$$|\gamma K\backslash\{0\} \cap \Lambda_i| \leq \left(\gamma\frac{2}{\lambda_1(K,\Lambda_i)} + 1\right)^{n_i} - 1 < \gamma^{n-1}3^{n-1}\frac{1}{\lambda_1(K,\Lambda_i)}. \tag{2-3}$$

Hence, for $\gamma \geq 1$, we have

$$\left|\gamma K\backslash\{0\} \cap \left(\bigcup_{i=1}^{s}\Lambda_i\right)\right| < \gamma^{n-1}3^{n-1}\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}. \tag{2-4}$$

Combining this bound with the upper bound (2-1) leads, for $\gamma \geq 1$, to the estimate

$$\begin{aligned}\Bigl|\gamma K\backslash\{0\}\cap\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\Bigr| &> \gamma^n\frac{\operatorname{vol}K}{2^{n-1}\det\Lambda}-2-\Bigl|\gamma K\backslash\{0\}\cap\Bigl(\bigcup_{i=1}^{s}\Lambda_i\Bigr)\Bigr|\\ &> \gamma^n\frac{\operatorname{vol}K}{2^{n-1}\det\Lambda}-\gamma^{n-1}3^{n-1}\Bigl(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\Bigr)-2\\ &= \frac{\operatorname{vol}K}{2^{n-1}\det\Lambda}(\gamma^n-\gamma^{n-1}\beta-\rho),\end{aligned}$$

where

$$\beta=6^{n-1}\frac{\det\Lambda}{\operatorname{vol}K}\Bigl(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\Bigr),\qquad \rho=2^n\frac{\det\Lambda}{\operatorname{vol}K}.$$

Hence, given β and ρ, we have to determine a $\gamma\geq 1$ such that $\gamma^n-\gamma^{n-1}\beta-\rho>0$. To this end let $\overline{\gamma}=\beta+\rho^{1/n}$. Then

$$\begin{aligned}\overline{\gamma}^n-\overline{\gamma}^{n-1}\beta &= (\beta+\rho^{1/n})^n-(\beta+\rho^{1/n})^{n-1}\beta\\ &= \rho^{1/n}(\beta+\rho^{1/n})^{n-1}>\rho^{1/n}\rho^{(n-1)/n}=\rho.\end{aligned}\tag{2-5}$$

Finally, we observe that

$$\overline{\gamma}>\rho^{1/n}=\Bigl(2^n\frac{\det\Lambda}{\operatorname{vol}K}\Bigr)^{1/n}\geq\lambda_1(K,\Lambda)=1,$$

by (1-1) and our assumption. Hence, $\overline{\gamma}>1$ and in view of (2-5) we have $\lambda_1\bigl(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\bigr)<\overline{\gamma}$, which by the definition of $\overline{\gamma}$ yields the desired bound of the theorem with respect to our normalization $\lambda_1(K,\Lambda)=1$. □

Compared to the bounds in (1-3) and (1-4), our formula uses only the successive minima and not the determinants of the forbidden sublattices which reflect more the structure of a lattice. However, instead of (2-2) one can use a Blichfeldt-type bound, proved in [Henze 2013], for o-symmetric convex bodies K with $\dim(K\cap\Lambda)=n$; namely, if $L_n(x)$ is the n-th Laguerre polynomial,

$$|K\cap\Lambda|\leq\frac{n!}{2^n}\frac{\operatorname{vol}K}{\det\Lambda}L_n(-2),$$

This leads to a bound on $\lambda_1\bigl(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\bigr)$ where the sum over $1/\lambda_1(K,\Lambda_i)$ is replaced by a sum over ratios of the type $\operatorname{vol}_{\dim H}(K\cap H)/\det(\Lambda_i\cap H)$ for certain lower-dimensional planes $H\subseteq\operatorname{lin}\Lambda_i$. In general, however, we have no control over the dimension of these hyperplanes H nor on the volume of the sections.

Theorem 1.1 can easily be extended inductively to higher restricted successive minima $\lambda_{j+1}\bigl(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\bigr)$, $1\leq j\leq n-1$, by avoiding, in addition, a j-dimensional lattice containing j linearly independent lattice points corresponding to the successive minima $\lambda_i\bigl(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\bigr)$, $1\leq i\leq j$.

Corollary 2.1. *Under the assumptions of Theorem 1.1 we have, for* $j = 1, \dots, n-1$,

$$\lambda_{j+1}\left(K, \Lambda\backslash \bigcup_{i=1}^{s} \Lambda_i\right) < 6^{n-1}\frac{\det\Lambda}{\lambda_1(K,\Lambda)^{n-2}\operatorname{vol}K}\left(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\right) + \left(\frac{3^j}{\lambda_1(K,\Lambda)^j}2^{n-1}\frac{\det\Lambda}{\operatorname{vol}K} + \left(2^n\frac{\det\Lambda}{\operatorname{vol}K}\right)^{\frac{n-j}{n}}\right)^{\frac{1}{n-j}}.$$

Proof. Let $z_i \in \lambda_i\big(K, \Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\big)\, K\cap\Lambda$, $1 \le i \le j$, be linearly independent, and let $\overline{\Lambda} = \Lambda \cap \operatorname{lin}\{z_1, \dots, z_j\}$. Then

$$\lambda_{j+1}\left(K, \Lambda\backslash \bigcup_{i=1}^{s}\Lambda_i\right) = \lambda_1\left(K, \Lambda\backslash\left(\bigcup_{i=1}^{s}\Lambda_i \cup \overline{\Lambda}\right)\right), \tag{2-6}$$

and we now follow the proof of Theorem 1.1. In particular, we assume $\lambda_1(K,\Lambda) = 1$. In addition to the upper bounds on $|\gamma K\backslash\{0\}\cap\Lambda_i|$, $1 \le i \le s$, in (2-3), we also use for $\gamma \ge \lambda_1(K,\overline{\Lambda}) \ge \lambda_1(K,\Lambda) = 1$ the bound

$$|\gamma K\backslash\{0\}\cap\overline{\Lambda}| < \left(\frac{2\gamma}{\lambda_1(K,\overline{\Lambda})}+1\right)^j \le 3^j\left(\frac{\gamma}{\lambda_1(K,\overline{\Lambda})}\right)^j. \tag{2-7}$$

Combining this bound with (2-1) leads for $\gamma \ge \lambda_1(K,\overline{\Lambda})$ to

$$\begin{aligned}
&\left|\gamma K\backslash\{0\} \cap \Lambda\backslash\left(\bigcup_{i=1}^{s}\Lambda_i\cup\overline{\Lambda}\right)\right| \\
&\qquad > \gamma^n\frac{\operatorname{vol}K}{2^{n-1}\det\Lambda} - 2 - \left|\gamma K\backslash\{0\} \cap \left(\bigcup_{i=1}^{s}\Lambda_i\right)\right| - |\gamma K\backslash\{0\}\cap\overline{\Lambda}| \\
&\qquad > \gamma^n\frac{\operatorname{vol}K}{2^{n-1}\det\Lambda} - 2 - \gamma^{n-1}3^{n-1}\left(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\right) - 3^j\left(\frac{\gamma}{\lambda_1(K,\overline{\Lambda})}\right)^j \\
&\qquad = \frac{\operatorname{vol}K}{2^{n-1}\det\Lambda}(\gamma^n - \gamma^{n-1}\beta - \gamma^j\alpha - \rho),
\end{aligned}$$

with

$$\beta = 6^{n-1}\frac{\det\Lambda}{\operatorname{vol}K}\left(\sum_{i=1}^{s}\frac{1}{\lambda_1(K,\Lambda_i)}\right), \quad \alpha = \frac{3^j}{\lambda_1(K,\overline{\Lambda})^j}2^{n-1}\frac{\det\Lambda}{\operatorname{vol}K}, \quad \rho = 2^n\frac{\det\Lambda}{\operatorname{vol}K}.$$

Now setting $\overline{\gamma} = \beta + \big(\alpha + \rho^{\frac{n-j}{n}}\big)^{\frac{1}{n-j}}$ we see as in the proof of Theorem 1.1 that

$$\begin{aligned}
\overline{\gamma}^n - \overline{\gamma}^{n-1}\beta - \overline{\gamma}^j\alpha - \rho &= \overline{\gamma}^j\big(\overline{\gamma}^{n-j} - \beta\overline{\gamma}^{n-j-1} - \alpha\big) - \rho \\
&> \overline{\gamma}^j\rho^{(n-j)/n} - \rho > 0.
\end{aligned} \tag{2-8}$$

Since $\overline{\gamma} > \beta + \rho^{1/n}$, which is, by the proof of Theorem 1.1, an upper bound on $\lambda_1(K,\overline{\Lambda})$, we also have $\overline{\gamma} > \lambda_1(K,\overline{\Lambda})$ and so we know $\lambda_{j+1}\big(K, \Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\big) < \overline{\gamma}$, by (2-8). By the definition of $\overline{\gamma}$ we get the required upper bound with respect to the normalization $\lambda_1(K, \Lambda) = 1$. □

An upper bound on $\lambda_j\bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\bigr)$ of a different kind involves the so-called covering radius $\mu(K, \Lambda)$ of a convex body $K \in \mathcal{K}_o^n$ and a lattice $\Lambda \in \mathcal{L}^n$, $\operatorname{rg}\Lambda = n$. This is the smallest positive number μ such that any translate of μK contains a lattice point:

$$\mu(K, \Lambda) = \min\bigl\{\mu > 0 : (t + \mu K) \cap \Lambda \neq \varnothing \text{ for all } t \in \mathbb{R}^n\bigr\}.$$

(see [Gruber and Lekkerkerker 1987, Chapter 2, Section 13]).

Proposition 2.2. *Under the assumptions of Theorem 1.1 we have*

$$\lambda_1\left(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\right) \leq (s+1)\mu(K, \Lambda),$$

and hence $\lambda_j\left(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\right) \leq (s+2)\mu(K, \Lambda)$ *for* $2 \leq j \leq n$.

Proof. Observe that on account of (2-6) the bound for $j \geq 2$ follows from the one for $\lambda_1\bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\bigr)$. For the proof in the case $j = 1$ let $H_i = \operatorname{lin}\Lambda_i$, $1 \leq i \leq s$, and for brevity we write $\bar{\mu}$ instead of $\mu(K, \Lambda)$. By Ball's solution [1991] of the affine plank problem for o-symmetric convex bodies, applied to $\bar{\mu}K$, we know that there exists a $t \in \mathbb{R}^n$ such that

$$\left(t + \frac{1}{s+1}\bar{\mu}K\right) \subset \bar{\mu}K \quad\text{and}\quad \operatorname{int}\left(t + \frac{1}{s+1}\bar{\mu}K\right) \cap H_i = \varnothing,\ 1 \leq i \leq s,$$

where $\operatorname{int}(\,\cdot\,)$ denotes the interior. Thus, for any $\epsilon > 0$ the body $(s+1+\epsilon)\bar{\mu}K$ contains a translate $t_\epsilon + \bar{\mu}K$ having no points in common with H_i, $1 \leq i \leq s$. Hence, together with the definition of the covering radius, we have $(t_\epsilon + \bar{\mu}K) \cap \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i \neq \varnothing$ and so $\lambda_1\bigl(K, \Lambda\setminus\bigcup_{i=1}^{s}\Lambda_i\bigr) \leq (s+1+\epsilon)\bar{\mu}$. By the arbitrariness of ϵ and the compactness of K the assertion follows. □

For a comparable uniform bound in the much more general adelic setting and, of course, with a completely different method see [Gaudron and Rémond 2012a, Proposition 3.2].

3. Avoiding full-dimensional sublattices

If the forbidden sublattices are full-dimensional we cannot argue as in the lower-dimensional case, since now the number of forbidden lattice points in $\lambda K \cap \bigcup_{i=1}^{s}\Lambda_i$ grows with the same order of magnitude with respect to λ as the number of points in $\lambda K \cap \Lambda$.

The tool we use in this full-dimensional case is the torus group $\mathbb{R}^n/\overline{\Lambda}$ for a certain lattice $\overline{\Lambda}$. For a more detailed discussion we refer to [Gruber 2007, Section 26].

We recall that this quotient of abelian groups is a compact topological group and we may identify $\mathbb{R}^n/\overline{\Lambda}$ with a fundamental parallelepiped P of $\overline{\Lambda}$:

$$\mathbb{R}^n/\overline{\Lambda} \sim P = \{\rho_1 b_1 + \cdots + \rho_n b_n : 0 \leq \rho_i < 1\},$$

where $b_1, \ldots, b_n$ form a basis of $\overline{\Lambda}$. Hence for $X \subset \mathbb{R}^n$, the set X modulo $\overline{\Lambda}$, $X/\overline{\Lambda}$, can be described as

$$X/\overline{\Lambda} = \{y \in P : \exists b \in \overline{\Lambda} \text{ such that } y + b \in X\} = (\overline{\Lambda} + X) \cap P,$$

and we can think of $\bar{X} \subseteq \mathbb{R}^n/\overline{\Lambda}$ as its image under inclusion into $\mathbb{R}^n$. In the same spirit we may identify addition $\oplus$ in $\mathbb{R}^n/\overline{\Lambda}$ with the corresponding operation in $\mathbb{R}^n$, i.e., for $\bar{X}_1, \bar{X}_2 \subset \mathbb{R}^n/\overline{\Lambda}$ we have

$$\bar{X}_1 \oplus \bar{X}_2 = ((\bar{X}_1 + \bar{X}_2) + \overline{\Lambda}) \cap P.$$

As $\mathbb{R}^n/\overline{\Lambda}$ is a compact abelian group, there is a unique Haar measure $\mathrm{vol}_T(\,\cdot\,)$ on it normalized to $\mathrm{vol}_T(\mathbb{R}^n/\overline{\Lambda}) = \det \overline{\Lambda}$, and for a "nice" measurable set $X \subset \mathbb{R}^n$ or $\bar{X} \subset \mathbb{R}^n/\overline{\Lambda}$ we have

$$\mathrm{vol}_T(X/\overline{\Lambda}) = \mathrm{vol}((\overline{\Lambda} + X) \cap P) \quad \text{and} \quad \mathrm{vol}_T(\bar{X}) = \mathrm{vol}((\overline{\Lambda} + \bar{X}) \cap P).$$

Regarding the volume of the sum of two sets $\bar{X}_1, \bar{X}_2 \subset \mathbb{R}^n/\overline{\Lambda}$ we have the following classical sum theorem of Kneser and Macbeath [Gruber 2007, Theorem 26.1]:

$$\mathrm{vol}_T(\bar{X}_1 \oplus \bar{X}_2) \geq \min\bigl\{\mathrm{vol}_T(\bar{X}_1) + \mathrm{vol}_T(\bar{X}_2),\ \det \overline{\Lambda}\bigr\}. \tag{3-1}$$

We also note that for an o-symmetric convex body $K \in \mathcal{K}_o^n$ and $\lambda \geq 0$ the set $\overline{\Lambda} + \lambda K$ forms a lattice packing, i.e., for any two different lattice points $\bar{a}, \bar{b} \in \overline{\Lambda}$ the translates $\bar{a} + \lambda K$ and $\bar{b} + \lambda K$ do not overlap if and only if $\lambda \leq \lambda_1(K, \overline{\Lambda})/2$. Hence we know that, for $0 \leq \lambda \leq \lambda_1(K, \overline{\Lambda})/2$,

$$\mathrm{vol}_T(\lambda K/\overline{\Lambda}) = \mathrm{vol}\bigl((\lambda K + \overline{\Lambda}) \cap P\bigr) = \lambda^n \mathrm{vol}\, K. \tag{3-2}$$

Furthermore, we also need a "torus version" of van der Corput's result (2-1):

Lemma 3.1. *Let $K \in \mathcal{K}_o^n$, $\Lambda \in \mathcal{L}^n$, $\mathrm{rg}\,\Lambda = n$ and let $\overline{\Lambda} \subsetneq \Lambda$ be a sublattice with $\mathrm{rg}\,\overline{\Lambda} = n$, and let $m \in \mathbb{N}$ with $m \det \Lambda < \det \overline{\Lambda}$. If $\mathrm{vol}_T\bigl(\tfrac{1}{2}K/\overline{\Lambda}\bigr) \geq m \det \Lambda$ then*

$$\#(K/\overline{\Lambda} \cap \Lambda) \geq m + 1;$$

i.e., K contains at least $m+1$ lattice points of Λ belonging to different cosets modulo $\overline{\Lambda}$.

Proof. By the compactness of K and the discreteness of lattices we may assume $\mathrm{vol}_T\bigl(\tfrac{1}{2}K/\overline{\Lambda}\bigr) > m \det \Lambda$. Let P be a fundamental parallelepiped of the lattice $\overline{\Lambda}$. Then by assumption we have for the measurable set $X = \bigl(\tfrac{1}{2}K + \overline{\Lambda}\bigr) \cap P$ that $\mathrm{vol}\, X > m \det \Lambda$. According to a result of van der Corput [Gruber and Lekkerkerker

1987, Section 6.1, Theorem 1] we know that there exists pairwise different $x_i \in X$, $1 \le i \le m+1$, such that $x_i - x_j \in \Lambda$. By the o-symmetry and convexity of K we have $(X - X) = (K + \overline{\Lambda}) \cap (P - P)$, and since $(P - P) \cap \overline{\Lambda} = \{0\}$ we conclude that

$$x_i - x_j \in (K + \overline{\Lambda}) \cap \Lambda \backslash \overline{\Lambda}, \quad i \neq j.$$

Hence the m points $x_i - x_1 \in K + \overline{\Lambda}$, $i = 2, \dots, m+1$, belong to different nontrivial cosets of Λ modulo $\overline{\Lambda}$ and thus $\#(K/\overline{\Lambda} \cap \Lambda) \ge m+1$, where the additional 1 counts the origin. □

We now state some simple facts on the intersection of full-dimensional sublattices.

Lemma 3.2. *Let* $\Lambda \in \mathscr{L}^n$, $\Lambda_i \subseteq \Lambda$, $1 \le i \le s$, $\operatorname{rg} \Lambda_i = \operatorname{rg} \Lambda = n$, *and let* $\overline{\Lambda} = \bigcap_{i=1}^{s} \Lambda_i$. *Then* $\overline{\Lambda} \in \mathscr{L}^n$ *with* $\operatorname{rg} \overline{\Lambda} = n$, *and*

$$\max_{1 \le i \le s} \det \Lambda_i \le \det \overline{\Lambda} \le (\det \Lambda)^{1-s} (\det \Lambda_1) \cdots (\det \Lambda_s).$$

Moreover, with $m = \sum_{i=1}^{s} \det \overline{\Lambda} / \det \Lambda_i - s + 1$ *we have*:

(i) *The union* $\bigcup_{i=1}^{s} \Lambda_i$ *is covered by at most m cosets of* Λ *modulo* $\overline{\Lambda}$.

(ii) *If* $\det \overline{\Lambda} / \det \Lambda \ge m+1$ *then* $\Lambda \neq \bigcup_{i=1}^{s} \Lambda_i$.

Proof. In order to show that $\overline{\Lambda}$ is a full-dimensional lattice it suffices to consider $s = 2$. Obviously, $\Lambda_1 \cap \Lambda_2$ is a discrete subgroup of Λ and it also contains n linearly independent points, e.g., $(\det \Lambda_2) a_1, \dots, (\det \Lambda_2) a_n$, where $a_1, \dots, a_n$ is a basis of Λ_1. Hence $\overline{\Lambda}$ is a full-dimensional lattice; see [Gruber and Lekkerkerker 1987, Section 3.2, Theorem 2]. The lower bound on $\det \overline{\Lambda}$ is clear by the inclusion $\overline{\Lambda} \subseteq \Lambda_i$, $1 \le i \le s$. For the upper bound we observe that two points $g, h \in \Lambda$ belong to different cosets modulo $\overline{\Lambda}$ if and only if g and h belong to different cosets of Λ modulo at least one Λ_i. There are $\det \Lambda_i / \det \Lambda$ many cosets for each i and so we get the upper bound.

For (i) we note that since Λ_i is the union of $\det \overline{\Lambda} / \det \Lambda_i$ many cosets modulo $\overline{\Lambda}$, the union is certainly covered by $\sum_{i=1}^{s} \det \overline{\Lambda} / \det \Lambda_i = m + s - 1$ many cosets of Λ modulo $\overline{\Lambda}$. But here we have counted the trivial coset at least s times. Part (ii) is a direct consequence of part (i). □

Lemma 3.2(ii) implies, in particular, that the union of two strict sublattices can never be the whole lattice. This is no longer true for three sublattices, as we see in the next example, which also shows that Lemma 3.2(ii) is not an equivalence.

Example 3.3. Let $\Lambda = \mathbb{Z}^2$, and let $\Lambda_1, \dots, \Lambda_4 \subset \mathbb{Z}^2$ be the sublattices

$$\Lambda_1 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_2 \equiv 0 \bmod 2\}, \quad \Lambda_2 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_1 \equiv 0 \bmod 2\},$$
$$\Lambda_3 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_2 \equiv 0 \bmod 3\}, \quad \Lambda_4 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_1 \equiv z_2 \bmod 2\}.$$

Then $\Lambda_1 \cup \Lambda_2 \cup \Lambda_4 = \Lambda$ but $\Lambda_1 \cup \Lambda_2 \cup \Lambda_3 \neq \Lambda$. Furthermore $\det \Lambda = 1$, $\det \Lambda_1 = \det \Lambda_2 = \det \Lambda_4 = 2$, $\det \Lambda_3 = 3$ and

$$\overline{\Lambda} = \Lambda_1 \cap \Lambda_2 \cap \Lambda_3 = \{(z_1, z_2)^\intercal \in \mathbb{Z}^2 : z_1 \equiv 0 \bmod 2,\ z_2 \equiv 0 \bmod 6\}$$

with $\det \overline{\Lambda} = 12$, while $\sum_{i=1}^{3} \frac{\det \overline{\Lambda}}{\det \Lambda_i} - 1 = 15$.

We now come to the proof of the full-dimensional case.

Proof of Theorem 1.2. Let $\Lambda_1, \dots, \Lambda_s$ be the full-dimensional forbidden sublattices of the given lattice Λ and let $\overline{\Lambda} = \bigcap_{i=1}^{s} \Lambda_i$. Let

$$m = \min\left\{ \sum_{i=1}^{s} \frac{\det \overline{\Lambda}}{\det \Lambda_i} - s + 1,\ \frac{\det \overline{\Lambda}}{\det \Lambda} \right\}.$$

Claim 1. *Let* $\lambda > 0$ *with* $\mathrm{vol}_T\big((\lambda \tfrac{1}{2} K)/\overline{\Lambda}\big) \geq m \det \Lambda$. *Then*

$$\lambda_1\left(K, \Lambda \setminus \bigcup_{i=1}^{s} \Lambda_i\right) \leq \lambda.$$

To verify the claim, we first assume

$$m = \sum_{i=1}^{s} \frac{\det \overline{\Lambda}}{\det \Lambda_i} - s + 1 < \frac{\det \overline{\Lambda}}{\det \Lambda}.$$

By Lemma 3.1, λK contains $m+1$ lattice points of Λ belonging to different cosets with respect to $\overline{\Lambda}$. By Lemma 3.2 (i), $\bigcup_{i=1}^{s} \Lambda_i$ is covered by at most m cosets of Λ modulo $\overline{\Lambda}$, and thus λK contains a lattice point of $\Lambda \setminus \bigcup_{i=1}^{s} \Lambda_i$.

Next suppose that $m = \det \overline{\Lambda} / \det \Lambda$. Then

$$\mathrm{vol}_T\big((\lambda \tfrac{1}{2} K)/\overline{\Lambda}\big) = \det \overline{\Lambda} = \mathrm{vol}_T(\mathbb{R}^n/\overline{\Lambda})$$

and, in particular, λK contains a representative of each coset of Λ modulo $\overline{\Lambda}$. By assumption there exists a coset containing a point $a \in \Lambda \setminus \bigcup_{i=1}^{s} \Lambda_i$, and hence all points of this coset, that is $a + \overline{\Lambda}$, lie in $\Lambda \setminus \bigcup_{i=1}^{s} \Lambda_i$.

This verifies the claim and it remains to compute a λ with

$$\mathrm{vol}_T\big((\lambda \tfrac{1}{2} K)/\overline{\Lambda}\big) \geq m \det \Lambda. \tag{3-3}$$

To this end we set $\lambda_1 = \lambda_1(K, \overline{\Lambda})$ and we write an arbitrary $\lambda > 0$ modulo λ_1 in the form $\lambda = \lfloor \lambda/\lambda_1 \rfloor \lambda_1 + \rho\lambda_1$, with $0 \leq \rho < 1$. Hence, in view of the sum theorem of Kneser and Macbeath (3-1) and the packing property (3-2) of λ_1 with respect to $\frac{1}{2}K$, we may write

$$
\begin{aligned}
\mathrm{vol}_T\big((\lambda\tfrac{1}{2}K)/\overline{\Lambda}\big) &= \mathrm{vol}_T\Big(\Big(\Big(\Big\lfloor\frac{\lambda}{\lambda_1}\Big\rfloor\frac{\lambda_1}{2}+\rho\frac{\lambda_1}{2}\Big)K\Big)/\overline{\Lambda}\Big)\\
&= \mathrm{vol}_T\Big(\underbrace{\Big(\frac{\lambda_1}{2}K\Big)/\overline{\Lambda}\oplus\cdots\oplus\Big(\frac{\lambda_1}{2}K\Big)/\overline{\Lambda}}_{\lfloor\lambda/\lambda_1\rfloor}\oplus\Big(\frac{\rho\lambda_1}{2}K\Big)/\overline{\Lambda}\Big)\\
&\geq \min\Big\{\Big(\Big\lfloor\frac{\lambda}{\lambda_1}\Big\rfloor+\rho^n\Big)\Big(\frac{\lambda_1}{2}\Big)^n\mathrm{vol}\,K,\det\overline{\Lambda}\Big\}.
\end{aligned}
$$

Thus, (3-3) is certainly satisfied for a $\bar{\lambda}$ with

$$
\Big(\Big\lfloor\frac{\bar{\lambda}}{\lambda_1}\Big\rfloor+\rho^n\Big)\Big(\frac{\lambda_1}{2}\Big)^n\mathrm{vol}\,K=\Big(\sum_{i=1}^{s}\frac{\det\overline{\Lambda}}{\det\Lambda_i}-s+1\Big)\det\Lambda. \tag{3-4}
$$

Using that

$$
\Big\lfloor\frac{\lambda}{\lambda_1}\Big\rfloor+\rho^n>\frac{\lambda-\lambda_1}{\lambda_1}, \tag{3-5}
$$

we find

$$
\lambda_1\Big(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\Big)\leq\bar{\lambda}<\frac{2^n\det\Lambda}{\lambda_1^{n-1}\mathrm{vol}\,K}\Big(\sum_{i=1}^{s}\frac{\det\overline{\Lambda}}{\det\Lambda_i}-s+1\Big)+\lambda_1. \qquad\square
$$

Remark 3.4. The bound in Theorem 1.2 can be slightly improved in lower dimensions by noticing that in (3-5) we may replace $(\lambda-\lambda_1)/\lambda_1$ by $\lambda/\lambda_1-\rho+\rho^n$. Since $\rho-\rho^n$ takes its maximum at $\rho=(1/n)^{1/(n-1)}$ we get in this way

$$
\begin{aligned}
&\lambda_1\Big(K,\Lambda\backslash\bigcup_{i=1}^{s}\Lambda_i\Big)\\
&\qquad\leq\frac{2^n\det\Lambda}{\lambda_1(K,\overline{\Lambda})^{n-1}\mathrm{vol}\,K}\Big(\sum_{i=1}^{s}\frac{\det\overline{\Lambda}}{\det\Lambda_i}-s+1\Big)+n^{-1/(n-1)}\,\frac{n-1}{n}\,\lambda_1(K,\overline{\Lambda}).
\end{aligned}
$$

There is a straightforward way to extend Theorem 1.2 to higher successive minima which we will first present in the special case $s=1$.

Corollary 3.5. *Under the assumptions of Corollary 1.3 we have, for* $1\leq i\leq n$,

$$
\lambda_i(K,\Lambda\backslash\Lambda_1)\leq\frac{2^n\det\Lambda}{\lambda_1(K,\Lambda_1)^{n-1}\mathrm{vol}\,K}+\lambda_1(K,\Lambda)+\lambda_i(K,\Lambda).
$$

Proof. By Corollary 1.3 it suffices to show $\lambda_i(K,\Lambda\backslash\Lambda_1)\leq\lambda_1(K,\Lambda\backslash\Lambda_1)+\lambda_i(K,\Lambda)$ for $i=2,\dots,n$. To this end let $a\in\lambda_1(K,\Lambda\backslash\Lambda_1)K\cap\Lambda\backslash\Lambda_1$ and let $b_1,\dots,b_n$ be linearly independent with $b_j\in\lambda_j(K,\Lambda)K\cap\Lambda$, $j=1,\dots,n$. Since not both b_j and $a+b_j$ can belong to the forbidden sublattice Λ_1 we can select from each pair b_j, $a+b_j$ one contained in $\Lambda\backslash\Lambda_1$, $1\leq j\leq n$. Let these points be denoted by $\bar{b}_j$, $j=1,\dots,n$. Then $a,\bar{b}_j\in\big(\lambda_1(K,\Lambda\backslash\Lambda_1)+\lambda_j(K,\Lambda)\big)K$, $1\leq j\leq n$.

Now choose k such that $a\notin\mathrm{lin}(\{b_1,\dots,b_n\}\backslash\{b_k\})$. Then the lattice points $a,\bar{b}_1,\dots\bar{b}_{k-1},\bar{b}_{k+1},\dots,\bar{b}_n$ are linearly independent and we are done. $\square$

For $s > 1$ the excluded substructure $\bigcup_{i=1}^{s} \Lambda_i$ is, in general, not a lattice anymore and so we cannot argue as above. Therefore, in this case, we choose the vectors b_j, $1 \le j \le n$, from the lattice $\overline{\Lambda} = \bigcap_{i=1}^{s} \Lambda_i$. Then for $a \in \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i$ we have

$$a, a + b_1, a + b_2, \dots, a + b_n \in \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i,$$

and analogously to the proof of Corollary 3.5 we get:

Corollary 3.6. *Under the assumptions of Theorem 1.2 we have, for* $1 \le i \le n$,

$$\lambda_i\left(K, \Lambda \backslash \bigcup_{i=1}^{s} \Lambda_i\right) \\ \le \frac{2^n \det \Lambda}{\lambda_1(K, \overline{\Lambda})^{n-1} \operatorname{vol} K} \left(\sum_{i=1}^{s} \frac{\det \overline{\Lambda}}{\det \Lambda_i} - s + 1\right) + \lambda_1(K, \overline{\Lambda}) + \lambda_i(K, \overline{\Lambda}).$$

Remark 3.7. It is also possible to extend lower-dimensional lattices to lattices of full rank by adjoining "sufficiently large" vectors, i.e., for each Λ_i of rank n_i choose linearly independent $z_{i,n_i+1}, \dots, z_{i,n} \in \Lambda \backslash \Lambda_i$ and consider the lattice $\overline{\Lambda}_i$ spanned by Λ_i and $z_{i,n_i+1}, \dots, z_{i,n}$. If $z_{i,j}$ are such that $\lambda_j(K, \overline{\Lambda}_i)$ is very large for $j > n_i$, one can apply the results from Section 3 to the collection $\overline{\Lambda}_i$, $1 \le i \le s$. However, the bounds obtained in this way are in general weaker, with one exception in the case $s = 1$ for the bound on $\lambda_1(K, \Lambda \backslash \Lambda_1)$. Here we get

$$\lambda_1(K, \Lambda \backslash \Lambda_1) \le \frac{2^n \det \Lambda}{\lambda_1(K, \Lambda_1)^{n-1} \operatorname{vol} K} + \lambda_1(K, \Lambda)$$

for $\Lambda_1 \subsetneq \Lambda$ with $\operatorname{rg} \Lambda_1 < n$, which improves on Theorem 1.1.

Acknowledgement

We thank the anonymous referee for very helpful and valuable comments, corrections and suggestions.

References

[Ball 1991] K. Ball, "The plank problem for symmetric bodies", *Invent. Math.* **104**:3 (1991), 535–543. MR 92c:52003 Zbl 0702.52003

[Ball and Pajor 1990] K. Ball and A. Pajor, "Convex bodies with few faces", *Proc. Amer. Math. Soc.* **110**:1 (1990), 225–231. MR 90m:52011 Zbl 0704.52003

[Bárány et al. 2001] I. Bárány, G. Harcos, J. Pach, and G. Tardos, "Covering lattice points by subspaces", *Period. Math. Hungar.* **43**:1-2 (2001), 93–103. MR 2002d:11076 Zbl 1062.11043

[Betke et al. 1993] U. Betke, M. Henk, and J. M. Wills, "Successive-minima-type inequalities", *Discrete Comput. Geom.* **9**:2 (1993), 165–175. MR 93j:52026 Zbl 0771.52007

[Bezdek and Hausel 1994] K. Bezdek and T. Hausel, "On the number of lattice hyperplanes which are needed to cover the lattice points of a convex body", pp. 27–31 in *Intuitive geometry* (Szeged, 1991), edited by K. Böröczky and G. Fejes Tóth, Colloq. Math. Soc. János Bolyai **63**, North-Holland, Amsterdam, 1994. MR 97a:52026 Zbl 0823.52006

[Bezdek and Litvak 2009] K. Bezdek and A. E. Litvak, "Covering convex bodies by cylinders and lattice points by flats", *J. Geom. Anal.* **19**:2 (2009), 233–243. MR 2011c:52016 Zbl 1175.52008

[Blömer and Naewe 2007] J. Blömer and S. Naewe, "Sampling methods for shortest vectors, closest vectors and successive minima", pp. 65–77 in *Automata, languages and programming*, edited by L. Arge et al., Lecture Notes in Comput. Sci. **4596**, Springer, Berlin, 2007. MR 2009k:68087 Zbl 1171.11328

[Bombieri and Vaaler 1983] E. Bombieri and J. Vaaler, "On Siegel's lemma", *Invent. Math.* **73**:1 (1983), 11–32. MR 85g:11049a Zbl 0533.10030

[Fukshansky 2006a] L. Fukshansky, "Integral points of small height outside of a hypersurface", *Monatsh. Math.* **147**:1 (2006), 25–41. MR 2006i:11074 Zbl 1091.11024

[Fukshansky 2006b] L. Fukshansky, "Siegel's lemma with additional conditions", *J. Number Theory* **120**:1 (2006), 13–25. MR 2007f:11026 Zbl 1192.11018

[Gaudron 2009] É. Gaudron, "Géométrie des nombres adélique et lemmes de Siegel généralisés", *Manuscripta Math.* **130**:2 (2009), 159–182. MR 2010i:11102 Zbl 1231.11076

[Gaudron and Rémond 2012a] É. Gaudron and G. Rémond, "Lemmes de Siegel d'évitement", *Acta Arith.* **154**:2 (2012), 125–136. MR 2945657 Zbl 1266.11080

[Gaudron and Rémond 2012b] É. Gaudron and G. Rémond, "Polarisations et isogénies", preprint, 2012, http://math.univ-bpclermont.fr/~gaudron/art13.pdf.

[Grötschel et al. 1993] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization*, 2nd ed., Algorithms and Combinatorics **2**, Springer, Berlin, 1993. MR 95e:90001 Zbl 0837.05001

[Gruber 2007] P. M. Gruber, *Convex and discrete geometry*, Grundlehren der Mathematischen Wissenschaften **336**, Springer, Berlin, 2007. MR 2008f:52001 Zbl 1139.52001

[Gruber and Lekkerkerker 1987] P. M. Gruber and C. G. Lekkerkerker, *Geometry of numbers*, 2nd ed., North-Holland Mathematical Library **37**, North-Holland, Amsterdam, 1987. MR 88j:11034 Zbl 0611.10017

[Henze 2013] M. Henze, "A Blichfeldt-type inequality for centrally symmetric convex bodies", *Monatsh. Math.* **170**:3-4 (2013), 371–379. MR 3055793 Zbl 06176422

[Vaaler 1979] J. D. Vaaler, "A geometric inequality with applications to linear forms", *Pacific J. Math.* **83**:2 (1979), 543–553. MR 81d:52007 Zbl 0465.52011

[Vaaler 2003] J. D. Vaaler, "The best constant in Siegel's lemma", *Monatsh. Math.* **140**:1 (2003), 71–89. MR 2004j:11072 Zbl 1034.11038

Received February 19, 2013. Revised May 17, 2013.

MARTIN HENK
martin.henk@ovgu.de

CARSTEN THIEL
carsten.thiel@gmail.com

FAKULTÄT FÜR MATHEMATIK
OTTO-VON-GUERICKE UNIVERSITÄT MAGDEBURG
UNIVERSITÄTSPLATZ 2
D-39106 MAGDEBURG
GERMANY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.355

RADIAL SOLUTIONS OF NON-ARCHIMEDEAN PSEUDODIFFERENTIAL EQUATIONS

ANATOLY N. KOCHUBEI

We consider a class of equations with the fractional differentiation operator D^α, $\alpha > 0$, for complex-valued functions $x \mapsto f(|x|_K)$ on a non-Archimedean local field K depending only on the absolute value $|\cdot|_K$. We introduce a right inverse I^α to D^α, such that the change of an unknown function $u = I^\alpha v$ reduces the Cauchy problem for an equation with D^α (for radial functions) to an integral equation whose properties resemble those of classical Volterra equations. This contrasts much more complicated behavior of D^α on other classes of functions.

1. Introduction

Pseudodifferential equations for complex-valued functions defined on a non-Archimedean local field are among the central objects of contemporary harmonic analysis and mathematical physics; see the monographs [Vladimirov et al. 1994; Kochubei 2001; Albeverio et al. 2010], and the survey [Zúñiga-Galindo 2011].

The simplest example is the fractional differentiation operator D^α, $\alpha > 0$, on the field $\mathbb{Q}_p$ of p-adic numbers (here p is a prime number). It can be defined as a pseudodifferential operator with the symbol $|\xi|_p^\alpha$ where $|\cdot|_p$ is the p-adic absolute value or, equivalently, as an appropriate convolution operator.

Already in this case, as it was first shown by Vladimirov (see [Vladimirov et al. 1994]), properties of the p-adic pseudodifferential operator are much more complicated than those of its classical counterpart. It suffices to say that, as an operator on $L_2(\mathbb{Q}_p)$, it has a point spectrum of infinite multiplicity. Considering a simple "formal" evolution equation with the operator D^α in the p-adic time variable t, Vladimirov [2003] noticed that such an equation does not possess a fundamental solution.

At the same time, it was found in [Kochubei 2008] that some of the evolution equations of the above kind behave reasonably, if one considers only solutions

This work was supported in part by Grant No. 01-01-12 of the National Academy of Sciences of Ukraine under the program of joint Ukrainian-Russian projects.
MSC2010: 11S80, 35S10.
Keywords: fractional differentiation operator, non-Archimedean local field, radial functions, Cauchy problem.

depending on $|t|_p$. This observation has led to the concept of a non-Archimedean wave equation possessing various properties resembling those of classical hyperbolic equations, up to the Huygens principle.

In this paper we consider the Cauchy problem for a class of equations like

$$D^\alpha u + a(|x|_p)u = f(|x|_p), \quad x \in \mathbb{Q}_p, \tag{1-1}$$

assuming that a solution is looked for in the class of radial functions, $u = u(|x|_p)$; the precise definition of D^α and assumptions on a, f are given below. This Cauchy problem is reduced to an integral equation resembling classical Volterra equations. It turns out that (1-1) and its generalizations considered on radial functions constitute p-adic counterparts of ordinary differential equations.

2. Preliminaries

2.1. ***Local fields.*** Let K be a non-Archimedean local field, that is, a nondiscrete totally disconnected locally compact topological field. It is well known that K is isomorphic either to a finite extension of the field $\mathbb{Q}_p$ of p-adic numbers (if K has characteristic 0), or to the field of formal Laurent series with coefficients from a finite field, if K has a positive characteristic. For a summary of main notions and results regarding local fields see, for example, [Kochubei 2001].

Any local field K is endowed with an absolute value $|\cdot|_K$, such that $|x|_K = 0$ if and only if $x = 0$, $|xy|_K = |x|_K \cdot |y|_K$, $|x + y|_K \le \max(|x|_K, |y|_K)$. Set

$$O = \{x \in K : |x|_K \le 1\}, \quad P = \{x \in K : |x|_K < 1\}.$$

Then O is a subring of K, and P is an ideal in O containing such an element β that $P = \beta O$. The quotient ring O/P is actually a finite field; denote by q its cardinality. We will always assume that the absolute value is normalized, that is $|\beta|_K = q^{-1}$. The normalized absolute value takes the values q^N, $N \in \mathbb{Z}$. Note that for $K = \mathbb{Q}_p$ we have $\beta = p$ and $q = p$; the p-adic absolute value is normalized.

Denote by $S \subset O$ a complete system of representatives of the residue classes from O/P. Any nonzero element $x \in K$ admits the canonical representation in the form of the convergent series

$$x = \beta^{-n}(x_0 + x_1\beta + x_2\beta^2 + \cdots) \tag{2-1}$$

where $n \in \mathbb{Z}$, $|x|_K = q^n$, $x_j \in S$, $x_0 \notin P$. For $K = \mathbb{Q}_p$, one may take $S = \{0, 1, \ldots, p-1\}$.

The additive group of any local field is self-dual; that is, if χ is a fixed nonconstant complex-valued additive character of K, then any other additive character can be written as $\chi_a(x) = \chi(ax)$, $x \in K$, for some $a \in K$. Below we assume that χ is a rank zero character, that is $\chi(x) \equiv 1$ for $x \in O$, while there exists such an element $x_0 \in K$ that $|x_0|_K = q$ and $\chi(x_0) \neq 1$.

This duality is used in the definition of the Fourier transform over K. Denoting by dx the Haar measure on the additive group of K (normalized in such a way that the measure of O equals 1) we write

$$\tilde{f}(\xi) = \int_K \chi(x\xi) f(x)\, dx, \quad \xi \in K,$$

where f is a complex-valued function from $L_1(K)$. As usual, the Fourier transform $\mathcal{F}$ can be extended from $L_1(K) \cap L_2(K)$ to a unitary operator on $L_2(K)$. If $\mathcal{F} f = \tilde{f} \in L_1(K)$, we have the inversion formula

$$f(x) = \int_K \chi(-x\xi) \tilde{f}(\xi)\, d\xi.$$

2.2. ***Integration formulas.*** As in real analysis, there are many well known formulas for integrals of complex-valued functions defined on subsets of a local field. There exist even tables of such integrals [Vladimirov 2003]. Note that formulas for integrals on $\mathbb{Q}_p$ and its subsets, as a rule, carry over to the general case, if one substitutes the normalized absolute value for $|\cdot|_p$ and q for p.

Here we collect some formulas used in this work. Let $n \in \mathbb{Z}$, $\alpha > 0$.

$$\int\limits_{|x|_K \le q^n} |x|_K^{\alpha-1}\, dx = \frac{1-q^{-1}}{1-q^{-\alpha}} q^{\alpha n}, \tag{2-2}$$

$$\int\limits_{|x|_K = q^n} |x-a|_K^{\alpha-1}\, dx = \frac{q-2+q^{-\alpha}}{q(1-q^{-\alpha})} |a|_K^{\alpha}, \quad |a|_K = q^n. \tag{2-3}$$

$$\int\limits_{|x|_K \le q^n} \log|x|_K\, dx = \Big(n - \frac{1}{q-1}\Big) q^n \log q. \tag{2-4}$$

$$\int\limits_{|x|_K = q^n} \log|x-a|_K\, dx = \Big[\Big(1-\frac{1}{q}\Big)\log|a|_K - \frac{\log q}{q-1}\Big]|a|_K, \quad |a|_K = q^n. \tag{2-5}$$

$$\int\limits_{|x|_K \le q^n} dx = q^n; \quad \int\limits_{|x|_K = q^n} dx = \Big(1-\frac{1}{q}\Big) q^n. \tag{2-6}$$

$$\int\limits_{\substack{|x|_K = q^n \\ x_0 = k_0}} dx = q^{n-1}, \quad 0 \neq k_0 \in S \tag{2-7}$$

(the restriction $x_0 = k_0$ is in the sense of the canonical representation (2-1)).

$$\int_{\substack{|x|_K=q^n\\ x_0\neq k_0}} dx=\Big(1-\frac{2}{q}\Big)q^n. \tag{2-8}$$

2.3. ***Test functions and distributions.*** A function $f: K\to\mathbb{C}$ is said to be locally constant, if there exists such an integer l that for any $x\in K$

$$f(x+x')=f(x),\quad \text{whenever } |x'|\le q^{-l}.$$

The smallest number l with this property is called the exponent of local constancy of the function f.

Let $\mathcal{D}(K)$ be the set of all locally constant functions with compact supports; it is a vector space over $\mathbb{C}$ with the topology of double inductive limit

$$\mathcal{D}(K)=\varinjlim_{N\to\infty}\ \varinjlim_{l\to\infty}\mathcal{D}_N^l$$

where $\mathcal{D}_N^l$ is the finite dimensional space of functions supported in the ball $B_N=\{x\in K:|x|\le q^N\}$ and having the exponents of local constancy $\le l$. The strong conjugate space $\mathcal{D}'(K)$ is called the space of Bruhat–Schwartz distributions.

The Fourier transform preserves the space $\mathcal{D}(K)$. Therefore the Fourier transform of a distribution defined by duality acts continuously on $\mathcal{D}'(K)$. As in the case of $\mathbb{R}^n$, there exists a well developed theory of distributions over local fields; it includes such topics as convolution, direct product, homogeneous distributions etc (see [Vladimirov et al. 1994; Kochubei 2001; Albeverio et al. 2010]). In connection with homogeneous distributions, it is useful to introduce the subspaces of $\mathcal{D}(K)$:

$$\Psi(K)=\{\psi\in\mathcal{D}(K):\psi(0)=0\},$$
$$\Phi(K)=\Big\{\varphi\in\mathcal{D}(K):\int_K\varphi(x)\,dx=0\Big\}.$$

The Fourier transform $\mathcal{F}$ is a linear isomorphism from $\Psi(K)$ onto $\Phi(K)$, thus also from $\Phi'(K)$ onto $\Psi'(K)$. The spaces $\Phi(K)$ and $\Phi'(K)$ are called the Lizorkin spaces (of the second kind) of test functions and distributions respectively; see [Albeverio et al. 2010]. Note that two distributions differing by a constant summand coincide as elements of $\Phi'(K)$.

3. Fractional differentiation and integration operators

3.1. ***Riesz kernels and operators generated by them.*** On a test function $\varphi\in\mathcal{D}(K)$, the fractional differentiation operator D^α, $\alpha>0$, is defined as

$$(D^\alpha\varphi)(x)=\mathcal{F}^{-1}\big[|\xi|_K^\alpha\big(\mathcal{F}(\varphi)\big)(\xi)\big](x). \tag{3-1}$$

However D^α does not act on the space $\mathfrak{D}(K)$, since the function $\xi \mapsto |\xi|_K^\alpha$ is not locally constant. On the other hand, $D^\alpha : \Phi(K) \to \Phi(K)$ and $D^\alpha : \Phi'(K) \to \Phi'(K)$; see [Albeverio et al. 2010], and that was a motivation to introduce these spaces.

The operator D^α can also be represented as a hypersingular integral operator:

$$(D^\alpha \varphi)(x) = \frac{1-q^\alpha}{1-q^{-\alpha-1}} \int_K |y|_K^{-\alpha-1}[\varphi(x-y)-\varphi(x)]\,dy \tag{3-2}$$

[Vladimirov et al. 1994; Kochubei 2001]. In contrast to (3-1), the expression in the right of (3-2) makes sense for wider classes of functions. Below we study this in detail for the case of radial functions.

The expression in (3-2) is in fact the convolution $f_{-\alpha} * \varphi$, where the Riesz kernel f_s, for complex $s \notin 1 + \frac{2\pi i}{\log q}\mathbb{Z}$, is defined first for $\operatorname{Re} s > 0$ as

$$f_s(x) = \frac{|x|_K^{s-1}}{\Gamma_K(s)}, \quad \Gamma_K(s) = \frac{1-q^{s-1}}{1-q^{-s}},$$

and then extended meromorphically to the remaining nonzero values of s as a distribution from $\mathfrak{D}'(K)$:

$$\langle f_s, \varphi\rangle = \frac{1-q^{-1}}{1-q^{s-1}}\varphi(0) + \frac{1-q^{-s}}{1-q^{s-1}}\Bigg[\int\limits_{|x|_K>1} \varphi(x)\frac{dx}{|x|_K^{1-s}} + \int\limits_{|x|_K\le 1} (\varphi(x)-\varphi(0))\frac{dx}{|x|_K^{1-s}}\Bigg],$$

For $s = 0$, we set $f_0(x) = \delta(x)$. For $s \in 1 + \frac{2\pi i}{\log q}\mathbb{Z}$, we define

$$f_s(x) = \frac{1-q}{\log q} \log |x|_K.$$

It is well known that $f_s * f_t = f_{s+t}$ in the sense of distributions from $\mathfrak{D}'(K)$, so long as $s, t, s+t \notin 1 + \frac{2\pi i}{\log q}\mathbb{Z}$. If these kernels are considered as distributions from $\Phi'(K)$, then $f_s * f_t = f_{s+t}$ for all $s, t \in \mathbb{C}$ [Albeverio et al. 2010]. In view of this identity, it is natural to define the operator $D^{-\alpha}$, $\alpha > 0$, setting

$$(D^{-\alpha}\varphi)(x) = (f_\alpha * \varphi)(x) = \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} \int_K |x-y|_K^{\alpha-1}\varphi(y)\,dy, \qquad \varphi \in \mathfrak{D}(K),\ \alpha \ne 1, \tag{3-3}$$

and

$$(D^{-1}\varphi)(x) = \frac{1-q}{q\log q} \int_K \log|x-y|_K \varphi(y)\,dy. \tag{3-4}$$

Then $D^\alpha D^{-\alpha} = I$ on $\mathfrak{D}(K)$, if $\alpha \ne 1$. This remains valid on $\Phi(K)$ also for $\alpha = 1$.

The notions and results above are well known; see [Vladimirov et al. 1994; Albeverio et al. 2010]. We now come to new phenomena, considering the case of radial functions.

3.2. ***Operators on radial functions.*** Let u be a radial function, that is $u = u(|x|_K)$, $x \in K$. (In order to make the notation concise, we identify the function $x \mapsto u(|x|_K)$ on K with the function $|x|_K \mapsto u(|x|_K)$ on $q^{\mathbb{Z}}$. This abuse of notation will not lead to confusion.)

Let us find an explicit expression for $D^\alpha u$, $\alpha > 0$. Below we write $d_\alpha = (1-q^\alpha)/(1-q^{-\alpha-1})$. For $x \in K$, we denote by x_0 the element from $S \subset O$ appearing in the representation (2-1).

Lemma 1. *If a function $u = u(|x|_K)$ is such that*

$$\sum_{k=-\infty}^{m} q^k \left|u(q^k)\right| < \infty, \quad \sum_{l=m}^{\infty} q^{-\alpha l}\left|u(q^l)\right| < \infty, \tag{3-5}$$

for some $m \in \mathbb{Z}$, then for each $n \in \mathbb{Z}$ the expression in the right-hand side of (3-2) *with $\varphi(x) = u(|x|_K)$ exists for $|x|_K = q^n$, depends only on $|x|_K$, and*

$$\begin{aligned}(D^\alpha u)(q^n) = d_\alpha\Big(1-\frac{1}{q}\Big)q^{-(\alpha+1)n}\sum_{k=-\infty}^{n-1} q^k u(q^k) + q^{-\alpha n-1}\frac{q^\alpha+q-2}{1-q^{-\alpha-1}}u(q^n)\\ + d_\alpha\Big(1-\frac{1}{q}\Big)\sum_{l=n+1}^{\infty} q^{-\alpha l}u(q^l).\end{aligned} \tag{3-6}$$

Proof. We find, using the ultrametric properties of the absolute value, that

$$(D^\alpha u)(x) = d_\alpha \int\limits_{|y|_K \ge |x|_K} |y|_K^{-\alpha-1}[u(|x-y|_K) - u(|x|_K)]\,dy.$$

If $|y|_K = |x|_K$ and $y_0 \ne x_0$, the integrand vanishes. Therefore, by (2-6),

$$\begin{aligned}(D^\alpha u)(x) &= d_\alpha \sum_{k=-\infty}^{n-1} \int_{|y-x|_K = q^k} |x|_K^{-\alpha-1}[u(q^k) - u(q^n)]\,dy\\ &\quad + d_\alpha \sum_{l=n+1}^{\infty} \int_{|y|_K = q^l} q^{-l(\alpha+1)}[u(q^k) - u(q^n)]\,dy\\ &= d_\alpha\Big(1-\frac{1}{q}\Big)q^{-(\alpha+1)n}\sum_{k=-\infty}^{n-1} q^k[u(q^k) - u(q^n)]\\ &\quad + d_\alpha\Big(1-\frac{1}{q}\Big)\sum_{l=n+1}^{\infty} q^{-\alpha l}[u(q^l) - u(q^n)].\end{aligned}$$

It is clear from this expression that $(D^\alpha u)(x)$, $|x|_K = q^n$, depends only on $|x|_K$. After elementary transformations we get (3-6). □

Definition. We say that the action $D^\alpha u$, $\alpha > 0$, on a radial function u is *defined in the strong sense* if the function u satisfies (3-5), so that $D^\alpha u(|x|_K)$, $|x|_K \neq 0$, is given by (3-6), and there exists the limit

$$D^\alpha u(0) \stackrel{\text{def}}{=} \lim_{x\to 0} D^\alpha u(|x|_K).$$

It is evident from (3-2) that D^α annihilates constant functions (recall that in $\Phi'(K)$ they are equivalent to zero). Therefore $D^{-\alpha}$ is not the only possible choice of the right inverse to D^α. In particular, we will use

$$(I^\alpha\varphi)(x) = (D^{-\alpha}\varphi)(x) - (D^{-\alpha}\varphi)(0). \tag{3-7}$$

This is defined initially for $\varphi \in \mathfrak{D}(K)$. It is seen from (3-3), (3-4), and the ultrametric property of the absolute value that

$$(I^\alpha\varphi)(x) = \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} \int_{|y|_K \le |x|_K} \bigl(|x-y|_K^{\alpha-1} - |y|_K^{\alpha-1}\bigr)\varphi(y)\,dy, \quad \alpha \neq 1, \tag{3-8}$$

and

$$(I^1\varphi)(x) = \frac{1-q}{q\log q} \int_{|y|_K \le |x|_K} \bigl(\log|x-y|_K - \log|y|_K\bigr)\varphi(y)\,dy. \tag{3-9}$$

In contrast to (3-3) and (3-4), in (3-8) and (3-9) the integrals are taken, for each fixed $x \in K$, over bounded sets.

Let us calculate $I^\alpha u$ for a radial function $u = u(|x|_K)$. Obviously, $(I^\alpha u)(0) = 0$ whenever I^α is defined.

Lemma 2. *Suppose that, for some $m \in \mathbb{Z}$,*

$$\sum_{k=-\infty}^{m} \max\bigl(q^k, q^{\alpha k}\bigr)\bigl|u(q^k)\bigr| < \infty \quad \text{if } \alpha \neq 1,$$

and

$$\sum_{k=-\infty}^{m} |k| q^k \bigl|u(q^k)\bigr| < \infty \qquad \text{if } \alpha = 1.$$

Then $I^\alpha u$ exists, it is a radial function, and for any $x \neq 0$, we have

$$(I^\alpha u)(|x|_K) = q^{-\alpha}|x|_K^\alpha u(|x|_K) + \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} \int_{|y|_K < |x|_K} \bigl(|x|_K^{\alpha-1} - |y|_K^{\alpha-1}\bigr)u(|y|_K)\,dy \tag{3-10}$$

if $\alpha \neq 1$, and

(3-11) $$(I^1u)(|x|_K) = \\ q^{-1}|x|_K u(|x|_K) + \frac{1-q}{q\log q}\int_{|y|_K<|x|_K}\big(\log|x|_K - \log|y|_K\big)u(|y|_K)\,dy.$$

Proof. It is sufficient to compute the integrals over the set $\{y \in K : |y|_K = |x|_K\}$, and that is done using the integration formulas (2-3) and (2-5). □

It follows from Lemma 2 that the function $I^\alpha u$ is continuous if, for example, u is bounded near the origin (see an estimate of the integral $I_{\alpha,0}$ in the proof of Theorem 1 below). If $|u(|x|_K)| \le C|x|_K^{-\varepsilon}$, as $|x|_K \ge 1$, then $|(I^\alpha u)(|x|_K)| \le C|x|_K^{\alpha-\varepsilon}$, as $|x|_K \ge 1$. Here and below we denote by C various (possibly different) positive constants.

It is easy to transform (3-10) and (3-11) further obtaining series involving $u(q^n)$.

Obviously, $D^\alpha I^\alpha = I$ on $\mathfrak{D}(K)$, if $\alpha \ne 1$, and on $\Phi(K)$, if $\alpha = 1$. Since by Lemma 1 and Lemma 2, the operators are defined in a straightforward sense for wider classes of functions, it is natural to look for conditions sufficient for this identity.

Lemma 3. *Suppose that for some $m \in \mathbb{Z}$,*

$$\sum_{k=-\infty}^{m} \max(q^k, q^{\alpha k})\big|v(q^k)\big| < \infty \quad \textit{and} \quad \sum_{l=m}^{\infty}\big|v(q^l)\big| < \infty \qquad \textit{if } \alpha \ne 1,$$
$$\sum_{k=-\infty}^{m} |k|q^k\,\big|v(q^k)\big| < \infty \qquad\qquad \textit{and} \quad \sum_{l=m}^{\infty} l\big|v(q^l)\big| < \infty \quad \textit{if } \alpha = 1.$$

Then there exists $(D^\alpha I^\alpha v)(|x|_K) = v(|x|_K)$ for any $x \ne 0$.

The proof consists of tedious but quite elementary calculations based on the integration formulas (2-2)–(2-8). A relatively nontrivial tool is the sum formula for the arithmetic-geometric progression (from [Gradshteyn and Ryzhik 1996, Formula 0.113]).

Using Lemma 3, we can consider the simplest Cauchy problem

$$D^\alpha u(|x|_K) = f(|x|_K), \quad u(0) = 0,$$

where f is a continuous function, such that

$$\sum_{l=m}^{\infty}\big|f(q^l)\big| < \infty, \text{ if } \alpha \ne 1, \quad \text{or} \quad \sum_{l=m}^{\infty} l\big|f(q^l)\big| < \infty, \text{ if } \alpha = 1.$$

The unique strong solution is $u = I^\alpha f$; the uniqueness follows from the fact that the equality $D^\alpha u = 0$ (in the sense of $\mathfrak{D}'(K)$) implies the equality $u = \text{const}$; see [Vladimirov et al. 1994] or [Kochubei 2001]. Therefore on radial functions, the operators D^α and I^α behave like the Caputo–Dzhrbashyan fractional derivative and the Riemann–Liouville fractional integral of real analysis (see, for example, [Kilbas et al. 2006]). However the next example illustrates a different behavior of the "fractional integral" in the non-Archimedean case.

Example. Let $f(|x|_K) \equiv 1$, $x \in K$. Then $(I^\alpha f)(|x|_K) \equiv 0$.

Proof. Let $|x|_K = q^n$. If $\alpha \neq 1$, then by (3-10), (2-2), and (2-6),

$$\begin{aligned}(I^\alpha f)(|x|_K) &= q^{-\alpha}|x|_K^\alpha + \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} \int_{|y|_K \le q^{n-1}} \bigl(|x|_K^{\alpha-1} - |y|_K^{\alpha-1}\bigr)\, dy \\ &= q^{-\alpha}|x|_K^\alpha + \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} \left[q^{n-1}|x|_K^{\alpha-1} - \frac{1-q^{-1}}{1-q^{-\alpha}} q^{\alpha(n-1)} \right] \\ &= q^{-\alpha}|x|_K^\alpha + \frac{1-q^{-\alpha}}{1-q^{\alpha-1}} |x|_K^\alpha \frac{q^{-1}-q^{-\alpha}}{1-q^{-\alpha}} = 0.\end{aligned}$$

If $\alpha = 1$, then by (3-11), (2-4), and (2-6),

$$\begin{aligned}(I^1 f)(|x|_K) &= q^{-1}|x|_K + \frac{1-q}{q\log q} \int_{|y|_K \le q^{n-1}} \bigl(\log|x|_K - \log|y|_K\bigr)\, dy \\ &= q^{-1}|x|_K + \frac{1-q}{q\log q} \left[q^{n-1}\log|x|_K - \Bigl(n-1-\frac{1}{q-1}\Bigr) q^{n-1}\log q \right] \\ &= |x|_K \left(q^{-1} + \frac{1-q}{q\log q}\Bigl(1+\frac{1}{q-1}\Bigr) q^{-1}\log q \right) = 0. \qquad \square\end{aligned}$$

Of course, these identities in the weaker sense of distributions from $\Phi'(K)$ are trivial, since the constant functions are identified with zero, I^α with $D^{-\alpha}$, and $D^\alpha D^{-\alpha} = I$.

On the other hand, the example shows that the condition of decay at infinity in Lemma 3 cannot be dropped.

4. Fractional differential equations

4.1. ***The Cauchy problem and an integral equation.*** In the class of radial functions $u = u(|x|_K)$, we consider the Cauchy problem

$$D^\alpha u + a(|x|_K)u = f(|x|_K), \quad x \in K, \tag{4-1}$$

$$u(0) = 0, \tag{4-2}$$

where a and f are continuous functions, that is, they have finite limits $a(0)$ and $f(0)$, as $x \to 0$.

Looking for a solution of the form $u = I^\alpha v$, where v is a radial function, we obtain formally an integral equation

$$v(|x|_K) + a(|x|_K)(I^\alpha v)(|x|_K) = f(|x|_K), \quad x \in K. \tag{4-3}$$

Let us study its solvability. Later we investigate, in what sense a solution of (4-3) corresponds to a solution of the Cauchy problem (4-1)–(4-2).

It follows from (4-3) that $v(0) = f(0)$. Suppose first that $\alpha \neq 1$. By Lemma 2, (4-3) can be written in the form

$$\begin{aligned}(4\text{-}3') \quad & \bigl[1+q^{-\alpha}a(|x|_K)|x|_K^{\alpha}\bigr]v(|x|_K) \\ & + c_\alpha a(|x|_K)\int_{|y|_K<|x|_K}\bigl(|x|_K^{\alpha-1}-|y|_K^{\alpha-1}\bigr)v(|y|_K)\,dy = f(|x|_K), \quad x \neq 0,\end{aligned}$$

where $c_\alpha = (1-q^{-\alpha})/(1-q^{\alpha-1})$.

Since a is continuous, there exists such $N \in \mathbb{Z}$ that

$$q^{-\alpha}a(|x|_K)|x|_K^{\alpha} < 1 \quad \text{for } |x|_K \leq q^N.$$

On the ball $B_N = \bigl\{x \in K : |x|_K \leq q^N\bigr\}$, the equation takes the form

$$v(|x|_K) + \int_{|y|_K<|x|_K} k_\alpha(x,y)v(|y|_K)\,dy = F(|x|_K) \tag{4-4}$$

where

$$k_\alpha(x,y) = \begin{cases}\bigl[1+q^{-\alpha}a(|x|_K)|x|_K^{\alpha}\bigr]^{-1}c_\alpha a(|x|_K)\bigl(|x|_K^{\alpha-1}-|y|_K^{\alpha-1}\bigr) & \text{if } x \neq 0,\\ 0 & \text{if } x = 0,\end{cases}$$

and

$$F(|x|_K) = \bigl[1+q^{-\alpha}a(|x|_K)|x|_K^{\alpha}\bigr]^{-1}f(|x|_K).$$

If we construct a solution of (4-4) on B_N, and if

$$a(|x|_K) \neq -q^{\alpha m} \quad \text{for any } x \in K,\ m \in \mathbb{Z}, \tag{4-5}$$

we will be able to construct a solution of (4-4), thus a solution of (4-3), successively for all $x \in K$.

If $\alpha = 1$, we use (3-11) and obtain in a similar way the equation (4-4) with

$$k_1(x,y) = \begin{cases}\dfrac{1-q}{q\log q}\bigl[1+q^{-1}a(|x|_K)|x|_K\bigr]^{-1}a(|x|_K)\bigl(\log|x|_K-\log|y|_K\bigr) & \text{if } x \neq 0,\\ 0 & \text{if } x = 0,\end{cases}$$

and

$$F(|x|_K) = \bigl[1+q^{-1}a(|x|_K)|x|_K\bigr]^{-1}f(|x|_K).$$

It is convenient to extend k_α (including the case $\alpha = 1$) by zero onto $B_N \times B_N$.

Theorem 1. *For each $\alpha > 0$, the integral equation* (4-4) *has a unique continuous solution on B_N.*

Proof. Let us consider the integral operator $\mathcal{K}$ appearing in (4-4) as an operator on the Banach space $C(B_N)$ of complex-valued continuous functions on B_N. By the theory of integral operators developed in sufficient generality in [Edwards 1965] (see Proposition 9.5.17), to prove that $\mathcal{K}$ is a compact operator, it suffices to check that, for any $x_0 \in B_N$,

$$\lim_{x \to x_0} \int_{B_N} |k_\alpha(x, y) - k_\alpha(x_0, y)|\, dy = 0. \tag{4-6}$$

The relation (4-6) is obvious for $x_0 \neq 0$, and also for $\alpha > 1$. For $x_0 = 0$, we have $k_\alpha(0, y) = 0$, and for $0 < \alpha < 1$, $|x|_K = q^n$, $n \geq N$, we get by (2-2) and (2-6) that

$$\begin{aligned}\int_{B_N} |k_\alpha(x, y)|\, dy &= \text{const} \int_{|y|_K \leq q^{n-1}} \bigl(|y|_K^{\alpha-1} - q^{n(\alpha-1)}\bigr)\, dy \\ &= \text{const}\Bigl(\frac{1-q^{-1}}{1-q^{-\alpha}} q^{\alpha(n-1)} - q^{n(\alpha-1)} q^{n-1}\Bigr) = \text{const}\, |x|_K^\alpha,\end{aligned}$$

which tends to 0 as $|x|_K \to 0$. For $\alpha = 1$, we use (2-4) and (2-6) to obtain that

$$\begin{aligned}\int_{B_N} |k_1(x, y)|\, dy &= \text{const} \int_{|y|_K \leq q^{n-1}} \bigl(\log q^n - \log |y|_K\bigr)\, dy \\ &= \text{const}\Bigl(nq^{n-1} \log q - \Bigl(n - 1 - \frac{1}{q-1}\Bigr) q^{n-1} \log q\Bigr) \\ &= \text{const}\, \frac{q}{q-1} q^{n-1} \log q = \text{const}\, \frac{\log q}{q-1} |x|_K,\end{aligned}$$

and this again tends to 0 as $|x|_K \to 0$.

Thus, $\mathcal{K}$ is compact, and by the Fredholm alternative [Edwards 1965, 9.10.3], our theorem will be proved if we show that $\mathcal{K}$ has no nonzero eigenvalues.

Suppose that $\mathcal{K}w = \lambda w$, $\lambda \neq 0$, for some $w \in C(B_N)$. We have $|w(y)| \leq C$,

$$|k_\alpha(x, y)| \leq M \bigl||x|_K^{\alpha-1} - |y|_K^{\alpha-1}\bigr|,$$

if $\alpha \neq 1$, and

$$|k_1(x, y)| \leq M(\log |x|_K - \log |y|_K),$$

if $\alpha = 1$, $|y|_K < |x|_K$.

In subsequent iterations we will deal with the integrals

$$\begin{aligned}I_{\alpha,m} &= \int_{|y|_K < |x|_K} \bigl||x|_K^{\alpha-1} - |y|_K^{\alpha-1}\bigr|\, |y|_K^{\alpha m}\, dy, \quad \alpha \neq 1, \\ I_{1,m} &= \int_{|y|_K < |x|_K} \bigl(\log |x|_K - \log |y|_K\bigr) |y|_K^m\, dy.\end{aligned}$$

If $\alpha > 1$, we find denoting $|x|_K = q^n$ and using (2-2) that

$$I_{\alpha,m} = |x|_K^{\alpha-1} \int_{|y|_K \le q^{n-1}} |y|_K^{\alpha m}\, dy - \int_{|y|_K \le q^{n-1}} |y|_K^{\alpha(m+1)-1}\, dy = d_{\alpha,m} |x|_K^{\alpha(m+1)}$$

where, for all $m = 0, 1, 2, \dots$

$$\begin{aligned} d_{\alpha,m} &= \frac{1-q^{-1}}{1-q^{-\alpha m-1}} q^{-\alpha m-1} - \frac{1-q^{-1}}{1-q^{-\alpha m-\alpha}} q^{-\alpha m-\alpha} \\ &= (1-q^{-1}) \frac{q^{\alpha-1}-1}{(1-q^{-\alpha m-1})(q^{\alpha m+\alpha}-1)} \le A q^{-\alpha m} \end{aligned}$$

for some $A > 0$. A similar result,

$$I_{\alpha,m} = d_{\alpha,m} |x|_K^{\alpha(m+1)}, \quad d_{\alpha,m} \le A q^{-\alpha m}, \quad m = 0, 1, 2, \dots \tag{4-7}$$

is obtained for $0 < \alpha < 1$, so that (4-7) holds for any $\alpha \ne 1$. If $\alpha = 1$, then the integral $I_{1,m}$ is evaluated as follows. We have

$$\begin{aligned} I_{1,m} &= \sum_{k=-\infty}^{n-1} \int_{|y|_K = q^k} \big(\log |x|_K - \log |y|_K\big) |y|_K^m\, dy \\ &= \Big(1 - \frac{1}{q}\Big) \log q \sum_{k=-\infty}^{n-1} (n-k) q^{k(m+1)} \\ &= \Big(1 - \frac{1}{q}\Big) \log q \sum_{\nu=1}^{\infty} \nu q^{(n-\nu)(m+1)} = d_{1,m} |x|_K^{\alpha(m+1)} \end{aligned}$$

where

$$d_{1,m} = \Big(1 - \frac{1}{q}\Big) \log q \sum_{\nu=1}^{\infty} \nu q^{-\nu(m+1)} = \Big(1 - \frac{1}{q}\Big) \log q \frac{q^{-m-1}}{(1-q^{-m-1})^2} \le A q^{-m}$$

(we have used [Gradshteyn and Ryzhik 1996, Identity 0.231.2]). Thus, we have proved (4-7) also for $\alpha = 1$.

Let us return to a function w satisfying the relation $\mathcal{H} w = \lambda w$, $\lambda \ne 0$. Using (4-7) (separately for $\alpha \ne 1$ and $\alpha = 1$) and iterating we find by induction that

$$|w(x)| \le C (M |\lambda|^{-1} A)^m \Big(\prod_{j=0}^{m} q^{-\alpha j} \Big) |x|_K^{\alpha m}, \quad m = 0, 1, 2, \dots, x \in B_N. \tag{4-8}$$

Since $\prod_{j=0}^{m} q^{-\alpha j} = q^{-\frac{\alpha}{2} m(m+1)}$, it follows from (4-8) that $w(x) \equiv 0$. □

4.2. ***Strong solutions.*** Below we assume that the inequality (4-5) is satisfied. Then, as we have mentioned, the solution v of (4-4) is automatically extended in a unique way from B_N onto K. The extended function v satisfies (4-3). Therefore the function $u = I^\alpha v$ satisfies (4-1) in the sense of distributions from Φ'. The initial condition (4-2) is satisfied automatically.

Let us find additional conditions on a and f, under which this construction gives a strong solution of the Cauchy problem (4-1)–(4-2). Note that, by Lemma 3 and Theorem 1, a strong solution is unique in the class of functions $u = I^\alpha v$ where v is a continuous radial function, such that $\sum_{l=m}^{\infty} |v(q^l)| < \infty$ for some $m \in \mathbb{Z}$.

Theorem 2. *Suppose that*

$$|a(|x|_K)| \le C|x|_K^{-\alpha-\varepsilon}, \quad |f(|x|_K)| \le C|x|_K^{-\varepsilon}, \quad \varepsilon > 0, C > 0, \tag{4-9}$$

as $|x|_K > 1$. *Then* $u = I^\alpha v$ *is a strong solution of the Cauchy problem* (4-1)–(4-2).

Proof. Let $v(|x|_K)$ be the solution of (4-3$'$) constructed above for all $x \in K$ (for $x = 0$, the integral in the right-hand side is assumed equal to zero). For $|x|_K \le q^N$ the existence of a solution v was obtained from the theory of compact operators; for larger values of $|x|_K$ we use successively (4-3$'$) itself. Denote

$$V_m = \sup_{|x|_K \le q^m} |v(q^m)|.$$

The sequence $\{V_m\}$ is nondecreasing.

As we assumed in Theorem 1 only the continuity of the coefficient a, we took N in such a way that the neighborhood $B_N = \{x : |x|_K \le q^N\}$ was sufficiently small. Here we assume (4-5), so that we can take any fixed integer N and obtain a solution v on B_N.

Consider the case where $\alpha \ne 1$. It follows from (4-5) and (4-9) that

$$\left|[1 + q^{-\alpha} a(|x|_K)\,|x|_K^\alpha]^{-1}\right| \le H$$

where $H > 0$ does not depend on $x \in K$. If $m \ge N$, then we find from (4-3$'$) and the above estimate for $I_{\alpha,0}$ that

$$|v(q^m)| \le c_\alpha d_{\alpha,0} H a(q^m) q^{\alpha m} V_{m-1} + H|f(g^m)|. \tag{4-10}$$

Let us choose $m_1 \ge N$ so big that

$$c_\alpha d_{\alpha,0} H a(q^m) q^{\alpha m} \le \tfrac{1}{2}, \quad H|f(g^m)| \le \tfrac{1}{2} V_{N-1},$$

as $m \ge m_1$ (that is possible due to (4-9)). Then it follows from (4-10) that $V_m \le V_{m-1}$, as $m \ge m_1$, hence that the function v is bounded on K.

Now we get from (4-3$'$) and the assumptions (4-9) that

$$\left|v(|x|_K)\right| \le C|x|_K^{-\varepsilon}, \quad |x|_K \ge 1, \tag{4-11}$$

$C > 0$. A similar reasoning works for $\alpha = 1$.

Taking into account the estimate (4-11) we find from Lemma 3 that

$$(D^\alpha I^\alpha v)(|x|_K) = v(|x|_K), \quad x \neq 0.$$

Therefore the function $u = I^\alpha v$ satisfies (4-1) for all $x \neq 0$. Since a, f, and u are continuous, the equation is satisfied in the strong sense. □

4.3. ***Generalizations.*** Instead of (4-2), one can consider an inhomogeneous initial condition $u(0) = u_0$, $u_0 \in \mathbb{C}$. Looking for a solution in the form $u = u_0 + I^\alpha v$, $v = v(|x|_K)$, we obtain the integral equation

$$v(|x|_K) + a(|x|_K)(I^\alpha v)(|x|_K) = f(|x|_K) - a(|x|_K)u_0,$$

which can be studied under the same assumptions.

All the above results carry over to the case of a matrix-valued coefficient $a(|x|_K)$ and vector-valued solutions. In this case, to obtain a strong solution, it is sufficient to demand that the spectrum of each matrix $a(|x|_K)$, $x \in K$, does not intersect the set $\{-q^N : N \in \mathbb{Z}\}$.

Acknowledgement

The author is grateful to M. M. Malamud for a helpful discussion.

References

[Albeverio et al. 2010] S. Albeverio, A. Y. Khrennikov, and V. M. Shelkovich, *Theory of p-adic distributions: Linear and nonlinear models*, London Mathematical Society Lecture Note Series **370**, Cambridge University Press, 2010. MR 2011f:46038 Zbl 1198.46001

[Edwards 1965] R. E. Edwards, *Functional analysis: Theory and applications*, Holt, Rinehart and Winston, New York, 1965. MR 36 #4308 Zbl 0182.16101

[Gradshteyn and Ryzhik 1996] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 5th ed., Academic Press, San Diego, 1996. MR 97c:00014 Zbl 0918.65001

[Kilbas et al. 2006] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and applications of fractional differential equations*, North-Holland Mathematics Studies **204**, Elsevier, Amsterdam, 2006. MR 2007a:34002 Zbl 1092.45003

[Kochubei 2001] A. N. Kochubei, *Pseudo-differential equations and stochastics over non-Archimedean fields*, Monographs and Textbooks in Pure and Applied Mathematics **244**, Marcel Dekker, New York, 2001. MR 2003b:35220 Zbl 0984.11063

[Kochubei 2008] A. N. Kochubei, "A non-Archimedean wave equation", *Pacific J. Math.* **235**:2 (2008), 245–261. MR 2009e:35305 Zbl 1190.35235

[Vladimirov 2003] V. S. Vladimirov, *Tables of integrals of complex-valued functions of p-adic arguments*, Current Problems in Mathematics **2**, Steklov Mathematical Institute, Moscow, 2003. In Russian; translated in ArXiv: math-ph/9911027. MR 2006a:11164 Zbl 1062.11003

[Vladimirov et al. 1994] V. S. Vladimirov, I. V. Volovich, and E. I. Zelenov, *p-adic analysis and mathematical physics*, Series on Soviet and East European Mathematics **1**, World Scientific, River Edge, NJ, 1994. MR 95k:11155 Zbl 0812.46076

[Zúñiga-Galindo 2011] W. A. Zúñiga-Galindo, "Local zeta functions and fundamental solutions for pseudo-differential operators over p-adic fields", *p-Adic Numbers Ultrametric Anal. Appl.* **3**:4 (2011), 344–358. MR 2854684 Zbl 06105091

Received February 20, 2013.

ANATOLY N. KOCHUBEI
INSTITUTE OF MATHEMATICS
NATIONAL ACADEMY OF SCIENCES OF UKRAINE
TERESHCHENKIVSKA 3
KIEV, 01601
UKRAINE
kochubei@i.com.ua

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.371

A JANTZEN SUM FORMULA FOR RESTRICTED VERMA MODULES OVER AFFINE KAC–MOODY ALGEBRAS AT THE CRITICAL LEVEL

JOHANNES KÜBEL

For a restricted Verma module of an affine Kac–Moody algebra at the critical level we describe the Jantzen filtration and calculate its character. This corresponds to the Jantzen sum formula of a baby Verma module over a modular Lie algebra. This also implies a new proof of the linkage principle which was already derived by Arakawa and Fiebig.

1. Introduction

To a simple complex Lie algebra $\mathfrak{g}$ with Cartan subalgebra $\mathfrak{h}$, one associates an affine Kac–Moody Lie algebra $\widehat{\mathfrak{g}}$ with Cartan subalgebra $\widehat{\mathfrak{h}}$. The root system R of $\mathfrak{g}$ can be embedded into the root system $\widehat{R}$ of $\widehat{\mathfrak{g}}$. Arakawa and Fiebig [2012a] introduced the category $\overline{\mathcal{O}}_c$ of restricted representations of $\widehat{\mathfrak{g}}$ at the critical level. Denote by $\mathcal{O}_c$ the direct summand of the usual highest weight category $\mathcal{O}$ over $\widehat{\mathfrak{g}}$ which consists of modules with critical level. Then $\overline{\mathcal{O}}_c$ is the subcategory of $\mathcal{O}_c$ on which those elements of the Feigin–Frenkel center act by zero which are homogeneous of degree unequal to zero. We call $\overline{\mathcal{O}}_c$ the *restricted category* $\mathcal{O}$ and its objects *restricted modules*.

Conjecturally, the restricted category $\overline{\mathcal{O}}_c$ should have the same structure as the representation category over a small quantum group or a modular Lie algebra described in [Andersen et al. 1994]. The standard modules in $\overline{\mathcal{O}}_c$, which should correspond to baby Verma modules in the representation categories of [Andersen et al. 1994], are the so called *restricted Verma modules*. They are the maximal restricted quotients of the ordinary Verma modules.

Towards the description of $\overline{\mathcal{O}}_c$, Arakawa and Fiebig [2012b] confirmed the above conjecture in the subgeneric case, and Frenkel [2005, Theorem 4.8] did so in the generic case. Andersen, Jantzen, and Soergel [Andersen et al. 1994, Chapter 6]

The author is supported by the DFG priority program 1388.
MSC2010: 17B10, 17B67.
Keywords: restricted Verma modules at the critical level, Jantzen sum formula, affine Kac–Moody algebras, critical representations of affine Kac–Moody algebras, Jantzen filtration, category $\mathcal{O}$.

and Kumar and Letzter [1997] computed a Jantzen sum formula for a baby Verma module $Z(\lambda)$ which describes the characters of the Jantzen filtration as an alternating sum formula of certain characters of baby Verma modules of weight "lower" than λ.

We deduce the analogous formula for restricted Verma modules at the critical level. To be more precise, let $\lambda \in \widehat{\mathfrak{h}}^*$ be a weight of critical level. We introduce the Jantzen filtration

$$\overline{\Delta}(\lambda) = \overline{\Delta}(\lambda)^0 \supset \overline{\Delta}(\lambda)^1 \supset \overline{\Delta}(\lambda)^2 \supset \cdots ,$$

and deduce the formula

$$\sum_{i>0} \operatorname{ch} \overline{\Delta}(\lambda)^i = \sum_{\alpha \in R(\lambda)^+} \Bigl(\sum_{i>0} (\operatorname{ch} \overline{\Delta}(\alpha \downarrow^{2i-1} \lambda) - \operatorname{ch} \overline{\Delta}(\alpha \downarrow^{2i} \lambda)) \Bigr).$$

Here $R(\lambda)^+ \subset R$ denotes the positive roots of the finite root system R which are integral on λ. The notation $\alpha \downarrow^i \lambda$ for $i > 0$ is defined inductively by $\alpha \downarrow (\alpha \downarrow^{i-1} \lambda)$, where $\alpha \downarrow \lambda = s_\alpha \cdot \lambda$ if $s_\alpha \cdot \lambda \leq \lambda$ and $\alpha \downarrow \lambda = s_{-\alpha+\delta} \cdot \lambda$ if $s_\alpha \cdot \lambda > \lambda$. Here $\delta \in \widehat{R}$ denotes the smallest positive imaginary root and $s_\alpha, s_{-\alpha+\delta}$ are the reflections corresponding to the real roots $\alpha, -\alpha+\delta$ of the affine Weyl group $\widehat{\mathcal{W}}$ with its dot-action on $\widehat{\mathfrak{h}}^*$.

The strategy to prove the Jantzen sum formula is to deduce the subgeneric cases first and then put these together to get the general result in a very similar manner as in [Jantzen 1979, Chapters 5.6 and 5.7]. To deduce the subgeneric case we use a result of [Arakawa and Fiebig 2012b] which states that for $\lambda \in \widehat{\mathfrak{h}}^*$ critical and subgeneric the maximal submodule of $\overline{\Delta}(\lambda)$ is isomorphic to the simple module $L(\alpha \downarrow \lambda)$ with highest weight $\alpha \downarrow \lambda$.

Arakawa and Fiebig [2012a] introduced projective objects in the restricted category $\overline{\mathcal{O}}_c$ and a BGGH-reciprocity to deduce the linkage principle for restricted Verma modules conjectured by Feigin and Frenkel. It states that if the simple module with highest weight $\mu \in \widehat{\mathfrak{h}}^*$ appears as a subquotient in a Jordan–Hölder series of $\overline{\Delta}(\lambda)$, where $\lambda \in \widehat{\mathfrak{h}}^*$ is critical, then $\mu \in \widehat{\mathcal{W}}(\lambda) \cdot \lambda$. Here $\widehat{\mathcal{W}}(\lambda)$ denotes the integral affine Weyl group of the root system $\widehat{R}(\lambda)$ corresponding to λ with its dot-action on $\widehat{\mathfrak{h}}^*$.

The linkage principle immediately follows from the Jantzen sum formula and thus gives an independent proof.

2. Preliminaries

In this chapter we shortly introduce the construction of an (untwisted) affine Kac–Moody algebra starting with a simple Lie algebra. We collect some facts about the root data, the Weyl group and the invariant bilinear form. The results and definitions in this section can be found in [Kac 1990] and [Kac and Kazhdan 1979].

Affine Kac–Moody algebras. Let $\mathfrak{g}$ be a simple Lie algebra with a Borel subalgebra $\mathfrak{b}$ and a Cartan subalgebra $\mathfrak{h}$. We denote by R the root system with positive roots R^+ and by Π the simple roots. Moreover, denote by $\mathcal{W}$ the finite Weyl group and by $\kappa : \mathfrak{g} \times \mathfrak{g} \to \mathbb{C}$ the Killing form.

We take a nonsplit central extension $\widetilde{\mathfrak{g}}$ of the loop algebra $\mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}]$. As a vector space, $\widetilde{\mathfrak{g}}$ is the direct sum $(\mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}]) \oplus \mathbb{C}c$, where c is a central element.

Adding a derivation operator d with the property $[d, \cdot] = t(\partial/\partial t)$, we get the affine Kac–Moody algebra $\widehat{\mathfrak{g}}$ associated to $\mathfrak{g}$. As a vector space, we have $\widehat{\mathfrak{g}} = (\mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}]) \oplus \mathbb{C}c \oplus \mathbb{C}d$ and the Lie bracket is given by

$$\begin{aligned}
[c, \widehat{\mathfrak{g}}] &= \{0\}, \\
[d, x \otimes t^n] &= n x \otimes t^n, \\
[x \otimes t^n, y \otimes t^m] &= [x, y] \otimes t^{m+n} + n\delta_{m,-n}\kappa(x, y)c,
\end{aligned}$$

where $x, y \in \mathfrak{g}$ and $n \in \mathbb{Z}$. The Borel subalgebra of $\widehat{\mathfrak{g}}$ corresponding to $\mathfrak{b} \subset \mathfrak{g}$ is given by

$$\widehat{\mathfrak{b}} = (\mathfrak{g} \otimes_{\mathbb{C}} t\mathbb{C}[t] + \mathfrak{b} \otimes_{\mathbb{C}} \mathbb{C}[t]) \oplus \mathbb{C}c \oplus \mathbb{C}d,$$

while the corresponding Cartan subalgebra of $\widehat{\mathfrak{g}}$ is given by

$$\widehat{\mathfrak{h}} = \mathfrak{h} \oplus \mathbb{C}c \oplus \mathbb{C}d.$$

Affine roots, Weyl groups and bilinear forms. For a vector space V we denote by $\langle \cdot, \cdot \rangle : V^* \times V \to \mathbb{C}$ the natural pairing with its dual space. Denote by $\leq$ the usual ordering on $\widehat{\mathfrak{h}}^*$; that is, $\lambda \leq \mu$ for $\lambda, \mu \in \widehat{\mathfrak{h}}^*$ if $\mu - \lambda$ can be expressed as a finite sum of positive roots. The projection $\widehat{\mathfrak{h}} = \mathfrak{h} \oplus \mathbb{C}c \oplus \mathbb{C}d \to \mathfrak{h}$ induces an embedding $\mathfrak{h}^* \subset \widehat{\mathfrak{h}}^*$. By this embedding we can consider all finite roots as elements of $\widehat{\mathfrak{h}}^*$. We define two weights $\Lambda_\circ, \delta \in \widehat{\mathfrak{h}}^*$ by the relations

$$\begin{aligned}
\langle \delta, \mathfrak{h} \oplus \mathbb{C}c \rangle &= \{0\}, \\
\langle \delta, d \rangle &= 1, \\
\langle \Lambda_\circ, \mathfrak{h} \oplus \mathbb{C}d \rangle &= \{0\}, \\
\langle \Lambda_\circ, c \rangle &= 1.
\end{aligned}$$

Then the roots of $\widehat{\mathfrak{g}}$ with respect to $\widehat{\mathfrak{h}}$ are given by $\widehat{R} = \widehat{R}^{\mathrm{re}} \cup \widehat{R}^{\mathrm{im}}$, where

$$\widehat{R}^{\mathrm{re}} := \{\alpha + n\delta \mid \alpha \in R \subset \widehat{R},\ n \in \mathbb{Z}\}$$

are the *real roots*, and

$$\widehat{R}^{\mathrm{im}} := \{n\delta \mid n \in \mathbb{Z}, n \neq 0\}$$

are the *imaginary roots*. The positive roots $\widehat{R}^+$, that is, the roots of $\widehat{\mathfrak{b}}$ with respect to $\widehat{\mathfrak{h}}$, are then given

$$\widehat{R}^+ = R^+ \cup \{\alpha + n\delta \mid \alpha \in R, n > 0\} \cup \{n\delta \mid n > 0\}.$$

Denote by θ the highest root of R. Then the set of simple affine roots is given by

$$\widehat{\Pi} = \Pi \cup \{-\theta + \delta\} \subset \widehat{R}^+.$$

For a real root $\alpha \in \widehat{R}^{\mathrm{re}}$ we denote by $\alpha^\vee \in \widehat{\mathfrak{h}}$ its *coroot* which is uniquely defined by the properties $\alpha^\vee \in [\widehat{\mathfrak{g}}_\alpha, \widehat{\mathfrak{g}}_{-\alpha}]$ and $\langle \alpha, \alpha^\vee \rangle = 2$.

We denote by $\widehat{\mathcal{W}} \subset \mathrm{Gl}(\widehat{\mathfrak{h}}^*)$ the *affine Weyl group* of the root system $\widehat{R}$, which is the subgroup generated by the reflections $s_\alpha : \widehat{\mathfrak{h}}^* \to \widehat{\mathfrak{h}}^*$, $\lambda \mapsto \lambda - \langle \lambda, \alpha^\vee \rangle \alpha$, where $\alpha \in \widehat{R}^{\mathrm{re}}$ is a real root. We can identify the finite Weyl group $\mathcal{W}$ with the subgroup of $\widehat{\mathcal{W}}$ generated by the reflections s_α corresponding to finite roots $\alpha \in R$. Then $\mathcal{W}$ stabilizes the subspace $\mathfrak{h}^* \subset \widehat{\mathfrak{h}}^*$.

Let $\bar{\rho} := \frac{1}{2} \sum_{\alpha \in R^+} \alpha$ be the half-sum of positive finite roots. We then set

$$\rho := \bar{\rho} + h^\vee \Lambda_\circ,$$

where $h^\vee$ is the dual Coxeter number of $\mathfrak{g}$. Then $\langle \rho, \alpha^\vee \rangle \neq 0$ for all $\alpha \in \widehat{R}^{\mathrm{re}}$ and $\langle \rho, c \rangle \neq 0$ as well. We can now define the ρ-shifted dot-action of $\widehat{\mathcal{W}}$ on $\widehat{\mathfrak{h}}^*$ by

$$w \cdot \lambda := w(\lambda + \rho) - \rho,$$

where $w \in \widehat{\mathcal{W}}$ and $\lambda \in \widehat{\mathfrak{h}}^*$.

The Killing form κ on the simple Lie algebra $\mathfrak{g}$ extends to a bilinear form $(\cdot \mid \cdot) : \widehat{\mathfrak{g}} \times \widehat{\mathfrak{g}} \to \mathbb{C}$ which is given by the equations

$$\begin{aligned} (x \otimes t^n \mid y \otimes t^m) &= \delta_{n,-m} \kappa(x, y), \\ (c \mid \mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}c) &= \{0\}, \\ (d \mid \mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}d) &= \{0\}, \\ (c \mid d) &= 1, \end{aligned}$$

for $x, y \in \mathfrak{g}$ and $m, n \in \mathbb{Z}$. It is nondegenerate, symmetric and invariant, i.e., $([x, y] \mid z) = (x \mid [y, z])$ for all $x, y, z \in \widehat{\mathfrak{g}}$. Furthermore, it induces a nondegenerate bilinear form on the affine Cartan subalgebra and thus an isomorphism $\nu : \widehat{\mathfrak{h}} \xrightarrow{\sim} \widehat{\mathfrak{h}}^*$ which coincides with the isomorphism $\mathfrak{h} \xrightarrow{\sim} \mathfrak{h}^*$ induced by the Killing form, when restricted to the finite Cartan subalgebra, and which sends c to δ and d to $\Lambda_\circ$. So the induced form on $\widehat{\mathfrak{h}}^*$ is given by

$$\begin{aligned} (\alpha \mid \beta) &= \kappa(\alpha, \beta), \\ (\Lambda_\circ \mid \mathfrak{h}^* \oplus \mathbb{C}\Lambda_\circ) &= \{0\}, \\ (\delta \mid \mathfrak{h}^* \oplus \mathbb{C}\delta) &= \{0\}, \\ (\Lambda_\circ \mid \delta) &= 1, \end{aligned}$$

for $\alpha, \beta \in \mathfrak{h}^*$ and κ the induced Killing form on $\mathfrak{h}^*$. The induced form is invariant under the linear action of the affine Weyl group, i.e.,

$$(w(\lambda) \mid w(\mu)) = (\lambda \mid \mu)$$

for $\lambda, \mu \in \widehat{\mathfrak{h}}^*$, $w \in \widehat{\mathcal{W}}$.

3. Verma modules

For a Lie algebra $\mathfrak{a}$ we denote by $U(\mathfrak{a})$ its universal enveloping algebra. For $\lambda \in \widehat{\mathfrak{h}}^*$ let $\mathbb{C}_\lambda$ be the one-dimensional representation of $U(\widehat{\mathfrak{b}})$ on which $\widehat{\mathfrak{h}}$ acts by the character λ and $[\widehat{\mathfrak{b}}, \widehat{\mathfrak{b}}]$ by zero. Then the *Verma module with highest weight* λ is defined by

$$\Delta(\lambda) := U(\widehat{\mathfrak{g}}) \otimes_{U(\widehat{\mathfrak{b}})} \mathbb{C}_\lambda.$$

It has a unique simple quotient, which we denote by $L(\lambda)$, and both modules are highest weight modules with highest weight λ.

Deformed Verma modules. Denote by $\widehat{S} := S(\widehat{\mathfrak{h}}) = U(\widehat{\mathfrak{h}})$ the symmetric algebra over the vector space $\widehat{\mathfrak{h}}$ and by $S = S(\mathfrak{h})$ the symmetric algebra over the vector space $\mathfrak{h}$. Then the projection $\widehat{\mathfrak{h}} \to \mathfrak{h}$ induces a homomorphism $\widehat{S} \to S$ and equips S with an $\widehat{S}$-algebra structure. We call a commutative, unital, noetherian $\widehat{S}$-algebra with structure morphism $\tau : \widehat{S} \to A$ a *deformation algebra*.

For a Lie algebra $\mathfrak{a}$ we set $\mathfrak{a}_A := \mathfrak{a} \otimes_{\mathbb{C}} A$. Then we can identify $(\widehat{\mathfrak{h}}_A)^* = \operatorname{Hom}_A(\widehat{\mathfrak{h}}_A, A)$ with $\widehat{\mathfrak{h}}^* \otimes_{\mathbb{C}} A$ and any weight $\lambda \in \widehat{\mathfrak{h}}^*$ induces a weight $\lambda \otimes 1 \in \widehat{\mathfrak{h}}_A^*$, which we simply denote by λ again. In this way, the composition $\widehat{\mathfrak{h}} \hookrightarrow \widehat{S} \xrightarrow{\tau} A$ induces the *canonical weight* $\tau \in \widehat{\mathfrak{h}}_A^*$.

For $\lambda \in \widehat{\mathfrak{h}}^*$ let A_λ be the $\widehat{\mathfrak{b}}_A$-module which is A as an A-module and on which $\widehat{\mathfrak{h}}$ acts via the character $\lambda + \tau$ and $[\widehat{\mathfrak{b}}, \widehat{\mathfrak{b}}]$ by zero. We then define the *deformed Verma module with highest weight* λ by

$$\Delta_A(\lambda + \tau) := U(\widehat{\mathfrak{g}}_A) \otimes_{U(\widehat{\mathfrak{b}}_A)} A_\lambda.$$

For an $\widehat{\mathfrak{h}}_A$-module M and $\lambda \in \widehat{\mathfrak{h}}^*$ we define the *deformed weight space* of λ by

$$M_\lambda := \{m \in M \mid Hm = (\lambda + \tau)(H)m \text{ for all } H \in \widehat{\mathfrak{h}}_A\}.$$

Then the deformed Verma module $\Delta_A(\lambda + \tau)$ decomposes as $\widehat{\mathfrak{h}}_A$-module into the direct sum of its weight spaces $\Delta_A(\lambda + \tau)_\mu$ with $\mu \in \widehat{\mathfrak{h}}^*$, such that $\Delta_A(\lambda + \tau)_\mu \neq 0$ implies $\mu \leq \lambda$.

If $A \to A'$ is a homomorphism of deformation algebras with structure maps $\tau : \widehat{S} \to A$ and $\tau' : \widehat{S} \to A \to A'$, then

$$\Delta_A(\lambda + \tau) \otimes_A A' \cong \Delta_{A'}(\lambda + \tau').$$

Note that for $\tau : \widehat{S} \to \mathbb{C}$, the surjection on the quotient $\mathbb{C} \cong \widehat{S}/\widehat{\mathfrak{h}}\widehat{S}$, we have $\Delta_{\mathbb{C}}(\lambda+\tau) \cong \Delta(\lambda)$.

Characters. Let $\mathbb{Z}[\widehat{\mathfrak{h}}^*] = \bigoplus_{\lambda\in\widehat{\mathfrak{h}}^*} \mathbb{Z}e^{\lambda}$ be the group algebra of $\widehat{\mathfrak{h}}^*$. We define a certain completion by

$$\widehat{\mathbb{Z}[\widehat{\mathfrak{h}}^*]} \subset \prod_{\lambda\in\widehat{\mathfrak{h}}^*} \mathbb{Z}e^{\lambda}$$

to be the subgroup of elements (c_λ) with the property that there exists a finite subset $\{\mu_1, \dots, \mu_n\} \subset \widehat{\mathfrak{h}}^*$ such that $c_\lambda \neq 0$ implies $\lambda \leq \mu_i$ for at least one i. Let $M \in \widehat{\mathfrak{g}}$-mod be semisimple over $\widehat{\mathfrak{h}}$ with the properties that each weight space M_λ is finite-dimensional and that there exists $\mu_1, \dots, \mu_n \in \widehat{\mathfrak{h}}^*$ such that $M_\lambda \neq 0$ implies $\lambda \leq \mu_i$ for at least one i. We define the *character* of M as element in $\widehat{\mathbb{Z}[\widehat{\mathfrak{h}}^*]}$ given by the formal sum

$$\operatorname{ch} M := \sum_{\lambda\in\widehat{\mathfrak{h}}^*} (\dim_{\mathbb{C}} M_\lambda) e^{\lambda}.$$

We define the *generalized Kostant partition function* $\mathcal{P} : \mathbb{Z}\widehat{R}^+ \to \mathbb{N}$ by

$$\mathcal{P}(\nu) := \begin{cases} \dim_{\mathbb{C}} \Delta(0)_\nu & \text{if } \nu \in \mathbb{N}\widehat{R}^+, \\ 0 & \text{otherwise.} \end{cases}$$

The name partition function comes from the combinatorial description of the formula

$$\operatorname{ch} \Delta(\lambda) = \prod_{\alpha\in\widehat{R}^+} (1 + e^{-\alpha} + e^{-2\alpha} + \cdots)^{\dim \widehat{\mathfrak{g}}_\alpha}$$

(compare [Kac 1990, Section 9.7]).

4. Restricted Verma modules

An equivalence relation. For a deformation algebra A with canonical weight $\tau : \widehat{\mathfrak{h}}_A \to A$, we extend the bilinear form $(\cdot\,|\,\cdot) : \widehat{\mathfrak{h}}^* \times \widehat{\mathfrak{h}}^* \to \mathbb{C}$ to an A-linear continuation $(\cdot\,|\,\cdot)_A : \widehat{\mathfrak{h}}_A^* \times \widehat{\mathfrak{h}}_A^* \to A$.

Let $A = \mathbb{K}$ be a field. For $\nu, \lambda \in \widehat{\mathfrak{h}}^*$ we write $\nu \preceq_{\mathbb{K}} \lambda$ if there exists $n \in \mathbb{N}$ and $\alpha \in \widehat{R}^+$ such that $2(\lambda+\rho+\tau \mid \alpha)_{\mathbb{K}} = n(\alpha \mid \alpha)_{\mathbb{K}}$ and $\nu = \lambda - n\alpha$. We now denote by $\preceq_{\mathbb{K}}$ the partial ordering on $\widehat{\mathfrak{h}}^*$ which is generated by such tuples $\nu \preceq_{\mathbb{K}} \lambda$. Then $\preceq_{\mathbb{K}}$ is a refinement of the usual ordering $\leq$ on $\widehat{\mathfrak{h}}^*$. We denote by $\sim_{\mathbb{K}}$ the equivalence relation on $\widehat{\mathfrak{h}}^*$ which is generated by $\preceq_{\mathbb{K}}$.

Let $L_{\mathbb{K}}(\lambda+\tau)$ be the unique simple quotient of $\Delta_{\mathbb{K}}(\lambda+\tau)$ and denote by $[\Delta_{\mathbb{K}}(\lambda+\tau) : L_{\mathbb{K}}(\mu+\tau)]$ the number of subquotients of a composition series of $\Delta_{\mathbb{K}}(\lambda+\tau)$ which are isomorphic to $L_{\mathbb{K}}(\mu+\tau)$.

Theorem 4.1 [Kac and Kazhdan 1979, Theorem 2]. *We have*

$$[\Delta_{\mathbb{K}}(\lambda+\tau) : L_{\mathbb{K}}(\mu+\tau)] \neq 0$$

if and only if $\mu \preceq_{\mathbb{K}} \lambda$.

Remark 4.2. From [Rocha-Caridi and Wallach 1982, Theorem 15] we know that $[\Delta_{\mathbb{K}}(\lambda+\tau) : L_{\mathbb{K}}(\mu+\tau)] \neq 0$ if and only if there exists an embedding $\Delta_{\mathbb{K}}(\mu+\tau) \hookrightarrow \Delta_{\mathbb{K}}(\lambda+\tau)$. Thus Theorem 4.1 also contains information about embeddings of Verma modules.

For $\lambda \in \widehat{\mathfrak{h}}^*$ we define the *integral roots* (with respect to λ and A) by

$$\widehat{R}_A(\lambda) := \{\alpha \in \widehat{R} \mid 2(\lambda+\rho+\tau \mid \alpha)_A \in \mathbb{Z}(\alpha \mid \alpha)_A\},$$

and the corresponding *integral Weyl group* by

$$\widehat{\mathcal{W}}_A(\lambda) := \langle s_\alpha \mid \alpha \in \widehat{R}_A(\lambda) \cap \widehat{R}^{\mathrm{re}} \rangle \subset \widehat{\mathcal{W}}.$$

We write $\widehat{R}_A(\lambda)^+ = \widehat{R}^+ \cap \widehat{R}_A(\lambda)$ and $\widehat{R}(\lambda) = \widehat{R}_{\mathbb{C}}(\lambda)$ in case $\tau : \widehat{S} \to \widehat{S}/\widehat{\mathfrak{h}}\widehat{S} \cong \mathbb{C}$ is the quotient map and similarly $\widehat{\mathcal{W}}(\lambda) = \widehat{\mathcal{W}}_{\mathbb{C}}(\lambda)$.

The critical level. For $\lambda \in \widehat{\mathfrak{h}}^*$ we define the *level* of λ to be the complex number $\lambda(c) \in \mathbb{C}$. If $\lambda \sim_{\mathbb{K}} \mu$, we have $\lambda(c) = \mu(c)$. Therefore, the equivalence class Λ of λ has a well-defined level, and we have $\nu(c) = (\nu \mid \delta)$ for all $\nu \in \widehat{\mathfrak{h}}^*$.

Lemma 4.3 [Arakawa and Fiebig 2012b, Lemma 4.2]. *For* $\Lambda \in \widehat{\mathfrak{h}}^*/\sim_{\mathbb{K}}$ *the following are equivalent.*

(1) *We have* $\lambda(c) = -\rho(c)$ *for some* $\lambda \in \Lambda$.

(2) *We have* $\lambda(c) = -\rho(c)$ *for all* $\lambda \in \Lambda$.

(3) *We have* $\lambda + \delta \in \Lambda$ *for all* $\lambda \in \Lambda$.

(4) *We have* $n\delta \in \widehat{R}_{\mathbb{K}}(\lambda)$ *for all* $n \neq 0$ *and* $\lambda \in \Lambda$.

We call $\mathrm{crit} := -\rho(c)$ the *critical level*.

Denote by $\widehat{\mathfrak{h}}^*_{\mathrm{crit}}$ the hyperplane which consists of all $\lambda \in \widehat{\mathfrak{h}}^*$ with $\lambda(c) = \mathrm{crit}$. Then for each $\lambda \in \widehat{\mathfrak{h}}^*_{\mathrm{crit}}$ we have $(\lambda+\rho \mid \delta) = 0$.

Restricted Verma modules. Let $\lambda \in \widehat{\mathfrak{h}}^*$ and $\tau : \widehat{S} \to A$ be a deformation algebra which is an integral domain. Denote by $Q(A)$ its field of fractions and assume that both structure maps factor through the restriction map $\widehat{S} \to S$. This implies that $\tau(c) = \tau(d) = 0$. We define $\Delta_A^-(\lambda+\tau)$ to be the submodule of $\Delta_A(\lambda+\tau)$ which is generated by the images of all homomorphisms $\Delta_A(\lambda - n\delta + \tau) \to \Delta_A(\lambda+\tau)$ for $n \in \mathbb{N}_{>0}$. Since $\tau(c) = \tau(d) = 0$, we have $(\tau \mid \delta) = 0$, and by Theorem 4.1 and its remark there is an injective map $\Delta_{Q(A)}(\lambda - n\delta + \tau) \hookrightarrow \Delta_{Q(A)}(\lambda+\tau)$ for every $n > 0$ and λ critical. But by our assumption on A, this also induces an injective map $\Delta_A(\lambda - n\delta + \tau) \hookrightarrow \Delta_A(\lambda + \tau)$. If λ is noncritical, we get $\Delta_A^-(\lambda + \tau) = \{0\}$. We now define the *restricted Verma module* as the quotient

$$\overline{\Delta}_A(\lambda + \tau) = \Delta_A(\lambda + \tau)/\Delta_A^-(\lambda + \tau).$$

As in the nonrestricted case, we omit the subscript of the restricted Verma modules if the deformation algebra is $\mathbb{C} \cong S/S\mathfrak{h}$. For example, we write $\overline{\Delta}(\lambda)$ instead of $\overline{\Delta}_{\mathbb{C}}(\lambda)$.

Remark 4.4. There is an alternative definition of restricted Verma modules. Denote by $V^{\text{crit}}(\mathfrak{g})$ the universal affine vertex algebra at the critical level and denote by $\mathfrak{z}$ its center. Then each smooth $[\widehat{\mathfrak{g}}, \widehat{\mathfrak{g}}]$-module M carries the structure of a graded $\mathfrak{z}$-module. By a theorem of Feigin and Frenkel [1992], $\mathfrak{z}$ yields an action on M by the graded polynomial ring generated by infinitely many homogeneous elements

$$\mathfrak{X} = \mathbb{C}[p_s^{(i)} \mid i = 1, \dots, l, s \in \mathbb{Z}] = \bigoplus_{n \in \mathbb{Z}} \mathfrak{X}_n.$$

Now a theorem of Frenkel and Gaitsgory [2006] shows that for any critical weight $\lambda \in \widehat{\mathfrak{h}}^*$ and $n < 0$ there is a surjective map $\mathfrak{X}_n \to \operatorname{Hom}_{\widehat{\mathfrak{g}}}(\Delta(\lambda + n\delta), \Delta(\lambda))$. Thus, the restricted Verma module $\overline{\Delta}(\lambda)$ coincides with the quotient

$$\Delta(\lambda)^{\text{res}} := \Delta(\lambda) / \sum_{n<0} \mathfrak{X}_n \Delta(\lambda).$$

However, we will not use this alternative description of restricted Verma modules in the rest of this paper.

Let $\bar{\cdot} : \widehat{\mathfrak{h}}^* \twoheadrightarrow \mathfrak{h}^*$ be the map induced by $\mathfrak{h} \hookrightarrow \widehat{\mathfrak{h}}$. For any subset $\Lambda \subset \widehat{\mathfrak{h}}^*$ we denote by $\bar{\Lambda} \subset \mathfrak{h}^*$ its image under $\bar{\cdot}$.

Definition 4.5. Let $\Lambda \in \widehat{\mathfrak{h}}^*_{\text{crit}}/\sim_{\mathbb{K}}$ be a critical equivalence class. We call Λ

(1) *generic* if $\bar{\Lambda} \subset \mathfrak{h}^*$ contains exactly one element;

(2) *subgeneric* if $\bar{\Lambda} \subset \mathfrak{h}^*$ contains exactly two elements.

We call any weight contained in a generic (subgeneric, resp.) equivalence class a generic (subgeneric, resp.) weight. If Λ is subgeneric, there is a weight $\bar{\lambda} \in \mathfrak{h}^*$ and a finite root $\alpha \in R$ such that $\bar{\Lambda} = \{\bar{\lambda}, s_\alpha \cdot \bar{\lambda}\}$.

Let $\lambda \in \widehat{\mathfrak{h}}^*_{\text{crit}}$ be a critical weight. Similarly to the integral roots of λ we now define the *finite integral root system* (with respect to λ and the deformation algebra A) by

$$R_A(\lambda) := \widehat{R}_A(\lambda) \cap R = \{\alpha \in R \mid 2(\lambda + \rho + \tau \mid \alpha)_A \in \mathbb{Z}(\alpha \mid \alpha)_A\},$$

and the *finite integral Weyl group* by

$$\mathcal{W}_A(\lambda) = \widehat{\mathcal{W}}_A(\lambda) \cap \mathcal{W}.$$

Again we write $R_A(\lambda)^+ = R^+ \cap R_A(\lambda)$ and $R(\lambda) = R_{\mathbb{C}}(\lambda)$ if the deformation algebra is $\mathbb{C}$. For $\lambda \in \widehat{\mathfrak{h}}^*_{\mathrm{crit}}$ and $\alpha \in R_A(\lambda)$, such that $s_\alpha \cdot \lambda \neq \lambda$, we have either $s_\alpha \cdot \lambda < \lambda$ or $s_{-\alpha+\delta} \cdot \lambda < \lambda$. We define $\alpha \downarrow \lambda$ to be the element in the set $\{s_\alpha \cdot \lambda, s_{-\alpha+\delta} \cdot \lambda\}$ which is smaller than λ. Furthermore, we define inductively $\alpha \downarrow^n \lambda := \alpha \downarrow (\alpha \downarrow^{n-1} \lambda)$. In case $s_\alpha \cdot \lambda = \lambda$ we have $\alpha \downarrow \lambda = \lambda$.

Now [2012b, Corollary 4.10] gives in our set up the following theorem:

Theorem 4.6. *Let $\lambda \in \widehat{\mathfrak{h}}^*_{\mathrm{crit}}$ and $\tau : \widehat{S} \to \mathbb{K}$ be a deformation algebra with $\mathbb{K}$ being a field. Assume that the structure map τ factors through S.*

(1) *If λ is generic, $\overline{\Delta}_{\mathbb{K}}(\lambda + \tau)$ is simple.*

(2) *If λ is subgeneric with $R_{\mathbb{K}}(\lambda) = \{\pm\alpha\}$, we have a short exact sequence*

$$L_{\mathbb{K}}(\alpha \downarrow \lambda) \hookrightarrow \overline{\Delta}_{\mathbb{K}}(\lambda) \twoheadrightarrow L_{\mathbb{K}}(\lambda).$$

Note that the term subgeneric implies $\lambda \neq \alpha \downarrow \lambda$.

5. The restricted Jantzen sum formula

Andersen et al. [1994, Chapter 6] established a Jantzen sum formula for baby Verma modules. It relates the sum of the characters of the Jantzen filtration to an alternating sum of characters of baby Verma modules with smaller highest weights. We deduce a similar formula for the restricted Verma modules at the critical level.

Theorem 5.1. *Let $\lambda \in \widehat{\mathfrak{h}}^*_{\mathrm{crit}}$. There is a filtration*

$$\overline{\Delta}(\lambda) = \overline{\Delta}(\lambda)^0 \supset \overline{\Delta}(\lambda)^1 \supset \overline{\Delta}(\lambda)^2 \supset \cdots$$

with these properties:

(1) *$\overline{\Delta}(\lambda)^1$ is the maximal submodule of $\overline{\Delta}(\lambda)$.*

(2) $\displaystyle\sum_{i>0} \mathrm{ch}\, \overline{\Delta}(\lambda)^i = \sum_{\alpha \in R(\lambda)^+} \Bigl(\sum_{i>0} (\mathrm{ch}\, \overline{\Delta}(\alpha \downarrow^{2i-1} \lambda) - \mathrm{ch}\, \overline{\Delta}(\alpha \downarrow^{2i} \lambda)) \Bigr).$

Note that the sum is taken over all finite, positive, integral roots $\alpha \in R(\lambda)^+$.

The Shapovalov determinant. Let $\sigma : \widehat{\mathfrak{g}} \to \widehat{\mathfrak{g}}$ be a Chevalley-involution and define $\widehat{\mathfrak{n}}_+ := \bigoplus_{\alpha>0} \widehat{\mathfrak{g}}_\alpha$ and $\widehat{\mathfrak{n}}_- := \bigoplus_{\alpha<0} \widehat{\mathfrak{g}}_\alpha$, where $\widehat{\mathfrak{g}}_\alpha$ is the root space of $\widehat{\mathfrak{g}}$ corresponding to the root $\alpha \in \widehat{R}$. There is a decomposition $U(\widehat{\mathfrak{g}}) = U(\widehat{\mathfrak{h}}) \oplus (\widehat{\mathfrak{n}}_- U(\widehat{\mathfrak{g}}) + U(\widehat{\mathfrak{g}})\widehat{\mathfrak{n}}_+)$, and we denote by $\beta : U(\widehat{\mathfrak{g}}) \to S(\widehat{\mathfrak{h}})$ the projection to the first summand of this decomposition.

The Shapovalov form is now defined as the bilinear pairing $F: U(\widehat{\mathfrak{g}}) \times U(\widehat{\mathfrak{g}}) \to S(\widehat{\mathfrak{h}})$ with $F(x, y) = \beta(\sigma(x)y)$. It is symmetric and contravariant, i.e., for $u, x, y \in U(\widehat{\mathfrak{g}})$ we have $F(\sigma(u)x, y) = F(x, uy)$. For $\eta \in \mathbb{N}\widehat{R}^+$ we denote by F_η the restriction of F to the weight space $U(\widehat{\mathfrak{n}}_-)_{-\eta}$. Recall the isomorphism $\nu : \widehat{\mathfrak{h}} \xrightarrow{\sim} \widehat{\mathfrak{h}}^*$ induced by the bilinear form $(\cdot\,|\,\cdot)$ on $\widehat{\mathfrak{h}}$ and define $h_\alpha := \nu^{-1}(\alpha)$ for any root $\alpha \in \widehat{R}$.

Theorem 5.2 [Kac and Kazhdan 1979, Theorem 1]. *The determinant of*

$$F_\eta : U(\hat{\mathfrak{n}}_-)_{-\eta} \times U(\hat{\mathfrak{n}}_-)_{-\eta} \to S(\hat{\mathfrak{h}})$$

is, up to multiplication with a nonzero complex number, given by the formula

$$\det F_\eta = \prod_{\alpha \in \hat{R}^+} \prod_{n=1}^{\infty} \left(h_\alpha + \rho(h_\alpha) - n \frac{(\alpha \mid \alpha)}{2} \right)^{\mathrm{mult}(\alpha) \cdot \mathcal{P}(\eta - n\alpha)},$$

where $\mathcal{P}$ is Kostant's partition function and $\mathrm{mult}(\alpha) := \dim_{\mathbb{C}}(\hat{\mathfrak{g}}_\alpha)$.

We equip the polynomial ring $\mathbb{C}[t]$ in one variable with two different structures of a deformation algebra. The first one is given by the map $\tau_1 : \hat{S} \twoheadrightarrow \mathbb{C}[t]$, where τ_1 is induced by the inclusion of the line $\mathbb{C}\rho \subset \hat{\mathfrak{h}}^*$. The second $\hat{S}$-module structure $\tau_2 : \hat{S} \twoheadrightarrow \mathbb{C}[t]$ is given by the inclusion $\mathbb{C}\bar{\rho} \subset \hat{\mathfrak{h}}^*$. Recall that we consider elements of $\mathfrak{h}^*$ as elements of $\hat{\mathfrak{h}}^*$ by the embedding from above. Furthermore, $\bar{\rho} \in \mathfrak{h}^*$ implies that τ_2 factors through the restriction map $\hat{S} \twoheadrightarrow S$. For a more intuitive notation, we follow [Jantzen 1979] and define $\Delta_{\mathbb{C}[t]}(\lambda + t\rho) := \Delta_{\mathbb{C}[t]}(\lambda + \tau_1)$ and $\Delta_{\mathbb{C}[t]}(\lambda + t\bar{\rho}) := \Delta_{\mathbb{C}[t]}(\lambda + \tau_2)$.

Note that for $\lambda \in \hat{\mathfrak{h}}^*_{\mathrm{crit}}$ critical, and since $\tau_2(c) = \tau_2(d) = t\bar{\rho}(c) = 0$, we can construct the restricted Verma module $\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})$. Let $\mathbb{C}(t)$ be the quotient field of $\mathbb{C}[t]$.

Lemma 5.3. *Let $\lambda \in \hat{\mathfrak{h}}^*_{\mathrm{crit}}$. Then $\overline{\Delta}_{\mathbb{C}(t)}(\lambda + t\bar{\rho}) = \overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho}) \otimes_{\mathbb{C}[t]} \mathbb{C}(t)$ is simple.*

Proof. If we prove that $R_{\mathbb{C}(t)}(\lambda) = \varnothing$, the lemma follows from Theorem 4.6. But since $(\bar{\rho}|\alpha) \neq 0$ for all $\alpha \in R^+$, we get $2(\lambda + \rho + t\bar{\rho} \mid \alpha)_{\mathbb{C}(t)} \notin \mathbb{Z}(\alpha \mid \alpha)_{\mathbb{C}(t)} \subset \mathbb{C}$ for all $\alpha \in R^+$. □

The Shapovalov form induces symmetric, contravariant bilinear forms on $\Delta_{\hat{S}}(\lambda + \epsilon')$ and $\overline{\Delta}_S(\lambda + \epsilon)$, where $\epsilon' \in \hat{\mathfrak{h}}^*_{\hat{S}}$ denote the canonical weight induced by $\hat{\mathfrak{h}} \hookrightarrow \hat{S}$ and $\epsilon \in \hat{\mathfrak{h}}^*_S$ denotes its composition with $\hat{S} \twoheadrightarrow S$. Moreover, it induces contravariant forms on all Verma modules $\Delta_{\mathbb{C}[t]}(\lambda + t\rho)$, $\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})$, $\Delta_{\mathbb{C}[t]}(\lambda + t\bar{\rho})$, $\Delta(\lambda)$ and $\overline{\Delta}(\lambda)$, which we have to deal with in the rest of this paper. The contravariance of the forms implies for $\Delta(\lambda)$ and $\overline{\Delta}(\lambda)$ that the radicals of the forms coincide with the maximal submodules of $\Delta(\lambda)$ and $\overline{\Delta}(\lambda)$.

Restricted Jantzen filtration. Let $(\cdot\,,\cdot)$ be the contravariant form on $\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})$ induced by the Shapovalov form. We first define a filtration on $\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})$ by

$$\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})^i := \{ m \in \overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho}) \mid (m, \overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})) \subset t^i \mathbb{C}[t] \}.$$

The *Jantzen filtration* on $\overline{\Delta}(\lambda)$ is then defined by

$$\overline{\Delta}(\lambda)^i := \mathrm{im}(\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})^i \hookrightarrow \overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho}) \twoheadrightarrow \overline{\Delta}(\lambda)),$$

where the second map is the specialization $t \mapsto 0$. In the same way we get the Jantzen filtration on $\Delta(\lambda)$ as it is defined in [Kac and Kazhdan 1979] using the deformed Verma module $\Delta_{\mathbb{C}[t]}(\lambda + t\rho)$.

Notation 5.4. Let $\mu \leq \lambda$. We denote the determinants of the contravariant bilinear forms on the μ-weight spaces $\Delta_{\mathbb{C}[t]}(\lambda + t\bar{\rho})_\mu$, $\Delta_{\mathbb{C}[t]}(\lambda + t\rho)_\mu$ and $\overline{\Delta}_{\mathbb{C}[t]}(\lambda + t\bar{\rho})_\mu$ by $D_{\lambda+t\bar{\rho}}(\mu + t\bar{\rho})$, $D_{\lambda+t\rho}(\mu + t\rho)$ and $\overline{D}_{\lambda+t\bar{\rho}}(\mu + t\bar{\rho})$.

For a polynomial $P \in \mathbb{C}[t]$ denote by $\operatorname{ord}_t(P)$ the natural number $n \in \mathbb{N}$ with $t^n \,|\, P$ but $t^{n+1} \nmid P$.

Lemma 5.5. *For the Jantzen filtrations of the μ-weight spaces of the nonrestricted and restricted Verma modules we have the formulas*

$$\sum_{i>0} \dim_{\mathbb{C}} \Delta(\lambda)^i_\mu = \operatorname{ord}_t(D_{\lambda+t\rho}(\mu + t\rho))$$

and

$$\sum_{i>0} \dim_{\mathbb{C}} \overline{\Delta}(\lambda)^i_\mu = \operatorname{ord}_t(\overline{D}_{\lambda+t\bar{\rho}}(\mu + t\bar{\rho})).$$

Proof. By Lemma 5.3 the Shapovalov form of the restricted deformed Verma module $\overline{\Delta}_{\mathbb{C}[t]}(\lambda)$ is nondegenerate. Thus we can apply [Jantzen 1979, Lemma 5.1] to get both formulas. □

We first want to describe the Jantzen filtration of a restricted Verma module with a highest weight, which is critical and subgeneric.

Proposition 5.6 [Kübel 2014, Lemma 11]. *Let $\lambda \in \widehat{\mathfrak{h}}^*_{\text{crit}}$ be subgeneric; that is, $R(\lambda) = \{\pm\alpha\}$ for a finite positive root $\alpha \in R^+$ and $\alpha \downarrow \lambda \neq \lambda$. Then the Jantzen filtration of $\overline{\Delta}(\lambda)$ is*

$$\overline{\Delta}(\lambda) \supset L(\alpha \downarrow \lambda) \supset 0,$$

and we have the alternating sum formula

$$\sum_{i>0} \operatorname{ch} \overline{\Delta}(\lambda)^i = \operatorname{ch} L(\alpha \downarrow \lambda) = \operatorname{ch} \overline{\Delta}(\alpha \downarrow \lambda) - \operatorname{ch} \overline{\Delta}(\alpha \downarrow^2 \lambda) + \operatorname{ch} \overline{\Delta}(\alpha \downarrow^3 \lambda) - \cdots.$$

Proof. The first part of the proposition is [Kübel 2014, Lemma 11]. The second part follows inductively from Theorem 4.6. □

Recall the canonical weight $\epsilon : \widehat{\mathfrak{h}}_S \to S$ induced by $\widehat{\mathfrak{h}} \subset \widehat{S} \twoheadrightarrow S$, and for $\nu \geq 0$ denote by $\overline{D}_{\epsilon-\rho}(\epsilon - \rho - \nu)$ the determinant of the contravariant form on $\overline{\Delta}_S(\epsilon - \rho)_{-\rho-\nu}$. Let $\phi : S \twoheadrightarrow \mathbb{C}[t]$ be the algebra homomorphism given by

$$\phi(H) := (\lambda + \rho)(H) + t\bar{\rho}(H),$$

for all $H \in \mathfrak{h}$. If $p \in S$ is a prime element and $a \in S$, we denote by $\mathrm{ord}_p(a)$ the integer $n \in \mathbb{N}$ such that $p^n | a$ but $p^{n+1} \nmid a$. By [Jantzen 1979, Chapter 5.6], we get for $a \in S$

$$\mathrm{ord}_t(\phi(a)) = \sum_p \mathrm{ord}_p(a)\, \mathrm{ord}_t(\phi(p)), \tag{5-1}$$

where p runs over all classes of associated prime elements of S. As in Lemma 5.3 we see that, for the quotient field $Q = Q(S)$ of S, the restricted Verma module $\overline{\Delta}_Q(\epsilon-\rho) \cong \overline{\Delta}_S(\epsilon-\rho) \otimes_S Q$ is simple. We conclude that $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu) \neq 0$ and also $\phi(\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu)) \neq 0$ for all $\nu \in \mathbb{N}\widehat{R}^+$. Combining (5-1) with Lemma 5.5, we get

$$\sum_{n>0} \mathrm{ch}\, \overline{\Delta}(\lambda)^n = e^\lambda \sum_p \mathrm{ord}_t(\phi(p)) \sum_{\nu \in \mathbb{N}\widehat{R}^+} \mathrm{ord}_p(\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu)) e^{-\nu}. \tag{5-2}$$

We are now able to prove the general case. We follow [Jantzen 1979, Chapter 5.7].

Proof of Theorem 5.1. If $\lambda \in \widehat{\mathfrak{h}}^*_{\mathrm{crit}}$ fulfills $\langle \lambda+\rho, \alpha^\vee \rangle \notin \mathbb{Z}\backslash\{0\}$ for any finite positive root $\alpha \in R^+$, then λ is a generic weight and $\overline{\Delta}(\lambda)$ is simple, by Theorem 4.6. The evaluation of the polynomial $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu) \in S$ at $\lambda+\rho$ for $\nu \in \mathbb{N}\widehat{R}^+$ can be viewed as the determinant of the contravariant form on the weight space $\overline{\Delta}(\lambda)_{\lambda-\nu}$ induced by the Shapovalov form. Since the weight spaces are orthogonal to each other according to the contravariant form, $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu)(\lambda+\rho)$ is unequal to zero for all $\nu \in \mathbb{N}\widehat{R}^+$. Otherwise we could construct a proper submodule of $\overline{\Delta}(\lambda)$, which would be a contradiction. $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu)$ decomposes into a product of linear factors, and it follows that all prime divisors of $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\nu)$ are of the form $\alpha^\vee - r$, where $\alpha \in R^+$ and $r \in \mathbb{Z}\backslash\{0\}$.

For $\alpha \in R^+$ and $r \in \mathbb{Z}$ we define $\nu_{\alpha,r} \in \widehat{\mathbb{Z}[\mathfrak{h}^*]}$ by

$$\nu_{\alpha,r} = \sum_{\eta \in \mathbb{N}\widehat{R}^+} \mathrm{ord}_{\alpha^\vee - r}(\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\eta)) e^{-\eta}.$$

Because $\langle \bar{\rho}, \alpha^\vee \rangle \neq 0$ for any $\alpha \in R^+$, the restriction of $\alpha^\vee - r$ to the curve $(\lambda+\rho) + \mathbb{C}\bar{\rho} \subset \widehat{\mathfrak{h}}^*$ is unequal to zero, i.e., in formulas we have

$$\phi(\alpha^\vee - r) = \langle \lambda+\rho, \alpha^\vee \rangle - r + t\langle \bar{\rho}, \alpha^\vee \rangle \neq 0.$$

If $r \neq \langle \lambda+\rho, \alpha^\vee \rangle$, then $\mathrm{ord}_t(\langle \lambda+\rho, \alpha^\vee \rangle - r + t\langle \bar{\rho}, \alpha^\vee \rangle) = 0$. However, for $r = \langle \lambda+\rho, \alpha^\vee \rangle$ we have $\mathrm{ord}_t(\langle \lambda+\rho, \alpha^\vee \rangle - r + t\langle \bar{\rho}, \alpha^\vee \rangle) = 1$. Now $\alpha^\vee - \langle \lambda+\rho, \alpha^\vee \rangle$ can only be a prime divisor of $\overline{D}_{\epsilon-\rho}(\epsilon-\rho-\eta)$ if $\alpha \in R(\lambda)^+$. Applying formula (5-2) we conclude that

$$\sum_{i>0} \mathrm{ch}\, \overline{\Delta}(\lambda)^i = \sum_{\alpha \in R(\lambda)^+} \nu_{\alpha, \langle \lambda+\rho, \alpha^\vee \rangle} e^\lambda. \tag{5-3}$$

Let $\alpha \in R(\lambda)^+$. Perturbing the weight λ in the hyperplane that contains λ and is parallel to the reflection hyperplane corresponding to α, we find a weight $\mu \in \widehat{\mathfrak{h}}^*_{\text{crit}}$ such that $\langle \mu+\rho, \alpha^\vee\rangle = n = \langle \lambda+\rho, \alpha^\vee\rangle$ and $\langle \mu+\rho, \beta^\vee\rangle \notin \mathbb{Z}$ for all $\beta \in R(\lambda)^+ \setminus \{\alpha\}$. Thus, we've found a subgeneric weight μ with $R(\mu) = \{\pm\alpha\}$ and $\nu_{\alpha,\langle\mu+\rho,\alpha^\vee\rangle} = \nu_{\alpha,\langle\lambda+\rho,\alpha^\vee\rangle}$. But by Proposition 5.6, the Jantzen filtration of $\overline{\Delta}(\mu)$ is given by

$$\overline{\Delta}(\mu) \supset L(\alpha\downarrow\mu) \supset 0.$$

We conclude using (5-3):

$$\nu_{\alpha,n} e^{\mu} = \operatorname{ch} L(\alpha\downarrow\mu) = \sum_{i>0} \bigl(\operatorname{ch} \overline{\Delta}(\alpha\downarrow^{2i-1}\mu) - \operatorname{ch} \overline{\Delta}(\alpha\downarrow^{2i}\mu)\bigr).$$

Now the choice of μ implies that $e^{\lambda-\mu} \operatorname{ch} \overline{\Delta}(\alpha\downarrow^n\mu) = \operatorname{ch} \overline{\Delta}(\alpha\downarrow^n\lambda)$. Thus, we conclude

$$\nu_{\alpha,n} e^{\lambda} = \sum_{i>0} \bigl(\operatorname{ch} \overline{\Delta}(\alpha\downarrow^{2i-1}\lambda) - \operatorname{ch} \overline{\Delta}(\alpha\downarrow^{2i}\lambda)\bigr).$$

Since we can apply this to any root $\beta \in R(\lambda)^+$ we can use (5-3) once more to get the formula in Theorem 5.1. □

As a consequence of Theorem 5.1 we get the linkage principle for restricted Verma modules at the critical level in the same way as in [Andersen et al. 1994, Chapter 6] or [Kumar and Letzter 1997, Theorem 10.3]. The linkage principle was already proved in [Arakawa and Fiebig 2012a] introducing restricted projective objects in the restricted category $\mathcal{O}$ over the Lie algebra $\widehat{\mathfrak{g}}$. Our proof, however, avoids the rather complicated construction of restricted projective objects.

Corollary 5.7 [Arakawa and Fiebig 2012a, Theorem 5.1]. *Let $\lambda \in \widehat{\mathfrak{h}}^*_{\text{crit}}$ and $\mu \in \widehat{\mathfrak{h}}^*$. Then $[\overline{\Delta}(\lambda) : L(\mu)] \neq 0$ implies $\mu \in \widehat{\mathcal{W}}(\lambda)\cdot\lambda$ and $\mu \leq \lambda$.*

Proof. The statement is obvious for $\lambda = \mu$, and it is also clear that $[\overline{\Delta}(\lambda) : L(\mu)] \neq 0$ implies $\mu \leq \lambda$. We use induction on $\lambda - \mu$ and assume $\mu < \lambda$. If $[\overline{\Delta}(\lambda) : L(\mu)] \neq 0$, then also $[\overline{\Delta}(\lambda)^1 : L(\mu)] \neq 0$ since $\overline{\Delta}(\lambda)^1 \subset \overline{\Delta}(\lambda)$ is the maximal submodule. But then the restricted Jantzen sum formula implies that $L(\mu)$ has to appear as a subquotient in some $\overline{\Delta}(\alpha\downarrow^n\lambda)$, where $\alpha \in R(\lambda)^+$ and $n > 0$. Our induction assumption then implies $\mu \in \widehat{\mathcal{W}}(\alpha\downarrow^n\lambda)\cdot(\alpha\downarrow^n\lambda)$, but the definition of $\alpha\downarrow\lambda$ implies $\widehat{\mathcal{W}}(\alpha\downarrow^n\lambda)\cdot(\alpha\downarrow^n\lambda) = \widehat{\mathcal{W}}(\lambda)\cdot(\lambda)$. □

Acknowledgements

I would like to thank Peter Fiebig for drawing my attention to the calculation of the above Jantzen sum formula and I am very grateful for his patience in explaining me the underlying theory. I also want to thank the referee for his valuable suggestions and comments to improve this article.

References

[Andersen et al. 1994] H. H. Andersen, J. C. Jantzen, and W. Soergel, *Representations of quantum groups at a pth root of unity and of semisimple groups in characteristic p: independence of p*, Astérisque **220**, Société Mathématique de France, Paris, 1994. MR 95j:20036 Zbl 0802.17009

[Arakawa and Fiebig 2012a] T. Arakawa and P. Fiebig, "The linkage principle for restricted critical level representations of affine Kac–Moody algebras", *Compos. Math.* **148**:6 (2012), 1787–1810. MR 2999304 Zbl 06147343

[Arakawa and Fiebig 2012b] T. Arakawa and P. Fiebig, "On the restricted Verma modules at the critical level", *Trans. Amer. Math. Soc.* **364**:9 (2012), 4683–4712. MR 2922606 Zbl 06191426

[Feigin and Frenkel 1992] B. Feigin and E. Frenkel, "Affine Kac–Moody algebras at the critical level and Gelfand–Dikii algebras", pp. 197–215 in *Infinite analysis, Part A* (Kyoto, 1991), edited by A. Tsuchiya et al., Adv. Ser. Math. Phys. **16**, World Scientific, River Edge, NJ, 1992. MR 93j:17049 Zbl 0925.17022

[Frenkel 2005] E. Frenkel, "Wakimoto modules, opers and the center at the critical level", *Adv. Math.* **195**:2 (2005), 297–404. MR 2006d:17018 Zbl 1129.17014

[Frenkel and Gaitsgory 2006] E. Frenkel and D. Gaitsgory, "Local geometric Langlands correspondence and affine Kac–Moody algebras", pp. 69–260 in *Algebraic geometry and number theory*, edited by V. Ginzburg, Progr. Math. **253**, Birkhäuser, Boston, MA, 2006. MR 2008e:17023 Zbl 1184.17011

[Jantzen 1979] J. C. Jantzen, *Moduln mit einem höchsten Gewicht*, Lecture Notes in Mathematics **750**, Springer, Berlin, 1979. MR 81m:17011 Zbl 0426.17001

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, 1990. MR 92k:17038 Zbl 0716.17022

[Kac and Kazhdan 1979] V. G. Kac and D. A. Kazhdan, "Structure of representations with highest weight of infinite-dimensional Lie algebras", *Adv. in Math.* **34**:1 (1979), 97–108. MR 81d:17004 Zbl 0427.17011

[Kübel 2014] J. Kübel, "Centers for the restricted category $\mathcal{O}$ at the critical level over affine Kac–Moody algebras", *Math. Z.* **276**:3-4 (2014), 1133–1149. MR 3175174

[Kumar and Letzter 1997] S. Kumar and G. Letzter, "Shapovalov determinant for restricted and quantized restricted enveloping algebras", *Pacific J. Math.* **179**:1 (1997), 123–161. MR 98g:17014 Zbl 0965.17006

[Rocha-Caridi and Wallach 1982] A. Rocha-Caridi and N. R. Wallach, "Projective modules over graded Lie algebras, I", *Math. Z.* **180**:2 (1982), 151–177. MR 83h:17018 Zbl 0467.17006

Received January 28, 2013. Revised July 15, 2013.

JOHANNES KÜBEL
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ERLANGEN
CAUERSTR. 11
91058 ERLANGEN
GERMANY

kuebel@mi.uni-erlangen.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.385

NOTES ON THE EXTENSION OF THE MEAN CURVATURE FLOW

YAN LENG, ENTAO ZHAO AND HAORAN ZHAO

In this paper, we present several new curvature conditions that assure the extension of the mean curvature flow on a finite time interval, which improve some known extension theorems.

1. Introduction

Let $F_0 : M^n \to N^{n+d}$ be a smooth isometric immersion from an n-dimensional closed (compact and without boundary) Riemannian manifold M to an $(n+d)$-dimensional Riemannian manifold N. Consider a one-parameter family of smooth isometric immersions $F : M \times [0, T) \to N$ satisfying

$$\begin{cases} \dfrac{\partial}{\partial t} F(x,t) = H(x,t), \\ F(x,0) = F_0(x), \end{cases} \tag{1-1}$$

where $H(x,t)$ is the mean curvature vector of $F_t(M)$ and $F_t(x) = F(x,t)$. Set $M_t = F_t(M)$. We call $F : M \times [0, T) \to N$ the mean curvature flow with initial value $F_0 : M \to N$.

The mean curvature flow is a (degenerate) quasilinear parabolic evolution equation, and one can obtain the short-time existence either by the Nash–Moser implicit function theorem or by the DeTurck trick to modify the mean curvature flow equation to a strongly parabolic equation. Without any special assumption on M_0, the mean curvature flow (1-1) will in general develop singularities in finite time, characterized by blowing up of the second fundamental form A.

Theorem 1.1 [Huisken 1984; 1986; Wang 2001]. *Suppose $T < \infty$ is the first singular time for a closed mean curvature flow in a Riemannian manifold with bounded geometry. Then we have*

$$\lim_{t\to T} \sup_{M_t} |A| = \infty.$$

Research supported by the National Natural Science Foundation of China, grants 11371315 and 11201416.

MSC2010: 53C44.

Keywords: Mean curvature flow, extension, blow-up method, volume expansion.

From Theorem 1.1, we see that if $\sup_{M_t\times[0,T)}|A|$ is bounded, then the mean curvature flow can be extended past the time T. Recently, Le and Šešum [2011] and Liu, Xu, Ye and Zhao [Liu et al. 2011; Xu et al. 2011a; 2011b] obtained some integral conditions to extend the mean curvature flow. Define a $(0,2)$-tensor B on M in a local orthonormal frame field by $B_{ij}=\langle H,h_{ij}\rangle$. Cooper obtained the following characterization of the singular time.

Theorem 1.2 [Cooper 2011]. *Suppose $T<\infty$ is the first singular time for a closed mean curvature flow in a Riemannian manifold with bounded geometry. Then we have*

$$\lim_{t\to T}\sup_{M_t}|B|=\infty.$$

Similarly, we see from Theorem 1.2 that if $\sup_{M_t\times[0,T)}|B|$ is bounded, then the mean curvature flow can be extended past the time T.

In the present paper, we make an improvement of Theorems 1.1 and 1.2 by considering the integral of $|B|$ on the time interval. More precisely, we prove the following theorem.

Theorem 1.3. *Let $F_t:M^n\to N^{n+d}$ be the mean curvature flow solution of closed submanifolds on a finite time interval $[0,T)$ and assume N has bounded geometry. If the function $f(x):=\int_0^T|B|(x,t)\,dt$ is continuous on M, then the mean curvature flow can be extended past the time T.*

By the dominated convergence theorem and Theorem 1.3, we obtain the following result, which recovers Theorems 1.1 and 1.2.

Theorem 1.4. *Suppose $T<\infty$ is the first singular time for a closed mean curvature flow in a Riemannian manifold with bounded geometry. Then we have*

$$\int_0^T\sup_{M_t}|B|(t)\,dt=\infty.$$

Analogous extension theorems for the Ricci flow have been proved recently [Wang 2012; He 2014]. Some general regularity results have been obtained by Cheeger, Haslhofer and Naber [Cheeger et al. 2013] and Ecker [2013], among others. To prove our theorems we combine the ideas in [Cooper 2011] and [He 2014]. First, by a suitable blow-up argument we get a minimal submanifold in Euclidean space. Second, we prove that the volume of geodesic balls in this minimal submanifold is less than the volume of geodesic balls with same radius. By the expansion for the volume of geodesic balls, we see that this minimal submanifold is in fact totally geodesic at the base point.

2. Preliminaries

Let M^n be an n-dimensional submanifold isometrically immersed in an $(n+d)$-dimensional Riemannian manifold N^{n+d}. Let A and H be the second fundamental form and the mean curvature vector of M in N, respectively. Define a $(0,2)$-tensor B from A and H by $B = \langle A, H\rangle$. Choose a local orthonormal frame field $\{e_A\}_{A=1}^{n+d}$ in N^{n+d} such that each e_i, $i = 1, \dots, n$, is tangent to M and let $\{\omega_A\}_{A=1}^{n+d}$ be the dual frame field of $\{e_A\}_{A=1}^{n+d}$. Then A, H and B can be written as

$$A = \sum_{i,j=1}^{n} \sum_{\alpha=n+1}^{n+d} h_{ij}^{\alpha}\, \omega_i \otimes \omega_j \otimes e_\alpha = \sum_{i,j} h_{ij}\, \omega_i \otimes \omega_j,\ h_{ij} = \sum_{\alpha=n+1}^{n+d} h_{ij}^{\alpha}\, e_\alpha,$$

$$H = \sum_{\alpha} H^{\alpha} e_\alpha,\ H^{\alpha} = \sum_{i=1}^{n} h_{ii}^{\alpha},$$

$$B = \sum_{i,j=1}^{n} B_{ij}\, \omega_i \otimes \omega_j,\ B_{ij} = \sum_{\alpha=n+1}^{n+d} H^{\alpha} h_{ij}^{\alpha}.$$

Let $F : M \times [0, T) \to N$ be a mean curvature flow solution with initial immersion $F_0 : M \to N$. Denote by $g(t)$ and $d\mu(t)$ the induced metric and the volume form on M. Under the mean curvature flow, $g(t)$ and $d\mu(t)$ satisfy the following evolution equations.

$$\frac{\partial}{\partial t} g(t) = -2B(t), \tag{2-1}$$

$$\frac{\partial}{\partial t} d\mu(t) = -|H|^2 d\mu(t). \tag{2-2}$$

3. Proof of Theorem 1.3

Now we give the proof of Theorem 1.3.

Proof. We argue by contradiction. Suppose that T is the maximal existence time. Then by Theorem 1.1 we see that $\lim_{t\to T} \sup_{M_t} |A| = \infty$. Choose a sequence of points $(O_i, t_i) \in M \times [0, T)$, $i = 1, 2, \dots$, such that $\lim_{i\to\infty} t_i = T$ and

$$|A|^2(O_i, t_i) = \max_{(x,t)\in M\times[0,t_i]} |A|^2(x, t) \to \infty \quad \text{as } i \to \infty.$$

Set $Q_i = |A|^2(O_i, t_i)$ and we suppose $Q_i \geq 1$ and $Q_i t_i \geq 1$. Denote by h the Riemannian metric on N. We consider the rescaled flows for $t \in [0, 1]$

$$F_i(t) = F\left(\frac{t-1}{Q_i} + t_i\right) : (M, g_i(t)) \longrightarrow (N, Q_i h),$$

where $g_i(t) = F_i(t)^*(Q_i h)$ is the induced metric on M. Then for each i, F_i is also a solution of the mean curvature flow on time interval [0, 1]. Denote by M_i the manifold M with metric $g_i(t)$. It follows from [Chen and He 2010] that there is a subsequence of $\{(M_i, g_i(t), O_i) : i = 1, 2, \dots\}$ which converges to a Riemannian manifold $(M_\infty, g_\infty(t), O_\infty)$, and the corresponding subsequence of immersions $F_i(t)$ converges to an immersion $F_\infty(t) : M_\infty \to \mathbb{R}^{n+d}$, $t \in [0, 1]$. Note that F_∞ is also a solution of the mean curvature flow on time interval [0, 1].

We first show that for any $t \in [0, 1]$, M_∞ is a minimal submanifold in $\mathbb{R}^{n+d}$. Let $B_\infty(\cdot\,, t)$ be the (0, 2)-tensor for $F_\infty(t)$. In fact, we prove the following:

Lemma 3.1. $B_\infty(t) = 0$ *for* $t \in [0, 1]$.

Proof. By the continuity assumption on

$$f(x) := \int_0^T |B|(x, t)\, dt$$

and the compactness of M, we can use elementary arguments to prove that

$$\lim_{t \to T} \int_t^T |B|(x, t)\, dt = 0.$$

First, we have

$$\begin{aligned} g_i(t) = F_i(t)^*(Q_i h) &= F\left(\frac{t-1}{Q_i} + t_i\right)^*(Q_i h) \\ &= Q_i F\left(\frac{t-1}{Q_i} + t_i\right)^*(h) = Q_i g\left(\frac{t-1}{Q_i} + t_i\right). \end{aligned}$$

Denote by $A_i(\cdot\,, t)$, $H_i(\cdot\,, t)$ and $B_i(\cdot\,, t)$ the second fundamental form, the mean curvature and the (0, 2)-tensor of $F_i(t)$, respectively. It is easy to see from the definition of second fundamental form that

$$A_i(\cdot\,, t) = A\left(\cdot\,, \frac{t-1}{Q_i} + t_i\right).$$

Since the mean curvature is the trace of the second fundamental form, we have

$$H_i(\cdot\,, t) = Q_i^{-1} H\left(\cdot\,, \frac{t-1}{Q_i} + t_i\right).$$

So for the (0, 2)-tensor we have

$$\begin{aligned} B_i(\cdot\,, t) &= \langle A_i(\cdot\,, t), H_i(\cdot\,, t)\rangle_{Q_i h} \\ &= \left\langle A\left(\cdot\,, \frac{t-1}{Q_i} + t_i\right), H\left(\cdot\,, \frac{t-1}{Q_i} + t_i\right)\right\rangle_h = B\left(\cdot\,, \frac{t-1}{Q_i} + t_i\right). \end{aligned}$$

From this we see that

$$
\begin{aligned}
|B_i(\,\cdot\,,t)|^2_{g_i(t)} &= \langle B_i(\,\cdot\,,t), B_i(\,\cdot\,,t)\rangle_{g_i(t)\otimes g_i(t)} \\
&= Q_i^{-2}\left\langle B\left(\,\cdot\,,\frac{t-1}{Q_i}+t_i\right), B\left(\,\cdot\,,\frac{t-1}{Q_i}+t_i\right)\right\rangle_{g((t-1)/Q_i+t_i)\otimes g((t-1)/Q_i+t_i)} \\
&= Q_i^{-2}\left|B\left(\,\cdot\,,\frac{t-1}{Q_i}+t_i\right)\right|^2_{g((t-1)/Q_i+t_i)}.
\end{aligned}
$$

For any $y \in M_\infty$, there are $y_i \in M$, $i = 1, 2, \dots$, such that $\lim_{i\to\infty} y_i = y$.

$$
\begin{aligned}
\int_0^1 |B|_{g_\infty(t)}(y,t)\,dt &= \lim_{i\to\infty}\int_0^1 |B_i|_{g_i(t)}(y_i,t)\,dt \\
&= \lim_{i\to\infty} Q_i^{-1}\int_0^1 |B|_{g((t-1)/Q_i+t_i)}\left(y_i, \frac{t-1}{Q_i}+t_i\right)dt \\
&= \lim_{i\to\infty}\int_{t_i-Q_i^{-1}}^{t_i} |B|_{g(s)}(y_i,s)\,ds \\
&= 0.
\end{aligned}
$$

Hence we have $B_\infty(t) = 0$ for each $t \in [0, 1]$. □

Lemma 3.2. *The induced metrics $g(t)$ on M are uniformly equivalent and converge pointwise as $t \to T$ to a continuous positive-definite metric $g(T)$.*

Proof. Under the assumption that $f(x) = \int_0^T |B|(x,t)\,dt$ is continuous, we see that $f(x)$ is bounded and for any $0 \le \tau \le \theta < T$

$$
\lim_{\tau\to\theta}\int_\tau^\theta |B|(x,t)\,dt = 0
$$

uniformly. Since $g(t)$ satisfies (2-1), we can carry out the same argument as in [Hamilton 1982] to prove the lemma. □

Let $\boldsymbol{B}_{g_\infty(1)}(O_\infty, r)$ be the geodesic ball of radius r centered at $O_\infty \in M_\infty$ with respect to the metric $g_\infty(1)$, and $\mathrm{Vol}_{g_\infty(1)}(\boldsymbol{B}_{g_\infty(1)}(O_\infty, r))$ be the volume of $\boldsymbol{B}_{g_\infty(1)}(O_\infty, r)$. Denote by ω_n the volume of the unit ball in $\mathbb{R}^n$.

Lemma 3.3. $$\mathrm{Vol}_{g_\infty(1)}(\boldsymbol{B}_{g_\infty(1)}(O_\infty, r)) \le \omega_n r^n.$$

Proof. Let $\boldsymbol{B}_{g_i(1)}(O_i, r)$ be the geodesic ball with radius r centered at $O_i \in M_i$ with respect to $g_i(t)$ and $\mathrm{Vol}_{g_i(1)}(\boldsymbol{B}_{g_i(1)}(O_i, r))$ the volume of $\boldsymbol{B}_{g_i(1)}(O_i, r)$. It is easy to see that

$$
\boldsymbol{B}_{g_i(1)}(O_i, r) = \boldsymbol{B}_{g(t_i)}(O_i, Q_i^{-1/2} r).
$$

Hence

$$\begin{aligned}\frac{\mathrm{Vol}_{g_\infty(1)}(\boldsymbol{B}_{g_\infty(1)}(O_\infty, r))}{r^n} &= \lim_{i\to\infty}\frac{\mathrm{Vol}_{g_i(1)}(\boldsymbol{B}_{g_i(1)}(O_i, r))}{r^n}\\ &= \lim_{i\to\infty}\frac{\mathrm{Vol}_{g_i(1)}(\boldsymbol{B}_{g(t_i)}(O_i, Q_i^{-1/2}r))}{r^n}\\ &= \lim_{i\to\infty}\frac{Q_i^{n/2}\mathrm{Vol}_{g(t_i)}(\boldsymbol{B}_{g(t_i)}(O_i, Q_i^{-1/2}r))}{r^n}\\ &= \lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_i)}(\boldsymbol{B}_{g(t_i)}(O_i, Q_i^{-1/2}r))}{(Q_i^{-1/2}r)^n}.\end{aligned}$$

From Lemma 3.2, we see that for any $\varepsilon > 0$, there is a positive constant δ such that if $t \geq t_0 > T-\delta$, then $(1-\varepsilon)g(t_0) \leq g(t) \leq (1+\varepsilon)g(t_0)$. We may pick t_is such that $t_i \geq t_0 \geq T-\delta$. From a lemma in [Cooper 2011; Glickenstein 2003] we see that

$$\begin{aligned}&\lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_i)}(\boldsymbol{B}_{g(t_i)}(O_i, Q_i^{-1/2}r))}{(Q_i^{-1/2}r)^n}\\ &\qquad\leq \lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_i)}(\boldsymbol{B}_{g(t_0)}(O_i, ((1-\varepsilon)Q_i)^{-1/2}r))}{(Q_i^{-1/2}r)^n}\\ &\qquad\leq \lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_0)}(\boldsymbol{B}_{g(t_0)}(O_i, ((1-\varepsilon)Q_i)^{-1/2}r))}{(Q_i^{-1/2}r)^n}\\ &\qquad= (1-\varepsilon)^{-n/2}\lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_0)}(\boldsymbol{B}_{g(t_0)}(O_i, ((1-\varepsilon)Q_i)^{-1/2}r))}{(((1-\varepsilon)Q_i)^{-1/2}r)^n}.\end{aligned}$$

Since $Q_i \to \infty$ as $i \to \infty$, we have

$$\lim_{i\to\infty}\frac{\mathrm{Vol}_{g(t_0)}(\boldsymbol{B}_{g(t_0)}(O_i, ((1-\varepsilon)Q_i)^{-1/2}r))}{(((1-\varepsilon)Q_i)^{-1/2}r)^n} = \omega_n.$$

Since ε is arbitrary, we see that

$$\frac{\mathrm{Vol}_{g_\infty(1)}(\boldsymbol{B}_{g_\infty(1)}(O_\infty, r))}{r^n} \leq \omega_n. \qquad \square$$

We continue the proof of Theorem 1.3. From the expansion formula for the volume of small balls (see Theorem 3.98 of [Gallot et al. 1987]) we have

$$\frac{\mathrm{Vol}_{g_\infty(1)}(\boldsymbol{B}_{g_\infty(1)}(O_\infty, r))}{\omega_n r^n} = 1 - \frac{R(O_\infty)}{6(n+2)}r^2 + o(r^2).$$

Here $R(O_\infty)$ is the scalar curvature at $O_\infty \in (M_\infty, g_\infty(1))$. From Lemma 3.3 we see that

$$R(O_\infty) \geq 0.$$

This combined with Lemma 3.1 implies that

$$|A|_\infty(O_\infty, 1) = 0.$$

However, it is seen from the point selecting process that

$$|A|_\infty(O_\infty, 1) = 1.$$

This is a contradiction, which completes the proof of Theorem 1.3. □

Theorem 3.4. *Let $F_t : M^n \to N^{n+d}$ be the mean curvature flow solution of closed submanifolds on a finite time interval $[0, T)$ and assume N has bounded geometry. Suppose $T < \infty$ is the first singular time. If the function $\int_0^T |A|(x, t)\, dt < +\infty$ is continuous on M, then we have*

$$\lim_{t\to T} \sup_{M_t} |H| = \infty.$$

Proof. We suppose $|H| \le C$ uniformly for all the existence time. Then

$$\left|\frac{\partial g}{\partial t}\right| = 2|B| \le 2|H||A| \le 2C|A|.$$

By the dominated convergence theorem, we know that $\int_0^T |A|(x, t)\, dt$ is continuous in x. Then by a similar argument as in the proof of Theorem 1.3, we get the conclusion. □

Remark 3.5. Theorem 3.4 recovers [Cooper 2011, Theorem 5.1].

Acknowledgement

The authors thank Professor Kefeng Liu and Professor Hongwei Xu for their help and discussions.

References

[Cheeger et al. 2013] J. Cheeger, R. Haslhofer, and A. Naber, “Quantitative stratification and the regularity of mean curvature flow”, *Geom. Funct. Anal.* **23**:3 (2013), 828–847. MR 3061773 Zbl 1277.53064

[Chen and He 2010] J. Chen and W. He, “A note on singular time of mean curvature flow”, *Math. Z.* **266**:4 (2010), 921–931. MR 2011j:53119 Zbl 1201.53075

[Cooper 2011] A. A. Cooper, “A characterization of the singular time of the mean curvature flow”, *Proc. Amer. Math. Soc.* **139**:8 (2011), 2933–2942. MR 2012d:53211 Zbl 1220.53080

[Ecker 2013] K. Ecker, “Partial regularity at the first singular time for hypersurfaces evolving by mean curvature”, *Math. Ann.* **356**:1 (2013), 217–240. MR 3038128 Zbl 1270.53084

[Gallot et al. 1987] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, Springer, Berlin, 1987. MR 88k:53001 Zbl 0636.53001

[Glickenstein 2003] D. Glickenstein, “Precompactness of solutions to the Ricci flow in the absence of injectivity radius estimates”, *Geom. Topol.* **7** (2003), 487–510. MR 2004k:53099 Zbl 1044.53048

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR 84a:53050 Zbl 0504.53034

[He 2014] F. He, "Remarks on the extension of the Ricci flow", *J. Geom. Anal.* **24**:1 (2014), 81–91. MR 3145915

[Huisken 1984] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR 86j:53097 Zbl 0556.53001

[Huisken 1986] G. Huisken, "Contracting convex hypersurfaces in Riemannian manifolds by their mean curvature", *Invent. Math.* **84**:3 (1986), 463–480. MR 87f:53066 Zbl 0589.53058

[Le and Šešum 2011] N. Q. Le and N. Sesum, "On the extension of the mean curvature flow", *Math. Z.* **267**:3-4 (2011), 583–604. MR 2012h:53153 Zbl 1216.53060

[Liu et al. 2011] K. F. Liu, H.-W. Xu, F. Ye, and E.-T. Zhao, "The extension and convergence of mean curvature flow in higher codimension", preprint, 2011. arXiv 1104.0971v1

[Wang 2001] M.-T. Wang, "Mean curvature flow of surfaces in Einstein four-manifolds", *J. Differential Geom.* **57**:2 (2001), 301–338. MR 2003j:53108 Zbl 1035.53094

[Wang 2012] B. Wang, "On the conditions to extend Ricci flow(II)", *Int. Math. Res. Not.* **2012**:14 (2012), 3192–3223. MR 2946223 Zbl 1251.53040

[Xu et al. 2011a] H.-W. Xu, F. Ye, and E.-T. Zhao, "Extend mean curvature flow with finite integral curvature", *Asian J. Math.* **15**:4 (2011), 549–556. MR 2012h:53158 Zbl 1242.53085

[Xu et al. 2011b] H.-W. Xu, F. Ye, and E.-T. Zhao, "The extension for mean curvature flow with finite integral curvature in Riemannian manifolds", *Sci. China Math.* **54**:10 (2011), 2195–2204. MR 2012k:53133 Zbl 1232.53056

Received March 25, 2013. Revised September 11, 2013.

YAN LENG
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU, 310027
CHINA
lengyan@cms.zju.edu.cn

ENTAO ZHAO
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU, 310027
CHINA
zhaoet@cms.zju.edu.cn

HAORAN ZHAO
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU, 310027
CHINA
zhaohaoran2010@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.393

HYPERSURFACES WITH PRESCRIBED ANGLE FUNCTION

HENRIQUE F. DE LIMA, ERALDO A. LIMA JR. AND ULISSES L. PARENTE

We deal with two-sided complete hypersurfaces immersed in a Riemannian product space, whose base is assumed to have sectional curvature bounded from below. In this setting, we obtain sufficient conditions which assure that such a hypersurface is a slice of the ambient space, provided that its angle function has some suitable behavior. Furthermore, we establish a natural relation between our results and the classical problem of describing the geometry of a hypersurface immersed in the Euclidean space through the behavior of its Gauss map.

1. Introduction and statements of the main results

Let $\psi : \Sigma^n \to \mathbb{M}^{n+1}$ be an immersion of an orientable Riemannian manifold Σ^n in a Riemannian space form $\mathbb{M}^{n+1}$ and let N be the unit normal vector field along Σ^n. When $\mathbb{M}^{n+1}$ is the Euclidean space $\mathbb{R}^{n+1}$ and ψ is the complete graph of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, the image $N(\Sigma)$ of its Gauss map is contained in an open hemisphere of the unit Euclidean sphere $\mathbb{S}^n$. The behavior of the Gauss map has deeper consequences for the immersion. For instance, one of the most celebrated theorems of the theory of minimal surfaces in $\mathbb{R}^3$ is Bernstein's theorem [1910], which establishes that the only complete minimal graphs in $\mathbb{R}^3$ are planes. This result was extended under the weaker hypothesis that the image of the Gauss map of Σ^2 lies in an open hemisphere of $\mathbb{S}^2$, as we can see in [Barbosa and do Carmo 1974].

Meanwhile, Osserman [1959] answered a conjecture due to Nirenberg, showing that if a complete minimal surface Σ^2 in $\mathbb{R}^3$ is not a plane, then its normals must be everywhere dense on the unit sphere $\mathbb{S}^2$. More generally, Fujimoto [1988] proved that if the Gaussian image misses more than four points, then it is a plane. On the other hand, Hoffman, Osserman and Schoen [Hoffman et al. 1982] showed that if a complete oriented surface Σ^2 with constant mean curvature in $\mathbb{R}^3$ is such that the image of its Gauss map $N(\Sigma)$ lies in some open hemisphere of $\mathbb{S}^2$, then Σ^2 is a

---

The first author is partially supported by CNPq, Brazil. The second author is partially supported by CAPES, Brazil.
MSC2010: primary 53C42; secondary 53C50, 53B30.
Keywords: Riemannian product spaces, two-sided complete hypersurfaces, mean curvature, height function, entire vertical graphs.

plane. Moreover, if $N(\Sigma)$ lies in a closed hemisphere, then Σ^2 is a plane or a right circular cylinder.

When the ambient space is a Riemannian product $\overline{M}^{n+1} = \mathbb{R} \times M^n$, the condition that the image of the Gauss map is contained in a closed hemisphere becomes that the angle function $\eta = \langle N, \partial_t \rangle$ does not change sign, as was observed in [Espinar and Rosenberg 2009]. Here, N denotes a unit normal vector field along a hypersurface $\psi : \Sigma^n \to \overline{M}^{n+1}$ and ∂_t stands for the unit vector field which determines on $\overline{M}^{n+1}$ a codimension-one foliation by totally geodesic slices $\{t\} \times M^n$. In this setting, our purpose in this work is to establish analogous results to those ones above described. In other words, we aim to give new satisfactory answers to the following question: *under what reasonable geometric restrictions on the angle function must a complete hypersurface immersed in a certain product space be a slice*?

We can truly say that one of the first remarkable results in this direction was the celebrated theorem of Bombieri, De Giorgi and Miranda [Bombieri et al. 1969], who proved that an entire minimal positive graph over $\mathbb{R}^n$ is a totally geodesic slice. Many other authors have approached problems in this branch. For instance, Rosenberg [2002] showed that when M^2 is a complete surface with nonnegative Gaussian curvature, an entire minimal graph in $\mathbb{R} \times M^2$ is totally geodesic. Hence, in this case, the graph is a horizontal slice or M^2 is a flat $\mathbb{R}^2$ and the graph is a tilted plane. Bérard and Sá Earp [2008] described all rotation hypersurfaces with constant mean curvature in $\mathbb{R} \times \mathbb{H}^n$, and used them as barriers to prove existence and characterization of certain vertical graphs with constant mean curvature and to give symmetry and uniqueness results for constant mean curvature compact hypersurfaces whose boundary is one or two parallel submanifolds in slices. Espinar and Rosenberg [2009] studied constant mean curvature surfaces in $\mathbb{R} \times M^2$, and classified them according to the infimum of the Gaussian curvature of their horizontal projection, under the assumption that the angle function does not change sign.

In [Aquino and Lima 2011] and [Lima and Parente 2012], we applied the well-known generalized maximum principle of Omori [1967] and Yau [1975], and an extension of it due to Akutagawa [1987], in order to obtain rigidity theorems concerning complete vertical graphs with constant mean curvature in $\mathbb{R} \times \mathbb{M}^n$. In [Lima 2014], the first author extended the technique developed in [Yau 1976] in order to investigate the rigidity of entire vertical graphs in a Riemannian product space $\mathbb{R} \times M^n$, whose base M^n is assumed to have Ricci curvature with strict sign. Under a suitable restriction on the norm of the gradient of the function u which determines such a graph $\Sigma^n(u)$, he proved that $\Sigma^n(u)$ must be a slice $\{t\} \times M^n$.

Now, motivated by the previous discussion, we will state our results. In what follows, $H_2 = 2/(n(n-1))S_2$ stands for the mean value of the second elementary symmetric function S_2 on the eigenvalues of the Weingarten operator A of the hypersurface Σ^n. Moreover, we recall that a hypersurface is said to be *two-sided* if its normal bundle is trivial, that is, if there is a globally defined unit normal vector field on it.

Theorem 1. *Let $\overline{M}^{n+1} = \mathbb{R} \times M^n$ be a Riemannian product space whose base M^n has sectional curvature K_M such that $K_M \geq -\kappa$ for some $\kappa > 0$, and let $\psi : \Sigma^n \to \overline{M}^{n+1}$ be a two-sided complete hypersurface with constant mean curvature H and H_2 bounded from below. Suppose that the angle function η of Σ^n is bounded away from zero and that its height function h satisfies one of the following conditions*:

$$|\nabla h|^2 \leq \frac{\alpha}{(n-1)\kappa}|A|^2 \tag{1-1}$$

for some constant $0 < \alpha < 1$; *or*

$$|\nabla h|^2 \leq \frac{n}{(n-1)\kappa}H^2. \tag{1-2}$$

Then, *Σ^n is a slice of $\overline{M}^{n+1}$.*

As a consequence of Example 10 given in Section 4, we cannot extend estimate (1-1) to the limit case $\alpha = 1$. On the other hand, taking into account estimate (1-2), when $M^n = \mathbb{R}^n$, we note that Theorem 1 reads as follows:

Corollary 2. *Let Σ^n be a two-sided complete hypersurface of $\mathbb{R}^{n+1}$ with constant mean curvature and scalar curvature bounded from below. If the closure of the image of the Gauss map of Σ^n is contained in an open hemisphere of $\mathbb{S}^n$, then Σ^n is minimal.*

Proceeding, we treat the case where the mean curvature H is not assumed to be constant, but is just assumed to not change sign along the hypersurface:

Theorem 3. *Let $\overline{M}^{n+1} = \mathbb{R} \times M^n$ be a Riemannian product space whose base M^n has sectional curvature bounded from below*, *and let $\psi : \Sigma^n \to \overline{M}^{n+1}$ be a two-sided complete hypersurface that lies between two slices of $\overline{M}^{n+1}$. Suppose the angle function η of Σ^n is bounded away from 1 or from -1. If H_2 is bounded from below and H is bounded and does not change sign on Σ^n*, *then* $\inf_\Sigma H = 0$. *In particular*, *if H is constant*, *then Σ^n is minimal.*

Thanks to the result of Osserman already cited, Theorem 3 yields:

Corollary 4. *The only two-sided complete constant mean curvature surfaces of $\mathbb{R}^3$ with Gaussian curvature bounded from below*, *lying between two planes and such that both poles of $\mathbb{S}^2$ are not in the closure of the image of the Gauss map*, *that are orthogonal to such planes*, *are planes of $\mathbb{R}^3$.*

On the other hand, Example 10 will show that the assumption that Σ^n lies between two slices of $\mathbb{R} \times M^n$ is necessary in Theorem 3 in order to conclude that the mean curvature of Σ^n cannot be globally bounded away from zero. Moreover, we observe that the horizontal circular cylinder $\mathscr{C} \subset \mathbb{R}^3$ satisfies almost all the hypothesis of Corollary 4, except the one which requires that neither pole of $\mathbb{S}^2$ orthogonal to $\mathscr{C}$ is in the closure of the image of the Gauss map N of $\mathscr{C}$. Actually, $\mathscr{C}$ is unbounded in all directions where N is isolated.

Rosenberg, Schulze and Spruck [Rosenberg et al. 2013] showed that an entire minimal graph with nonnegative height function in a product space $\mathbb{R} \times M^n$, whose base M^n is a complete Riemannian manifold having nonnegative Ricci curvature and with sectional curvature bounded from below, must be a slice. Consequently, Theorem 3 yields:

Corollary 5. *Let M^n be a complete Riemannian manifold with nonnegative Ricci curvature and whose sectional curvature is bounded from below. Let $\Sigma^n(u) = \{(u(x), x) : x \in M^n\} \subset \mathbb{R} \times M^n$ be the entire graph of a nonnegative smooth function $u : M^n \to \mathbb{R}$, with H constant and H_2 bounded from below. If u is bounded, then $u \equiv t_0$ for some $t_0 \in \mathbb{R}$.*

Again from Theorem 3, this time combined with Theorem 1.2 of [Rosenberg et al. 2013], we obtain:

Corollary 6. *Let M^n be a parabolic complete Riemannian manifold with bounded sectional curvature. Let $\Sigma^n(u) = \{(u(x), x) : x \in M^n\} \subset \mathbb{R} \times M^n$ be the entire graph of a smooth function $u : M^n \to \mathbb{R}$, with H constant and H_2 bounded from below. If u is bounded, then $u \equiv t_0$ for some $t_0 \in \mathbb{R}$.*

In the situation of Theorem 3, we saw that a constant mean curvature hypersurface satisfying the hypotheses of the theorem must be minimal. Theorem 1 suggests an interesting, related question: If a constant mean curvature hypersurface trapped between two planes is a graph, and the closure of the image of the Gauss map does not contain either pole, must the hypersurface be trivial? Osserman's theorem asserts that the hypersurface is indeed a plane when the ambient space is $\mathbb{R}^3$. When the ambient space is a product whose base has nonnegative Ricci curvature and sectional curvature bounded from below, Corollary 5 also gives a positive answer for this question, provided that the hypersurface is already a graph of a bounded and nonnegative function, while Corollary 6 deals with the parabolic case using the conformal invariance of parabolicity.

The proofs of Theorems 1 and 3 are given in Section 3.

2. Preliminaries

We consider an $(n+1)$-dimensional product space $\overline{M}^{n+1}$ of the form $\mathbb{R} \times M^n$, where M^n is an n-dimensional connected Riemannian manifold and $\overline{M}^{n+1}$ is endowed with the standard product metric

$$\langle \ , \ \rangle = \pi_{\mathbb{R}}^*(dt^2) + \pi_M^*(\langle \ , \ \rangle_M),$$

where $\pi_{\mathbb{R}}$ and π_M denote the canonical projections from $\mathbb{R} \times M^n$ onto each factor and $\langle\ ,\ \rangle_M$ is the Riemannian metric on M^n. For simplicity, we will just write $\overline{M}^{n+1} = \mathbb{R} \times M^n$ and $\langle\ ,\ \rangle = dt^2 + \langle\ ,\ \rangle_M$. For a fixed $t_0 \in \mathbb{R}$, we say that $M^n_{t_0} = \{t_0\} \times M^n$ is a *slice* of $\overline{M}^{n+1}$. It is not difficult to prove that such a slice of $\overline{M}^{n+1}$ is a totally geodesic hypersurface (see, for instance, [O'Neill 1983]).

Throughout this paper, we will deal with two-sided complete hypersurfaces $\psi : \Sigma^n \to \mathbb{R} \times M^n$. Let $\overline{\nabla}$ and ∇ denote the Levi-Civita connections in $\mathbb{R} \times M^n$ and Σ^n, respectively. The Gauss and Weingarten formulas for ψ are respectively

$$\overline{\nabla}_X Y = \nabla_X Y + \langle AX, Y\rangle N \tag{2-1}$$

and

$$AX = -\overline{\nabla}_X N \tag{2-2}$$

where $X, Y \in \mathfrak{X}(\Sigma)$ are tangent vector fields, and $A : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ is the Weingarten operator of Σ^n with respect to its orientation (unit normal vector field) N.

We will consider two particular functions naturally attached to such a hypersurface Σ^n: the (vertical) height function $h = (\pi_{\mathbb{R}})|_\Sigma$ and the angle function $\eta = \langle N, \partial_t\rangle$. Since Σ^n is assumed to be two-sided its angle function η is globally defined.

A simple computation shows that the gradient of $\pi_{\mathbb{R}}$ on $\mathbb{R} \times M^n$ is given by

$$\overline{\nabla}\pi_{\mathbb{R}} = \langle\overline{\nabla}\pi_{\mathbb{R}}, \partial_t\rangle\partial_t = \partial_t. \tag{2-3}$$

Consequently, from (2-3) we have that the gradient of h on Σ^n is

$$\nabla h = (\overline{\nabla}\pi_{\mathbb{R}})^\top = \partial_t^\top = \partial_t - \eta N, \tag{2-4}$$

where $(\ )^\top$ denotes the tangential component of a vector field in $\mathfrak{X}(\overline{M}^{n+1})$ along Σ^n. Hence, from (2-4) we get the relation

$$|\nabla h|^2 = 1 - \eta^2, \tag{2-5}$$

where $|\ |$ denotes the norm of a vector field on Σ^n. From Proposition 7.35 of [O'Neill 1983] we have

$$\overline{\nabla}_X \partial_t = 0 \tag{2-6}$$

for every $X \in \mathfrak{X}(\Sigma)$. Thus, from (2-4) and (2-6) we get

$$\nabla_X(\nabla h) = \nabla_X(\partial_t^\top) = \eta AX \tag{2-7}$$

for every tangent vector field $X \in \mathfrak{X}(\Sigma)$. Therefore, the Laplacian on Σ^n of the height function is given by

$$\Delta h = nH\eta, \tag{2-8}$$

where $H = (1/n)\,\mathrm{tr}(A)$ is the mean curvature of Σ^n relative to N. Moreover, as a particular case of Proposition 3.1 of [Caminha and Lima 2009], we obtain a useful formula for the Laplacian on Σ^n of the angle function η:

Lemma 7. *Let $\psi : \Sigma^n \to \mathbb{R} \times M^n$ be a hypersurface with orientation N, and let $\eta = \langle N, \partial_t \rangle$ be its angle function. If Σ^n has constant mean curvature H, then*

$$\Delta\eta = -(\mathrm{Ric}_M(N^*, N^*) + |A|^2)\eta,$$

where Ric_M denotes the Ricci curvature of the base M^n, N^* is the projection of the unit normal vector field N onto the base M^n and $|A|$ is the Hilbert–Schmidt norm of the shape operator A.

On the other hand, as in [O'Neill 1983], the curvature tensor R of a hypersurface $\psi : \Sigma^n \to \mathbb{R} \times M^n$ is given by

$$R(X, Y)Z = \nabla_{[X,Y]}Z - [\nabla_X, \nabla_Y]Z,$$

where $[\ ,\ ]$ denotes the Lie bracket and $X, Y, Z \in \mathfrak{X}(\Sigma)$. A well known fact is that, using (2-1) and (2-2), we can describe the curvature tensor R of the hypersurface Σ^n in terms of the shape operator A and the curvature tensor $\bar{R}$ of $\mathbb{R} \times M^n$ by the so-called Gauss equation given by

$$R(X, Y)Z = (\bar{R}(X, Y)Z)^\top + \langle AX, Z\rangle AY - \langle AY, Z\rangle AX \tag{2-9}$$

for tangent vector fields $X, Y, Z \in \mathfrak{X}(\Sigma)$.

To close this section, we recall the generalized maximum principle of Omori [1967] and Yau [1975], which will be the main analytical tool used in proving to prove our Bernstein-type results:

Lemma 8. *Let Σ^n be an n-dimensional complete Riemannian manifold whose Ricci curvature is bounded from below, and $f : \Sigma^n \to \mathbb{R}$ a smooth function which is bounded from below on Σ^n. Then there is a sequence of points (p_k) in Σ^n such that*

$$\lim_{k\to\infty} f(p_k) = \inf_\Sigma f, \quad \lim_{k\to\infty} |\nabla f(p_k)| = 0 \quad \text{and} \quad \liminf_{k\to\infty} \Delta f(p_k) \geq 0.$$

3. Proofs of Theorems 1 and 3

Proof of Theorem 1. Since we are assuming that η is bounded away from zero, we can suppose that $\eta > 0$ and, consequently, $\inf \eta > 0$. From Lemma 7, we have

$$\Delta\eta = -(\mathrm{Ric}_M(N^*, N^*) + |A|^2)\eta. \tag{3-1}$$

Since we are also assuming that the sectional curvature K_M of the base M^n is such that $K_M \geq -\kappa$ for some $\kappa > 0$, with a straightforward computation we get

$$\mathrm{Ric}_M(N^*, N^*) \geq -(n-1)\kappa|N^*|^2 = -(n-1)\kappa(1-\eta^2),$$

where N^* stands for the component of N tangent to M^n. Then, from (2-5) and (3-1) we obtain

$$\Delta\eta \leq -(|A|^2 - (n-1)\kappa|\nabla h|^2)\eta. \tag{3-2}$$

Thus, if we assume that the height function of Σ^n satisfies the hypothesis (1-1), from (1-1) and (3-2) we have

$$\Delta\eta \leq -(1-\alpha)|A|^2\eta. \tag{3-3}$$

On the other hand, we claim that the Ricci curvature of Σ^n is bounded from below. Therefore we can apply Lemma 8 to the function η, obtaining a sequence of points $p_k \in \Sigma^n$ such that $\liminf_{k\to\infty} \Delta\eta(p_k) \geq 0$ and $\lim_{k\to\infty} \eta(p_k) = \inf_{p\in\Sigma} \eta(p)$. Consequently, since we are assuming that the Weingarten operator A is bounded on Σ^n, from (3-3), up to a subsequence, we get

$$0 \leq \liminf_{k\to\infty} \Delta\eta(p_k) \leq -(1-\alpha) \lim_{k\to\infty} |A|^2(p_k) \inf_{p\in\Sigma} \eta(p) \leq 0.$$

Thus, we obtain that $\lim_{k\to\infty} |A|(p_k) = 0$ and, from (1-1), $\lim_{k\to\infty} |\nabla h|(p_k) = 0$. Hence, from (2-5) we conclude that $\inf_{p\in\Sigma} \eta(p) = 1$ and, consequently, $\eta \equiv 1$. Therefore, Σ is a slice.

It just remains to prove our claim that the Ricci curvature of Σ^n is bounded from below. For this, let us consider $X \in \mathfrak{X}(\Sigma)$ and a local orthonormal frame $\{E_1, \dots, E_n\}$ of $\mathfrak{X}(\Sigma)$. Then, it follows from the Gauss equation (2-9) that

$$\mathrm{Ric}_\Sigma(X, X) = \sum_i \langle \bar{R}(X, E_i)X, E_i\rangle + nH\langle AX, X\rangle - \langle AX, AX\rangle. \tag{3-4}$$

Thus, taking into account once more the lower bound of the sectional curvature of the base M^n, we have

$$\langle \bar{R}(X, E_i)X, E_i\rangle \geq -\kappa\big(\langle X^*, X^*\rangle_{M^n}\langle E_i^*, E_i^*\rangle_{M^n} - \langle X^*, E_i^*\rangle^2_{M^n}\big), \tag{3-5}$$

where $X^* = X - \langle X, \partial_t\rangle\partial_t$ and $E_i^* = E_i - \langle E_i, \partial_t\rangle\partial_t$ are the projections of the tangent vector fields X and E_i onto M^n, respectively. Then, adding up the relation (3-5) we get

$$\begin{aligned}\sum_i \langle \bar{R}(X, E_i)X, E_i\rangle &\geq -\kappa\big((n-1)|X|^2 - |\nabla h|^2|X|^2 - (n-2)\langle X, \nabla h\rangle^2\big)\\ &\geq -\kappa(n-1)|X|^2.\end{aligned}$$

Therefore, from (3-4), and using the Cauchy–Schwarz inequality, we have that the Ricci curvature of Σ^n satisfies the lower estimate

$$\mathrm{Ric}_\Sigma(X, X) \geq -\big((n-1)\kappa - |A||A - nHI|\big)|X|^2 \tag{3-6}$$

for all $X \in \mathfrak{X}(\Sigma)$. Therefore, taking into account that

$$|A|^2 = n^2H^2 - n(n-1)H_2, \tag{3-7}$$

our restrictions on H and H_2 guarantee that the Ricci curvature tensor of Σ^n is bounded from below and, hence, we conclude the first part of the proof of Theorem 1.

Now, let us suppose that the height function of Σ^n satisfies the hypothesis (1-2). In this case, from (3-2) and (3-7) we obtain

$$\Delta\eta \le -n(n-1)(H^2 - H_2)\eta. \tag{3-8}$$

Consequently, in a similar way as in the previous case, we can apply Lemma 8 in order to obtain a sequence of points $p_k \in \Sigma^n$ such that

$$0 \le \liminf_{k\to\infty} \Delta\eta(p_k) \le -n(n-1)\liminf_{k\to\infty}(H^2 - H_2)(p_k)\inf_{p\in\Sigma}\eta(p) \le 0.$$

Hence, up to a subsequence, $\lim_{k\to\infty}(H^2 - H_2)(p_k) = 0$. Moreover, since H is assumed to be constant, we get from (3-7) that

$$\lim_{k\to\infty}|A|^2(p_k) = nH^2.$$

Now we recall that $|A|^2 = \sum_i \kappa_i^2$, where the κ_i are the eigenvalues of A. Thus, up to taking a subsequence, for all $1 \le i \le n$ we have that $\lim_k \kappa_i(p_k) = \kappa_i^*$ for some $\kappa_i^* \in \mathbb{R}$. Motivated by this fact, we set

$$\frac{n(n-1)}{2}\overline{H}_2 = \sum_{i<j}\kappa_i^*\kappa_j^*,$$

and we note that $H = \frac{1}{n}\sum_i \kappa_i^*$. Thus $H^2 = \overline{H}_2$ and $\kappa_i^* = H$ for all $1 \le i \le n$. So, let $\{e_i\}$ be a local orthonormal frame of eigenvectors associated to the eigenvalues $\{\kappa_i\}$ of A. We can write $\nabla h = \sum_i \lambda_i e_i$, where the λ_i are continuous functions on Σ^n.

On the other hand, from (2-4) and (2-6) we have

$$X(\eta) = -\langle A(X), \partial_t\rangle = -\langle X, A(\partial_t^\top)\rangle = -\langle X, A(\nabla h)\rangle$$

for all $X \in \mathfrak{X}(\Sigma)$. Thus,

$$\nabla\eta = -A(\nabla h). \tag{3-9}$$

By applying Lemma 8 once more to the function η, from (3-9) we then get

$$\begin{aligned} 0 = \lim_k |A(\nabla h)|^2(p_k) &= \sum_i \lim_k(\kappa_i^2\lambda_i^2)(p_k) \\ &= \sum_i(\kappa_i^*)^2\lim_k\lambda_i^2(p_k) = H^2\sum_i\lim_k\lambda_i^2(p_k), \end{aligned}$$

up to taking a subsequence. If $H = 0$, from hypothesis (1-1) we have immediately that Σ^n is a slice. If $H^2 > 0$, then for all $1 \le i \le n$ we have $\lim_k \lambda_i(p_k) = 0$. Thus, $\lim_k |\nabla h|(p_k) = 0$ and, from (2-5),

$$\inf_{p\in\Sigma} \eta(p) = \lim_{k\to\infty} \eta(p_k) = 1.$$

Therefore, $\eta = 1$ on Σ^n, and hence Σ^n is a slice. □

Proof of Theorem 3. Note that, as in the proof of Theorem 1, our restrictions on the sectional curvature of the base M^n and the hypothesis on the mean curvatures H and H_2 guarantee that the Ricci curvature of Σ^n is bounded from below.

Now, suppose for instance that $H \ge 0$ on Σ^n. Thus, since Σ^n lies between two slices of $\mathbb{R} \times M^n$, from (2-8) and Lemma 8 we obtain a sequence of points $p_k \in \Sigma^n$ such that

$$0 \ge \limsup_{k\to\infty} \Delta h(p_k) = n \limsup_{k\to\infty} (H\eta)(p_k).$$

From (2-5) we also have

$$0 = \lim_{k\to\infty} |\nabla h|(p_k) = 1 - \lim_{k\to\infty} \eta^2(p_k).$$

Thus, if we suppose, for instance, that -1 is not in the closure of the image of η, we get $\lim\limits_{k\to\infty} \eta(p_k) = 1$. Consequently,

$$0 \ge \limsup_{k\to\infty} \Delta h(p_k) = n \limsup_{k\to\infty} H(p_k) \ge 0,$$

and, hence, we conclude that

$$\limsup_{k\to\infty} H(p_k) = 0.$$

If $H \le 0$, from (2-8) and (2-5) we can once more apply Lemma 8 in order to obtain a sequence $q_k \in \Sigma^n$ such that $0 \le \liminf_{k\to\infty} \Delta h(q_k) = n \liminf_{k\to\infty} (H\eta)(q_k)$, and, supposing once more that -1 is not in the closure of the image of η, we get

$$0 \le \liminf_{k\to\infty} \Delta h(p_k) = n \liminf_{k\to\infty} H(p_k) \le 0.$$

Consequently, we have

$$\liminf_{k\to\infty} H(p_k) = 0.$$

Therefore, in this case we also conclude that $\inf_\Sigma H = 0$. □

4. Entire vertical graphs in $\mathbb{R} \times M^n$

We recall that a *vertical graph* over a connected domain Ω of a complete Riemannian manifold M^n is determined by a smooth function $u \in C^\infty(\Omega)$, and is given by

$$\Sigma^n(u) = \{(u(x), x) : x \in \Omega\} \subset \mathbb{R} \times M^n.$$

From the product metric on the ambient space, $\Sigma^n(u)$ induces on Ω the metric

$$\langle\ ,\ \rangle = du^2 + \langle\ ,\ \rangle_{M^n}. \tag{4-1}$$

A vertical graph $\Sigma^n(u)$ is said to be *entire* if $\Omega = M^n$. Now, when the base M^n is complete, any entire vertical graph $\Sigma^n(u)$ in the product space $\mathbb{R} \times M^n$ is complete, because such a graph is properly immersed in $\mathbb{R} \times M^n$, which is obviously complete if M^n is. (Alternatively one can argue as follows: the Cauchy–Schwarz inequality and (4-1) give

$$\langle X, X\rangle = \langle X, X\rangle_{M^n} + \langle Du, X\rangle_{M^n}^2 \geq (1 + |Du|^2)\langle X, X\rangle_{M^n}$$

for every tangent vector field X on Σ^n. Hence, $\langle X, X\rangle \geq \langle X, X\rangle_{M^n}$. This implies that $L \geq L_{M^n}$, where L and L_{M^n} denote the length of a curve on $\Sigma^n(u)$ with respect to the Riemannian metrics $\langle\ ,\ \rangle$ and $\langle\ ,\ \rangle_{M^n}$; the completeness of $\Sigma^n(u)$ follows.)

Let $\Sigma^n(u) = \{(u(x), x) : x \in M^n\} \subset \mathbb{R} \times M^n$ be an entire vertical graph. The function $g : \mathbb{R} \times M^n \to \mathbb{R}$ given by $g(t, x) = t - u(x)$ is such that $\Sigma^n(u) = g^{-1}(0)$. Moreover, for all tangent vector fields X on $\mathbb{R} \times M^n$,

$$X(g) = \langle X, \partial_t\rangle \partial_t(g) + X^*(g) = \langle \partial_t - Du, X\rangle,$$

where $X^* = X - \langle X, \partial_t\rangle\partial_t$ is the projection of X onto the base M^n and Du is the gradient of u in M^n. Thus,

$$\overline{\nabla} g(u(x), x) = \partial_t|_{(u(x),x)} - Du(x) \quad \text{for all } x \in M^n.$$

Hence, the unit vector field

$$N(x) = \frac{1}{\sqrt{1 + |Du|^2}}(\partial_t|_{(u(x),x)} - Du(x)), \quad x \in M^n \tag{4-2}$$

gives an orientation for $\Sigma^n(u)$ such that $\eta > 0$ on it. Consequently, taking into account (2-5), from (4-2) we get

$$|\nabla h|^2 = \frac{|Du|^2}{1 + |Du|^2}. \tag{4-3}$$

Let us study the shape operator A of $\Sigma^n(u)$ with respect to the orientation given by (4-2). For any $X \in \mathfrak{X}(\Sigma(u))$, since $X = \langle Du, X\rangle_{M^n}\partial_t + X^*$, we have

$$AX = -\overline{\nabla}_X N = -\langle Du, X\rangle \overline{\nabla}_{\partial_t} N - \overline{\nabla}_X N. \tag{4-4}$$

Consequently, from (4-2) and (4-4), and with the aid of Proposition 7.35 of [O'Neill 1983], we verify that

$$AX = \frac{1}{\sqrt{1 + |Du|^2}} D_X Du + \frac{\langle D_X Du, Du\rangle}{(1 + |Du|^2)^{3/2}} Du, \tag{4-5}$$

where D denotes the Levi-Civita connection in M^n with respect to its metric $\langle\ ,\ \rangle_{M^n}$. Consequently, the mean curvature of $\Sigma^n(u)$ is given by

$$nH = \operatorname{Div} \frac{Du}{\sqrt{1+|Du|^2}}, \tag{4-6}$$

where Div stands for the divergence on the base M^n.

Remark 9. Salavessa [1989] showed that when the base M^n is complete noncompact, an entire graph $\Sigma^n(u)$ in $\mathbb{R}\times M^n$ with constant mean curvature H is minimal provided that the Cheeger constant $\mathfrak{h}(M)$ of the base M^n vanishes. We recall that

$$\mathfrak{h}(M) = \inf_D \frac{A(\partial D)}{V(D)},$$

where D ranges over all open submanifolds of M^n with compact closure in M^n and smooth boundary, and where $V(D)$, $A(\partial D)$ are the volume of D and the area of ∂D, respectively, relative to the metric of M^n.

Returning to the context of Theorem 1, we observe the condition that the angle function η of the hypersurface Σ^n is bounded away from zero assures that Σ^n is, in fact, an entire vertical graph $\Sigma^n(u)$ for some smooth function $u : M^n \to \mathbb{R}$. Consequently, considering the case that there exists a hypersurface with positive constant mean curvature H, and supposing that (1-2) holds, from (4-3) and (4-6) we see that Salavessa's argument allows us to get

$$\begin{aligned} nHV(D) &\le \int_D nH\,dV = \int_D \operatorname{Div} \frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\,dV \\ &= \oint_{\partial D} \left\langle \frac{Du}{\sqrt{1+|Du|^2}}, \nu \right\rangle dA \le \sqrt{\frac{n}{(n-1)\kappa}}\,HA(\partial D), \end{aligned}$$

where ν is the outward unit normal of ∂D. This yields the following lower estimate for the Cheeger constant of the base M^n:

$$\sqrt{n(n-1)\kappa} \le \mathfrak{h}(M).$$

Furthermore, recalling the stability operator $\mathscr{L} = -\Delta - \operatorname{Ric}(N,N) - |A|^2$, a constant mean curvature hypersurface Σ^n is said to be stable if

$$\int_\Sigma (\mathscr{L}f)f \ge 0 \quad \text{for all } f \in C_0^2(\Sigma). \tag{4-7}$$

We also note that under the stated hypothesis of Theorem 1, the hypersurface is a slice and therefore $\operatorname{Ric}(\partial_t,\partial_t) = 0$ and $|A|^2 \equiv 0$. Hence, in this case from (4-7) we see that the minimal hypersurface is stable.

We close our paper by presenting a suitable example of a nontrivial complete vertical graph $\Sigma^2(u)$ with constant mean curvature in the product space $\mathbb{R}\times\mathbb{H}^2$, which is directly related to the hypothesis of Theorems 1 and 3 (see the comments in Section 1).

Example 10. We consider the upper half-plane model for the two-dimensional hyperbolic space $\mathbb{H}^2$; that is, $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, endowed with the complete metric $\langle\ ,\ \rangle_{\mathbb{H}^2} = (1/y^2)(dx^2 + dy^2)$.

In this setting, let us define the smooth function $u : \mathbb{H}^2 \to \mathbb{R}$ by $u(x, y) = a \ln y$, $a \in \mathbb{R}$, and consider the entire vertical graph

$$\Sigma^2(u) = \{(a \ln y, x, y) : y > 0\} \subset \mathbb{R} \times \mathbb{H}^2.$$

We have $Du(x, y) = (0, ay)$ and hence $|Du(x, y)|^2 = a^2$. Moreover, the height function h of $\Sigma^2(u)$ satisfies

$$|\nabla h|^2 = \frac{|Du|^2}{1 + |Du|^2} = \frac{a^2}{1 + a^2}.$$

Thus, from (2-5) we have that the angle function η of $\Sigma^2(u)$ with respect to the orientation (4-2) is given by

$$\eta = \frac{1}{\sqrt{1 + |a|^2}}.$$

Consequently, by using that $\mathrm{Div} = \mathrm{Div}_0 - (2/y)dy$, where Div_0 denotes the divergent on $\mathbb{R}^2$, with a straightforward computation we verify that

$$2Hr^3 = r^2 y^2 \Delta_0 u - y^3(yQ(u) + u_y |D_0 u|_0^2), \tag{4-8}$$

where Δ_0, D_0 and $|\ |_0$ stand for the Laplacian, the gradient and the norm in the canonical Euclidean metric, $r = \sqrt{1 + |Du|^2} = \sqrt{1 + a^2}$ and

$$Q(u) = u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy}.$$

Thus, replacing $u(x, y) = a \ln y$ in (4-8), we obtain

$$H = \frac{a}{2\sqrt{1 + a^2}}$$

and, since η is a positive constant, from Lemma 7 we get

$$0 = \Delta\eta = -(|A|^2 - |\nabla h|^2)\eta, \tag{4-9}$$

and, hence,

$$|\nabla h|^2 = |A|^2.$$

Furthermore, from (3-7) we easily see that $H_2 = 0$ on $\Sigma^2(u)$. But, $H_2 = \kappa_1 \kappa_2$, where κ_1, κ_2 denote the eigenvalues of A. Therefore, considering $\kappa_2 = 0$ and using that $H = (\kappa_1 + \kappa_2)/2 = \kappa_1/2$, we obtain that $\kappa_1 = a/\sqrt{1 + a^2}$.

Finally, according to the stability criteria given in (4-7), from (4-9) we also conclude that $\Sigma^2(u)$ constitutes a nontrivial example of a stable surface in $\mathbb{R} \times \mathbb{H}^2$. Consequently, concerning the context of Theorem 1, we see that the stability of the hypersurface cannot alone guarantee the uniqueness result.

Acknowledgements

The authors would like to thank the referee for giving several valuable suggestions and comments which allowed them to improve this paper.

References

[Akutagawa 1987] K. Akutagawa, "On spacelike hypersurfaces with constant mean curvature in the de Sitter space", *Math. Z.* **196**:1 (1987), 13–19. MR 88h:53052 Zbl 0611.53047

[Aquino and Lima 2011] C. P. Aquino and H. F. de Lima, "On the rigidity of constant mean curvature complete vertical graphs in warped products", *Differential Geom. Appl.* **29**:4 (2011), 590–596. MR 2012f:53073 Zbl 1219.53056

[Barbosa and do Carmo 1974] J. L. Barbosa and M. do Carmo, "Stable minimal surfaces", *Bull. Amer. Math. Soc.* **80** (1974), 581–583. MR 49 #1307 Zbl 0286.53004

[Bérard and Sa Earp 2008] P. Bérard and R. Sa Earp, "Examples of H-hypersurfaces in $\mathbb{H}^n \times \mathbb{R}$ and geometric applications", *Mat. Contemp.* **34** (2008), 19–51. MR 2011b:53134 Zbl 1203.53053

[Bernstein 1910] S. Bernstein, "Sur les surfaces définies au moyen de leur courbure moyenne ou totale", *Ann. Sci. École Norm. Sup.* (3) **27** (1910), 233–256. MR 1509123

[Bombieri et al. 1969] E. Bombieri, E. De Giorgi, and M. Miranda, "Una maggiorazione a priori relativa alle ipersuperfici minimali non parametriche", *Arch. Rational Mech. Anal.* **32** (1969), 255–267. MR 40 #1898 Zbl 0184.32803

[Caminha and Lima 2009] A. Caminha and H. F. de Lima, "Complete vertical graphs with constant mean curvature in semi-Riemannian warped products", *Bull. Belg. Math. Soc. Simon Stevin* **16**:1 (2009), 91–105. MR 2010b:53107 Zbl 1160.53362

[Espinar and Rosenberg 2009] J. M. Espinar and H. Rosenberg, "Complete constant mean curvature surfaces and Bernstein type theorems in $M^2 \times \mathbb{R}$", *J. Differential Geom.* **82**:3 (2009), 611–628. MR 2010m:53015 Zbl 1180.53062

[Fujimoto 1988] H. Fujimoto, "On the number of exceptional values of the Gauss maps of minimal surfaces", *J. Math. Soc. Japan* **40**:2 (1988), 235–247. MR 89b:53013 Zbl 0629.53011

[Hoffman et al. 1982] D. A. Hoffman, R. Osserman, and R. Schoen, "On the Gauss map of complete surfaces of constant mean curvature in $\mathbb{R}^3$ and $\mathbb{R}^4$", *Comment. Math. Helv.* **57**:4 (1982), 519–531. MR 84f:53004 Zbl 0512.53008

[Lima 2014] H. F. de Lima, "Entire vertical graphs in Riemannian product spaces", *Quaest. Math.* (2014).

[Lima and Parente 2012] H. F. de Lima and U. L. Parente, "A Bernstein type theorem in $\mathbb{R} \times \mathbb{H}^n$", *Bull. Braz. Math. Soc. (N.S.)* **43**:1 (2012), 17–26. MR 2909921 Zbl 1276.53067

[Omori 1967] H. Omori, "Isometric immersions of Riemannian manifolds", *J. Math. Soc. Japan* **19** (1967), 205–214. MR 35 #6101 Zbl 0154.21501

[O'Neill 1983] B. O'Neill, *Semi-Riemannian geometry: with applications to relativity*, Pure and Applied Mathematics **103**, Academic Press, New York, 1983. MR 85f:53002 Zbl 0531.53051

[Osserman 1959] R. Osserman, "Proof of a conjecture of Nirenberg", *Comm. Pure Appl. Math.* **12** (1959), 229–232. MR 21 #4436 Zbl 0086.36202

[Rosenberg 2002] H. Rosenberg, "Minimal surfaces in $\mathbb{M}^2 \times \mathbb{R}$", *Illinois J. Math.* **46**:4 (2002), 1177–1195. MR 2004d:53015 Zbl 1036.53008

[Rosenberg et al. 2013] H. Rosenberg, F. Schulze, and J. Spruck, "The half-space property and entire positive minimal graphs in $M \times \mathbb{R}$", *J. Differential Geom.* **95**:2 (2013), 321–336. MR 3128986 Zbl 06218509

[Salavessa 1989] I. M. C. Salavessa, "Graphs with parallel mean curvature", *Proc. Amer. Math. Soc.* **107**:2 (1989), 449–458. MR 90a:53072 Zbl 0681.53031

[Yau 1975] S. T. Yau, "Harmonic functions on complete Riemannian manifolds", *Comm. Pure Appl. Math.* **28** (1975), 201–228. MR 55 #4042 Zbl 0291.31002

[Yau 1976] S. T. Yau, "Some function-theoretic properties of complete Riemannian manifold and their applications to geometry", *Indiana Univ. Math. J.* **25**:7 (1976), 659–670. MR 54 #5502 Zbl 0335.53041

Received April 30, 2013. Revised February 28, 2014.

HENRIQUE F. DE LIMA
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CAMPINA GRANDE-PB
58429-970
BRAZIL
henrique@dme.ufcg.edu.br

ERALDO A. LIMA JR.
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO CEARÁ
FORTALEZA-CE
60356-000
BRAZIL
eraldojrps@hotmail.com

ULISSES L. PARENTE
FACULDADE DE EDUCAÇÃO, CIÊNCIAS E LETRAS DO SERTÃO CENTRAL
UNIVERSIDADE ESTADUAL DO CEARÁ
QUIXADÁ-CE
63900-000
BRAZIL
ulisseslimaparente@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.407

EXISTENCE OF NONPARAMETRIC SOLUTIONS FOR A CAPILLARY PROBLEM IN WARPED PRODUCTS

JORGE H. LIRA AND GABRIELA A. WANDERLEY

We prove that there exist solutions for a nonparametric capillary problem in a wide class of Riemannian manifolds endowed with a Killing vector field. In other terms, we prove the existence of Killing graphs with prescribed mean curvature and prescribed contact angle along its boundary. These results may be useful for modeling stationary hypersurfaces under the influence of a nonhomogeneous gravitational field defined over an arbitrary Riemannian manifold.

1. Introduction

Let M be an $(n+1)$-dimensional Riemannian manifold endowed with a Killing vector field Y. Suppose that the distribution orthogonal to Y is of constant rank and integrable. Given an integral leaf P of that distribution, let $\Omega \subset P$ be a bounded domain with regular boundary $\Gamma = \partial\Omega$. We suppose for simplicity that Y is complete. In this case, let $\vartheta : \mathbb{R} \times \bar{\Omega} \to M$ be the flow generated by Y with initial values in M. In geometric terms, the ambient manifold is a warped product $M = P \times_{1/\sqrt{\gamma}} \mathbb{R}$, where $\gamma = 1/\|Y\|^2$.

The Killing graph of a differentiable function $u : \bar{\Omega} \to \mathbb{R}$ is the hypersurface $\Sigma \subset M$ parametrized by the map

$$X(x) = \vartheta(u(x), x), \quad x \in \bar{\Omega}.$$

The Killing cylinder K over Γ is in turn defined by

$$K = \{\vartheta(s, x) : s \in \mathbb{R},\ x \in \Gamma\}. \tag{1}$$

The height function with respect to the leaf P is measured by the arc length parameter ς of the flow lines of Y; that is,

$$\varsigma = \frac{1}{\sqrt{\gamma}} s.$$

Fixing these notations, we are able to formulate a capillary problem in this geometric

MSC2010: 53C42, 53C21.
Keywords: capillary, mean curvature, Killing graphs.

context which models stationary graphs under a gravity force whose intensity depends on the point in the space. More precisely, given a *gravitational potential* $\Psi \in C^{1,\alpha}(\overline{\Omega} \times \mathbb{R})$ we define the functional

$$\mathscr{A}[u] = \int_\Sigma \left(1 + \int_0^{u/\sqrt{\gamma}} \Psi(x, s(\varsigma))\, \mathrm{d}\varsigma\right) \mathrm{d}\Sigma. \tag{2}$$

The volume element $\mathrm{d}\Sigma$ of Σ is given by

$$\frac{1}{\sqrt{\gamma}}\sqrt{\gamma + \|\nabla u\|^2}\, \mathrm{d}\sigma,$$

where $\mathrm{d}\sigma$ is the volume element in P. In what follows we denote

$$W = \sqrt{\gamma + \|\nabla u\|^2}.$$

The first variation formula of this functional may be deduced as follows. Given an arbitrary function $v \in C_c^\infty(\Omega)$ we compute

$$\begin{aligned}
\frac{d}{d\tau}\bigg|_{\tau=0} & \mathscr{A}[u + \tau v] \\
&= \int_\Omega \left(\frac{1}{\sqrt{\gamma}} \frac{\langle \nabla u, \nabla v\rangle}{\sqrt{\gamma + \|\nabla u\|^2}} + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))\, v\right) \sqrt{\sigma}\, \mathrm{d}x \\
&= \int_\Omega \left(\mathrm{div}\left(\frac{1}{\sqrt{\gamma}} \frac{\nabla u}{W} v\right) - \mathrm{div}\left(\frac{1}{\sqrt{\gamma}} \frac{\nabla u}{W}\right) v + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))\, v\right) \sqrt{\sigma}\, \mathrm{d}x \\
&\quad - \int_\Omega \left(\frac{1}{\sqrt{\gamma}} \mathrm{div}\left(\frac{\nabla u}{W}\right) - \frac{1}{\sqrt{\gamma}} \left\langle \frac{\nabla \gamma}{2\gamma}, \frac{\nabla u}{W}\right\rangle - \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))\right) v \sqrt{\sigma}\, \mathrm{d}x,
\end{aligned}$$

where $\sqrt{\sigma}\, \mathrm{d}x$ is the volume element $\mathrm{d}\sigma$ expressed in terms of local coordinates in P. The differential operators div and ∇ are respectively the divergence and gradient in P with respect to the metric induced from M.

We conclude that stationary functions satisfy the capillary-type equation

$$\mathrm{div}\left(\frac{\nabla u}{W}\right) - \left\langle \frac{\nabla \gamma}{2\gamma}, \frac{\nabla u}{W}\right\rangle = \Psi. \tag{3}$$

Notice that a Neumann boundary condition arises naturally from this variational setting: given a $C^{2,\alpha}$ function $\Phi : K \to (-1, 1)$, we impose the prescribed angle condition

$$\langle N, \nu\rangle = \Phi \tag{4}$$

along $\partial\Sigma$, where

$$N = \frac{1}{W}(\gamma Y - \vartheta_* \nabla u) \tag{5}$$

is the unit normal vector field along Σ satisfying $\langle N, Y\rangle > 0$ and ν is the unit normal vector field along K pointing into the Killing cylinder over Ω.

Equation (3) is the prescribed mean curvature equation for Killing graphs. A general existence result for solutions of the Dirichlet problem for this equation may be found in [Dajczer et al. 2008] and [Dajczer and de Lira 2012]. There the authors used local perturbations of the Killing cylinders as barriers for obtaining height and gradient estimates. However this kind of barrier is not suitable to obtain a priori estimates for solutions of Neumann problems. Indeed, these barriers depend on Dirichlet boundary data and do not involve any a priori information about the prescribed contact angle. It turns out that for Dirichlet boundary conditions the slope of the graph along the boundary is controlled in terms of the height of the graph.

For that reason we now consider local perturbations of the graph itself, adapted from the original approach by N. Korevaar [1988] and its extension by M. Calle and L. Shahriyari [2011].

Following these two sources we suppose that the data Ψ and Φ satisfy

(i) $|\Psi| + \|\bar{\nabla}\Psi\| \leq C_\Psi$ in $\bar{\Omega} \times \mathbb{R}$,

(ii) $\langle \bar{\nabla}\Psi, Y\rangle \geq \beta > 0$ in $\bar{\Omega} \times \mathbb{R}$,

(iii) $\langle \bar{\nabla}\Phi, Y\rangle \leq 0$,

(iv) $(1 - \Phi^2) \geq \beta'$,

(v) $|\Phi| + \|\nabla\Phi\| + \|\nabla^2\Phi\| \leq C_\Phi$ in K,

for some positive constants C_Ψ, C_Φ, β and β', where $\bar{\nabla}$ denotes the Riemannian connection in M. Assumption (ii) is classically referred to as the *positive gravity* condition. Even in the Euclidean space, it seems to be an essential assumption in order to obtain a priori height estimates. A very geometric discussion about this issue may be found in [Concus and Finn 1974]. Condition (iii) is the same as in [Calle and Shahriyari 2011] and [Korevaar 1988] since in those references N is chosen in such a way that $\langle N, Y\rangle > 0$.

The main result in this paper is the following:

Theorem 1. *Let Ω be a bounded $C^{3,\alpha}$ domain in P. Suppose that $\Psi \in C^{1,\alpha}(\bar{\Omega} \times \mathbb{R})$ and $\Phi \in C^{2,\alpha}(K)$ with $|\Phi| \leq 1$ satisfy conditions* (i)–(v) *above. Then there exists a unique solution $u \in C^{3,\alpha}(\bar{\Omega})$ of the capillary problem* (3)–(4).

We observe that $\Psi = nH$, where H is the mean curvature of Σ calculated with respect to N. Therefore Theorem 1 establishes the existence of Killing graphs with prescribed mean curvature Ψ and prescribed contact angle with K along the boundary. Since the Riemannian product $P \times \mathbb{R}$ corresponds to the particular case where $\gamma = 1$, our result extends the main existence theorem in [Calle and Shahriyari 2011]. Space forms constitute other important examples of the kind of warped products we are considering. In particular, we encompass the case of Killing graphs over totally geodesic hypersurfaces in the hyperbolic space $\mathbb{H}^{n+1}$.

In Section 2, we prove a priori height estimates for solutions of (3)–(4) based on the method presented in [Uraltseva 1973]. These height estimates are one of the main steps for using the well-known continuity method in order to prove Theorem 1. At this respect, we refer the reader to the classical references [Concus and Finn 1974], [Gerhardt 1976] and [Simon and Spruck 1976].

Section 3 contains the proof of interior and boundary gradient estimates. There we follow closely a method due to Korevaar [1988] for graphs in the Euclidean spaces and extended by Calle and Shahriyari [2011] for Riemannian products. Finally the classical continuity method is applied to (3)–(4) in Section 4 for proving the existence result.

2. Height estimates

In this section, we use a technique developed by N. Uraltseva [1973] (see also [Ladyzhenskaya and Uraltseva 1964] and [Gilbarg and Trudinger 2001] for classical references on the subject) in order to obtain a height estimate for solutions of the capillary problem (3)–(4). This estimate requires the *positive gravity* assumption (ii) stated in the introduction.

Proposition 2. *Set*

$$\beta = \inf_{\Omega\times\mathbb{R}} \langle \bar\nabla\Psi, Y\rangle \quad \textit{and} \quad \mu = \sup_{\Omega} \Psi(x,0). \tag{6}$$

Suppose that $\beta > 0$. Then any solution u of (3)–(4) *satisfies*

$$|u(x)| \le \frac{\sup_\Omega \|Y\|}{\inf_\Omega \|Y\|}\frac{\mu}{\beta} \tag{7}$$

for all $x \in \bar\Omega$.

Proof. Fix an arbitrary real number k with

$$k > \frac{\sup_\Omega \|Y\|}{\inf_\Omega \|Y\|}\frac{\mu}{\beta}. \tag{8}$$

Suppose that the superlevel set

$$\Omega_k = \{x \in \Omega : u(x) > k\}$$

has nonzero Lebesgue measure. Define $u_k : \Omega \to \mathbb{R}$ as

$$u_k(x) = \max\{u(x) - k, 0\}.$$

From the variational formulation we have

$$\begin{aligned}0 &= \int_{\Omega_k} \left(\frac{1}{\sqrt{\gamma}} \frac{\langle \nabla u, \nabla u_k \rangle}{\sqrt{\gamma + \|\nabla u\|^2}} + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))\, u_k \right) \sqrt{\sigma}\, \mathrm{d}x \\ &= \int_{\Omega_k} \left(\frac{1}{\sqrt{\gamma}} \frac{\|\nabla u\|^2}{W} + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))(u-k) \right) \sqrt{\sigma}\, \mathrm{d}x \\ &= \int_{\Omega_k} \left(\frac{1}{\sqrt{\gamma}} \frac{W^2 - \gamma}{W} + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))(u-k) \right) \sqrt{\sigma}\, \mathrm{d}x \\ &= \int_{\Omega_k} \left(\frac{W}{\sqrt{\gamma}} - \frac{\sqrt{\gamma}}{W} + \frac{1}{\sqrt{\gamma}} \Psi(x, u(x))(u-k) \right) \sqrt{\sigma}\, \mathrm{d}x.\end{aligned}$$

However

$$\Psi(x, u(x)) = \Psi(x, 0) + \int_0^{u(x)} \frac{\partial \Psi}{\partial s}\, \mathrm{d}s \geq -\mu + \beta u(x).$$

Since $\sqrt{\gamma}/W \leq 1$ we conclude that

$$|\Omega_k| - |\Omega_k| - \mu \int_{\Omega_k} \frac{1}{\sqrt{\gamma}}(u-k) + \beta \int_{\Omega_k} \frac{1}{\sqrt{\gamma}} u(u-k) \leq 0,$$

where $|\Omega_k|$ is the Lebesgue measure of Ω_k. Hence we have

$$\beta \int_{\Omega_k} \frac{1}{\sqrt{\gamma}} u(u-k) \leq \mu \int_{\Omega_k} \frac{1}{\sqrt{\gamma}}(u-k).$$

It follows that

$$\beta k \inf_{\Omega} \|Y\| \int_{\Omega_k} (u-k) \leq \mu \sup_{\Omega} \|Y\| \int_{\Omega_k} (u-k).$$

Since $|\Omega_k| \neq 0$ we have

$$k \leq \frac{\sup_\Omega \|Y\|}{\inf_\Omega \|Y\|} \frac{\mu}{\beta},$$

which contradicts the choice of k. We conclude that

$$|\Omega_k| = 0 \quad \text{for all } k \geq \frac{\sup_\Omega \|Y\|}{\inf_\Omega \|Y\|} \frac{\mu}{\beta}.$$

This implies that

$$u(x) \leq \frac{\sup_\Omega \|Y\|}{\inf_\Omega \|Y\|} \frac{\mu}{\beta}$$

for all $x \in \overline{\Omega}$. A lower estimate may be deduced in a similar way. □

Remark 3. The construction of geometric barriers similar to those in [Concus and Finn 1974] is also possible at least in the case where P is endowed with a rotationally invariant metric and Ω is contained in a normal neighborhood of a pole of P.

3. Gradient estimates

Let Ω' be a subset of Ω and define

$$\Sigma' = \{\vartheta(u(x), x) : x \in \Omega'\} \subset \Sigma \tag{9}$$

to be the graph of $u|_{\Omega'}$. Let $\mathcal{O}$ be an open subset in M containing Σ'. We consider a vector field $Z \in \Gamma(TM)$ with bounded C^2 norm and supported in $\mathcal{O}$. Hence there exists $\varepsilon > 0$ such that the local flow $\Xi : (-\varepsilon, \varepsilon) \times \mathcal{O} \to M$ generated by Z is well-defined. We also suppose that

$$\langle Z(y), \nu(y)\rangle = 0 \tag{10}$$

for any $y \in K \cap \mathcal{O}$. This implies that the flow line of Z passing through a point $y \in K \cap \mathcal{O}$ is entirely contained in K.

We define a variation of Σ by a one-parameter family of hypersurfaces Σ_τ, $\tau \in (-\varepsilon, \varepsilon)$, parametrized by $X_\tau : \overline{\Omega} \to M$, where

$$X_\tau(x) = \Xi(\tau, \vartheta(u(x), x)), \quad x \in \overline{\Omega}. \tag{11}$$

It follows from the implicit function theorem that there exist $\Omega_\tau \subset P$ and $u_\tau : \overline{\Omega}_\tau \to \mathbb{R}$ such that Σ_τ is the graph of u_τ. Moreover, $\Omega_\tau \subset \Omega$.

Hence given a point $y \in \Sigma$, denote $y_\tau = \Xi(\tau, y) \in \Sigma_\tau$. It follows that there exists $x_\tau \in \Omega_\tau$ such that $y_\tau = \vartheta(u_\tau(x_\tau), x_\tau)$. Then we denote by $\hat{y}_\tau = \vartheta(u(x_\tau), x_\tau)$ the point in Σ in the flow line of Y passing through y_τ. The vertical separation between y_τ and $\hat{y}_\tau$ is by definition the function $s(y, \tau) = u_\tau(x_\tau) - u(x_\tau)$.

Lemma 4. *For any $\tau \in (-\varepsilon, \varepsilon)$, let A_τ and H_τ be, respectively, the Weingarten map and the mean curvature of the hypersurface Σ_τ calculated with respect to the unit normal vector field N_τ along Σ_τ which satisfies $\langle N_\tau, Y\rangle > 0$. Denote $H = H_0$ and $A = A_0$. If $\zeta \in C^\infty(\mathcal{O})$ and $T \in \Gamma(T\mathcal{O})$ are defined by*

$$Z = \zeta N_\tau + T \tag{12}$$

with $\langle T, N_\tau\rangle = 0$ then

(i) $\partial s/\partial\tau|_{\tau=0} = \langle Z, N\rangle W$,

(ii) $\overline{\nabla}_Z N|_{\tau=0} = -AT - \nabla^\Sigma \zeta$,

(iii) $\partial H/\partial\tau|_{\tau=0} = \Delta_\Sigma \zeta + (\|A\|^2 + \mathrm{Ric}_M(N, N))\zeta + \langle \overline{\nabla}\Psi, Z\rangle$,

where $W = \langle Y, N_\tau\rangle^{-1} = (\gamma + \|\nabla u_\tau\|^2)^{-1/2}$. The operators ∇^Σ and Δ_Σ are, respectively, the intrinsic gradient operator and the Laplace–Beltrami operator in Σ with respect to the induced metric. Moreover, $\overline{\nabla}$ and Ric_M denote, respectively, the Riemannian covariant derivative and the Ricci tensor in M.

Proof. (i) Let $(x^i)_{i=1}^n$ be a set of local coordinates in $\Omega \subset P$. Differentiating (11) with respect to τ we obtain

$$X_{\tau *}\frac{\partial}{\partial \tau} = Z|_{X_\tau} = \zeta N_\tau + T.$$

On the other hand differentiating both sides of

$$X_\tau(x) = \vartheta(u_\tau(x_\tau), x_\tau)$$

with respect to τ we have

$$\begin{aligned} X_{\tau *}\frac{\partial}{\partial \tau} &= \left(\frac{\partial u_\tau}{\partial \tau} + \frac{\partial u_\tau}{\partial x^i}\frac{\partial x_\tau^i}{\partial \tau}\right)\vartheta_* Y + \frac{\partial x_\tau^i}{\partial \tau}\vartheta_*\frac{\partial}{\partial x^i} \\ &= \frac{\partial u_\tau}{\partial \tau}\vartheta_* Y + \frac{\partial x_\tau^i}{\partial \tau}\left(\vartheta_*\frac{\partial}{\partial x^i} + \frac{\partial u_\tau}{\partial x^i}\vartheta_* Y\right). \end{aligned}$$

Since the term between parenthesis after the second equality is a tangent vector field in Σ_τ we conclude that

$$\frac{\partial u_\tau}{\partial \tau}\langle Y, N_\tau\rangle = \left\langle X_{\tau *}\frac{\partial}{\partial \tau}, N_\tau\right\rangle = \zeta,$$

and it follows that

$$\frac{\partial u_\tau}{\partial \tau} = \zeta W$$

and

$$\frac{\partial s}{\partial \tau} = \frac{\partial}{\partial \tau}(u_\tau - u) = \frac{\partial u_\tau}{\partial \tau} = \zeta W.$$

(ii) Now we have

$$\begin{aligned} \langle \bar\nabla_Z N_\tau, X_*\partial_i\rangle &= -\langle N_\tau, \bar\nabla_Z X_*\partial_i\rangle = -\langle N_\tau, \bar\nabla_{X_*\partial_i} Z\rangle = -\langle N_\tau, \bar\nabla_{X_*\partial_i}(\zeta N + T)\rangle \\ &= -\langle N_\tau, \bar\nabla_{X_*\partial_i} T\rangle - \langle N_\tau, \bar\nabla_{X_*\partial_i}\zeta N_\tau\rangle = -\langle A_\tau T, X_*\partial_i\rangle - \langle \nabla^\Sigma \zeta, X_*\partial_i\rangle, \end{aligned}$$

for any $1 \le i \le n$. It follows that

$$\bar\nabla_Z N = -AT - \nabla^\Sigma \zeta.$$

(iii) This is a well-known formula whose proof may be found in a number of references, such as [Gerhardt 2006]. □

For further reference, we point out that the comparison principle [Gilbarg and Trudinger 2001] when applied to (3)–(4) may be stated in geometric terms as follows. Fix τ, and let $x \in \bar\Omega'$ be a point of maximal vertical separation $s(\cdot, \tau)$. If x is an interior point we have

$$\nabla u_\tau(x, \tau) - \nabla u(x) = \nabla s(x, \tau) = 0,$$

which implies that the graphs of the functions u_τ and $u + s(x, \tau)$ are tangent at their common point $y_\tau = \vartheta(u_\tau(x), x)$. Since the graph of $u + s(x, \tau)$ is obtained from Σ only by a translation along the flow lines of Y we conclude that the mean curvatures of these two graphs are the same at corresponding points. Since the graph of $u + s(x, \tau)$ is locally above the graph of u_τ we conclude that

$$H(\hat{y}_\tau) \geq H_\tau(y_\tau). \tag{13}$$

If $x \in \partial\Omega \subset \partial\Omega'$, we have

$$\langle \nabla u_\tau, \nu \rangle|_x - \langle \nabla u, \nu \rangle|_x = \langle \nabla s, \nu \rangle \leq 0,$$

since ν points toward Ω. This implies that

$$\langle N, \nu \rangle|_{y_\tau} \geq \langle N, \nu \rangle|_{\hat{y}_\tau}. \tag{14}$$

3.1. ***Interior gradient estimate.***

Proposition 5. *Let $B_R(x_0) \subset \Omega$, where $R < \operatorname{inj} P$. Then there exists a constant $C > 0$ depending on β, C_Ψ, Ω and K such that*

$$\|\nabla u(x)\| \leq C \frac{R^2}{R^2 - d^2(x)}, \tag{15}$$

where $d = \operatorname{dist}(x_0, x)$ in P.

Proof. Fix $\Omega' = B_R(x_0) \subset \Omega$. We consider the vector field Z given by

$$Z = \zeta N, \tag{16}$$

where ζ is a function to be defined later. Fix $\tau \in [0, \varepsilon)$, and let $x \in B_R(x_0)$ be a point where the vertical separation $s(\cdot\,, \tau)$ attains a maximum value.

If $y = \vartheta(u(x), x)$ it follows that

$$H_\tau(y_\tau) - H_0(y) = \left.\frac{dH_\tau}{d\tau}\right|_{\tau=0} \tau + o(\tau). \tag{17}$$

However, the comparison principle implies that $H_0(\hat{y}_\tau) \geq H_\tau(y_\tau)$. By Lemma 1(iii) we conclude that

$$H_0(\hat{y}_\tau) - H_0(y) \geq \left.\frac{dH_\tau}{d\tau}\right|_{\tau=0} \tau + o(\tau) = (\Delta_\Sigma \zeta + \|A\|^2 \zeta + \operatorname{Ric}_M(N, N)\zeta)\tau + o(\tau).$$

Since $\hat{y}_\tau = \vartheta(-s(y, \tau), y_\tau)$, we have

$$\left.\frac{d\hat{y}_\tau}{d\tau}\right|_{\tau=0} = -\frac{ds}{d\tau}\vartheta_* \frac{\partial}{\partial s} + \frac{\partial y_\tau^i}{\partial \tau}\vartheta_* \frac{\partial}{\partial x^i} = -\frac{ds}{d\tau} Y + \left.\frac{dy_\tau}{d\tau}\right|_{\tau=0} = -\frac{ds}{d\tau} Y + Z(y). \tag{18}$$

Hence using Lemma 1(i) and (16) we have

$$\left.\frac{d\hat{y}_\tau}{d\tau}\right|_{\tau=0} = -\zeta W Y + \zeta N. \tag{19}$$

On the other hand, for each $\tau \in (-\varepsilon, \varepsilon)$ there exists a smooth $\xi : (-\varepsilon, \varepsilon) \to TM$ such that

$$\hat{y}_\tau = \exp_y \xi(\tau).$$

Hence we have

$$\left.\frac{d\hat{y}_\tau}{d\tau}\right|_{\tau=0} = \xi'(0).$$

With a slight abuse of notation we denote $\Psi(s, x)$ by $\Psi(y)$, where $y = \vartheta(s, x)$. It results that

$$\begin{aligned} H_0(\hat{y}_\tau) - H_0(y) &= \Psi(x_\tau, u(x_\tau)) - \Psi(x, u(x)) = \Psi(\exp_y \xi_\tau) - \Psi(y) \\ &= \langle \bar{\nabla}\Psi|_y, \xi'(0)\rangle\tau + o(\tau). \end{aligned}$$

However,

$$\langle \bar{\nabla}\Psi, \xi'(0)\rangle = \zeta\langle \bar{\nabla}\Psi, N - WY\rangle = -\zeta W \frac{\partial \Psi}{\partial s} + \zeta\langle \bar{\nabla}\Psi, N\rangle. \tag{20}$$

We conclude that

$$-\zeta W \frac{\partial \Psi}{\partial s}\tau + \zeta\langle \bar{\nabla}\Psi, N\rangle\tau + o(\tau) \geq (\Delta_\Sigma \zeta + \|A\|^2\zeta + \mathrm{Ric}_M(N, N)\zeta)\tau + o(\tau).$$

Suppose that

$$W(x) > \frac{C + \|\bar{\nabla}\Psi\|}{\beta} \tag{21}$$

for a constant $C > 0$ to be chosen later. Hence we have

$$(\Delta_\Sigma \zeta + \mathrm{Ric}_M(N, N)\zeta)\tau + C\zeta\tau \leq o(\tau).$$

Following [Calle and Shahriyari 2011] and [Korevaar 1988] we choose

$$\zeta = 1 - \frac{d^2}{R^2},$$

where $d = \mathrm{dist}(x_0, \cdot)$. It follows that

$$\nabla^\Sigma \zeta = -\frac{2d}{R^2}\nabla^\Sigma d,$$

and

$$\Delta_\Sigma \zeta = -\frac{2d}{R^2}\Delta_\Sigma d - \frac{2}{R^2}\|\nabla^\Sigma d\|^2.$$

However, using the fact that P is totally geodesic and that $[Y, \bar{\nabla} d] = 0$, we have

$$\begin{aligned}\Delta_\Sigma d &= \Delta_M d - \langle \bar{\nabla}_N \bar{\nabla} d, N\rangle + nH\langle \bar{\nabla} d, N\rangle \\ &= \Delta_P d - \left\langle \nabla_{\nabla u/W} \nabla d, \frac{\nabla u}{W}\right\rangle - \gamma^2\langle Y, N\rangle^2\langle \bar{\nabla}_Y \bar{\nabla} d, Y\rangle + nH\langle \bar{\nabla} d, N\rangle.\end{aligned}$$

Let $\pi : M \to P$ be the projection defined by $\pi(\vartheta(s, x)) = x$. Then

$$\pi_* N = -\frac{\nabla u}{W}.$$

We denote

$$\pi_* N^\perp = \pi_* N - \langle \pi_* N, \nabla d\rangle \nabla d.$$

If $\mathcal{A}_d$ and $\mathcal{H}_d$ denote, respectively, the Weingarten map and the mean curvature of the geodesic ball $B_d(x_0)$ in P we conclude that

$$\Delta_\Sigma d = n\mathcal{H}_d - \langle \mathcal{A}_d(\pi_* N^\perp), \pi_* N^\perp\rangle + \gamma\langle Y, N\rangle^2 \kappa + nH\langle \bar{\nabla} d, N\rangle,$$

where

$$\kappa = -\gamma\langle \bar{\nabla}_Y \bar{\nabla} d, Y\rangle$$

is the principal curvature of the Killing cylinder over $B_d(x_0)$ relative to the principal direction Y. Therefore we have

$$|\Delta_\Sigma d| \le C_1\Big(C_\Psi, \sup_{B_R(x_0)} (\mathcal{H}_d + \kappa), \sup_{B_R(x_0)} \gamma\Big)$$

in $B_R(x_0)$. Hence setting

$$C_2 = \sup_{B_R(x_0)} \mathrm{Ric}_M,$$

we fix

$$C = \max\Big\{2(C_1 + C_2), \sup_{\mathbb{R}\times\Omega} \|\bar{\nabla}\Psi\|\Big\}. \tag{22}$$

With this choice we conclude that

$$C\zeta \le \frac{o(\tau)}{\tau},$$

a contradiction. This implies that

$$W(x) \le \frac{C - \|\bar{\nabla}\Psi\|}{\beta}. \tag{23}$$

However,

$$\zeta(z)W(z) + o(\tau) = s(X(z), \tau) \le s(X(x), \tau) = \zeta(x)W(x) + o(\tau)$$

for any $z \in B_R(x_0)$. It follows that

$$W(z) \leq \frac{R^2 - d^2(z)}{R^2 - d^2(x)} W(x) + o(\tau) \leq \frac{R^2}{R^2 - d^2(x)} \frac{C - \|\overline{\nabla}\Psi\|}{\beta} + o(\tau) \leq \widetilde{C} \frac{R^2}{R^2 - d^2(x)},$$

for very small $\varepsilon > 0$. □

Remark 6. If Ω satisfies the interior sphere condition for a uniform radius $R > 0$, we conclude that

$$W(x) \leq \frac{C}{d_\Gamma(x)}, \tag{24}$$

for $x \in \Omega$, where $d_\Gamma(x) = \mathrm{dist}(x, \Gamma)$.

3.2. ***Boundary gradient estimates.*** Now we establish boundary gradient estimates using another local perturbation of the graph, which this time has also tangential components.

Proposition 7. *Let $x_0 \in P$ and $R > 0$ such that $3R < \mathrm{inj}\, P$. Denote by Ω' the subdomain $\Omega \cap B_{2R}(x_0)$. Then there exists a positive constant C, depending only on $R, \beta, \beta', C_\Psi, C_\Phi, \Omega, K$, such that*

$$W(x) \leq C, \tag{25}$$

for all $x \in \overline{\Omega}'$.

Proof. Now we consider the subdomain $\Omega' = \Omega \cap B_{2R}(x_0)$. We define

$$Z = \eta N + X, \tag{26}$$

where

$$\eta = \alpha_0 v + \alpha_1 d_\Gamma,$$

and α_0 and α_1 are positive constants to be chosen and d_Γ is a smooth extension of the distance function $\mathrm{dist}(\,\cdot\,, \Gamma)$ to Ω' with $\|\nabla d_\Gamma\| \leq 1$ and

$$v = 4R^2 - d^2,$$

where $d = \mathrm{dist}(x_0, \cdot\,)$. Moreover,

$$X = \alpha_0 \Phi(v\nu - d_\Gamma \nabla v).$$

In this case we have

$$\zeta = \eta + \langle X, N\rangle = \alpha_0 v + \alpha_1 d_\Gamma + \alpha_0 \Phi(v\langle N, \nu\rangle - d_\Gamma \langle N, \nabla v\rangle).$$

Fix $\tau \in [0, \varepsilon)$, and let $x \in \overline{\Omega}'$ be a point where the maximal vertical separation between Σ and Σ_τ is attained. We first suppose that $x \in \mathrm{int}(\partial\Omega' \cap \partial\Omega)$. In this case, setting $y_\tau = \vartheta(u_\tau(x), x) \in \Sigma_\tau$ and $\hat{y}_\tau = \vartheta(u(x), x) \in \Sigma$, it follows from the comparison principle that

$$\langle N_\tau, \nu\rangle|_{y_\tau} \geq \langle N, \nu\rangle|_{\hat{y}_\tau}. \tag{27}$$

Note that $\hat{y}_\tau \in \partial\Sigma$. Moreover, since $Z|_{K\cap\mathcal{O}}$ is tangent to K there exists $y \in \partial\Sigma$ such that

$$y = \Xi(-\tau, y_\tau).$$

We claim that

$$\left|\left\langle \overline{\nabla}\langle N_\tau, \nu\rangle, \frac{dy_\tau}{d\tau}\Big|_{\tau=0}\right\rangle\right| \leq \alpha_1(1-\Phi^2) + \widetilde{C}\alpha_0, \tag{28}$$

for some positive constant $\widetilde{C} = C(C_\Phi, K, \Omega, R)$.

Hence (4) implies that

$$\langle N, \nu\rangle|_{\hat{y}_\tau} - \langle N, \nu\rangle|_y = \Phi(\hat{y}_\tau) - \Phi(y) = \tau\left\langle \overline{\nabla}\Phi, \frac{d\hat{y}_\tau}{d\tau}\Big|_{\tau=0}\right\rangle + o(\tau).$$

Therefore

$$\langle N, \nu\rangle|_{y_\tau} - \langle N, \nu\rangle|_y \geq \tau\left\langle \overline{\nabla}\Phi, \frac{d\hat{y}_\tau}{d\tau}\Big|_{\tau=0}\right\rangle + o(\tau).$$

On the other hand we have

$$\langle N, \nu\rangle|_{y_\tau} - \langle N, \nu\rangle|_y = \tau\left\langle \overline{\nabla}\langle N, \nu\rangle, \frac{dy_\tau}{d\tau}\Big|_{\tau=0}\right\rangle + o(\tau).$$

We conclude that

$$\tau\left\langle \overline{\nabla}\langle N, \nu\rangle, \frac{dy_\tau}{d\tau}\Big|_{\tau=0}\right\rangle \geq \tau\left\langle \overline{\nabla}\Phi, \frac{d\hat{y}_\tau}{d\tau}\Big|_{\tau=0}\right\rangle + o(\tau).$$

Hence we have

$$\alpha_1(1-\Phi^2)\tau + \widetilde{C}\alpha_0\tau \geq \tau\left\langle \overline{\nabla}\Phi, \frac{d\hat{y}_\tau}{d\tau}\Big|_{\tau=0}\right\rangle + o(\tau).$$

It follows from (28) that

$$\alpha_1(1-\Phi^2) + \widetilde{C}\alpha_0 \geq -\zeta W\langle \overline{\nabla}\Phi, Y\rangle + \zeta\langle \overline{\nabla}\Phi, N\rangle + o(\tau)/\tau.$$

Since

$$\langle \overline{\nabla}\Phi, Y\rangle = \frac{\partial\Phi}{\partial s} \leq 0,$$

we conclude that

$$W(x) \leq C(C_\Phi, \beta', K, \Omega, R). \tag{29}$$

We now prove the claim. For that, observe that Lemma 1(ii) implies that

$$\begin{aligned}\langle N,\nu\rangle|_{y_\tau}-\langle N,\nu\rangle|_y &= \tau\frac{\partial}{\partial\tau}\Big|_{\tau=0}\langle N_\tau,\nu\rangle|_{y_\tau}+o(\tau)\\ &= \tau(\langle N,\overline{\nabla}_Z\nu\rangle|_y-\langle AT+\nabla^\Sigma\zeta,\nu\rangle|_y)+o(\tau).\end{aligned}$$

Since $Z|_y\in T_yK$, it follows that

$$\langle N,\nu\rangle|_{y_\tau}-\langle N,\nu\rangle|_y=-\tau(\langle A_KZ,N\rangle|_y+\langle AT+\nabla^\Sigma\zeta,\nu\rangle|_y)+o(\tau),$$

where A_K is the Weingarten map of K with respect to ν. We conclude that

$$-\tau(\langle A_KZ,N\rangle|_y+\langle AT+\nabla^\Sigma\zeta,\nu\rangle|_y)\ge\tau\Big\langle\overline{\nabla}\Phi,\frac{d\hat{y}_\tau}{d\tau}\Big|_{\tau=0}\Big\rangle+o(\tau),\tag{30}$$

where

$$\nu^T=\nu-\langle N,\nu\rangle N.$$

We have

$$\langle\nabla^\Sigma\zeta+AT,\nu^T\rangle=\alpha_0\langle\nabla v,\nu^T\rangle+\alpha_1\langle\nabla^\Sigma d_\Gamma,\nu^T\rangle+\langle\nabla^\Sigma\langle X,N\rangle,\nu^T\rangle+\langle AT,\nu^T\rangle.$$

We compute

$$\begin{aligned}\langle\nabla^\Sigma\langle X,N\rangle,\nu^T\rangle &= \alpha_0(v\langle N,\nu\rangle-d_\Gamma\langle N,\nabla v\rangle)\langle\overline{\nabla}\Phi,\nu^T\rangle\\ &\quad+\alpha_0\Phi\big(\langle\nabla v,\nu^T\rangle\langle N,\nu\rangle+v(\langle\overline{\nabla}_{\nu^T}N,\nu\rangle+\langle N,\overline{\nabla}_{\nu^T}\nu\rangle)\\ &\qquad-\langle\nabla d_\Gamma,\nu^T\rangle\langle N,\nabla v\rangle-d_\Gamma(\langle\overline{\nabla}_{\nu^T}N,\nabla v\rangle+\langle N,\overline{\nabla}_{\nu^T}\nabla v\rangle)\big).\end{aligned}$$

Hence we have at y that

$$\begin{aligned}\langle\nabla^\Sigma\langle X,N\rangle,\nu^T\rangle &= \alpha_0(v\Phi-d_\Gamma\langle N,\nabla v\rangle)\langle\overline{\nabla}\Phi,\nu^T\rangle\\ &\quad+\alpha_0\Phi\big(\langle\nabla v,\nu^T\rangle\Phi+v(-\langle A\nu^T,\nu^T\rangle+\langle N,\overline{\nabla}_\nu\nu\rangle\\ &\qquad-\langle N,\nu\rangle\langle N,\overline{\nabla}_N\nu\rangle)-\langle\nu,\nu^T\rangle\langle N,\nabla v\rangle\\ &\qquad\quad-d_\Gamma(-\langle A\nu^T,\nabla v\rangle+\langle N,\overline{\nabla}_\nu\nabla v\rangle-\langle N,\nu\rangle\langle N,\overline{\nabla}_N\nabla v\rangle)\big).\end{aligned}$$

Therefore we have

$$\begin{aligned}\langle\nabla^\Sigma\langle X,N\rangle,\nu^T\rangle &= \alpha_0(v\Phi-d_\Gamma\langle N,\nabla v\rangle)\langle\overline{\nabla}\Phi,\nu^T\rangle\\ &\quad+\alpha_0\Phi\big(\langle\nabla v,\nu^T\rangle\Phi-v(\langle A\nu^T,\nu^T\rangle+\langle N,\nu\rangle\langle N,\overline{\nabla}_N\nu\rangle)\\ &\qquad-\langle\nu,\nu^T\rangle\langle N,\nabla v\rangle\\ &\qquad\quad+d_\Gamma(\langle A\nu^T,\nabla v\rangle-\langle N,\overline{\nabla}_\nu\nabla v\rangle+\langle N,\nu\rangle\langle N,\overline{\nabla}_N\nabla v\rangle)\big).\end{aligned}$$

It follows that

$$
\begin{aligned}
\langle \nabla^\Sigma \zeta + AT, \nu^T\rangle = \langle AT, \nu^T\rangle &+ \alpha_0\langle \nabla v, \nu^T\rangle + \alpha_1\langle \nu, \nu^T\rangle \\
&+ \alpha_0(v\Phi - d_\Gamma\langle N, \nabla v\rangle)\langle \overline{\nabla}\Phi, \nu^T\rangle \\
&\quad + \alpha_0\Phi\big(\langle \nabla v, \nu^T\rangle\Phi - v(\langle A\nu^T, \nu^T\rangle + \langle N, \nu\rangle\langle N, \overline{\nabla}_N \nu\rangle) \\
&\qquad - \langle \nu, \nu^T\rangle\langle N, \nabla v\rangle \\
&\qquad\quad + d_\Gamma(\langle A\nu^T, \nabla v\rangle - \langle N, \overline{\nabla}_\nu \nabla v\rangle + \langle N, \nu\rangle\langle N, \overline{\nabla}_N \nabla v\rangle)\big).
\end{aligned}
$$

However,

$$
\langle AT, \nu^T\rangle = \langle A\nu^T, X\rangle = \alpha_0\Phi v\langle A\nu^T, \nu^T\rangle - \alpha_0\Phi d_\Gamma\langle A\nu^T, \nabla v\rangle.
$$

Hence we have

$$
\begin{aligned}
\langle \nabla^\Sigma \zeta + AT, \nu^T\rangle = \alpha_0\langle \nabla v, \nu^T\rangle + \alpha_1\langle \nu, \nu^T\rangle &+ \alpha_0(v\Phi - d_\Gamma\langle N, \nabla v\rangle)\langle \overline{\nabla}\Phi, \nu^T\rangle \\
+ \alpha_0\Phi\big(\langle \nabla v, \nu^T\rangle\Phi &- v\Phi\langle N, \overline{\nabla}_N \nu\rangle - \langle \nu, \nu^T\rangle\langle N, \nabla v\rangle \\
- d_\Gamma(\langle N, \overline{\nabla}_\nu \nabla v\rangle &- \langle N, \nu\rangle\langle N, \overline{\nabla}_N \nabla v\rangle)\big).
\end{aligned}
$$

Since $d_\Gamma(y) = 0$, we have

$$
\begin{aligned}
\langle \nabla^\Sigma \zeta + AT, \nu^T\rangle &= \alpha_0\langle \nabla v, \nu^T\rangle + \alpha_1\langle \nu, \nu^T\rangle + \alpha_0 v\Phi\langle \overline{\nabla}\Phi, \nu^T\rangle \\
&\quad + \alpha_0\Phi\big(\langle \nabla v, \nu^T\rangle\Phi - v\Phi\langle N, \overline{\nabla}_N \nu\rangle - \langle \nu, \nu^T\rangle\langle N, \nabla v\rangle\big).
\end{aligned}
$$

Rearranging terms we obtain

$$
\begin{aligned}
\langle \nabla^\Sigma \zeta + AT, \nu^T\rangle &= \alpha_1(1 - \langle N, \nu\rangle^2) + \alpha_0\langle \nabla v, \nu^T\rangle(1 + \Phi^2) + \alpha_0 v\Phi\langle \overline{\nabla}\Phi, \nu^T\rangle \\
&\quad - \alpha_0\Phi\big(v\Phi\langle N, \overline{\nabla}_N \nu\rangle + (1 - \langle N, \nu\rangle^2)\langle N, \nabla v\rangle\big).
\end{aligned}
$$

Therefore there exists a constant $C = C(\Phi, K, \Omega, R)$ such that

$$
|\langle \nabla^\Sigma \zeta + AT, \nu^T\rangle| \le \alpha_1(1 - \Phi^2) + C\alpha_0. \tag{31}
$$

Since $d_\Gamma(y) = 0$, it holds that

$$
|\langle A_K Z, N\rangle| = \|A_K\|\|Z\| \le \|A_K\|(\eta + \|X\|) \le 4R^2\alpha_0\|A_K\|(1 + \Phi),
$$

from which we conclude that

$$
\left|\left\langle \overline{\nabla}\langle N_\tau, \nu\rangle, \frac{dy_\tau}{d\tau}\Big|_{\tau=0}\right\rangle\right| \le \alpha_1(1 - \Phi^2) + \widetilde{C}\alpha_0, \tag{32}
$$

for some constant $\widetilde{C}(C_\Phi, K, \Omega, R) > 0$.

Now we suppose that $x \in \overline{\partial\Omega' \cap \Omega}$. In this case we have $v(x) = 0$. Then $\eta = \alpha_1 d_\Gamma$ and

$$
X = -\alpha_0\Phi d_\Gamma \nabla v
$$

at x. Thus

$$\zeta = \eta + \langle X, N\rangle = \alpha_1 d_\Gamma + 2\alpha_0 \Phi d d_\Gamma \langle \nabla d, N\rangle.$$

Moreover, we have

$$W(x) \le \frac{C}{d_\Gamma(x)}$$

(see Remark 6). It follows that

$$\zeta W \le C(\alpha_1 + 2\alpha_0 \Phi d\langle \nabla d, N\rangle) \le C(\alpha_1 + 4R\alpha_0\Phi). \tag{33}$$

We conclude that

$$W(x) \le C(C_\Phi, K, \Omega, R). \tag{34}$$

Now we consider the case when $x \in \Omega \cap \Omega'$. In this case we have

$$\begin{aligned}\Delta_\Sigma \zeta = {}& \alpha_0 \Delta_\Sigma v + \alpha_1 \Delta_\Sigma d_\Gamma + \alpha_0 \Delta_\Sigma \Phi(v\langle N, \nu\rangle - d_\Gamma \langle N, \nabla v\rangle) \\ & + \alpha_0 \Phi\big(\Delta_\Sigma v \langle N, \nu\rangle + v\Delta_\Sigma \langle N, \nu\rangle + 2\langle \nabla^\Sigma v, \nabla^\Sigma \langle N, \nu\rangle\rangle - \Delta_\Sigma d_\Gamma \langle N, \nabla v\rangle \\ & - d_\Gamma \Delta_\Sigma \langle N, \nabla v\rangle - 2\langle \nabla^\Sigma d_\Gamma, \nabla^\Sigma \langle N, \nabla v\rangle\rangle\big) + 2\alpha_0\big\langle \nabla^\Sigma \Phi, \nabla^\Sigma v\langle N, \nu\rangle \\ & + v\nabla^\Sigma \langle N, \nu\rangle - \nabla^\Sigma d_\Gamma \langle N, \nabla v\rangle - d_\Gamma \nabla^\Sigma \langle N, \nabla v\rangle\big\rangle.\end{aligned}$$

Notice that given an arbitrary vector field U along Σ, we have

$$\langle \nabla^\Sigma \langle N, U\rangle, V\rangle = -\langle AU^T, V\rangle + \langle N, \overline{\nabla}_V U\rangle$$

for any $V \in \Gamma(T\Sigma)$. Here, U^T denotes the tangential component of U. Hence using Codazzi's equation we obtain

$$\Delta_\Sigma \langle N, U\rangle \le \langle \overline{\nabla}(nH), U^T\rangle + \mathrm{Ric}_M(U^T, N) + C\|A\|,$$

for a constant C depending on $\overline{\nabla} U$ and $\overline{\nabla}^2 U$. Hence using (3) we conclude that

$$\Delta_\Sigma \langle N, U\rangle \le \langle \overline{\nabla}\Psi, U^T\rangle + \widetilde{C}\|A\|, \tag{35}$$

where $\widetilde{C}$ is a positive constant depending on $\overline{\nabla} U$, $\overline{\nabla}^2 U$ and Ric_M.

We also have

$$\begin{aligned}\Delta_\Sigma d_\Gamma &= \Delta_P d_\Gamma + \gamma\langle \overline{\nabla}_Y \overline{\nabla} d, Y\rangle - \langle \overline{\nabla}_N \overline{\nabla} d_\Gamma, N\rangle + nH\langle \overline{\nabla} d_\Gamma, N\rangle \\ &\le C_0 \Psi + C_1,\end{aligned}$$

where C_0 and C_1 are positive constants depending on the second fundamental form of the Killing cylinders over the equidistant sets $d_\Gamma = \delta$ for small values of δ. Similar estimates also hold for $\Delta_\Sigma d$ and then for $\Delta_\Sigma v$.

We conclude that

$$\Delta_\Sigma \zeta \ge -\widetilde{C}_0 - \widetilde{C}_1 \|A\|, \tag{36}$$

where $\widetilde{C}_0$ and $\widetilde{C}_1$ are positive constants depending only on Ω, K, Ric_M, and $|\Phi| + \|\nabla\Phi\| + \|\nabla^2\Phi\|$.

Now proceeding similarly as in the proof of Proposition 5, we observe that Lemma 1(iii) and the comparison principle yield

$$
\begin{aligned}
H_0(\hat{y}_\tau) - H_0(y) &\geq \frac{dH_\tau}{d\tau}\bigg|_{\tau=0} \tau + o(\tau) \\
&= (\Delta_\Sigma \zeta + \|A\|^2\zeta + \mathrm{Ric}_M(N, N)\zeta)\tau + \tau\langle \bar{\nabla}\Psi, T\rangle + o(\tau).
\end{aligned}
$$

However,

$$H_0(\hat{y}_\tau) - H_0(y) = \langle \bar{\nabla}\Psi|_y, \xi'(0)\rangle\tau + o(\tau).$$

Using (18) we have

$$\langle \bar{\nabla}\Psi, \xi'(0)\rangle = \langle \bar{\nabla}\Psi, Z - \zeta WY\rangle = \langle \bar{\nabla}\Psi, Z\rangle - \zeta W\frac{\partial\Psi}{\partial s}.$$

We conclude that

$$-\zeta W\frac{\partial\Psi}{\partial s}\tau + \zeta\langle \bar{\nabla}\Psi, N\rangle\tau + o(\tau) \geq (\Delta_\Sigma\zeta + \|A\|^2\zeta + \mathrm{Ric}_M(N, N)\zeta)\tau + o(\tau).$$

Suppose that

$$W > \frac{C + \|\bar{\nabla}\Psi\|}{\beta}, \tag{37}$$

for a constant $C > 0$ as in (22). Hence we have

$$(\Delta_\Sigma\zeta + |A|^2\zeta + \mathrm{Ric}_M(N, N)\zeta)\tau + C\zeta\tau \leq o(\tau).$$

We conclude that

$$-C_0 - C_1|A| + C_2\|A\|^2 + C \leq \frac{o(\tau)}{\tau},$$

a contradiction. It follows from this contradiction that

$$W(x) \leq \frac{C + \|\bar{\nabla}\Psi\|}{\beta}. \tag{38}$$

Now, proceeding as in the end of the proof of Proposition 5, we use the estimate for $W(x)$ in each one of the three cases for obtaining a estimate for W in Ω'. $\square$

4. Proof of Theorem 1

We use the classical continuity method for proving Theorem 1. For details, we refer the reader to [Gerhardt 1976] and [Ladyzhenskaya and Uraltseva 1964]. For any

$\tau \in [0, 1]$, we consider the Neumann boundary problem $\mathcal{N}_\tau$ of finding $u \in C^{3,\alpha}(\overline{\Omega})$ such that

$$\mathcal{F}[\tau, x, u, \nabla u, \nabla^2 u] = 0, \tag{39}$$

$$\left\langle \frac{\nabla u}{W}, \nu \right\rangle + \tau \Phi = 0, \tag{40}$$

where $\mathcal{F}$ is the quasilinear elliptic operator defined by

$$\mathcal{F}[x, u, \nabla u, \nabla^2 u] = \operatorname{div}\left(\frac{\nabla u}{W}\right) - \left\langle \frac{\nabla \gamma}{2\gamma}, \frac{\nabla u}{W} \right\rangle - \tau \Psi. \tag{41}$$

Since the coefficients of the first and second order terms do not depend on u, it follows that

$$\frac{\partial \mathcal{F}}{\partial u} = -\tau \frac{\partial \Psi}{\partial u} \leq -\tau \beta < 0. \tag{42}$$

We define $\mathcal{I} \subset [0, 1]$ as the subset of values of $\tau \in [0, 1]$ for which the Neumann boundary problem $\mathcal{N}_\tau$ has a solution. Since $u = 0$ is a solution for $\mathcal{N}_0$, it follows that $\mathcal{I} \neq \varnothing$. Moreover, the implicit function theorem (see [Gilbarg and Trudinger 2001, Chapter 17]) implies that $\mathcal{I}$ is open in view of (42). Finally, the height and gradient a priori estimates we obtained in Sections 2 and 3 are independent of $\tau \in [0, 1]$. This implies that (3) is uniformly elliptic. Moreover, we may ensure the existence of some $\alpha_0 \in (0, 1)$ for which there exists a constant $C > 0$ independent of τ such that

$$|u_\tau|_{1,\alpha_0,\overline{\Omega}} \leq C.$$

Redefine $\alpha = \alpha_0$. Thus, this fact, Schauder elliptic estimates and the compactness of $C^{3,\alpha_0}(\overline{\Omega})$ in $C^3(\overline{\Omega})$ imply that $\mathcal{I}$ is closed. It follows that $\mathcal{I} = [0, 1]$.

The uniqueness follows from the comparison principle for elliptic PDEs. We point out that a more general uniqueness statement — comparing a nonparametric solution with a general hypersurface with the same mean curvature and contact angle at corresponding points — is also valid. It is a consequence of a flux formula coming from the existence of a Killing vector field in M. We refer the reader to [Dajczer et al. 2008] for further details.

This finishes the proof of the Theorem 1.

References

[Calle and Shahriyari 2011] M. Calle and L. Shahriyari, "Existence of a capillary surface with prescribed contact angle in $M \times \mathbb{R}$", preprint, 2011. arXiv 1012.5490v2

[Concus and Finn 1974] P. Concus and R. Finn, "On capillary free surfaces in a gravitational field", *Acta Math.* **132** (1974), 207–223. MR 58 #32327c Zbl 0382.76005

[Dajczer and de Lira 2012] M. Dajczer and J. H. de Lira, "Conformal Killing graphs with prescribed mean curvature", *J. Geom. Anal.* **22**:3 (2012), 780–799. MR 2927678 Zbl 1261.53058

[Dajczer et al. 2008] M. Dajczer, P. A. Hinojosa, and J. H. de Lira, "Killing graphs with prescribed mean curvature", *Calc. Var. Partial Differential Equations* **33**:2 (2008), 231–248. MR 2009m:53154 Zbl 1152.53046

[Gerhardt 1976] C. Gerhardt, "Global regularity of the solutions to the capillarity problem", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:1 (1976), 157–175. MR 58 #29199 Zbl 0338.49008

[Gerhardt 2006] C. Gerhardt, *Curvature problems*, Series in Geometry and Topology **39**, International Press, Somerville, MA, 2006. MR 2007j:53001 Zbl 1131.53001

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, Berlin, 2001. MR 2001k:35004 Zbl 1042.35002

[Korevaar 1988] N. J. Korevaar, "Maximum principle gradient estimates for the capillary problem", *Comm. Partial Differential Equations* **13**:1 (1988), 1–31. MR 89d:35061 Zbl 0659.35042

[Ladyzhenskaya and Uraltseva 1964] O. A. Ladyzhenskaya and N. N. Uraltseva, *Линейные и квазилинейные уравнения эллиптического типа*, Nauka, Moscow, 1964. Translated as *Linear and quasilinear elliptic equations*, Mathematics in Science and Engineering **46**, Academic Press, New York, 1968. MR 35 #1955 Zbl 0143.33602

[Simon and Spruck 1976] L. Simon and J. Spruck, "Existence and regularity of a capillary surface with prescribed contact angle", *Arch. Rational Mech. Anal.* **61**:1 (1976), 19–34. MR 58 #7339 Zbl 0361.35014

[Uraltseva 1973] N. N. Uraltseva, "Nonlinear boundary value problems for equations of minimal-surface type", *Proc. Steklov Inst. Math.* **116** (1973), 227–237. Zbl 0269.35039

Received May 2, 2013. Revised November 26, 2013.

JORGE H. LIRA
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DO PICI, BLOCO 914
FORTALEZA, CEARÁ
60455-760
BRAZIL
jorge.h.lira@gmail.com

GABRIELA A. WANDERLEY
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DO PICI, BLOCO 914
FORTALEZA, CEARÁ
60455-760
BRAZIL
gwanderley@mat.ufc.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.425

A COUNTEREXAMPLE TO THE SIMPLE LOOP CONJECTURE FOR PSL(2, ℝ)

KATHRYN MANN

In this note, we give an explicit counterexample to the simple loop conjecture for representations of surface groups into $\mathrm{PSL}(2,\mathbb{R})$. Specifically, we use a construction of DeBlois and Kent to show that for any orientable surface with negative Euler characteristic and genus at least 1, there are uncountably many nonconjugate, noninjective homomorphisms of its fundamental group into $\mathrm{PSL}(2,\mathbb{R})$ that kill no simple closed curve (nor any power of a simple closed curve). This result is not new — work of Louder and Calegari for representations of surface groups into $\mathrm{SL}(2,\mathbb{C})$ applies to the $\mathrm{PSL}(2,\mathbb{R})$ case, but our approach here is explicit and elementary.

1. Introduction

The simple loop conjecture, proved by Gabai [1985], states that any noninjective homomorphism from a closed surface group to another closed surface group has an element represented by a simple closed curve in the kernel. It has been conjectured that the result still holds if the target is replaced by the fundamental group of an orientable 3-manifold (see the problem list in [Kirby 1997]). Although special cases have been proved (e.g., [Hass 1987; Rubinstein and Wang 1998]), the general hyperbolic case is still open.

Minsky [2000] asked whether the conjecture holds if the target group is instead $\mathrm{SL}(2,\mathbb{C})$. This was answered in the negative with the following theorem.

Theorem 1.1 [Cooper and Manning 2011]. *Let Σ be a closed orientable surface of genus $g \geq 4$. Then there is a homomorphism $\rho : \pi_1(\Sigma) \to \mathrm{SL}(2,\mathbb{C})$ such that*:

(1) *ρ is not injective.*

(2) *If $\rho(\alpha) = \pm I$, then α is not represented by a simple closed curve.*

(3) *If $\rho(\alpha)$ has finite order, then $\rho(\alpha) = I$.*

The third condition implies in particular that no power *of a simple closed curve lies in the kernel.*

MSC2010: primary 57M05; secondary 57N16.
Keywords: simple loop conjecture, surface group, character variety, representation, representation space, $\mathrm{PSL}(2,\mathbb{R})$.

Inspired by this, we ask whether a similar result holds for $\mathrm{PSL}(2,\mathbb{R})$, this being an intermediate case between Gabai's result for surface groups and Cooper and Manning's for $\mathrm{SL}(2,\mathbb{C})$. Techniques of Cooper and Manning's proof do not seem to carry over directly to the $\mathrm{PSL}(2,\mathbb{R})$ case — their work involves both a dimension count on the $\mathrm{SL}(2,\mathbb{C})$ character variety and a proof showing that a specific subvariety is irreducible and smooth on a dense subset, and complex varieties and their real points generally behave quite differently. However, we will show here with different methods that an analog to Theorem 1.1 does hold for $\mathrm{PSL}(2,\mathbb{R})$.

While this note was in progress, we learned of work of Louder and Calegari (independently in [Louder 2011] and [Calegari 2013]) that can also be applied to answer our question in the affirmative. Louder shows the simple loop conjecture is false for representations into limit groups, and Calegari gives a practical way of verifying no simple closed curves lie in the kernel of a noninjective representation using stable commutator length and the Gromov norm.

The difference here is that our construction is entirely elementary. We use an explicit representation found in [DeBlois and Kent 2006] (which uses work from [Goldman 1988] and [Shalen 1979]), and we verify by elementary means that this representation is noninjective and kills no simple closed curve. Our end result parallels that of Cooper and Manning but also include surfaces with boundary and all genera at least 1:

Theorem 1.2. *Let Σ be an orientable surface of negative Euler characteristic and of genus $g \geq 1$, possibly with boundary. Then there is a homomorphism $\rho : \pi_1(\Sigma) \to \mathrm{SL}(2,\mathbb{R})$ such that*:

(1) *ρ is not injective.*

(2) *If $\rho(\alpha) = \pm I$, then α is not represented by a simple closed curve.*

(3) *In fact, if α is represented by a simple closed curve, then $\rho(\alpha^k) \neq I$ for any $k \in \mathbb{Z}$.*

Moreover, there are uncountably many nonconjugate representations satisfying (1) *through* (3).

In the case of a nonorientable surface, the appropriate target group is $\mathrm{PGL}(2,\mathbb{R})$, as the fundamental group of a nonorientable hyperbolic surface can be represented as a lattice in $\mathrm{PGL}(2,\mathbb{R})$. This again gives an intermediate case between the simple loop conjecture for representations into surface groups and into $\mathrm{PSL}(2,\mathbb{C})$. We have the following direct generalization of Theorem 1.2, with essentially the same proof.

Theorem 1.3. *Let Σ be a nonorientable surface of negative Euler characteristic and of nonorientable genus $g \geq 2$ that is not the punctured Klein bottle nor the closed nonorientable genus-*3 *surface. Then there are uncountably many representations $\rho : \pi_1(\Sigma) \to \mathrm{PGL}(2,\mathbb{R})$ satisfying conditions* (1) *through* (3) *of Theorem 1.2.*

See Section 3 for a comment on the exceptional cases of the punctured Klein bottle and the closed, nonorientable genus-3 surface.

2. Proof of Theorem 1.2

We describe a family of (noninjective) representations constructed in [DeBlois and Kent 2006] based on a construction from [Goldman 1988]. We will then show that this family contains infinitely many nonconjugate representations with no simple closed curve in the kernel.

Let Σ be an orientable surface of genus $g \geq 1$ and negative Euler characteristic, possibly with boundary. Assume for the moment that Σ is not the once-punctured torus — Theorem 1.2 for this case will follow easily later on.

Let $c \subset \Sigma$ be a simple closed curve separating Σ into a genus-1 subsurface with single boundary component c and a genus-$(g-1)$ subsurface with one or more boundary components. Let Σ_A denote the genus-$(g-1)$ subsurface and Σ_B the genus-1 subsurface. Finally, we let $A = \pi_1(\Sigma_A)$ and $B = \pi_1(\Sigma_B)$, so that $\pi_1(\Sigma) = A *_C B$, where C is the infinite cyclic subgroup generated by the element $[c]$ represented by the curve c. We assume that the basepoint for $\pi_1(\Sigma)$ lies on c.

Let $x \in B$ and $y \in B$ be generators such that $B = \langle x, y \rangle$, and that the curve c represents the commutator $[x, y]$. See Figure 1.

Fix α and β in $\mathbb{R} \setminus \{0, \pm 1\}$, and following [DeBlois and Kent 2006] define $\phi_B : B \to \mathrm{SL}(2, \mathbb{R})$ by

$$\phi_B(x) = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \quad \phi_B(y) = \begin{pmatrix} \beta & 1 \\ 0 & \beta^{-1} \end{pmatrix}.$$

We have then

$$\phi_B([x, y]) = \begin{pmatrix} 1 & \beta(\alpha^2 - 1) \\ 0 & 1 \end{pmatrix},$$

so $\phi_B([x, y])$ is invariant under conjugation by the matrix $\lambda_t := \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$. Projecting these matrices to PSL(2, $\mathbb{R}$) gives a representation $B \to \mathrm{PSL}(2, \mathbb{R})$ that is upper

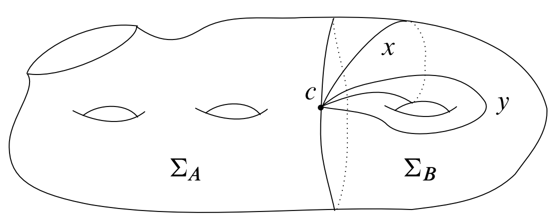


Figure 1. Decomposition of Σ and curves representing generators x and y for B.

triangular, hence solvable, and therefore noninjective. Abusing notation, we let ϕ_B denote this representation.

Now let $\phi_A : A \to \mathrm{PSL}(2,\mathbb{R})$ be Fuchsian such and that the image of the boundary curve c under ϕ_A agrees with $\phi_B([x,y])$. That such a representation exists is standard — Σ_A has negative Euler characteristic and therefore admits a complete hyperbolic structure. The image of $[c]$ under the corresponding Fuchsian representation is a parabolic element of $\mathrm{PSL}(2,\mathbb{R})$, so after conjugation we may assume that it is equal to $\phi_B([x,y])$, since $\beta(\alpha^2-1) \neq 0$.

Finally, we combine ϕ_A with conjugates of ϕ_B to get a one-parameter family of representations $\phi_t : \pi_1(\Sigma) \to \mathrm{PSL}(2,\mathbb{R})$ as follows. For $t \in \mathbb{R}$ and $g \in \pi_1(\Sigma) = A *_C B$, let

$$\phi_t(g) = \begin{cases} \phi_A(g) & \text{if } g \in A, \\ \lambda_t \circ \phi_B(g) \circ (\lambda_t)^{-1} & \text{if } g \in B. \end{cases}$$

This representation is well-defined because $\phi_B([x,y]) = \phi_A([x,y])$, and is invariant under conjugation by λ_t.

Our next goal is to show that for appropriate choice of α, β, and t, the representation ϕ_t satisfies the criteria in Theorem 1.2. The main difficulty will be checking that no element representing a simple closed curve is of finite order. To do so, we employ a stronger form of Lemma 2 from [DeBlois and Kent 2006]. This trick originally comes from the proof of Proposition 1.3 in [Shalen 1979].

Lemma 2.1. *Suppose $w \in A *_C B$ is a word of the form $w = a_1 b_1 a_2 b_2 \cdots a_l b_l$, with $a_i \in A$ and $b_i \in B$ for $1 \le i \le l$. Assume that for each i, the matrix $\phi_0(a_i)$ has a nonzero $(2,1)$-entry and $\phi_0(b_i)$ is hyperbolic. If t is transcendental over the entry field of $\phi_0(A *_C B)$, then $\phi_t(w)$ is not of finite order.*

By the *entry field* of a group Γ of matrices, we mean the field generated over $\mathbb{Q}$ by the collection of all entries of matrices in Γ.

Remark 2.2. Lemma 2 of [DeBlois and Kent 2006] is a proof that $\phi_t(w)$ is not the *identity*, under the assumptions of Lemma 2.1. We use some of their work in our proof.

Proof of Lemma 2.1. DeBlois and Kent show by a straightforward induction (we leave it as an exercise) that under the hypotheses of Lemma 2.1, the entries of $\phi_t(w)$ are polynomials in t such that the degree of the $(2,2)$-entry is l, the degree of the $(1,2)$-entry is at most l, and the other entries have degree at most $l-1$. Now, suppose that $\phi_t(w)$ is finite order. Then it is conjugate to a matrix of the form $\left(\begin{smallmatrix} u & v \\ -v & u \end{smallmatrix}\right)$, where $u = \cos\theta$ and $v = \sin\theta$ for θ a rational multiple of π. In particular, it follows from the de Moivre formula for sine and cosine that u and v are algebraic.

Now suppose that the matrix conjugating $\phi_t(w)$ to $\left(\begin{smallmatrix} u & v \\ -v & u \end{smallmatrix}\right)$ has entries a_{ij}. Then we have

$$\phi_t(w) = \begin{pmatrix} u - (a_{12}a_{22} - a_{11}a_{21})v & (a_{12}^2 a_{11}^2)v \\ -(a_{22}^2 a_{21}^2)v & u + (a_{12}a_{22} + a_{11}a_{21})v \end{pmatrix}.$$

Looking at the $(2,2)$-entry we see that $a_{12}a_{22} + a_{11}a_{21}$ must be a polynomial in t of degree l. But this means that the $(1,1)$-entry is also a polynomial in t of degree l, contradicting DeBlois and Kent's calculation. This proves the lemma.

□

To complete our construction, choose any $t \in \mathbb{R}$ that is transcendental over the entry field of $\phi_0(A *_C B)$. We want to show that no power of an element representing a simple closed curve lies in the kernel of ϕ_t. To this end, consider any word w in $A *_C B$ that has a simple closed curve as a representative. There are three cases to check.

Case i: w is a word in A alone. In this case $\phi_t(w)$ is not finite order, since $\phi_t(A)$ is Fuchsian and therefore injective.

Case ii: w is a word in B alone. Theorem 5.1 of [Birman and Series 1984] states that w can be represented by a simple closed curve only if it has one of the following forms after cyclic reduction:

1. $w = x^{\pm 1}$ or $w = y^{\pm 1}$.
2. $w = [x^{\pm 1}, y^{\pm 1}]$.
3. Up to replacing x with x^{-1}, y with y^{-1}, and interchanging x and y, there is some $n \in \mathbb{Z}^+$ such that $w = x^{n_1} y x^{n_2} y \cdots x^{n_s} y$, where $n_i \in \{n, n+1\}$.

The heuristic for Case 3 of the Birman–Series theorem is shown in Figure 2 — if w is represented by a simple closed curve and terminates with $x^{n_s} y$, this forces the

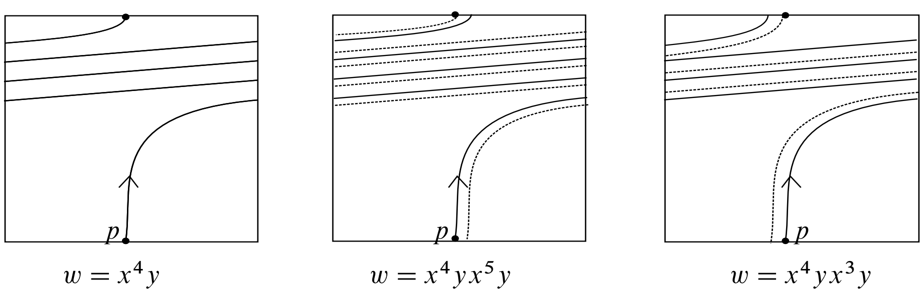


Figure 2. Simple closed curves on the once punctured torus. Assume the puncture is at the vertex, x is represented by a horizontal loop oriented from left to right, and y is a vertical loop oriented from bottom to top.

rest of the curve representing w to wind around the punctured torus in a set pattern. The figure shows the behavior for $n_s = 4$.

By construction, no word of type 1, 2 or 3 above is finite order, provided that $\alpha^s \beta^k \neq 1$ for any integers s and k other than zero — indeed, we only need to check words of type 3, and these necessarily have trace equal to $\alpha^s \beta^k + \alpha^{-s} \beta^{-k}$ for some $s, k \neq 0$. Since cyclic reduction corresponds to conjugation, no word in B has finite order image.

Note also that, in particular, under the condition that $\alpha^s \beta^k \neq 1$ for $s, k \neq 0$, all type 3 words are hyperbolic. We will use this fact again later on.

Case iii: general case. If w is a word including both A and B, we claim that it can be written in a form where Lemma 2.1 applies. To write it this way, take a simple curve γ on Σ that represents w and has a minimal number of (geometric) intersections with c. We can write γ as a concatenation of simple arcs $\gamma = \gamma_1 \delta_1 \gamma_2 \delta_2 \cdots \gamma_n \delta_n$, with $\gamma_i \subset \Sigma_A$ and $\delta_i \subset \Sigma_B$. Since we chose γ to have a minimal number of intersections with c, no arc γ_i (or δ_i) is homotopic in Σ_A (respectively in Σ_B) to a segment of c — if it were, we could apply an isotopy of Σ supported in a neighborhood of the disc bounded by the arc and the segment of c to push the arc across c and reduce the total number of intersections.

Now choose a proper segment c' of c that contains the basepoint p and all endpoints of all γ_i and δ_i, and close each of the arcs γ_i and δ_i into a simple loop by attaching a segment of c'. If $a_i \in A$ and $b_i \in B$ are represented by the loops γ_i and δ_i, then $a_1 b_1 a_2 b_2 \cdots a_n b_n = w$ in $\pi_1(\Sigma)$.

Since no arc γ_i or δ_i was homotopic to a segment of c, no a_i or b_i is represented by a power of $[c]$ in $\pi_1(\Sigma)$. We claim that in this case $a_1 b_1 a_2 b_2 \cdots a_n b_n$ satisfies the hypotheses of Lemma 2.1. Indeed, since ϕ_A is Fuchsian, the only elements with a nonzero $(2,1)$-entry are powers of $[c]$, and the Birman–Series classification of simple closed curves on Σ_b implies that the only simple closed curves which are *not* hyperbolic represent $[c]$ or $[c]^{-1}$.

It remains only to remark that the representation ϕ_t is noninjective and that, by choosing appropriate parameters, we can produce uncountably many nonconjugate representations. Noninjectivity follows immediately since $\phi_t(B)$ is solvable, so the restriction of ϕ_t to B is noninjective. Now, for any fixed α and β (satisfying the requirement that $\alpha^s \beta^k \neq 1$ for all integers s, k), varying t among transcendentals over the entry field of $\phi_0(A *_C B)$ produces uncountably many nonconjugate representations that are all noninjective, but have no power of a simple closed curve in the kernel. This concludes the proof of Theorem 1.2, assuming that the surface was not the punctured torus.

The punctured torus case is now immediate: any representation of the form of ϕ_B where $\alpha^s \beta^k \neq 1$ for any integers s and k is noninjective and our work above shows that no element represented by a simple closed curve has finite order. Fixing α and varying β produces uncountably many nonconjugate representations.

3. Nonorientable surfaces

Recall that the *genus* of a nonorientable surface Σ is defined to be the number of $\mathbb{R}\mathrm{P}^2$-summands in a decomposition of the surface as $\Sigma = \mathbb{R}\mathrm{P}^2 \# \mathbb{R}\mathrm{P}^2 \# \cdots \# \mathbb{R}\mathrm{P}^2$. A closed, nonorientable genus-g surface has Euler characteristic $\chi = 2 - g$.

Let Σ be a nonorientable surface of negative Euler characteristic and nonorientable genus $g \geq 2$ that is not the punctured Klein bottle nor the closed nonorientable genus-3 surface. The same strategy as in the orientable case can then be used to produce uncountably many noninjective representations $\pi_1(\Sigma) \to \mathrm{PGL}(2, \mathbb{R})$ such that no power of a simple closed curve lies in the kernel. In detail, our assumptions on Σ imply that we may decompose Σ along a (2-sided) curve c into a genus-1 orientable surface Σ_B with one boundary component and a nonorientable surface Σ_A of negative Euler characteristic.

We define ϕ_B exactly as in the orientable case, but now consider the matrices as elements of $\mathrm{PGL}(2, \mathbb{R})$ rather than $\mathrm{PSL}(2, \mathbb{R})$. We let $\phi_A : \pi_1(\Sigma_A) \to \mathrm{PGL}(2, \mathbb{R})$ be a discrete, faithful representation such that $\phi_A([c]) = \phi_B([c])$. As in the case of the orientable surface, we may take this to be a representation corresponding to a complete hyperbolic structure on Σ. Define $\phi_t : \pi_1(\Sigma) \to \mathrm{PGL}(2, \mathbb{R})$ by "gluing together" ϕ_A with a conjugate of ϕ_B by λ_t exactly as in the orientable case. The proof now carries through verbatim, for none of the topological arguments that we used required orientability of Σ_A. We also reassure the reader (who may be unfamiliar with lattices in $\mathrm{PGL}(2, \mathbb{R})$) that powers of $\phi_A([c])$ are indeed the only elements of the image of ϕ_A with 0 as the $(2, 1)$-entry.

This strategy does not cover the case of the punctured Klein bottle, which cannot be decomposed with a T^2-summand, nor the closed nonorientable genus-3 surface, which decomposes as $T^2 \# \mathbb{R}\mathrm{P}^2$. It would be interesting to try to cover this case in a manner analogous to the punctured torus case of Theorem 1.2 by providing a classification of simple closed curves on these surfaces. Indeed (as the referee has pointed out) the punctured Klein bottle case is not too difficult. The closed, nonorientable genus-3 surface case appears to be more challenging.

Acknowledgements

I am grateful to the referee for bringing the nonorientable case to my attention, for suggesting improvements to Case iii of the main proof, and for providing additional context and references. I am also indebted to Benson Farb, who introduced me to the simple loop conjecture and much of the background material that went into this paper.

References

[Birman and Series 1984] J. S. Birman and C. Series, "An algorithm for simple curves on surfaces", *J. London Math. Soc.* (2) **29**:2 (1984), 331–342. MR 85m:57002 Zbl 0507.57006

[Calegari 2013] D. Calegari, "Certifying incompressibility of noninjective surfaces with scl", *Pacific J. Math.* **262**:2 (2013), 257–262. MR 3069061 Zbl 06189535

[Cooper and Manning 2011] D. Cooper and J. F. Manning, "Non-faithful representations of surface groups into $\mathrm{SL}(2,\mathbb{C})$ which kill no simple closed curve", preprint, 2011. To appear in *Geom. Dedicata*. arXiv 1104.4492v1

[DeBlois and Kent 2006] J. DeBlois and I. Kent, Richard P., "Surface groups are frequently faithful", *Duke Math. J.* **131**:2 (2006), 351–362. MR 2007b:57001 Zbl 1109.57002

[Gabai 1985] D. Gabai, "The simple loop conjecture", *J. Differential Geom.* **21**:1 (1985), 143–149. MR 86m:57013 Zbl 0556.57007

[Goldman 1988] W. M. Goldman, "Topological components of spaces of representations", *Invent. Math.* **93**:3 (1988), 557–607. MR 89m:57001 Zbl 0655.57019

[Hass 1987] J. Hass, "Minimal surfaces in manifolds with S^1 actions and the simple loop conjecture for Seifert fibered spaces", *Proc. Amer. Math. Soc.* **99**:2 (1987), 383–388. MR 88e:57015 Zbl 0627.57008

[Kirby 1997] R. Kirby, ed., "Problems in low-dimensional topology", pp. 35–473 in *Geometric topology* (Athens, GA, 1993), vol. 2, edited by W. H. Kazez, AMS/IP Stud. Adv. Math. **2**, Amer. Math. Soc., Providence, RI, 1997. MR 1470751 Zbl 0888.57014

[Louder 2011] L. Louder, "Simple loop conjecture for limit groups", preprint, 2011. To appear in Israel J. Math. arXiv 1106.1350v1

[Minsky 2000] Y. N. Minsky, "Short geodesics and end invariants", *Sūrikaisekikenkyūsho Kōkyūroku* 1153 (2000), 1–19. MR 1805224 Zbl 0968.57500 arXiv math/0006002

[Rubinstein and Wang 1998] J. H. Rubinstein and S. Wang, "π_1-injective surfaces in graph manifolds", *Comment. Math. Helv.* **73**:4 (1998), 499–515. MR 99h:57039 Zbl 0916.57001

[Shalen 1979] P. B. Shalen, "Linear representations of certain amalgamated products", *J. Pure Appl. Algebra* **15**:2 (1979), 187–197. MR 80e:20011 Zbl 0401.20024

Received November 27, 2012. Revised September 4, 2013.

KATHRYN MANN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CHICAGO
5734 SOUTH UNIVERSITY AVENUE
CHICAGO, IL 60637
UNITED STATES

mann@math.uchicago.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.433

TWISTED ALEXANDER POLYNOMIALS OF 2-BRIDGE KNOTS FOR PARABOLIC REPRESENTATIONS

TAKAYUKI MORIFUJI AND ANH T. TRAN

In this paper we show that the twisted Alexander polynomial associated to a parabolic representation determines fiberedness and genus of a wide class of 2-bridge knots. As a corollary we give an affirmative answer to a conjecture of Dunfield, Friedl and Jackson for infinitely many hyperbolic knots.

1. Introduction

The twisted Alexander polynomial was introduced by Lin [2001] for knots in the 3-sphere and by Wada [1994] for finitely presentable groups. It is a generalization of the classical Alexander polynomial and gives a powerful tool in low-dimensional topology. A theory of twisted Alexander polynomials has developed rapidly over the past ten years. One of the most important aspects is the determination of fiberedness [Friedl and Vidussi 2011b] and genus (the Thurston norm) [Friedl and Vidussi 2012] of knots by the collection of the twisted Alexander polynomials corresponding to all finite-dimensional representations. For literature on other applications and related topics, we refer to the survey paper [Friedl and Vidussi 2011a].

Let K be a knot in S^3 and G_K its knot group. Namely it is the fundamental group of the complement of K in S^3, $G_K = \pi_1(S^3 \backslash K)$. In this paper, we consider the twisted Alexander polynomial $\Delta_{K,\rho}(t) \in \mathbb{C}[t^{\pm 1}]$ associated to a parabolic representation $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$. A typical example is the holonomy representation $\rho_0 : G_K \to \mathrm{SL}_2(\mathbb{C})$ of a hyperbolic knot K, which is a lift of a discrete faithful representation $\bar{\rho}_0 : G_K \to \mathrm{PSL}_2(\mathbb{C}) \cong \mathrm{Isom}^+(\mathbb{H}^3)$ such that $\mathbb{H}^3/\bar{\rho}_0(G_K) \cong S^3 \backslash K$ where $\mathbb{H}^3$ denotes the upper half space model of hyperbolic 3-space (see [Thurston 1997]). Dunfield, Friedl and Jackson [Dunfield et al. 2012] numerically computed the twisted Alexander polynomial $\mathcal{T}_K(t) = \Delta_{K,\rho_0}(t)$, which is called the hyperbolic torsion polynomial, for all hyperbolic knots of 15 or fewer crossings. Based on these huge computations, they conjectured that the hyperbolic torsion polynomial determines the knot genus and, moreover, the knot is fibered if and only if $\mathcal{T}_K(t)$ is a monic polynomial. This conjecture is nice because it would imply the fiberedness and genus of a knot is determined by the twisted Alexander polynomial associated

MSC2010: 57M27.
Keywords: 2-bridge knot, parabolic representation, twisted Alexander polynomial.

to a single representation. However it is widely open except for the hyperbolic twist knots [Morifuji 2012].

The purpose of this paper is to show that the above conjecture is true for a wide class of 2-bridge knots. Since 2-bridge knots are alternating, their fiberedness and genus can be determined by the Alexander polynomial [Crowell 1959; Murasugi 1958a; 1958b]. However there seems to be no a priori reason that the same must be true for the hyperbolic torsion polynomial.

For a prime p and an integer a between 1 and $p-1$, we say that a is a primitive root modulo p if it is a generator of the cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$. Let $\mathcal{P}_2$ be the set of all odd primes p such that 2 is a primitive root modulo p. Note that all primes $p=2q+1$ such that q is a prime $\equiv 1 \pmod 4$ are contained in $\mathcal{P}_2$; see, for example, [LeVeque 1977, Theorem 5.6].

Theorem 1.1. *Let K be the knot $J(k,2n)$ as in Figure 1, where $k>0$ and $n\in\mathbb{Z}$. For all hyperbolic knots K, the hyperbolic torsion polynomial $\mathcal{T}_K(t)$ determines the genus of K. Moreover for $k=2m+1$, $k=2$ (twist knot), or $k=2m$ and $|4mn-1|\in\mathcal{P}_2$, the knot $J(k,2n)$ is fibered if and only if $\mathcal{T}_K(t)$ is monic.*

As mentioned above, the holonomy representation ρ_0 is parabolic, so that Theorem 1.1 is an immediate corollary of the following theorem.

Theorem 1.2. *Let $\rho: G_K\to \mathrm{SL}_2(\mathbb{C})$ be a parabolic representation of $K=J(k,2n)$. Then*:

(1) *$\Delta_{K,\rho}(t)$ determines the genus of $J(k,2n)$.*

(2) *$\Delta_{K,\rho}(t)$ determines the fiberedness of $J(k,2n)$ if $k=2m+1$, $k=2$ (twist knot), or $k=2m$ and $|4mn-1|\in\mathcal{P}_2$.*

Remark 1.3. (1) Suppose $k=2m$ and $n>0$. Then $4mn-1\in\mathcal{P}_2$ if $4mn-1$ is a prime and $2mn-1$ is a prime $\equiv 1 \pmod 4$.

(2) It is known that the conjugacy classes of parabolic representations into $\mathrm{SL}_2(\mathbb{C})$ of the knot $J(2m,2n)$ can be described as the zero locus of an integral polynomial in one variable. The condition $|4mn-1|\in\mathcal{P}_2$ in Theorem 1.2 (hence Theorem 1.1) assures the irreducibility over $\mathbb{Z}$ of this polynomial, see Section 5. We do not know whether Theorem 1.2(2) holds true for every integer m and n.

This paper is organized as follows. In Section 2, we study nonabelian representations of the knot $J(k,2n)$ and give an explicit formula of the defining equation of the representation space. In Section 3, we investigate parabolic representations of $J(k,2n)$. In Section 4, we quickly review the definition of the twisted Alexander polynomial and some related work on fiberedness and genus of knots. In particular, we calculate the coefficients of the highest- and lowest-degree terms of the twisted Alexander polynomial associated to a nonabelian representation of $J(k,2n)$ and

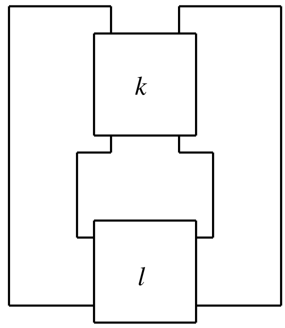


Figure 1. The knot $K = J(k,l)$. Here $k > 0$ and $l = 2n$ $(n \in \mathbb{Z})$ denote the numbers of half twists in each box. Positive numbers correspond to right-handed twists and negative numbers correspond to left-handed twists.

give the proof of Theorem 1.2(1). In Section 5, we discuss the fibering problem and prove Theorem 1.2(2).

2. Non-abelian representations

Let $K = J(k,l)$ be the knot as in Figure 1. Note that $J(k,l)$ is a knot if and only if kl is even, and is the trivial knot if $kl = 0$. Furthermore, $J(k,l) \cong J(l,k)$ and $J(-k,-l)$ is the mirror image of $J(k,l)$. Hence, in the following, we consider $K = J(k,2n)$ for $k > 0$ and $|n| > 0$. When $k = 2$, $J(2,2n)$ is the twist knot.

In this section we explicitly calculate the defining equation of the nonabelian representation space of $J(k,2n)$.

By [Hoste and Shanahan 2004] the knot group of $K = J(k,2n)$ is presented by $G_K = \langle a, b \mid w^n a = b w^n \rangle$, where

$$w = \begin{cases} (ba^{-1})^m (b^{-1}a)^m & \text{for } k = 2m, \\ (ba^{-1})^m ba (b^{-1}a)^m & \text{for } k = 2m+1. \end{cases}$$

Let $\{S_i(z)\}_i$ be the sequence of Chebyshev polynomials, defined by $S_0(z) = 1$, $S_1(z) = z$, and $S_{i+1}(z) = zS_i(z) - S_{i-1}(z)$ for all positive integers i.

The following lemmas are standard; see, for example, [Tran 2013a, Lemma 2.4] and [Tran 2013b, Lemma 3.2].

Lemma 2.1. *One has* $S_i^2(z) - zS_i(z)S_{i-1}(z) + S_{i-1}^2(z) = 1$.

Lemma 2.2. *Suppose the sequence* $\{M_i\}_i$ *of* 2×2 *matrices satisfies the recurrence relation* $M_{i+1} = zM_i - M_{i-1}$ *for all integers* i*. Then*

$$M_i = S_{i-1}(z)M_1 - S_{i-2}(z)M_0, \tag{2-1}$$

$$M_i = S_i(z)M_0 - S_{i-1}(z)M_{-1}. \tag{2-2}$$

A representation $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ is called nonabelian if $\rho(G_K)$ is a nonabelian subgroup of $\mathrm{SL}_2(\mathbb{C})$. Taking conjugation if necessary, we can assume that ρ has the form

$$\rho(a) = A = \begin{bmatrix} s & 1 \\ 0 & s^{-1} \end{bmatrix} \quad \text{and} \quad \rho(b) = B = \begin{bmatrix} s & 0 \\ 2-y & s^{-1} \end{bmatrix}, \tag{2-3}$$

where $(s, y) \in \mathbb{C}^* \times \mathbb{C}$ satisfies the matrix equation $W^n A - B W^n = 0$. Here $W = \rho(w)$. It can be easily checked that $y = \operatorname{tr} AB^{-1}$ holds. Let $x = \operatorname{tr} A = \operatorname{tr} B = s + s^{-1}$.

Lemma 2.3. *One has*

$$WA - BW = \begin{bmatrix} 0 & \alpha_k(x, y) \\ (y-2)\alpha_k(x, y) & 0 \end{bmatrix},$$

where

$$\alpha_k(x, y) = \begin{cases} 1 - (y + 2 - x^2) S_{m-1}(y)\big(S_{m-1}(y) - S_{m-2}(y)\big) & \text{for } k = 2m, \\ 1 + (y + 2 - x^2) S_{m-1}(y)\big(S_m(y) - S_{m-1}(y)\big) & \text{for } k = 2m+1. \end{cases}$$

Proof. Recall that by the Cayley–Hamilton theorem, $M^{i+1} = (\operatorname{tr} M) M^i - M^{i-1}$ for all matrices $M \in \mathrm{SL}_2(\mathbb{C})$ and all integers i.

If $k = 2m$ then by applying (2-1) twice, we have

$$\begin{aligned} WA &= (BA^{-1})^m (B^{-1}A)^m A \\ &= S_{m-1}^2(y) BA^{-1}B^{-1}AA - S_{m-1}(y) S_{m-2}(y)(BA^{-1}A + B^{-1}AA) + S_{m-2}^2(y) A. \end{aligned}$$

Similarly,

$$\begin{aligned} BW &= B(BA^{-1})^m (B^{-1}A)^m \\ &= S_{m-1}^2(y) BBA^{-1}B^{-1}A - S_{m-1}(y) S_{m-2}(y)(BBA^{-1} + BB^{-1}A) + S_{m-2}^2(y) B. \end{aligned}$$

Hence, by direct calculations using (2-3), we obtain

$$\begin{aligned} WA - BW &= S_{m-1}^2(y)(BA^{-1}B^{-1}AA - BBA^{-1}B^{-1}A) + S_{m-2}^2(y)(A - B) \\ &\qquad - S_{m-1}(y) S_{m-2}(y)(BA^{-1}A - BBA^{-1} + B^{-1}AA - BB^{-1}A) \\ &= \begin{bmatrix} 0 & \alpha_k(x, y) \\ (y-2)\alpha_k(x, y) & 0 \end{bmatrix}, \end{aligned}$$

where

$$\alpha_k(x, y) = (s^{-2} + 1 + s^2 - y) S_{m-1}^2(y) - (s^{-2} + s^2) S_{m-1}(y) S_{m-2}(y) + S_{m-2}^2(y).$$

Since $S_{m-1}^2(y) - y S_{m-1}(y) S_{m-2}(y) + S_{m-2}^2(y) = 1$ (by Lemma 2.1) and $x = s + s^{-1}$,

$$\alpha_k(x, y) = 1 - (y + 2 - x^2) S_{m-1}(y)\big(S_{m-1}(y) - S_{m-2}(y)\big).$$

If $k = 2m+1$ then by applying (2-2) twice, we have

$$\begin{aligned}
WA &= (BA^{-1})^m BA(B^{-1}A)^m A \\
&= S_m^2(y)BAA - S_m(y)S_{m-1}(y)((BA^{-1})^{-1}BAA + BA(B^{-1}A)^{-1}A) \\
&\qquad\qquad + S_{m-1}^2(y)(BA^{-1})^{-1}BA(B^{-1}A)^{-1}A \\
&= S_m^2(y)BAA - S_m(y)S_{m-1}(y)(A^3 + B^2A) + S_{m-1}^2(y)ABA.
\end{aligned}$$

Similarly,

$$\begin{aligned}
BW &= B(BA^{-1})^m BA(B^{-1}A)^m \\
&= S_m^2(y)BBA - S_m(y)S_{m-1}(y)(B(BA^{-1})^{-1}BA + BBA(B^{-1}A)^{-1}) \\
&\qquad\qquad + S_{m-1}^2(y)B(BA^{-1})^{-1}BA(B^{-1}A)^{-1} \\
&= S_m^2(y)BBA - S_m(y)S_{m-1}(y)(BA^2 + B^3) + S_{m-1}^2(y)BAB.
\end{aligned}$$

Hence, by direct calculations using (2-3), we obtain

$$\begin{aligned}
WA - BW &= S_m^2(y)(BAA - BBA) + S_{m-1}^2(y)(ABA - BAB) \\
&\qquad\qquad - S_m(y)S_{m-1}(y)(A^3 - BA^2 + B^2A - B^2) \\
&= \begin{bmatrix} 0 & \alpha_k(x,y) \\ (y-2)\alpha_k(x,y) & 0 \end{bmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\alpha_k(x,y) &= S_m^2(y) - (s^{-2}+s^2)S_m(y)S_{m-1}(y) + (s^{-2}+1+s^2-y)S_{m-1}^2(y) \\
&= 1 + (y+2-x^2)S_{m-1}(y)(S_m(y) - S_{m-1}(y)).
\end{aligned}$$

This completes the proof of Lemma 2.3. □

The proof of the following lemma is similar to that of Lemma 2.3.

Lemma 2.4. *One has*

$$\operatorname{tr} W = \begin{cases} 2 + (y-2)(y+2-x^2)S_{m-1}^2(y) & \text{for } k = 2m, \\ x^2 - y - (y-2)(y+2-x^2)S_m(y)S_{m-1}(y) & \text{for } k = 2m+1. \end{cases}$$

We are now ready to calculate the expression $W^nA - BW^n$ as follows. Let $\lambda = \operatorname{tr} W$.

Proposition 2.5. *One has*

$$W^nA - BW^n = \begin{bmatrix} 0 & S_{n-1}(\lambda)\alpha_k(x,y) - S_{n-2}(\lambda) \\ (y-2)(S_{n-1}(\lambda)\alpha_k(x,y) - S_{n-2}(\lambda)) & 0 \end{bmatrix}.$$

Proof. By applying (2-1) and Lemma 2.3, we have

$$\begin{aligned} W^nA - BW^n &= S_{n-1}(\lambda)(WA - BW) - S_{n-2}(\lambda)(A - B) \\ &= S_{n-1}(\lambda)\begin{bmatrix} 0 & \alpha_k(x,y) \\ (y-2)\alpha_k(x,y) & 0 \end{bmatrix} - S_{n-2}(\lambda)\begin{bmatrix} 0 & 1 \\ y-2 & 0 \end{bmatrix}. \end{aligned}$$

The proposition follows. □

Proposition 2.5 implies that the assignment (2-3) gives a nonabelian representation $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ if and only if $(s, y) \in \mathbb{C}^* \times \mathbb{C}$ satisfies the equation

$$\phi_{k,2n}(x, y) := S_{n-1}(\lambda)\alpha_k(x, y) - S_{n-2}(\lambda) = 0,$$

where $\alpha_k(x, y)$ and $\lambda = \operatorname{tr} W$ are given by the formulas in Lemmas 2.3 and 2.4 respectively.

The polynomial $\phi_{k,2n}(x, y)$ is also known as the Riley polynomial [Riley 1984; Tkhang 1993] of $J(k, 2n)$.

3. Parabolic representations

A representation $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ is called parabolic if the meridian μ of K is sent to a parabolic element (i.e., $\operatorname{tr}\rho(\mu) = 2$) of $\mathrm{SL}_2(\mathbb{C})$ and $\rho(G_K)$ is nonabelian.

Let $K = J(k, 2n)$. In this section we will show that if $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ is a parabolic representation of the form

$$\rho(a) = A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \rho(b) = B = \begin{bmatrix} 1 & 0 \\ 2-y & 1 \end{bmatrix},$$

where y is a real number satisfying the equation $\phi_{k,2n}(2, y) = 0$, then $y > 2$.

Lemma 3.1. *Suppose $x = 2$. Then*

$$\alpha_k^2(x, y) - \alpha_k(x, y)\lambda + 1 = \begin{cases} (y-2)^3 S_{m-1}^4(y) & \textit{for } k = 2m, \\ (y-2)((y-2)S_m(y)S_{m-1}(y) + 1)^2 & \textit{for } k = 2m+1. \end{cases}$$

Proof. If $k = 2m$ then $\alpha_k(x, y) = 1 - (y + 2 - x^2)S_{m-1}(y)(S_{m-1}(y) - S_{m-2}(y))$ and $\lambda = 2 + (y-2)(y+2-x^2)S_{m-1}^2(y)$ by Lemmas 2.3 and 2.4. Hence, by direct calculations using $x = 2$, we have

$$\begin{aligned} &\alpha_k^2(x, y) - \alpha_k(x, y)\lambda + 1 \\ &= \left(-1 - S_{m-1}^2(y) + yS_{m-1}^2(y) - yS_{m-1}(y)S_{m-2}(y) + S_{m-2}^2(y)\right)(y-2)^2 S_{m-1}^2(y). \end{aligned}$$

Since $S_{m-1}^2(y) - yS_{m-1}(y)S_{m-2}(y) + S_{m-2}^2(y) = 1$ (by Lemma 2.1), we obtain

$$\alpha_k^2(x, y) - \alpha_k(x, y)\lambda + 1 = (y-2)^3 S_{m-1}^4(y).$$

If $k=2m+1$ then $\alpha_k(x,y)=1+(y+2-x^2)S_{m-1}(y)\,(S_m(y)-S_{m-1}(y))$ and $\lambda=x^2-y-(y-2)(y+2-x^2)S_m(y)S_{m-1}(y)$ by Lemmas 2.3 and 2.4. Hence, by direct calculations using $x=2$, we have

$$\begin{aligned}\alpha_k^2(x,y)-\alpha_k(x,y)\lambda+1=(y-2)\big(&1+(2-y)S_{m-1}^2(y)+(y-2)S_{m-1}^4(y)\\&+2(y-2)S_{m-1}(y)S_m(y)+(2-y)yS_{m-1}^3(y)S_m(y)\\&+(y^2-3y+2)S_{m-1}^2(y)S_m^2(y)\big).\end{aligned}$$

By replacing $yS_{m-1}^3(y)S_m(y)=S_{m-1}^2(y)(S_{m-1}^2(y)+S_m^2(y)-1)$ in this equality, we obtain

$$\alpha_k^2(x,y)-\alpha_k(x,y)\lambda+1=(y-2)\left((y-2)S_m(y)S_{m-1}(y)+1\right)^2,$$

as claimed. □

Proposition 3.2. *Suppose y is a real number satisfying the equation $\phi_{k,2n}(2,y)=0$. Then $y>2$.*

Proof. Suppose $\phi_{k,2n}(x,y)=0$. Then $S_{n-1}(\lambda)\alpha_k(x,y)=S_{n-2}(\lambda)$. Hence

$$\begin{aligned}1&=S_{n-1}^2(\lambda)-\lambda S_{n-1}(\lambda)S_{n-2}(\lambda)+S_{n-2}^2(\lambda)\\&=\big(\alpha_k^2(x,y)-\alpha_k(x,y)\lambda+1\big)\,S_{n-1}^2(\lambda).\end{aligned}$$

If we also suppose that $x=2$ and y is a real number, then the above equality implies that $\alpha_k^2(x,y)-\alpha_k(x,y)\lambda+1>0$. By Lemma 3.1, we must have $y>2$. □

4. Twisted Alexander polynomials

In this section we explicitly calculate the coefficients of the highest- and lowest-degree terms of the twisted Alexander polynomial associated to a nonabelian representation of $J(k,2n)$ and give the proof of Theorem 1.2(1).

Twisted Alexander polynomials. For a knot group $G_K=\pi_1(S^3\backslash K)$, we choose and fix a Wirtinger presentation

$$G_K=\langle a_1,\dots,a_q\mid r_1,\dots,r_{q-1}\rangle.$$

Then the abelianization homomorphism $f:G_K\to H_1(S^3\backslash K;\mathbb{Z})\cong\mathbb{Z}=\langle t\rangle$ is given by $f(a_1)=\cdots=f(a_q)=t$. Here we specify a generator t of $H_1(S^3\backslash K;\mathbb{Z})$ and denote the sum in $\mathbb{Z}$ multiplicatively. Let us consider a linear representation $\rho:G_K\to\mathrm{SL}_2(\mathbb{C})$.

The maps ρ and f naturally induce two ring homomorphisms $\tilde\rho:\mathbb{Z}[G_K]\to M(2,\mathbb{C})$ and $\tilde f:\mathbb{Z}[G_K]\to\mathbb{Z}[t^{\pm1}]$ respectively, where $\mathbb{Z}[G_K]$ is the group ring of G_K and $M(2,\mathbb{C})$ is the matrix algebra of degree 2 over $\mathbb{C}$. Then $\tilde\rho\otimes\tilde f$ defines a ring homomorphism $\mathbb{Z}[G_K]\to M(2,\mathbb{C}[t^{\pm1}])$. Let F_q denote the free group on generators $a_1, \ldots, a_q$ and $\Phi : \mathbb{Z}[F_q] \to M(2, \mathbb{C}[t^{\pm 1}])$ the composition of the surjection $\mathbb{Z}[F_q] \to \mathbb{Z}[G_K]$ induced by the presentation of G_K and the map $\tilde{\rho} \otimes \tilde{f} : \mathbb{Z}[G_K] \to M(2, \mathbb{C}[t^{\pm 1}])$.

Consider the $(q-1) \times q$ matrix M whose (i, j)-component is the 2×2 matrix

$$\Phi\Big(\frac{\partial r_i}{\partial a_j}\Big) \in M(2, \mathbb{Z}[t^{\pm 1}]),$$

where $\partial/\partial a$ denotes the free differential calculus. For $1 \le j \le q$, let us denote by M_j the $(q-1) \times (q-1)$ matrix obtained from M by removing the j-th column. We regard M_j as a $2(q-1) \times 2(q-1)$ matrix with coefficients in $\mathbb{C}[t^{\pm 1}]$. Then Wada's twisted Alexander polynomial of a knot K associated to a representation $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ is defined to be a rational function

$$\Delta_{K,\rho}(t) = \frac{\det M_j}{\det \Phi(1 - a_j)}$$

and moreover well-defined up to a factor t^{2n} $(n \in \mathbb{Z})$. It is also known that if two representations ρ, ρ' are conjugate, then $\Delta_{K,\rho}(t) = \Delta_{K,\rho'}(t)$ holds. See [Wada 1994] and [Goda et al. 2005] for details.

Remark 4.1. Let $\rho : G_K \to \mathrm{SL}_2(\mathbb{C})$ be a nonabelian representation.

(1) The twisted Alexander polynomial $\Delta_{K,\rho}(t)$ associated to ρ is always a Laurent polynomial for any knot K [Kitano and Morifuji 2005].

(2) The twisted Alexander polynomial is reciprocal; that is, $\Delta_{K,\rho}(t) = t^i \Delta_{K,\rho}(t^{-1})$ for some $i \in \mathbb{Z}$ [Hillman et al. 2010; Friedl et al. 2012].

(3) If K is a fibered knot, then $\Delta_{K,\rho}(t)$ is a monic polynomial for every nonabelian representation ρ [Goda et al. 2005]. It is also known that the converse holds for alternating knots [Kim and Morifuji 2012, Remark 4.2].

(4) If K is a knot of genus g, then $\deg(\Delta_{K,\rho}(t)) \le 4g - 2$ [Friedl and Kim 2006]. Moreover if K is fibered, then the equality holds [Kitano and Morifuji 2005].

We say the twisted Alexander polynomial $\Delta_{K,\rho}(t)$ determines the knot genus $g(K)$ if $\deg(\Delta_{K,\rho}(t)) = 4g(K) - 2$ holds. For a hyperbolic knot K, the hyperbolic torsion polynomial $\mathcal{T}_K(t)$ is defined to be $\Delta_{K,\rho_0}(t)$ for the holonomy representation $\rho_0 : G_K \to \mathrm{SL}_2(\mathbb{C})$. We note that it is normalized so that $\mathcal{T}_K(t) = \mathcal{T}_K(t^{-1})$ holds.

Proof of part (1) *of Theorem 1.2.* It is known that the genus of $J(k, 2n)$, where $k > 1$ and $|n| > 0$, is 1 if k is even, and is $|n|$ if k is odd. Moreover, the genus of $J(1, 2n)$ (the $(2, 2n-1)$-torus knot) is $n - 1$ if $n > 0$ and is $-n$ if $n < 0$.

We first consider the case $n > 0$. Let $r = w^n a w^{-n} b^{-1}$, where w is as defined in Section 2. By direct calculations, we have

$$\frac{\partial r}{\partial a} = w^n\Big(1 + (1-a)(w^{-1} + \cdots + w^{-n})\frac{\partial w}{\partial a}\Big), \tag{4-1}$$

where, for $k=2m$,

$$\frac{\partial w}{\partial a}=-\left(ba^{-1}+\cdots+(ba^{-1})^m\right)+(ba^{-1})^m\left(1+b^{-1}a+\cdots+(b^{-1}a)^{m-1}b^{-1}\right),$$

and, for $k=2m+1$,

$$\frac{\partial w}{\partial a}=-\left(ba^{-1}+\cdots+(ba^{-1})^m\right)+(ba^{-1})^m b\left(1+ab^{-1}+\cdots+(ab^{-1})^m\right).$$

Suppose $\rho: G_K\to \mathrm{SL}_2(\mathbb{C})$ is a nonabelian representation given by (2-3). Then the twisted Alexander polynomial of K associated to ρ is

$$\Delta_{K,\rho}(t)=\frac{\det\Phi\left(\frac{\partial r}{\partial a}\right)}{\det\Phi(1-b)}=\frac{\det\Phi\left(\frac{\partial r}{\partial a}\right)}{1-tx+t^2}.$$

The case $J(2m,2n),\ n>0$. From (4-1) we have

$$\det\Phi\left(\frac{\partial r}{\partial a}\right)=\left|I+(I-tA)(W^{-1}+\cdots+W^{-n})V\right|,$$

where I is the 2×2 identity matrix and

$$V=-(BA^{-1}+\cdots+(BA^{-1})^m)+(BA^{-1})^m(I+B^{-1}A+\cdots+(B^{-1}A)^{m-1})t^{-1}B^{-1}.$$

The next lemma follows easily.

Lemma 4.2. *The highest- and lowest-degree terms of* $\det\Phi\left(\frac{\partial r}{\partial a}\right)$ *are respectively*

$$\left|A(W^{-1}+\cdots+W^{-n})(BA^{-1}+\cdots+(BA^{-1})^m)\right|t^2$$

and

$$\left|(W^{-1}+\cdots+W^{-n})(BA^{-1})^m(I+B^{-1}A+\cdots+(B^{-1}A)^{m-1})B^{-1}\right|t^{-2}.$$

Let $\{T_i(z)\}_i$ be the sequence of Chebyshev polynomials defined by $T_0(z)=2$, $T_1(z)=z$ and $T_{i+1}(z)=zT_i(z)-T_{i-1}(z)$ for all integers i. Recall that $y=\operatorname{tr}AB^{-1}$ and $\lambda=\operatorname{tr}W$.

Proposition 4.3. *The highest- and lowest-degree terms of* $\det\Phi\left(\frac{\partial r}{\partial a}\right)$ *are respectively*

$$\frac{T_n(\lambda)-2}{\lambda-2}\,\frac{T_m(y)-2}{y-2}\,t^2\quad\text{and}\quad\frac{T_n(\lambda)-2}{\lambda-2}\,\frac{T_m(y)-2}{y-2}\,t^{-2}.$$

Proof. Let $\beta_\pm$ be the roots of $z^2-\lambda z+1$ and $\gamma_\pm$ the roots of z^2-yz+1. Lemma 4.2 implies that the highest- and lowest degree terms of $\det\Phi(\partial r/\partial a)$ are respectively

$$(1+\beta_++\cdots+\beta_+^{n-1})(1+\beta_-+\cdots+\beta_-^{n-1})\\ \times(1+\gamma_++\cdots+\gamma_+^{m-1})(1+\gamma_-+\cdots+\gamma_-^{m-1})t^2$$

and

$$(1+\beta_+ + \cdots + \beta_+^{n-1})(1+\beta_- + \cdots + \beta_-^{n-1}) \\ \times (1+\gamma_+ + \cdots + \gamma_+^{m-1})(1+\gamma_- + \cdots + \gamma_-^{m-1})t^{-2}.$$

Proposition 4.3 then follows from Lemma 4.4 below. □

Lemma 4.4. $(1+\beta_+ + \cdots + \beta_+^{n-1})(1+\beta_- + \cdots + \beta_-^{n-1}) = \dfrac{T_n(\lambda)-2}{\lambda-2} \in \mathbb{Z}[\lambda].$

Proof. The left-hand side is equal to

$$\frac{(\beta_+^n-1)(\beta_-^n-1)}{(\beta_+-1)(\beta_--1)} = \frac{\beta_+^n+\beta_-^n-2}{\beta_++\beta_--2} = \frac{T_n(\lambda)-2}{\lambda-2}.$$

The lemma follows. □

Proposition 4.3 implies that the highest- and lowest-degree terms of the twisted Alexander polynomial

$$\Delta_{K,\rho}(t) = \frac{\det \Phi\left(\dfrac{\partial r}{\partial a}\right)}{1-tx+t^2}$$

are respectively $U_{m,n}(y)t^0$ and $U_{m,n}(y)t^{-2}$, where

$$U_{m,n}(y) = \frac{T_n(\lambda)-2}{\lambda-2}\,\frac{T_m(y)-2}{y-2}.$$

Hence to prove Theorem 1.2(1) for $J(2m, 2n)$, $n > 0$, we only need to show that the coefficients of these terms are nonzero under the assumption that $\phi_K(2, y) = S_{n-1}(\lambda)\alpha_k(2, y) - S_{n-2}(\lambda) = 0$ (because the roots of this equation correspond to the parabolic representations). To this end, we show that at $x = 2$ the polynomials $\phi_K(2, y)$ and $U_{m,n}(y)$ do not have any common zero $y \in \mathbb{C}$ (in fact, if they have a common zero, the highest- and lowest-degree terms vanish at $x = 2$). It is equivalent to show that at $x = 2$ these polynomials are relatively prime in $\mathbb{C}[y]$.

Recall that $\lambda = \operatorname{tr} W = (y-2)(y+2-x^2)S_{m-1}^2(y)+2$ and

$$\alpha_k(x, y) = 1-(y+2-x^2)S_{m-1}(y)\big(S_{m-1}(y)-S_{m-2}(y)\big).$$

The next two lemmas will complete the proof of Theorem 1.2(1) for $J(2m, 2n)$, $n > 0$.

Lemma 4.5. *Suppose* $x = 2$. *Then* $\gcd\left(\phi_K(x, y), \dfrac{T_n(\lambda)-2}{\lambda-2}\right) = 1$ *in* $\mathbb{C}[y]$.

Proof. It is equivalent to show that at $x = 2$, $\phi_K(x, y)$ and $(T_n(\lambda)-2)/(\lambda-2)$ do not have any common root $y \in \mathbb{C}$.

Suppose $T_n(\lambda) = 2$ and $\lambda \neq 2$ then $\beta_+^n = \beta_-^n = 1$ and $\beta_+ \neq 1$. If $\beta_+ \neq -1$ then $S_{n-1}(\lambda) = (\beta_+^n - \beta_-^n)/(\beta_+-\beta_-) = 0$ and $S_{n-2}(\lambda) = (\beta_+^{n-1}-\beta_-^{n-1})/(\beta_+-\beta_-) = -1$; hence $\phi_K(x, y) = 1 \neq 0$.

If $\beta_+ = -1$ (in this case n must be even) then $\lambda = -2$. It implies that $S_{n-1}(\lambda) = -n$, and $S_{n-2}(\lambda) = n-1$. Hence $\phi_K(x, y) = -n\,\alpha_k(x, y) - (n-1)$.

Suppose $\phi_K(x, y) = 0$. Then $\alpha_k(x, y) = 1/n - 1$. We have

$$(y-2)\big(y+2-x^2\big)S_{m-1}^2(y) = \lambda - 2 = -4,$$
$$(y+2-x^2)\left(S_{m-1}^2(y) - S_{m-1}(y)S_{m-2}(y)\right) = 1 - \alpha_k(x, y) = 2 - \frac{1}{n}.$$

Thus $(y+2-x^2)S_{m-1}^2(y) = \dfrac{-4}{y-2}$, $(y+2-x^2)S_{m-1}(y)S_{m-2}(y) = -\Big(\dfrac{4}{y-2}+2-\dfrac{1}{n}\Big)$.

Since

$$S_{m-1}^4(y) - yS_{m-1}^2(y)(S_{m-1}(y)S_{m-2}(y)) + (S_{m-1}(y)S_{m-2}(y))^2 = S_{m-1}^2(y)$$

(by Lemma 2.1), we must have

$$\frac{16}{(y-2)^2} - \frac{4y}{y-2}\left(\frac{4}{y-2} + 2 - \frac{1}{n}\right) + \left(\frac{4}{y-2} + 2 - \frac{1}{n}\right)^2 = \frac{-4(y+2-x^2)}{y-2},$$

that is, $y = 2 + 4n^2x^2$.

If $x \in \mathbb{R}$ then $y = 2 + 4n^2x^2 \in \mathbb{R}$ and

$$-4 = (y-2)(y+2-x^2)S_{m-1}^2(y) = 4n^2x^2(4 + (4n^2-1)x^2)S_{m-1}^2(y) \geq 0,$$

a contradiction. Hence $\phi_K(x, y) \neq 0$ when $\dfrac{T_n(\lambda)-2}{\lambda-2} = 0$ and $x \in \mathbb{R}$. The lemma follows. □

Lemma 4.6. *Suppose $x = 2$. Then* $\gcd\Big(\phi_K(x, y), \dfrac{T_m(y)-2}{y-2}\Big) = 1$ *in* $\mathbb{C}[y]$.

Proof. Suppose $T_m(y) = 2$ and $y \neq 2$ then $\gamma_+^m = \gamma_-^m = 1$ and $\gamma_+ \neq 1$. If $\gamma_+ \neq -1$ then $S_{m-1}(y) = (\gamma_+^m - \gamma_-^m)/(\gamma_+ - \gamma_-) = 0$, hence $\lambda = 2$ and $\alpha_k(x, y) = 1$. This implies that $\phi_K(x, y) = S_{n-1}(2) - S_{n-2}(2) = 1 \neq 0$.

If $\gamma_+ = -1$ (in this case m must be even) then $y = -2$. We have

$$\lambda = (y-2)(y+2-x^2)S_{m-1}^2(y) + 2 = 4m^2x^2 + 2$$

and

$$\alpha_k(x, y) = 1 - (y+2-x^2)S_{m-1}(y)(S_{m-1}(y) - S_{m-2}(y)) = m(2m-1)x^2 + 1.$$

This implies that

$$\phi_K(x, y) = S_{n-1}(\lambda)\alpha_k(x, y) - S_{n-2}(\lambda) = (m(2m-1)x^2+1)S_{n-1}(\lambda) - S_{n-2}(\lambda).$$

If $x \in \mathbb{Z}$ then $\lambda = 4m^2x^2 + 2 \in \mathbb{Z}$ is even. This means that $\phi_K(x, y)$ is odd, since

$$\phi_K(x, y) \equiv S_{n-1}(\lambda) - S_{n-2}(\lambda) \pmod 2.$$

Thus $\phi_K(x, y) \neq 0$ when $\dfrac{T_m(y)-2}{y-2} = 0$ and $x \in \mathbb{Z}$. The lemma follows. □

The case $J(2m+1, 2n),\ n > 0$. From (4-1) we have

$$\det \Phi\Big(\frac{\partial r}{\partial a}\Big) = \big|I + (I - tA)(t^{-2}W^{-1} + \cdots + t^{-2n}W^{-n})V\big|\, t^{4n},$$

where

$$V = -(BA^{-1} + \cdots + (BA^{-1})^m) + t(BA^{-1})^m B(I + AB^{-1} + \cdots + (AB^{-1})^m).$$

We first consider the case $m = 0$ (in this case we must have $n > 1$ so that K is a nontrivial knot). Then $W = BA$ and

$$\begin{aligned}\det \Phi\Big(\frac{\partial r}{\partial a}\Big) &= \big|I + (I - tA)(t^{-2}W^{-1} + \cdots + t^{-2n}W^{-n})tB\big|\, t^{4n} \\ &= \big|(t^{-2}W^{-1} + \cdots + t^{-2n}W^{-n})tB - tA(t^{-4}W^{-2} + \cdots + t^{-2n}W^{-n})tB\big|\, t^{4n}.\end{aligned}$$

This implies that the highest- and lowest-degree terms of $\det \Phi(\partial r/\partial a)$ are $|t^{-1}W^{-1}B|\, t^{4n} = t^{4n-2}$ and $|t^{1-2n}W^{-n}B|\, t^{4n} = t^2$, respectively. Hence the highest- and lowest degree terms of $\Delta_{K,\rho}(t)$ are t^{4n-4} and t^2, respectively. Since the genus of $J(1, 2n)$, where $n > 1$, is $n - 1$, we complete the proof of Theorem 1.2(1) for $J(1, 2n),\ n > 1$.

We now consider the case $m > 0$. In this case, we have the following.

Lemma 4.7. (1) *The highest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\begin{aligned}&\big|I - AW^{-1}(BA^{-1})^m B(I + AB^{-1} + \cdots + (AB^{-1})^m)\big|\, t^{4n} \\ &\qquad\qquad = \big|I + BA^{-1} + \cdots + (BA^{-1})^{m-1}\big|\, t^{4n}.\end{aligned}$$

(2) *The lowest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\big|-W^{-n}(BA^{-1} + \cdots + (BA^{-1})^m)\big|\, t^0 = \big|I + BA^{-1} + \cdots + (BA^{-1})^{m-1}\big|\, t^0.$$

Lemmas 4.7 and 4.4 imply the following.

Proposition 4.8. *The highest- and lowest-degree terms of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *are respectively*

$$\frac{T_m(y) - 2}{y - 2}\, t^{4n} \quad \text{and} \quad \frac{T_m(y) - 2}{y - 2}\, t^0.$$

Proposition 4.8 implies that the highest- and lowest-degree terms of $\Delta_{K,\rho}(t)$ are $(T_m(y) - 2)/(y - 2)t^{4n-2}$ and $(T_m(y) - 2)/(y - 2)t^0$, respectively. Hence to prove Theorem 1.2(1) for $J(2m+1, 2n)$, where $m, n > 0$, we only need to show that at $x = 2$ (parabolic representation) the polynomials $\phi_K(x, y) = S_{n-1}(\lambda)\alpha_k(x, y) - S_{n-2}(\lambda)$ and $(T_m(y) - 2)/(y - 2)$ are relatively prime in $\mathbb{C}[y]$.

Recall that $\lambda = \operatorname{tr} W = x^2 - y - (y-2)(y+2-x^2)S_m(y)S_{m-1}(y)$ and

$$\alpha_k(x, y) = 1 + (y+2-x^2)S_{m-1}(y)(S_m(y) - S_{m-1}(y)).$$

The next lemma will complete the proof of Theorem 1.2(1) for $J(2m+1, 2n)$, where $m, n > 0$.

Lemma 4.9. *Suppose $x = 2$. Then* $\gcd\Bigl(\phi_K(x, y), \dfrac{T_m(y)-2}{y-2}\Bigr) = 1$ *in* $\mathbb{C}[y]$.

Proof. Suppose $T_m(y) = 2$ and $y \neq 2$, then $\gamma_+^m = \gamma_-^m = 1$ and $\gamma_+ \neq 1$. If $\gamma_+ \neq -1$ then $S_{m-1}(y) = 0$ and $S_m(y) = 1$; hence $\lambda = x^2 - y$ and $\alpha_k(x, y) = 1$. This implies that $\phi_K(x, y) = S_{n-1}(\lambda) - S_{n-2}(\lambda)$.

Since $\gamma_+^m = 1$, we have $y = \gamma_+ + \gamma_+^{-1} = 2\cos(2\pi j/m)$ for some $0 < j < m$. If $\phi_K(x, y) = S_{n-1}(\lambda) - S_{n-2}(\lambda) = 0$ then $\lambda = 2\cos\bigl((2j'-1)\pi/(2n-1)\bigr)$ for some $1 \le j' \le n-1$; see [Le and Tran 2012, Lemma 4.13], for example. Hence

$$x^2 = y + \lambda = 2\biggl(\cos\frac{2\pi j}{m} + \cos\frac{(2j'-1)\pi}{2n-1}\biggr) < 4.$$

If $\gamma_+ = -1$ (in this case m must be even) then $y = -2$. We have

$$\lambda = -(y-2)(y+2-x^2)S_m(y)S_{m-1}(y) + x^2 - y = (2m+1)^2x^2 + 2$$

and

$$\alpha_k(x, y) = 1 + (y+2-x^2)S_{m-1}(y)(S_m(y) - S_{m-1}(y)) = m(2m+1)x^2 + 1.$$

If x is an even integer then $\lambda = (2m+1)^2x^2 + 2$ is an even integer and $\alpha_k(x, y) = m(2m+1)x^2 + 1$ is an odd integer. Hence

$$\phi_K(x, y) = S_{n-1}(\lambda)\alpha_k(x, y) - S_{n-2}(\lambda) \equiv S_{n-1}(\lambda) - S_{n-2}(\lambda) \pmod 2$$

is odd and so is nonzero.

In both cases, we obtain $\phi_K(x, y) \neq 0$ when $\dfrac{T_m(y)-2}{y-2} = 0$ and x is an even integer at least 2. The lemma follows. $\square$

Next we consider the case $n < 0$. We put $l = -n$ ($l > 0$). For $r = w^n a w^{-n} b^{-1} = w^{-l} a w^l b^{-1}$, we have

$$\frac{\partial r}{\partial a} = \frac{\partial w^{-l}}{\partial a} + w^{-l}\Bigl(1 + a\frac{\partial w^l}{\partial a}\Bigr) = w^{-l}\Bigl(1 - (1-a)\bigl(1 + w + \cdots + w^{l-1}\bigr)\frac{\partial w}{\partial a}\Bigr). \tag{4-2}$$

<u>The case $J(2m, 2n)$, $n < 0$.</u> From (4-2) we have

$$\det \Phi\Bigl(\frac{\partial r}{\partial a}\Bigr) = \bigl|I - (I - tA)(I + W + \cdots + W^{l-1})V\bigr|,$$

where

$$V = -(BA^{-1} + \cdots + (BA^{-1})^m) + (BA^{-1})^m(I + B^{-1}A + \cdots + (B^{-1}A)^{m-1})t^{-1}B^{-1}.$$

Lemma 4.10. (1) *The highest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\big|-A(I+W+\cdots+W^{l-1})(BA^{-1}+\cdots+(BA^{-1})^m)\big|\,t^2.$$

(2) *The lowest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\big|-(I+W+\cdots+W^{l-1})(BA^{-1})^m(I+B^{-1}A+\cdots+(B^{-1}A)^{m-1})B^{-1}\big|\,t^{-2}.$$

We can apply a similar argument to that of the parallel case with $n>0$ (page 441) to conclude that $\Delta_{K,\rho}(t)$, for ρ parabolic, determines the knot genus in this case.

The case $J(2m+1,2n),\ n<0$. From (4-2) we have

$$\begin{aligned}\det \Phi\Big(\frac{\partial r}{\partial a}\Big) &= \big|t^{-2l}W^{-l}\big(I-(I-tA)(I+t^2W+\cdots+t^{2(l-1)}W^{l-1})V\big)\big|\\ &= \big|I-(I-tA)(I+t^2W+\cdots+t^{2(l-1)}W^{l-1})V\big|\,t^{-4l},\end{aligned}$$

where

$$V=-(BA^{-1}+\cdots+(BA^{-1})^m)+t(BA^{-1})^mB(I+AB^{-1}+\cdots+(AB^{-1})^m).$$

Lemma 4.11. (1) *The highest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\begin{aligned}&\big|AW^{l-1}(BA^{-1})^mB(I+AB^{-1}+\cdots+(AB^{-1})^m)\big|\,t^0\\ &\qquad\qquad = |I+AB^{-1}+\cdots+(AB^{-1})^m|\,t^0.\end{aligned}$$

(2) *The lowest-degree term of* $\det \Phi\Big(\dfrac{\partial r}{\partial a}\Big)$ *is*

$$\big|I+BA^{-1}+\cdots+(BA^{-1})^m\big|\,t^{-4l}.$$

We can apply a similar argument to that of the parallel case with $n>0$ (page 444) to conclude again that $\Delta_{K,\rho}(t)$, for ρ parabolic, determines the knot genus in this case.

The case analysis starting on page 441 covers all possibilities. Theorem 1.2(1) follows immediately. □

5. The fibering problem

In this section we study some properties of the parabolic representation spaces of 2-bridge knots and give the proof of Theorem 1.2(2).

Parabolic representations of 2-bridge knots. Consider the 2-bridge knot $K=\mathfrak{b}(p,q)$, where $p>q\geq 1$ are relatively prime. The knot group G_K has a presentation $G_K=\langle a,b\mid wa=bw\rangle$, where $w=a^{\varepsilon_1}b^{\varepsilon_2}\cdots a^{\varepsilon_{p-2}}b^{\varepsilon_{p-1}}$ and $\varepsilon_j=(-1)^{\lfloor jq/p\rfloor}$ (see, e.g., [Burde and Zieschang 2003]).

Let $\phi_K(x, y)$ be the defining equation for the nonabelian representations into $\mathrm{SL}_2(\mathbb{C})$ of G_K, where $x = \mathrm{tr}\,\rho(a) = \mathrm{tr}\,\rho(b)$ and $y = \mathrm{tr}\,\rho(ab^{-1})$. Then $\phi_K(2, y)$ is the defining equation for the parabolic representations. It is known that $\phi_K(2, y) \in \mathbb{Z}[y]$ is a monic polynomial of degree $d = (p-1)/2$; see [Riley 1984; Tkhang 1993].

We want to study the irreducibility of $\phi_K(2, y) \in \mathbb{Z}[y]$.

Lemma 5.1. *One has* $\phi_K(2, y) = S_d(y) + S_{d-1}(y)$ *in* $\mathbb{Z}_2[y]$.

Proof. The proof is similar to that of [Le and Tran 2012, Proposition A.2].

Suppose ρ is a parabolic representation. Let $A = \rho(a)$, $B = \rho(b)$ and $W = \rho(w)$. Taking conjugation if necessary, we can assume that

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 2-y & 1 \end{bmatrix}, \tag{5-1}$$

where $y = \mathrm{tr}\, AB^{-1} \in \mathbb{C}$ satisfies the matrix equation $WA - BW = 0$.

By the Cayley–Hamilton theorem applying for matrices in $\mathrm{SL}_2(\mathbb{C})$ we have $A + A^{-1} = \mathrm{tr}(A)I = 2I = 0 \pmod 2$, that is, $A^{-1} = A \pmod 2$. Similarly, $B^{-1} = B \pmod 2$. This implies that $W = A^{\varepsilon_1} B^{\varepsilon_2} \dots A^{\varepsilon_{2d-1}} B^{\varepsilon_{2d}} = (AB)^d \pmod 2$. By applying (2-2), we have

$$\begin{aligned} WA + BW &= (AB)^d A + B(AB)^d \\ &= S_d(y)(A+B) + S_{d-1}(y)(B^{-1} + A^{-1}) \\ &= (S_d(y) + S_{d-1}(y))\,(A+B) \pmod 2, \end{aligned}$$

where $A + B = \left[\begin{smallmatrix} 0 & 1 \\ y & 0 \end{smallmatrix}\right] \pmod 2$. Hence $\phi_K(2, y) = S_d(y) + S_{d-1}(y)$ in $\mathbb{Z}_2[y]$. □

Recall from the Introduction that $\mathcal{P}_2$ is the set of all odd primes p such that 2 is a primitive root modulo p.

Lemma 5.2. *Suppose* $p \in \mathcal{P}_2$. *Then* $S_d(y) + S_{d-1}(y) \in \mathbb{Z}_2[y]$ *is irreducible.*

Proof. Let $y = u + u^{-1}$. Then

$$S_d(y) + S_{d-1}(y) = \frac{u^{d+1} + u^{-(d+1)}}{u + u^{-1}} + \frac{u^d + u^{-d}}{u + u^{-1}} = u^{-d}\,\frac{1 + u^{2d+1}}{1+u}.$$

Suppose $p \in \mathcal{P}_2$. We will show that $(1+u^p)/(1+u) \in \mathbb{Z}_2[u]$ is irreducible. This will imply that $S_d(y) + S_{d-1}(y) \in \mathbb{Z}_2[y]$ is irreducible.

We have $(1+u^p)/(1+u) = u^{p-1} + \cdots + u + 1$ is the p-th-cyclotomic polynomial $C_p(u) \in \mathbb{Z}_2[u]$ (since p is an odd prime). It is well known that $C_p(u) \in \mathbb{Z}_2[u]$ is irreducible if $p \in \mathcal{P}_2$; see for example, [Roman 2006, Theorem 11.2.8]. The lemma follows. □

Proposition 5.3. *Suppose* $p \in \mathcal{P}_2$. *Then* $\phi_K(2, y) \in \mathbb{Z}[y]$ *is irreducible.*

Proof. By Lemma 5.1, $\phi_K(2, y) = S_d(y) + S_{d-1}(y) \in \mathbb{Z}_2[y]$. Since $p \in \mathcal{P}_2$, the polynomial $S_d(y) + S_{d-1}(y) \in \mathbb{Z}_2[y]$ is irreducible by Lemma 5.2. This implies that $\phi_K(2, y)$ is irreducible in $\mathbb{Z}_2[y]$. Since $\phi_K(2, y) \in \mathbb{Z}[y]$ is a monic polynomial in y, it is irreducible in $\mathbb{Z}[y]$. □

Proof of part (2) *of Theorem 1.2.* It is known that $J(k, 2n)$ is fibered only for the trivial knot $J(k, 0)$, the trefoil knot $J(2, 2)$, the figure eight knot $J(2, -2)$, the knots $J(1, 2n)$ for any n, and the knots $J(3, 2n)$ for $n > 0$.

The case $J(2m, 2n)$, $m > 1$. We will apply Proposition 5.3 to study the fibering problem for $K = J(2m, 2n)$.

Let $p = |4mn - 1|$ then it is known that $\phi_K(2, y)$ has degree $(p - 1)/2$. By Proposition 5.3, the polynomial $\phi_K(2, y) \in \mathbb{Z}[y]$ is irreducible if $p \in \mathcal{P}_2$.

Proposition 5.4. *Suppose $m > 1$ and $p = |4mn - 1| \in \mathcal{P}_2$. Then $\Delta_{K,\rho}(t)$ is nonmonic for every parabolic representation ρ.*

Proof. We only need to consider the case $n > 0$. The case $n < 0$ is similar.

Suppose ρ is a parabolic representation, that is, $x = 2$. Since $k = 2m$, by Proposition 4.3 the coefficient of the highest-degree term of $\Delta_{K,\rho}(t)$ is $h(y) = (T_n(\lambda) - 2)/(\lambda - 2) \times (T_m(y) - 2)/(y - 2)$, an integer polynomial in y of degree $(n - 1)(2m) + (m - 1) = 2mn - (m + 1) < 2mn - 1 = (p - 1)/2$.

Since $p \in \mathcal{P}_2$, the polynomial $\phi_K(2, y) \in \mathbb{Z}[y]$ is irreducible. This implies that $\phi_K(2, y)$ does not divide $h(y) - 1$ in $\mathbb{Z}[y]$. Hence $h(y) \neq 1$ when $\phi_K(2, y) = 0$. The proposition follows. □

Twist knots $J(2, 2n)$. For $K = J(2, 2n)$ we have $\lambda = y^2 - yx^2 + 2x^2 - 2$ and $\phi_K(x, y) = -(y + 1 - x^2)S_{n-1}(\lambda) - S_{n-2}(\lambda)$. Suppose ρ is a nonabelian representation. By Proposition 4.3 the coefficient of the highest-degree term of $\Delta_{K,\rho}(t)$ is $(T_n(\lambda)-2)/(\lambda-2)$. We want to show that for $|n| > 1$, we have $(T_n(\lambda)-2)/(\lambda-2) \neq 1$ when $\phi_K(x, y) = 0$ and $x = 2$. This will imply that for any parabolic representation ρ, $\Delta_{K,\rho}(t)$ is monic if and only if $|n| = 1$.

Lemma 5.5. *If $x = 2$, then* $\gcd\left(\phi_K(2, y), \dfrac{T_n(\lambda)-2}{\lambda-2} - 1\right) = 1$ *in $\mathbb{C}[y]$ for $|n| > 1$.*

Proof. We only need to consider the case $n > 1$. The case $n < -1$ is similar.

Suppose $T_n(\lambda) = \lambda$ and $\lambda \neq 2$. Then $\beta_+^n + \beta_-^n = \beta_+ + \beta_-$, i.e., $\beta_+^{n-1} = 1$ or $\beta_+^{n+1} = 1$. It implies that $\lambda = -2$, or $\lambda = 2\cos 2j\pi/(n-1)$ for some $1 \leq j \leq n-2$ and $j \neq (n-1)/2$, or $\lambda = 2\cos 2j\pi/(n+1)$ for some $1 \leq j \leq n$ and $j \neq (n+1)/2$.

Case 1: $\lambda = -2$ (in this case n must be odd). By similar arguments as in the proof of Lemma 4.5, we have $\phi_K(x, y) \neq 0$ if $x \in \mathbb{R}$.

Case 2: $\lambda = 2\cos 2j\pi/(n-1)$ for some $1 \leq j \leq n-2$ and $j \neq (n-1)/2$. Then $S_{n-1}(\lambda) = 1$ and $S_{n-2}(\lambda) = 0$, hence $\phi_K(x, y) = -(y + 1 - x^2)$.

Suppose $\phi_K(x, y) = 0$. Then $y = x^2 - 1$ and $\lambda = y^2 - yx^2 + 2x^2 - 2 = x^2 - 1$. This cannot occur if $x^2 - 1 \geq 2$, since $\lambda < 2$. Hence $\phi_K(x, y) \neq 0$ if $x^2 \geq 3$.

Case 3: $\lambda = 2\cos 2j\pi/(n+1)$ for some $1 \leq j \leq n$ and $j \neq (n+1)/2$. Then $S_{n-1}(\lambda) = -1$ and $S_{n-2}(\lambda) = -\lambda$, hence $\phi_K(x, y) = y + 1 - x^2 + \lambda$.

Suppose $\phi_K(x, y) = 0$. Then $y = x^2 - \lambda - 1$ and $\lambda = y^2 - yx^2 + 2x^2 - 2 = \lambda^2 + \lambda(2 - x^2) + x^2 - 1$, that is, $\lambda^2 - \lambda(x^2 - 1) + x^2 - 1 = 0$. This equation does not have any real solution λ if $1 < x^2 < 5$. Hence $\phi_K(x, y) \neq 0$ if $1 < x^2 < 5$.

In all cases, $\phi_K(x, y) \neq 0$ when $(T_n(\lambda) - 2)/(\lambda - 2) = 1$ and $3 \leq x^2 < 5$. The lemma follows. □

Remark 5.6. Lemma 5.5 gives a proof of [Morifuji 2012, Theorem 1.2] that does not use the irreducibility of $\phi_{J(2,2n)}(2, y) \in \mathbb{Z}[y]$ proved in [Hoste and Shanahan 2001].

The case $J(2m+1, 2n)$. Let $K = J(2m+1, 2n)$. Suppose ρ is a nonabelian representation. By Proposition 4.8 and Lemma 4.11, the coefficient of the highest-degree term of $\Delta_{K,\rho}(t)$ is

$$\frac{T_m(y) - 2}{y - 2} \quad \text{if } n > 0 \qquad \text{and} \qquad \frac{T_{m+1}(y) - 2}{y - 2} \quad \text{if } n < 0.$$

We want to show that for $m > 1$, we have $(T_m(y)-2)/(y-2) \neq 1$ when $\phi_K(2, y) = 0$. This will imply that for any parabolic representation ρ, $\Delta_{K,\rho}(t)$ is monic if and only if $K = J(1, 2n)$, or $K = J(3, 2n)$ and $n > 0$.

The key point of the proof of the following lemma is to apply Proposition 3.2.

Lemma 5.7. *If* $x = 2$, *then* $\gcd\left(\phi_K(2, y), \dfrac{T_m(y)-2}{y-2} - 1\right) = 1$ *in* $\mathbb{C}[y]$ *for* $m > 1$.

Proof. Suppose $T_m(y) = y$ and $y \neq 2$. Then $\gamma_+^m + \gamma_-^m = \gamma_+ + \gamma_-$, i.e., $\gamma_+^{m-1} = 1$ or $\gamma_+^{m+1} = 1$. This implies that $y = -2$, or $y = 2\cos 2j\pi/(m-1)$ for some $1 \leq j \leq m-2$ and $j \neq (m-1)/2$, or $y = 2\cos 2j\pi/(m+1)$ for some $1 \leq j \leq m$ and $j \neq (m+1)/2$. In all cases, $y \in \mathbb{R}$ and $y < 2$. Proposition 3.2 then implies that $\phi_K(2, y) \neq 0$. The lemma follows. □

The case analysis on the last two pages cover all possibilities, showing part (2) of Theorem 1.2. This completes the proof of the theorem. □

Acknowledgements

Morifuji was partially supported by Grant-in-Aid for Scientific Research 23540076 from the Ministry of Education, Culture, Sports, Science and Technology, Japan. Tran thanks T. T. Q. Le for helpful discussions. Both authors would like to thank the referee for comments and suggestions.

References

[Burde and Zieschang 2003] G. Burde and H. Zieschang, *Knots*, 2nd ed., de Gruyter Studies in Mathematics **5**, de Gruyter, Berlin, 2003. MR 2003m:57005 Zbl 1009.57003

[Crowell 1959] R. Crowell, "Genus of alternating link types", *Ann. of Math.* (2) **69** (1959), 258–275. MR 20 #6103b Zbl 0111.35803

[Dunfield et al. 2012] N. M. Dunfield, S. Friedl, and N. Jackson, "Twisted Alexander polynomials of hyperbolic knots", *Exp. Math.* **21**:4 (2012), 329–352. MR 3004250 Zbl 1266.57008

[Friedl and Kim 2006] S. Friedl and T. Kim, "The Thurston norm, fibered manifolds and twisted Alexander polynomials", *Topology* **45**:6 (2006), 929–953. MR 2007g:57020 Zbl 1105.57009

[Friedl and Vidussi 2011a] S. Friedl and S. Vidussi, "A survey of twisted Alexander polynomials", pp. 45–94 in *The mathematics of knots*, edited by M. Banagl and D. Vogel, Contrib. Math. Comput. Sci. **1**, Springer, Heidelberg, 2011. MR 2012f:57024 Zbl 1223.57012

[Friedl and Vidussi 2011b] S. Friedl and S. Vidussi, "Twisted Alexander polynomials detect fibered 3-manifolds", *Ann. of Math.* (2) **173**:3 (2011), 1587–1643. MR 2012f:57025 Zbl 1231.57020

[Friedl and Vidussi 2012] S. Friedl and S. Vidussi, "The Thurston norm and twisted Alexander polynomials", preprint, 2012. To appear in *J. Reine Angew. Math.* arXiv 1204.6456

[Friedl et al. 2012] S. Friedl, T. Kim, and T. Kitayama, "Poincaré duality and degrees of twisted Alexander polynomials", *Indiana Univ. Math. J.* **61**:1 (2012), 147–192. MR 3029395 Zbl 1273.57009

[Goda et al. 2005] H. Goda, T. Kitano, and T. Morifuji, "Reidemeister torsion, twisted Alexander polynomial and fibered knots", *Comment. Math. Helv.* **80**:1 (2005), 51–61. MR 2005m:57008 Zbl 1066.57008

[Hillman et al. 2010] J. A. Hillman, D. S. Silver, and S. G. Williams, "On reciprocality of twisted Alexander invariants", *Algebr. Geom. Topol.* **10**:2 (2010), 1017–1026. MR 2011j:57021 Zbl 1200.57005

[Hoste and Shanahan 2001] J. Hoste and P. D. Shanahan, "Trace fields of twist knots", *J. Knot Theory Ramifications* **10**:4 (2001), 625–639. MR 2002b:57012 Zbl 1003.57014

[Hoste and Shanahan 2004] J. Hoste and P. D. Shanahan, "A formula for the A-polynomial of twist knots", *J. Knot Theory Ramifications* **13**:2 (2004), 193–209. MR 2005c:57006 Zbl 1057.57010

[Kim and Morifuji 2012] T. Kim and T. Morifuji, "Twisted Alexander polynomials and character varieties of 2-bridge knot groups", *Internat. J. Math.* **23**:6 (2012), 1250022, 24. MR 2925470 Zbl 1255.57013

[Kitano and Morifuji 2005] T. Kitano and T. Morifuji, "Divisibility of twisted Alexander polynomials and fibered knots", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **4**:1 (2005), 179–186. MR 2006e:57006 Zbl 1117.57004

[Le and Tran 2012] T. T. Q. Le and A. T. Tran, "On the AJ conjecture for knots", preprint, 2012. arXiv 1111.5258

[LeVeque 1977] W. J. LeVeque, *Fundamentals of number theory*, Addison-Wesley, Reading, MA, 1977. MR 58 #465 Zbl 0368.10001

[Lin 2001] X. S. Lin, "Representations of knot groups and twisted Alexander polynomials", *Acta Math. Sin.* (*Engl. Ser.*) **17**:3 (2001), 361–380. MR 2003f:57018 Zbl 0986.57003

[Morifuji 2012] T. Morifuji, "On a conjecture of Dunfield, Friedl and Jackson", *C. R. Math. Acad. Sci. Paris* **350**:19-20 (2012), 921–924. MR 2990904 Zbl 1253.57007

[Murasugi 1958a] K. Murasugi, "On the genus of the alternating knot, I", *J. Math. Soc. Japan* **10** (1958), 94–105. MR 20 #6103a Zbl 0084.19301

[Murasugi 1958b] K. Murasugi, "On the genus of the alternating knot, II", *J. Math. Soc. Japan* **10** (1958), 235–248. MR 20 #6103a Zbl 0106.16701

[Riley 1984] R. Riley, "Nonabelian representations of 2-bridge knot groups", *Quart. J. Math. Oxford Ser.* (2) **35**:138 (1984), 191–208. MR 85i:20043 Zbl 0549.57005

[Roman 2006] S. Roman, *Field theory*, 2nd ed., Graduate Texts in Mathematics **158**, Springer, New York, 2006. MR 2006e:12001 Zbl 1172.12001

[Thurston 1997] W. P. Thurston, *Three-dimensional geometry and topology, I*, Princeton Mathematical Series **35**, Princeton University Press, 1997. MR 97m:57016 Zbl 0873.57001

[Tkhang 1993] L. T. K. Tkhang, "Varieties of representations and their subvarieties of cohomology jumps for knot groups", *Mat. Sb.* **184**:2 (1993), 57–82. In Russian; translated in *Russ. Acad. Sci. Sb. Math.* **78**:1 (1994), 187–209. MR 94a:57016 Zbl 0836.57004

[Tran 2013a] A. T. Tran, "The universal character ring of some families of one-relator groups", *Algebr. Geom. Topol.* **13**:4 (2013), 2317–2333. MR 3073918 Zbl 06185351

[Tran 2013b] A. T. Tran, "The universal character ring of the $(-2, 2m+1, 2n)$-pretzel link", *Int. J. Math.* **24**:8 (2013), 1350063, 13. Zbl 06220065

[Wada 1994] M. Wada, "Twisted Alexander polynomial for finitely presentable groups", *Topology* **33**:2 (1994), 241–256. MR 95g:57021 Zbl 0822.57006

Received December 21, 2012. Revised August 22, 2013.

TAKAYUKI MORIFUJI
DEPARTMENT OF MATHEMATICS
KEIO UNIVERSITY
HIYOSHI CAMPUS
YOKOHAMA 223-8521
JAPAN
morifuji@z8.keio.jp

ANH T. TRAN
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
100 MATH TOWER
231 WEST 18TH AVENUE
COLUMBUS, OH 43210-1174
UNITED STATES
tran.350@osu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.453

SCHWARZIAN DIFFERENTIAL EQUATIONS ASSOCIATED TO SHIMURA CURVES OF GENUS ZERO

FANG-TING TU

Let $X_0^D(N)$, where $(D, N) = 1$, denote the Shimura curve associated to an Eichler order of level N, in an indefinite quaternion algebra over $\mathbb{Q}$ of discriminant D. Let $W_{D,N}$ be the group of all Atkin–Lehner involutions on $X_0^D(N)$ and W_D the subgroup consisting of Atkin–Lehner involutions w_m with $m \mid D$. In this paper, we will determine Schwarzian differential equations associated to Shimura curves $X_0^D(N)/W_D$ of genus zero in the cases where there exists a squarefree integer $M > 1$ such that $X_0^D(M)/W_D$ is of genus zero.

1. Introduction

Let B be an indefinite quaternion algebra of discriminant D over $\mathbb{Q}$. For an Eichler order $\mathbb{O}$ of level N, $(D, N) = 1$, in B, we let $X_0^D(N)$ denote the Shimura curve associated to $\mathbb{O}$. For each divisor m of DN with $(m, DN/m) = 1$, we let w_m denote the Atkin–Lehner involution on $X_0^D(N)$ and $W_{D,N}$ be the group of all Atkin–Lehner involutions. We also let the subgroup of $W_{D,N}$ consisting of w_m, $m \mid D$, be denoted by W_D. (We refer the reader to [Alsina and Bayer 2004; Elkies 1998] for general definitions and properties of Shimura curves.)

The notion of Shimura curves generalizes that of classical modular curves, which correspond to the case $B = M(2, \mathbb{Q})$ with $D = 1$. Many properties and theories about classical modular curves can be extended to the case of Shimura curves. However, because of the lack of cusps in the case $D \neq 1$, there have been very few explicit methods for general Shimura curves. One of the few methods uses differential equations satisfied by automorphic forms and automorphic functions. (See [Bayer and Travesa 2007; Elkies 1998; Yang 2013b; 2004].) The idea is that even though it is difficult to explicitly construct automorphic functions that can be put into practical use, the Schwarzian differential equations associated to automorphic functions in the case of Shimura curves of genus zero can often be determined using analytic information about the automorphic functions and coverings between Shimura curves. (See Section 2 for the definition and properties of Schwarzian differential equations.)

MSC2010: primary 11F12; secondary 11G18.
Keywords: automorphic form, modular form, Shimura curves, Schwarzian differential equation.

Then one can use the solutions of the Schwarzian differential equations in place of automorphic forms to study properties of automorphic forms. For example, Yang [2013b] devised a method to determine Hecke eigenforms in the spaces of automorphic forms, expressed in terms of solutions of Schwarzian differential equations. In [Tu and Yang 2013], we obtained several new algebraic transformations of ${}_2F_1$-hypergeometric functions by interpreting identities among hypergeometric functions as identities among automorphic forms on different Shimura curves.

In view of the significance of Schwarzian differential equations, it is important to determine the Schwarzian differential equation for each of the Shimura curves $X_0^D(N)/G$, $G < W_{D,N}$, of genus zero. Elkies [1998] worked out the Schwarzian equations on $X_0^{10}(1)/W_{10}$, $X_0^{14}(1)/W_{14}$, and $X_0^{15}(1)/W_{15}$. Bayer and Travesa [2007] computed all the Schwarzian differential equations for the Shimura curves $X_0^6(1)/G$ with $G < W_6$. Yang [2013b] also gave Schwarzian differential equations on $X_0^6(1)/W_6$ and $X_0^{10}(1)/W_{10}$ from the properties of the automorphic derivatives. (See Section 2.)

In this paper, we will consider the cases $X_0^D(N)/W_D$ when there exists an integer $M > 1$ such that $X_0^D(M)/W_D$ has genus zero. The reason for this restriction is that we need additional information from coverings between Shimura curves of genus zero in order to completely determine the differential equations. (Note that in [Yang 2013b], a covering between Shimura curves of different levels is also needed in order to compute Hecke operators.) In the process, we also need to work out equations for some Shimura curves of genus one and hyperelliptic Shimura curves, which are useful in determining the covering maps between Shimura curves. As a byproduct of our computation of coverings $X_0^D(N)/W_D \to X_0^D(1)/W_D$, we can also determine the values of Hauptmoduln at several CM-points.

A possible future work related to Schwarzian differential equations is Ramanujan-type series for Shimura curves. A typical example of Ramanujan-type identities for the classical modular curves is

$$\sum_{n=0}^{\infty} \frac{(6n+1)(1/2)_n^3}{(n!)^3}\left(\frac{1}{4}\right)^n = \frac{4}{\pi},$$

where $(a)_n = a(a+1)\cdots(a+n-1)$ is the Pochhammer symbol. Yang [2013a] gave several Ramanujan-type formulae for the Shimura curve $X_0^6(1)/W_6$. He conjectured that the general Ramanujan-type identities for Shimura curves are

$$\sum_{n=0}^{\infty}(R_1 n + R_2)A_n t_0^n = R_3 \frac{\pi}{\Omega_d^2},$$

where R_1, R_2, $R_3 \in \overline{\mathbb{Q}}$, $\sum_0^\infty A_n t^n$ is the expansion of a meromorphic automorphic form of weight 2 with respect to a Hauptmodul t of a Shimura curve of genus zero such that t takes value 0 at a CM-point of discriminant d, and t_0 is the value of t at some CM-point of discriminant $d' \neq d$. The number Ω_d is the period of an elliptic curve E over $\overline{\mathbb{Q}}$ with CM by an imaginary quadratic number field of discriminant d. In the same article, he also gave some numerical results of p-adic analogues of these Ramanujan-type identities. It is natural to expect that those p-adic identities should be related to the p-adic periods of elliptic curves with CM. In this paper, in support of his conjecture, we will numerically obtain Ramanujan-type identities for $X_0^{14}(1)/W_{14}$ using our Schwarzian differential equation. However, we are not able to give a rigorous proof at present.

The rest of the paper is organized as follows. In Section 2, we will review the definition of properties of Schwarzian differential equations. In Section 3, we determine all Shimura curves $X_0^D(N)/W_D$ of genus 0, $N > 1$. In Section 4, we will find explicit coverings of $X_0^D(N)/W_D \to X_0^D(1)/W_D$. The equations for Shimura curves and the methods to obtain them given in [González and Rotger 2004; 2006; Molina 2012] are important here. The explicit coverings will be used later. In Section 5, we will work out Schwarzian differential equations and examples for Ramanujan-type identities from the Shimura curve $X_0^{14}(1)/W_{14}$.

From now on, for simplicity of statements, all Shimura curves mentioned below are assumed not to be classical modular curves.

2. Schwarzian differential equations

Let $t(\tau)$ be a nonconstant automorphic function on a Shimura curve X. It is straightforward to verify that $t'(\tau)$ is a meromorphic automorphic form of weight 2 on X and that the *Schwarzian derivative*

$$\{t, \tau\} := \frac{t'''(\tau)}{t'(\tau)} - \frac{3}{2}\left(\frac{t''(\tau)}{t'(\tau)}\right)^2$$

is a meromorphic automorphic form of weight 4 on X. Thus, the ratio of $\{t, \tau\}$ and $t'(\tau)^2$ is an automorphic function on X. In particular, if X has genus zero and $t(\tau)$ is a Hauptmodul, that is, if t generates the field of automorphic functions on X, then

$$Q(t) := -\frac{\{t, \tau\}}{2t'(\tau)^2}$$

is a rational function of t. In [Bayer and Travesa 2007], given a thrice-differentiable function f of z, the function

$$D(f, z) := -\frac{\{f, z\}}{2f'(z)^2}$$

is called the *automorphic derivative* associated to f.

Now the relation $2Q(t)t'(\tau)^2 + \{t, \tau\} = 0$ can also be written as

$$\frac{d^2}{dt(\tau)^2} t'(\tau)^{1/2} + Q(t)t'(\tau)^{1/2} = 0.$$

In other words, if we consider $t'(\tau)^{1/2}$ as a function of t, then $t'(\tau)^{1/2}$ is a solution of the differential equation

$$(\dagger) \qquad \frac{d^2}{dt^2} f + Q(t) f = 0.$$

Definition 1. The differential equation (†) is called the *Schwarzian differential equation* associated to $t(\tau)$.

The significance of Schwarzian differential equations can be seen from the following result.

Proposition 2 [Yang 2013b]. *Assume that a Shimura curve X has genus zero with elliptic points $\tau_1, \dots, \tau_r$ of orders $e_1, \dots, e_r$, respectively. Let $t(\tau)$ be a Hauptmodul of X and set $a_i = t(\tau_i)$, $i = 1, \dots, r$. For a positive even integer $k \geq 4$, let*

$$d_k = \dim S_k(X) = 1 - k + \sum_{j=1}^{r} \left\lfloor \frac{k}{2}\left(1 - \frac{1}{e_j}\right) \right\rfloor,$$

$S_k(X)$ being the space of automorphic forms of weight k on X. A basis for $S_k(X)$ is

$$t'(\tau)^{k/2} t(\tau)^j \prod_{\substack{j=1 \\ a_j \neq \infty}}^{r} \bigl(t(\tau) - a_j\bigr)^{-\lfloor k(1-1/e_j)/2 \rfloor}, \qquad j = 0, \dots, d_k - 1.$$

In other words, if we can determine the Schwarzian differential equation associated to a Hauptmodul on a Shimura curve, then we can express automorphic forms of any even weight k on this Shimura curve in terms of solutions of the differential equation. This provides a concrete space that we can use to study properties of automorphic forms. For example, Yang [2013b] demonstrated how to compute Hecke operators on these spaces.

Now the upshot is that it is often possible to determine a Schwarzian differential equation without constructing a Hauptmodul first. This is especially true when a Shimura curve of genus zero has three elliptic points. This is due to the well-known fact that a second-order Fuchsian differential equation with precisely three singularities is uniquely determined its local exponents at the three points. For general Shimura curves, the following properties of Schwarzian differential equations and automorphic derivatives are very useful in determining the differential equations.

Proposition 3. *Assume that $X(\mathbb{O})$ has genus zero with elliptic points $\tau_1, \dots, \tau_r$ of order $e_1, \dots, e_r$, respectively. Let $t(\tau)$ be a Hauptmodul of $X(\mathbb{O})$ and set $a_i = t(\tau_i)$, $i = 1, \dots, r$. Then the automorphic derivative $Q(t) = D(t, \tau)$ is equal to*

$$Q(t) = \frac{1}{4} \sum_{\substack{j=1 \\ a_j \neq \infty}}^{r} \frac{1 - 1/e_j^2}{(t - a_j)^2} + \sum_{\substack{j=1 \\ a_j \neq \infty}}^{r} \frac{B_j}{t - a_j}$$

for some constants B_j. Moreover, if $a_j \neq \infty$ for all j, then the constants B_j satisfy

$$\sum_{j=1}^{r} B_j = \sum_{j=1}^{r} \bigl(a_j B_j + \tfrac{1}{4}(1 - 1/e_j^2)\bigr) = \sum_{j=1}^{r} \bigl(a_j^2 B_j + \tfrac{1}{2} a_j (1 - 1/e_j^2)\bigr) = 0.$$

Also, if $a_r = \infty$, then the B_j satisfy

$$\sum_{j=1}^{r-1} B_j = 0, \quad \sum_{j=1}^{r-1} \bigl(a_j B_j + \tfrac{1}{4}(1 - 1/e_j^2)\bigr) = \tfrac{1}{4}\bigl(1 - 1/e_r^2\bigr).$$

Proposition 4 [Yang 2013b]. *Automorphic derivatives have the following properties.*

(1) $D((az+b)/(cz+d), z) = 0$ *for all* $\left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right) \in \mathrm{GL}(2, \mathbb{C})$.

(2) $D(g \circ f, z) = D(g, f(z)) + D(f, z)/(dg/df)^2$.

Proposition 5. *Let $t(\tau)$ be a Hauptmodul for a Shimura curve X of genus* 0. *Let $R(x) \in \mathbb{C}(x)$ be the rational function such that the automorphic derivative $Q(t) = D(t, \tau)$ is equal to $R(z)$. Assume that γ is an element of* $\mathrm{SL}(2, \mathbb{R})$ *normalizing the order $\mathbb{O}$ associated to X and let σ be the automorphism of X induced by γ. If $\sigma : t \mapsto (at+b)/(ct+d)$, then $R(x)$ satisfies*

$$\frac{(ad-bc)^2}{(cx+d)^4} R\left(\frac{ax+b}{cx+d}\right) = R(x).$$

Proof. We shall compute $D(t(\gamma\tau), \tau)$ in two ways. By Proposition 4, we have

$$D(t(\gamma\tau), \tau) = D\left(\frac{at(\tau)+b}{ct(\tau)+d}, t(\tau)\right) + \frac{D(t(\tau), \tau)}{(dt(\gamma\tau)/dt(\tau))^2} = 0 + \frac{(ct+d)^4 R(t)}{(ad-bc)^2}.$$

On the other hand, by the same proposition, we also have

$$D(t(\gamma\tau), \tau) = D(t(\gamma\tau), \gamma\tau) + \frac{D(\gamma\tau, \tau)}{(dt(\gamma\tau)/d\gamma\tau)^2} = R(t(\gamma\tau)) = R\left(\frac{at+b}{ct+d}\right).$$

Comparing the two expressions, we get the formula. □

3. Shimura curves of genus zero

In this section, we will determine all pairs of integers (D, N), $D, N > 1$, such that $X_0^D(N)/W_D$ has genus 0. As explained in the introduction, we will need explicit coverings $X_0^D(N)/W_D \to X_0^D(1)/W_D$ in order to determine Schwarzian differential equations.

To describe the genus formula for $X_0^D(N)/W_D$, we need to recall the definition of CM-points first. Let B be a quaternion algebra of discriminant D over $\mathbb{Q}$ and $\mathbb{O}$ an Eichler order of level N in B. Fix an embedding ι of B into $M(2, \mathbb{R})$. Let K be an imaginary quadratic field and R an order of discriminant $d_R = f^2 d_K$ in K. Following Eichler, we say an embedding $\phi : R \to \mathbb{O}$ is *optimal* if $\phi(K) \cap \mathbb{O} = \phi(R)$. Now the action of the set $\iota \circ \phi(R \backslash \{0\}) \subset \mathrm{GL}^+(2, \mathbb{R})$ on the upper half-plane $\mathbb{H}$ fixes precisely one point τ_ϕ. Such a point is called a *CM-point* (point with complex multiplication) of discriminant d_R. We denote the set of CM-points of discriminant d_R, up to $\mathbb{O}_1^*$-equivalence, by $\mathrm{CM}(d_R)$.

Lemma 6 [Ogg 1983]. *Assume that m is a squarefree divisor of DN such that $(m, DN/m) = 1$. Then the set of the fixed points of an Atkin–Lehner involution w_m, $m > 1$, on $X_0^D(N)$ is*

$$\begin{cases} \mathrm{CM}(-4) \cup \mathrm{CM}(-8) & \text{if } m = 2, \\ \mathrm{CM}(-m) \cup \mathrm{CM}(-4m) & \text{if } m \equiv 3 \bmod 4, \\ \mathrm{CM}(-4m) & \text{otherwise}. \end{cases}$$

We remark that in the case m is not squarefree, the fixed points of w_m will still be CM-points, but the description is complicated. (In general, they will be a proper subset of $\bigcup_{f^2 \mid 4m} \mathrm{CM}(-4m/f^2)$.)

From this lemma, it is easy to determine the number of elliptic points on $X_0^D(N)/G$ for any subgroup G of $W_{D,N}$ such that m is squarefree for any w_m in G.

Lemma 7. *Let G be a nontrivial subgroup of the group $W_{D,N}$ of Atkin–Lehner involutions on $X_0^D(N)$ such that m is squarefree for any $w_m \in G$. Then the only possible orders of elliptic points on $X_0^D(N)/G$ are 2, 3, 4, and 6.*

(1) *If $w_2 \in G$, then the number of elliptic points of order 2 on $X_0^D(N)/G$ is*

$$\frac{2}{|G|} \begin{cases} \displaystyle\sum_{\substack{w_m \in G \\ m \neq 1}} (\#\mathrm{CM}(-4m) + \#\mathrm{CM}(-m)) - \#\mathrm{CM}(-3) & \text{if } w_3 \in G, \\ \displaystyle\sum_{\substack{w_m \in G \\ m \neq 1}} (\#\mathrm{CM}(-4m) + \#\mathrm{CM}(-m)) & \text{if } w_3 \notin G. \end{cases}$$

If $w_2 \notin G$, then the number is $(\#\mathrm{CM}(-4)+2A)/|G|$, *where A is*

$$\begin{cases} \sum\limits_{\substack{w_m\in G\\ m\neq 1}} (\#\mathrm{CM}(-4m)+\#\mathrm{CM}(-m)) - \#\mathrm{CM}(-3) & \text{if } w_3 \in G,\\ \sum\limits_{\substack{w_m\in G\\ m\neq 1}} (\#\mathrm{CM}(-4m)+\#\mathrm{CM}(-m)) & \text{if } w_3 \notin G. \end{cases}$$

(*If $-m$ is not a discriminant, we simply set* $\#\mathrm{CM}(-m)=0$.)

(2) *If $w_3 \in G$, then there are no elliptic points of order* 3 *on $X_0^D(N)/G$. If $w_3 \notin G$, then the number of elliptic points of order* 3 *is* $\#\mathrm{CM}(-3)/|G|$.

(3) *If $w_2 \notin G$, then there are no elliptic points of order* 4 *on $X_0^D(N)/G$. If $w_2 \in G$, then the number of elliptic points of order* 4 *is* $2\,\#\mathrm{CM}(-4)/|G|$.

(4) *If $w_3 \notin G$, then there are no elliptic points of order* 6 *on $X_0^D(N)/G$. If $w_3 \in G$, then the number of elliptic points of order* 6 *is* $2\,\#\mathrm{CM}(-3)/|G|$.

Proof. The fact that only 2, 3, 4, and 6 can be the orders of elliptic points on $X_0^D(N)/G$ is well-known.

Let $w_m \in G$. By Lemma 6, the fixed points of w_m consist of CM(-4), CM$(-m)$, or CM$(-4m)$, depending on m. If $m \neq 1, 3$, then points in CM$(-4m)$ or CM$(-m)$ are fixed only by w_m and every other Atkin–Lehner involution other than w_1 permutes them. Thus, there are totally $|G|/2$ points in CM$(-4m)$ or CM$(-m)$ that are mapped to the same point in the covering $X_0^D(N) \to X_0^D(N)/G$. For points in CM(-4), which constitute elliptic points of order 2 on $X_0^D(N)$, they are also fixed by w_2. Thus, if $w_2 \in G$, then there are $2\,\#\mathrm{CM}(-4)/|G|$ elliptic points of order 4 on $X_0^D(N)/G$. If $w_2 \notin G$, points in CM(-4) contribute another $\#\mathrm{CM}(-4)/|G|$ elliptic points of order 2 on $X_0^D(N)/G$. For points in CM(-3), which are elliptic points of order 3 on $X_0^D(N)$, they are also fixed by w_3. If $w_3 \in G$, then they become elliptic points of order 6 on $X_0^D(N)/G$ and there are $2\,\#\mathrm{CM}(-3)/|G|$ such points. If $w_3 \notin G$, then they remain elliptic points of order 3. There are $\#\mathrm{CM}(-3)/|G|$ such points. Summarizing, we get the lemma. □

In view of these lemmas, a formula for the genus of $X_0^D(N)/G$, $G < W_{D,N}$, will involve the numbers of CM-points of certain discriminants. The general formula for the number of CM-points of an arbitrary discriminant is complicated to state. (See [Alsina and Bayer 2004; Ogg 1983].) For the goal of this section, we only need to know the number of CM-points of discriminant -3, d_K, or $4d_K$ in the case $d_K \equiv 1 \bmod 4$, for $K = \mathbb{Q}(\sqrt{-m})$ with $m \mid D$.

Lemma 8 [Ogg 1983]. *For $m \mid D$ or $m = 3$, let d_K denote the discriminant of the field $K = \mathbb{Q}(\sqrt{-m})$. We have*

$$\#\mathrm{CM}(d_K) = h(d_K)\begin{cases} 0 & \text{if } p^2 \mid N \text{ for some } p \mid d_K,\\ \prod\limits_{p\mid D}\left(1-\left(\frac{d_K}{p}\right)\right)\prod\limits_{p\mid N}\left(1+\left(\frac{d_K}{p}\right)\right) & \text{otherwise.}\end{cases}$$

Also, for $m \mid D$ with $m \equiv 3 \bmod 4$, we have

$$\#\mathrm{CM}(4d_K) = \delta h(4d_K) \begin{cases} 0 & \text{if } 2 \mid D, \\ \prod_{p \mid D} \left(1 - \left(\frac{4d_K}{p}\right)\right) \prod_{p \mid N} \left(1 + \left(\frac{4d_K}{p}\right)\right) & \text{if } 2 \nmid D, \end{cases}$$

where, when $m \equiv 7 \bmod 8$,

$$\delta = \begin{cases} 6 & \text{if } 8 \mid N, \\ 4 & \text{if } 4 \mid N, \\ 2 & \text{if } 2 \mid N, \\ 1 & \text{if } 2 \nmid N, \end{cases}$$

and when $m \equiv 3 \bmod 8$,

$$\delta = \begin{cases} 0 & \text{if } 8 \mid N, \\ 2 & \text{if } 2 \mid N \text{ or } 4 \mid N, \\ 1 & \text{if } 2 \nmid N. \end{cases}$$

Here $h(d)$ is the class number of the imaginary quadratic order of discriminant d.

Proof. These formulas are just special cases of Theorems 1 and 2 of [Ogg 1983]. □

Lemma 9. *The complete list of integers (D, N) with $D, N > 1$ such that the Shimura curve $X_0^D(N)/W_D$ has genus zero, is*

$$(6,5),\ (6,7),\ (6,13),\ (10,3),\ (10,7),\ (14,3),\ (14,5),$$
$$(15,2),\ (15,4),\ (21,2),\ (26,3),\ (35,2),\ (39,2).$$

Proof. Let Γ be a congruence Fuchsian subgroup of $\mathrm{SL}(2,\mathbb{R})$. (See [Katok 1992] for the definition of a congruence Fuchsian subgroup; the groups considered here are all congruence Fuchsian subgroups.) A famous result of Selberg [1965] stated that if Γ is a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$, then the first eigenvalue λ_1 of the Laplace operator on the space of square-integrable function on $\Gamma\backslash\mathbb{H}$ is not less than $3/16$. By combining this result with the Jacquet–Langlands correspondence, Vignéras [1983] showed that the same inequality also holds for congruence Fuchsian subgroups coming from indefinite quaternion algebras over $\mathbb{Q}$ of discriminant not equal to 1.

On the other hand, Zograf [1991] showed that if the area $A(\Gamma\backslash\mathbb{H})$ is at least $16(g(\Gamma)+1)$, then $\lambda_1 < 4(g(\Gamma)+1)/A(\Gamma\backslash\mathbb{H})$. Here $g(\Gamma)$ denotes the genus of Γ and the area is normalized such that $A(\mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}) = 1/6$. Combining Selberg's inequality and Zograf's result, one sees that if a congruence Fuchsian subgroup has genus 0, then the area must be less than $64/3$.

Now recall from [Shimizu 1965] that the area of $X_0^D(N)$ is given by

$$\frac{DN}{6} \prod_{p \mid D} \left(1 - \frac{1}{p}\right) \prod_{p \mid N} \left(1 + \frac{1}{p}\right).$$

This immediately shows that if the number of prime factors of D is at least 6, then the genus of $X_0^D(N)/W_D$ cannot be 0 for any $N \geq 2$. Also, if $D = pq$ is a product of two primes such that $(p-1)(q-1) > 512/3$, then $X_0^D(N)/W_D$ must have a positive genus for any $N \geq 2$. A similar bounds exists for the case D has 4 prime factors. This leaves finitely many cases to check.

Now recall that the genus of a Shimura curve X is given by

$$g(X) = 1 + \frac{A(X)}{2} - \frac{1}{2}\sum_{i=1}^{r}\left(1 - \frac{1}{e_i}\right),$$

where the sum runs through all elliptic points with e_i being their respective orders. For $X = X_0^D(N)/W_D$, by Lemma 7, we have

$$\begin{aligned} g(X) = 1 + \frac{A(X)}{2} - \frac{1}{4}\sum_{\substack{m\mid D\\ m\neq 1,3}} \frac{1}{2^{r-1}}(\#\mathrm{CM}(-4m) + \#\mathrm{CM}(-m)) \\ - \begin{cases} \dfrac{1}{4\cdot 2^r}\,\#\mathrm{CM}(-4) & \text{if } 2\nmid D, \\ \dfrac{3}{8\cdot 2^{r-1}}\,\#\mathrm{CM}(-4) & \text{if } 2\mid D \end{cases} \\ - \begin{cases} \dfrac{1}{3\cdot 2^r}\,\#\mathrm{CM}(-3) & \text{if } 3\nmid D, \\ \left(\dfrac{1}{4\cdot 2^{r-1}}\,\#\mathrm{CM}(-12) + \dfrac{5}{12\cdot 2^{r-1}}\,\#\mathrm{CM}(-3)\right) & \text{if } 3\mid D, \end{cases} \end{aligned}$$

where r is the number of prime divisors of D. (Of course, if d is not a discriminant, then we simply let $\mathrm{CM}(d)$ be the empty set.)

Using the Selberg–Zograf bound, the genus formula in the paragraph above and Lemma 8, we check case by case that the pairs of integers given in the lemma are the only cases where $X_0^D(N)/W_D$, $N > 1$, has genus zero. □

We now tabulate all Shimura curves $X_0^D(M)/W_D$ of genus 0 for integers D that appear in the lemma. We will also give a description of their elliptic points. We wish to determine the Schwarzian differential equations for these curves. Here v_j denotes the number of elliptic points of order j on $X_0^D(M)/W_D$. Here we also let $\mathrm{CM}(-m)$ denote the set of points on $X_0^D(N)/W_D$ that are the image of CM-points of discriminants $-m$ under the covering $X_0^D(N) \to X_0^D(N)/W_D$. The number n in $\mathrm{CM}(-m)^{\times n}$ means the number of elements in $\mathrm{CM}(-m)$ is n. If $n = 1$, we omit this annotation.

4. Coverings of Shimura curves

The goal of this section is to obtain explicit coverings of $X_0^D(N)/W_D \to X_0^D(1)/W_D$ for pairs of D and N given in Lemma 9. That is, we wish to find a Hauptmodul t_1

D, N	v_2, v_3, v_4, v_6	elliptic points
6, 1	1, 0, 1, 1	CM(−3), CM(−4), CM(−24)
6, 5	2, 0, 2, 0	$\mathrm{CM}(-4)^{\times 2}$, $\mathrm{CM}(-24)^{\times 2}$
6, 7	2, 0, 0, 2	$\mathrm{CM}(-3)^{\times 2}$, $\mathrm{CM}(-24)^{\times 2}$
6, 13	0, 0, 2, 2	$\mathrm{CM}(-3)^{\times 2}$, $\mathrm{CM}(-4)^{\times 2}$
10, 1	3, 1, 0, 0	CM(−3), CM(−8), CM(−20), CM(−40)
10, 3	4, 1, 0, 0	CM(−3), $\mathrm{CM}(-8)^{\times 2}$, $\mathrm{CM}(-20)^{\times 2}$
10, 7	4, 2, 0, 0	$\mathrm{CM}(-3)^{\times 2}$, $\mathrm{CM}(-20)^{\times 2}$, $\mathrm{CM}(-40)^{\times 2}$
14, 1	3, 0, 1, 0	CM(−4), CM(−8), $\mathrm{CM}(-56)^{\times 2}$
14, 3	6, 0, 0, 0	$\mathrm{CM}(-8)^{\times 2}$, $\mathrm{CM}(-56)^{\times 4}$
14, 5	4, 0, 2, 0	$\mathrm{CM}(-4)^{\times 2}$, $\mathrm{CM}(-56)^{\times 4}$
15, 1	3, 0, 0, 1	CM(−3), CM(−12), CM(−15), CM(−60)
15, 2	6, 0, 0, 0	$\mathrm{CM}(-12)^{\times 2}$, $\mathrm{CM}(-15)^{\times 2}$, $\mathrm{CM}(-60)^{\times 2}$
15, 4	8, 0, 0, 0	$\mathrm{CM}(-12)^{\times 2}$, $\mathrm{CM}(-15)^{\times 2}$, $\mathrm{CM}(-60)^{\times 4}$
21, 1	5, 0, 0, 0	CM(−4), CM(−7), CM(−28), $\mathrm{CM}(-84)^{\times 2}$
21, 2	7, 0, 0, 0	CM(−4), $\mathrm{CM}(-7)^{\times 2}$, $\mathrm{CM}(-28)^{\times 2}$, $\mathrm{CM}(-84)^{\times 2}$
26, 1	5, 0, 0, 0	CM(−8), CM(−52), $\mathrm{CM}(-104)^{\times 3}$
26, 3	8, 0, 0, 0	$\mathrm{CM}(-8)^{\times 2}$, $\mathrm{CM}(-104)^{\times 6}$
35, 1	6, 0, 0, 0	CM(−7), CM(−28), CM(−35), $\mathrm{CM}(-140)^{\times 3}$
35, 2	10, 0, 0, 0	$\mathrm{CM}(-7)^{\times 2}$, $\mathrm{CM}(-28)^{\times 2}$, $\mathrm{CM}(-140)^{\times 6}$
39, 1	6, 0, 0, 0	$\mathrm{CM}(-52)^{\times 2}$, $\mathrm{CM}(-39)^{\times 2}$, $\mathrm{CM}(-156)^{\times 2}$
39, 2	10, 0, 0, 0	$\mathrm{CM}(-52)^{\times 2}$, $\mathrm{CM}(-39)^{\times 4}$, $\mathrm{CM}(-156)^{\times 4}$

Table 1. All Shimura curves $X_0^D(M)/W_D$ of genus 0 for integers D appearing in Lemma 9.

of $X_0^D(1)/W_D$, a Hauptmodul t_N of $X_0^D(N)/W_D$, and the relation between them. Of course, there are infinitely many choices for t_1 and t_N. For $X_0^D(N)/W_D$, we will choose t_N such that the Atkin–Lehner involution w_N acts by $w_N : t_N \mapsto -t_N$. This will make the determination of Schwarzian differential equation simpler.

***Case D* = 6.** In the case $D = 6$, all the coverings $X_0^6(N)/W_6 \to X_0^6(1)/W_6$, $N = 5, 7, 13$, are already given in [Elkies 1998]. Here we just modify the t_N in [Elkies 1998] such that the new t_N satisfies $w_N : t_N \mapsto -t_N$.

Lemma 10 [Elkies 1998]. (1) *There is a Hauptmodul t_1 for $X_0^6(1)/W_6$ that takes values* 0, 1, *and* ∞ *at the CM-points of discriminants* -24, -4, *and* -3, *respectively.*

(2) *There is a Hauptmodul* $t = t_5$ *for* $X_0^6(5)/W_6$ *that takes values* $\pm i/8$ *and* $\pm\sqrt{-6}/3$ *at the CM-points of discriminants* -4 *and* -24, *respectively. The relation between* t_1 *and* t *is*

$$t_1 = \frac{(2+3t^2)(34-117t+1824t^2)^2}{125(1+6t)^6} = 1 + \frac{27(1+64t^2)(3-7t)^4}{125(1+6t)^6}.$$

The Atkin–Lehner involution w_5 *acts by* $w_5 : t \mapsto -t$.

(3) *There is a Hauptmodul* $t = t_7$ *for* $X_0^6(7)/W_6$ *that takes values* $\pm\sqrt{-3}/9$ *and* $\pm\sqrt{-6}/8$ *at the CM-points of discriminants* -3 *and* -24, *respectively. The relation between* t_1 *and* t *is*

$$t_1 = -\frac{(3+32t^2)(78-396t+1963t^2-12312t^3)^2}{4(1+27t^2)(3+10t)^6}.$$

The Atkin–Lehner involution w_7 *acts by* $w_7 : t \mapsto -t$.

(4) *There is a Hauptmodul* $t = t_{13}$ *for* $X_0^6(13)/W_6$ *that takes values* $\pm 4\sqrt{-3}/9$ *and* $\pm 3i/4$ *at the CM-points of discriminants* -3 *and* -4, *respectively. The relation between* t_1 *and* t *is*

$$t_1 = 1 - \frac{27(9+16t^2)(144-98t+246t^2-161t^3)^4}{16(16+27t^2)(30+3t+55t^2)^6}.$$

The Atkin–Lehner involution w_{13} *acts by* $w_{13} : t \mapsto -t$.

Proof. Elkies [1998] showed that explicit coverings of $X_0^6(N)/W_6 \to X_0^6(1)/W_6$, $N = 5, 7, 13$, are given by

$$t_1 = 1 + 135s^4 + 324s^5 + 540s^6, \qquad w_5 : s \mapsto \frac{42-55s}{55+300s},$$
$$t_1 = -\frac{(4s^2+4s+25)(2s^3-3s^2+12s-2)^2}{108(7s^2-8s+37)}, \quad w_7 : s \mapsto \frac{116-9s}{9+20s},$$

and

$$t_1 = \frac{(s^7-50s^6+63s^5-5040s^4+783s^3-168426s^2-6831s-1864404)^2}{4(7s^2+2s+247)(s^2+39)^6}$$

with

$$w_{13} : s \mapsto \frac{5s+72}{2s-5},$$

respectively. Choosing t such that

$$s = \frac{7t-3}{30t+5}, \quad s = \frac{-29t+6}{10t+3}, \quad s = \frac{-8t+9}{2t+1},$$

respectively, we get the lemma. □

Case $D = 10$. The covering $X_0^{10}(3)/W_{10} \to X_0^{10}(1)/W_{10}$ has also been given in [Elkies 1998]. Here we mainly work on the case $N = 7$.

Lemma 11. (1) *There is a Hauptmodul t_1 for $X_0^{10}(1)/W_{10}$ that takes values* 0, ∞, 2, *and* 27 *at the CM-points of discriminants* -3, -8, -20, *and* -40, *respectively.*

(2) *There is a Hauptmodul $t = t_3$ for $X_0^{10}(3)/W_{10}$ that takes values* 0, $\pm 1/4\sqrt{-2}$, $\pm 1/\sqrt{-5}$ *at the CM-points of discriminants* -3, -8, *and* -20, *respectively. The relation between t_1 and t is*

$$t_1 = \frac{108t(1-2t)^3}{(1+32t^2)(1+7t)^2} = 2 - \frac{2(1+5t^2)(1-20t)^2}{(1+32t^2)(1+7t)^2}.$$

The Atkin–Lehner involution w_3 acts by $w_3 : t \mapsto -t$.

(3) *There is a Hauptmodul $t = t_7$ for $X_0^{10}(7)/W_{10}$ that takes values* $\pm 1/3\sqrt{-3}$, $\pm 1/2\sqrt{-5}$, *and* $\pm\sqrt{-10}/16$ *at the CM-points of discriminants* -3, -20, *and* -40, *respectively. The relation between t_1 and t is*

$$t_1 = \frac{8(1+27t^2)(2-3t+44t^2)^3}{7(1+4t+55t^2+102t^3+736t^4)^2}.$$

The Atkin–Lehner involution w_7 acts by $w_7 : t \mapsto -t$.

Proof. In [Elkies 1998], it is shown that an explicit covering $X_0^{10}(3)/W_{10} \to X_0^{10}(1)/W_{10}$ is given by

$$t_1 = \frac{216(s-1)^3}{(s+1)^2(9s^2-10s+17)}$$

with $w_3 : s \mapsto 10/9 - s$. Let t be the Hauptmodul of $X_0^{10}(1)/W_{10}$ with

$$s = \frac{2}{9t} + \frac{5}{9}.$$

Then the relation of t_1 and t and the action of w_3 are given as in the lemma.

We next consider the case $N = 7$. According to Theorem 3.4 of [González and Rotger 2006], an equation for $X_0^{10}(7)$ is given by

$$y^2 = -27x^4 - 40x^3 + 6x^2 + 40x - 27. \tag{1}$$

The actions of the Atkin–Lehner involutions on this model of $X_0^{10}(7)$ are given by

$$w_{70} : (x, y) \mapsto (x, -y), \quad w_5 : (x, y) \mapsto \left(-\frac{1}{x}, -\frac{y}{x^2}\right),$$

and

$$w_{10} : (x, y) \mapsto \left(\frac{2x+1}{x-2}, \frac{5y}{(x-2)^2}\right).$$

Since CM(-20) are fixed points under the action of w_5, their coordinates on (1) are $(i, \pm 2\sqrt{5}(1+2i))$ and $(-i, \pm 2\sqrt{5}(1-2i))$. Likewise, we find that CM(-40) have coordinates $(2+\sqrt{5}, \pm 8\sqrt{-10}(2+\sqrt{5}))$ and $(2-\sqrt{5}, \pm 8\sqrt{-10}(2-\sqrt{5}))$. Furthermore, from the method of [González and Rotger 2006], we know that the two points at infinity are CM-points of discriminant -3. Thus, the coordinates of CM(-3) are ∞, $(0, \pm 3\sqrt{-3})$, $(2, \pm 15\sqrt{-3})$, and $(-1/2, \pm 15\sqrt{-3}/4)$.

From (1), we can obtain an equation $w^2+27z^2+40z+20=0$ for $X_0^{10}(7)/\langle w_{10}\rangle$, where the covering $X_0^{10}(7) \to X_0^{10}(7)/\langle w_{10}\rangle$ is given by

$$(x, y) \mapsto (w, z) = \left(\frac{y}{x-2}, \frac{x^2+1}{x-2}\right).$$

In this equation for $X_0^{(10)}(7)/\langle w_{10}\rangle$, the actions of the Atkin–Lehner involutions are given by

$$w_{70} = w_7 : (w, z) \mapsto (-w, z), \quad w_2 = w_5 : (w, z) \mapsto \left(\frac{w}{2z+1}, \frac{-z}{2z+1}\right).$$

The coordinates of CM(-3) are the two points at ∞ and $(\pm 3\sqrt{-3}/2, -1/2)$. Also, the coordinates of CM(-20) are $(\pm 2\sqrt{-5}, 0)$, and the coordinates of CM(-40) are $(\pm 8\sqrt{-2}(2+\sqrt{5}), 4+2\sqrt{5})$ and $(\pm 8\sqrt{-2}(2-\sqrt{5}), 4-2\sqrt{5})$.

Now set $t = (z+1)/w$. We can check that t is invariant under w_2 and that $(w, z) \mapsto t = (z+1)/w$ is 2-to-1. Thus, t is a Hauptmodul of $X_0^{10}(7)/W_{10}$. The coordinates of the CM-points of discriminants -3, -20, and -40 are $\pm 1/3\sqrt{-3}$, $\pm 1/2\sqrt{-5}$, and $\pm\sqrt{-10}/16$, respectively. It follows that the relation between t_1 and t is

$$t_1 = \frac{A(1+27t^2)(1+a_1t+a_2t^2)^3}{(1+b_1t+b_2t^2+b_3t^3+b_4t^4)^2}$$

with

$$\begin{aligned} A(1+27t^2)(1+a_1t+a_2t^2)^3 - 2(1+b_1t+b_2t^2+b_3t^3+b_4t^4)^2 \\ = B(1+20t^2)(1+c_1t+c_2t^2+c_3t^3)^2, \\ A(1+27t^2)(1+a_1t+a_2t^2)^3 - 27(1+b_1t+b_2t^2+b_3t^3+b_4t^4)^2 \\ = C\left(1+\tfrac{128}{5}t^2\right)(1+d_1t+d_2t^2+d_3t^3)^2 \end{aligned}$$

for some constants A, B, C, a_j, b_j, c_j, and d_j. Comparing the coefficients, we get

$$t_1 = \frac{8(1+27t^2)(2-3t+44t^2)^3}{7(1+4t+55t^2+102t^3+736t^4)^2}$$

(or the same expression with t replaced by $-t$). This proves the lemma. □

***Case D* = 14.** The case $D = 14$ is also worked out in [Elkies 1998]. Here we only need to make a change of variable so that w_N acts by $w_N : t_N \to -t_N$.

Lemma 12 [Elkies 1998]. (1) *There is a Hauptmodul t_1 for $X_0^{14}(1)/W_{14}$ that takes values* ∞, 0, *and* $(-13 \pm 7\sqrt{-7})/32$ *at CM-points of discriminants* -4, -8, *and* -56, *respectively.*

(2) *There is a Hauptmodul $t = t_3$ for $X_0^{14}(3)/W_{14}$ that takes values $\pm 1/\sqrt{-2}$ and $(\pm 9\sqrt{-7} \pm 4\sqrt{-14})/49$ at CM-points of discriminants -8 and -56, respectively. The relation between t_1 and t is*

$$t_1 = \frac{4(1+2t^2)(1-5t)^2}{9(1+t)^4}.$$

The Atkin–Lehner involution w_3 acts by $w_3 : t \mapsto -t$.

(3) *There is a Hauptmodul $t = t_5$ for $X_0^{14}(5)/W_{14}$ that takes values $\pm i/4$ and $(\pm 5\sqrt{-7} \pm 4\sqrt{-14})/7$ at CM-points of discriminants -4 and -56, respectively. The relation between t_1 and t is*

$$t_1 = -\frac{5(1-t+17t^2-13t^3)^2}{(1+16t^2)(1+3t)^4}.$$

The Atkin–Lehner involution w_5 acts by $w_5 : t \mapsto -t$.

Proof. In [Elkies 1998], it is shown that explicit coverings $X_0^{14}(N)/W_{14} \to X_0^{14}(1)/W_{14}$ can be given by

$$t_1 = \tfrac{1}{27}(s^4 + 2s^3 + 9s^2), \quad w_3 : s \mapsto \frac{5-2s}{2+s}$$

and

$$t_1 = -\frac{(256s^3 + 224s^2 + 232s + 217)^2}{50000(s^2+1)}, \quad w_5 : s \mapsto \frac{24-7s}{7+24s},$$

respectively. Choosing t with

$$s = \frac{1-5t}{1+t}, \quad s = \frac{3-16t}{4+12t},$$

respectively, we get the lemma. □

***Case D* = 15.** An explicit covering $X_0^{15}(2)/W_{15} \to X_0^{15}(1)/W_{15}$ is given in [Elkies 1998]. Here we only need make a change of variable so w_N acts by $w_N : t_N \to -t_N$.

Lemma 13. (1) *There is a Hauptmodul for $X_0^{15}(1)/W_{15}$ that takes values* ∞, 0, 81, *and* 1 *at CM-points of discriminants* -3, -12, -15, *and* -60, *respectively.*

(2) *There is a Hauptmodul* t_2 *for* $X_0^{15}(2)/W_{15}$ *that takes values* ± 1, $\pm\sqrt{-15}/3$, *and* $\pm 1/5$ *at CM-points of discriminant* -12, -15, *and* -60, *respectively. The relation between* t_1 *and* t_2 *is*

$$t_1 = \frac{27(1-t_2)(1-3t_2)^2}{2(1+t_2)^3} = 1 + \frac{(1-5t_2)(5-7t_2)^2}{2(1+t_2)^3} = 81 - \frac{27(1+5t_2)(5+3t_2^2)}{2(1+t_2)^3}.$$

The Atkin–Lehner involution w_2 *acts by* $w_2 : t_2 \mapsto -t_2$.

(3) *There is a Hauptmodul* t_4 *for* $X_0^{15}(4)/W_{15}$ *that takes values* $\pm 1/\sqrt{-3}$, $\pm\sqrt{-15}/5$, *and* $(\pm 1 \pm \sqrt{-15})/8$ *at CM-points of discriminants* -12, -15, *and* -60, *respectively. The relation between* t_4 *and* t_2 *is*

$$t_2 = \frac{5t_4^2 + 2t_4 + 1}{7t_4^2 - 2t_4 + 3}.$$

Proof. In [Elkies 1998], an explicit covering $X_0^{15}(2)/W_{15} \to X_0^{15}(1)/W_{15}$ is given by

$$t_1 = \tfrac{1}{4}s(s-3)^2, \quad w_2 : s \mapsto \frac{36}{s}.$$

Choosing a Hauptmodul t for $X_0^{15}(2)/W_{15}$ with

$$s = \frac{6-6t}{1+t},$$

we establish the claim about $X_0^{15}(2)/W_{15}$.

For the covering map $X_0^{15}(4)/W_{15} \to X_0^{15}(2)/W_{15}$, it is clear that one of the CM-points of discriminant -12 on $X_0^{15}(2)/W_{15}$ becomes two CM-points of discriminant -12 on $X_0^{15}(4)/W_{15}$, and the other is ramified. To determine the ramification data of this covering completely, we need to consider the optimal embeddings of the quadratic orders of the field $\mathbb{Q}(\sqrt{-15})$ into the Eichler order of level 2 and the Eichler order of level 4 in the quaternion algebra B over $\mathbb{Q}$ with discriminant 15 at the finite place $p = 2$.

Let $R_1 = \mathbb{Z} + \mathbb{Z}\alpha$, $p_1(x) = x^2 + x + 4$ be the irreducible polynomial of α over $\mathbb{Q}$, and $R_2 = \mathbb{Z} + \mathbb{Z}\beta$, $p_2(x) = x^2 + 15$ be the irreducible polynomial of β over $\mathbb{Q}$. Up to conjugation, we may assume that in the localization $M(2, \mathbb{Q}_2)$ of B at the finite place 2, the Eichler orders $\mathbb{O}_2$, $\mathbb{O}_4$ of level 2 and 4 are

$$\mathbb{O}_2 = \begin{pmatrix} \mathbb{Z}_2 & \mathbb{Z}_2 \\ 2\mathbb{Z}_2 & \mathbb{Z}_2 \end{pmatrix}, \quad \mathbb{O}_4 = \begin{pmatrix} \mathbb{Z}_2 & \mathbb{Z}_2 \\ 4\mathbb{Z}_2 & \mathbb{Z}_2 \end{pmatrix},$$

respectively. Then the inequivalent optimal embeddings of R_1 into $\mathbb{O}_2$ can be given by sending α to

$$A_{-15,1} = \begin{pmatrix} 0 & -1 \\ 4 & -1 \end{pmatrix} \quad \text{and} \quad A_{-15,2} = \begin{pmatrix} -1 & -1 \\ 4 & 0 \end{pmatrix};$$

the inequivalent optimal embeddings of R_2 into $\mathbb{O}_2$ can be given by sending β to

$$A_{-60,1} = \begin{pmatrix} 1 & -1 \\ 16 & -1 \end{pmatrix} \quad \text{and} \quad A_{-60,2} = \begin{pmatrix} 1 & -8 \\ 2 & -1 \end{pmatrix}.$$

The inequivalent optimal embeddings of R_1 and R_2 into $\mathbb{O}_4$ are given by

$$B_{-15,1} = \begin{pmatrix} 0 & -1 \\ 4 & -1 \end{pmatrix} \quad \text{and} \quad B_{-15,2} = \begin{pmatrix} -1 & -1 \\ 4 & 0 \end{pmatrix},$$

and

$$B_{-60,1} = \begin{pmatrix} 1 & -1 \\ 16 & -1 \end{pmatrix} \quad \text{and} \quad B_{-60,2} = \begin{pmatrix} -1 & -1 \\ 16 & 1 \end{pmatrix},$$

$$B_{-60,3} = \begin{pmatrix} 1 & -4 \\ 4 & -1 \end{pmatrix} \quad \text{and} \quad B_{-60,4} = \begin{pmatrix} -1 & -4 \\ 4 & 1 \end{pmatrix},$$

respectively. Furthermore, we can check the embeddings sending β to $B_{-60,3}$, $B_{-60,4}$ give optimal embeddings of R_1 into $\mathbb{O}_2$, and the matrices $B_{-60,1}$, $B_{-60,2}$, and $A_{-60,1}$ are conjugate to each other in $\mathbb{O}_2$.

According this information, we can conclude that each CM-point of discriminant -15 on $X_0^{15}(2)/W_{15}$ becomes one CM-point of discriminant -15 and one CM-point of discriminant -60 on $X_0^{15}(4)/W_{15}$. One of the CM-points of discriminant -60 on $X_0^{15}(2)/W_{15}$ becomes two CM-points of discriminant -60 on $X_0^{15}(4)/W_{15}$, and the other CM-points of discriminant -60 on $X_0^{15}(2)/W_{15}$ is ramified.

We now suppose that the covering $X_0^{15}(4)/W_{15} \to X_0^{15}(2)/W_{15}$ is given by

$$t_2 = \frac{a_2 t^2 + a_1 t + a_3}{t^2 + b_1 t + b_2},$$

where $t = t_4$ is a Hauptmodul for $X_0^{15}(4)/W_{15}$. Since the Atkin–Lehner involution w_2 switches the two CM-points of discriminant -12 on $X_0^{15}(2)/W_{15}$, we may assume that the CM-point of discriminant -12 having coordinate 1 is a ramified point. According to the ramification data and the fields of definition of these CM-points, without loss the generality, we may assume that t has repeated roots 1 when $t_2 = 1$, and assume that the CM-points of discriminant -12 of $X_0^{15}(4)/W_{15}$ that lie above the unramified CM-point of discriminant -12 of $X_0^{15}(2)/W_{15}$ are $\pm 1/\sqrt{-3}$. Therefore, we have

$$t_2 = \frac{(2a-3)t^2 + (3a-1)t + 1 - 2a}{t^2 + (1-3a)t + a},$$

for some constant a. From the information of the CM-points of discriminant -60,

$$t_2^2 - 1 = \frac{(t-c)^2(t^2 + c_1 t + c_2)}{(t^2 + (1-3a)t + a)^2},$$

and the roots of $t^2 + c_1 t + c_2$ are in the field $\mathbb{Q}(\sqrt{-3}, \sqrt{5})$, we can deduce that

$$t_2 = \frac{5t^2 + 2t + 1}{7t^2 - 2t + 3}.$$

We get the lemma. □

***Case D* = 21.** We will need an equation for some Atkin–Lehner quotient of $X_0^{21}(2)$ in order to determine the coordinates of elliptic points on $X_0^{21}(2)$.

Lemma 14. *An equation for* $X_0^{21}(2)/\langle w_{21}\rangle$ *is* $y^2 = (x+12)(x^2-7x+28)$. *Moreover, the action of the Atkin–Lehner involution* $w_3 = w_7$ *on this curve is given by the map* $(x, y) \mapsto (x, -y)$. *Also, the two rational points* ∞ *and* $(-12, 0)$ *are the CM-points of discriminant* -28, *and the other two 2-torsion points* $((7 \pm 3\sqrt{-7})/2, 0)$ *are the CM-points of discriminant* -7.

Proof. We follow the methods of [González and Rotger 2006]. The Shimura curve $X_0^{21}(2)/\langle w_{21}\rangle$ has genus 1. By Lemma 5.10 of that paper, the two CM-points of discriminant -28 are $\mathbb{Q}$-rational points on this curve. Thus, $X_0^{21}(2)/\langle w_{21}\rangle$ is an elliptic curve over $\mathbb{Q}$. Now in the space $S_2(\Gamma_0(42))^{21\text{-new}}$ the unique Hecke eigenform with $+$-eigenvalue for w_{21} is coming from the newform space of $S_2(\Gamma_0(42))$. Therefore, the elliptic curve $X_0^{21}(2)/\langle w_{21}\rangle$ has conductor 42. Using the Cherednik–Drinfeld theory of p-adic uniformization of Shimura curves, we find that the types of singular fibers at primes of bad reduction of $X_0^{21}(2)/\langle w_{21}\rangle$ agree with those of the elliptic curve 42A1, in Cremona's notation. The global minimal model of the elliptic curve 42A1 is $y^2 + xy + y = x^3 + x^2 - 4x + 5$. With a simple change of variables, we write it as $y^2 = (x + 12)(x^2 - 7x + 28)$.

Now the covering $X_0^{21}(2)/\langle w_{21}\rangle \to X_0^{21}(2)/W_{21}$ is ramified at the two CM-points of discriminant -7 and the two CM-points of discriminant -28. If we let one of the CM-points of discriminant -28 be the point at infinity, then an equation for $X_0^{21}(2)/\langle w_{21}\rangle$ is of the form $y^2 = f(x)$ for some polynomial $f(x) = x^3 + \cdots$ of degree 3 in $\mathbb{Q}[x]$ with the Atkin–Lehner involution w_3 acting by $(x, y) \mapsto (x, -y)$. Up to a transformation of the form $x \mapsto ax + b$, this polynomial $f(x)$ must be the polynomial $(x + 12)(x^2 - 7x + 28)$. This proves the lemma. □

Remark 15. According to Cremona's table of elliptic curves [1997], the elliptic curve 42A1 has 8 rational points. Thus, $X_0^{21}(2)/\langle w_{21}\rangle$ also has 8 $\mathbb{Q}$-rational points. Two of them are the CM-points of discriminant -28 mentioned above. The rest of $\mathbb{Q}$-rational points consist of two CM-points of discriminant -4 and four CM-points of discriminant -16.

Lemma 16. *There is a Hauptmodul* t_1 *for* $X_0^{21}(1)/W_{21}$ *that takes values* 49, 0, ∞, *and* $(47 \pm 8\sqrt{-3})/7$ *at CM-points of discriminants* -4, -7, -28, *and* -84, *respectively.*

Also, there is a Hauptmodul $t = t_2$ *for* $X_0^{21}(2)/W_{21}$ *that takes values* $0, \pm 1/3\sqrt{-7}$, ± 1, *and* $\pm 1/3\sqrt{-3}$ *at CM-points of discriminants* -4, -7, -28, *and* -84, *respectively. The relation between* t_1 *and* t *is*

$$t_1 = \frac{49(1+t)(1+63t^2)}{(1-t)(1-15t)^2} = 49 + \frac{1568t(1-3t)^2}{(1-t)(1-15t)^2}.$$

The Atkin–Lehner involution w_2 *acts by* $w_2 : t \mapsto -t$.

Proof. According to [González and Rotger 2006], an equation for $X_0^{21}(1)$ is given by $y^2 = -7x^4 + 94x^2 - 343$ with the actions of the Atkin–Lehner involutions given by

$$w_3 : (x, y) \mapsto (-x, -y), \quad w_7 : (x, y) \mapsto (-x, y), \quad w_{21} : (x, y) \mapsto (x, -y).$$

The Atkin–Lehner involution w_7 fixes the two points at ∞ and $(0, \pm 7\sqrt{-7})$. Since the equation has a symmetry $(x, y) \mapsto (7/x, 7y/x^2)$, we might as well assume that the two points $(0, \pm 7\sqrt{-7})$ are the CM-points of discriminant -7 and the two points at infinity are the CM-points of discriminant -28. Moreover, the four points with $y = 0$ correspond to the four CM-points of discriminant -84.

Since w_3 acts by $(x, y) \to (-x, -y)$, an equation for $X_0^{21}(1)/\langle w_3 \rangle$ is $y^2 = -7x^3 + 94x^2 - 343x$, where the covering $X_0^{21}(1) \to X_0^{21}(1)/\langle w_3 \rangle$ is given by $(x, y) \mapsto (x^2, xy)$. Then $t_1 = x$ generates the function field of X_0^{21}/W_{21}. The values of t_1 at the CM-points of discriminants -7, -28, and -84 are 0, ∞, and $(47 \pm 8\sqrt{-3})/7$, respectively. The value of t_1 at the CM-point of discriminant -4 will be determined later.

By Lemma 14, an equation $X_0^{21}(2)/\langle w_{21} \rangle$ is $y^2 = (x+12)(x^2 - 7x + 28)$ with the Atkin–Lehner involution $w_3 = w_7$ acting by $(x, y) \to (x, -y)$. Thus, $s = x$ generates the function field of $X_0^{21}(2)/W_{21}$. According to the lemma, the values of s at the CM-points of discriminant -7 are $(7 \pm 3\sqrt{-7})/2$ and those at CM-points of discriminant -28 are -12 and ∞. The Atkin–Lehner involution w_2 switches the two CM-points of discriminant -28. It also switches the two CM-points of discriminant -7. (Note that in general, w_2 can send a CM-point of discriminant $-d$ on $X_0^D(N)/G$ to a CM-point of discriminant $-4d$ and vice versa. Here because w_2 is defined over $\mathbb{Q}$, it must send a $\mathbb{Q}$-rational point to another $\mathbb{Q}$-rational point.) This information suffices to determine w_2 in terms of s. We find

$$w_2 : s \mapsto \frac{-12s + 112}{s + 12}.$$

Choosing a new Hauptmodul

$$t = \frac{4 - s}{28 + s},$$

we have $w_2 : t \mapsto -t$. The new coordinates of CM-points of discriminants -7 and -28 are $\pm 1/3\sqrt{-7}$ and ± 1, respectively. Also, since w_2 fixes the unique CM-point of discriminant -4, we find that the CM-point of discriminant -4 has coordinate 0. We now determine the relation between t_1 and t.

Replacing t by $-t$ if necessary, we may assume that the CM-point of discriminant -28 of $X_0^{21}(2)/W_{21}$ that lies above the CM-point of discriminant -7 of $X_0^{21}(1)/W_{21}$ is -1. Then

$$t_1 = \frac{A(1+t)(1+63t^2)}{(1-t)(1-at)^2}$$

for some constants A and a. Since $X_0^{21}(2)/W_{21} \to X_0^{21}(1)/W_{21}$ is also ramified at the CM-points of discriminant -84, the discriminant of the polynomial

$$A(1+t)(1+63t^2) - B(1-t)(1-at)^2$$

in t must be divisible by the polynomial $7B^2 - 94B + 343$. This gives us two conditions on A and a. Solving them for A and a, we find that the only legitimate values for A and a are $A = 49$ and $a = 15$. Because t has value 0 at the CM-point of discriminant -4 on $X_0^{21}(2)/W_{21}$, the CM-point of -4 on $X_0^{21}(1)/W_{21}$ has coordinate 49. This proves the lemma. □

***Case D* = 26.** We first recall a lemma of González and Rotger.

Lemma 17 [González and Rotger 2004, Proposition 2.1]. *Let C be a hyperelliptic curve of genus 2 defined over a field k of characteristic not equal to 2 or 3 and let w be its hyperelliptic involution. Assume that the group of automorphisms of C over k contains a subgroup $\langle u_1, u_2 = u_1 \cdot w\rangle$ isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$ and denote by C_i the elliptic quotient $C/\langle u_i\rangle$. If the two elliptic curves*

$$E_1 : y^2 = x^3 + A_1 x + B_1, \quad E_2 : y^2 = x^3 + A_2 x + B_2$$

are isomorphic to C_1 and C_2 over k, respectively, then C admits a hyperelliptic equation of the form $y^2 = ax^6 + bx^4 + cx^2 + d$, where $a \in k^*$, $b \in k$ are solutions of

$$\begin{aligned} 27a^3 B_2 &= 2A_1^3 + 27B_1^2 + 9A_1 B_1 b + 2A_1^2 b^2 - B_1 b^3, \\ 9a^2 A_2 &= -3A_1^2 + 9B_1 b + A_1 b^2, \end{aligned}$$

$c = (3A_1 + b^2)/(3a)$, $d = (27B_1 + 9A_1 b + b^3)/(27a^2)$, *and the involution u_1 on C is given by $(x, y) \mapsto (-x, y)$.*

Lemma 18. *The Shimura curves* $X_1 : X_0^{26}(3)/\langle w_2, w_3\rangle$, $X_2 : X_0^{26}(3)/\langle w_2, w_{39}\rangle$, *and* $X_3 : X_0^{26}(3)/\langle w_6, w_{13}\rangle$ *are elliptic curves over* $\mathbb{Q}$ *with defining equations*

$$\begin{aligned} X_1 &: y^2 = x^3 - 3403x - 83834, \\ X_2 &: y^2 = x^3 - 43x + 166, \\ X_3 &: y^2 = x^3 + 621x + 9774. \end{aligned}$$

Moreover, on the equation for X_1, *the point at* ∞ *is the CM-point of discriminant* -312, *and the involution* $(x, y) \mapsto (x, -y)$ *is the Atkin–Lehner involution* $w_{13} = w_{26} = w_{39} = w_{78}$. *On the equation for* X_2, *the point at* ∞ *is the CM-point of discriminant* -24 *and the involution* $(x, y) \mapsto (x, -y)$ *is the Atkin–Lehner involution* $w_3 = w_6 = w_{13} = w_{26}$. *On the equation for* X_3, *the point at* ∞ *is the CM-point of discriminant* -8 *and the involution* $(x, y) \mapsto (x, -y)$ *is the Atkin–Lehner involution* $w_2 = w_3 = w_{26} = w_{39}$. *In all three cases, the* 2*-torsion points are the CM-points of discriminant* -104 *on their respective curves.*

Proof. The fact that the three curves in the lemma have genus one can be verified either by using the genus formula, together with Lemmas 6, 7, and 8, or by counting the dimensions of subspaces of $S_2(\Gamma_0(78))^{26\text{-new}}$ with appropriate eigenvalues for the Atkin–Lehner involutions. We omit the details.

On X_1, there is a unique CM-point of discriminant -312, which must be a $\mathbb{Q}$-rational point. Thus, X_1 is an elliptic curve over $\mathbb{Q}$. Likewise, X_2 and X_3 have unique CM-points of discriminants -24 and -8, respectively. They are also elliptic curves over $\mathbb{Q}$.

Observe that all cusp forms in $S_2(\Gamma_0(78))^{26\text{-new}}$ having -1 eigenvalue for w_2 are from the cusp form of level 26 corresponding to the isogeny class 26B of elliptic curves in Cremona's notation. Thus, X_1 and X_2 are isomorphic to either 26B1 or 26B2. Similarly, we find that the one-dimensional subspace of $S_2(\Gamma_0(78))^{26\text{-new}}$ that has eigenvalue $+1$ for both w_6 and w_{13} comes from the cusp form associated to 26A. Using the Cerednik–Drinfeld theory to compute the types of singular fibers at primes 2 and 13, we see that X_1 is isomorphic to the elliptic curve 26B2, X_2 is isomorphic to 26B1, and X_3 is isomorphic to 26A3. If we put the CM-point of discriminant -312 on X_1, that of discriminant -24 on X_2, and that of discriminant -8 on X_3 at ∞, respectively, and require that the Atkin–Lehner involutions w_{13}, w_3, and w_2 act by $(x, y) \to (x, -y)$ on the three curves, respectively, we get the equations for the three curves. □

Lemma 19. (1) *An equation for the curve* $X_0^{26}(3)/\langle w_2\rangle$ *is*

$$y^2 = -\tfrac{2197}{3}x^6 - 362x^4 - 55x^2 - \tfrac{8}{3}$$

with the actions of the Atkin–Lehner involutions given by

$$w_3 : (x, y) \mapsto (-x, y), \quad w_{13} : (x, y) \mapsto (x, -y).$$

On this model, the two CM-points of discriminant -312 *are the two points at infinity, and the two CM-points of discriminant* -24 *are* $(0, \pm 2\sqrt{-6}/3)$.

(2) *An equation for the curve* $X_0^{26}(3)/\langle w_6\rangle$ *is*

$$y^2 = \tfrac{2197}{72}x^6 - \tfrac{699}{8}x^4 - \tfrac{225}{8}x^2 - \tfrac{81}{8}$$

with the actions of the Atkin–Lehner involutions given by

$$w_2 : (x, y) \mapsto (-x, y), \quad w_{26} : (x, y) \mapsto (x, -y).$$

On this model, the two CM-points of discriminant -312 *are the two points at infinity, and the two CM-points of discriminant* -8 *are* $(0, \pm 9\sqrt{-2}/4)$.

(3) *An equation for* $X_0^{26}(3)/\langle w_{39}\rangle$ *is*

$$y^2 = \tfrac{8}{9}x^6 + 9x^4 - 18x^2 + 81$$

with the actions of the Atkin–Lehner involutions given by

$$w_2 : (x, y) \mapsto (-x, y), \quad w_6 : (x, y) \mapsto (x, -y).$$

On this model, the two CM-points of discriminant -24 *are the two points at infinity, and the two CM-points of discriminant* -8 *are* $(0, \pm 9)$.

Moreover, on each of these three curves, there are six CM-points of discriminant -104. *Their coordinates are* $(\alpha_j, 0)$, $j = 1, \dots, 6$, *where* α_j *are the zeros of their respective polynomials of degree* 6.

Proof. We apply Proposition 2.1 of [González and Rotger 2004], cited as Lemma 17 above, with $C = X_0^{26}(3)/\langle w_2\rangle$, w_{13}, $u_1 = w_3$, $u_2 = w_{39}$, $A_1 = -3403$, $B_1 = -83834$, $A_2 = -43$, and $B_2 = 166$. We find an equation for $X_0^{26}(3)/\langle w_2\rangle$ is

$$y^2 = -\tfrac{2197}{3}x^6 - 362x^4 - 55x^2 - \tfrac{8}{3}$$

with the Atkin–Lehner involutions given by

$$w_3 : (x, y) \mapsto (-x, y), \quad w_{13} : (x, y) \mapsto (x, -y).$$

Since the CM-points of discriminant -24 are fixed by the involution $w_6 = w_3 : (x, y) \mapsto (-x, y)$, we see that their coordinates are $(0, \pm 2\sqrt{-6}/3)$. Likewise, the CM-points of discriminant -312 are the fixed points of $w_{78} = w_{39} : (x, y) \mapsto (-x, -y)$, so they are the two points at infinity. Also, the CM-points of discriminant -104 are the fixed points of $w_{26} = w_{13} : (x, y) \mapsto (x, -y)$. Their coordinates are $(\alpha_j, 0)$, $j = 1, \dots, 6$, where α_j are the zeros of $-2197x^6/3 - 362x^4 - 55x^2 - 8/3$.

The equations of the other two curves are obtained in the same way. □

Lemma 20. *Let* $y^2 = -2197x^6/3 - 362x^4 - 55x^2 - 8/3$ *be the equation for* $X_0^{26}(3)/\langle w_2\rangle$ *given in the previous lemma. Then the coordinates of the four CM-points of discriminant* -8 *are* $(\pm 1/2\sqrt{-2}, \pm 3/16\sqrt{-2})$.

Proof. By Lemma 19, an equation for $X_0^{26}(3)/\langle w_2\rangle$ is $y^2 = -2197x^6/3 - 362x^4 - 55x^2 - 8/3$ with $w_3 : (x, y) \mapsto (-x, y)$ and $w_{13} : (x, y) \mapsto (x, -y)$. Thus, if we let $t_1 = x^2$, then t_1 is a Hauptmodul for $X_0^{26}(3)/W_{26,3}$. Likewise, if we let t_2 be the function x^2 in the equation $y^2 = 2197x^6/72 - 699x^4/8 - 225x^2/8 - 81/8$ for $X_0^{26}(3)/\langle w_6\rangle$, then t_2 is also a Hauptmodul for $X_0^{26}(3)/W_{26,3}$. It follows that $t_1 = (at_2 + b)/(ct_2 + d)$ for some a, b, c, d.

Now observe that the values of t_1 and t_2 at the CM-point of discriminant -312 are both ∞. Thus, $t_1 = at_2 + b$ for some a and b. The values of t_1 and t_2 at the CM-points of discriminant -104 are the zeros of $f_1(z) = -2197z^3/3 - 362z^2 - 55z - 8/3$ and $f_2(z) = 2197z^3/72 - 699z^2/8 - 225z/8 - 81/8$, respectively. Therefore, the constants a and b must satisfy $f_1(az+b) = Af_2(z)$ for some constant A. Comparing the coefficients, we find $A = 1/576$, $a = -1/24$ and $b = -1/8$. Since the value of t_2 at the CM-point of discriminant -8 is 0, the value of t_1 at the same point is $-1/8$, which implies that the four CM-points of discriminant -8 on $X_0^{26}(3)/\langle w_2\rangle$ have coordinates $(\pm 1/(2\sqrt{-2}), \pm 3/(16\sqrt{-2}))$ on the equation $y^2 = -2197x^6/3 - 362x^4 - 55x^2 - 8/3$ for $X_0^{26}(3)/\langle w_2\rangle$. □

Lemma 21. *There is a Hauptmodul* t_1 *for* $X_0^{26}(1)/W_{26}$ *that takes values* ∞, 0, *and the three zeros of* $-2x^3 + 19x^2 - 24x - 169$ *at the CM-point of discriminant* -8, *the CM-point of discriminant* -52, *and three CM-points of discriminant* -104, *respectively. Also, there is a Hauptmodul* $t = t_3$ *for* $X_0^{26}(3)/W_{26}$ *that takes values* $\pm 1/(2\sqrt{-2})$ *and the six zeros of* $-2197x^6/3 - 362x^4 - 55x^2 - 8/3$ *at the two CM-points of discriminant* -8 *and the six CM-points of discriminant* -104, *respectively. Moreover, the relation between* t_1 *and* t *and the action of* w_3 *on* t *are given by*

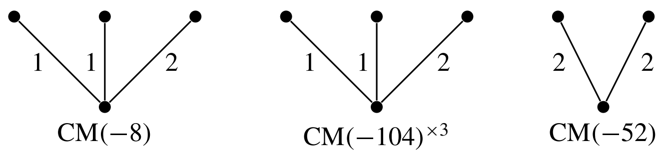
$$t_1 = -\frac{3(1 + t + 10t^2)^2}{(1 + 8t^2)(1 - t)^2}, \quad w_3 : t \mapsto -t.$$

Proof. According to Theorem 3.1 of [González and Rotger 2004], an equation for $X_0^{26}(1)$ is $y^2 = -2x^6 + 19x^4 - 24x^2 - 169$. In fact, the method used in that paper to deduce this equation also shows that the Atkin–Lehner involutions act by $w_{13} : (x, y) \mapsto (-x, y)$ and $w_{26} : (x, y) \mapsto (x, -y)$. Then the two points $(0, \pm 13\sqrt{-1})$ are the CM-points of discriminant -52, the two points at infinity are the fixed points of $w_2 : (x, y) \mapsto (-x, -y)$, that is, the two CM-points of discriminant -8, and the six points $(\alpha_j, 0)$, $j = 1, \dots, 6$, are the six CM-points of discriminant -104, where α_j are the zeros of $-2x^6 + 19x^4 - 24x^2 - 169$. Thus, $t_1 = x^2$ is a Hauptmodul of $X_0^{26}(1)/W_{26}$ with values ∞, 0, the zeros of

$-2x^3+19x^2-24x-169$ at the CM-point of discriminant -8, the CM-point of discriminant -52, and the three CM-points of discriminant -104 on $X_0^{26}(1)/W_{26}$.

On the other hand, Lemmas 19 and 20 show that if we let t be the x in the equation $y^2=-2197x^6/3-362x^4-55x^2-8/3$ for $X_0^{26}(3)/\langle w_2\rangle$, then t is a Hauptmodul for $X_0^{26}(3)/W_{26}$ that takes values $\pm 1/(2\sqrt{-2})$ at the two CM-points of discriminant -8 and β_j, $j=1,\dots,6$, at the six CM-points of discriminant -104, where β_j are the six zeros of $-2197x^6/3-362x^4-55x^2-8/3$. It is clear that w_3 acts on t by $w_3: t\mapsto -t$.

The relation between t_1 and t is simple to determine. From the table at the end of Section 3, we know that the covering $X_0^{26}(3)/W_{26}\to X_0^{26}(1)/W_{26}$ is ramified precisely at the CM-points of discriminants -8, -52, and -104 of $X_0^{26}(1)/W_{26}$ with ramification types given by



It follows that

$$t_1=\frac{A(1+a_1t+a_2t^2)^2}{(1+8t^2)(1+bt)^2}$$

for some constants A, a_1, a_2, and b such that

$$\begin{aligned}&-2f^3+19f^2g-24fg^2-169g^3\\&\qquad=B(-2197t^6/3-362t^4-55t^2-8/3)(1+c_1t+c_2t^2+c_3t^3)^2\end{aligned}$$

for some constants B, c_1, c_2, and c_3, where $f=A(1+8t^2)(1+at)^2$ and $g=(1+b_1t+b_2t^2)^2$. Comparing the coefficients, we find

$$t_1=-\frac{3(1+t+10t^2)^2}{(1+8t^2)(1-t)^2}\quad\text{or}\quad t_1=-\frac{3(1-t+10t^2)^2}{(1+8t^2)(1+t)^2}.$$

Both are valid, since the action of w_3 sends one to the other. This gives us the lemma. $\square$

Case $D=35$.

Lemma 22. *An equation for $X_0^{35}(1)/\langle w_5\rangle$ is*

$$y^2=-(x+12)(7x+4)(x^3+4x^2+144x+80)$$

with the action $w_7=w_{35}$ given by $w_7:(x,y)\mapsto(x,-y)$. The coordinates of the CM-points of discriminants -7, -28, -35, and -140 are $(-12,0)$, $(-4/7,0)$, ∞, and $(\alpha_j,0)$, respectively, where α_j are the three roots of $x^3+4x^2+144x+80$.

An equation for $X_0^{35}(2)/\langle w_7\rangle$ *is*

$$-2y^2=(x^3+3x^2+11x+25)(x^3-3x^2+11x-25)$$

with the actions of $w_2=w_{14}$ *and* $w_5=w_{35}$ *given by* $w_2:(x,y)\mapsto(-x,-y)$ *and* $w_5:(x,y)\mapsto(x,-y)$. *The coordinates of the CM-points of discriminants* -7, -8, -140, *and* -280 *are* $(\pm\sqrt{-7},\pm 8)$, *two points at* ∞, $(\beta_j,0)$, $j=1,\dots,6$, *and* $(0,\pm 25/\sqrt{-2})$, *respectively, where* β_j *are the six roots of the polynomial* $(x^3+3x^2+11x+25)(x^3-3x^2+11x-25)$.

Proof. In Section 10.4 of [2012], Molina showed that an equation for $X_0^{35}(1)/\langle w_5\rangle$ is

$$y^2=-x(9x+4)(4x+1)(172x^3+176x^2+60x+7),$$

where $w_7:(x,y)\mapsto(x,-y)$ and the points $(0,0)$, $(-4/9,0)$, $(-1/4,0)$, and $(\gamma_j,0)$, $j=1,\dots,3$, are the CM-points of discriminant -7, -28, -35, and -140, respectively. Here γ_j are the zeros of $172x^3+176x^2+60x+7$. Setting

$$(x,y)=\left(-\frac{x'+12}{4x'+28},\frac{5y'}{16(x'+7)^3}\right),$$

we get the equation in our lemma. The reason for this change of variable is the Shimura curve $X_0^{35}(1)/\langle w_7\rangle$ has genus 1 and the unique CM-point of discriminant -35 is a $\mathbb{Q}$-rational point. Thus, it is an elliptic curve over $\mathbb{Q}$. Computing the singular fibers at primes of bad reduction, we find that it is isomorphic to the elliptic curve 35A1, which, after a change of variables, has an equation $y^2=x^3+4x^2+144x+80$. If we choose a Weierstrass equation for $X_0^{35}(1)/\langle w_7\rangle$ by requiring that the CM-point of discriminant -35 is the point at infinity and that w_5 acts by $(x,y)\to(x,-y)$, then up to a transformation of the form $x\to ax+b$, this Weierstrass equation must be $y^2=x^3+4x^2+144x+80$ and the three 2-torsion points $(\alpha_j,0)$ must be the three CM-points of discriminant -140. In view of this equation for $X_0^{35}(1)/\langle w_7\rangle$, we make the above change of variables for $X_0^{35}(1)/\langle w_5\rangle$.

We now consider the Shimura curve $X_0^{35}(2)/\langle w_7\rangle$. It is bielliptic with elliptic quotients $C_1:X_0^{35}(2)/\langle w_7,w_{10}\rangle$ and $C_2:X_0^{35}(2)/\langle w_2,w_7\rangle$. Here C_1 is an elliptic curve over $\mathbb{Q}$ because it has a unique CM-point of discriminant -8 and another two $\mathbb{Q}$-rational point coming from CM(-7). Likewise, C_2 is an elliptic curve over $\mathbb{Q}$ because C_2 has a unique CM-point of discriminant -280. By considering the eigenvalues of the Atkin–Lehner involutions associated to the eigenforms in $S_2(\Gamma_0(70))^{35\text{-new}}$, we find that both C_1 and C_2 fall in the isogeny class 35A, in Cremona's notation. Furthermore, by considering its singular fibers at primes of bad reduction using the Cerednik–Drinfeld theory, we find that C_1 is isomorphic to the elliptic curve 35A3 and C_2 is isomorphic to 35A2. We take $y^2=x^3-1728x+30672$

and $y^2 = x^3 - 170208x - 28273968$ to be (nonminimal) equations for 35A3 and 35A2, respectively.

Now if we choose a Weierstrass equation for C_1 by requiring that the CM-point of discriminant -8 is the infinity point and that the Atkin–Lehner involution w_2 acts by $(x, y) \mapsto (x, -y)$, then by a suitable transformation $x \mapsto ax + b$, the equation must be $y^2 = x^3 - 1728x + 30672$. Similarly, if we put the CM-point of discriminant -280 at infinity and require that w_5 acts by $(x, y) \mapsto (x, -y)$, then an equation for C_2 is $y^2 = x^3 - 170208x - 28273968$. Applying Lemma 17, we find an equation for $X_0^{35}(2)/\langle w_7\rangle$ is

$$y^2 = -\tfrac{9}{2}(x^6 + 13x^4 - 29x^2 - 625) = -\tfrac{9}{2}(x^3 + 3x^2 + 11x + 25)(x^3 - 3x^2 + 11x - 25).$$

Replacing y by $3y$, we get the equation

$$-2y^2 = (x^3 + 3x^2 + 11x + 25)(x^3 - 3x^2 + 11x - 25) \tag{2}$$

as claimed in the lemma. According to Lemma 17, the Atkin–Lehner involutions act by

$$w_{10} : (x, y) \mapsto (-x, y), \quad w_5 : (x, y) \mapsto (x, -y), \quad w_2 : (x, y) \mapsto (-x, -y).$$

Since the CM-points of discriminant -8, -140, and -280 on $X_0^{35}(2)/\langle w_7\rangle$ are fixed points of w_2, w_5, and w_{10}, respectively, we find that their coordinates are the two points at infinity, $(\beta_j, 0)$, $j = 1, \ldots, 6$, and $(0, \pm 25/\sqrt{-2})$, respectively, where β_j are the zeros of the polynomial on the right-hand side of (2).

To determine the coordinates of the four CM-points of discriminant -7, we observe that the curve $C_1 : X_0^{35}(2)/\langle w_7, w_{10}\rangle$ has exactly three $\mathbb{Q}$-rational points since it is isomorphic to the elliptic curve 35A3, which has precisely three $\mathbb{Q}$-rational points. Since we already know that C_1 has three $\mathbb{Q}$-rational points consisting of CM(-8) and CM(-7), any $\mathbb{Q}$-rational point of C_1 that is the CM-point of discriminant -8 will be a CM-point of discriminant -7. From the model $-2y^2 = x^6 + 13x^4 - 29x^2 - 625$ for $X_0^{35}(2)/\langle w_7\rangle$, we see that $-2y^2 = x^3 + 13x^2 - 29x - 625$ is also an equation for $X_0^{35}/\langle w_7, w_{10}\rangle$. On this model, the point at infinity is the CM-point of discriminant -8. Thus, the 3-torsion points $(-7, \pm 8)$ are the CM-points of discriminant -7 on $X_0^{35}(2)/\langle w_7, w_{10}\rangle$. This in turn implies that the four CM-points of discriminant -7 on $X_0^{35}(2)/\langle w_7\rangle$ have coordinates $(\pm\sqrt{-7}, \pm 8)$. This completes the proof of the lemma. $\square$

Lemma 23. *There is a Hauptmodul t_1 for $X_0^{35}(1)/W_{35}$ that takes values -12, $-4/7$, ∞, and the three zeros of $x^3 + 4x^2 + 144x + 80$ at the CM-points of discriminants -7, -28, -35, and -140, respectively. Also, there is also a Hauptmodul t for $X_0^{35}(2)/W_{35}$ that takes values $\pm\sqrt{-7}$, ± 5, the six zeros of*

$$(x^3 + 3x^2 + 11x + 25)(x^3 - 3x^2 + 11x - 25),$$

and 0 *at the CM-points of discriminants* -7, -8, -140, *and* -280, *respectively. Moreover*, *the relation between* t_1 *and* t *is*

$$t_1 = -\frac{2(t-1)(t^2-6t+25)}{t^3+3t^2+11t+25}$$

and the Atkin–Lehner involution w_2 *on* t *is given by* $w_2 : t \mapsto -t$.

Proof. The existence of Hauptmoduln with the described values at CM-points follows immediately from Lemma 22. The fact that w_2 acts on t by $w_2 : t \mapsto -t$ also follows from the same lemma. We now determine the relation between Hauptmoduln.

The CM-point of discriminant -35 on $X_0^{35}(1)/W_{35}$ splits completely in the covering $X_0^{35}(2)/W_{35} \to X_0^{35}(1)/W_{35}$ and the three points lying above it are CM-points of discriminant -140 on $X_0^{35}(2)/W_{35}$. Replacing t by $-t$ if necessary, we may assume that the coordinates of these three points are the three zeros of $x^3+3x^2+11x+25$. Considering CM-points of discriminant -7, we have

$$t_1 + 12 = \frac{A(t^2+7)(t-a)}{t^3+3t^2+11t+25} \tag{3}$$

for some constants A and a. The point $t = a$ is a CM-point of discriminant -28. Thus, the point $t = -a$ is the other CM-point of discriminant -28 and this point lies above the CM-point of discriminant -28 on $X_0^{35}(1)/W_{35}$. Therefore, we have

$$t_1 + \tfrac{4}{7} = \frac{B(t+a)(t-b)^2}{t^3+3t^2+11t+25} \tag{4}$$

for some constants B and b. Comparing (3) and (4), we find $A = 10$, $B = -10/7$, $a = -5$, and $b = 3$. It follows that

$$t_1 = -\frac{2(t-1)(t^2-6t+25)}{t^3+3t^2+11t+25}.$$

To check the correctness, we observe that the point t with $t^3-3t^2+11t-25$ lies above CM-points of discriminant -140 on $X_0^{35}(1)/W_{35}$. Thus, if we write $t_1^3+4t_1^2+144t_1+80$ as a rational function of t, then $t^3-3t^2+11t-25$ should divide its numerator. Indeed, we find

$$t_1^3+4t_1^2+144t_1+80 = -\frac{200(t^3-t^2+11t-25)(t^3-t^2-5t-35)^2}{(t^3+3t^2+11t+25)^3}$$

as expected. This proves the lemma. □

***Case D* = 39.**

Lemma 24. *An equation for* $X_0^{39}(1)/\langle w_{13}\rangle$ *is*

$$y^2 = -(7x^2+23x+19)(x^2+x+1)$$

with $w_3 = w_{39} : (x, y) \mapsto (x, -y)$. *Moreover, the coordinates of the CM-points of discriminants* -52, -39, *and* -156 *are* $(\pm 2i, \pm\sqrt{13}(3+2i))$, $((-1\pm\sqrt{-3})/2, 0)$, *and* $((-23\pm\sqrt{-3})/14, 0)$, *respectively.*

Proof. By [Molina 2012], an equation for $X_0^{39}(1)$ is

$$y^2 = -(7x^4 + 79x^3 + 311x^2 + 497x + 277)(x^4 + 9x^3 + 29x^2 + 39x + 19)$$

with $w_{39} : (x, y) \mapsto (x, -y)$. Moreover, the coordinates of the CM-points of discriminants -39 and -156 are $(\alpha_j, 0)$ and $(\beta_j, 0)$, $j = 1, \dots, 4$, respectively, where α_j are the zeros of $x^4 + 9x^3 + 29x^2 + 39x + 19$ and β_j are the zeros of $7x^4 + 79x^3 + 311x^2 + 497x + 277$. Substituting x by $x - 2$, we obtain an equation

$$y^2 = -(7x^4 + 23x^3 + 5x^2 - 23x + 7)(x^4 + x^3 - x^2 - x + 1) \tag{5}$$

with smaller coefficients. This hyperelliptic curve has an obvious automorphism $(x, y) \mapsto (-1/x, y/x^4)$. We will show that this is the Atkin–Lehner involution w_{13}.

The Atkin–Lehner w_{13} permutes the CM-points of discriminant -39. It also permutes the CM-points of discriminant -156. Therefore, if w_{13} maps (x, y) to $((ax+b)/(cx+d), Cy/(cx+d)^4)$, then the constants a, b, c, and d must satisfy

$$(cx+d)^4 f_j\left(\frac{ax+b}{cx+d}\right) = C_j f_j(x)$$

for $f_1(x) = 7x^4 + 23x^3 + 5x^2 - 23x + 7$ and $f_2(x) = x^4 + x^3 - x^2 - x - 1$. We find w_{13} maps (x, y) to either $(-1/x, y/x^4)$ or $(-1/x, -y/x^4)$. The latter has no fixed points, so we conclude that w_{13} maps (x, y) to $(-1/x, y/x^4)$.

Now it is easy to show that $Y = y/x^2$ and $X = x - 1/x$ generate the function field of $X_0^{39}(1)/\langle w_{13}\rangle$. The relation between X and Y is also easy to find. It is

$$Y^2 = -(7X^2 + 23X + 19)(X^2 + X + 1), \tag{6}$$

which gives us an equation for $X_0^{39}(1)/\langle w_{13}\rangle$. The coordinates of the CM-points of discriminants -39 and -156 on $X_0^{39}(1)/\langle w_{13}\rangle$ are $((-1\pm\sqrt{-3})/2, 0)$ and $((-23\pm\sqrt{-3})/14, 0)$, respectively.

To find the coordinates of the CM-points of discriminant -52 on $X_0^{39}(1)/\langle w_{13}\rangle$, we first consider the CM-points of the same discriminant on $X_0^{39}(1)$. Since these points on $X_0^{39}(1)$ are the fixed points of w_{13} and on (5), the Atkin–Lehner involution w_{13} acts by $(x, y) \mapsto (-1/x, y/x^4)$, we find that the coordinates of the CM-points of discriminant -52 on (5) are $(\pm i, \pm\sqrt{13}(3+2i))$. This implies that the CM-points of discriminant -52 on $X_0^{39}(1)/\langle 13\rangle$ are $(\pm 2i, \pm\sqrt{13}(3+2i))$. The proof of the lemma is complete. $\square$

Lemma 25. *There is a Hauptmodul t_1 on $X_0^{39}(1)/W_{39}$ that takes values*

$$\pm 2i, \quad \frac{-1\pm\sqrt{-3}}{2}, \quad \frac{-23\pm\sqrt{-3}}{14}$$

at the CM-points of discriminants -52, -39, *and* -156, *respectively. Also, there is a Hauptmodul t on $X_0^{39}(2)/W_{39}$ that takes values*

$$\pm 3i, \quad \frac{\pm 2\sqrt{-3}\pm\sqrt{-39}}{3}, \quad \pm 1\pm 2\sqrt{-3}$$

at the CM-points of discriminants -52, -39, *and* -156, *respectively. Moreover, the relation between t_1 and t is*

$$t_1 = -\frac{2(t^3+t^2+11t+3)}{(t^2+7)(t+3)}$$

and the Atkin–Lehner involution w_2 on t is $w_2 : t \mapsto -t$.

Proof. The existence of t_1 with the described properties follows from the previous lemma. Now let $s_1 = (t_1-2i)/(t_1+2i)$ so that s_1 takes values 0 and ∞ at the two CM-points of discriminant -52. Then the values of s_1 at the two CM-points of discriminant -156 are the zeros of

$$(9+46i)x^2+94x+(9-46i). \tag{7}$$

The covering $X_0^{39}(2)/W_{39} \to X_0^{39}(1)/W_{39}$ is ramified at $\mathrm{CM}(-52)\cup\mathrm{CM}(-156)$ of $X_0^{39}(1)/W_{39}$. There is a Hauptmodul s of $X_0^{39}(2)/W_{39}$ such that

$$s_1 = \frac{As(1-s)^2}{(1-as)^2}$$

for some complex numbers A and a. That is, s is determined by the property that it takes values 0 and 1 at the two points lying above the point $s_1 = 0$ with the point $s = 1$ having a ramification index 2 and value ∞ at the point lying above $s_1 = \infty$ with ramification index 1.

Now the condition that the CM-points of discriminant -156 are ramified implies that the discriminant of

$$As(1-s)^2 - x(1-as)^2$$

as a polynomial in s must be divisible by the polynomial in (7). This gives two relations between A and a. Solving them for A and a, we find that the only legitimate choice is $A = 9-46i$ and $a = 13$. Then we have

$$t_1 = \frac{2i(s_1+1)}{-s_1+1} = \frac{4394is^3+(-15548-5746i)s^2+(2392+3926i)s-92+18i}{(13s-3+2i)(-169s^2+(416+624i)s+5-12i)}.$$

Let t be the Hauptmodul of $X_0^{39}(2)/W_{39}$ with

$$s=-\frac{3+2i}{13}\frac{(5+i)t+3-15i}{(5-i)t+3+15i}.$$

Then we have

$$t_1=-\frac{2(t^3+t^2+11t+3)}{(t+3)(t^2+7)}.$$

The values of t at CM(−52), CM(−39), and CM(−156) can be read off from

$$t_1^2+4=\frac{8(t^2+9)(t^2+2t+5)^2}{(t+3)^2(t^2+7)^2},$$

$$t_1^2+t_1+1=\frac{(t^2+2t+13)(3t^4+34t^2+27)}{(t+3)^2(t^2+7)^2},$$

and

$$7t_1^2+23t_1+19=\frac{(t^2-2t+13)(t^2-6t+21)^2}{(t+3)^2(t^2+7)^2},$$

respectively. To determine the action of w_2 on t, we recall that w_2 switches the two points in CM(−52). It also exchanges the two zeros of $x^2+2x+13$, corresponding to the two points in CM(−156) that lie above the CM-points of discriminant −39 on $X_0^{39}(1)/W_{39}$, with the two zeros of $x^2-2x+13$, corresponding to the other two points in CM(−156) that lie above the CM-points of discriminant −156 on $X_0^{39}(1)/W_{39}$. From this information, we can deduce that $w_2: t\mapsto -t$. □

5. Main results

5.1. *Schwarzian differential equations.*

Theorem. *Let Hauptmoduln for $X_0^D(N)/W_D$ be as in the lemmas. Then the automorphic derivatives associated to them are as follows. For $(D,N)=(6,1)$,*

$$Q(t)=\frac{108-113t+140t^2}{576t^2(1-t)^2}.$$

For $(D,N)=(6,5)$,

$$Q(t)=-\frac{15(23-456t^2+1608t^4)}{2(2+3t^2)^2(1+64t^2)^2}.$$

For $(D,N)=(6,7)$,

$$Q(t)=-\frac{3(267+6480t^2+64352t^4)}{4(1+27t^2)^2(3+32t^2)^2}.$$

For $(D,N)=(6,13)$,

$$Q(t) = -\frac{3(12492 + 43272t^2 + 37541t^4)}{(9+16t^2)^2(16+27t^2)^2}.$$

For $(D, N) = (10, 1)$,

$$Q(t) = \frac{3t^4 - 119t^3 + 3157t^2 - 7296t + 10368}{16t^2(t-2)^2(t-27)^2}.$$

For $(D, N) = (10, 3)$,

$$Q(t) = \frac{8 - 303t^2 - 1200t^4 - 95840t^6}{36t^2(1+32t^2)^2(1+5t^2)^2}.$$

For $(D, N) = (10, 7)$,

$$Q(t) = -\frac{655 + 62410t^2 + 2237231t^4 + 35817920t^6 + 216522240t^8}{(1+27t^2)^2(1+20t^2)^2(5+128t^2)^2}.$$

For $(D, N) = (14, 1)$,

$$Q(t) = \frac{192 + 440t + 43t^2 + 1036t^3 + 960t^4}{16t^2(8+13t+16t^2)^2}.$$

For $(D, N) = (14, 3)$,

$$Q(t) = -\frac{3(497 - 1988t^2 + 31494t^4 + 141436t^6 + 139601t^8)}{2(1+2t^2)^2(7+226t^2+343t^4)^2}.$$

For $(D, N) = (14, 5)$,

$$Q(t) = -\frac{623 + 16772t^2 + 55178t^4 - 853468t^6 + 97503t^8}{(1+16t^2)^2(7+114t^2+7t^4)^2}.$$

For $(D, N) = (15, 1)$,

$$Q(t) = \frac{177147 - 244944t + 244242t^2 - 3680t^3 + 35t^4}{144t^2(1-t)^2(81-t)^2}.$$

For $(D, N) = (15, 2)$,

$$Q(t) = \frac{3(385 + 5500t^2 - 2042t^4 + 35196t^6 - 2175t^8)}{4(1-t)^2(1+t)^2(1-5t)^2(1+5t)^2(5+3t^2)^2}.$$

For $(D, N) = (15, 4)$,

$$Q(t) = -\frac{9(14 + 271t^2 + 2024t^4 + 7746t^6 + 19895t^{10} + 16674t^8 + 10720t^{12})}{4(4t^2-t+1)^2(4t^2+t+1)^2(3t^2+1)^2(5t^2+3)^2}.$$

For $(D, N) = (21, 1)$,

$$Q(t) = \frac{21(40353607 - 17647350t + 3561369t^2 - 477652t^3 + 31833t^4 - 630t^5 + 7t^6)}{16t^2(49-t)^2(343-94t+7t^2)^2}.$$

For $(D, N) = (21, 2)$,

$$Q(t) = \frac{3(1 - 69t^2 - 4086t^4 + 23670t^6 + 6043653t^8 + 6781887t^{10})}{16t^2(1-t)^2(1+t)^2(1+27t^2)^2(1+63t^2)^2}.$$

For $(D, N) = (26, 1)$,

$$Q(t) = \frac{85683 + 15210t + 16694t^2 - 9480t^3 + 1363t^4 - 170t^5 + 12t^6}{16t^2(169+24t-19t^2+2t^3)^2}.$$

For $(D, N) = (26, 3)$,

$$Q(t) = -\frac{6(85+3528t^2+60543t^4+552448t^6+2850579t^8+7990200t^{10}+9677785t^{12})}{(1+8t^2)^2(8+165t^2+1086t^4+2197t^6)^2}.$$

For $(D, N) = (35, 1)$,

$$Q(t) = Q_1(t)/16(t+12)^2(7t+4)^2(t^3+4t^2+144t+80)^2,$$

where

$$Q_1(t) = 666427392t + 1132800t^4 + 181420032 - 753984t^5 + 24576t^6 + 147t^8 + 659096576t^2 + 85540864t^3 + 3808t^7.$$

For $(D, N) = (35, 2)$,

$$Q(t) = Q_1(t)/4(t^2+7)^2(t^2-25)^2(t^6+13t^4-29t^2-625)^2,$$

where

$$Q_1(t) = 2842805000t^2 + 91524600t^6 - 2082286t^8 - 217416t^{10} + 54644t^{12} + 3784t^{14} + 19t^{16} - 992578125 + 1017474100t^4.$$

For $(D, N) = (39, 1)$,

$$Q(t) = \frac{-3Q_1(t)}{4(4+t^2)^2(1+t+t^2)^2(19+23t+7t^2)^2},$$

where

$$Q_1(t) = 2596 + 7104t + 9692t^2 + 12348t^3 + 13149t^4 + 9522t^5 + 4367t^6 + 1086t^7 + 97t^8.$$

For $(D, N) = (39, 2)$,

$$Q(t) = \frac{-9Q_1(t)}{4(9+t^2)^2(13+2t+t^2)^2(13-2t+t^2)^2(27+34t^2+3t^4)^2},$$

where

$$Q_1(t) = 419253003 + 119984328t^2 + 89200020t^4 + 43676088t^6 + 10194786t^8 + 1272824t^{10} + 87380t^{12} + 3080t^{14} + 43t^{16}.$$

For these results, we take the Schwarzian differential equations on $X_0^{14}(1)/W_{14}$, $X_0^{14}(3)/W_{14}$, and $X_0^{14}(5)/W_{14}$ as examples for the proofs.

Proof. In Lemma 12, we see that there is a Hauptmodul t_1 on $X_0^{14}(1)/W_{14}$ with value ∞ at the elliptic point of order 4 and values 0 and $(-13\pm 7\sqrt{-7})/32$ at the elliptic points of order 2. According to Proposition 3, the automorphic derivative $Q(t_1)$ associate to t_1 is

$$Q(t_1) = \frac{3}{16} - \frac{21+16B}{52t} + \frac{3(512t^2+416t-87)}{(16t^2+13t+8)^2} + \frac{4(21t+B(16t+13))}{13(16t^2+13t+8)},$$

for some constant B. We now use the covering $X_0^{14}(3)/W_{14} \to X_0^{14}(1)/W_{14}$ to determine the constant B. More precisely, according to Proposition 4, we have the relation between $Q(t_1)$ and the automorphic derivative $Q(t)$ associative to a Hauptmodul t of $X_0^{14}(3)/W_{14}$,

$$Q(t) = D(t_1, t) + Q(t_1)/(dt_1/dt)^2.$$

Note that there is a Hauptmodul t for $X_0^{14}(3)/W_{14}$ that takes values $\pm 1/\sqrt{-2}$, $(\pm 9\sqrt{-7}\pm 4\sqrt{-14})/49$ at the 6 elliptic points of order 6. Thus, the automorphic derivative $Q(t)$ is

$$Q(t) = \frac{3(2t^2-1)}{4(2t^2+1)^2} + \frac{3(18335t^2+38759t^4+117649t^6-791)}{4(7+226t^2+343t^4)^2} + \frac{343(686C_4t^3+109C_3t^2+109C_4t+109C_5)}{436(7+226t^2+343t^4)} - \frac{1372C_4t+981+218C_3}{436(2t^2+1)}$$

for some constants C_3, C_4, and C_5. Also, the action of the Atkin–Lehner involution w_3 is $w_3 : t \mapsto -t$. Thus, by Proposition 5, we can get the value $C_4 = 0$.

From the relations

$$t_1 = \frac{4(1+2t^2)(1-5t)^2}{9(1+t)^4} \quad \text{and} \quad Q(t) = D(t_1, t) + \frac{Q(t_1)}{(dt_1/dt)^2}$$

we find that

$$B = -\tfrac{373}{512}, \; C_3 = -\tfrac{91}{9}, \; \text{and } C_5 = -\tfrac{1301}{3087}.$$

For the case of $X_0^{14}(5)/X_{14}$, the chosen Hauptmodul t takes values $\pm i/4$ at the elliptic points of order 4, $(\pm 5\sqrt{-7}\pm 4\sqrt{-14})/7$ at the elliptic points of order 2, and the action of Atkin–Lehner involution w_5 is $t \mapsto -t$. Therefore, the automorphic derivative associative to t is

$$Q(t) = \frac{15(16t^2-1)}{2(16t^2+1)^2} + \frac{3(49t^6+399t^4+6351t^2-399)}{4(7t^4+114t^2+7)^2} - \frac{39+8B_1}{2(16t^2+1)} + \frac{7(B_1t^2+B_2)}{4(7t^4+114t^2+7)},$$

for some constants B_1 and B_2. From the relation

$$t_1 = -\frac{5(1-t+17t^2-13t^3)^2}{(1+16t^2)(1+3t)^4}$$

and Proposition 4, we can conclude that

$$Q(t) = -\frac{97503t^8 - 853468t^6 + 55178t^4 + 16772t^2 + 623}{(16t^2+1)^2(7t^4+114t^2+7)^2}. \qquad \square$$

5.2. ***Ramanujan-type formulae.*** Recall that if E is an elliptic curve defined over $\overline{\mathbb{Q}}$, which has CM by an imaginary quadratic field K of discriminant d, then up to an algebraic factor, the period of E can be expressed by

$$\Omega_d = \sqrt{\pi} \prod_{0<a<|d|} \Gamma\left(\frac{a}{|d|}\right)^{w_d \chi_d(a)/4h_d},$$

where w_d is the number of roots of unity in K, χ_d is the Kronecker character $\left(\frac{d}{\cdot}\right)$ associated to K, and h_d is the class number of K. Yang [2013a] contributes many Ramanujan-type series. For example,

$$\sum_{n=0}^{\infty} \left(74480n + \tfrac{6860}{3}\right) \frac{(1/12)_n(1/4)_n(5/12)_n}{(1/2)_n(3/4)_n n!} \left(\frac{-7^4}{3375}\right)^n = 7^3\sqrt{5}\sqrt[4]{3375}\frac{4\pi}{\sqrt[4]{12}\Omega_{-4}^2},$$

which is related to the period of an elliptic curve with CM by $\mathbb{Q}(\sqrt{-1})$. The power series

$$\sum_{n=0}^{\infty} \frac{(1/12)_n(1/4)_n(5/12)_n}{(1/2)_n(3/4)_n n!} t^n$$

mentioned above is the hypergeometric function

$${}_3F_2\left(\tfrac{1}{12}, \tfrac{1}{4}, \tfrac{5}{12}; \tfrac{1}{2}, \tfrac{3}{4}, t\right) = {}_2F_1\left(\tfrac{1}{24}, \tfrac{5}{24}; \tfrac{3}{4}; t\right)^2.$$

Note that the function ${}_2F_1\left(\tfrac{1}{24}, \tfrac{5}{24}; \tfrac{3}{4}; t\right)$ is related to the Schwarzian differential equation associated to the Hauptmodul t of $X_0^6(1)/W_6$ that takes values 0, 1, and ∞ at the CM-points of discriminants -4, -24, and -3, respectively. Yang also gave other similar identities related to Ω_{-4}, and also the Ramanujan-type series related to Ω_{-3} for the curve $X_0^6(1)/W_6$.

Yang [2013a] guesses that, in general, we can use the t-series expansion of a meromorphic form to obtain Ramanujan-type identities, which are related to certain periods of elliptic curves with CM. That is, we may have

$$\sum_{n=0}^{\infty}(R_1 n+R_2)A_n t_0^n = R_3\frac{\pi}{\Omega_d^2},$$

where R_1, R_2, $R_3 \in \overline{\mathbb{Q}}$, $\sum_0^\infty A_n t^n$ is the expansion of a meromorphic automorphic form of weight 2 with respect to a Hauptmodul t of a Shimura curve of genus zero such that t takes value 0 at a CM-point of discriminant d, and t_0 is the value of t at some CM-point of discriminant $d' \neq d$. To be more precise, let g_1 and g_2 be two linearly independent solutions of a given Schwarzian differential equation associated to a Shimura curve of genus 0. Write $g_1^2 = \sum_0^\infty A_n t^n$ and $g_2^2 = \sum_0^\infty B_n t^n$; then we expect that

$$\sum_{n=0}^{\infty}(R_1 n+R_2)\, A_n t_0^n = R_3\frac{\pi}{\Omega_d^2},$$
$$\sum_{n=0}^{\infty}(R_1 n+R_2+R_1/a)\, B_n t_0^n = R_3\frac{\Omega_d^2}{\pi},$$

for certain positive integer a. We remark that the series also converges p-adically for primes $p \mid M$ while $t_0 = M/N$. The p-adic numbers to which they converge should be related to the p-adic periods of certain elliptic curves with CM. Yang also gave some numerical examples of the p-adic analogues for the Ramanujan-type series obtained from $X_0^6(1)/W_6$. Here, let us see some numerical examples coming from $X_0^{14}(1)/W_{14}$.

From the Lemma 12, we know that there is a Hauptmodul t for $X_0^{14}(1)/W_{14}$ that takes values ∞, 0, and $(-13 \pm 7\sqrt{-7})/32$ at CM-points of discriminants -4, -8, and -56, respectively. The t-series expansions of two linearly independent solutions of the Schwarzian differential equation associated to t (see Theorem),

$$\frac{d^2}{dt^2}f + Q(t)f = 0, \quad Q(t) = \frac{192+440t+43t^2+1036t^3+960t^4}{16t^2(8+13t+16t^2)^2},$$

are

$$g_1 = t^{1/4}\left(1+\tfrac{23}{64}t+\tfrac{1867}{8192}t^2-\tfrac{955937}{2621440}t^3+\tfrac{157030847}{671088640}t^4+\tfrac{3694251053}{42949672960}t^5+\dots\right) \text{ and}$$
$$g_2 = t^{3/4}\left(1+\tfrac{23}{192}t+\tfrac{3149}{24576}t^2-\tfrac{434593}{1572864}t^3+\tfrac{264972083}{1207959552}t^4+\tfrac{39014127761}{850403524608}t^5+\dots\right).$$

The Hauptmodul t takes value $t_0 = -13/81$ at the CM-points of discriminants -91 (this is given in [Elkies 1998]). We now let

$$\sum_{n=0}^{\infty} A_n = t^{-1/2}g_1^2, \quad \sum_{n=0}^{\infty} B_n = t^{-3/2}g_2^2,$$

and

$$C = \frac{81}{2548}\frac{\Gamma(5/8)\Gamma(7/8)}{\Gamma(1/8)\Gamma(3/8)} = \frac{81}{2548}\Omega_{-8}^2/\pi.$$

Then

$$\left(\sum_{n=0}^{\infty} R_1 n + R_2\right) A_n t_0^n = \tfrac{847}{18} 13^{3/4} 3C, \tag{8}$$

$$\left(\sum_{n=0} \infty R_1 n + R_1 + R_2\right) B_n t_0^n = \tfrac{847}{18} 13^{1/4} 27C^{-1}. \tag{9}$$

If we choose a Hauptmodul t that takes values 0, ∞, and $(-39 \pm 21\sqrt{-7})/16$ at CM-points of discriminant -4, -8, and -56, respectively, the Schwarzian differential equation associated to t is given by

$$\frac{d^2}{dt^2} f + Q(t) f = 0, \quad Q(t) = \frac{3(64t^4 + 440t^3 + 129t^2 + 9324t + 25920)}{16t^2(8t^2 + 39t + 144)^2},$$

and its two linearly independent solutions are

$$g_1 = t^{3/8}\left(1 + \tfrac{131}{2304}t + \tfrac{21631}{3538944}t^2 - \tfrac{49745249}{29896998912}t^3 + \tfrac{16603576771}{91843580657664}t^4 + \dots\right),$$
$$g_2 = t^{5/8}\left(1 + \tfrac{131}{3840}t + \tfrac{8923}{1966080}t^2 - \tfrac{257758957}{176664084480}t^3 + \tfrac{646181570409}{9226105147883520}t^4 + \dots\right).$$

The Hauptmodul t takes value $t_0 = 27/200$ at the CM-points of discriminants -168. Let

$$\sum_{n=0}^{\infty} C_n = t^{-3/4} g_1^2, \quad \sum_{n=0}^{\infty} D_n = t^{-5/4} g_2^2.$$

We have

$$\sum_{n=0}^{\infty} (R_1 n + R_2)\, C_n t_0^n = \frac{810000}{11^8} 27^{3/4} 200^{1/4} C,$$
$$\sum_{n=0}^{\infty} (R_1 n + R_2 + R_1/2)\, D_n t_0^n = \frac{810000}{11^8} 27^{1/4} 200^{3/4} C^{-1}$$

with $R_1 = 2904$, $R_2 = 12$, where

$$C = \frac{\Gamma(3/4)^2}{\Gamma(1/4)^2}\left(\frac{196}{3}\right)^{1/4} = \left(\frac{196}{3}\right)^{1/4} \Omega_{-4}^2/\pi.$$

Let $\Gamma_p(\cdot)$ stand for the p-adic Gamma function. The numerical results checked for 70 p-adic digits yield that

$$\sum_{n=0}^{\infty} (R_1 n + R_2)\, C_n t_0^n = \frac{2^4 \cdot 11^8}{9} \left(27^3 200 \frac{98\Gamma_3(1/4)}{27\Gamma_3(3/4)} \right)^{1/4},$$

$$\sum_{n=0}^{\infty} (R_1 n + R_2 + R_1/2)\, D_n t_0^n = \frac{2^4 \cdot 11^8}{9} \left(27 \cdot 200^3 \frac{27\Gamma_3(3/4)}{98\Gamma_3(1/4)} \right)^{1/4},$$

hold 3-adically with $R_1 = 29040$ and $R_2 = 120$.

For the numbers $\sum n A_n t_0^n$, $\sum A_n t_0^n$, $\sum n B_n t_0^n$, and $\sum B_n t_0^n$, after numerical computation, we find that the equalities

$$\left(\sum_{n=0}^{\infty} (11011n + 7290) A_n t_0^n \right)^2 = 3^3 \cdot 7 \cdot 137 \cdot 1571 \frac{\Gamma_{13}(5/8)\Gamma_{13}(7/8)}{2\Gamma_{13}(1/8)\Gamma_{13}(3/8)},$$

$$\left(\sum_{n=0}^{\infty} (11011n + 75897)\, B_n t_0^n \right)^2 = 3^{12} \cdot 7 \cdot 11^4 \frac{\Gamma_{13}(1/8)\Gamma_{13}(3/8)}{8\Gamma_{13}(5/8)\Gamma_{13}(7/8)},$$

hold 13-adically.

Acknowledgements

The author would like to thank Professor Y. Yang of the National Chiao Tung University for his help with the exposition of this paper. The examples of the Ramanujan-type series covering p-adically given in Section 5.2 are his. This work is partially supported by National Center for Theoretical Sciences of Taiwan, and Waseda University of Japan.

References

[Alsina and Bayer 2004] M. Alsina and P. Bayer, *Quaternion orders, quadratic forms, and Shimura curves*, CRM Monograph Series **22**, Amer. Math. Soc., Providence, RI, 2004. MR 2005k:11226 Zbl 1073.11040

[Bayer and Travesa 2007] P. Bayer and A. Travesa, "Uniformizing functions for certain Shimura curves, in the case $D = 6$", *Acta Arith.* **126**:4 (2007), 315–339. MR 2008d:11055 Zbl 1158.11031

[Cremona 1997] J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, 1997. MR 99e:11068 Zbl 0872.14041

[Elkies 1998] N. D. Elkies, "Shimura curve computations", pp. 1–47 in *Algorithmic number theory* (Portland, OR, 1998), edited by J. P. Buhler, Lecture Notes in Comput. Sci. **1423**, Springer, Berlin, 1998. MR 2001a:11099 Zbl 1010.11030

[González and Rotger 2004] J. González and V. Rotger, "Equations of Shimura curves of genus two", *Int. Math. Res. Not.* **2004**:14 (2004), 661–674. MR 2004k:11095 Zbl 1081.11041

[González and Rotger 2006] J. González and V. Rotger, "Non-elliptic Shimura curves of genus one", *J. Math. Soc. Japan* **58**:4 (2006), 927–948. MR 2007k:11093 Zbl 1123.11019

[Katok 1992] S. Katok, *Fuchsian groups*, University of Chicago Press, Chicago, IL, 1992. MR 93d:20088 Zbl 0753.30001

[Molina 2012] S. Molina, "Equations of hyperelliptic Shimura curves", *Proc. Lond. Math. Soc.* (3) **105**:5 (2012), 891–920. MR 2997041 Zbl 06111534

[Ogg 1983] A. P. Ogg, "Real points on Shimura curves", pp. 277–307 in *Arithmetic and geometry, I*, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, Boston, 1983. MR 85m:11034 Zbl 0531.14014

[Selberg 1965] A. Selberg, "On the estimation of Fourier coefficients of modular forms", pp. 1–15 in *Proc. Sympos. Pure Math., VIII*, edited by A. Whiteman, Amer. Math. Soc., Providence, R.I., 1965. MR 32 #93 Zbl 0142.33903

[Shimizu 1965] H. Shimizu, "On zeta functions of quaternion algebras", *Ann. of Math.* (2) **81** (1965), 166–193. MR 30 #1998 Zbl 0201.37903

[Tu and Yang 2013] F.-T. Tu and Y. Yang, "Algebraic transformations of hypergeometric functions and automorphic forms on Shimura curves", *Trans. Amer. Math. Soc.* **365**:12 (2013), 6697–6729. MR 3105767 Zbl 06218200

[Vignéras 1983] M.-F. Vignéras, "Quelques remarques sur la conjecture $\lambda_1 \geq \frac{1}{4}$", pp. 321–343 in *Seminar on number theory* (Paris, 1981/1982), edited by M.-J. Bertin, Progr. Math. **38**, Birkhäuser, Boston, MA, 1983. MR 85c:11049 Zbl 0523.10015

[Yang 2004] Y. Yang, "On differential equations satisfied by modular forms", *Math. Z.* **246**:1-2 (2004), 1–19. MR 2005b:11049 Zbl 1108.11040

[Yang 2013a] Y. Yang, "Ramanujan-type identities for Shimura curves", preprint, 2013. arXiv 1301.3344

[Yang 2013b] Y. Yang, "Schwarzian differential equations and Hecke eigenforms on Shimura curves", *Compos. Math.* **149**:1 (2013), 1–31. MR 3011876 Zbl 06132104

[Zograf 1991] P. Zograf, "A spectral proof of Rademacher's conjecture for congruence subgroups of the modular group", *J. Reine Angew. Math.* **414** (1991), 113–116. MR 92d:11041 Zbl 0709.11031

Received December 1, 2013. Revised March 27, 2014.

FANG-TING TU
NATIONAL CENTER FOR THEORETICAL SCIENCES, MATHEMATICS DIVISION
NATIONAL TSING HUA UNIVERSITY
NO. 101, SECTION 2, KUANG FU ROAD
HSINCHU 30043
TAIWAN

ft12@math.cts.nthu.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 269, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.269.491

POLYNOMIAL INVARIANTS OF WEYL GROUPS FOR KAC–MOODY GROUPS

ZHAO XU-AN AND JIN CHUNHUA

We prove that the ring of polynomial invariants of Weyl group for an indecomposable and indefinite Kac–Moody Lie algebra is generated by the invariant symmetric bilinear form or is trivial depending on whether A is symmetrizable or not. The result was conjectured by Moody and assumed by Kac. As an application, we discuss the rational homotopy types of Kac–Moody groups and their flag manifolds.

1. Introduction

Let $A = (a_{ij})$ be an $n \times n$ integer matrix satisfying:

(1) For each i, $a_{ii} = 2$.

(2) For $i \neq j$, $a_{ij} \leq 0$.

(3) If $a_{ij} = 0$, then $a_{ji} = 0$.

Then A is called a Cartan matrix.

Let h be the real vector space with basis $\Pi^\vee = \{\alpha_1^\vee, \alpha_2^\vee, \dots, \alpha_n^\vee\}$, and denote the dual basis of $\Pi^\vee$ in the dual space h^* by $\{\omega_1, \omega_2, \dots, \omega_n\}$; that is, $\omega_i(\alpha_j^\vee) = \delta_{ij}$ for $1 \leq i, j \leq n$. Let $\Pi = \{\alpha_1, \dots, \alpha_n\} \subset h^*$ be given by $\langle \alpha_i^\vee, \alpha_j \rangle = a_{ij}$ for all i, j; then $\alpha_i = \sum_{j=1}^n a_{ji}\omega_j$. Note that if the Cartan matrix A is singular, then $\{\alpha_i \mid 1 \leq i \leq n\}$ is not a basis of h^*. Π and $\Pi^\vee$ are called the simple root system and simple coroot system associated to the Cartan matrix A, and $\alpha_i, \alpha_i^\vee, \omega_i$, $1 \leq i \leq n$ are the simple roots, simple coroots and fundamental dominant weights respectively.

By [Kac 1968] and [Moody 1968], for each Cartan matrix A, there is a Lie algebra $g(A)$ associated to A, which is called the Kac–Moody Lie algebra.

The Kac–Moody Lie algebra $g(A)$ is generated by $\alpha_i^\vee, e_i, f_i$, $1 \leq i \leq n$ over $\mathbb{C}$, with the defining relations

(1) $[\alpha_i^\vee, \alpha_j^\vee] = 0$,

The authors are supported by National Science Foundation of China, 11171025.
MSC2010: primary 17C99; secondary 55N45.
Keywords: Cartan matrix, Weyl group, polynomial invariants of Weyl group, Kac–Moody Group, flag manifold, cohomology of Kac–Moody groups.

(2) $[e_i, f_j] = \delta_{ij}\alpha_i^\vee$,

(3) $[\alpha_i^\vee, e_j] = a_{ij}e_j, \quad [\alpha_i^\vee, f_j] = -a_{ij}f_j$,

(4) $\mathrm{ad}(e_i)^{-a_{ij}+1}(e_j) = 0,\ 1 \le i \ne j \le n$,

(5) $\mathrm{ad}(f_i)^{-a_{ij}+1}(f_j) = 0,\ 1 \le i \ne j \le n$.

Kac and Peterson [1983; 1985] (see also [Kac 1985a]) constructed the Kac–Moody group $G(A)$ with Lie algebra $g(A)$. In this paper, for convenience we consider the quotient Lie algebra of $g(A)$ modulo its center $c(g(A))$ and the associated simply connected group $G(A)$ modulo $C(G(A))$. We still use the same symbols $g(A)$ and $G(A)$ and call them the Kac–Moody Lie algebra and the Kac–Moody group.

Cartan matrices and their associated Kac–Moody Lie algebras and Kac–Moody groups are divided into three types:

(1) Finite type, if A is positive definite. In this case, $G(A)$ is just the simply connected complex semisimple Lie group with Cartan matrix A.

(2) Affine type, if A is positive semidefinite and has rank $n-1$.

(3) Indefinite type otherwise.

A Cartan matrix A is called hyperbolic if all the proper principal submatrices of A are of finite or affine type. A Cartan matrix A is called symmetrizable if there exist an invertible diagonal matrix D and a symmetric matrix B such that $A = DB$. Also, $g(A)$ is called a symmetrizable Kac–Moody Lie algebra if A is symmetrizable.

The Weyl group $W(A)$ associated to a Cartan matrix A is the group generated by the Weyl reflections $\sigma_i : h^* \to h^*$ with respect to $\alpha_i^\vee$, for $1 \le i \le n$, where $\sigma_i(\alpha) = \alpha - \alpha(\alpha_i^\vee)\alpha_i$. $W(A)$ has a Coxeter presentation

$$W(A) = \langle \sigma_1, \dots, \sigma_n \mid \sigma_i^2 = e, 1 \le i \le n;\ (\sigma_i\sigma_j)^{m_{ij}} = e, 1 \le i < j \le n \rangle,$$

where $m_{ij} = 2, 3, 4, 6$ or ∞ as $a_{ij}a_{ji} = 0, 1, 2, 3$ or ≥ 4, respectively. The action of σ_i on fundamental dominant weights is given by $\sigma_i(\omega_j) = \omega_j - \omega_j(\alpha_i^\vee)\alpha_i = \omega_j - \delta_{ji}\alpha_i$. For details see [Kac 1983; Humphreys 1990].

The action of the Weyl group $W(A)$ on h^* induces an action of $W(A)$ on the polynomial ring $\mathbb{Q}[h^*] \cong \mathbb{Q}[\omega_1, \dots, \omega_n]$. If for each $\sigma \in W(A)$, $\sigma(f) = f$, then $f \in \mathbb{Q}[h^*]$ is called a $W(A)$-invariant polynomial. Since $W(A)$ is generated by σ_i, $1 \le i \le n$, f is a $W(A)$-invariant polynomial if and only if $\sigma_i(f) = f$ for $1 \le i \le n$. The $W(A)$-invariant polynomials form a ring, called the ring of $W(A)$ polynomial invariants, denoted by $I(A)$.

The invariant theory of Weyl groups has been a significant topic since the 1950s. It has important applications in the homology of Lie groups and their classifying spaces. Motivated by that study, Chevalley showed that the ring of invariants of a finite Weyl group is a polynomial algebra. A comprehensive study of the polynomial invariants was undertaken by Bourbaki, Solomon, Springer and Steinberg, etc.

Moody [1978] proved the following:

Theorem. *Let A be an indecomposable and symmetrizable $n \times n$ Cartan matrix whose associated invariant bilinear form ψ is nondegenerate and of signature* $(n-1, 1)$. *Then the ring of $W(A)$ polynomial invariants is* $\mathbb{Q}[\psi]$.

In the same paper, Moody further said: "*We conjecture that it is in fact true for all Weyl groups arising from nonsingular Cartan matrices of nonfinite type*."

Kac [1985b] also assumed that for an indecomposable and indefinite Cartan matrix, the ring of $W(A)$ polynomial invariants is $\mathbb{Q}[\psi]$ or trivial depending on whether A is symmetrizable or not.

In this paper, we prove the following:

Theorem. *Let A be an indecomposable and indefinite Cartan matrix A. If A is symmetrizable*, *then* $I(A) = \mathbb{Q}[\psi]$; *if A is nonsymmetrizable*, *then* $I(A) = \mathbb{Q}$.

The content of this paper is as follows. In Section 2, we discuss the general results about the polynomial invariants of Weyl group $W(A)$. In Sections 3 and 4, we consider the rank 2 and hyperbolic cases, respectively. The main theorem is proved in Section 5. In Section 6, we consider the applications of the theorem in determining the rational homotopy type of Kac–Moody groups and their flag manifolds.

2. Rings of polynomial invariants of Weyl groups: general case

In this section, we discuss some general properties of the ring of invariants of Weyl groups.

Lemma 2.1. *If a Cartan matrix A is the direct sum of Cartan matrices A_1, A_2, then $I(A)$ is isomorphic to the tensor product $I(A_1) \otimes I(A_2)$.*

So we only consider indecomposable Cartan matrices.

Lemma 2.2. *Write $f \in I(A)$ as $f = \sum_{i=0}^{l} f_i \in I(A)$, where f_i is the degree i homogeneous component of f. Then $f_i \in I(A)$.*

So $I(A)$ is a graded ring: $I(A) = \bigoplus_{i=0}^{\infty} I^i(A)$, where $I^i(A)$ is the subspace of homogeneous polynomials of degree i in $I(A)$. To determine the ring $I(A)$, we only need to consider homogeneous invariant polynomials.

Lemma 2.3. *For an indecomposable Cartan matrix A of affine or indefinite type, the orbit $\{\sigma(\omega) \mid \sigma \in W(A)\}$ of a nonzero element ω in the Tits cone $\left\{\sum_{i=1}^{n} \lambda_i \omega_i \mid \lambda_i \geq 0\right\}$ is an infinite set.*

Proof. If $\omega \neq 0$ is in the Tits cone, then we can assume $\omega = \sum_{i=1}^{n} \lambda_i \omega_i$, $\lambda_i \geq 0$. Let $S = \{1, 2, \ldots, n\}$, $J = \{i \in S \mid \lambda_i = 0\}$, $W_J(A)$ be the subgroup of $W(A)$ generated by $\{\sigma_i \mid i \in J\}$ and $G_J(A)$ be the parabolic subgroup of $G(A)$ corresponding to the Weyl group $W_J(A)$. Since $\{\sigma(\omega) \mid \sigma \in W(A)\} \cong W(A)/W_J(A)$ indexes the Schubert varieties of the generalized flag manifold $F_J(A) = G(A)/G_J(A)$, the lemma follows from the fact that the number of Schubert varieties in $F_J(A)$ is infinite for affine and indefinite type. For reference, see [Kumar 2002]. □

Corollary 2.4. *Let $f \in I(A)$ be a homogeneous invariant polynomial and $\omega \neq 0$ be in the Tits cone. If $\omega \mid f$, then $f = 0$.*

Proof. If $\omega \mid f$, then for any $\sigma \in W(A)$, $\sigma(\omega) \mid \sigma(f) = f$. Since $\{\sigma(\omega) \mid \sigma \in W(A)\}$ is an infinite set, if $f \neq 0$, this contradicts the condition that the degree of f is finite. □

Lemma 2.5. *For a Cartan matrix A, $I^1(A) = \{0\}$.*

Proof. Suppose $f = \sum_{i=1}^{n} \lambda_i \omega_i \in I^1(A)$; then for each j, $\sigma_j(f) = f - \lambda_j \alpha_j = f$. Since $\alpha_j \neq 0$, $\lambda_j = 0$. Hence $f = 0$. □

Lemma 2.6. *For a Cartan matrix A, we may write any homogeneous degree 2 polynomial f in the form $\sum_{i,j=1}^{n} \lambda_{ij} \omega_i \omega_j$, where $\lambda_{ij} = \lambda_{ji}$. Then $f \in I^2(A)$ if and only if $\partial f / \partial \omega_j = \frac{1}{2}(\partial^2 f / \partial \omega_j^2) \alpha_j$ for $1 \leq j \leq n$; that is, $2\lambda_{ij} = a_{ij} \lambda_{jj}$ for $1 \leq i, j \leq n$.*

Proof. If f is an invariant polynomial, then for each j,

$$\sigma_j(f) = f(\omega_1, \ldots, \omega_j - \alpha_j, \ldots, \omega_n) = f(\omega) - \frac{\partial f}{\partial \omega_j} \alpha_j + \frac{1}{2} \frac{\partial^2 f}{\partial \omega_j^2} \alpha_j^2 = f.$$

This is equivalent to $\dfrac{\partial f}{\partial \omega_j} = \dfrac{1}{2} \dfrac{\partial^2 f}{\partial \omega_j^2} \alpha_j$. That is, $2 \sum\limits_{i=1}^{n} \lambda_{ij} \omega_i = \lambda_{jj} \alpha_j = \lambda_{jj} \sum\limits_{i=1}^{n} a_{ij} \omega_i$; i.e., $2\lambda_{ij} = a_{ij} \lambda_{jj}$. □

Lemma 2.6 can be generalized:

Lemma 2.7. *Let A be a Cartan matrix. Then f is a degree l invariant polynomial if and only if for $1 \leq j \leq n$,*

$$\frac{\partial f}{\partial \omega_j} - \frac{1}{2!} \frac{\partial^2 f}{\partial \omega_j^2} \alpha_j + \cdots + (-1)^l \frac{1}{l!} \frac{\partial^l f}{\partial \omega_j^l} \alpha_j^{l-1} = 0. \tag{1}$$

Lemma 2.8. *An $n \times n$ Cartan matrix $A = \{a_{ij}\}$ is symmetrizable if and only if there exist nonzero $d_1, d_2, \ldots, d_n$ such that $a_{ij} d_j = a_{ji} d_i$ for all i, j.*

Proof. Suppose A is symmetrizable. Then there exist an invertible diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ and a symmetric matrix B such that $A = DB$; that is, $a_{ij} = d_i b_{ij}$ for all i, j. So $a_{ij}/d_i = b_{ij} = b_{ji} = a_{ji}/d_j$. Equivalently, $a_{ij}d_j = a_{ji}d_i$.

If there exist nonzero $d_1, d_2, \ldots, d_n$ such that $a_{ij}d_j = a_{ji}d_i$ for all i, j, let $D = \mathrm{diag}(d_1, \ldots, d_n)$ and $B = (b_{ij})_{n\times n} = (a_{ij}/d_i)_{n\times n}$; then, $A = DB$. Therefore A is symmetrizable. □

Corollary 2.9. *If A is an indecomposable Cartan matrix, then* $\dim I^2(A) = 1$ *or* 0 *depending on whether A is symmetrizable or not. And if A is symmetrizable, then $I^2(A)$ is spanned by the invariant bilinear form ψ, which is unique up to a constant.*

Proof. If $\dim I^2(A) > 0$, choose $f(\omega) = \sum_{i,j=1}^{n} \lambda_{ij}\omega_i\omega_j \in I^2(A)$, $f(\omega) \neq 0$. By permuting the simple roots, we can assume that there exists an integer $k > 0$ such that $\lambda_{11}, \ldots, \lambda_{kk} \neq 0$ but $\lambda_{k+1,k+1}, \ldots, \lambda_{nn} = 0$. If $i \leq k$ and $j > k$, by Lemma 2.6 $2\lambda_{ij} = a_{ij}\lambda_{jj}$, therefore $\lambda_{ij} = 0$. By $0 = 2\lambda_{ij} = 2\lambda_{ji} = a_{ji}\lambda_{ii}$, we get $a_{ji} = 0$ for all $i \leq k$, $j > k$. Since A is indecomposable, k must equal n. Let $d_i = \lambda_{ii}$ for $1 \leq i \leq n$; then $a_{ij}d_j = a_{ji}d_i$ for all i, j. This means that A is symmetrizable.

If $\dim I^2(A) = 0$, then A is nonsymmetrizable (if A is symmetrizable, then the Killing form gives an element in $I^2(A)$).

Since A is indecomposable, for all i, j, the ratios $\lambda_{ij} : \lambda_{jj}$ and $\lambda_{ii} : \lambda_{jj}$ are determined by A. Therefore, if $\dim I^2(A) > 0$, then $\dim I^2(A) = 1$. □

Below, for an indecomposable and symmetrizable Cartan matrix A, we always fix a nonzero $\psi \in I^2(A)$.

3. Rings of polynomial invariants of Weyl groups, $n = 2$ case

A 2×2 Cartan matrix is of the form

$$A_{a,b} := \begin{pmatrix} 2 & -a \\ -b & 2 \end{pmatrix}.$$

We say that $A_{a,b}$ is of finite, affine or indefinite type if $ab < 4$, $ab = 4$ or $ab > 4$, respectively. The action of reflections $\sigma_1, \sigma_2 \in W(A)$ on h^* is given by

$$\sigma_1(\omega_1) = -\omega_1 + b\omega_2, \quad \sigma_1(\omega_2) = \omega_2, \quad \sigma_2(\omega_1) = \omega_1, \quad \sigma_2(\omega_2) = -\omega_2 + a\omega_1.$$

Lemma 3.1. *The Weyl group $W_{a,b}$ of a Cartan matrix $A_{a,b}$ is the dihedral group D_m, where $m = 2, 3, 4, 6$ or ∞ when $ab = 0, 1, 2, 3$ or ≥ 4, respectively. If $A_{a,b}$ is of affine or indefinite type, then the ring of polynomial invariants of the Weyl group $W_{a,b}$ is $I(A_{a,b}) = \mathbb{Q}[\psi]$.*

Proof. For a Cartan matrix $A_{a,b}$ of affine or indefinite type, $ab \neq 0$. Since A is indecomposable and symmetrizable, $\dim I^2(A) = 1$ and $I^2(A)$ is spanned by $\psi = a\omega_1^2 - ab\omega_1\omega_2 + b\omega_2^2$. Suppose $f(\omega) = \sum_{i=0}^{l} \lambda_i \omega_1^i \omega_2^{l-i}$ is a degree l homogeneous invariant polynomial; then

$$\begin{aligned}\sigma_2(f) &= \sum_{i=0}^{l} \lambda_i \omega_1^i (-\omega_2 + a\omega_1)^{l-i} \\ &= \sum_{i=0}^{l} \sum_{j=0}^{l-i} (-1)^j \lambda_i \binom{l-i}{j} \omega_1^i \omega_2^j (a\omega_1)^{l-i-j} \\ &= \sum_{j=0}^{l} \sum_{i=0}^{l-j} (-1)^j \lambda_i \binom{l-i}{j} a^{l-i-j} \omega_1^{l-j} \omega_2^j \\ &= \sum_{j=0}^{l} \Bigl(\sum_{i=0}^{j} (-1)^{l-j} \lambda_i \binom{l-i}{l-j} a^{j-i} \Bigr) \omega_1^j \omega_2^{l-j}.\end{aligned}$$

So $\sigma_2(f) = f$ is equivalent to

$$\lambda_j = \sum_{i=0}^{j} (-1)^{l-j} \lambda_i \binom{l-i}{l-j} a^{j-i}, \quad 0 \le j \le l. \tag{2}$$

Letting $j = 0$, we get $\lambda_0 = (-1)^l \lambda_0$. So $\lambda_0 = 0$ or l is even.

(1) If $\lambda_0 = 0$, then $\omega_1 \mid f$. By Corollary 2.4, $f = 0$.

(2) If l is even, suppose $l = 2m$. There exists a constant λ such that $f - \lambda\psi^m$ is an invariant polynomial and $\omega_1 \mid (f - \lambda\psi^m)$, hence $f = \lambda\psi^m$.

This proves the lemma. □

4. Some results about hyperbolic Cartan matrices

Moody [1978] proved that for each indecomposable and symmetrizable hyperbolic Cartan matrix A, the ring of polynomial invariants $I(A)$ is $\mathbb{Q}[\psi]$, where ψ is the invariant symmetric bilinear form. So in this section we only consider nonsymmetrizable Cartan matrices.

Indecomposable $n \times n$ hyperbolic Cartan matrices exist only for $n \le 10$, and their number is finite for $3 \le n \le 10$. There are lists of hyperbolic Cartan matrices in [Wan 1991] and [Carbone et al. 2010].

Lemma 4.1. *Let A be an indecomposable and nonsymmetrizable hyperbolic Cartan matrix with $n \ge 4$.*

(C1) *The Dynkin diagram of A forms a circle. That is, $a_{ij} \ne 0$ if and only if $|i - j| = 0, 1$ or $n - 1$.*

(C2) *If $|i - j| = 1$ or $n - 1$, then $a_{ij} = -1$ or $a_{ji} = -1$.*

The lemma is proved by direct checking of the lists.

Remark 4.2. The lemma is not true for the case $n = 3$.

Lemma 4.3. *Let A be an indecomposable and nonsymmetrizable hyperbolic Cartan matrix with $n = 3$. Then A contains a 2×2 principal submatrix of affine type, or all the 2×2 principal submatrices of A are of finite type. In the latter case, A satisfies conditions* (C1) *and* (C2).

Feingold and Nicolai [2004] proved the following theorem:

Theorem 4.4. *Let $g(A)$ be the Kac–Moody Lie algebra associated to a symmetrizable Cartan matrix $A = (a_{ij})_{n\times n}$ which is generated by $\alpha_i^\vee, e_i, f_i, 1 \le i \le n$, and let $\beta_1, \dots, \beta_m$ be a set of positive real roots of $g(A)$ such that $\beta_i - \beta_j$ is not a root for $1 \le i \ne j \le m$. Let E_i, F_i be root vectors in the one-dimensional root spaces corresponding to the positive real roots β_i and negative real roots $-\beta_i$, respectively, and let $H_i = [E_i, F_i]$. Then the Lie subalgebra of g generated by $\{E_i, F_i, H_i \mid 1 \le i \le m\}$ is a regular Kac–Moody subalgebra with Cartan matrix $B = (b_{ij})_{n\times n} = (2(\beta_j, \beta_i)/(\beta_i, \beta_i))_{n\times n}$.*

In the formulas below, subscripts not in the range from 1 to n are to be taken modulo n in that range. By using the ideas in the theorem of Feingold and Nicolai, we can prove the following lemma:

Lemma 4.5. *Let A be an $n \times n$ Cartan matrix satisfying conditions* (C1) *and* (C2) *in Lemma 4.1, with simple root system $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, then $\beta_i = \alpha_{i+1} + \alpha_{i+2}$, $1 \le i \le n$ is a set of positive real roots of $g(A)$, and $\beta_i - \beta_j$ for $i \ne j$ are not roots. Let $\alpha_i^\vee$, e_i, f_i, $1 \le i \le n$ be the generators of $g(A)$, $E_i = [e_{i+1}, e_{i+2}]$, $F_i = -[f_{i+1}, f_{i+2}]$ and $H_i = [E_i, F_i]$. Then H_i, E_i, F_i, $1 \le i \le n$ generate a full rank regular Kac–Moody subalgebra with simple root system $\{\beta_1, \beta_2, \dots, \beta_n\}$ and Cartan matrix $B = (b_{ij}) = (\beta_j(H_i))_{n\times n}$.*

Proof. For $\beta_i = \alpha_{i+1} + \alpha_{i+2}$,

$$\begin{aligned}
H_i &= [E_i, F_i] \\
&= -[[e_{i+1}, e_{i+2}], [f_{i+1}, f_{i+2}]] \\
&= -\big[[[e_{i+1}, e_{i+2}], f_{i+1}], f_{i+2}\big] - [f_{i+1}, [[e_{i+1}, e_{i+2}], f_{i+2}]] \\
&= \big[[[f_{i+1}, e_{i+1}], e_{i+2}], f_{i+2}\big] + [[e_{i+1}, [f_{i+1}, e_{i+2}]], f_{i+2}] \\
&\quad + [f_{i+1}, [[f_{i+2}, e_{i+1}], e_{i+2}]] + \big[f_{i+1}, [e_{i+1}, [f_{i+2}, e_{i+2}]]\big] \\
&= -[[\alpha_{i+1}^\vee, e_{i+2}], f_{i+2}] - [f_{i+1}, [e_{i+1}, \alpha_{i+2}^\vee]] \\
&= -(a_{i+1,i+2}\alpha_{i+2}^\vee + a_{i+2,i+1}\alpha_{i+1}^\vee).
\end{aligned}$$

Then $[H_i, E_i] = -2(a_{i+1,i+2} + a_{i+2,i+1} + a_{i+1,i+2}a_{i+2,i+1})E_i$.

Since, for each $1 \le i \le n$,

$$-2(a_{i+1,i+2}+a_{i+2,i+1}+a_{i+1,i+2}a_{i+2,i+1}) = 2(1-(a_{i+1,i+2}+1)(a_{i+2,i+1}+1)) = 2,$$

a routine check shows that $\beta_j(H_i) \le 0$. Therefore the matrix B with $b_{ij} = \beta_j(H_i)$ is a Cartan matrix. The Serre relations for H_i, E_i, F_i, $1 \le i \le n$ from B are checked as in [Feingold and Nicolai 2004], so H_i, E_i, F_i, $1 \le i \le n$ generate a Kac–Moody Lie algebra $g(B)$ inside $g(A)$. □

We can't say Lemma 4.5 is a corollary of the previous theorem, since we don't know whether the theorem is true for nonsymmetrizable Cartan matrices A. So we must prove Lemma 4.5 by direct computation.

Corollary 4.6. *Let A be an $n \times n$ indecomposable and nonsymmetrizable hyperbolic Cartan matrix. Then in $g(A)$ there is a full rank indecomposable and nonhyperbolic regular indefinite Kac–Moody subalgebra $g(B)$.*

This corollary is proved by using Lemma 4.5 and checking the list of indecomposable, nonsymmetrizable hyperbolic Cartan matrices. We have written a computer program to do the checking. The computation results show that except for the hyperbolic Lie algebras labeled 131, 132, 133, 137, 139, 141 in the list in [Carbone et al. 2010], all the subalgebras we constructed are nonsymmetrizable.

Below is a simple example:

Example 4.7. For the hyperbolic Cartan matrix

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -2 & -1 & 2 \end{pmatrix},$$

we obtain a regular subalgebra $g(B)$ of $g(A)$ with simple roots

$$\beta_1 = \alpha_2 + \alpha_3, \quad \beta_2 = \alpha_3 + \alpha_1, \quad \beta_3 = \alpha_1 + \alpha_2,$$

and its Cartan matrix is

$$B = \begin{pmatrix} 2 & -2 & -2 \\ -3 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

It is nonsymmetrizable and indefinite, but nonhyperbolic.

5. Proof of the main theorem

Some preparatory lemmas. Let A be an $n \times n$ Cartan matrix and $S = \{1, 2, \ldots, n\}$. For $J \subset S$, let A_J be the principal submatrix $(a_{ij})_{i,j \in J}$ corresponding to J. Then $A' = A_{S-\{n\}}$ is the upper-left $(n-1) \times (n-1)$ principal submatrix of A. Let h' be the subspace of h spanned by $\alpha_1^\vee, \ldots, \alpha_{n-1}^\vee$ and h'^* the subspace of h^* spanned by $\omega_1, \ldots, \omega_{n-1}$; then $h = h' \oplus \mathbb{R}\alpha_n^\vee$ and $h^* = h'^* \oplus \mathbb{R}\omega_n$. Let $\alpha_i \in h^*$, $1 \le i \le n$ and $\alpha'_i \in h'^*$, $1 \le i \le n-1$ be the simple roots of Cartan matrices A and A'

respectively, and σ_i, $1 \le i \le n$ and σ'_i, $1 \le i \le n-1$ be the Weyl reflections in h^* and h'^* respectively. For $1 \le i \ne j \le n-1$, $\alpha_i = \alpha'_i + a_{ni}\omega_n$, $\sigma_i(\omega_j) = \sigma'_i(\omega_j)$, and $\sigma_i(\omega_i) = \omega_i - \alpha_i = \sigma'_i(\omega_i) - a_{ni}\omega_n$.

Lemma 5.1. *Let $\omega = (\omega_1, \dots, \omega_{n-1}, \omega_n)$ and $\omega' = (\omega_1, \dots, \omega_{n-1})$. If $f(\omega)$ is a degree l invariant polynomial under the action of $\sigma_1, \dots, \sigma_{n-1}$ and $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$, with each $f_i(\omega')$ a degree i homogeneous polynomial in $S(h'^*)$, then $f_l(\omega')$ is invariant under the action of $\sigma'_1, \dots, \sigma'_{n-1}$.*

Proof. For $k \ne n$,

$$\begin{aligned} f(\omega) = \sigma_k(f(\omega)) &= \sum_{i=0}^{l} \sigma_k(f_i(\omega'))\omega_n^{l-i} \\ &= \sum_{i=0}^{l} f_i(\sigma_k(\omega_1), \sigma_k(\omega_2), \dots, \sigma_k(\omega_k), \dots, \sigma_k(\omega_{n-1}))\omega_n^{l-i} \\ &= \sum_{i=0}^{l} f_i(\sigma'_k(\omega_1), \sigma'_k(\omega_2), \dots, \sigma'_k(\omega_k) - a_{nk}\omega_n, \dots, \sigma'_k(\omega_{n-1}))\omega_n^{l-i}. \end{aligned}$$

Setting $\omega_n = 0$, we get $f_l(\omega') = f_l(\sigma'_k(\omega')) = \sigma'_k(f_l(\omega'))$. □

Corollary 5.2. *If $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i} \in I(A)$, then $f_l(\omega') \in I(A')$.*

Lemma 5.3. *If the degree l polynomial $f(\omega)$ is invariant under the action of $\sigma_1, \dots, \sigma_{n-1}$ and $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$, then for $k \ne n$ and $0 \le i \le l$,*

$$\sigma'_k(f_i(\omega')) = \sum_{j=0}^{l-i} \frac{(-a_{nk})^j}{j!} \frac{\partial^j f_{i+j}(\omega')}{(\partial\omega_k)^j}. \tag{3}$$

Proof. Continuing the calculation of $f(w)$ as in Lemma 5.1, for $k \ne n$, we get

$$\begin{aligned} &\sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i} \\ &= f(\omega) = \sum_{i=0}^{l} f_i(\sigma'_k(\omega_1), \sigma'_k(\omega_2), \dots, \sigma'_k(\omega_k) - a_{nk}\omega_n, \dots, \sigma'_k(\omega_{n-1}))\omega_n^{l-i} \\ &= \sum_{i=0}^{l}\sum_{j=0}^{i} \frac{1}{j!}\frac{\partial^j f_i}{(\partial\omega_k)^j}(\sigma'_k(\omega_1), \sigma'_k(\omega_2), \dots, \sigma'_k(\omega_k), \dots, \sigma'_k(\omega_{n-1}))(-a_{nk}\omega_n)^j\omega_n^{l-i} \\ &= \sum_{i=0}^{l}\sum_{j=0}^{i} \frac{(-a_{nk})^j}{j!}\frac{\partial^j f_i}{(\partial\omega_k)^j}(\sigma'_k(\omega_1), \sigma'_k(\omega_2), \dots, \sigma'_k(\omega_k), \dots, \sigma'_k(\omega_{n-1}))\omega_n^{l-i+j} \\ &= \sum_{i=0}^{l}\left(\sum_{j=0}^{l-i} \frac{(-a_{nk})^j}{j!}\frac{\partial^j f_{i+j}}{(\partial\omega_k)^j}(\sigma'_k(\omega'))\right)\omega_n^{l-i}. \end{aligned}$$

By comparing the coefficients of ω_n^{l-i} in the two sides, we get

$$f_i(\omega') = \sum_{j=0}^{l-i} \frac{(-a_{nk})^j}{j!} \frac{\partial^j f_{i+j}}{(\partial\omega_k)^j}(\sigma'_k(\omega')).$$

Acting by σ'_k on both sides, we prove the lemma. □

In the following, given an $n \times n$ Cartan matrix A, we denote its upper-left $(n-1) \times (n-1)$ principal submatrix by A'.

Lemma 5.4. *Let A be an indefinite $n \times n$ Cartan matrix. If both A and A' are indecomposable and symmetrizable, then the restriction of the invariant symmetric bilinear form ψ to h' gives an invariant symmetric bilinear form ψ'.*

The proof is obvious by checking $\psi|_{h'} \neq 0$ and $\psi' = \psi|_{h'}$ is invariant under the action of $\sigma'_1, \dots, \sigma'_{n-1}$.

Lemma 5.5. *Let f be a $W(A)$-invariant polynomial and $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. Then for $1 \le j \le l$,*

$$f_j(\omega') = \sum_{i=0}^{j}(-1)^{l-i} f_i(\omega')\binom{l-i}{l} - j\omega_n'^{\,j-i}, \tag{4}$$

where $\omega'_n = \sum_{j\neq n} a_{jn}\omega_j$.

Proof. We have $\sigma_n(\omega_n) = \omega_n - \alpha_n = -\omega_n - \sum_{j\neq n} a_{jn}\omega_j = -\omega_n - \omega'_n$, and

$$\begin{aligned} f = \sigma_n(f) &= \sum_{i=0}^{l} f_i(\omega')\sigma_n(\omega_n^{l-i}) \\ &= \sum_{i=0}^{l} f_i(\omega')(-\omega_n - \sum_{j\neq n} a_{jn}\omega_j)^{l-i} \\ &= \sum_{i=0}^{l} f_i(\omega') \sum_{j=0}^{l-i}(-1)^{l-i}\binom{l-i}{j}\omega_n^j\omega_n'^{\,l-i-j} \\ &= \sum_{j=0}^{l}\Bigl[\sum_{i=0}^{l-j}(-1)^{l-i} f_i(\omega')\binom{l-i}{j}\omega_n'^{\,l-i-j}\Bigr]\omega_n^j \\ &= \sum_{j=0}^{l}\Bigl[\sum_{i=0}^{j}(-1)^{l-i} f_i(\omega')\binom{l-i}{l} - j\omega_n'^{\,j-i}\Bigr]\omega_n^{l-j}. \end{aligned}$$

By comparing the coefficients of ω_n^{l-j}, we prove the lemma. □

Remark 5.6. In fact, equations (3) and (4) are just corollaries of (1) applied to $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$.

Lemma 5.7. *Let f be a $W(A)$-invariant polynomial of degree l and let $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. If $l = 2m$, then for each $i \le m-1$, there exist constants $a_j^i(l)$, $0 \le j \le i$, depending on l, such that $f_{2i+1} = \sum_{j=0}^{i} a_j^i(l) f_{2(i-j)}\omega_n'^{2j+1}$. If $l = 2m+1$, then $f_0 = 0$, and for each $i \le m$, there exist constants $b_j^i(l)$, $1 \le j \le i$, such that $f_{2i} = \sum_{j=1}^{i} b_j^i(l) f_{2(i-j)+1}\omega_n'^{2j-1}$. And the coefficients $a_j^i(l)$ and $b_j^i(l)$ can be computed.*

Proof. Letting $j = 0$ in (4), we get $f_0 = (-1)^l f_0$. So there are two cases:

Case 1: l is even. Let $j = 1$ in (4). Then $f_1 = -f_1 + \binom{l}{1} f_0\omega_n'$. That is,

$$f_1 = \frac{1}{2}\binom{l}{1} f_0\omega_n'. \tag{5}$$

For $j = 2$, we get $f_2 = f_2 - \binom{l-1}{1} f_1\omega_n' + \binom{l}{2} f_0\omega_n'^2$; equivalently, $f_1 = \frac{1}{2}\binom{l}{1} f_0\omega_n'$. For $j = 3$, we get

$$f_3 = -f_3 + \binom{l-2}{1} f_2\omega_n' - \binom{l-1}{2} f_1\omega_n'^2 + \binom{l}{3} f_0\omega_n'^3.$$

Substituting (5), we get

$$f_3 = \frac{1}{2}\binom{l-2}{1} f_2\omega_n' - \frac{1}{4}\binom{l}{3} f_0\omega_n'^3. \tag{6}$$

Continuing this procedure, we've proved the lemma when l is even.

Case 2: l is odd. Then $f_0 = 0$, and the proof is similar to the previous case. □

Corollary 5.8. *Let f be a $W(A)$-invariant polynomial and $f(\omega) = \sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. Then $\omega_n' \mid f_{l-1}(\omega')$.*

Computation motivates us to make the following conjecture:

Conjecture. If l is even, then equation (4) for $j = 2k$ can be derived from the set of equations for $j = 0, 1, 2, \dots, 2k-1$. If l is odd, Equation (4) for $j = 2k-1$ can be derived from the set of equations for $j = 0, 1, 2, \dots, 2k-2$.

The conjecture is verified for $k \le 3$.

Lemma 5.9. *Let A be an $n \times n$ Cartan matrix. If $f(\omega), g(\omega) \in S(h^*)$ satisfy $\sigma_k(f(\omega)) - f(\omega) = \sigma_k(g(\omega)) - g(\omega)$ for each $1 \le k \le n$, then $f - g \in I(A)$.*

The proof is trivial.

Lemma 5.10. *Let A be an indefinite $n \times n$ Cartan matrix and A' its upper-left $(n-1) \times (n-1)$ principal submatrix. If the ring of $W(A')$ polynomial invariants $I(A')$ is equal to $\mathbb{Q}[\psi']$ and $l = 2m$, then for each $W(A)$-invariant polynomial*

$f(\omega)=\sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$ of degree l, there exists a constant k such that $f_l(\omega')=k\psi'^m$ and $f_{l-1}(\omega')=km\psi'^{m-1}\omega_n^*$, where $\omega_n^*=\sum_{k\neq n}\lambda_{kk}a_{nk}\omega_k$.

Proof. By Corollary 5.2, $f_l(\omega')\in I(A')$. Since $I(A')=\mathbb{Q}[\psi']$, there exists k such that $f_l(\omega')=k\psi'^m$. In (3), letting $j=l-1$, we get for $1\le k\le n-1$ that

$$\sigma_k'(f_{l-1}(\omega'))-f_{l-1}(\omega')=-a_{nk}\frac{\partial f_l}{\partial\omega_k}.$$

Let $g(\omega')=km\psi'^{m-1}\omega_n^*$; then it is easy to check

$$-a_{nk}\frac{\partial f_l}{\partial\omega_k}=\sigma_k'(g(\omega'))-g(\omega'),$$

so $\sigma_k'(f_{l-1}(\omega'))-f_{l-1}(\omega')=\sigma_k'(g(\omega'))-g(\omega')$. Applying Lemma 5.9 to f_{l-1}, g for the Cartan matrix A', we get $f_{l-1}(\omega')-g(\omega')\in I^{l-1}(A')$. But $I^{l-1}(A')=I^{2m-1}(A')=\{0\}$, hence $f_{l-1}(\omega')=g(\omega')$. □

Proof of three propositions.

Proposition 5.11. *Let A be an $n\times n$ indecomposable and indefinite Cartan matrix. If $I(A')=\mathbb{Q}$, then $I(A)=\mathbb{Q}$.*

Proof. Let f be a $W(A)$-invariant polynomial and $f(\omega)=\sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. Then by Corollary 5.2, $f_l(\omega')\in I(A')$, so $f_l(\omega')=0$.

For $i=l-1$, Equation (3) is

$$\sigma_k'(f_{l-1}(\omega'))=f_{l-1}(\omega')-a_{nk}\frac{\partial f_l}{\partial\omega_k}(\omega').$$

Substituting $f_l(\omega')=0$ in the above equation, we get $\sigma_k'(f_{l-1}(\omega'))=f_{l-1}(\omega')$, so $f_{l-1}(\omega')=0$. Continuing this procedure, we show that $f_i(\omega')=0$ for all $i>0$ and f_0 is a constant. Hence $f(\omega)=f_0\omega_n^l$. By Corollary 2.4, $f=0$. □

Proposition 5.12. *Let A be an $n\times n$ symmetrizable and indefinite Cartan matrix. If $I(A')=\mathbb{Q}[\psi']$, then $I(A)=\mathbb{Q}[\psi]$.*

Proof. Let f be a $W(A)$-invariant polynomial and $f(\omega)=\sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. Then by Corollary 5.2, $f_l(\omega')\in I(A')$, so $f_l(\omega')=0$ or there exists $\lambda\neq 0$ such that $f_l=\lambda\psi'^m$. If $f_l=0$, then $\omega_n\mid f$, so $f=0$. If $f_l=\lambda\psi'^m$, then by Lemma 5.4 we can assume $\psi|_{h'}=\psi'$, so $f-\lambda\psi^m$ is a $W(A)$-invariant polynomial and $\omega_n\mid(f-\lambda\psi^m)$. Hence $f=\lambda\psi^m$. □

Proposition 5.13. *Let A be an $n\times n$ indecomposable and nonsymmetrizable Cartan matrix. If A' is symmetrizable and $I(A')=\mathbb{Q}[\psi']$, then $I(A)=\mathbb{Q}$.*

Proof. Let f be a $W(A)$-invariant polynomial and $f(\omega)=\sum_{i=0}^{l} f_i(\omega')\omega_n^{l-i}$. Suppose $\psi'=\sum_{i,j=1}^{n-1}\lambda_{ij}\omega_i\omega_j$ with $\lambda_{ij}=\lambda_{ji}$ for all $1\le i,j\le n-1$.

If l is even, suppose $l=2m$. We prove $f_l=0$ first. Suppose $f_l\neq 0$; then by Lemma 5.10, there exists $k \neq 0$ such that $f_l = k\psi'^m$ and $f_{l-1} = km\psi'^{m-1}\omega_n^*$. By Corollary 5.8, $\omega_n' \mid f_{l-1} = km\psi'^{m-1}\omega_n^*$. So $\omega_n' \mid \psi'$ or $\omega_n' \mid \omega_n^*$. Since ψ' is $W(A)$-invariant and $-\omega_n'$ is in the Tits cone, by Corollary 2.4 $\omega_n' \mid \psi'$ is impossible. Therefore $\omega_n' \mid \omega_n^*$. Because A is indecomposable, both ω_n' and ω_n^* are not 0. Therefore there exists a constant $d_n \neq 0$ such that $\omega_n^* = d_n\omega_n'$. But $\omega_n' = \sum_{j\neq n} a_{jn}\omega_j$ and $\omega_n^* = \sum_{j\neq n} \lambda_{jj}a_{nj}\omega_j$, so $a_{jn}d_n = a_{nj}\lambda_{jj}$. Let $d_i = \lambda_{ii}$, $1 \leq i \leq n-1$; since A' is symmetrizable, by Lemma 2.8 we know $a_{ij}d_j = a_{ji}d_i$ for all $i, j \leq n-1$. Combining with $a_{jn}d_n = a_{nj}\lambda_{jj}$, we get $a_{ij}d_j = a_{ji}d_i$ for all $i, j \leq n$. This shows A is symmetrizable, contradicting our assumption, so $f_l = 0$.

If l is odd, then $f_l \in I^l(A')$ also implies $f_l = 0$.

If $f_l = 0$, then the remaining procedure of the proof is similar to the proof of Proposition 5.12. □

Proof of the main theorem. To prove the main theorem we need the following lemma:

Lemma 5.14. *Let A be a nonhyperbolic, indecomposable and indefinite Cartan matrix. Then there exists an integer k, $1 \leq k \leq n$ such that $A_{S-\{k\}}$ is an indecomposable and indefinite Cartan matrix.*

Proof. Since the Cartan matrix A is nonhyperbolic, there exists an integer k, $1 \leq k \leq n$ such that $A_{S-\{k\}}$ is indefinite. If $A_{S-\{k\}}$ is indecomposable, the lemma is proved. If $A_{S-\{k\}}$ is decomposable, then the Dynkin diagram of $A_{S-\{k\}}$ is split into r connected subdiagrams $\Gamma_1, \ldots, \Gamma_r$, with $r > 1$, and there is an s_0, $1 \leq s_0 \leq r$, such that the principal submatrix corresponding to Γ_{s_0} is indefinite. Since A is indecomposable, the simple root α_k is connected to all Γ_s, $1 \leq s \leq r$.

We find a connected subdiagram Γ_s, $s \neq s_0$. There must be a vertex $\alpha_{k'}$ of Γ_s such that the subdiagram $\Gamma_s - \{\alpha_{k'}\}$ is connected (note that we can choose a vertex α from a connected finite graph Γ and the resulted subgraph $\Gamma - \{\alpha\}$ is still connected). It is obvious that $A_{S-\{k'\}}$ is indecomposable and indefinite. □

Now we can prove the main theorem:

Theorem. *Let A be an $n \times n$ indecomposable and indefinite Cartan matrix A. If A is symmetrizable, then $I(A) = \mathbb{Q}[\psi]$; if A is nonsymmetrizable, then $I(A) = \mathbb{Q}$.*

Proof. We prove this theorem by induction on n. For $n = 2$, this is Lemma 3.1.

Suppose this theorem is true for all $(n-1) \times (n-1)$ indecomposable and indefinite Cartan matrices.

For an $n \times n$ indecomposable and indefinite Cartan matrix A, if A is not hyperbolic, then by Lemma 5.14 we can find an $(n-1) \times (n-1)$ principal submatrix A' which is both indecomposable and indefinite. Without loss of generality, we can assume A' is the upper-left $(n-1) \times (n-1)$ principal submatrix.

Then by considering the symmetrizability of A' and A, there are three cases:

(1) Both A' and A are nonsymmetrizable.

(2) Both A' and A are symmetrizable.

(3) A' is symmetrizable and A is nonsymmetrizable.

The proof for these three cases are dealt with by combining the induction assumption and Proposition 5.11, Proposition 5.12, and Proposition 5.13 respectively.

So the theorem has been proven when A is nonhyperbolic. For the hyperbolic case, if A is symmetrizable, the proof is given in [Moody 1978]; if A is nonsymmetrizable, it is proved in the following:

Proposition 5.15. *For an $n \times n$ indecomposable, nonsymmetrizable hyperbolic Cartan matrix A, $I(A) = \mathbb{Q}$.*

Proof. For a Cartan matrix A with $n \geq 4$, by Corollary 4.6 we can find an $n \times n$ indecomposable, nonhyperbolic and indefinite Cartan matrix B such that the root system associated to B is a sub-root system of the root system associated to A, and the Weyl group $W(B)$ is a subgroup of $W(A)$. Therefore $I(A) \subset I(B)$.

If B is nonsymmetrizable, then by combining Lemma 5.14, Proposition 5.11 or Proposition 5.13 and the same induction procedure, we can prove $I(B) = \mathbb{Q}$. Hence $I(A) = \mathbb{Q}$.

If B is symmetrizable, then by combining Lemma 5.14 and Proposition 5.12, we prove $I(B) = \mathbb{Q}[\psi_B]$. To prove $I(A) = \mathbb{Q}$, it is sufficient to show the ψ_B^m, $m \geq 1$ are not $W(A)$ invariants.

Suppose ψ_B^m is a $W(A)$-invariant polynomial. If m is odd, we get $\psi_B = (\psi_B^m)^{1/m}$ is $W(A)$-invariant. If m is even, similarly we get for each $\sigma \in W(A)$ that $\sigma(\psi_B) = \psi_B$ or $-\psi_B$. But $\sigma(\psi_B) = -\psi_B$ is impossible (a symmetric bilinear form $\psi = \sum_{i,j=1}^{n} \lambda_{ij}\omega_i\omega_j$ with all the λ_{ii}, $1 \leq i \leq n$ having the same sign can't be linearly transformed to $-\psi$). So we get $\sigma(\psi_B) = \psi_B$. Therefore ψ_B is a $W(A)$-invariant polynomial. Since A is nonsymmetrizable, this is impossible. Hence $I(A) = \mathbb{Q}$.

For the $n = 3$ case, there are two possibilities. If A contains a 2×2 principal submatrix A' of affine type, then by combining Lemma 3.1 and Proposition 5.13, we show $I(A) = \mathbb{Q}$. If all the 2×2 principal submatrices of A are of finite type, then A satisfies conditions (C1) and (C2). So we can find an indecomposable, nonhyperbolic and indefinite Cartan matrix B such that $g(B)$ is a regular subalgebra of $g(A)$. By a similar method as for $n \geq 4$, we can also prove $I(A) = \mathbb{Q}$. This proves the proposition and with it the theorem. □

6. Applications to rational homotopy types of Kac–Moody groups and their flag manifolds of indefinite type

For the Kac–Moody Lie algebra $g(A)$, there is the Cartan decomposition $g(A) = h \oplus \sum_{\alpha \in \Delta} g_\alpha$, where h is the Cartan subalgebra and Δ is the root system of $g(A)$.

Let $b = h \oplus \sum_{\alpha\in\Delta^+} g_\alpha$ be the Borel subalgebra; then b corresponds to a Borel subgroup $B(A)$ in the Kac–Moody group $G(A)$. The homogeneous space $F(A) = G(A)/B(A)$ is called the flag manifold of $G(A)$. By [Kumar 2002], $F(A)$ is an ind-variety.

The cohomologies of Kac–Moody groups and their flag manifolds of finite and affine types are extensively studied. For reference see [Pontryagin 1935; Hopf 1941; Borel 1953a; 1953b; 1954; Bott and Samelson 1955; Bott 1956; Milnor and Moore 1965; Chevalley 1994]. But for the indefinite type, little is known.

The rational cohomology rings of Kac–Moody groups and their flag manifolds are also considered in [Kac 1985b] and [Kumar 1985]. The essentially new part of our work is that we study the properties of $P_A(q)$ and derive the explicit formula for i_k. For details see [Chunhua 2010; Chunhua and Xu-an 2012; Xu-an et al. 2013].

For a Kac–Moody group $G(A)$, $H^*(G(A))$ is a locally finite free graded commutative algebra over $\mathbb{Q}$. Let the odd-dimensional free generators of $H^*(G(A))$ be $y_1, \dots, y_l$ and the even-dimensional free generators be $z_1, \dots, z_k, \dots$. By [Kac 1985b; Kitchloo 1998], $l < n$. Denote the number of degree k generators of $H^*(G(A))$ by i_k; then the Poincaré series of $G(A)$ is

$$P_G(q) = \prod_{k=1}^{\infty} \frac{(1-q^{2k-1})^{i_{2k-1}}}{(1-q^{2k})^{i_{2k}}}.$$

The Poincaré series $P_G(q)$ determines the isometry type of the cohomology ring $H^*(G(A))$ and the rational homotopy type of $G(A)$.

Let $BB(A)$ be the classifying space of the Borel subgroup $B(A)$ and $j : F(A) \to BB(A)$ the classifying map of the principal $B(A)$-bundle $\pi : G(A) \to F(A)$. Denote the cohomology generators of $H^*(BB(A))$ by $\omega_1, \dots, \omega_n$, $\deg \omega_i = 2$. A routine computation on the Leray–Serre spectral sequences of the fibration $G(A) \xrightarrow{\pi} F(A) \xrightarrow{j} BB(A)$ shows

$$H^*(F(A)) \cong E_3^{*,*} \cong \mathbb{Q}[\omega_1, \dots, \omega_n]/\langle f_j \mid 1 \le j \le l\rangle \otimes \mathbb{Q}[z_1, \dots, z_k, \dots],$$

where each f_j corresponds to the differential of y_j and the collection of such f_j generates the ring $I(A)$ of $W(A)$ polynomial invariants.

By [Xu-an et al. 2013], there is the following theorem:

Theorem 6.1. *Let $P_A(q)$ be the Poincaré series of a flag manifold $F(A)$. Then the sequence $i_2 - i_1, i_4 - i_3, \dots, i_{2k} - i_{2k-1}, \dots$ can be derived from $P_A(q)$. In fact we can recover $P_A(q)$ from the sequence $i_2 - i_1, i_4 - i_3, \dots, i_{2k} - i_{2k-1}, \dots$.*

But to determine the rational homotopy type of $G(A)$, we need to determine the sequence $i_1, i_2, \dots, i_k, \dots$. So in addition to the Poincaré series $P_A(q)$, we need more ingredients. Note the number of generators of $I(A)$ of degree k is just the integer i_{2k-1}. So if we can determine the degrees of all the generators in $I(A)$, then we can determine the sequence $i_1, i_3, \dots, i_{2k-1}, \dots$. And the main theorem of this paper fills the gap. Now we have:

Theorem 6.2. *For an indecomposable and indefinite Cartan matrix A, $i_{2k-1} = 0$ for all $k > 0$ except for $k = 2$. And for $k = 2$, if A is symmetrizable, $i_3 = 1$; if A is nonsymmetrizable, $i_3 = 0$.*

Setting $\epsilon(A) = 1$ or 0 depending on whether A is symmetrizable or not as in [Kac 1985b], we get:

Theorem 6.3. *The sequence $i_1, i_2, i_3, \dots, i_k, \dots$ is determined by the Poincaré series $P_A(q)$ and $\epsilon(A)$.*

Theorem 6.4. *For an indecomposable and indefinite Cartan matrix A, the rational homotopy types of $G(A)$ are determined by the Poincaré series $P_A(q)$ and $\epsilon(A)$.*

Kumar [1985] proved that for a Kac–Moody Lie algebra $g(A)$, the Lie algebra cohomology $H^*(g(A), \mathbb{C})$ is given by $H^*(G(A)) \otimes \mathbb{C}$. So we also computed the cohomology of a Kac–Moody Lie algebra $g(A)$ with trivial coefficient.

For a Kac–Moody group $G(A)$, $i_1 = i_2 = 0$. And we have:

Corollary 6.5. *For an indecomposable and nonsymmetrizable indefinite Cartan matrix A, $G(A)$ is a 3-connected space.*

Corollary 6.6. *The dimension of the odd rational homotopy group $\pi_{\text{odd}}(G(A))$ of an indefinite Kac–Moody group $G(A)$ is 1 or 0 depending on whether A is symmetrizable or not.*

Theorem 6.7. *For an indecomposable and indefinite Cartan matrix A, if A is symmetrizable, then*

$$H^*(G(A)) \cong \Lambda_{\mathbb{Q}}(y_3) \otimes \mathbb{Q}[z_1, \dots, z_k, \dots]$$

and

$$H^*(F(A)) \cong \mathbb{Q}[\omega_1, \dots, \omega_n]/\langle\psi\rangle \otimes \mathbb{Q}[z_1, \dots, z_k, \dots].$$

If A is nonsymmetrizable, then

$$H^*(G(A)) \cong \mathbb{Q}[z_1, \dots, z_k, \dots]$$

and

$$H^*(F(A)) \cong \mathbb{Q}[\omega_1, \dots, \omega_n] \otimes \mathbb{Q}[z_1, \dots, z_k, \dots],$$

where $\deg z_k \geq 4$ *is even for all k and can be determined from the Poincaré series $P_A(q)$ and $\epsilon(A)$.*

Note that the Poincaré series $P_A(q)$ can be computed easily by an inductive procedure. See [Chunhua 2010; Chunhua and Xu-an 2012] for details. So in principle the computation of rational homotopy types is solved for all indecomposable and indefinite Kac–Moody groups, whether they are symmetrizable or not.

Since Kac–Moody groups and their flag manifolds are products of indecomposable Kac–Moody groups and indecomposable Kac–Moody flag manifolds, by combining the known results for finite and affine types, we have determined the rational homotopy types of all Kac–Moody groups and their flag manifolds. Since $G(A)$ and $F(A)$ are rational formal (see [Sullivan 1977; Kumar 2002]), the rational homotopy groups and rational minimal model of the corresponding Kac–Moody group $G(A)$ and its flag manifold $F(A)$ can be directly computed from Theorem 6.7.

Theorem 6.8. *For an $n \times n$ indecomposable and indefinite Cartan matrix A satisfying $a_{ij}a_{ji} \geq 4$ for all $1 \leq i, j \leq n$, the rational homotopy type of $G(A)$ is determined by $\epsilon(A)$.*

Since there are a large number of Cartan matrices satisfying the condition of Theorem 6.8, this assertion may seem very surprising. But the proof is very simple. It is derived from the equality

$$P_A(q) = \frac{1+q}{1-(n-1)q}.$$

See [Chunhua 2010; Chunhua and Xu-an 2012] for explicit computations of $P_A(q)$.

It deserves to be mentioned that for a 3×3 nonsymmetrizable Cartan matrix A with $a_{ij}a_{ji} \geq 4$ for all i, j, the Kac–Moody group $G(A)$ is a 5-connected space.

For an indecomposable and symmetrizable Cartan matrix A, let p, q, r be the dimensions of the positive, negative and zero vector subspaces of the invariant bilinear form ψ, and set $\tau(A) = (p, q, r)$.

Theorem 6.9. *For an indecomposable and indefinite Cartan matrix A, if A is symmetrizable, then the cohomology ring $H^*(F(A), \mathbb{C})$ is determined by $P_A(q)$ and $\tau(A)$. If $g(A)$ is nonsymmetrizable, then the cohomology ring $H^*(F(A), \mathbb{C})$ is determined by $P_A(q)$.*

This is obtained from the Theorem 6.7 and the classification of real quadratic forms.

Acknowledgements

We would like to thank Professors Feingold and Nicolai for clarifying some facts about Theorem 3.1 in their paper. The referee of our paper has given a detailed list of misprints and errors in the previous version of this paper. It was a great help for us in improving our paper. We wish to express our thanks and respect to him or her.

References

[Borel 1953a] A. Borel, "Homology and cohomology of compact connected Lie groups", *Proc. Nat. Acad. Sci. U. S. A.* **39** (1953), 1142–1146. MR 15,505c Zbl 0052.40401

[Borel 1953b] A. Borel, "Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts", *Ann. of Math.* (2) **57** (1953), 115–207. MR 14,490e Zbl 0052.40001

[Borel 1954] A. Borel, "Sur l'homologie et la cohomologie des groupes de Lie compacts connexes", *Amer. J. Math.* **76** (1954), 273–342. MR 16,219b Zbl 0056.16401

[Bott 1956] R. Bott, "An application of the Morse theory to the topology of Lie-groups", *Bull. Soc. Math. France* **84** (1956), 251–281. MR 19,291a Zbl 0073.40001

[Bott and Samelson 1955] R. Bott and H. Samelson, "The cohomology ring of G/T", *Proc. Nat. Acad. Sci. U. S. A.* **41** (1955), 490–493. MR 17,182f Zbl 0064.25903

[Carbone et al. 2010] L. Carbone, S. Chung, L. Cobbs, R. McRae, D. Nandi, Y. Naqvi, and D. Penta, "Classification of hyperbolic Dynkin diagrams, root lengths and Weyl group orbits", *J. Phys. A* **43**:15 (2010), 155209, 30. MR 2011e:17014 Zbl 1187.81156

[Chevalley 1994] C. Chevalley, "Sur les décompositions cellulaires des espaces G/B", pp. 1–23 in *Algebraic groups and their generalizations: classical methods* (University Park, PA, 1991), edited by W. J. Haboush and B. J. Parshall, Proc. Sympos. Pure Math. **56**, Amer. Math. Soc., Providence, RI, 1994. MR 95e:14041 Zbl 0824.14042

[Chunhua 2010] J. Chunhua, *Poincaré series of Kac–Moody Lie algebras*, master's thesis, Beijing Normal University, 2010.

[Chunhua and Xu-an 2012] J. Chunhua and Z. Xu-an, "On the Poincaré series of Kac–Moody Lie algebras", preprint, 2012. arXiv 1210.0648v1

[Feingold and Nicolai 2004] A. J. Feingold and H. Nicolai, "Subalgebras of hyperbolic Kac–Moody algebras", pp. 97–114 in *Kac–Moody Lie algebras and related topics*, edited by N. Sthanumoorthy and K. C. Misra, Contemp. Math. **343**, Amer. Math. Soc., Providence, RI, 2004. MR 2005d:17032 Zbl 1050.17021

[Hopf 1941] H. Hopf, "Über die Topologie der Gruppen–Mannigfaltigkeiten und ihre Verallgemeinerungen", *Ann. of Math.* (2) **42** (1941), 22–52. MR 3,61b Zbl 0025.09303

[Humphreys 1990] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics **29**, Cambridge University Press, Cambridge, 1990. MR 92h:20002 Zbl 0725.20028

[Kac 1968] V. G. Kac, "Simple irreducible graded Lie algebras of finite growth", *Izv. Akad. Nauk SSSR Ser. Mat.* **32** (1968), 1323–1367. In Russian. MR 41 #4590 Zbl 0222.17007

[Kac 1983] V. G. Kac, *Infinite-dimensional Lie algebras: an introduction*, Progress in Mathematics **44**, Birkhäuser, Boston, 1983. MR 86h:17015 Zbl 0537.17001

[Kac 1985a] V. G. Kac, "Constructing groups associated to infinite-dimensional Lie algebras", pp. 167–216 in *Infinite-dimensional groups with applications* (Berkeley, CA, 1984), edited by V. G. Kac, Math. Sci. Res. Inst. Publ. **4**, Springer, New York, 1985. MR 87c:17024 Zbl 0614.22006

[Kac 1985b] V. G. Kac, "Torsion in cohomology of compact Lie groups and Chow rings of reductive algebraic groups", *Invent. Math.* **80**:1 (1985), 69–79. MR 86m:57041 Zbl 0566.57028

[Kac and Peterson 1983] V. G. Kac and D. H. Peterson, "Regular functions on certain infinite-dimensional groups", pp. 141–166 in *Arithmetic and geometry, Volume II*, edited by M. Artin and J. Tate, Progr. Math. **36**, Birkhäuser, Boston, 1983. MR 86b:17010 Zbl 0578.17014

[Kac and Peterson 1985] V. G. Kac and D. H. Peterson, "Defining relations of certain infinite-dimensional groups", pp. 165–208 in *The mathematical heritage of Élie Cartan* (Lyon, 1984), Astérisque (numéro hors série), Société Mathématique de France, Paris, 1985. MR 87j:22027 Zbl 0625.22014

[Kitchloo 1998] N. R. Kitchloo, *Topology of Kac–Moody groups*, Ph.D. thesis, Massachusetts Institute of Technology, Ann Arbor, MI, 1998, http://search.proquest.com/docview/304473679. MR 2716803

[Kumar 1985] S. Kumar, "Rational homotopy theory of flag varieties associated to Kac–Moody groups", pp. 233–273 in *Infinite-dimensional groups with applications* (Berkeley, CA, 1984), edited by V. G. Kac, Math. Sci. Res. Inst. Publ. **4**, Springer, New York, 1985. MR 87c:17026 Zbl 0612.22009

[Kumar 2002] S. Kumar, *Kac–Moody groups, their flag varieties and representation theory*, Progress in Mathematics **204**, Birkhäuser, Boston, 2002. MR 2003k:22022 Zbl 1026.17030

[Milnor and Moore 1965] J. W. Milnor and J. C. Moore, "On the structure of Hopf algebras", *Ann. of Math.* (2) **81** (1965), 211–264. MR 30 #4259 Zbl 0163.28202

[Moody 1968] R. V. Moody, "A new class of Lie algebras", *J. Algebra* **10** (1968), 211–230. MR 37 #5261 Zbl 0191.03005

[Moody 1978] R. V. Moody, "Polynomial invariants of isometry groups of indefinite quadratic lattices", *Tôhoku Math. J.* (2) **30**:4 (1978), 525–535. MR 80a:20047 Zbl 0395.10032

[Pontryagin 1935] L. S. Pontryagin, "Sur les nombres de Betti des groupes de Lie", *C. R. Acad. Sci., Paris* **200** (1935), 1277–1280. Zbl 0011.10501

[Sullivan 1977] D. Sullivan, "Infinitesimal computations in topology", *Inst. Hautes Études Sci. Publ. Math.* 47 (1977), 269–331. MR 58 #31119 Zbl 0374.57002

[Wan 1991] Z. X. Wan, *Introduction to Kac–Moody algebra*, World Scientific Publishing Co., Teaneck, NJ, 1991. MR 93f:17042 Zbl 0760.17023

[Xu-an et al. 2013] Z. Xu-an, J. Chunhua, and Z. Jimin, "Poincaré series and rational cohomology rings of Kac–Moody groups and their flag manifolds", preprint, 2013. arXiv 1301.2647

Received March 12, 2013. Revised November 17, 2013.

ZHAO XU-AN
DEPARTMENT OF MATHEMATICS
BEIJING NORMAL UNIVERSITY
KEY LABORATORY OF MATHEMATICS AND COMPLEX SYSTEMS, MINISTRY OF EDUCATION
BEIJING, 100875
CHINA
zhaoxa@bnu.edu.cn

JIN CHUNHUA
ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
BEIJING, 100190
CHINA
jinch@amss.ac.cn

CONTENTS

Volume 269, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 269 No. 2 June 2014

0030-8730(201406)269:2;1-Z