Pacific Journal of Mathematics

Volume 220 No. 1 May 2005

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982) F. Wolf (1904–1989)

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

Regular subscription rate: $400.00 a year (10 issues). Special rate: $200.00 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94707-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840 is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94707-0163.

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

ORTHOGONAL FUNCTIONS IN H^∞

CHRISTOPHER J. BISHOP

We construct examples of H^∞ functions f on the unit disk such that the push-forward of Lebesgue measure on the circle is a radially symmetric measure μ_f in the plane, and we characterize which symmetric measures can occur in this way. Such functions have the property that $\{f^n\}$ is orthogonal in H^2, and provide counterexamples to a conjecture of W. Rudin, independently disproved by Carl Sundberg. Among the consequences is that there is an f in the unit ball of H^∞ such that the corresponding composition operator maps the Bergman space isometrically into a closed subspace of the Hardy space.

1. Introduction

Let H^∞ denote the algebra of bounded holomorphic functions on the unit disk $\mathbb{D}$, let $\mathcal{U}$ be the closed unit ball of H^∞ and let $\mathcal{U}_0 = \{f \in \mathcal{U} : f(0) = 0\}$. If $f \in H^\infty$ then it has radial boundary values (which we also call f) almost everywhere on the unit circle $\mathbb{T}$. We say that f is *orthogonal* if the sequence of powers $\{f^n : n = 0, 1, \dots\}$ is orthogonal, that is, if

$$\int_{\mathbb{T}} f^n \bar{f}^m \, d\theta = 0$$

whenever $n \neq m$. In this paper we will characterize orthogonal functions in H^∞ in terms of the Borel probability measure $\mu_f(E) = |f^{-1}(E)|$, where $|\cdot|$ denotes Lebesgue measure on $\mathbb{T}$, normalized to have mass 1. We will also determine exactly which measures arise in this way. We say a measure is *radial* if $\mu(E) = \mu(e^{i\theta} E)$ for $-\infty < \theta < \infty$ and every measurable set E. We will prove:

Theorem 1.1. *If $f \in \mathcal{U}_0$ then $\{f^n : n = 0, 1, \dots\}$ is an orthogonal sequence if and only if μ_f is a radial probability measure supported in the closed unit disk and*

MSC2000: primary 30H05; secondary 30D35, 30D55, 47B38.
Keywords: Rudin's conjecture, orthogonal functions, cut-and-paste construction, composition operators, radial measures, Nevalinna function, harmonic measure, Bergmann space, Hardy space.
The author is partially supported by NSF Grant DMS 01-03626.

satisfying

$$\int_{|z|\leq 1} \log \frac{1}{|z|}\, d\mu_f(z) < \infty. \tag{1–1}$$

Moreover, given any measure μ satisfying these conditions there exists $f \in \mathcal{U}_0$ such that $\mu = \mu_f$.

The result is motivated by the observation that if f is an inner function (that is, $f \in H^\infty$ and $|f| = 1$ almost everywhere on $\mathbb{T}$) with $f(0) = 0$ then μ_f is normalized Lebesgue measure on $\mathbb{T}$ (Lemma 2.3) and f is orthogonal since, if $m > n$,

$$\int_{\mathbb{T}} f^n \bar{f}^m d\theta = \int_{\mathbb{T}} f^{n-m} d\theta = 2\pi f^{n-m}(0) = 0.$$

At a 1988 MSRI conference Walter Rudin asked if the converse is true, that is, are multiples of inner functions the only orthogonal bounded holomorphic functions on the disk? In other words, is normalized Lebesgue measure on the circle the only radial measure which can occur as a μ_f? Our characterization shows that many other symmetric measures can occur and hence provide counterexamples to Rudin's "orthogonality conjecture". The conjecture was independently disproved by Carl Sundberg [2003].

The simplest example of a measure satisfying Theorem 1.1 (other than Lebesgue measure on a circle) is to take μ to be Lebesgue measure on the union of two circles $\{z : |z| = \frac{1}{2}\} \cup \{z : |z| = 1\}$, normalized to give each mass $\frac{1}{2}$. The corresponding function f is orthogonal by the theorem, but is clearly not inner since $|f| = \frac{1}{2}$ on a subset of $\mathbb{T}$ of positive measure.

A more interesting example of a radial measure satisfying (1–1) is normalized area measure on the disk. Thus there is an $f \in \mathcal{U}_0$ such that μ_f is normalized area measure. We will show (Lemma 6.1) that for any holomorphic g on the disk, and $f \in \mathcal{U}_0$ orthogonal,

$$\|g \circ f\|_{H^p}^p = \int_{\mathbb{D}} |g|^p d\mu_f + \mu_f(\mathbb{T}) \|g\|_{H^p}^p, \tag{1–2}$$

and hence:

Corollary 1.2. *There is an $f \in \mathcal{U}_0$ such that for any analytic g on $\mathbb{D}$, g is in the Bergman space A^p, if and only if $g \circ f$ is in the Hardy space H^p, and the norms are equal.*

Thus the subspace M_f spanned by the powers of f in H^2 is isomorphic to the Bergman space, and multiplication by f on M_f is isomorphic to multiplication by z on the Bergman space. Since both spaces are Hilbert spaces, of course one is isomorphic to a subspace of the other, but it is perhaps a little surprising that this isomorphism can be accomplished with a composition operator. Similar statements can be made for Bergman spaces with respect to radial weights $w\,dx\,dy = d\mu$ of finite mass which satisfy (1–1).

More generally, it would be interesting to know for which pair of spaces X, Y, of analytic functions on $\mathbb{D}$, there is an $f \in \mathcal{U}_0$ such that $g \in X$ if and only if $g \circ f \in Y$, and to characterize such f's when they exist. The latter problem is interesting even when $X = Y$ (for example, see [Cima and Hansen 1990]). In Corollary 6.3 we characterize orthogonal functions with this property when $X = Y = H^p$ (it is true if and only if $\mu_f(\mathbb{T}) > 0$). In particular, all inner functions have this property (as claimed in [Cima and Hansen 1990]).

Paul Bourdon has pointed out that (1–2) implies that orthogonal functions f where $\mu_f(\mathbb{T}) > 0$ give examples of composition operators with closed range. See [Cima et al. 1974/75] and [Zorboska 1994] for characterizations of such functions.

The radial symmetry of a "Rudin counterexample" has also been noted by Paul Bourdon [1997a]. He showed that f is orthogonal if and only if the Nevanlinna counting function,

$$N_f(w) = \sum_{f(z)=w} \log \frac{1}{|z|}$$

is almost everywhere constant on each circle centered on the origin. He also showed that the answer to Rudin's question is "yes" if f is univalent, and that if f is orthogonal, the closure of the range of f is a disk (since the range of f equals the set where N_f is positive). The Nevanlinna function N_f is related to μ_f by the formula

$$N_f(w) = \log \frac{1}{|w|} - \int \log \frac{1}{|z-w|} d\mu_f(z)$$

(except possibly on a set of logarithmic capacity zero). This is due to W. Rudin [1967] but we shall give a proof for completeness (Lemma 3.1).

Corollary 1.3. *If $f \in \mathcal{U}_0$ is nonconstant and orthogonal then $N_f(w) = N(|w|)$ for all w outside an exceptional set of zero logarithmic capacity, where*

$$N(r) = \int_r^1 \frac{1-\mu(t)}{t}\, dt$$

for some increasing function μ on $[0, 1]$ such that $\mu(0) = 0$ and $\mu(1) = 1$, and $\int_0^1 \mu(t)\, dt/t < \infty$ (in fact, $\mu(r) = \mu_f(D(0, r))$). Moreover, for every such N there is an $f \in \mathcal{U}_0$ such that $N_f(w) = N(|w|)$ except possibly on a set of logarithmic capacity zero.

The first part of this is due to Paul Bourdon [1997b]. The condition on N in the previous result has many equivalent formulations; for example, it holds if and only if $M(r) = N(e^r)$ on $(-\infty, 0]$ is concave up, has $M(0) = 0$ and $\sup_{r<0} M(r) + r < \infty$, or if $N(|z|)$ is subharmonic on $\mathbb{D} \setminus \{0\}$ and $N(|z|) + \log|z|$ is bounded above.

The behavior of the composition operator $C_f : g \to g \circ f$ can often be expressed in terms of N_f, for example, see [Shapiro 1987; Smith 1996; Smith and Yang 1998]. The result above provides radial examples with any desired rate of decay faster than $1-r$ as $r \to 1$.

If f is orthogonal, then $f(0) = 2\pi \int f\, d\theta = 0$, so f cannot be an outer function. However, our construction can be modified to give:

Corollary 1.4. *There is an orthogonal f such that $f(z)/z$ is a nonconstant outer function.*

Thus, not only are there orthogonal functions which are not inner, there are examples with only the most trivial possible inner factor. I do not know whether there is an example where $f(z)/z$ is bounded away from zero on $\mathbb{D}$ or which symmetric measures μ are of the form μ_f with $f(z)/z$ outer.

One can also construct examples with other properties. For example, $f \in \mathcal{U}_0$ is said to be in the hyperbolic little Bloch class $\mathcal{B}_0^h$ if

$$\lim_{|z|\to 1} \frac{(1-|z|^2)|f'(z)|}{1-|f(z)|^2} = 0.$$

(This is contained in the usual little Bloch space, where only the numerator is required to go to zero.) We will show (Lemma 5.2) that if g is inner and $f \in H^\infty$ then $\mu_{f\circ g} = \mu_f$. Thus taking g to be an inner function in the hyperbolic little Bloch class (which exists by a result of Wayne Smith [1998] and independently of Aleksandrov, Anderson and Nicolau [Aleksandrov et al. 1999]; also see [Cantón 1998]), we can deduce:

Corollary 1.5. *Any of the measures in Theorem 1.1 is μ_f for some $f \in \mathcal{B}_0^h$.*

Cima, Korenblum and Stessin [Cima et al. 1993] also identified symmetric properties of orthogonal functions and showed the answer to Rudin's question is "yes" if f is Hölder of order $\alpha > \frac{1}{2}$ on $\mathbb{T}$. I do not know if there exists any (noninner) orthogonal function which is continuous up to the boundary, but expect that it might be possible to build one by modifying the construction in this paper. If there is a continuous orthogonal function, it would be very interesting to know if the result of Cima, Korenblum and Stessin is sharp, and if not, what the best modulus of continuity for such a function could be. What other natural conditions on an orthogonal function imply that it is actually inner?

The remaining sections are organized as follows:

Section 2: We describe some elementary properties of μ_f and prove it is radial if and only if f is orthogonal.

Section 3: We prove Corollary 1.3 (given Theorem 1.1).

Section 4: We prove some results concerning the convergence of μ_f.

Section 5: We prove Corollary 1.5 (given Theorem 1.1).

Section 6: We prove Corollary 1.2 (given Theorem 1.1).

Section 7: We construct a symmetric μ_f which is supported on two circles.

Section 8: We construct all examples supported in $\left\{\frac{1}{2} \le |z| \le 1\right\}$.

Section 9: We complete the proof of Theorem 1.1.

Section 10: We prove Corollary 1.4.

2. Elementary properties of μ_f

We begin by recalling a few simple facts about analytic functions f and their corresponding measures μ_f. Many of these are well known but we include them for the convenience of the reader.

Lemma 2.1. *If $f \in H^\infty$ then μ_f satisfies*

$$\int \log \frac{1}{|z|} \, d\mu_f(z) < \infty.$$

Proof. If f has a zero of order n at the origin, then $g(z) = f(z)/z^n$ is holomorphic on the unit disk and $|g| = |f|$ on $\mathbb{T}$, hence $\mu_g(A) = \mu_f(A)$ for any annulus $A = \{z : r_1 \le |z| \le r_2\}$. Thus

$$\int \varphi(z) \, d\mu_f(z) = \int \varphi(z) \, d\mu_g(z)$$

for any radial function φ. Using Fatou's lemma and the fact that $\log |g(z)|^{-1}$ is superharmonic on the disk (see [Garnett 1981, page 35]), we deduce

$$\begin{aligned}\int \log \frac{1}{|z|} \, d\mu_f(z) &= \int \log \frac{1}{|z|} \, d\mu_g(z) = \frac{1}{2\pi} \int \log |g(e^{i\theta})|^{-1} \, d\theta \\ &= \frac{1}{2\pi} \int \lim_{r \to 1} \log |g(re^{i\theta})|^{-1} \, d\theta \\ &\le \frac{1}{2\pi} \lim_{r \to 1} \int \log |g(re^{i\theta})|^{-1} \, d\theta \le \log |g(0)|^{-1} < \infty. \qquad \square\end{aligned}$$

A similar estimate is true for other points, for example,

$$\int \log \frac{1}{|z - a|} \, d\mu_f(z) < \infty.$$

In particular, this implies the well-known fact that the set where f has radial limit a must have measure zero.

Given an arc $I \subset \mathbb{T}$ we define the *Carleson box* with *base* I to be

$$Q = Q_I = \{z \in \mathbb{D} : z/|z| \in I, \, 1 - |z| \le |I|\}.$$

A positive measure μ is a *Carleson measure* if there exists a $C < \infty$ such that $\mu(Q_I) \le C|I|$, for every arc $I \in \mathbb{D}$.

Lemma 2.2. *If $f \in \mathcal{U}_0$ then μ_f is a Carleson measure with constant independent of f.*

Proof. Define $\varphi(z) = \omega(z, Q, \mathbb{D} \setminus Q)$ for $z \in \mathbb{D} \setminus Q$ and $\varphi(z) = 1$ for $z \in Q$. It is easy to see that $\omega(z, I, \mathbb{D}) \ge M^{-1} > 0$ for every $z \in \partial Q \cap \mathbb{D}$ and some $M < \infty$ (independent of I and $z \in \partial Q$), so the maximal principle implies

$$\varphi(0) \le M\omega(0, I, \mathbb{D}) \le M|I|.$$

Let $f_r(z) = f(rz)$. Note that $\lim_{r\to1} \varphi(f(rx)) = \varphi(f(x))$ for almost every $x \in \mathbb{T}$, because φ is continuous on the closed disk except at two points, and the set where f has a radial limit equal to one of these has measure zero (by the remark following Lemma 2.1). So by the Lebesgue dominated convergence theorem,

$$\mu_f(Q) \le \int \varphi \, d\mu_f = \frac{1}{2\pi} \int \varphi \circ f \, d\theta = \lim_{r\to1} \frac{1}{2\pi} \int \varphi \circ f_r \, d\theta. \tag{2–1}$$

Since φ is superharmonic on $\mathbb{D}$, it follows that $\varphi \circ f$ is too, so the right-hand side of (2–1) is at most $\varphi(f(0)) = \varphi(0) \le M|I|$. □

If $f(0) \ne 0$ then μ_f is still a Carleson measure, but with norm depending on $|f(0)|$.

One can think of the previous lemma as a weak version of the Littlewood subordination principle: that if f is an analytic self-map of the disk then $g \in H^p$ implies $g \circ f \in H^p$ (with smaller or equal norm). Formally, this implies that if $f(0) = 0$, then

$$\int |g|^p \, d\mu_f \le \frac{1}{2\pi} \int_{\mathbb{T}} |g \circ f|^p \, d\theta = \|g \circ f\|^p_{H^p} \le \|g\|^p_{H^p}$$

for every $g \in H^p$. This implies that $d\mu_f$ is a Carleson measure with norm independent of f (see, for example, [Garnett 1981, Theorem I.5.6]).

The following result appears in many places (for example, [Löwner 1923; Nordgren 1968, Lemma 1; Rudin 1980, page 405; Tsuji 1959, Theorem VIII.30]) and is sometimes called "Löwner's lemma". See [Fernández et al. 1996] and its references for various generalizations.

Lemma 2.3. *If f is an inner function such that $f(0) = 0$, then μ_f is normalized Lebesgue measure on the unit circle.*

Proof. It is enough to check that $\mu_f(I) = |I|$ for arcs. Let I be an arc on the unit circle and let $\varphi(z) = \omega(z, I, \mathbb{D})$. Then $\varphi \circ f$ is bounded and harmonic, and takes radial boundary values 1 and 0 almost everywhere (1 almost everywhere that f has radial limit in I, and 0 almost everywhere that f has radial limit outside I). Thus

$$|I| = \varphi(0) = \varphi(f(0)) = \frac{1}{2\pi}\int_{f^{-1}(I)} d\theta = \mu_f(I). \qquad \square$$

As noted before, the following lemma is similar to results in [Bourdon 1997a] and [Cima et al. 1993].

Lemma 2.4. *Suppose $f \in H^\infty$. Then the measure μ_f is radial if and only if $\{f^n\}$ is orthogonal.*

Proof. If μ_f is radial, it can be written so that

$$\int g(z)\,d\mu_f(z) = \int_0^{2\pi}\int_0^\infty g(re^{i\theta})\,d\theta\,d\nu(r)$$

for every $g \in C_c(\mathbb{R}^2)$, the set of continuous functions of compact support defined on $\mathbb{R}^2$, and for some measure ν on $(0,\infty)$. Thus

$$\int_{\mathbb{T}} f^n \bar{f}^m\,d\theta = \int_{\mathbb{C}} z^n\bar{z}^m\,d\mu_f(z) = \int_0^\infty\int_0^{2\pi} r^{n+m}e^{i(n-m)\theta}\,d\theta\,d\nu(r) = 0$$

if $n \neq m$, so f is orthogonal. Conversely, if f is orthogonal, then μ_f satisfies

$$\int_{\mathbb{C}} z^n\bar{z}^m\,d\mu_f(z) = \int_0^{2\pi}\int_0^\infty r^{n+m}e^{i(n-m)\theta}\,d\mu_f(re^{i\theta}) = 0$$

for $n \neq m$. Thus

$$\int_{\mathbb{C}} P(z,\bar{z})\,d\mu_f(z) = \int_{\mathbb{D}} \sum_n a_{n,n}r^{2n}\,d\mu_f(z)$$

for any polynomial $P(z,\bar{z}) = \sum_{n,m} a_{n,m}z^n\bar{z}^m$ in z and $\bar{z}$, and hence

$$\int_{\mathbb{C}} P(\lambda z,\bar{\lambda}\bar{z})\,d\mu_f(z) = \int_{\mathbb{C}} P(z,\bar{z})\,d\mu_f(z)$$

for any $|\lambda| = 1$. Since polynomials in z and $\bar{z}$ are dense in the continuous functions on the closed unit disk, we deduce that

$$\int_{\mathbb{D}} g(z)\,d\mu_f(z) = \int_{\mathbb{D}} g(\lambda z)\,d\mu_f(z)$$

for any $g \in C_c(\mathbb{R}^2)$ and any $|\lambda| = 1$. This implies μ_f is radial. $\square$

The following lemma greatly simplifies the construction of the basic example, where μ_f is supported on two circles. It says that if we can construct an example where μ_f is radial on the smaller circle, then it automatically looks like Lebesgue measure on the larger one.

Lemma 2.5. *Suppose f lies in $\mathcal{U}_0$, and μ_f is supported on the circles $C_{1/2} \cup C_1 = \{|z| = \frac{1}{2}\} \cup \{|z| = 1\}$. If μ_f restricted to $C_{1/2}$ is a multiple of Lebesgue 1-dimensional measure, then so is μ_f restricted to C_1.*

Proof. Suppose u is any bounded harmonic function on $\mathbb{D}$. Then $v(z) = u(f(z))$ is also bounded and harmonic on $\mathbb{D}$ and $u(0) = v(0)$. Thus

$$\begin{aligned} u(0) = v(0) &= \frac{1}{2\pi} \int_0^{2\pi} u(f(e^{i\theta}))\, d\theta = \int u(z)\, d\mu_f(z) \\ &= \int_{C_{1/2}} u(z)\, d\mu_f(z) + \int_{C_1} u(z)\, d\mu_f(z) \\ &= \mu_f(C_{1/2}) u(0) + \int_{C_1} u(z)\, d\mu_f(z). \end{aligned}$$

Hence $\int_{C_1} u\, d\mu_f = \mu_f(C_1) u(0)$ for any bounded harmonic function u on $\mathbb{D}$. This easily implies that μ_f restricted to C_1 is a multiple of Lebesgue measure on C_1. □

The same proof gives the following generalization of Lemma 2.5.

Lemma 2.6. *Suppose $f \in \mathcal{U}_0$. Then μ_f restricted to the unit circle is of the form $\frac{1}{2\pi}(1 - g(\theta))\, d\theta$, where g is the balayage of μ_f onto the circle, that is,*

$$g(\theta) = \int_{\mathbb{D}} P_z(\theta)\, d\mu_f(z),$$

where $P_z(\theta)$ is the Poisson kernel for $\mathbb{D}$ with respect to the point z.

3. The Nevanlinna counting function

For $f \in H^\infty$, the Nevanlinna counting function is defined to be

$$N_f(w) = \sum_{f(z)=w} \log \frac{1}{|z|}.$$

If $f \in \mathcal{U}_0$ then $N_f(w) \le \log |w|^{-1}$. Clearly this is just the Green's function for the Riemann surface associated to f (projected to the plane by summing over sheets). Since μ_f is the projection of harmonic measure for the Riemann surface, the following is analogous to the standard result for Green's functions of planar domains. Let $\Delta = \partial_x^2 + \partial_y^2$ denote the Laplacian and let δ_0 be the Dirac mass at the origin.

Lemma 3.1 [Rudin 1967]. *If $f \in \mathcal{U}_0$ then $\Delta N_f = -\delta_0 + \mu_f$ in the sense of distributions, and*

$$N_f(w) = \log \frac{1}{|w|} - \int \log \frac{1}{|z - w|}\, d\mu_f(z) \tag{3–1}$$

for all w, except possibly for an exceptional set E of logarithmic capacity zero where "$<$" holds.

The exceptional set is required. For example, if f is the universal covering map of $\mathbb{D}$ minus a compact set E of zero logarithmic capacity, f is an inner function, μ_f is normalized Lebesgue measure on the circle and $N_f(z) = \chi_{\mathbb{D}\setminus E}(z) \log |z|^{-1}$.

Proof. For $0 < r < 1$, let $f_r(z) = f(rz)$ and let $\gamma_r = f_r(\mathbb{T})$. If we choose r so that f' never vanishes on the circle of radius r, then γ_r is a smooth curve and it is easy to check using Green's theorem that $\Delta N_{f_r} = -\delta_0 + \mu_{f_r}$. To see that (3–1) holds for μ_{f_r}, note that both sides of the equation have the same distributional Laplacian, so they differ by a harmonic function. N_{f_r} vanishes outside the unit disk by definition, and the right side of (3–1) vanishes there because μ_{f_r} evaluates harmonic functions at 0. Hence the difference between the left and right sides is the constant zero function.

For any smooth φ with compact support,

$$\int N_{f_r} \Delta\varphi \, dx\, dy = -\varphi(0) + \int \varphi \, d\mu_{f_r}.$$

We shall see later that μ_{f_r} weakly converges to μ_f (Corollary 4.4), and clearly $N_{f_r} \nearrow N_f$ as $r \nearrow 1$. Thus taking $r \to 1$ and applying the monotone convergence theorem we get

$$\int N_f \Delta\varphi \, dx\, dy = -\varphi(0) + \int \varphi \, d\mu_f.$$

This proves the first claim of the lemma. Next we verify (3–1).

We already know that if we replace f by f_r then we have equality in (3–1) for all z and as $r \to 1$, and we know $N_{f_r}(z) \nearrow N_f(z)$ for all z. Thus the question reduces to whether

$$U_r(w) \to U_1(w) \text{ as } r \to 1 \tag{3–2}$$

for all w except a set E of logarithmic capacity zero, where

$$U_r(w) = \int \log \frac{1}{|z-w|} \, d\mu_{f_r}(z).$$

Note that U_r is decreasing in r, by the superharmonicity of $\log |f|^{-1}$, and that U_1 is bounded below by $-\log 2$, since $|z-w| < 2$ for points in the unit disk.

To prove that (3–2) holds, we follow the proof of Frostman's theorem (see [Garnett 1981, Theorem II.6.4], for example). Suppose σ is a measure such that $V(z) = \int \log |z-w|^{-1} \, d\sigma(z)$ is bounded. It suffices to show $\sigma(E) = 0$. By Fatou's lemma

$$\lim_{r\to 1} \int \log \frac{1}{|z-w|} \, d\mu_{f_r}(z) \geq \int \log \frac{1}{|z-w|} \, d\mu_f(z),$$

so $\lim_{r\to 1} U_r(w) \geq U_1(w)$ for all w. On the other hand, by Fatou's lemma, Fubini's theorem and the Lebesgue dominated convergence theorem,

$$\begin{aligned}\int_E \lim_{r\to 1} U_r(w)\,d\sigma(w) &\leq \lim_{r\to 1} \int_E U_r(w)\,d\sigma(w)\\ &= \lim_{r\to 1} \frac{1}{2\pi}\int_0^{2\pi} V(f(re^{i\theta}))\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} V(f(e^{i\theta}))\,d\theta\\ &= \frac{1}{2\pi}\int_0^{2\pi}\int_E \log\frac{1}{|f(e^{i\theta})-w|}\,d\sigma(w)\,d\theta = \int_E U_1(z)\,d\sigma(w).\end{aligned}$$

Thus we must have $\lim_{r\to 1} U_r(w) = U_1(w)$ except on a set of zero σ measure. $\square$

Lemma 3.1 clearly implies that μ_f is radial if and only if N_f is (except for the exceptional set). Thus we see that $\{f^n\}$ is an orthogonal sequence if and only if μ_f is radial, if and only if N_f is radial, except on a set of logarithmic capacity zero. This gives an alternate approach to the results of Bourdon [1997a].

We can also compute exactly which radial functions can occur as N_f for some $f \in \mathcal{U}_0$. Note that

$$\frac{1}{2\pi}\int_0^{2\pi} \log\frac{1}{|re^{i\theta}-w|}\,d\theta = \begin{cases} \log(1/|w|) & \text{if } r \leq |w|,\\ \log(1/r) & \text{if } r \geq |w|.\end{cases}$$

Thus if μ_f is radial and we set $\mu(r) = \mu_f(D(0,r))$, then

$$N_f(w) = \log\frac{1}{|w|} - \int \log\frac{1}{|z-w|}\,d\mu_f(z) = \int_{|w|}^1 \frac{1-\mu(r)}{r}\,dr.$$

Moreover, the integral condition

$$\int_{\mathbb{D}} \log\frac{1}{|z|}\,d\mu < \infty$$

becomes

$$\int_0^1 \mu(r)\frac{dr}{r} < \infty.$$

Thus Theorem 1.1 implies the following corollary.

Corollary 3.2. *Suppose $N(r) = \int_r^1 (1-\mu(t))\,dt/t$ for some increasing function μ such that $\int_0^1 \mu(r)\,dr/r < \infty$, with $\mu(0)=0$ and $\mu(1)=1$. Then there is an $f \in \mathcal{U}_0$ such that $N_f(z) = N(|z|)$ except on a set of zero logarithmic capacity.*

For example, if μ_f is normalized area measure on the unit disk then $\mu(r) = r^2$ and $N_f(z) = \log 1/r - (1-r) \approx (1-r)^2$ as $r \to 1$.

4. Weak* convergence of μ_f

We will obtain the functions f in Theorem 1.1 by a "cut and paste" construction of the corresponding Riemann surface. What this means is that we shall build a sequence of nested Riemann surfaces $R_0 \subset R_1 \subset R_2 \subset \cdots \subset \bigcup R_n = R$ by identifying subdomains of the unit disk along common boundary arcs. The projection of R into the unit disk is a bounded holomorphic function on R, and hence R must be hyperbolic, that is, its universal covering space is the unit disk $\mathbb{D}$. The desired map will be the covering map $f : \mathbb{D} \to R$ followed by the projection into the disk and the corresponding measure μ_f is simply the harmonic measure for the surface R, projected into the plane. In fact, we shall abuse notation and consider the covering map $f : \mathbb{D} \to R$ as actually mapping into the complex numbers (that is, we identify the covering map and this map followed by the projection into the plane). By a similar abuse we shall think of harmonic measure on R and the corresponding projected measure μ_f as the same. Similarly, we will fix a point in R_0 which projects to 0 and call it 0 as well. All our covering maps will be chosen to map 0 in the disk to 0 on the surface. See [Bishop 1993] and [Stephenson 1988], where a similar procedure has been used in different problems.

The main point we must be careful about is to show that the harmonic measure for R is the limit of the measures for R_n. To see that there might be a problem in general, consider what can happen when the surfaces are not nested. For example, R_n is the unit disk minus the points $\left\{z_k = \frac{1}{2}\exp(i2\pi k 2^{-n}) : k = 1, \ldots, 2^n\right\}$. Then the universal covering map $f_n : \mathbb{D} \to R_n$ is an inner function (the isolated boundary points do not have any harmonic measure, so all the measure lives on the part of the boundary above the unit circle) and hence μ_{f_n} is Lebesgue measure on the unit circle. However, one can show (with some work) that $f_n(z) \to \frac{1}{2}z$ uniformly on compact sets of $\mathbb{D}$, so that μ_f is Lebesgue measure on the circle of radius $\frac{1}{2}$. However, if the Riemann surfaces are nested by (increasing) inclusion, then we will show the corresponding measures converge weak*, that is,

$$\lim_{n\to\infty} \int g\, d\mu_n = \int g\, d\mu$$

for any $g \in C_c(\mathbb{R}^2)$.

Lemma 4.1. *Suppose $\epsilon > 0$ and $D(0,\epsilon) = R_0 \subset R_1 \subset \cdots$ are obtained by identifying subdomains of the unit disk along boundary arcs. Let $R = \bigcup_{n=1}^{\infty} R_n$. Choose covering maps $f_n : \mathbb{D} \to R_n$ and $f : \mathbb{D} \to R$ so that $f_n(0) = f(0) = 0$. Then μ_{f_n} converges weak* to μ_f on the closed unit disk.*

The easiest way to see this is using Brownian motion; we shall first sketch such a proof and then give a more classical proof without using Brownian motion.

Let $\mathcal{W}$ be the Wiener space of continuous paths in $\mathbb{C}$ starting at the origin. If R is a Riemann surface constructed as above then we can think of the paths as taking values in R and for each path $w \in \mathcal{W}$, we define the stopping time t_w as the first time t such that $w(t) \notin R$. Then $w \to t_w$ is measurable and the harmonic measure for R is simply the push-forward of Wiener measure on $\mathcal{W}$ under the map given by $w \to w(t_w)$. Given a sequence of nested surfaces $R_0 \subset R_1 \subset \cdots$ as in the lemma, we get a corresponding sequence of maps $g_n : \mathcal{W} \to \mathbb{C}$. Moreover, if $R = \bigcup_n R_n$ and $g : \mathcal{W} \to \mathbb{C}$ is the corresponding map, then $g(w) = \lim_n g_n(w)$; this is because the inclusions imply that for any continuous path in the plane, the first time it leaves R is the limit of the first time it left R_n. Thus for any bounded, continuous function φ on the plane, $\varphi(g_n(w)) \to \varphi(g(w))$ for all w, so the Lebesgue dominated convergence theorem implies that

$$\int_{\mathcal{W}} \varphi(g(w))\, dw = \lim_{n\to\infty} \int_{\mathcal{W}} \varphi(g_n(w))\, dw,$$

which is the desired weak* convergence.

The sketch above is simple and explains why the result is true, but uses the existence of Wiener measure and deep connections between it and harmonic measure. It therefore seems desirable to provide a second proof which uses only function theory. Moreover, we will need some corollaries of the following classical proof for our applications to composition operators.

Let $\{R_n\}$, R, $\{f_n\}$ and f be as in the lemma and let $\Omega_n = f^{-1}(R_n) \subset \mathbb{D}$. Then $\Omega_0 \subset \Omega_1 \subset \cdots$ and $\bigcup_n \Omega_n = \mathbb{D}$. Let ω_n be the harmonic measure for Ω_n with respect to the origin and let φ be any continuous function on the plane. We want to show that

$$\lim_{n\to\infty} \int \varphi(f(z))\, d\omega_n(z) = \int_{\mathbb{T}} \varphi(f(e^{i\theta}))\, d\theta/2\pi.$$

We start by proving the much easier fact that ω_n converges weak* to normalized Lebesgue measure on the circle. (Since f need not be continuous up to the boundary, $\varphi \circ f$ need not be continuous either, so weak* convergence of ω_n is not, by itself, enough to prove weak* convergence of μ_{f_n}.)

Lemma 4.2. *If $\{0\} \in \Omega_0 \subset \Omega_1 \subset \cdots$ is a sequence of subdomains such that $\bigcup_n \Omega_n = \mathbb{D}$, and $\omega_n = \omega(0, \cdot, \Omega_n)$ is the corresponding harmonic measure with respect to the origin, then $\{\omega_n\}$ converges weak* to (normalized) Lebesgue measure on $\mathbb{T}$. Moreover, the measures ω_n are all Carleson with a uniform constant.*

Proof. The Carleson condition follows from Lemma 2.2 applied to the covering map onto Ω_n, so we need only prove weak* convergence. Since $\bigcap_n \big(\overline{\mathbb{D}} \setminus \Omega_n\big) = \mathbb{T}$, there is a sequence $\{r_n\} \nearrow 1$ such that $D_n = \{z : |z| < r_n\} \subset \Omega_n \subset \mathbb{D}$. Suppose that $I \subset \mathbb{T}$ is an open arc and let $Q = \{z \in \mathbb{D} : z/|z| \in I,\ 1 - |z| \leq |I|\}$ be the corresponding

Carleson square. To show ω_n converges weak* to normalized Lebesgue measure, it is clearly enough to show that $\omega_n(Q) \to |I|$.

Let $U_n = D_n \cup Q$ and $V_n = \mathbb{D} \setminus (Q \setminus D_n)$. Then $\omega(0, I, U_n) \to |I|$. To see this, first note that $\omega(0, I, U_n) \leq |I|$ follows immediately from the maximum principle applied to $\omega(z, I, U_n)$ on U_n. For the other direction, suppose that $J \subset I$ is a proper subinterval and note that

$$\omega(0, I, U_n) \geq \omega(0, J, U_n) = |J| - \int_{\partial U_n \setminus I} \int_J P_z(\theta)\, d\theta\, d\omega(0, \cdot, U_n),$$

and that $\int_J P_z(\theta)\, d\theta \to 0$ as $n \to \infty$ for $z \in \partial U_n \setminus I$. Thus the Lebesgue dominated convergence theorem implies $\liminf \omega(0, I, U_n) \geq |J|$. Since this holds for any proper subinterval J, we see that $\omega(0, I, U_n) \to |I|$ as desired. A similar argument shows that $\omega(0, Q \cap V_n, V_n) \to |I|$ as $n \to \infty$.

Thus, by the monotonicity of harmonic measure,

$$\omega_n(Q) \geq \omega(0, \partial\Omega_n \cap Q, U_n) \geq \omega(0, I, U_n) \to |I|,$$

and so $\liminf_n \omega_n(Q) \geq |I|$. On the other hand,

$$\omega_n(Q) \leq \omega(0, Q \cap \partial V_n, V_n) \to |I|,$$

which implies that $\omega_n(Q) \to |I|$. This proves the lemma. □

Lemma 4.3. *Suppose g is a bounded, continuous function on $\mathbb{D}$ which has nontangential limit $g(x)$ almost everywhere on $\mathbb{T}$, and that ν_n is a sequence of probability measures on $\overline{\mathbb{D}}$ which converge weak* to (normalized) Lebesgue measure on the circle and which are all Carleson measures with a uniform constant. Then*

$$\lim_{n\to\infty} \int_{\overline{\mathbb{D}}} g(z)\, d\nu_n(z) = \frac{1}{2\pi} \int_{\mathbb{T}} g(e^{i\theta})\, d\theta.$$

Proof. We may assume that $\|g\|_\infty = 1$. Fix some $\epsilon > 0$. Since g has nontangential limits almost everywhere, given almost any $x \in \mathbb{T}$ there is a $\delta(x) > 0$ such that if I is any interval containing x with length less than $\delta(x)$, then g is within $\epsilon/2$ of $g(x)$ on the top half of the corresponding Carleson box Q. Fix a complex number a, and a $\delta > 0$, and assume that $E_a = \{x \in \mathbb{T} : |g(x) - a| \leq \epsilon/2, \delta(x) > \delta\}$ has positive Lebesgue measure. Using the Lebesgue density theorem choose a dyadic interval I of length less than δ so that $|I \cap E_a| \geq (1-\epsilon)|I|$ and let Q_k be the collection of maximal dyadic subsquares with bases $\{I_k\} \subset I$ such that $|g(z) - a| > \epsilon$ for some z in the top half of Q_k. Let Q be the Carleson square with base I and let $W = Q \setminus \bigcup Q_k$. Then g is within ϵ of a constant on W, and $|\partial W \cap I| \geq (1-\epsilon)|I|$.

We claim that for these domains $\lim_n \nu_n(W) = |\partial W \cap \mathbb{T}|$. To prove this, we use the weak* convergence of $\{\nu_n\}$ to deduce

$$\begin{aligned}
\lim_{n\to\infty} \int_W d\nu_n &= \lim_{n\to\infty} \Big(\nu_n(Q) - \sum_k \nu_n(Q_k) \Big) \\
&= |I| - \lim_{n\to\infty} \sum_k \nu_n(Q_k) \\
&= |I| - \sum_k \lim_{n\to\infty} \nu_n(Q_k) \\
&= |I| - \sum |I_k| \\
&= |\partial W \cap \mathbb{T}|,
\end{aligned}$$

where we used the Lebesgue dominated convergence theorem on the sequence space ℓ^1 to interchange the limit and the infinite sum (our assumption that the measures are uniformly Carleson implies that $\nu_n(Q_k) \le C|I_k|$, independent of n; this gives the ℓ^1 upper bound).

Moreover, the intervals I with these properties form a Vitali cover of $\mathbb{T}$ (see, for example, [Wheeden and Zygmund 1977, Section 7.3]), so we can form a disjoint cover of almost every point of $\mathbb{T}$ using such intervals. Thus we can construct a finite number of disjoint domains $W_j = Q_j \setminus \bigcup_k Q_k^j$, where

(1) Q_j is a Carleson square with base I_j and $|\partial W_j \cap I_j| \ge (1-\epsilon)|I_j|$,

(2) g is within ϵ of a constant c_j on each W_j,

(3) $\sum_j |\partial W_j \cap \mathbb{T}| \ge 1 - \epsilon$.

Let $W = \bigcup_j W_j$ be this finite union. The weak* convergence of $\{\nu_n\}$ implies that

$$\limsup_{n\to\infty} \nu_n(\mathbb{D} \setminus W) \le \epsilon,$$

and so if $\|g\|_\infty \le 1$,

$$\begin{aligned}
\left| \lim_{n\to\infty} \int g\, d\nu_n - \int_{\mathbb{T}} g\, d\theta/2\pi \right| &\le \lim_{n\to\infty} \left| \int_W g\, d\nu_n - \frac{1}{2\pi} \int_{\partial W \cap \mathbb{T}} g\, d\theta \right| \\
&\qquad + \int_{\mathbb{D}\setminus W} |g|\, d\nu_n + \frac{1}{2\pi} \int_{\mathbb{T}\setminus \partial W} |g|\, d\theta \\
&\le C\epsilon \sum_j |\partial W_j \cap \mathbb{T}| + 2|\mathbb{T} \setminus \cup \partial W_j| \\
&\le C\epsilon.
\end{aligned}$$

Letting $\epsilon \to 0$ proves Lemma 4.3 and thus completes our function-theoretic proof of Lemma 4.1. □

A very special (and easier) case of Lemma 4.3 is:

Corollary 4.4. *Suppose* $f \in \mathcal{U}_0$ *and let* $f_r(z) = f(rz)$ *for* $r < 1$. *Then* μ_{f_r} *converges weak* to* μ_f *as* $r \to 1$.

Corollary 4.5. *If* f *is inner and* $f(0) = 0$ *then* μ_{f_r} *converges weak* to normalized Lebesgue measure on* $\mathbb{T}$.

5. A change of variables

The following result was suggested by Paul Bourdon and simplifies certain arguments from an earlier version of the paper.

Lemma 5.1. *Suppose* g *is a positive, continuous function on* $\mathbb{D}$ *and has nontangential boundary values almost everywhere on* $\mathbb{T}$. *Then, for any* $f \in \mathcal{U}$,

$$\int g(z)\, d\mu_f(z) = \frac{1}{2\pi} \int_0^{2\pi} g(f(e^{i\theta}))\, d\theta.$$

The integral on the left requires some interpretation since g is not necessarily continuous on the support of μ_f. On the interior of the disk, g is continuous and positive so the integral is well defined (possibly infinite). On the circle, μ_f is absolutely continuous with respect to Lebesgue measure and the boundary values of g are Borel, so the integral on the circle is also well defined.

Proof. Using the monotone convergence theorem we can reduce to the case when g is bounded (just truncate and let the truncation tend to ∞). So assume g is bounded by M. For any $\epsilon > 0$ we can easily construct a sawtooth region W so that $|\mathbb{T} \cap \partial W| > 1 - \epsilon$ and g extends continuously to the closure of W. Thus we can write $g = (g - h) + h$ where h is continuous, bounded by M and $g - h$ is zero on W. The lemma is true for continuous functions by the definition of μ_f, and

$$\int (g - h)\, d\mu_f \le 2M\mu_f(\overline{\mathbb{D}} \setminus W) \le 2MC\epsilon,$$

since μ_f is Carleson with a uniform constant. Similarly

$$\int (g - h) \circ f(e^{i\theta})\, d\theta \le 2MC\epsilon,$$

so taking $\epsilon \to 0$ proves the lemma. □

The following lemma is now immediate.

Lemma 5.2. *If* $g \in H^\infty$ *and* f *is inner with* $f(0) = 0$ *then* $\mu_g = \mu_{g \circ f}$.

The hyperbolic little Bloch space, $\mathcal{B}_0^h$, is defined to be the space of those holomorphic maps $f \in \mathcal{U}$ such that

$$\lim_{|z|\to 1} \frac{(1-|z|^2)|f'(z)|}{1-|f(z)|^2} = 0,$$

and is contained in the usual little Bloch space, $\mathcal{B}_0$. Schwarz's inequality implies the left side is bounded by 1 for any analytic self-map of the disk, and from this it is easy to verify that g and f are both holomorphic self-maps of the disk, and f is hyperbolic little Bloch then so is $g \circ f$. It is far from obvious that there is an inner function in the hyperbolic little Bloch space, but they do exist (see [Aleksandrov et al. 1999; Cantón 1998; Smith 1998]). This and Lemma 5.2 thus imply:

Corollary 5.3. *If $g \in \mathcal{U}$, then there is an $f \in \mathcal{B}_0^h$ such that $\mu_f = \mu_g$.*

Recall that the Hardy space, H^p , is the set of holomorphic functions g such that

$$\|g\|_{H^p} = \lim_{r\to 1} \left(\frac{1}{2\pi} \int_0^{2\pi} |g(re^{i\theta})|^p \, d\theta \right)^{1/p} < \infty.$$

Such a function has radial boundary values almost everywhere on $\mathbb{T}$, which we also denote by g. If we know $g \in H^p$ for $p > 1$, then the radial maximal function of g is in L^p and so on can use the dominated convergence theorem to deduce that

$$\|g\|_{H^p} = \frac{1}{2\pi} \int_0^{2\pi} |g(e^{i\theta})|^p \, d\theta.$$

In general, however, the right-hand side might be finite but g might not be in H^p (there exist nonzero holomorphic functions on the disk that have radial value zero almost everywhere, and hence are not in H^p). If $f \in \mathcal{U}$ then μ_f restricted to $\mathbb{T}$ is absolutely continuous with respect to Lebesgue measure, so $\int_{\overline{\mathbb{D}}} |g|^p \, d\mu_f$ makes sense.

As another application of Lemma 5.1 we can show

Lemma 5.4. *Suppose $g \in H^p$ on the unit disk and $f \in \mathcal{U}_0$. Then for any $0 < p < \infty$,*

$$\|g \circ f\|_{H^p}^p = \lim_{r\to 1} \int_{\mathbb{D}} |g|^p \, d\mu_{f_r} = \int_{\overline{\mathbb{D}}} |g|^p \, d\mu_f.$$

Proof. The first equality is the definition of the H^p norm, so we only have to prove the second. If $g \in H^p$ and $f \in \mathcal{U}_0$ then by a result of Ryff [1966], $g \circ f \in H^p$ with smaller or equal norm. Thus $|g|^p$ is positive, continuous function on the disk which has nontangential boundary values almost everywhere, so Lemma 5.1 shows that

$$\int |g(z)|^p \, d\mu_f = \frac{1}{2\pi} \int_0^{2\pi} |g(f(e^{i\theta}))|^p \, d\theta,$$

and since we already know $g \circ f \in H^p$, we can deduce that the right-hand side equals $\|g \circ f\|_{H^p}$. $\square$

6. Mapping the Bergman space into the Hardy space

For our applications to composition operators, we need a version of Lemma 5.4 that works without the assumption that $g \in H^p$. The proof given above doesn't work in general because if g is not in H^p we can't say that $\|g\|_{H^p} = \int_0^{2\pi} |g|^p \, d\theta/2\pi$. In fact, we will not even assume g has boundary values on the circle, so this integral is not necessarily defined.

Lemma 6.1. *Suppose g is holomorphic on the open unit disk, $f \in \mathcal{U}_0$ and μ_f is radial. Then, for any $0 < p < \infty$,*

$$\|g \circ f\|_{H^p}^p = \lim_{r \to 1} \int_{\mathbb{D}} |g|^p \, d\mu_{f_r} = \int_{\mathbb{D}} |g|^p \, d\mu_f + \mu_f(\mathbb{T}) \|g\|_{H^p}^p. \tag{6–1}$$

Proof. Let $g_s(z) = g(sz)$ for $0 < s < 1$. First, we want to show that, for any $0 < p < \infty$,

$$\lim_{s \to 1} \int |g(sz)|^p \, d\mu_f = \int_{\mathbb{D}} |g(z)|^p \, d\mu_f + \mu_f(\mathbb{T}) \|g\|_{H^p}^p, \tag{6–2}$$

Since g is holomorphic, $|g|^p$ is subharmonic for $0 < p < \infty$ (see, for example, [Garnett 1981, page 35]) and hence $m(r) = \frac{1}{2\pi} \int |g(re^{i\theta})|^p \, d\theta$, is defined on $[0, 1)$ and is an increasing function of r [Garnett 1981, Corollary I.6.6]. Therefore we can extend it to be defined at $r = 1$ by $\|g\|_{H^p}^p = m(1) = \lim_{r \to 1} m(r)$. Thus $m_s(r) \equiv m(sr)$ increases to $m(r)$ as $s \to 1$ for all $r \in [0, 1]$. Let ν be the measure on $[0, 1]$ defined by $\nu(E) = \mu_f(\{z : |z| \in E\})$. Since μ_f is radial we have

$$\int \varphi \, d\mu_f = \frac{1}{2\pi} \int_0^1 \int_0^{2\pi} \varphi(re^{i\theta}) \, d\theta \, d\nu(r).$$

Thus by the monotone convergence theorem,

$$\lim_{s \to 1} \int |g_s|^p \, d\mu_f = \lim_{s \to 1} \int m_s(r) \, d\nu = \int_{[0,1]} m(r) \, d\nu = \int_{\mathbb{D}} |g|^p \, d\mu_f + \mu_f(\mathbb{T}) m(1).$$

This is (6–2).

We will break the proof of (6–1) into three cases.

Case 1: $\int_{\mathbb{D}} |g|^p \, d\mu_f = \infty$.

For any $M > 0$ choose $0 < t < 1$ so that $\int_{|z|<t} |g|^p \, d\mu_f > 2M$ and write $|g|^p = g_1 + g_2$ where g_1 and g_2 are nonnegative, $g_1 = |g|^p$ on $|z| < t$, and g_1 is continuous and compactly supported in $\mathbb{D}$. Then

$$\int |g|^p \, d\mu_{f_r} \geq \int g_1 \, d\mu_{f_r} > \frac{1}{2} \int g_1 \, d\mu_f \geq M$$

if r is close enough to 1. Thus $\int |g|^p \, d\mu_{f_r} \to \infty = \int |g|^p \, d\mu_f$.

Case 2: $\int_{\mathbb{D}} |g|^p \, d\mu_f < \infty$ and $\mu_f(\mathbb{T}) = 0$.

Since μ_{f_r} converges weak* to μ_f,

$$\lim_{r\to 1} \int |g_s|^p \, d\mu_{f_r} = \int |g_s|^p \, d\mu_f$$

for any fixed $s < 1$. Since $g_s(f(z))$ is holomorphic on the open disk, $|g_s(f(z))|^p$ is subharmonic. Thus $\int |g_s|^p \, d\mu_{f_r}$ is increasing in r, and hence

$$\int |g_s|^p \, d\mu_{f_r} \le \int |g_s|^p \, d\mu_f.$$

Now take $s \to 1$. For r fixed, μ_{f_r} is compactly supported in the disk, so $|g_s|^p$ is uniformly bounded on its support and hence the left-hand side converges to $\int |g|^p \, d\mu_{f_r}$. Condition (6–2) implies the right-hand side converges to $\int |g|^p \, d\mu_f$. Thus

$$\int |g|^p \, d\mu_{f_r} \le \int |g|^p \, d\mu_f$$

for all $r < 1$.

Fix $\epsilon > 0$ and choose $0 < t < 1$ so that $\int_{t<|z|<1} |g|^p \, d\mu_f < \epsilon$. Write $|g|^p = g_1 + g_2$ as in Case 1. Thus $\int g_2 \mu_f < \epsilon$. Also, if r is close enough to 1 then, by weak* convergence,

$$\left| \int g_1 \, d\mu_f - \int g_1 \, d\mu_{f_r} \right| < \epsilon.$$

Thus

$$\int g_2 \, d\mu_{f_r} \le \left| \int g_1 \, d\mu_f - \int g_1 \, d\mu_{f_r} \right| + \int g_2 \, d\mu_f \le 2\epsilon.$$

Hence

$$\begin{aligned} \left| \int |g|^p \, d\mu_f - \int |g|^p \, d\mu_{f_r} \right| &\le \int g_2 \, d\mu_{f_r} + \left| \int g_1 \, d\mu_f - \int g_1 \, d\mu_{f_r} \right| + \int g_2 \, d\mu_f \\ &\le 4\epsilon, \end{aligned}$$

if r is close enough to 1.

Case 3: $\int_{\mathbb{D}} |g|^p \, d\mu_f < \infty$ and $\mu_f(\mathbb{T}) > 0$.

If $\lim_{r\to 1} \int |g|^p \, d\mu_{f_r} = \infty$ then by the subharmonicity of $|g \circ f|^p$ we see that $\int |g|^p \, d\mu_f = \infty$, so (6–1) holds. Thus we may assume that $\lim_{r\to 1} \int |g|^p \, d\mu_{f_r} < \infty$, that is, we may assume that $g \circ f \in H^p$, and hence that $|g(f(z))|^p$ has a harmonic majorant u on $\mathbb{D}$ (see [Garnett 1981, Lemma II.1.1]).

First we show that $g \in H^p$. For $0 < r < 1$ let $D_r = D(0, r)$. Let Ω_r be the component of $f^{-1}(D_r)$ which contains the origin, and let ω_r be the harmonic measure on Ω_r with respect to the origin. Let ν_r be the push-forward of ω_r under the map f. Then clearly ν_r is supported on $\overline{D}_r$ and $\nu_r(E) \le \mu_f(E)$ for any $E \subset D_r$. By Lemma 2.6, ν_r on $C_r = \partial D_r$ must be $\frac{1}{2\pi}\,d\theta$ minus the balayage of ν_r restricted to D_r. Since $\nu_r \le \mu_f$, this means that ν_r on C_r is at least $\frac{1}{2\pi}\,d\theta$ minus the balayage of μ_f restricted to D_r. Since μ_f is radial, its balayage onto C_r is also radial, that is, equal to $\frac{1}{2\pi}\mu_f(D_r)\,d\theta \le \frac{1}{2\pi}(1-\mu_f(\mathbb{T}))\,d\theta$. Thus $\nu_r \ge \frac{1}{2\pi}\mu_f(\mathbb{T})\,d\theta$ on C_r. Hence, for any g holomorphic on $\mathbb{D}$,

$$\frac{1}{2\pi}\int_0^{2\pi} |g(re^{i\theta})|^p\,d\theta \le \frac{1}{\mu_f(\mathbb{T})}\int |g|^p\,d\nu_r = \frac{1}{\mu_f(\mathbb{T})}\int |g\circ f|^p\,d\omega_r.$$

Thus, if u is a harmonic majorant of $|g\circ f|^p$ on $\mathbb{D}$,

$$\frac{1}{2\pi}\int_0^{2\pi} |g(re^{i\theta})|^p\,d\theta \le \frac{1}{\mu_f(\mathbb{T})}\int u\,d\omega_r = \frac{u(0)}{\mu_f(\mathbb{T})} < \infty.$$

In other words, $g \in H^p$ and thus (6–1) follows from Lemma 5.4. □

Recall that the Bergman space A^p is defined as the set of holomorphic functions g on the disk $\mathbb{D}$ such that

$$\|g\|_{A^p} = \Big(\frac{1}{\pi}\int_{\mathbb{D}} |g|^p\,dx\,dy\Big)^{1/p} < \infty.$$

Corollary 6.2. *If $f \in H^\infty$ such that $d\mu_f = \frac{1}{\pi}\chi_{\mathbb{D}}\,dx\,dy$, then any function g, analytic on the disk, is in the Bergman space if and only if $g\circ f$ is in the Hardy space, and $\|g\|_{A^p} = \|g\circ f\|_{H^p}$, that is, the composition operator $C_f : A^p \to H^p$ is an isometry.*

Proof. Using Lemma 6.1 we see that

$$\begin{aligned}\|g\circ f\|_{H^p} &= \lim_{r\to 1}\Big(\frac{1}{2\pi}\int_0^{2\pi} |g(f(re^{i\theta}))|^p\,d\theta\Big)^{1/p}\\ &= \lim_{r\to 1}\Big(\int |g|^p\,d\mu_{f_r}\Big)^{1/p} = \Big(\int |g|^p\,d\mu_f\Big)^{1/p} = \|g\|_{A^p}.\end{aligned}$$ □

This corollary may seem a little surprising, since functions in H^p have nontangential limits almost everywhere, whereas those in A^p need not, but since f has almost all of its boundary values in the interior of the disk, this is not a contradiction. Of course, it still remains to show (see Section 9) that there is an $f \in H^\infty$ such that μ_f is area measure.

Corollary 6.2 obviously holds for any weighted Bergman space where the weight is a radial measure of finite mass satisfying the integral condition (1–1) in Theorem 1.1. If instead of an isometry, we merely want $\|g\|_{A_p} \simeq \|g\circ f\|_{H^2}$ we could take a much bigger class of functions f, for example, $\mu_f = w\,dx\,dy$ for some weight w which is bounded above and below on an annulus $\{r < |z| < 1\}$. Constructing such examples only needs the techniques of Section 8, not the full proof of Theorem 1.1.

Similarly, by appropriate choices of μ_f one can construct composition operators on H^p which satisfy conditions like

$$\|C_f(g)\|^p_{H^p} = \tfrac{1}{2}\|g\|^p_{H^p} + \tfrac{1}{2}\|g\|^p_{A^p} \quad \text{or} \quad \|C_f(g)\|^p_{H^p} = \tfrac{1}{2}\|g\|^p_{H^p} + \tfrac{1}{2}\|g_{1/2}\|^p_{H^p}.$$

In [Cima and Hansen 1990], a function f is said to have property $(*)$ relative to H^p if $g \circ f \in H^p$ implies that $g \in H^p$, for any holomorphic g on $\mathbb{D}$. Paul Bourdon has pointed out that for general $f \in \mathcal{U}$, the condition $\mu_f(\mathbb{T}) = 0$ implies condition $(*)$, which implies $N_f(z) = o(1-|z|)$ which, by J. Shapiro's theorem [1987], implies that C_f is compact and hence does not have a bounded right inverse. Since f is nonconstant, C_f is 1-to-1 and so does not have closed range (this is a consequence of the open mapping theorem, for example [Rudin 1973, Corollary 2.12c]). Thus C_f does not have property $(*)$, since any function in $\overline{C_f(H^p)} \setminus C_f(H^p)$ is an H^p function without an H^p preimage. Lemma 6.1 clearly implies the following corollary.

Corollary 6.3. *If $f \in \mathcal{U}_0$ is orthogonal, then f has property $(*)$ relative to H^p if and only if $\mu_f(\mathbb{T}) > 0$.*

Proof. If $\mu_f(\mathbb{T}) > 0$ then the argument in Case 3 of the proof of Lemma 6.1 shows that $g \circ f \in H^p$ implies $g \in H^p$. Thus f has property $(*)$ with respect to H^p. □

A special case of Corollary 6.3 is when $\mu_f(\mathbb{T}) = 1$, that is, all inner functions have property $(*)$. It would be very interesting to have a similar characterization of property $(*)$ for general functions in $\mathcal{U}_0$.

7. An example of μ_f supported on two circles

In this section we will construct an $f \in H^\infty$ so that μ_f is supported on the union of two circles $C_{1/2}$ and C_1 (where $C_r = \{z : |z| = r\}$) and is a multiple of Lebesgue measure on each. This example suffices to disprove Rudin's orthogonality conjecture, and introduces the estimates and techniques needed for the general case of Theorem 1.1. In the next section we will show that any radial probability measure supported in $\{\frac{1}{2} \le |z| \le 1\}$ can occur as a μ_f, and in Section 9 we will do the general case of measures supported on $\mathbb{D}$.

Based on Lemmas 4.1 and 2.5, it suffices to build an increasing sequence of Riemann surfaces $\{R_n\}$ so that the corresponding maps $\{f_n\}$ satisfy $f_n(0) = 0$, that μ_{f_n} is supported on the two circles $C_{1/2} \cup C_1$, and that μ_{f_n} restricted to $C_{1/2}$ is of the form $\frac{1}{2\pi} g_n(\theta)\, d\theta$, where g_n converges uniformly to a positive constant.

We start by taking $f_0(z) = \frac{1}{2}z$, that is, f_0 is the (trivial) Riemann mapping from $\mathbb{D}$ to the disk $R_0 = \{|z| < 1/2\}$. The corresponding measure $\mu_0 = \mu_{f_0}$ is normalized Lebesgue measure on the circle $C_{1/2}$, that is, $\frac{1}{2\pi} g_0(\theta)\, d\theta$ where $g_0(\theta) = 1$.

Now we describe the idea of the construction of R_1 (we will give the details later). First we replace R_0 with a slightly smaller disk, S_1. We divide the boundary of S_1 into a large number of alternating intervals which we call type I and type J. Along each type I interval we attach a copy of a certain Riemann surface with boundary over $C_{1/2}$ (attaching different copies to different intervals) and along each type J interval we attach copies of certain surfaces with boundary over C_1. This gives the surface R_1. With appropriate choices of the parameters involved we can show that, with high probability, the Brownian paths which first hit ∂S_1 at a type I interval go on to hit the part of ∂R_1 over $C_{1/2}$ and the paths which hit the J intervals go on to hit ∂R_1 over C_1. Thus we have "rerouted" a certain fraction of the harmonic measure on $C_{1/2}$ out to C_1. By choosing various parameters correctly, we can make the harmonic measure over $C_{1/2}$ in R_1 be close to any multiple of Lebesgue measure we want (as long as the total mass is less than 1). The resulting measure may not be radial but, by iterating the construction with variable size barriers, we can make harmonic measure as close to a multiple of Lebesgue measure as we wish, obtaining a radial measure in the limit.

Now we give the construction of R_1 in more detail. Choose δ_1 very small and let $S_1 = D(0, r_1)$, where $r_1 = \frac{1}{2} - \delta_1$. Obviously harmonic measure on S_1 is just normalized Lebesgue measure on its boundary. Choose a large integer m_1 and points $\{z_j : j = 1, \dots, m_1\}$ equally spaced on the circle C_{r_1}. Choose a continuous function $0 < \eta(x) < 1$ on C_{r_1}, let I_j be an arc of ∂S_1 of angle measure $\eta(z_j)2\pi/m_1$ centered at z_j, and let $\{J_j\}$ be the complementary arcs. For the first step of the construction we can take $\eta(x) = \eta_1$ to be a constant for simplicity, but in later steps we will have to use nonconstant η's.

Fix some $0 < \tau_1 < 1$ and, for each arc of the form I_j with endpoints $\{p, q\}$, choose a countable collection of points $E = \{w_k^j\} \subset I_j$, accumulating only at the endpoints of I_j, so that for any $z \in I_j$

$$\operatorname{dist}(z, E) \le \tau_1 \operatorname{dist}(z, \{p, q\}). \tag{7–1}$$

Let the components of $I_j \setminus E$ be denoted $\{I_k^j\}$. For each I_k^j, consider the (infinitely connected) planar domain $\mathbb{D} \setminus E$ and the universal cover of the domain. Take a copy of the arc I_k^j in the universal cover; it is on the boundary of a simply connected domain D in the universal cover which covers $D(0, \frac{1}{2})$. The arc cuts the universal cover into two components and we let R_k^j denote the component which does not contain D. For each interval I_k^j, we attach a copy of R_k^j to S_1 along the arc I_k^j.

For the intervals $\{J_j\}$ we follow the same procedure, defining a set $E \subset J_j$ and sub intervals $\{J_k^j\}$, but replacing $D(0, \frac{1}{2})$ with $D(0, 1)$. That is, we attach a component of the universal cover of $D(0, 1) \setminus E$, cut along J_k^j. Doing this for all j and k gives the surface R_1. The harmonic measure for R_1 is now supported on $C_{1/2} \cup C_1$, (the rest of the ideal boundary covers a countable set, so has zero measure) so we only need to check that it is still close to radial on $C_{1/2}$.

Now we want to discuss the two main estimates for describing the harmonic measure of R_1. The first says that a continuous convolution of the Poisson kernel is well approximated by a discrete version if the sample points are sufficiently close together. The second says that the harmonic measure of I intervals is small when viewed from a J interval, and vice versa.

Suppose $D(0, r)$ is a disk and g is a continuous function on a smaller circle C_s, $s < r$. The balayage of g onto the circle C_r is

$$Bg(\theta) = \int_0^{2\pi} g(se^{it}) P_{se^{it}}(\theta)\, dt,$$

where $P_z(\theta)$ is the Poisson kernel for $D(0, r)$ with respect to the point z.

Lemma 7.1. *With the intervals $\{I_j\}$ defined as above, and $F = \bigcup_j I_j$, for any continuous $0 < g < 1$ on the circle C_s*

$$B(g\chi_F)(\theta) = \int_F g(se^{it}) P_{se^{it}}(\theta)\, dt \to B(g\eta)(\theta),$$

uniformly as $m_1 \to \infty$.

Proof. Let K_j be the interval on C_s, centered at z_j, of angle measure $2\pi/m_1$ (choose them to be half-open, so that they form a disjoint cover of the circle). Define piecewise constant functions $a(x)$ and $b(x)$ on C_{r_1} by

$$a(x) = \sum_j \chi_{K_j}(x)\eta(z_j), \quad b(x, \theta) = \sum_j \chi_{K_j}(x) g(z_j) P_{z_j}(\theta),$$

and let

$$A(m_1) = \|\eta(z) - a(z)\|_\infty, \quad B(m_1) = \|g(x)P_x(\theta) - b(x, \theta)\|_\infty.$$

It is clear that, by uniform continuity, both quantities tend to zero as $m_1 \to \infty$. Thus by using the fact that $\chi_F(x) - a(x)$ has mean value zero on each interval K_j where $b(x, \theta)$ is constant in x we get

$$\begin{aligned}
&|B(g\chi_F)(\theta) - B(g\eta)(\theta)| \\
&\quad = \left| \int_0^{2\pi} (g(se^{it}) P_{se^{it}}(\theta) - b(se^{it}, \theta) + b(se^{it}, \theta))(\chi_F(se^{it}) - \eta(se^{it}))\, dt \right| \\
&\quad \le B(m_1) \int_0^{2\pi} \left|(\chi_F - \eta(se^{it})\right| dt + \int_0^{2\pi} b(se^{it}, \theta) \left|a(se^{it}) - \eta(se^{it})\right| dt \\
&\quad \le 2\pi B(m_1) + A(m_1) \max |b|.
\end{aligned}$$

This clearly tends to zero as $m_1 \to \infty$, as desired. □

Now for the second estimate. We want to show that the harmonic measure of $C_{1/2}$ is much larger than that of C_1 with respect to a point $z \in I_k^j$.

Lemma 7.2. *Suppose that $z \in I_k^j$, and suppose that γ is a circular arc in S_1 with endpoints in the corresponding set E such that* $\operatorname{dist}(\gamma, z) \simeq \operatorname{dist}(z, \{p, q\})$ *(with constants independent of τ_1), and which separates z from all the J-intervals. Let Ω be the component of $R_1 \setminus \gamma$ which contains z. Then $\omega(z, \gamma, \Omega) \to 0$ as τ_1 does.*

Proof. Standard estimates of hyperbolic metric imply that γ is within a bounded hyperbolic distance of a geodesic in R_1, and that the hyperbolic distance from γ to z is at least $C \log \tau_1^{-1}$. Lifted to the disk, this implies the harmonic measure of γ with respect to z is $\le \exp(C \log \tau_1) \le \tau_1^\alpha$, for some $\alpha > 0$, as desired. Obviously, the same estimate holds if we reverse the rôles of the I and J intervals. □

The previous result has a simple explanation in terms of Brownian motion. Consider a Brownian motion on the Riemann surface started at z and run until it either hits γ or leaves R_1. The path will only hit γ if it stays on the correct sheet of R_1, but this is extremely unlikely because it will cross the arc I_j many times and each time it has a certain chance (which is large if τ is small) of becoming "tangled" and ending up on the wrong sheet.

We can now show that the harmonic measure of R_1 on the circle $C_{1/2}$ can be taken as close to a multiple of Lebesgue measure as we wish (depending on our choices of m_1, τ_1 and η). The harmonic measure of R_1 on the circle $C_{1/2}$ will be the balayage of the harmonic measure of S_1 restricted to the I intervals, with an error bounded by $C\tau_1^\alpha$. The harmonic measure is (normalized) angle measure restricted to the I-intervals. Thus if m_1 is large enough, the harmonic measure on $C_{1/2}$ will be of the form $\frac{1}{2\pi} g_1(x)\, d\theta$, with g_1 as close to a constant as we wish. Take $\frac{1}{2} + \frac{1}{10} \le g_1(x) \le \frac{1}{2} + \frac{3}{10}$, to be concrete.

Now suppose we have constructed R_{n-1}. To construct R_n, we follow the method above. We start passing to a subsurface $S_n \subset R_{n-1}$ where the boundary circles over $C_{1/2}$ are replaced by boundaries over $C_{1/2-\delta_n}$. The parameter δ_n is chosen so small that every component of $R_{n-1} \backslash S_n$ is a regular cover of the annulus $\{\frac{1}{2} - \delta_n < |z| < \frac{1}{2}\}$ (which will be possible by the construction of R_{n-1}) and so that harmonic measure μ_{S_n} on S_n is very close to harmonic measure on R_{n-1}, say

$$\left| \int \varphi d(\mu_{S_n} - \mu_{R_{n-1}}) \right| \le 2^{-n} \tag{7–2}$$

for every smooth φ with gradient bounded by n.

As before we choose m_n equally spaced points $\{z_j^n\}$ on $C_n = C_{\frac{1}{2}-\delta_n}$ and define intervals $\{I_j^n\}$ of C_n, centered at these points, of angle measure $2\pi\eta_n(z_j^n)/m_n$, where

$$\eta_n(x) = \left(\frac{1}{2} + \frac{2}{10^n}\right) / g_{n-1}(x).$$

The complementary intervals are denoted $\{J_j^n\}$. We choose a very small τ_n and sets E in each interval which satisfies (7–1) with τ_n. We then attach copies $D(0, \frac{1}{2}) \setminus E$

to the copies of the I intervals in ∂S_n and copies of $D(0, 1) \setminus E$ to the J intervals. Then if we choose δ_n and τ_n small enough and m_n large enough, we can get the harmonic measure of R_n over $C_{1/2}$ to be $g_n(x)\, d\theta/2\pi$ with g_n as close to $g_{n-1}\eta_n$ as we wish, say

$$\frac{1}{2} + \frac{1}{10^n} \leq g_n \leq \frac{1}{2} + \frac{3}{10^n}.$$

Continuing in this way we can clearly construct a sequence $\{R_n\}$ of Riemann surfaces so that the harmonic measures over $C_{1/2}$ converge to a multiple of Lebesgue measure. This almost finishes the proof, except that the surfaces $\{R_n\}$ are not nested by inclusion. However, the subsurfaces $\{S_n\}$ constructed as part of the induction are nested and their union is also R. Hence their harmonic measures converge to that of R. By (7–2), the weak* limit for the measures on $\{S_n\}$ and $\{R_n\}$ must be the same, so we are done.

The same proof shows that we can build an $f \in H^\infty$ so that $\mu_f|_{C_{1/2}} = \frac{1}{2\pi} g\, d\theta$ for any continuous g with $0 \leq g < 1$ (or any g which is the decreasing limit of such functions). Similarly, the circle can be replaced by any smooth curve γ, and g by a continuous function such that $g\, ds \leq d\omega(0, \cdot, \mathbb{D} \setminus \gamma)$.

The construction in this section clearly generalizes as follows.

Lemma 7.3. *Suppose R is a Riemann surface built by attaching subdomains of $\mathbb{D}$ along boundary arcs. Let Π denote the corresponding projection of R into the plane. Suppose $\Pi(\partial R)$ hits C_r and there is a $\delta > 0$ such that every component of $\Pi^{-1}(C_r)$ in ∂R is the boundary of a domain in R which is a regular cover of the annulus $\{r - \delta < |z| < r\}$ (or $\{r < |z| < r + \delta\}$). Suppose the harmonic measure of R over C_r projects to a measure of the form $\frac{1}{2\pi} g\, d\theta$ on C_r, where $0 < g < 1$. Choose $s < r$ (or $s > r$) very close to r. Suppose we are given N functions $\{\eta_k\}$ such that $0 < \eta_k < 1$. Choose a large integer m and choose mN equally spaced points $\{z_i\}$ on C_s. Let I_j^k be the interval of length $2\pi \eta_k(z_{k+jN})/mN$ centered at z_{k+jN}. Let J_i denote the components of $C_s \setminus \bigcup_{j,k} I_j^k$. Choose a small τ and choose sets E satisfying (7–1) in every interval. For $k = 0, \dots, N$, choose $s_k < s < r_k$. For each arc in ∂R projecting to I_j^k attach a copy of $A_k \setminus E = \{s_k < |z| < r_k\} \setminus E$. To each arc projecting to a J_i attach a copy of $A_0 \setminus E = \{s_0 < |z| < r_0\} \setminus E$. If s is close enough to r, if m is large enough and if τ is small enough, then the projected harmonic measure of the new surface S on $\partial S \setminus R$ is as close to $\sum_k B_k(\eta_k g)$ as we wish, where B_k denotes balayage from C_s onto ∂A_k.*

For the proof of Theorem 1.1, we can always take $s_k = 0$, that is, we can attach disks instead of annuli. Only for the proof of Corollary 1.4 will we have to attach proper annuli.

8. Theorem 1.1 on an annulus

In this section we will show that any radial probability measure μ supported in the annulus $\{z : \frac{1}{2} \le |z| \le 1\}$ is of the form μ_f for some $f \in \mathcal{U}_0$, and in the next section we will extend this to the general case.

First some notation. For $0 < r < s < 1$ let $A(r, s) = \{z : r \le |z| < s\}$. When $s = 1$, we let $A(r, 1) = \{z : r \le |z| \le 1\}$. For $0 < r < 1$, let $\mu(r) = \mu(A(0, r))$. Let $r^0 = \frac{1}{2}$, let $r_0^1 = \frac{1}{2}$, let $r_1^1 = \frac{3}{4}$ and, more generally, let $r_k^n = \frac{1}{2} + k2^{-n-1}$ for $k = 0, \dots, 2^n - 1$. Let $\mu_k^n = \mu\big(A\big(r_k^n, r_{k+1}^n\big)\big)$, and let $C_k^n = C_{r_k^n}$.

By rescaling, we may assume that $\mathbb{T} \subset \operatorname{supp}(\mu) \subset \overline{\mathbb{D}}$ and hence that $\mu_{2^n-1}^n$ is positive for all n.

We will construct a sequence $R_0 \subset R_1 \subset \cdots$ of Riemann surfaces, such that the corresponding measure μ_n is supported on the union of 2^n circles, $\bigcup_{k=0}^{2^n-1} C_k^n$. On C_k^n the measure μ_n will have the form $\frac{1}{2\pi} g_k^n \, d\theta$ where

$$\mu_k^n < g_k^n \le \mu_k^n + \epsilon_n \tag{8–1}$$

for $k = 0, \dots, 2^n - 2$ and any $\epsilon_n > 0$ we choose, and for $k = 2^n - 1$ we have

$$\mu_{2^{n+1}-2}^{n+1} < g_k^n \le \mu_{2^n-1}^n. \tag{8–2}$$

Recall that since μ_n is a probability measure, if it gives too much mass to the first $2^n - 1$ annuli, then it must give too little to the last one. It is obvious that such measures $\{\mu_n\}$ converge weak* to μ, so by the argument at the end of the previous section, the μ_f corresponding to the limiting surface $R = \bigcup_n R_n$ must equal μ.

Thus it only remains to construct the surfaces. As in the previous section we start with $R_0 = D\big(0, \frac{1}{2}\big)$. To construct R_1, we will proceed exactly as in the previous section, except that instead of redirecting harmonic measure to the unit circle, we send it to the circle $C_{3/4}$. The estimates are all the same so we can obtain a surface R_1 such that the corresponding μ_1 is supported on $C_{1/2} \cup C_{3/4}$ and is of the form $\frac{1}{2\pi} g_0^1 \, d\theta$ on C_0^1 and $\frac{1}{2\pi} g_1^1 \, d\theta$ on C_1^1 where

$$\mu_0^1 < g_0^1 < \mu_0^1 + \epsilon_1 \quad \text{and} \quad \mu_2^2 < g_1^1 < \mu_1^1,$$

for any $\epsilon_1 > 0$ we choose.

To construct R_{n+1} for $n \ge 1$, we just make one small change. The mass on the outermost circle $C_{2^n-1}^n$ is redistributed to itself, $C_{2^n-1}^n = C_{2^{n+1}-2}^{n+1}$, and to the outermost circle of the next stage, $C_{2^{n+1}-1}^{n+1}$. The mass of any other circle C_j^n is redistributed to three circles; itself, $C_j^n = C_{2j}^{n+1}$, the next circle out in the next generation, C_{2j+1}^{n+1} and the outermost circle of the next generation, $C_{2^{n+1}-1}^{n+1}$.

To do this we let $\tilde{C}_j^n$ be the circle of radius $r_j^n - \delta_n$, where $\delta_n < 2^{-n-10}$ is chosen so small that the harmonic measure on S_n (the subsurface of R_n bounded by the lifts of $\tilde{C}_j^n$ which contain 0 and hence contain S_{n-1}) is as close as we wish to harmonic measure on R_{n-1}, that is, it satisfies (7–2). We now just apply the construction of Lemma 7.3, with $N=2$, $s_0=s_1=s_2=0$, $r_0=r^{n+1}_{2j}$, $r_1=r^{n+1}_{2j+1}$ and $r_2=r^{n+1}_{2^{n+1}-1}$. More precisely, suppose that we have two continuous functions η_1 and η_2 defined on $\tilde{C}^n_j$, such that $\eta_1+\eta_2<2$, together with m_n equidistributed points $\{z_j\}$ on ∂S_n, and choose intervals centered at these points. However, instead of having two types of intervals, we will have three: $\{I_j\}$ of angle measure $2\pi\eta_1(\theta)/m_n$ centered at z_j for j even, $\{K_j\}$ of angle measure $2\pi\eta_2(\theta)/m_n$ centered at z_j for j odd, and the remaining intervals $\{J_j\}$. We choose a very small τ_n and a countable set E in each interval which satisfies (7–1). Then along type I intervals we attach a copy of the universal cover of $D(0,r^n_j)\setminus E$, along the type K intervals we attach the universal cover of $D(0,r^{n+1}_{2j+1})\setminus E$, and along the type J intervals we attach that of $D(0,r^{n+1}_{2^{n+1}-1})\setminus E$. Then if we take m_n large enough and δ_n and τ_n small enough, the harmonic measure of the surface R_{n+1} over C^n_j will be as close to the balayage of $\eta_1 g^n_j$ onto C^n_j as we wish and the harmonic measure over C^{n+1}_{2j+1} will be as close to the balayage of $\eta_2 g^n_j$ onto that circle as we wish, independent of what changes we make at circles other than C^n_j.

Now do a similar construction around each circle C^n_j, for $j=0,\dots,2^n-2$. At the outermost circle $C^n_{2^n-1}$, we redirect the measure to only two circles: itself and the outermost circle of the next generation, $C^{n+1}_{2^{n+1}-1}$. By construction, condition (8–1) holds with any constant ϵ_n we want. Then by Lemma 2.6, μ_n on the outermost circle must be normalized Lebesgue measure minus the balayage of the measures on the inner circles. Since these measures have total mass as close to, but larger than,

$$\mu\big(A\big(\tfrac{1}{2},r^n_{2^n-1}\big)\big)=\sum_{j=0}^{2^n-2}\mu^n_j$$

r as we wish, the mass of the outermost circle is as close to, but smaller than, $\mu\big(A\big(r^n_{2^n-1},1\big)\big)=\mu^n_{2^n-1}$. Moreover, since the measures on the inner circles are as close to radial as we wish, so is their balayage onto the outermost circle and hence so is μ_f restricted to the outermost circle (this condition defines our choice of ϵ_n). This gives condition (8–2). The proof is completed by taking limits just as before.

9. Theorem 1.1 on the whole disk

To complete the proof of Theorem 1.1 we need to show how to obtain any measure satisfying (1–1). As in the last section we can assume $\mathbb{T}\subset\operatorname{supp}(\mu)\subset\overline{\mathbb{D}}$. We can also simplify the situation slightly by observing that it is enough to assume that most of the mass of μ lives away from the origin, that is,

$$\int\log\frac{1}{|z|}\,d\mu\le\delta. \tag{9–1}$$

This is because for $f \in H^\infty$ the measure μ_{f^d} is the push-forward under $z \to z^d$ of the measure μ_f and so

$$\int \log \frac{1}{|z|}\, d\mu_f = \frac{1}{d} \int \log \frac{1}{|z|}\, d\mu_{f^d}.$$

By taking d large we can make the right-hand side as small as we wish. Thus for any μ on the disk satisfying (1–1), it suffices to construct an f corresponding to the pull-back of μ under z^d, that is, it suffices to consider only measures satisfying (9–1) for any $\delta > 0$ we choose.

Start by taking $R_0 = D\big(0, \frac{1}{4}\big)$. Let $r_n = 2^{-n}$ for $n = 0, 1, 2, \dots$ and let $\mu_n = \mu(A(r_n, r_{n-1}))$. Then

$$\sum_{n>2} (n-1)(\log 2)\mu_n \le \int \log \frac{1}{|z|}\, d\mu \le \delta, \tag{9–2}$$

so

$$\mu_n \le \frac{\delta}{(\log 2)(n-1)} \le \frac{\delta'}{n}, \tag{9–3}$$

where δ' is as small as we wish.

We need two simple facts about harmonic measure on an annulus.

Lemma 9.1. *Suppose $A = \{z : s < |z| < r\}$ and $s < t < r$. Then $\omega(z, C_s, A) = u_{s,r}(z) = (\log|z| - \log r)/(\log s - \log r)$ for any z with $|z| = t$.*

Proof. This is immediate since the given function is harmonic in A, equals 1 on C_s and equals 0 on C_r. □

Lemma 9.2. *Suppose s, t, r and A are as in Lemma 9.1. Then if $t \ge 2s$, there is an $M < \infty$, independent of s, t and r, such that for $|z| = t$, $\omega(z, \cdot, A)$ restricted to C_s has the form $\frac{1}{2\pi} g\, d\theta$ and g satisfies $\max_{C_s} g \le M \min_{C_s} g$.*

Proof. Recall that harmonic measure on ∂A is the normal derivative of Green's function G with pole at z. Let $t' = \frac{2}{3}t > s$. By Harnack's inequality there is an M such that $\max_{C_{t'}} G \le M \min_{C_{t'}} G$, and hence there is a constant C such that

$$C(1 - u_{s,t'}) \le G \le MC(1 - u_{s,t'}),$$

on $\{s < |z| < t'\}$. Since the normal derivative of $u_{s,r'}$ is constant on C_s (since u is radial), this implies the normal derivative of G on C_s is trapped between two constants A and MA, as desired. □

Consider the annulus $A_n = \{z : 2^{-n} < |z| < 2^{-1}, n = 3, 4, \dots\}$ and a point z such that $|z| = \frac{1}{3}$. The two previous results imply that there is a constant B such that harmonic measure for A on the circle $C_{2^{-n}}$ is of the form $\frac{1}{2\pi} g\, d\theta$ where $g \ge B/n$ for $n \ge 3$. By (9–2) we can assume μ is chosen so that $\sum_n n\mu_n \le (2B)^{-1}$.

Thus $\sum_n Bn\mu_n \le \frac{1}{2}$, and hence it is possible to choose a collection of disjoint, adjacent intervals $\{I_n : n = 2, 3, 4 \dots\}$ on $C_{1/4}$, of angle measure $4\pi n\mu_n/B$. In each interval I_n choose a countable set E_n satisfying the "thickness" condition (7–1) with some τ_n, and attach to I_n a copy of the universal cover of $A_{n+1} \setminus E_n$. The resulting Riemann surface has harmonic measures supported over the union of circles $\bigcup_n C_{2^{-n}}$ for $n = 1, 3, 4, 5, \dots$ and, moreover, if we choose $\tau_n \to 0$ quickly enough, the harmonic measure of the circles corresponding to $n = 3, 4, 5 \dots$ is of the form $\frac{1}{2\pi} g_n\, d\theta$ with $g_n > \mu_{n-1}$, but might not be close to radial.

For each such circle $C_{2^{-n}}$, choose I and J intervals in the usual way and attach copies of $D\big(0, \frac{1}{2}\big) \setminus E$ and $D\big(0, 2^{-n}\big) \setminus E$ respectively. As we have seen before, we can choose η, m and τ so that the harmonic measure $\frac{1}{2\pi} g_n\, d\theta$ on $C_{2^{-n}}$ is as close to (but larger than) μ_n as we wish. Using Lemma 2.6, the harmonic measure of $C_{1/2}$ will be as close to (but less than) μ_1 as we wish and, in particular, it is larger than $\mu\big(\big\{\frac{1}{2} \le |z| < \frac{3}{4}\big\}\big)$ (this is where we use the assumption that $\mathbb{T}$ is in the support of μ).

The rest of the proof is now the same as the previous section. On each annulus we redistribute the harmonic measure from the circle into the annulus, sending any "extra" measure to the outermost circle, $C_{1-2^{-n}}$. In the limit, we obtain the desired measure μ.

10. An example which is almost an outer function

In this section we will construct an orthogonal function f whose only inner factor is the required zero at 0, that is, $f(z)/z$ is outer. We will construct f so that 0 is the only zero of f; thus $f(z)/z$ has no Blaschke factor. In order to prove it has no singular inner factor, recall that if $f(z)/z = gh$ with g outer and h a nontrivial singular inner function, then

$$\log |f|^{-1} = \log |g|^{-1} + \log |h|^{-1},$$

and that the first term on the right is the Poisson integral of its boundary values on $\mathbb{T}$, but that the second term is the Poisson integral of a singular measure on $\mathbb{T}$ and has boundary value zero almost everywhere on $\mathbb{T}$. Let

$$H_\epsilon = \{z \in \mathbb{D} : |h(z)| < \epsilon\} \quad \text{and} \quad F_\epsilon = \{z \in \mathbb{D} : |f(z)| < \epsilon\}.$$

Since $\log |h(0)|^{-1} = \log(1/\epsilon)\omega(0, H_\epsilon, \mathbb{D} \setminus H_\epsilon)$, we deduce that

$$\omega(0, H_\epsilon, \mathbb{D} \setminus H_\epsilon) \ge C/\log(1/\epsilon),$$

where $C = \log |h(0)|^{-1}$ and consequently, since $H_\epsilon \subset F_\epsilon$,

$$\omega(0, F_\epsilon, \mathbb{D} \setminus F_\epsilon) \ge C/\log(1/\epsilon). \tag{10–1}$$

We will construct R so that the harmonic measure of $\{z \in R \setminus D(0, \frac{1}{2}) : |z| \le 2^{-n}\}$ has harmonic measure (in R, with respect to 0) less than λ^n for some $\lambda < 1$. This contradicts (10–1), so the covering map has no singular inner factor.

Since we have already seen several constructions of this type in great detail, I will only sketch the construction. Start with $R_0 = D\big(0, \frac{1}{2}\big)$. Divide $C_{1/2}$ into a finite collection of intervals $\{I_n\}$ and in each choose a set E satisfying (7–1). Along each interval attach a copy of $\{\frac{1}{4} < |z| < 1\} \setminus E$. This gives R_2.

Lemmas 9.1 and 9.2 imply that harmonic measure of R_2 over $C_{1/4}$ is of the form $\frac{1}{2\pi} g\, d\theta$ where the max of g is bounded by a universal constant times the minimum. Thus there is a constant $c < \min(g)$ and a $\lambda < 1$ such that

$$\int (g - c)\, d\theta \le \lambda \int g\, d\theta.$$

In other words, we can truncate g to be a constant and still retain a fixed fraction of the harmonic measure.

Now do the standard construction of I and J intervals on $C_{\frac{1}{4}+\delta}$, attaching copies of $\big\{\frac{1}{8} < |z| < 1\big\}$ and $\big\{\frac{1}{4} < |z| < 1\big\}$ respectively, so that the new harmonic measure on $C_{1/4}$ is very close to radial (say within ϵ_1 of constant) and has mass at least $(1 - \lambda)$ times the previous mass.

At the next stage we do the construction near both circles $C_{1/4}$ and $C_{1/8}$. At $C_{1/8}$ we repeat the process of the previous paragraph, making the harmonic measure above $C_{1/8}$ as close to radial as we wish, while retaining at least $(1 - \lambda)$ of the total mass, transferring the excess to C_1 and $C_{1/16}$. On $C_{1/4}$ we only make the measure within ϵ_2 of constant (while losing at most ϵ_1 of the mass), the excess being transferred to $C_{1/8}$ and C_1.

We now iterate the process in the obvious way. At stage n we have a surface R_n which only covers the origin once, and such that the harmonic measure is supported on the circles $\{C_{2^{-k}}\}$, with the k-th circle getting mass at most λ^k. Thus the same is true for the limiting measure μ, and hence the harmonic measure of the set $\{z \in R \setminus R_0 : |z| < 2^{-n}\}$ has harmonic measure less than $C\lambda^n$ in R. This proves that $f(z)/z$ is outer.

Acknowledgements

I thank Joseph Cima for helpful conversations about Rudin's problem and composition operators. This paper first appeared as preprint in March 1998 and I am extremely grateful to Paul Bourdon for his carefully reading of that version and the many helpful comments, suggestions and references he provided. Also thanks to Carl Sundberg for discussing his results with me. I am indebted to the referee for a careful and thoughtful reading of the manuscript and numerous suggestions which improved it.

References

[Aleksandrov et al. 1999] A. B. Aleksandrov, J. M. Anderson, and A. Nicolau, "Inner functions, Bloch spaces and symmetric measures", *Proceedings London Math. Soc.* (3) **79**:2 (1999), 318–352. MR 2000g:46029 Zbl 01463593

[Bishop 1993] C. J. Bishop, "An indestructible Blaschke product in the little Bloch space", *Publ. Mat.* **37**:1 (1993), 95–109. MR 94j:30032 Zbl 0810.30024

[Bourdon 1997a] P. S. Bourdon, "Rudin's orthogonality problem and the Nevanlinna counting function", *Proc. Amer. Math. Soc.* **125**:4 (1997), 1187–1192. MR 98b:30034 Zbl 0866.30028

[Bourdon 1997b] P. S. Bourdon, "Rudin's orthogonality problem and the Nevanlinna counting function, II", 1997. Unpublished notes.

[Cantón 1998] A. Cantón, "Singular measures and the little Bloch space", *Publ. Mat.* **42**:1 (1998), 211–222. MR 99g:30046 Zbl 0916.30032

[Cima and Hansen 1990] J. A. Cima and L. J. Hansen, "Space-preserving composition operators", *Michigan Math. J.* **37**:2 (1990), 227–234. MR 91m:47042 Zbl 0715.30028

[Cima et al. 1974/75] J. A. Cima, J. Thomson, and W. Wogen, "On some properties of composition operators", *Indiana Univ. Math. J.* **24** (1974/75), 215–220. MR 50 #2979 Zbl 0276.47038

[Cima et al. 1993] J. A. Cima, B. Korenblum, and M. Stessin, "Composition isometries and Rudin's problems", preprint, 1993.

[Fernández et al. 1996] J. L. Fernández, D. Pestana, and J. M. Rodríguez, "Distortion of boundary sets under inner functions, II", *Pacific J. Math.* **172**:1 (1996), 49–81. MR 97b:30035 Zbl 0847.32005

[Garnett 1981] J. B. Garnett, *Bounded analytic functions*, Pure and Applied Mathematics **96**, Academic Press Inc., New York, 1981. MR 83g:30037 Zbl 0469.30024

[Löwner 1923] K. Löwner, "Untersuchungen über schlichte konforme Abbildungen des Einheitskreises, I", *Math. Ann.* **89** (1923), 103–121. JFM 49.0714.01

[Nordgren 1968] E. A. Nordgren, "Composition operators", *Canad. J. Math.* **20** (1968), 442–449. MR 36 #6961 Zbl 0161.34703

[Rudin 1967] W. Rudin, "A generalization of a theorem of Frostman", *Math. Scand* **21** (1967), 136–143 (1968). MR 38 #3463 Zbl 0185.33301

[Rudin 1973] W. Rudin, *Functional analysis*, McGraw-Hill, New York, 1973. MR 51 #1315 Zbl 0185.33301

[Rudin 1980] W. Rudin, *Function theory in the unit ball of* $\mathbf{C}^n$, Grundlehren der Mathematischen Wissenschaften **241**, Springer, New York, 1980. MR 82i:32002 Zbl 0495.32001

[Ryff 1966] J. V. Ryff, "Subordinate H^p functions", *Duke Math. J.* **33** (1966), 347–354. MR 33 #289 Zbl 0148.30205

[Shapiro 1987] J. H. Shapiro, "The essential norm of a composition operator", *Ann. of Math.* (2) **125**:2 (1987), 375–404. MR 88c:47058 Zbl 0642.47027

[Smith 1996] W. Smith, "Composition operators between Bergman and Hardy spaces", *Trans. Amer. Math. Soc.* **348**:6 (1996), 2331–2348. MR 96i:47056 Zbl 0857.47020

[Smith 1998] W. Smith, "Inner functions in the hyperbolic little Bloch class", *Michigan Math. J.* **45**:1 (1998), 103–114. MR 2000e:30070 Zbl 0976.30018

[Smith and Yang 1998] W. Smith and L. Yang, "Composition operators that improve integrability on weighted Bergman spaces", *Proc. Amer. Math. Soc.* **126**:2 (1998), 411–420. MR 98d:47070 Zbl 0892.47031

[Stephenson 1988] K. Stephenson, "Construction of an inner function in the little Bloch space", *Trans. Amer. Math. Soc.* **308**:2 (1988), 713–720. MR 89k:30031 Zbl 0654.30024

[Sundberg 2003] C. Sundberg, "Measures induced by analytic functions and a problem of Walter Rudin", *J. Amer. Math. Soc.* **16**:1 (2003), 69–90. MR 2003i:30049 Zbl 1012.30022

[Tsuji 1959] M. Tsuji, *Potential theory in modern function theory*, Maruzen Co. Ltd., Tokyo, 1959. MR 22 #5712 Zbl 0322.30001

[Wheeden and Zygmund 1977] R. L. Wheeden and A. Zygmund, *Measure and integral*, Marcel Dekker Inc., New York, 1977. MR 58 #11295 Zbl 0362.26004

[Zorboska 1994] N. Zorboska, "Composition operators with closed range", *Trans. Amer. Math. Soc.* **344**:2 (1994), 791–801. MR 94k:47050 Zbl 0813.47037

Received 1 May 2002. Revised 1 October 2002.

CHRISTOPHER J. BISHOP
MATHEMATICS DEPARTMENT
SUNY AT STONY BROOK
STONY BROOK
NEW YORK 11794-3651
bishop@math.sunysb.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

BASES OF QUANTIZED ENVELOPING ALGEBRAS

BANGMING DENG AND JIE DU

We give a systematic description of many monomial bases for a specified quantized enveloping algebra and of many integral monomial bases for the associated Lusztig $\mathbb{Z}[v, v^{-1}]$-form. The relations among monomial bases, PBW bases and canonical bases are also discussed.

1. Introduction

Let $\mathfrak{g}$ be a (complex) semisimple Lie algebra and let $\boldsymbol{U}^+$ be the positive part of its associated quantized enveloping algebra $\boldsymbol{U} = \boldsymbol{U}_v(\mathfrak{g})$ over $\mathbb{Q}(v)$ with a Drinfeld–Jimbo presentation in the generators $E_i, F_i, K_i^{\pm 1}$ $(i \in I = [1, n])$. We denote by U^+ the Lusztig form of $\boldsymbol{U}^+$, that is, U^+ is generated by all the divided powers $E_i^{(m)}$ over $\mathscr{Z} := \mathbb{Z}[v, v^{-1}]$. Let Ω be the set of words on the alphabet I and, for $w = i_1^{e_1} i_2^{e_2} \cdots i_m^{e_m} \in \Omega$ with $i_{j-1} \neq i_j$ for all j, put $E_w = E_{i_1}^{e_1} \cdots E_{i_m}^{e_m}$ and $\mathfrak{m}^{(w)} = E_{i_1}^{(e_1)} \cdots E_{i_m}^{(e_m)}$. Further, let Λ denote the set of all functions from the set of positive roots of $\mathfrak{g}$ to nonnegative integers.

Certain monomial bases of the form $\mathfrak{m}^{(w)}$ have been introduced for U^+ in [Lusztig 1990, 7.8] and [Ringel 1995, Theorem 1′] for the simply laced case, and in [Chari and Xi 1999] in general, and are used in the elementary construction of canonical bases. In this paper, we present a systematic way to sort out bases from the monomials E_w for $\boldsymbol{U}^+$ and from the monomials $\mathfrak{m}^{(w)}$ for U^+, and relate them to PBW bases and canonical bases. The main result is:

Theorem 1.1. *Assume that $\mathfrak{g}$ is simply laced. There is a partition $\Omega = \bigcup_{\lambda \in \Lambda} \Omega_\lambda$ such that, by choosing an arbitrary word $w_\lambda \in \Omega_\lambda$ for every $\lambda \in \Lambda$, the set $\{E_{w_\lambda}\}_{\lambda \in \Lambda}$ of monomials forms a basis for $\boldsymbol{U}^+$. If all words w_λ are chosen to be distinguished* (see Section 5), *the set $\{\mathfrak{m}^{(w_\lambda)}\}_{\lambda \in \Lambda}$ forms a $\mathscr{Z}$-basis for U^+.*

We shall see from Remarks 6.5 that the monomial bases given in [Lusztig 1990], [Ringel 1995] and [Reineke 2001a, 4.2] can be obtained in this systematic description by a selection of the representatives w_λ. The assumption of simply laced types

MSC2003: 17B37, 16G20.
Keywords: quantized enveloping algebra, Ringel–Hall algebra, generic extension, monomial basis, canonical basis.
Supported partially by the NSF of China (Grant no. 10271014), the TRAPOYP, the Doctoral Program of Higher Education, and the Australian Research Council.

is made so that we may directly use the theory of quiver representations. See also [Deng and Du 2005] for a similar result in the affine $\mathfrak{sl}_n$ case. It is natural to expect that a similar result holds in the nonsimply laced case and to relate this theory to Kashiwara's crystal bases defined by using the monomials in Kashiwara operators [1991].

The main ingredients for the proof are Ringel's Hall algebra theory [Ringel 1995], the monoidal structure [Reineke 2001b] on the set $\mathcal{M}$ of isoclasses of finite-dimensional representations of a Dynkin quiver Q and the Bruhat–Chevalley type partial ordering on orbits in an affine space. These will be discussed separately in Sections 2, 3 and 4. Distinguished words are introduced and investigated in Section 5 and we prove the main result in Section 6. As an application of the theory, we mention an elementary construction [Reineke 2001b, §6] of the canonical bases for U^+ as the counterpart of a similar construction for the Hecke algebra in [Kazhdan and Lusztig 1979]. This construction uses the same order as the one used in the geometric construction, involving perverse sheaf and intersection cohomology theories. Finally, more explicit results on distinguished words are worked out for the case of type A in Section 7.

Throughout, k denotes a finite field unless otherwise specified. Let $q_k = |k|$. All modules are finite-dimensional over k. If M is a module, nM, $n \geqslant 0$, denotes the direct sum of n copies of M. Further, by $[M]$ we denote the class of modules isomorphic to M, i.e., the isoclass of M. For modules $M, N_1, \ldots, N_t$, let $F^M_{N_1 \cdots N_t}$ denote the number of filtrations

$$M = M_0 \supset M_1 \supset \cdots \supset M_{t-1} \supset M_t = 0$$

such that $M_{i-1}/M_i \cong N_i$ for all $1 \leqslant i \leqslant t$.

2. Ringel–Hall algebras of Dynkin quivers

Let $Q = (I, Q_1)$ be a quiver, i.e., a finite directed graph, where $I = Q_0$ is the set of vertices $\{1, 2, \ldots, n\}$ and Q_1 is the set of arrows. If $\rho \in Q_1$ is an arrow from tail i to head j, we write $h(\rho)$ for j and $t(\rho)$ for i. Thus we obtain functions $h, t : Q_1 \to I$. A vertex $i \in I$ is called a sink if there is no arrow ρ with $t(\rho) = i$, and a source if there is no ρ with $h(\rho) = i$.

Let kQ be the path algebra of Q. A (finite-dimensional) *representation* V of Q, consisting of a set of finite-dimensional vector spaces V_i for each $i \in I$ and a set of linear transformations $V_\rho : V_{t(\rho)} \to V_{h(\rho)}$ for each $\rho \in Q_1$, is identified with a (left) kQ-module. We call $\mathbf{dim}\, V := (\dim V_1, \ldots, \dim V_n)$ the *dimension vector* of V and $\ell(V) := \sum_{i=1}^n \dim V_i$ the *length* of V (also called the dimension of V). If

Q contains no oriented cycles, there are exactly n pairwise nonisomorphic simple kQ-modules $S_1, \dots, S_n$ corresponding bijectively to the vertices of Q.

From now on, we assume that Q is a Dynkin quiver, that is, a quiver whose underlying graph is a (simply laced) Dynkin graph. By Gabriel's theorem [1972], there is a bijection between the set of isoclasses of indecomposable kQ-modules and a positive system Φ^+ of the root system Φ associated with Q. For any $\beta \in \Phi^+$, let $M(\beta) = M_k(\beta)$ denote the corresponding indecomposable kQ-module. By the Krull–Remak–Schmidt theorem, every kQ-module M is isomorphic to

$$M(\lambda) = M_k(\lambda) := \bigoplus_{\beta \in \Phi^+} \lambda(\beta) M_k(\beta),$$

for some function $\lambda : \Phi^+ \to \mathbb{N}$. Thus the isoclasses of kQ-modules are indexed by the set

$$\Lambda = \{\lambda : \Phi^+ \to \mathbb{N}\} \cong \mathbb{N}^{|\Phi^+|}.$$

By a result of Ringel [1990], for $\lambda, \mu_1, \dots, \mu_m$ in Λ, there is a polynomial $\varphi^{\lambda}_{\mu_1 \cdots \mu_m}(q) \in \mathbb{Z}[q]$, called a Hall polynomial, such that for any finite field k of q_k elements

$$\varphi^{\lambda}_{\mu_1 \cdots \mu_m}(q_k) = F^{M_k(\lambda)}_{M_k(\mu_1) \cdots M_k(\mu_m)}.$$

Let $\mathcal{A} = \mathbb{Z}[q]$ be the integral polynomial ring in the indeterminate q. The generic (untwisted) Ringel–Hall algebra $\mathcal{H} = \mathcal{H}_q(Q)$ of Q over $\mathcal{A}$ is by definition the free $\mathcal{A}$-module having basis $\{u_\lambda \mid \lambda \in \Lambda\}$, and satisfying the multiplicative relations

$$u_\mu u_\nu = \sum_{\lambda \in \Lambda} \varphi^{\lambda}_{\mu\nu}(q) u_\lambda.$$

We sometimes write $u_\lambda = u_{[M(\lambda)]}$ in order to make certain calculations in term of modules. For $i \in I$, we set $u_i = u_{[S_i]}$. Clearly, $\mathcal{H}$ admits a natural $\mathbb{N}^n$-grading by dimension vectors.

Following [Ringel 1993b], we can twist the multiplication of the Ringel–Hall algebra to obtain the positive part $\boldsymbol{U}^+$ of a quantized enveloping algebra.

Let $\mathcal{Z} = \mathbb{Z}[v, v^{-1}]$, where v is an indeterminate with $v^2 = q$. The *twisted* Ringel–Hall algebra $\mathcal{H}^\star = \mathcal{H}^\star_v(Q)$ of Q is by definition the free $\mathcal{Z}$-module having basis $\{u_\lambda = u_{[M(\lambda)]} \mid \lambda \in \Lambda\}$ and satisfying the multiplication rules

$$u_\mu \star u_\nu = v^{\langle \mu, \nu \rangle} u_\mu u_\nu = v^{\langle \mu, \nu \rangle} \sum_{\lambda \in \Lambda} \varphi^{\lambda}_{\mu\nu}(v^2) u_\lambda,$$

where $\langle \mu, \nu \rangle = \dim_k \mathrm{Hom}_{kQ}(M(\mu), N(\nu)) - \dim_k \mathrm{Ext}^1_{kQ}(M(\mu), N(\nu))$ is the Euler form associated with the quiver Q. Note that, if we define the bilinear form $\langle -, - \rangle :$ $\mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{Z}$ by

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_{i \in I} a_i b_i - \sum_{\rho \in Q_1} a_{t(\rho)} b_{h(\rho)},$$

where $\boldsymbol{a} = (a_1, \cdots, a_n)$, $\boldsymbol{b} = (b_1, \ldots, b_n)$, then

$$\langle \mu, \nu \rangle = \big\langle \mathbf{dim}\, M(\mu), \mathbf{dim}\, M(\nu) \big\rangle.$$

For each $m \geqslant 1$, set $[m] = (v^m - v^{-m})/(v - v^{-1})$ and $[m]^! = [1][2] \cdots [m]$. We define, for each $i \in I$, the divided powers

$$u_i^{(\star m)} = \frac{u_i^{\star m}}{[m]^!} \quad \text{and} \quad E_i^{(m)} = \frac{E_i^m}{[m]^!},$$

in $\mathcal{H}^\star$ and $\boldsymbol{U}^+$, respectively. Here $u_i^{\star m} = \underbrace{u_i \star \cdots \star u_i}_{m} = v^{m(m-1)/2} u_i^m$.

Proposition 2.1 [Ringel 1995, §7]. *The algebra $\mathcal{H}^\star$ is generated by all $u_i^{(\star m)}$, for $i \in I$, $m \geqslant 1$. There is a natural isomorphism*

$$\Psi : U^+ \xrightarrow{\sim} \mathcal{H}^\star, \ E_i^{(m)} \mapsto u_i^{(\star m)} \quad (i \in I, m \geqslant 1).$$

We shall identify U^+ with $\mathcal{H}^\star$ under this isomorphism.

3. Generic extensions and the monoid $\mathcal{M}$

In this section, we collect some recent results on generic extensions for quiver representations over an *algebraically closed* field k.

Fix $\boldsymbol{d} = (d_i)_i \in \mathbb{N}^n$ and define the affine space

$$R(\boldsymbol{d}) = R(Q, \boldsymbol{d}) := \prod_{\alpha \in Q_1} \mathrm{Hom}_k(k^{d_{t(\alpha)}}, k^{d_{h(\alpha)}}) \cong \prod_{\alpha \in Q_1} k^{d_{h(\alpha)} \times d_{t(\alpha)}}.$$

Thus a point $x = (x_\alpha)_\alpha$ of $R(\boldsymbol{d})$ determines a representation $V(x)$ of Q. The algebraic group $\mathrm{GL}(\boldsymbol{d}) = \prod_{i=1}^n \mathrm{GL}_{d_i}(k)$ acts on $R(\boldsymbol{d})$ by conjugation:

$$(g_i)_i \cdot (x_\alpha)_\alpha = (g_{h(\alpha)} x_\alpha g_{t(\alpha)}^{-1})_\alpha.$$

The $\mathrm{GL}(\boldsymbol{d})$-orbits $\mathcal{O}_x$ in $R(\boldsymbol{d})$ correspond bijectively to the isoclasses $[V(x)]$ of representations of Q with dimension vector $\boldsymbol{d}$.

The stabilizer $\mathrm{GL}(\boldsymbol{d})_x = \{g \in \mathrm{GL}(\boldsymbol{d}) \mid gx = x\}$ of x is the group of automorphisms of $M := V(x)$ which is Zariski-open in $\mathrm{End}_{kQ}(M)$ and has dimension equal to $\dim \mathrm{End}_{kQ}(M)$. It follows that the orbit $\mathcal{O}_M := \mathcal{O}_x$ of M has dimension

$$\dim \mathcal{O}_M = \dim \mathrm{GL}(\boldsymbol{d}) - \dim \mathrm{End}_{kQ}(M).$$

Lemma 3.1 [Reineke 2001b]. *Let Q be a Dynkin quiver. For $x \in R(\boldsymbol{d}_1)$ and $y \in R(\boldsymbol{d}_2)$, let $\mathcal{E}(\mathcal{O}_x, \mathcal{O}_y)$ be the set of all $z \in R(\boldsymbol{d})$ where $\boldsymbol{d} = \boldsymbol{d}_1 + \boldsymbol{d}_2$ such that $V(z)$ is an extension of some $M \in \mathcal{O}_x$ by some $N \in \mathcal{O}_y$. Then $\mathcal{E}(\mathcal{O}_x, \mathcal{O}_y)$ is irreducible.*

Given representations M, N of Q, consider the extensions

$$0 \to N \to E \to M \to 0$$

of M by N. By the lemma, there is a unique (up to isomorphism) such extension G with $\dim \mathbb{O}_G$ maximal (i.e., with $\dim \mathrm{End}_{kQ}(G)$ minimal). We call G the *generic extension* of M by N, denoted by $M * N$.

For two representations M, N, we say that M degenerates to N, or that N is a degeneration of M, and write $[N] \leqslant [M]$ (or simply $N \leqslant M$), if $\mathbb{O}_N \subseteq \overline{\mathbb{O}}_M$, the closure of $\mathbb{O}_M$. Note that $N < M \iff \mathbb{O}_N \subseteq \overline{\mathbb{O}}_M \backslash \mathbb{O}_M$.

Remark 3.2. The relation $\leqslant$ on the isoclasses is independent of the field k. This is seen from the following equivalence proved in [Bongartz 1996, Proposition 3.2]:

$$N \leqslant M \iff \dim \mathrm{Hom}(X, N) \geqslant \dim \mathrm{Hom}(X, M) \text{ for all } X \tag{3–1}$$

and the fact that the dimension $\dim \mathrm{Hom}(X, Y)$ is the same over any field. Thus we may simply define a (characteristic-free) partial order on Λ by

$$\lambda \leqslant \mu \iff M_k(\lambda) \leqslant M_k(\mu).$$

for any given (algebraically closed) field k.

The first part of the following result is well-known (see, for example, [Bongartz 1996, 1.1]) and the other parts are proved in [Reineke 2001b].

Theorem 3.3. (1) *If* $0 \to N \to E \to M \to 0$ *is exact and nonsplit, then* $M \oplus N < E$.

(2) *Let* M, N, X *be representations of* Q. *Then* $X \leqslant M * N$ *if and only if there exist* $M' \leqslant M$, $N' \leqslant N$ *such that* X *is an extension of* M' *by* N'. *In particular,* $M' \leqslant M,\ N' \leqslant N \implies M' * N' \leqslant M * N$.

(3) *Let* $\mathcal{M}$ *be the set of isoclasses of* kQ*-modules and define a multiplication* $*$ *on* $\mathcal{M}$ *by* $[M] * [N] = [M * N]$ *for any* $[M], [N] \in \mathcal{M}$. *Then* $\mathcal{M}$ *is a monoid with identity* $1 = [0]$ *and the multiplication* $*$ *preserves the induced partial ordering on* $\mathcal{M}$.

(4) $\mathcal{M}$ *is generated by the simple modules* $[S_i]$, $i \in I$.

Let Ω be the set of words in the alphabet $I = \{1, \ldots, n\}$. For $w = i_1 i_2 \cdots i_m \in \Omega$, let $\wp(w) \in \Lambda$ be the element defined by

$$[S_{i_1}] * \cdots * [S_{i_m}] = [M(\wp(w))]. \tag{3–2}$$

Thus we obtain a map $\wp : \Omega \to \Lambda$. The theorem shows that $\wp$ is surjective and induces a partition $\Omega = \bigcup_{\lambda \in \Lambda} \Omega_\lambda$ with $\Omega_\lambda = \wp^{-1}(\lambda)$. Each Ω_λ is called a *fibre* of $\wp$.

By Remark 3.2, if we set $\lambda * \mu := M(\lambda * \mu) \cong M(\lambda) * M(\mu)$ for $\lambda, \mu \in \Lambda$, the element $\lambda * \mu$ is well-defined, independent of the field k. Note that the multiplication $*$ on Λ depends on the orientation of Q.

4. The poset Λ

In this section we look at some properties of the poset $(\Lambda, \leqslant)$, where $\leqslant$ is defined in Remark 3.2.

For $w = i_1 i_2 \cdots i_m \in \Omega$ and $\lambda \in \Lambda$, let φ_w^λ denote the Hall polynomial $\varphi_{\mu_1\cdots\mu_m}^\lambda$, where $M(\mu_r) \cong S_{i_r}$. Thus, for a finite field k,

$$\varphi_w^\lambda(q_k) = F_{S_{i_1 k}\cdots S_{i_m k}}^{M_k(\lambda)}$$

is the number of composition series of $M_k(\lambda)$:

$$M_k(\lambda) = M_0 \supset M_1 \supset \cdots \supset M_{m-1} \supset M_m = 0$$

with $M_{j-1}/M_j \cong S_{i_j k}$. Such a composition series is called a composition series of *type* w.

The following lemma is a bit stronger than [Deng and Du 2005, 6.2].

Lemma 4.1. *Let $w \in \Omega$ and $\mu \geqslant \lambda$ in Λ. Then $\varphi_w^\mu \neq 0$ implies $\varphi_w^\lambda \neq 0$.*

Proof. Let $w = i_1 i_2 \cdots i_m$ and $w' = i_2 \cdots i_m$. We apply induction on m. If $m = 1$ then $\mu \geqslant \lambda$ forces $M(\mu) = M(\lambda)$ and the result is clear. Now assume $m > 1$. If $\varphi_w^\mu \neq 0$, then $\varphi_w^\mu(q_k) \neq 0$ for some finite field k. Thus $M_k(\mu)$ has a submodule $M_k' \cong M_k(\mu')$ having a composition series of type w'. Hence $\varphi_{w'}^{\mu'} \neq 0$, since $\varphi_{w'}^{\mu'}(q_k) \neq 0$. Base change to the algebraic closure $\bar{k}$ of k gives an exact sequence over $\bar{k}$

$$0 \longrightarrow M' \longrightarrow M(\mu) \longrightarrow S_{i_1} \longrightarrow 0,$$

where we have dropped the subscripts $\bar{k}$. Thus

$$M(\lambda) \leqslant M(\mu) \leqslant S_{i_1} * M'.$$

By Theorem 3.3(2), there exist modules N', N'' such that $M(\lambda)$ is an extension of N' by N'' and $N' \leqslant M'$, $N'' \leqslant S_{i_1}$. So we obtain an exact sequence (over $\bar{k}$)

$$0 \longrightarrow N' \xrightarrow{f} M(\lambda) \xrightarrow{g} N'' \longrightarrow 0.$$

Now the condition $N' \leqslant M'$ means $\lambda' \leqslant \mu'$ where $N' \cong M(\lambda')$. Since $\varphi_{w'}^{\mu'} \neq 0$, it follows from induction that $\varphi_{w'}^{\lambda'} \neq 0$, that is, N' has a composition series of type w'. On the other hand, since S_{i_1} is simple, $N'' \leqslant S_{i_1}$ implies $N'' \cong S_{i_1}$. Therefore, $M(\lambda)$ has a composition series of type w, and consequently, $\varphi_w^\lambda \neq 0$. $\square$

We now relate the partial order $\leqslant$ to certain nonzero Hall polynomials.

Theorem 4.2. *Let $\lambda, \mu \in \Lambda$. Then $\lambda \leqslant \mu$ if and only if there exists a word $w \in \wp^{-1}(\mu)$ with $\varphi_w^\lambda \neq 0$.*

Proof. Suppose $\lambda \leqslant \mu$. Since $\wp$ is surjective, $\mu = \wp(w)$ for some $w \in \Omega$. By (3–2), we see that $\varphi_w^{\wp(w)} \neq 0$. Thus Lemma 4.1 implies $\varphi_w^\lambda \neq 0$, as required.

Conversely, let $w = i_1 i_2 \cdots i_m \in \Omega$, $\lambda \in \Lambda$, and suppose $\varphi_w^\lambda \neq 0$. We use induction on m to prove that $\lambda \leqslant \wp(w)$. If $m = 1$, there is nothing to prove. Let $m > 1$ and $w' = i_2 \cdots i_m$ and assume $\lambda' \leqslant \wp(w')$ whenever $\varphi_{w'}^{\lambda'} \neq 0$. Since $\varphi_w^\lambda \neq 0$, there is a finite field k (of any given characteristic) such that $\varphi_w^\lambda(q_k) \neq 0$. Thus there is a submodule M_k' of $M_k(\lambda)$ having a composition series of type w'. This implies $\varphi_{w'}^{\lambda'} \neq 0$ where $M_k(\lambda') \cong M_k'$. By induction, we have $\lambda' \leqslant \wp(w')$.

On the other hand, base change to the exact sequence

$$0 \longrightarrow M_k' \longrightarrow M_k(\lambda) \longrightarrow S_{i_1 k} \longrightarrow 0$$

yields an exact sequence over $\bar{k}$

$$0 \longrightarrow M' \longrightarrow M(\lambda) \longrightarrow S_{i_1} \longrightarrow 0.$$

(Here again we dropped the subscripts $\bar{k}$.) By Theorem 3.3(2) we obtain

$$M(\lambda) \leqslant S_{i_1} * M(\lambda') \leqslant S_{i_1} * M(\wp(w')) = M(\wp(w)).$$

Therefore, $\lambda \leqslant \wp(w)$. □

5. Distinguished words

Let $w = i_1 i_2 \cdots i_m$ be a word in Ω. Then w can be uniquely expressed in the *tight form* $w = j_1^{e_1} j_2^{e_2} \cdots j_t^{e_t}$, where $e_r \geqslant 1$, $1 \leqslant r \leqslant t$, and $j_r \neq j_{r+1}$ for $1 \leqslant r \leqslant t-1$. Following [Ringel 1993a, 2.3], a filtration

$$M = M_0 \supset M_1 \supset \cdots \supset M_{t-1} \supset M_t = 0$$

of a module is called a *reduced* filtration of type w if $M_{r-1}/M_r \cong e_r S_{j_r}$ for all $1 \leqslant r \leqslant t$. Any reduced filtration of M of type w can be refined to a composition series of M of type w. Conversely, given a composition series of M of type w, there is a unique reduced filtration of M of type w such that the given composition series is a refinement of this reduced filtration. By $\gamma_w^\lambda(q)$ we denote the Hall polynomial $\varphi_{\mu_1 \cdots \mu_t}^\lambda(q)$, where $M(\mu_r) = e_r S_{j_r}$. Thus, for a finite field k of q_k elements, $\gamma_w^\lambda(q_k)$ is the number of the reduced filtrations of $M_k(\lambda)$ of type w. A word w is called *distinguished* if $\gamma_w^{\wp(w)}(q) = 1$; this is the case if and only if, for some algebraically closed field k, $M_k(\wp(w))$ has a unique reduced filtration of type w. See [Deng and Du 2005, §5].

Example 5.1. Let $w = j_1{}^{e_1} j_2{}^{e_2} \cdots j_t{}^{e_t}$ be in the tight form. If $j_1, \ldots, j_t$ are pairwise distinct and satisfy

$$\mathrm{Ext}^1_{kQ}(S_{j_r}, S_{j_s}) \neq 0 \implies r < s,$$

then $F^M_{N_1 \cdots N_t} = 0$ or 1 for every kQ-module M, where $N_r = e_r S_{j_r}$. Thus w is distinguished.

Distinguished words will be used in the construction of integral monomial bases for the Lusztig form. The following lemma shows that these words are somehow evenly distributed.

Lemma 5.2. *Each fibre of $\wp$ contains at least one distinguished word.*

Proof. This follows directly from [Reineke 2001a, Lemma 4.5]. For completeness, we present here the construction of such distinguished words.

By $\mathcal{I}$ we denote the set of the isoclasses of indecomposable representations of Q. Let $\mathcal{I}_*$ be a *directed* partition of $\mathcal{I}$ [Reineke 2001a, §4], that is, a partition of the set $\mathcal{I}$ into subsets $\mathcal{I}_1, \dots, \mathcal{I}_m$ such that

(a) $\mathrm{Ext}^1_{kQ}(M, N) = 0$ for all M, N in the same part $\mathcal{I}_r$,

(b) $\mathrm{Ext}^1_{kQ}(M, N) = 0 = \mathrm{Hom}_{kQ}(N, M)$ if $M \in \mathcal{I}_r$, $N \in \mathcal{I}_s$, where $1 \leqslant r < s \leqslant m$.

Then, for each $\lambda \in \Lambda$, we have a unique decomposition

$$M(\lambda) = \bigoplus_{r=1}^{m} M_r,$$

where all the summands of M_r belong to $\mathcal{I}_r$, $1 \leqslant r \leqslant m$. Thus

$$\mathrm{Hom}_{kQ}(M_r, M_s) \neq 0 \implies r \leqslant s. \tag{5–1}$$

Further, since Q is a Dynkin quiver, we can order the vertices of Q in a sequence $i_1, i_2, \dots, i_n$ such that, for each $1 < j \leqslant n$, i_j is a sink in the full subquiver of Q with vertices $\{i_1, \dots, i_{j-1}, i_j\}$. Equivalently, $i_1, i_2, \dots, i_n$ are ordered to satisfy

$$\mathrm{Ext}^1_{kQ}(S_{i_j}, S_{i_l}) \neq 0 \implies j < l. \tag{5–2}$$

Let $\boldsymbol{d}^{(r)} = (d_1^{(r)}, \dots, d_n^{(r)}) = \mathbf{dim}\, M_r$, for $1 \leqslant r \leqslant m$, and set

$$w_r = \underbrace{i_1 \cdots i_1}_{d^{(r)}_{i_1}} \cdots\cdots \underbrace{i_n \cdots i_n}_{d^{(r)}_{i_n}}$$

and $w_\lambda = w_1 \cdots w_m \in \Omega$. Then [Reineke 2001a, Lemma 4.5] implies that $\wp(w_\lambda) = \lambda$ and $\gamma^\lambda_{w_\lambda}(q) = 1$, that is, w_λ is distinguished. □

We call the distinguished words constructed above *directed* distinguished words (with respect to the given directed partition $\mathcal{I}_*$).

We mention a special case of directed partitions $\mathcal{I}_*$ where each part $\mathcal{I}_r$ contains only one isoclass. This case is equivalent to ordering the indecomposable modules $M(\beta_1), M(\beta_2), \dots$ such that

$$\mathrm{Hom}_{kQ}(M(\beta_r), M(\beta_s)) \neq 0 \implies r \leqslant s. \tag{5–3}$$

Note that monomial bases associated to these special directed distinguished words have been constructed in [Lusztig 1990] and [Ringel 1995]; see Remarks 6.5 below.

The following example shows that a fibre of $\wp$ could contain many words other than directed distinguished ones.

Example 5.3. Let Q denote the quiver

$$\begin{tikzcd} 1 \arrow[dr] & 2 \arrow[d] & 3 \arrow[dl] \\ & 4 & \end{tikzcd}$$

Let $\lambda \in \Lambda$ be such that $M(\lambda)$ is the indecomposable kQ-module with dimension vector $(1, 1, 1, 2)$. Then $\wp^{-1}(\lambda)$ contains 12 words

$$1234^2,\ 1324^2,\ 2134^2,\ 2314^2,\ 3124^2,\ 3214^2,$$
$$12434,\ 13424,\ 21434,\ 23414,\ 31424,\ 32414$$

all distinguished. From the structure of the Auslander–Reiten quiver of kQ, one sees easily that the first 6 words are directed distinguished, but the last 6 are not.

6. Monomial and integral monomial bases

For $m \geqslant 1$, let $[\![m]\!]^! = [\![1]\!][\![2]\!] \cdots [\![m]\!]$, where $[\![e]\!] = (q^e - 1)/(q-1)$. Then $[\![m]\!] = v^{m-1}[m]$ and $[\![m]\!]^! = v^{m(m-1)/2}[m]^!$.

Lemma 6.1. *Let $w \in \Omega$ be a word with the tight form $j_1^{e_1} j_2^{e_2} \cdots j_t^{e_t}$. Then, for each $\lambda \in \Lambda$,*

$$\varphi_w^\lambda(q) = \gamma_w^\lambda(q) \prod_{r=1}^{t} [\![e_r]\!]^!.$$

In particular, $\varphi_w^{\wp(w)}(q) = \prod_{r=1}^t [\![e_r]\!]^!$ if w is distinguished.

Proof. The result follows from the definition of a distinguished word and the fact that the number of composition series of eS_i is $[\![e]\!]^!$ (see [Ringel 1993b, 8.2]). $\square$

To each word $w = i_1 i_2 \cdots i_m \in \Omega$, we associate a monomial

$$u_w = u_{i_1} u_{i_2} \cdots u_{i_m} \in \mathcal{H}.$$

Theorem 4.2 and Lemma 6.1 give:

Proposition 6.2. *For each $w \in \Omega$ with the tight form ${j_1}^{e_1} {j_2}^{e_2} \cdots {j_t}^{e_t}$, we have*

$$u_w = \sum_{\lambda \leqslant \wp(w)} \varphi_w^\lambda(q) u_\lambda = \prod_{r=1}^{t} [\![e_r]\!]^! \sum_{\lambda \leqslant \wp(w)} \gamma_w^\lambda(q) u_\lambda. \tag{6–1}$$

Moreover, the coefficients appearing in the sum are all nonzero.

This improves [Ringel 1995, Theorem 1, p. 96] in two ways: it generalizes the formula from certain directed distinguished words to all words, and it replaces the lexicographical order by the Bruhat type partial order $\leqslant$.

For any commutative ring $\mathscr{A}'$ which is an $\mathscr{A}$-algebra and any $\mathscr{A}$-module M, let $M_{\mathscr{A}'} = \mathscr{A}' \otimes_{\mathscr{A}} M$ denote the $\mathscr{A}'$-module obtained from M by base change to $\mathscr{A}'$.

Theorem 6.3. *For every $\lambda \in \Lambda$, choose an arbitrary word $w_\lambda \in \wp^{-1}(\lambda)$. The set $\{u_{w_\lambda} \mid \lambda \in \Lambda\}$ is a $\mathbb{Q}(q)$-basis of $\mathscr{H}_{\mathbb{Q}(q)}$. If all the w_λ are chosen to be distinguished, then this set is an $\mathscr{A}_{(q-1)}$-basis of $\mathscr{H}_{\mathscr{A}_{(q-1)}}$ where $\mathscr{A}_{(q-1)}$ denotes the localization of $\mathscr{A}$ at the maximal ideal generated by $q-1$.*

Proof. This follows from Proposition 6.2 and the fact that $\varphi_{w_\lambda}^{\wp(w_\lambda)}$ is invertible in $\mathscr{A}_{(q-1)}$ if w_λ is distinguished. □

Let $\mathfrak{g} = \mathfrak{n}_- \oplus \mathfrak{h} \oplus \mathfrak{n}_+$ be the Lie algebra over $\mathbb{Q}$ of type Q with generators e_i, f_i, h_i. Let $\mathfrak{U}(\mathfrak{g})$ be the universal enveloping algebra of $\mathfrak{g}$. Define monomials e_w similarly for $w \in \Omega$ in $\mathfrak{U}(\mathfrak{n}_+)$.

Corollary 6.4. *For every $\lambda \in \Lambda$, choose an arbitrary distinguished word $w_\lambda \in \wp^{-1}(\lambda)$. The set $\{e_{w_\lambda} \mid \lambda \in \Lambda\}$ is a $\mathbb{Q}$-basis of $\mathfrak{U}(\mathfrak{n}_+)$.*

Proof. The result follows from the isomorphism $\mathscr{H}_{\mathscr{A}'}/(q-1)\mathscr{H}_{\mathscr{A}'} \cong \mathfrak{U}(\mathfrak{n}_+)$, where $\mathscr{A}' = \mathscr{A}_{(q-1)}$, and Theorem 6.3. □

Proof of Theorem 1.1. For each $w = i_1 i_2 \cdots i_m \in \Omega$ we have

$$u_{i_1} \star \cdots \star u_{i_m} = v^{\varepsilon(w)} u_w,$$

where

$$\varepsilon(w) = \sum_{1 \leqslant r < s \leqslant m} \langle \mathbf{dim}\, S_{i_r}, \mathbf{dim}\, S_{i_s} \rangle.$$

Let, for $w = j_1^{e_1} \cdots j_t^{e_t}$ in tight form,

$$\mathfrak{m}^{(w)} := E_{j_1}^{(e_1)} \cdots E_{j_t}^{(e_t)} = \left(\prod_{r=1}^{t} [e_r]^! \right)^{-1} u_{j_1}^{\star e_1} \star \cdots \star u_{j_t}^{\star e_t}.$$

Since $\prod_{r=1}^{t} [e_r]^! = v^{-\delta(w)} \prod_{r=1}^{t} [\![e_r]\!]^!$, where $\delta(w) = \sum_{r=1}^{t} e_r(e_r - 1)/2$, it follows from Proposition 6.2 that

$$\mathfrak{m}^{(w)} = \left(\prod_{r=1}^{t} [\![e_r]\!]^! \right)^{-1} v^{\delta(w)+\varepsilon(w)} u_w = v^{\delta(w)+\varepsilon(w)} \sum_{\lambda \leqslant \wp(w)} \gamma_w^\lambda(v^2) u_\lambda. \tag{6–2}$$

Together with Proposition 2.1 and Theorem 6.3, this implies Theorem 1.1 with $\Omega_\lambda = \wp^{-1}(\lambda)$ for all $\lambda \in \Lambda$. □

Remarks 6.5. (a) It is clear, from the definition, that the monomial basis $\{E^{(M)}\}$ constructed in [Reineke 2001a, Theorem 4.2] involves only directed distinguished words w_λ.

(b) As a special case of [Reineke 2001a, Theorem 4.2], the monomial bases constructed in [Lusztig 1990, 7.8; Ringel 1995, pp. 101–2] involve only those directed distinguished words defined with respect to the special directed partition $\mathcal{I}_*$ satisfying conditions (5–3) and (5–2); see [Ringel 1995, Theorem 1] and [Lusztig 1990, 4.12(c), 4.13].[1]

We now look briefly at the elementary and algebraic construction of the canonical basis for U^+ [Reineke 2001b, §6]. Note that the elementary constructions given in, e.g., [Lusztig 1990; Kashiwara 1991; Ringel 1995; Chari and Xi 1999] used a finer order than the one used in the geometric construction. We now use the same order which has an algebraic interpretation (3–1).

For each $\lambda \in \Lambda$, set

$$\tilde{u}_\lambda = v^{-\dim M(\lambda)+\dim \operatorname{End}(M(\lambda))} u_\lambda.$$

Then, by Proposition 2.1, U^+ is $\mathcal{Z}$-free with basis $\mathcal{E} = \{\tilde{u}_\lambda : \lambda \in \Lambda\}$. Note that $U^+ = \bigoplus_{\boldsymbol{d}} U^+_{\boldsymbol{d}}$ is $\mathbb{N}I$-graded according to the dimension vectors, and each $U^+_{\boldsymbol{d}}$ is $\mathcal{Z}$-free with basis $\mathcal{E} \cap U^+_{\boldsymbol{d}} = \{\tilde{u}_\lambda : \lambda \in \Lambda_{\boldsymbol{d}}\}$. Clearly, each $\Lambda_{\boldsymbol{d}}$ together with $\leqslant$ is a poset.

Define a ring homomorphism $\iota : U^+ \to U^+$ by setting $\iota(E_i^{(m)}) = E_i^{(m)}$ and $\iota(v) = v^{-1}$. Clearly, ι preserves the grading of U^+. Write, for any $\tilde{u}_\lambda \in U^+_{\boldsymbol{d}}$,

$$\iota(\tilde{u}_\mu) = \sum_\lambda r_{\lambda,\mu} \tilde{u}_\lambda. \tag{6–3}$$

By [Lusztig 1990, 9.10] (see [Du 1994] for more details), the existence of the canonical bases for $U^+_{\boldsymbol{d}}$ follows from the property

$$r_{\lambda,\lambda} = 1,\ r_{\lambda,\mu} = 0 \quad \text{unless } \lambda \leqslant \mu. \tag{6–4}$$

of the coefficients $r_{\lambda,\mu}$. We use (6–2) to derive (6–4). We first calculate $\delta(w)+\varepsilon(w)$ for directed distinguished words; compare [Ringel 1995, Lemma, p. 102].

Lemma 6.6. *We have for any directed distinguished word $w \in \Omega$*

$$\delta(w) + \varepsilon(w) = -\dim M(\wp(w)) + \dim \operatorname{End} M(\wp(w)).$$

Proof. Let $w \in \Omega$ be a directed distinguished word. Then, by definition, there is a directed partition $\mathcal{I}_*$ of $\mathcal{I}$ and a $\lambda \in \Lambda$ such that w has the form $w = w_\lambda = w_1 \cdots w_m$

[1] It seems to us that condition (5–2) was implicitly used in [Lusztig 1990, 7.2], though it was not explicitly stated in the paper.

with

$$w_r = \underbrace{i_1 \cdots i_1}_{d_{i_1}^{(r)}} \cdots\cdots \underbrace{i_n \cdots i_n}_{d_{i_n}^{(r)}},$$

where $M(\lambda) = M_1 \oplus M_2 \oplus \cdots \oplus M_m$, $\boldsymbol{d}^{(r)} = (d_1^{(r)}, \ldots, d_n^{(r)}) = \mathbf{dim}\, M_r$ for $1 \leqslant r \leqslant m$, and the sequence $i_1, i_2, \ldots, i_n$ of vertices are ordered to satisfy (5–2). Clearly,

$$\delta(w) = \sum_{r=1}^{m} \sum_{j=1}^{n} \frac{d_{i_j}^{(r)}\big(d_{i_j}^{(r)} - 1\big)}{2}.$$

Since $\langle \mathbf{dim}\, S_{i_j}, \mathbf{dim}\, S_{i_l} \rangle = 0$ for $j > l$ and $\mathrm{Ext}^1(M_r, M_s) = 0$ for all $1 \leqslant r \leqslant s \leqslant m$, we obtain, for each $1 \leqslant r \leqslant m$,

$$\begin{aligned}
\varepsilon(w_r) &= \sum_{j=1}^{n} \frac{d_{i_j}^{(r)}(d_{i_j}^{(r)} - 1)}{2} \langle \mathbf{dim}\, S_{i_j}, \mathbf{dim}\, S_{i_j} \rangle + \sum_{1 \leqslant j < l \leqslant n} \big\langle \mathbf{dim}\, d_{i_j}^{(r)} S_{i_j}, \mathbf{dim}\, d_{i_l}^{(r)} S_{i_l} \big\rangle \\
&= \langle \mathbf{dim}\, M_r, \mathbf{dim}\, M_r \rangle - \sum_{j=1}^{n} \frac{\big(d_{i_j}^{(r)}\big)^2}{2} - \sum_{j=1}^{n} \frac{d_{i_j}^{(r)}}{2} \\
&= \dim \mathrm{End}(M_r) - \sum_{j=1}^{n} \frac{d_{i_j}^{(r)}\big(d_{i_j}^{(r)} + 1\big)}{2}
\end{aligned}$$

and therefore,

$$\begin{aligned}
\varepsilon(w) &= \sum_{r=1}^{m} \varepsilon(w_r) + \sum_{1 \leqslant r < s \leqslant m} \langle \mathbf{dim}\, M_r, \mathbf{dim}\, M_s \rangle \\
&= \sum_{r=1}^{m} \varepsilon(w_r) + \sum_{1 \leqslant r < s \leqslant m} \dim \mathrm{Hom}(M_r, M_s).
\end{aligned}$$

Noting that $\mathrm{Hom}(M_r, M_s) = 0$ for $r > s$, we finally obtain

$$\begin{aligned}
\delta(w) + \varepsilon(w) &= \sum_{r=1}^{m} \dim \mathrm{End}(M_r) + \sum_{1 \leqslant r < s \leqslant m} \dim \mathrm{Hom}(M_r, M_s) - \sum_{r=1}^{m} \sum_{j=1}^{n} d_{i_j}^{(r)} \\
&= \dim \mathrm{End}(M(\lambda)) - \dim M(\lambda).
\end{aligned}$$

This completes the proof. □

By Lemma 6.6 and (6–2), any directed distinguished word w satisfies

(6–5) $$\mathfrak{m}^{(w)} = \tilde{\mathfrak{u}}_{\wp(w)} + \sum_{\lambda < \wp(w)} f_{\lambda, \wp(w)} \tilde{\mathfrak{u}}_{\lambda},$$

where $0 \neq f_{\lambda,\wp(w)} \in \mathscr{Z}$. If we fix a representative set $\Lambda' = \{w_\lambda : \lambda \in \Lambda\}$, where $w_\lambda \in \Omega_\lambda$, consisting of directed distinguished words, the relation above implies that, for any $\mu \in \Lambda$,

$$\tilde{\mathfrak{u}}_\mu \in \mathfrak{m}^{(w_\mu)} + \sum_{\lambda<\mu} \mathscr{Z}\mathfrak{m}^{(w_\lambda)}.$$

Restricting to $\Lambda_{\boldsymbol{d}}$, where $\boldsymbol{d}$ is a fixed dimension vector, we obtain the transition matrix $(f_{\lambda,\mu})_{\lambda,\mu\in\Lambda_{\boldsymbol{d}}}$. This matrix has an inverse $(g_{\lambda,\mu})_{\lambda,\mu\in\Lambda_{\boldsymbol{d}}}$ satisfying $g_{\lambda,\lambda} = 1$ and $g_{\lambda,\mu} = 0$ unless $\lambda \leqslant \mu$. Thus

$$\tilde{\mathfrak{u}}_\mu = \mathfrak{m}^{(w_\mu)} + \sum_{\lambda<\mu} g_{\lambda,\mu}\mathfrak{m}^{(w_\lambda)}.$$

Applying ι, we obtain by (6–5)

$$\iota(\tilde{\mathfrak{u}}_\mu) = \mathfrak{m}^{(w_\mu)} + \sum_{\lambda<\mu} \bar{g}_{\lambda,\mu}\mathfrak{m}^{(w_\lambda)} = \tilde{\mathfrak{u}}_\mu + \sum_{\lambda<\mu} r_{\lambda,\mu}\tilde{\mathfrak{u}}_\lambda. \tag{6–6}$$

This proves that the coefficients in (6–3) satisfy (6–4). Thus the corresponding canonical basis $\{\mathfrak{c}_\lambda\}_{\lambda\in\Lambda}$ is uniquely defined.

Remarks 6.7. (a) The canonical basis defined above is the same as Lusztig's canonical basis. This is because the basis $\mathscr{E}$ is a PBW type basis (see [Ringel 1996, Theorem 7]). We also note that, as in the Hecke algebra case [Kazhdan and Lusztig 1979; 1980], the partial order used in this construction is the same as the one used in the geometric construction (see [Lusztig 1990, §9]).

(b) The relation (6–6) is derived via directed distinguished words. However, it can be used to prove the following result,[2] which generalizes the formula given in Lemma 6.6 to all distinguished words. Thus we may also use nondirected distinguished words in the construction above to obtain canonical bases.

Proposition 6.8. *For any distinguished word $w \in \Omega$, we have*

$$\delta(w) + \varepsilon(w) = -\dim M(\wp(w)) + \dim \operatorname{End} M(\wp(w)).$$

Proof. Let $w \in \Omega$ be distinguished. By (6–2), we have

$$\mathfrak{m}^{(w)} = v^s \tilde{\mathfrak{u}}_{\wp(w)} + \sum_{\lambda<\wp(w)} h_{\lambda,\wp(w)}\tilde{\mathfrak{u}}_\lambda, \tag{6–7}$$

where $s = \delta(w) + \varepsilon(w) + \dim M(\wp(w)) - \dim \operatorname{End} M(\wp(w))$ and $0 \neq h_{\lambda,\wp(w)} \in \mathscr{Z}$ for $\lambda < \wp(w)$. By applying ι to (6–7), we deduce from (6–6) that

$$\iota(\mathfrak{m}^{(w)}) = v^{-s}\tilde{\mathfrak{u}}_{\wp(w)} + \sum_{\lambda<\wp(w)} d_{\lambda,\wp(w)}\tilde{\mathfrak{u}}_\lambda$$

[2]We thank the referee for pointing out the proof.

for some $d_{\lambda,\wp(w)} \in \mathcal{Z}$. Since $\iota(\mathfrak{m}^{(w)}) = \mathfrak{m}^{(w)}$, equating coefficients yields $v^s = v^{-s}$. This implies $s = 0$, that is,

$$\delta(w) + \varepsilon(w) = -\dim M(\wp(w)) + \dim \operatorname{End} M(\wp(w)). \qquad \square$$

7. The type A case

We now give a combinatorial description of the map $\wp : \Omega \to \Lambda$ for the linear quiver

$$Q = A_n : 1 \longrightarrow 2 \longrightarrow \cdots \longrightarrow n-1 \longrightarrow n.$$

We also give an explicit description of the distinguished words in this case. Since A_n is a subquiver of a cyclic quiver, the results obtained below and their proofs are similar to (or even simpler than) those given in [Deng and Du 2005], and the proofs will mostly be omitted.

It is known that, for $1 \leqslant i \leqslant j \leqslant n$, there is a unique (up to isomorphism) indecomposable kA_n-module M_{ij} with top S_i and of length $j - i + 1$, and all M_{ij}, $1 \leqslant i \leqslant j \leqslant n$, form a complete set of nonisomorphic indecomposable kA_n-modules. By Gabriel's theorem, each M_{ij} corresponds to a positive root β_{ij}. Thus $\Phi^+ = \{\beta_{ij} \mid 1 \leqslant i \leqslant j \leqslant n\}$. For each map $\lambda \in \Lambda$, we set $\lambda_{ij} = \lambda(\beta_{ij})$. First, we have the following positivity result, which can be proved by counting and induction on the length of w (compare [Deng and Du 2005, Proposition 9.1]).

Proposition 7.1. *For each $w \in \Omega$ and each $\lambda \in \Lambda$, the polynomial φ_w^λ lies in $\mathbb{N}[q]$.*

Now, for each $i \in I$, we define a map $\sigma_i : \Lambda \to \Lambda$ as follows. For $\lambda \in \Lambda$, if S_{i+1} is not a summand of $M(\lambda)/\operatorname{rad} M(\lambda)$ (i.e., $\lambda_{i+1,l} = 0$ for all l), then $\sigma_i\lambda$ is obtained by adding 1 to λ_{ii} so that $M(\sigma_i\lambda) = M(\lambda) \oplus S_i$; otherwise, $\sigma_i\lambda$ is defined by

$$(\sigma_i\lambda)_{rs} = \begin{cases} \lambda_{rs} & \text{if } (r,s) \neq (i,j), (i+1,j), \\ \lambda_{ij} + 1 & \text{if } (r,s) = (i,j), \\ \lambda_{i+1,j} - 1 & \text{if } (r,s) = (i+1,j), \end{cases}$$

where j is the maximal index with $\lambda_{i+1,j} \neq 0$. We have the following (compare [Deng and Du 2005, Proposition 3.7]).

Proposition 7.2. *Let $i \in I$ and $\lambda \in \Lambda$. Then $S_i * M(\lambda) \cong M(\sigma_i\lambda)$. Therefore $\wp(w) = \sigma_{i_1} \cdots \sigma_{i_m}(0)$ for any $w = i_1 \cdots i_m \in \Omega$.*

Let $w = j_1^{e_1} j_2^{e_2} \cdots j_t^{e_t} \in \Omega$ be in the tight form. For each $0 \leqslant r \leqslant t$, we put $w_r = j_{r+1}^{e_{r+1}} \cdots j_t^{e_t}$ and $\lambda^{(r)} = \wp(w_r)$. In particular, $w_0 = w$ and $w_t = 1$. Further, for $r \geqslant 1$, we have

$$\lambda^{(r-1)} = \wp(w_{r-1}) = \underbrace{\sigma_{j_r} \cdots \sigma_{j_r}}_{e_r}(\lambda^{(r)}).$$

The following result gives a combinatorial description of distinguished words (compare [Deng and Du 2005, 5.5]).

Proposition 7.3. *Let $w = j_1^{e_1} j_2^{e_2} \cdots j_t^{e_t} \in \Omega$ and $\lambda^{(r)}$, with $0 \leqslant r \leqslant t$, be given as above. Then w is distinguished if and only if, for each $1 \leqslant r \leqslant t$, either $\lambda^{(r)}_{j_r j} = 0$ for all $j_r \leqslant j \leqslant n$, or $e_r \leqslant \sum_{a=l_r+1}^{n} \lambda^{(r)}_{j_r+1,a}$ where l_r is the maximal index for which $\lambda^{(r)}_{j_r l_r} \neq 0$.*

Proof. Using a similar argument as in [Deng and Du 2005, Theorem 5.5], one can show that w is distinguished if and only if, for each $1 \leqslant r \leqslant t$, $M(\lambda^{(r-1)})$ admits a unique submodule isomorphic to $M(\lambda^{(r)})$. However, the latter condition is equivalent to the described combinatorial condition, as shown in [Deng and Du 2005, Lemma 5.4]. $\square$

Acknowledgment

The authors thank the universities of New South Wales and Virginia for their hospitality during the writing of the paper, and Brian Parshall for his comments on an early version of the paper.

References

[Bongartz 1996] K. Bongartz, "On degenerations and extensions of finite-dimensional modules", *Adv. Math.* **121**:2 (1996), 245–287. MR 98e:16012 Zbl 0862.16007

[Chari and Xi 1999] V. Chari and N. Xi, "Monomial bases of quantized enveloping algebras", pp. 69–81 in *Recent developments in quantum affine algebras and related topics* (Raleigh, NC, 1998), edited by N. Jing and K. C. Misra, Contemp. Math. **248**, Amer. Math. Soc., Providence, RI, 1999. MR 2001c:17023 Zbl 1054.17503

[Deng and Du 2005] B. Deng and J. Du, "Monomial bases for quantum affine $\mathfrak{sl}_n$", *Adv. Math.* **191**:2 (2005), 276–304. MR 2103214 Zbl 02128861

[Du 1994] J. Du, "IC bases and quantum linear groups", pp. 135–148 in *Algebraic groups and their generalizations: quantum and infinite-dimensional 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 95d:17010

[Gabriel 1972] P. Gabriel, "Unzerlegbare Darstellungen, I", *Manuscripta Math.* **6** (1972), 71–103. Correction, ibid. **6** (1972), 309. MR 48 #11212 Zbl 0232.08001

[Kashiwara 1991] M. Kashiwara, "On crystal bases of the Q-analogue of universal enveloping algebras", *Duke Math. J.* **63**:2 (1991), 465–516. MR 93b:17045 Zbl 0739.17005

[Kazhdan and Lusztig 1979] D. Kazhdan and G. Lusztig, "Representations of Coxeter groups and Hecke algebras", *Invent. Math.* **53**:2 (1979), 165–184. MR 81j:20066 Zbl 0499.20035

[Kazhdan and Lusztig 1980] D. Kazhdan and G. Lusztig, "Schubert varieties and Poincaré duality", pp. 185–203 in *Geometry of the Laplace operator* (Honolulu, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, R.I., 1980. MR 84g:14054 Zbl 0461.14015

[Lusztig 1990] G. Lusztig, "Canonical bases arising from quantized enveloping algebras", *J. Amer. Math. Soc.* **3**:2 (1990), 447–498. MR 90m:17023 Zbl 0703.17008

[Reineke 2001a] M. Reineke, "Feigin's map and monomial bases for quantized enveloping algebras", *Math. Z.* **237**:3 (2001), 639–667. MR 2002f:17029 Zbl 1038.17008

[Reineke 2001b] M. Reineke, "Generic extensions and multiplicative bases of quantum groups at $q = 0$", *Represent. Theory* **5** (2001), 147–163. MR 2002c:17029 Zbl 1050.17015

[Ringel 1990] C. M. Ringel, "Hall algebras", pp. 433–447 in *Topics in algebra, part 1: Rings and representations of algebras* (Warsaw, 1988), edited by S. Balcerzyk et al., Banach Center Publ. **26**, PWN, Warsaw, 1990. MR 93f:16027 Zbl 0778.16004

[Ringel 1993a] C. M. Ringel, "The composition algebra of a cyclic quiver: Towards an explicit description of the quantum group of type $\tilde{A}_n$", *Proc. London Math. Soc.* (3) **66**:3 (1993), 507–537. MR 94g:16013 Zbl 0797.16014

[Ringel 1993b] C. M. Ringel, "Hall algebras revisited", pp. 171–176 in *Quantum deformations of algebras and their representations* (Ramat-Gan and Rehovot, 1991/1992), edited by A. Joseph and S. Shnider, Israel Math. Conf. Proc. **7**, Bar-Ilan Univ., Ramat Gan, 1993. MR 94k:16021 Zbl 0852.17009

[Ringel 1995] C. M. Ringel, "The Hall algebra approach to quantum groups", pp. 85–114 in *XI Escuela Latinoamericana de Matemáticas* (Mexico City, 1993), edited by X. Gómez-Mont et al., Aportaciones Mat. Comun. **15**, Soc. Mat. Mexicana, Mexico City, 1995. MR 96m:17034

[Ringel 1996] C. M. Ringel, "PBW-bases of quantum groups", *J. Reine Angew. Math.* **470** (1996), 51–88. MR 97d:17009 Zbl 0840.17010

Received August 15, 2003. Revised June 10, 2004.

BANGMING DENG
DEPARTMENT OF MATHEMATICS
BEIJING NORMAL UNIVERSITY
BEIJING 100875
CHINA
dengbm@bnu.edu.cn

JIE DU
SCHOOL OF MATHEMATICS
UNIVERSITY OF NEW SOUTH WALES
UNSW SYDNEY NSW 2052
AUSTRALIA
j.du@unsw.edu.au
http://www.maths.unsw.edu.au/~jied

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

FLAT MODULES AND LIFTING OF FINITELY GENERATED PROJECTIVE MODULES

ALBERTO FACCHINI, DOLORS HERBERA AND ISKHAK SAKHAJEV

We introduce nets in rings, which turn out to describe right flat modules and left flat modules over a fixed ring R at the same time. As an application we prove that for a finitely generated projective right $R/J(R)$-module P, there exists a finitely generated flat right R-module M with $M/MJ(R)$ isomorphic to P if and only if there exists a projective left R-module P' with $P'/J(R)P'$ isomorphic to the dual of P.

1. Introduction

Although there is a close relation between finitely generated projective right R-modules and finitely generated projective left R-modules given by the duality $\mathrm{Hom}_R(-, R)$, there does not seem to be such an evident relation between finitely generated flat right R-modules and finitely generated flat left R-modules. In this paper we define an algebraic object that allows us to describe right flat modules and left flat modules at the same time. We call this algebraic object a net, because its definition recalls the definition of nets encountered in topology. Our concept finds its origin in [Vasconcelos 1969, proof of Theorem 2.1], and was implicitly used in [Lazard 1974; Sakhaev 1987; 1993; 1996]. As an application of our theory, we study how projective modules over the ring $R/J(R)$ lift to projective or flat modules over R. For instance, we find that for a finitely generated projective right $R/J(R)$-module P, there exists a finitely generated *flat right* R-module M with $M/MJ(R)$ isomorphic to P if and only if there exists a *projective left* R-module P' with $P'/J(R)P'$ isomorphic to the dual $\mathrm{Hom}_{R/J(R)}(P, R/J(R))$ of P (Theorem 7.1).

MSC2000: 16D40.
Keywords: flat modules, projective covers.
The research of the first author was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (Italy) and by Departament d'Universitats, Recerca i Societat de la Informació (Generalitat de Catalunya, Spain). This paper was written during a sabbatical year at the Centre de Recerca Matemàtica (Barcelona). The first author acknowledges the kind hospitality received. The research of the second author was partially supported by the DGI and the European Regional Development Fund, jointly, through Project BFM2002-01390, and by the Comissionat per Universitats i Recerca of the Generalitat de Catalunya.

The paper is organized as follows. In the next two sections we give our basic definitions and constructions. We define nets in rings and show how it is possible to associate to each net both a flat cyclic right module and a flat cyclic left module. In Section 4 we prove that this construction allows us to describe all flat right or left modules.

In Section 5 we give a couple of examples. The first one is the flat module introduced in [Bass 1960]. The second one is based on [Sakhaev 1987; 1993; 1996] and is the key tool in the last two sections to study finitely generated flat modules that are projective modulo the Jacobson radical.

Our rings are associative and have an identity. Modules are unital. For every module M_R, we denote by $\mathscr{L}(M_R)$ the set of all submodules of M_R. The Jacobson radical of a ring R is denoted by $J(R)$.

2. Nets in rings

In this section we introduce the concepts that will be used freely throughout the paper.

Let A be a set with a transitive relation $<$. (We denote the relation by $<$, not $\leq$, to stress that it is not necessarily reflexive.)

Definition 2.1. A *net* in A is a pair (Λ, φ), where

(1) Λ is a nonempty partially ordered set, without a greatest element, without a least element, and with Λ upward directed and downward directed (that is, for each pair λ, μ in Λ there exist ν and ξ in Λ such that $\lambda \leq \nu$, $\mu \leq \nu$, $\xi \leq \lambda$ and $\xi \leq \mu$);

(2) $\varphi : \Lambda \to A$ is a strictly increasing map, that is, for every $\lambda, \mu \in \Lambda$, $\lambda \leq \mu$ and $\lambda \neq \mu$ implies $\varphi(\lambda) < \varphi(\mu)$.

For every $\lambda, \mu \in \Lambda$, we shall write $\lambda < \mu$ whenever $\lambda \leq \mu$ and $\lambda \neq \mu$. Whenever (Λ, φ) is a net in A, we will usually write a_λ instead of $\varphi(\lambda)$. The standard notation for the net will be $(a_\lambda)_{\lambda \in \Lambda}$.

Let S be a ring. Let $<$ be the relation on S defined by $s < t$ if $ts = s$ for $s, t \in S$.

Proposition 2.2. *Let S be a ring with the relation just defined.*

(i) *The relation $<$ is transitive.*

(ii) *If $s, t \in S$ and t is idempotent, then $s < t$ if and only if $sS \subseteq tS$.*

(iii) *For every $s, t \in S$, $s < t$ and $t < s$ if and only if s and t are both idempotent and $sS = tS$. In particular, for every $s \in S$, $s < s$ if and only if s is idempotent.*

(iv) *For every $s, t \in S$, $s < t$ in S if and only if $1 - t < 1 - s$ in the opposite ring S^{op} of S.*

Proof. Properties (i), (ii) and (iv) are trivial. For (iii), suppose that $s < t$ and $t < s$. Then $ts = s$ and $st = t$, so that $s^2 = s(ts) = ts = s$. By symmetry, t also is idempotent. Now (iii) follows from (ii). □

Let $(s_\lambda)_{\lambda\in\Lambda}$ be a net in a ring S with the transitive relation $<$ just defined. Then:

(1) From $s_\mu s_\lambda = s_\lambda$ it follows that $s_\lambda S \subseteq s_\mu S$ whenever $\lambda \le \mu$, so that $(s_\lambda S)_{\lambda\in\Lambda}$ is a net in the set $\mathscr{L}(S_S)$ with the transitive relation $\subseteq$.

(2) The canonical projections $S/s_\lambda S \to S/s_\mu S$ give a direct system of right S-modules over the upward directed set Λ. We shall denote the direct limit $S/\bigcup_{\lambda\in\Lambda} s_\lambda S$ of this direct system by $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$, and call it the *upper limit* of the net $(s_\lambda)_{\lambda\in\Lambda}$.

(3) From Proposition 2.2(iv) it follows that $(1-s_\lambda)_{\lambda\in\Lambda^{\mathrm{op}}}$ is a net in S^{op} defined on the opposite partially ordered set Λ^{op} of Λ. Thus in the ring S we have that $S(1-s_\mu) \subseteq S(1-s_\lambda)$ for $\lambda \le \mu$ in Λ, so that the canonical projections $S/S(1-s_\mu) \to S/S(1-s_\lambda)$ give a direct system of left S-modules over Λ^{op} (Λ^{op} is upward directed because Λ is downward directed). The direct limit of this direct system of left S-modules is $S/\bigcup_{\lambda\in\Lambda} S(1-s_\lambda)$. We shall denote it by $\underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$, and call it the *lower limit* of the net $(s_\lambda)_{\lambda\in\Lambda}$. It coincides with the upper limit of the net $(1-s_\lambda)_{\lambda\in\Lambda^{\mathrm{op}}}$, which is a right S^{op}-module.

Proposition 2.3. *Let $(s_\lambda)_{\lambda\in\Lambda}$ be a net in a ring S. Then*:

(i) *The upper limit $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$ is a cyclic flat right S-module.*

(ii) *The exact sequence*

$$0 \to \bigcup_{\lambda\in\Lambda} s_\lambda S \to S \to \overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \to 0$$

is pure, and $\bigcup_{\lambda\in\Lambda} s_\lambda S$ is a flat right ideal of S.

(iii) *The upper limit $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$ is projective if and only if there exists $\lambda_0 \in \Lambda$ such that $s_{\lambda_0} S = s_\lambda S$ for any $\lambda \in \Lambda$, $\lambda \ge \lambda_0$. In this case, $s_\lambda^2 = s_\lambda$ for any $\lambda \in \Lambda$, $\lambda > \lambda_0$.*

(iv) *The lower limit $\underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$ is a cyclic flat left S-module.*

(v) *The exact sequence*

$$0 \to \bigcup_{\lambda\in\Lambda} S(1-s_\lambda) \to S \to \underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \to 0$$

is pure, and $\bigcup_{\lambda\in\Lambda} S(1-s_\lambda)$ is a flat left ideal of S.

(vi) *The lower limit $\underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$ is projective if and only if there exists $\mu_0 \in \Lambda$ such that $S(1-s_{\mu_0}) = S(1-s_\lambda)$ for any $\lambda \in \Lambda$, $\lambda \le \mu_0$. In this case, $s_\lambda^2 = s_\lambda$ for any $\lambda \in \Lambda$, $\lambda < \mu_0$.*

Proof. In order to show that $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} = S/\bigcup_{\lambda\in\Lambda} s_\lambda S$ is flat, it is enough to prove that $\left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) L = \left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) \cap L$ for any left ideal L [Anderson and Fuller 1992, Lemma 19.18]. The inclusion $\left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) L \subseteq \left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) \cap L$ always holds. If $x \in \left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) \cap L$, then $x = s_\mu y$ for suitable $\mu \in \Lambda$ and $y \in S$. As Λ does not have a greatest element, there exists $\nu > \mu$, so that $x = s_\mu y = s_\nu s_\mu y = s_\nu x \in \left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) L$. This shows (i).

Statement (ii) follows from (i), because every short exact sequence that ends with a flat module is pure.

To prove (iii), assume that $\lambda_0 \in \Lambda$ is such that $s_{\lambda_0} S = s_\lambda S$ for $\lambda \in \Lambda$, $\lambda \geq \lambda_0$. Then, for every $\lambda > \lambda_0$, there exists $a \in S$ such that $s_\lambda = s_{\lambda_0} a = s_\lambda s_{\lambda_0} a = s_\lambda^2$. Thus $\bigcup_{\lambda\in\Lambda} s_\lambda S$ is generated by an idempotent, hence it is a direct summand of S. Conversely, if $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}$ is projective, then $\bigcup_{\lambda\in\Lambda} s_\lambda S$ is principal, so that there is a λ_0 with $s_{\lambda_0} S = s_\lambda S$ for every $\lambda \geq \lambda_0$.

The proofs of statements (iv) to (vi) are similar. □

Notice that every countable partially ordered set Λ satisfying condition (1) of Definition 2.1 contains an upward and downward cofinal subset order-isomorphic to the ordered set $\mathbb{Z}$. Thus we can always suppose $\Lambda = \mathbb{Z}$ for every countably infinite net.

Examples 2.4. Let S be a ring, and let Λ be a partially ordered set satisfying condition (1) of Definition 2.1.

(1) Let $e \in S$ be an idempotent. Then the constant map $\Lambda \to S$ defined by $\lambda \mapsto e$ for every $\lambda \in \Lambda$ is a net whose upper limit is the projective right module $S/eS \cong (1-e)S$ and whose lower limit is the projective left module $S/S(1-e) \cong Se$.

(2) More generally, let $\varphi : \Lambda \to S$ be a net such that, for every $\lambda \in \Lambda$, $\varphi(\lambda) = e_\lambda$ is an idempotent of S. Equivalently, $\{e_\lambda\}_{\lambda\in\Lambda}$ is a family of, not necessarily distinct, idempotents of S such that $e_\lambda S \subseteq e_\mu S$ for any pair $\lambda < \mu$ in Λ. The upper limit of this net is $S/\bigcup_{\lambda\in\Lambda} e_\lambda S$ and the lower limit is $S/\bigcup_{\lambda\in\Lambda} S(1-e_\lambda)$. Moreover, the upper limit is projective if and only if the family $\{e_\lambda S\}_{\lambda\in\Lambda}$ has a greatest element, $e_{\lambda_0} S$ say, and in this case $\overline{\lim}_S\,(e_\lambda)_{\lambda\in\Lambda} \cong (1-e_{\lambda_0})S$. Dually, the lower limit is projective if and only if $\{e_\lambda S\}_{\lambda\in\Lambda}$ has a least element, $e_{\lambda_1} S$ say, and then $\underline{\lim}_S\,(e_\lambda)_{\lambda\in\Lambda} \cong S e_{\lambda_1}$.

3. Tensoring nets with bimodules

Now we study how elements of nets act on bimodules producing interesting pure exact sequences.

Proposition 3.1. *Let R and S be rings, let ${}_SM_R$ be an S-R-bimodule, and let ${}_RN_S$ be an R-S-bimodule. Assume $(s_\lambda)_{\lambda\in\Lambda}$ is a net in the ring S. Then*:

(i) $(s_\lambda M)_{\lambda\in\Lambda}$ *is a net in* $\mathcal{L}(M_R)$ *with the transitive relation* $\subseteq$, *and* $(M/s_\lambda M)_{\lambda\in\Lambda}$ *is a directed system of right* R*-modules.*

(ii) *There is an exact sequence*

$$0 \to \left(\bigcup_{\lambda\in\Lambda} s_\lambda S\right) \otimes_S M \to S \otimes_S M \to \left(\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}\right) \otimes_S M \to 0,$$

which is a pure sequence of right R*-modules.*

(iii) $\varinjlim M/s_\lambda M \cong M/\sum_{\lambda\in\Lambda} s_\lambda M \cong \overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \otimes_S M.$

(iv) *The module* M_R *is flat if and only if both* $M/\sum_{\lambda\in\Lambda} s_\lambda M$ *and* $\sum_{\lambda\in\Lambda} s_\lambda M$ *are flat.*

(v) $(N(1-s_\lambda))_{\lambda\in\Lambda}$ *is a net in* $\mathcal{L}({}_RN)$ *with the transitive relation* $\subseteq$, *and*

$$(N/N(1-s_\lambda))_{\lambda\in\Lambda}$$

is a directed system of left R*-modules.*

(vi) *There is an exact sequence*

$$0 \to N \otimes_S \left(\bigcup_{\lambda\in\Lambda} S(1-s_\lambda)\right) \to N \otimes_S S \to N \otimes_S \left(\underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}\right) \to 0,$$

which is a pure sequence of left R*-modules.*

(vii) $\varinjlim N/N(1-s_\lambda) \cong N/\sum_{\lambda\in\Lambda} N(1-s_\lambda) \cong N \otimes_S \underline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda}.$

(viii) *The left module* ${}_RN$ *is flat if and only if both* $N/\sum_{\lambda\in\Lambda} N(1-s_\lambda)$ *and* $\sum_{\lambda\in\Lambda} N(1-s_\lambda)$ *are flat.*

Proof. (i) follows easily from the fact that $(s_\lambda)_{\lambda\in\Lambda}$ is a net in S. (ii) follows from Proposition 2.3(ii) and the associativity of tensor product.

Let $\{e_\lambda\}_{\lambda\in\Lambda}$ be the canonical basis of the free right S-module $S^{(\Lambda)}$. Setting $f(e_\lambda) = s_\lambda$ we obtain from Proposition 2.3(ii) an exact sequence

$$S^{(\Lambda)} \xrightarrow{f} S \to \overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \to 0.$$

Tensoring this exact sequence with M, we get that $\overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \otimes_S M$ is isomorphic to the cokernel of $f \otimes_S 1_M : S^{(\Lambda)} \otimes_S M \to S \otimes_S M$. Thus (iii) follows from (ii).

To prove (iv) notice that the sequence

$$0 \to \bigcup_{\lambda\in\Lambda} s_\lambda S \otimes_S M \cong \sum_{\lambda\in\Lambda} s_\lambda M \to S \otimes_S M \cong M \to \overline{\lim}_S\,(s_\lambda)_{\lambda\in\Lambda} \otimes_S M \to 0$$

is pure.

The proof of statements (v) to (viii) is similar. □

In the next examples we apply Proposition 3.1 to Examples 2.4(2).

Examples 3.2. Let M_R be an arbitrary right module over a ring R. Let S be the endomorphism ring $\mathrm{End}(M_R)$, so that ${}_SM_R$ is a bimodule.
(1) As in Examples 2.4(2), let $(e_\lambda)_{\lambda\in\Lambda}$ be a net of idempotents of S. Then, for each $\lambda \in \Lambda$, $e_\lambda M$ is a direct summand of M and $e_\lambda S_S \cong \mathrm{Hom}_R(M, e_\lambda M)_S$. In view of Proposition 2.3, $K = \bigcup_{\lambda\in\Lambda} \mathrm{Hom}_R(M, e_\lambda M)_S$ is a pure flat right ideal of S and S/K is a cyclic flat S-module. By Proposition 3.1, we obtain a pure exact sequence of right R-modules

$$0 \to \sum_{\lambda\in\Lambda} e_\lambda M \to M \to M \Big/ \sum_{\lambda\in\Lambda} e_\lambda M \to 0.$$

(2) Nets as in (1) can be also constructed directly from a suitable family of direct summands of M. Let Λ' be a nonempty, upward directed and downward directed subset of $\mathscr{L}(M_R)$ whose elements are direct summands of M_R. Let $\Lambda = \Lambda' \times \mathbb{Z}$ be partially ordered with the lexicographic order, so that Λ is upward directed and downward directed and does not have a greatest element and a least element. For every $\lambda \in \Lambda'$ fix an idempotent $e_\lambda \in S$ with image λ. Let $\varphi : \Lambda \to S$ be defined by $\varphi : (\lambda, n) \mapsto e_\lambda$ for every $(\lambda, n) \in \Lambda$. Then $(e_\lambda)_{\lambda\in\Lambda}$ is a net of idempotents of S.

(3) Assume $M = \bigoplus_{\alpha\in A} M_\alpha$. Let Λ be the set of all finite subsets of A. For each subset λ of A, let $M_\lambda = \bigoplus_{\alpha\in\lambda} M_\alpha$, and let e_λ be the idempotent endomorphism of M with image M_λ and kernel $M_{A\setminus\lambda}$. Then

$$K = \sum_{\lambda\in\Lambda} e_\lambda S = \sum_{\lambda\in\Lambda} \mathrm{Hom}_R(M, e_\lambda M) = \bigoplus_{\alpha\in A} e_{\{\alpha\}} S$$

is a pure and projective right ideal of S. Note that, if M_α is nonzero for every $\alpha \in A$, then $K = S$ if and only if A is finite, but in any case $S/K \otimes_S M \cong M/\bigoplus_{\alpha\in A} M_\alpha = 0$. The set Λ has a least element $\varnothing$, and, when A is finite, a greatest element A. However, taking $\Lambda' = \Lambda \times \mathbb{Z}$ with the lexicographic order, we obtain a partially ordered set with the properties required for index sets of nets.

4. All flat right modules and all flat left modules arise from suitable nets

Let I be a nonempty set, and let R be a ring. Let $F_R = \{ f : I \times \{1\} \to R \mid f((i, 1)) = 0$ for almost all $i \in I\}$. Then F_R is a free right R-module isomorphic to $R_R^{(I)}$, and we will rather think of it as the right R-module of all columns indexed by I, with entries in R, and at most finitely many nonzero entries. Let $\{e_i \mid i \in I\}$ be the canonical basis of F_R.

Let $F^0 = \{ f : \{1\} \times I \to R \mid f((1, i)) = 0$ for almost all $i \in I\}$. Then F^0 is a free left R-module isomorphic to ${}_RR^{(I)}$, and we will think of it as the set of all rows indexed by I, with entries in R, and at most finitely many nonzero entries. Also denote by $\{e_i \mid i \in I\}$ the canonical basis of ${}_RF^0$.

Let $\mathbb{RCF}(I, R)$ denote the ring of all square matrices indexed by $I \times I$ with only a finite number of nonzero entries in each row and each column. Then $\mathbb{RCF}(I, -)$ is a functor of the category of associative rings with identity into itself. Let $B(I, R)$ be the set of all square matrices indexed by $I \times I$, with entries from R, with at most finitely many nonzero entries. The set $B(I, R)$ is a two-sided ideal in $\mathbb{RCF}(I, R)$. If I is finite of cardinality n, $\mathbb{RCF}(I, R) = B(I, R)$ is the ring of all $n \times n$ square matrices over R.

Let $S(I, -)$ be a subfunctor of $\mathbb{RCF}(I, -)$ with the following property: for every ring R, the subring $S(I, R)$ of $\mathbb{RCF}(I, R)$ contains $B(I, R)$ (and contains the identity of $\mathbb{RCF}(I, R)$). For instance, $S(I, R)$ could be the ring $\mathbb{RCF}(I, R)$ itself; or the subring $B(I, R) + 1_{\mathbb{RCF}(I,R)} \cdot R$, where $1_{\mathbb{RCF}(I,R)} \cdot R$ is the set of all scalar matrices; or $S(I, R) = B(I, R) + 1_{\mathbb{RCF}(I,R)} \cdot \mathbb{Z}$.

From now on, in this section, we specialize nets to the rings $S = S(I, R)$. Notice that F is an S-R-bimodule and F^0 is an R-S-bimodule.

Let $(A_\lambda)_{\lambda\in\Lambda}$ be a net in S. We can apply Proposition 3.1 and obtain a flat right R-module $\overline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \otimes_S F \cong F_R / \bigcup_{\lambda\in\Lambda} A_\lambda F_R$ with presentation

$$0 \to \bigcup_{\lambda\in\Lambda} A_\lambda S \otimes_S F \cong \sum_{\lambda\in\Lambda} A_\lambda F \to S \otimes_S F \cong F \to \overline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \otimes_S F \to 0$$

and a flat left R-module $F^0 \otimes_S \underline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \cong F^0 / \bigcup_{\lambda\in\Lambda} F^0(1 - A_\lambda)$ with presentation

$$\begin{aligned} 0 \to F^0 \otimes_S \bigcup_{\lambda\in\Lambda} S(1 - A_\lambda) \cong \sum_{\lambda\in\Lambda} F^0(1 - A_\lambda) \to F^0 \otimes_S S \cong F^0 \to \\ \to F^0 \otimes_S \underline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \to 0. \end{aligned}$$

In the following theorem we show that all flat right R-modules and all flat left R-modules arise in this way from a net in $S = S(I, R)$ for a suitable set I.

Theorem 4.1. *Let $F_R \to M_R$ be an epimorphism of the free right R-module $F_R \cong R_R^{(I)}$ onto a flat right R-module M_R. Then there exists a net $(A_\lambda)_{\lambda\in\Lambda}$ in $S = S(I, R)$ with $A_\lambda \in B(I, R)$ for every $\lambda \in \Lambda$ and $\overline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \otimes_S F \cong M_R$. Dually, let ${}_RF^0 \to {}_RN$ be an epimorphism of the free left R-module ${}_RF^0 \cong {}_RR^{(I)}$ onto a flat left R-module ${}_RN$. Then there exists a net $(B_\lambda)_{\lambda\in\Lambda}$ in $S = S(I, R)$ with $1 - B_\lambda \in B(I, R)$ for every $\lambda \in \Lambda$ and $F^0 \otimes_S \underline{\lim}_S\,(B_\lambda)_{\lambda\in\Lambda} \cong {}_RN$.*

Proof. For the proof, we need the following result, which is a corollary of a theorem due to O. Villamayor [Lam 1999, Theorem 4.23].

Proposition 4.2. *Let $\psi : F_R \to M_R$ be an epimorphism of the free right R-module $F_R \cong R_R^{(I)}$ onto a flat right R-module M_R. Then for any finitely generated submodule C of* $\ker\psi$ *there exists $A \in B(I, R)$ such that $\psi(AF_R) = 0$ and $Ax = x$ for every $x \in C$.*

Proof. The proof of [Lam 1999, Theorem 4.23 (1)⇒(2)] shows that for any $c \in \ker\psi$ there exists $\vartheta \in \mathrm{Hom}(F, \ker\psi)$ with $\vartheta(c) = c$ such that $\vartheta(e_i) = 0$ for almost all $i \in I$. The proof by induction of [Lam 1999, Theorem 4.23, (2) ⇒ (3)] shows that for any $c_1, \dots, c_n \in \ker\psi$ there exists $\vartheta \in \mathrm{Hom}(F, \ker\psi)$ with $\vartheta(c_j) = c_j$ for all $j = 1, \dots, n$ and such that $\vartheta(e_i) = 0$ for almost all $i \in I$. If $c_1, \dots, c_n$ generate the submodule C of $\ker\psi$, then the matrix A associated to ϑ with respect to the basis $\{e_i \mid i \in I\}$ has the required properties. □

We are now ready for the proof of Theorem 4.1. Let $\psi : F_R \to M_R$ be an epimorphism of F_R onto a flat right R-module M_R, and let K be the kernel of ψ.

Suppose that K is not finitely generated. Let G be a set of generators of K. Let $\mathcal{P}_{\mathrm{fin}}(G)$ denote the set of all finite subsets of G, partially ordered by set inclusion. Let $\mathbb{Z}^-$ be the set of negative integers with its usual order, and let Λ be the disjoint union of $\mathbb{Z}^-$ and $\mathcal{P}_{\mathrm{fin}}(G)$. Define $z < H$ for every $z \in \mathbb{Z}^-$ and every $H \in \mathcal{P}_{\mathrm{fin}}(G)$, so that Λ turns out to be upward directed and downward directed, without a greatest element and without a least element. In order to define a net $\{A_\lambda \mid \lambda \in \Lambda\}$ in S, first of all set $A_z = 0$ for $z \in \mathbb{Z}^-$. Then define, for each $H \in \mathcal{P}_{\mathrm{fin}}(G)$, a matrix $A_H \in B(I, R)$ by induction on the cardinality $|H|$ of H. For $H = \varnothing$, set $A_\varnothing = 0$. Let $H \in \mathcal{P}_{\mathrm{fin}}(G)$ with $|H| > 0$ and suppose that $A_{H'}$ has already been defined for every $H' \in \mathcal{P}_{\mathrm{fin}}(G)$ with $|H'| < |H|$. Since H has only finitely many proper subsets, the submodule C of K generated by H and by all $A_{H'}F_R$ when H' ranges in the set of all proper subsets of H is a finitely generated submodule of K. By Proposition 4.2, there exists $A_H \in B(I, R)$ such that $A_H x = x$ for every $x \in C$ and $A_H F_R \subseteq K$. This completes the definition of the matrix A_H. Notice that $A_{H'}F_R \subseteq C$, so that $A_H A_{H'} = A_{H'}$ whenever $H' \subset H$. Thus $(A_\lambda)_{\lambda\in\Lambda}$ is a net with the required properties.

If the module K is finitely generated, there is a finite subset J of I with $K \subseteq \bigoplus_{i\in J} e_i R$. Now M_R is flat and

$$M_R \cong F_R/K \cong \Bigl(\bigoplus_{i\in J} e_i R/K\Bigr) \oplus \Bigl(\bigoplus_{i\in I\setminus J} e_i R\Bigr).$$

Thus $\bigl(\bigoplus_{i\in J} e_i R\bigr)/K$ is flat and finitely presented, hence projective. It follows that K is a direct summand of $\bigoplus_{i\in J} e_i R$. Thus K is a direct summand of F_R and there is an idempotent endomorphism ε of F_R with image K. Let A be the matrix associated to ε with respect to the basis $\{e_i \mid i \in I\}$ of F_R. Then the partially ordered set $\mathbb{Z}$ of the integers with the matrices $A_z = A$ for every $z \in \mathbb{Z}$ form a net in S with upper limit $\overline{\lim}_S (A_z)_{z\in\mathbb{Z}} \otimes_S F \cong F/\sum_{z\in\mathbb{Z}} A_z F \cong M_R$; cf. Examples 2.4(1). This concludes the case of K finitely generated.

Dually, let ${}_RF^0 \to {}_RN$ be an epimorphism of ${}_RF^0$ onto a flat left R-module ${}_RN$. Passing to the opposite ring R^{op} of R, one has an epimorphism $F \to N$ of the free right R^{op}-module F onto the flat right R^{op}-module N. By applying to this epimorphism the first part of the statement, which we have just proved, we see that there exists a net $(C_\lambda)_{\lambda\in\Lambda}$ in $S(I, R^{\mathrm{op}})$ with $\overline{\lim}\,(C_\lambda)_{\lambda\in\Lambda}\otimes F \cong F/\sum_{\lambda\in\Lambda} C_\lambda F \cong N$ as right R^{op}-modules. In particular, the C_λ's belong to $S(I, R^{\mathrm{op}})$, and $C_\mu C_\lambda = C_\lambda$ whenever $\lambda, \mu \in \Lambda$ and $\lambda < \mu$. Viewing these objects as left R-modules again and remarking that transposition is an isomorphism $tr : S(I, R^{\mathrm{op}}) \to (S(I, R))^{\mathrm{op}}$, we have that the C_λ^{tr}'s belong to $S(I, R)$, that $C_\lambda^{\mathrm{tr}} C_\mu^{\mathrm{tr}} = C_\lambda^{\mathrm{tr}}$ in $S(I, R)$ whenever $\lambda < \mu$, and ${}_RN \cong {}_RF^0/\sum_{\lambda\in\Lambda}({}_RF^0)C_\lambda^{\mathrm{tr}}$. From $C_\lambda^{\mathrm{tr}} C_\mu^{\mathrm{tr}} = C_\lambda^{\mathrm{tr}}$ for $\lambda < \mu$, we obtain that $(1 - C_\lambda^{\mathrm{tr}})(1 - C_\mu^{\mathrm{tr}}) = 1 - C_\mu^{\mathrm{tr}}$ in $S = S(I, R)$ for $\lambda < \mu$. Denoting the set Λ with the inverse order by Λ^{op}, we see that there is a net $(1 - C_\lambda^{\mathrm{tr}})_{\lambda\in\Lambda^{\mathrm{op}}}$ in S and

$$F^0 \otimes_S \underline{\lim}_S (1 - C_\lambda^{\mathrm{tr}})_{\lambda\in\Lambda^{\mathrm{op}}} \cong {}_RF^0 \Big/ \sum_{\lambda\in\Lambda} {}_RF^0 C_\lambda^{\mathrm{tr}} \cong {}_RN. \qquad \square$$

Remark 4.3. Let S and S' be rings. A ring homomorphism $f : S \to S'$ induces for every net $(s_\lambda)_{\lambda\in\Lambda}$ in S a net $(f(s_\lambda))_{\lambda\in\Lambda}$ in S'.

For instance, a ring homomorphism $g : R \to R'$ induces a ring homomorphism $\tilde{g} = S(I, g) : S = S(I, R) \to S' = S(I, R')$. If $(A_\lambda)_{\lambda\in\Lambda}$ is a net in S, then

$$\big(\overline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda} \otimes_S F_R\big) \otimes_R R' \cong \overline{\lim}_{S'}\,(\tilde{g}(A_\lambda))_{\lambda\in\Lambda} \otimes_{S'} F'_{R'} \tag{1}$$

and

$$R' \otimes_R \big({}_RF^0 \otimes_S \underline{\lim}_S\,(A_\lambda)_{\lambda\in\Lambda}\big) \cong {}_{R'}(F')^0 \otimes_{S'} \underline{\lim}_{S'}\,(\tilde{g}(A_\lambda))_{\lambda\in\Lambda}, \tag{2}$$

where we have denoted by $F'_{R'}$ the free right R'-module of rank $|I|$ and by ${}_{R'}(F')^0$ the free left R'-module of same rank.

We will be particularly interested in the case in which $g : R \to R/J(R)$ is the canonical projection.

5. Two examples of flat modules

As a first example, we shall consider the flat module F_R/G introduced in the seminal paper [Bass 1960]. Fix a sequence a_n $(n \geq 1)$ of elements of a given ring R, let F_R be the free right R-module with basis $\{e_n \mid n \geq 1\}$, and let G be the submodule of F_R generated by the elements $y_n = e_n - e_{n+1}a_n$, for $n \geq 1$. It is known that G is a free R-module with basis $\{y_n \mid n \geq 1\}$ and F_R/G is a flat module [Bass 1960; Anderson and Fuller 1992, Lemma 28.1]. The module F_R/G is projective, that is, G is a direct summand of F_R, if and only if all the descending chains $Ra_n \supseteq Ra_{n+1}a_n \supseteq Ra_{n+2}a_{n+1}a_n \supseteq \cdots$, for $n \geq 1$, are stationary [Azumaya 1987, Theorem 26].

Let S be the ring $S(\mathbb{Z}^+, R)$, where $\mathbb{Z}^+$ denotes the set of all positive integers. Let $(A_n)_{n\in\mathbb{Z}}$ be the net in S defined by $A_n = 0$ for $n \le 0$, and

$$A_n = \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & \\ \hline -a_n \dots a_2 a_1 & -a_n \dots a_2 & \cdots & -a_n & \\ & & & & 0 \\ & 0 & & & \end{array}\right)$$

for $n \ge 1$. An easy computation shows that $A_n^2 = A_n$ for any $n \in \mathbb{Z}$. In particular, $A_n F_R$ is a direct summand of F_R.

Proposition 5.1. *For every $n \ge 1$ the right R-module $A_n F_R$ is the free submodule of G generated by $y_1, \dots, y_n$.*

Proof. We must show that the right R-module $A_n F_R$, generated by

$$e_1 - e_{n+1} a_n \dots a_2 a_1, \dots, e_n - e_{n+1} a_n$$

coincides with the right module generated by $y_1 = e_1 - e_2 a_1, \dots, y_n = e_n - e_{n+1} a_n$.

$$\begin{aligned} e_i - e_{n+1} a_n \dots a_{i+1} a_i &= (e_i - e_{i+1} a_i) + (e_{i+1} - e_{i+2} a_{i+1}) a_i \\ &\quad + (e_{i+2} - e_{i+3} a_{i+2}) a_{i+1} a_i + \cdots + (e_n - e_{n+1} a_n) a_{n-1} a_{n-2} \dots a_i. \end{aligned}$$

Conversely, for $i < n$, we see that $y_i = e_i - e_{i+1} a_i = (e_i - e_{n+1} a_n \dots a_{i+1} a_i) - (e_{i+1} - e_{n+1} a_n \dots a_{i+1}) a_i$. $\square$

Thus $\sum_{n\in\mathbb{Z}} A_n F_R = G$ and $\varinjlim_S (A_\lambda)_{\lambda\in\Lambda} \otimes_S F \cong F_R/G$.

Let $E = \mathrm{End}_R(F_R)$. Let $K_1 = \bigcup_{n=1}^\infty \mathrm{Hom}_R(F, \sum_{j=1}^n e_j R)$, and let $K_2 = \bigcup_{n=1}^\infty \mathrm{Hom}_R(F, A_n F)$. In view of Examples 3.2, K_1 and K_2 are pure right ideals of E. Note that they are also projective [Lazard 1969, Théorème 3.2].

Proposition 5.2. *The cyclic right E-modules E/K_1 and E/K_2 are flat and non-isomorphic. If the elements of the sequence a_n $(n \ge 1)$ belong to $J(R)$, then $E/K_1 \otimes_E E/J(E) \cong E/K_2 \otimes_E E/J(E)$.*

Proof. Applying the functor $- \otimes_E F$ to the pure exact sequence

$$0 \to K_1 \to E \to E/K_1 \to 0,$$

it follows that $E/K_1 \otimes_E F = 0$ (Examples 3.2). While applying the functor $- \otimes_E F$ to the pure exact sequence

$$0 \to K_2 \to E \to E/K_2 \to 0,$$

it follows that $K_2 \otimes_E F \cong G$, hence $E/K_2 \otimes_E F \cong F/G$.

If a_n $(n \geq 1)$ is a sequence of elements in $J(R)$ and $g : R \to R/J(R)$ denotes the canonical projection, then $\tilde{g}(K_1) = \tilde{g}(K_2)$, so

$$E/K_1 \otimes_E E/J(E) \cong E/K_2 \otimes_E E/J(E)$$

in view of Remark 4.3. □

The isomorphism $f : F \to G$ defined by $f(e_n) = y_n$ for every $n \geq 1$ induces an isomorphism between the projective ideals K_1 and K_2.

In the next proposition we give an example that was our initial motivation to define nets. We construct a countable net whose upper limit is nontrivial if and only and only if its lower limit is nontrivial; that is, the net produces a nontrivial right flat module if and only if it produces a nontrivial left flat module. This idea will be further developed and applied in the proof of Theorem 7.1.

Proposition 5.3. *Let S be a ring. Let s and u be elements of S such that u is invertible and $s^2 = us$. For every $m \in \mathbb{Z}$, set $s_m = u^{-m}(u^{-1}s)u^m$. Let $I = \sum_{m\in\mathbb{Z}} s_m S$, and let $L = \sum_{m\in\mathbb{Z}} S(1 - s_m)$. Then:*

(i) $(s_m)_{m\in\mathbb{Z}}$ *is a net.*

(ii) *The right ideal I and the left ideal L are projective. The right S-module S/I and the left S-module S/L are flat.*

(iii) *There exists $m \in \mathbb{Z}$ such that $s_m^2 = s_m$ if and only if $su^{-1}s = s$, if and only if $s_m^2 = s_m$ for all $m \in \mathbb{Z}$.*

(iv) *The right ideal I is finitely generated if and only if the left ideal L is finitely generated, if and only if $su^{-2}s = u^{-1}s$.*

Proof. (i) Direct computation shows that $s_m = s_n s_m$ for $m < n$.

(ii) Note that $S/I = \overline{\lim}_S\, (s_m)_{m\in\mathbb{Z}}$ and $S/L = \underline{\lim}_S\, (s_m)_{m\in\mathbb{Z}}$. By Proposition 2.3, S/I is a flat right module and S/L is a flat left module. Since, by [Lazard 1969, Théorème 3.2], countably presented flat modules have projective dimension 1, I and L are projective.

To prove (iii), observe that $s_m^2 = s_m$ if and only if

$$u^{-m}(u^{-1}s)u^m \cdot u^{-m}(u^{-1}s)u^m = u^{-m}(u^{-1}s)u^m,$$

if and only if $su^{-1}s = s$, as claimed. As this condition does not depend on m, if one s_m is idempotent all must be idempotent.

(iv) First observe that the identity $s_m s_{m+1} = s_{m+1}$ holds for some m if and only if $su^{-2}s = u^{-1}s$. As this condition does not depend on m, this happens if and only if $s_m s_{m+1} = s_{m+1}$ for all m.

Since always $s_m S \subseteq s_{m+1} S$ for every m, $su^{-2}s = u^{-1}s$ implies that $s_m S = s_{m+1} S$ for all m. Hence I is principal.

Conversely, if I_S is finitely generated, then S/I is flat and finitely presented, hence projective, so by Proposition 2.3(iii) there exists m such that $s_m^2 = s_m$ and $I = s_m S$. Then $s_m S = s_{m+1} S$ implies $s_m s_{m+1} = s_{m+1}$, hence $su^{-2}s = u^{-1}s$.

Similar arguments show the statement for L. □

The symmetry in the conclusions of Proposition 5.3 can be explained through the following lemma, which is an observation based on [Zöschinger 1981, Satz 1.2]. See also [Puninski 2004, Section 3].

Lemma 5.4. *Let S be a ring, and let $s \in S$. There exists a unit u such that $s^2 = us$ if and only if there exists $t \in S$ such that $ts = 0$ and $s + t$ is a unit. In this situation, there exists a unit $v \in S$ such that $t^2 = tv$.*

Proof. Assume there exists a unit u such that $s^2 = us$. Then $t = u - s$ satisfies the required properties. Conversely, if there exists $t \in S$ such that $ts = 0$ and $s + t$ is a unit, then taking $u = s + t$ we have that $us = s^2$. Note that then also $tu = t^2$. □

6. Lifting projective modules modulo the Jacobson radical

In this section and the next we apply the theory developed earlier to the lifting of finitely generated projective modules modulo the Jacobson radical.

For every right (left) R-module M_R (${}_RN$), let $M^* = \mathrm{Hom}_R(M_R, R_R)$ ($N^* = \mathrm{Hom}_R({}_RN, {}_RR)$) denote the dual of the module M_R (${}_RN$), which is a left (right) R-module. This defines a duality, that is, a contravariant equivalence, between the full subcategory of finitely generated projective right R-modules and the full subcategory of finitely generated projective left R-modules.

Consider a direct sum decomposition $P \oplus Q = (R/J(R))^n$ of the free right $R/J(R)$-module $(R/J(R))^n$, so that P and Q are two projective right $R/J(R)$-modules. It is easy to see that there exists a finitely generated projective right R-module M_R such that $M/MJ(R) \cong P$ if and only if there exists a finitely generated projective right R-module Q'_R such that $Q'/Q'J(R) \cong Q$, if and only if there exists a finitely generated projective left R-module ${}_RN$ such that $N/J(R)N \cong \mathrm{Hom}_R(Q, R/J(R))$, if and only if there exists a finitely generated projective left R-module ${}_RP'$ such that $P'/J(R)P' \cong \mathrm{Hom}_R(P, R/J(R))$. (To prove this, let M_R be a finitely generated projective right R-module such that $P \cong M/MJ(R)$. Then $M/MJ(R) \oplus Q \cong (R/J(R))^n$. By [Anderson and Fuller 1992, Lemma 17.17] there exists a finitely generated projective right R-module Q' such that $Q'/Q'J(R) \cong Q$ and $M \oplus Q' \cong R^n$. Take $N = \mathrm{Hom}_R(Q', R)$.)

In this section we consider the problem of lifting these projective $R/J(R)$-modules to projective R-modules, not necessarily finitely generated. We need the following result of Bergman [Jøndrup 1976, Lemma 2.2].

Proposition 6.1. *Let Q and Q' be projective right R-modules, and let $\varphi : Q' \to Q$ be a homomorphism. If the mapping $\overline{\varphi} : Q'/Q'J(R) \to Q/QJ(R)$ induced by φ is a pure monomorphism, then φ is a pure monomorphism.*

Proof. First choose an R-module P' such that $Q' \oplus P'$ is free, then an R-module P such that $(Q \oplus P') \oplus P$ is free. Let $\varepsilon : P' \to P' \oplus P$ denote the embedding. Substituting $\varphi : Q' \to Q$ with $\varphi \oplus \varepsilon : Q' \oplus P' \to Q \oplus P' \oplus P$, we may suppose that Q and Q' are free. In order to show that φ is a pure monomorphism, fix a finitely generated free direct summand N' of Q'. Let N be a finitely generated free direct summand of Q containing $\varphi(N')$. Let $f : N \to Q$ and $f' : N' \to Q'$ be the inclusions, and $g : Q \to N$, $g' : Q' \to N'$ be homomorphisms such that $gf = 1_N$ and $g'f' = 1_{N'}$. If $\varphi|_{N'} : N' \to N$ denotes the restriction of $\varphi : Q' \to Q$, then $f\varphi|_{N'} = \varphi f'$. If $^{-}$ denotes reduction modulo $J(R)$, then $\overline{f'}$ is a pure monomorphism, so that $\overline{\varphi}\,\overline{f'}$ is a pure monomorphism. From $f\varphi|_{N'} = \varphi f'$, it follows that $\overline{\varphi|_{N'}}$ is a pure monomorphism. Thus the cokernel of $\overline{\varphi|_{N'}}$ is a flat finitely presented module, that is, a projective finitely generated module. In particular, $\overline{\varphi|_{N'}}$ is a split monomorphism. Let $h : N \to N'$ be a homomorphism such that $1_{N'/N'J(R)} = \overline{h\varphi|_{N'}}$. Then $h\varphi|_{N'}$ is an automorphism of N', so that $\varphi|_{N'}$ is a split monomorphism. In particular, φ is injective, and $\varphi(N')$ is a direct summand of Q for every finitely generated free direct summand N' of Q. As $\varphi(Q')$ is the directed union of all these direct summands $\varphi(N')$, $\varphi(Q')$ is a pure submodule of Q. $\square$

Corollary 6.2. *Let R be a ring with the property that for every projective right $R/J(R)$-module P there exists a projective right R-module Q with $Q/QJ(R) \cong P$. For every flat right $R/J(R)$-module M of projective dimension $pd_{R/J(R)}(M) \leq 1$ there exists a flat right R-module N of projective dimension $pd_R(N) \leq 1$ with $N/NJ(R) \cong M$. Moreover, if M is finitely generated, then N can also be chosen finitely generated.*

Proof. Apply Proposition 6.1 to a presentation

$$0 \longrightarrow Q'/Q'J(R) \xrightarrow{\overline{\varphi}} Q/QJ(R) \longrightarrow M \longrightarrow 0$$

of the $R/J(R)$-module M with Q and Q' projective R-modules. $\square$

The hypothesis of Corollary 6.2 applies to all rings R for which every projective right $R/J(R)$-module is free, and to all exchange rings R.

Proposition 6.3. *Let R be a ring and X a set. Let $P \oplus Q = (R/J(R))^{(X)}$ be a direct sum decomposition of the free right $R/J(R)$-module $(R/J(R))^{(X)}$ as a direct sum of two projective right $R/J(R)$-modules P and Q, and let $\pi : (R/J(R))^{(X)} \to P$ be the projection with kernel Q. The following statements are equivalent:*

(i) *There exist a flat right R-module M_R of projective dimension at most* 1, *an epimorphism $\psi : R^{(X)} \to M_R$ and an isomorphism $\alpha : M_R/M_R J(R) \to P$ such that $\alpha \circ (\psi \otimes R/J(R)) = \pi$.*

(ii) *There exists a projective right R-module Q'_R such that $Q'/Q'J(R) \cong Q$.*

Proof. (i)$\Rightarrow$(ii) Let M_R, ψ and α have the properties stated in (i). We will show that the projective module $Q' = \ker \psi$ has the property required in (ii). From the exact sequence

$$0 \longrightarrow Q' \longrightarrow R^{(X)} \xrightarrow{\psi} M_R \longrightarrow 0,$$

we get the exact sequence

$$0 \longrightarrow Q'/Q'J(R) \longrightarrow (R/J(R))^{(X)} \xrightarrow{\psi\otimes R/J(R)} M_R/M_R J(R) \longrightarrow 0.$$

Thus $Q = \ker \pi = \ker(\alpha \circ (\psi \otimes R/J(R))) = \ker(\psi \otimes R/J(R)) \cong Q'/Q'J(R)$.

(ii)$\Rightarrow$(i) Let Q'_R be a projective R-module such that $Q'/Q'J(R) \cong Q$. Let $\rho : Q' \to Q$ be an epimorphism with kernel $Q'J(R)$. Denote by

$$\varepsilon : Q \to (R/J(R))^{(X)}$$

the embedding, which is a split monomorphism. As $\varepsilon\rho : Q'_R \to (R/J(R))^{(X)}$ factors through the canonical projection of $R^{(X)}$ onto $(R/J(R))^{(X)}$, there is a commutative diagram

$$\begin{tikzcd} Q'_R \arrow[r, "\varphi"] \arrow[d, "\rho"'] & R^{(X)} \arrow[d] \\ Q \arrow[r, "\varepsilon"] & (R/J(R))^{(X)} \end{tikzcd}$$

By Proposition 6.1 the mapping φ is a pure monomorphism, so that its cokernel M_R is a flat module of projective dimension ≤ 1. Let $\psi : R^{(X)} \to M_R$ be the canonical projection. Applying $- \otimes_R R/J(R)$ to the pure exact sequence

$$0 \longrightarrow Q'_R \xrightarrow{\varphi} R^{(X)} \xrightarrow{\psi} M_R \to 0,$$

we obtain an exact sequence that is the upper row of the commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & Q'_R/Q'_R J(R) \arrow[r] \arrow[d] & (R/J(R))^{(X)} \arrow[r, "\psi\otimes R/J(R)"] \arrow[d] & M_R/M_R J(R) \arrow[r] & 0 \\ 0 \arrow[r] & Q \arrow[r] & (R/J(R))^{(X)} \arrow[r, "\pi"] & P \arrow[r] & 0. \end{tikzcd}$$

Here the two vertical arrows are isomorphisms (the vertical arrow on the right is the identity). Thus there is an isomorphism $\alpha : M_R/M_R J(R) \to P$ that completes the commutative diagram; that is, $\alpha \circ (\psi \otimes R/J(R)) = \pi$. $\square$

7. Lifting finitely generated projective modules

Theorem 7.1. *Let $P \oplus Q = (R/J(R))^n$ be a direct sum decomposition of the finitely generated free right $R/J(R)$-module $(R/J(R))^n$ as a direct sum of two projective right $R/J(R)$-modules P and Q. Then the following statements are equivalent*:

(i) *There exists a finitely generated flat right R-module M_R such that $M/MJ(R)$ is isomorphic to P.*

(ii) *There exists a finitely generated, countably presented, flat right R-module M_R such that $M/MJ(R) \cong P$.*

(iii) *There exists a projective right R-module Q'_R such that $Q'/Q'J(R) \cong Q$.*

(iv) *There exists a finitely generated flat left R-module ${}_RN$ such that $N/J(R)N \cong$* $\mathrm{Hom}_R(Q, R/J(R))$.

(v) *There exists a finitely generated, countably presented, flat left R-module ${}_RN$ such that $N/J(R)N \cong$* $\mathrm{Hom}_R(Q, R/J(R))$.

(vi) *There exists a projective left R-module ${}_RP'$ such that $P'/J(R)P'$ is isomorphic to* $\mathrm{Hom}_R(P, R/J(R))$.

Proof. Suppose that (i) holds. Let M_R be a finitely generated flat right R-module such that $M/MJ(R) \cong P$, where $P \oplus Q = (R/J(R))^n$. Let $\alpha : M/MJ(R) \to P$ be an isomorphism, and let $\pi : (R/J(R))^n \to P$ be the projection with kernel Q. The onto mapping $\alpha^{-1}\pi : (R/J(R))^n \to M/MJ(R)$ can be lifted to a homomorphism of right R-modules $\psi : R^n_R \to M_R$, which is necessarily onto by Nakayama's Lemma. Let $K = \ker\psi$ and consider the pure exact sequence of right R-modules

$$0 \longrightarrow K \xrightarrow{\varepsilon} R^n_R \xrightarrow{\psi} M_R \longrightarrow 0.$$

Tensoring by $R/J(R)$, this induces the exact sequence

$$(3) \qquad 0 \longrightarrow K/KJ(R) \xrightarrow{\overline{\varepsilon}} (R/J(R))^n \xrightarrow{\alpha^{-1}\pi} M/MJ(R) \longrightarrow 0,$$

which splits because $M/MJ(R) \cong P$ is projective. Moreover, $Q = \ker\pi = \ker(\alpha^{-1}\pi) \cong K/KJ(R)$. As (3) splits, there exists a left inverse

$$\overline{\varphi} : (R/J(R))^n \to K/KJ(R)$$

of $\overline{\varepsilon}$ with kernel isomorphic to $M/MJ(R) \cong P$. As R^n is projective, $\overline{\varphi}$ can be lifted to a map $\varphi : R^n \to K$ making the diagram

$$\begin{tikzcd}
K \arrow[r, "\varepsilon"] \arrow[d] & R^n_R \arrow[r, "\varphi"] \arrow[d] & K \arrow[d] \\
K/KJ(R) \arrow[r, "\overline{\varepsilon}"] & (R/J(R))^n \arrow[r, "\overline{\varphi}"] & K/KJ(R),
\end{tikzcd}$$

commute. In this diagram the vertical arrows are the natural projections. By Proposition 4.2, there exists $\omega : R^n \to K$ such that $\varepsilon\omega$ is the identity over $\varepsilon\varphi(R^n)$. Equivalently, $\omega\varepsilon\varphi = \varphi$. Denote by $\overline{\omega} : (R/J(R))^n \to K/KJ(R)$ the induced homomorphism. Then $\overline{\omega\varepsilon\varphi} = \overline{\varphi}$. As $\overline{\varphi\varepsilon}$ is the identity mapping, $\overline{\omega - \varphi\varepsilon\omega} = \overline{\omega - \omega} = 0$, so that $(\omega - \varphi\varepsilon\omega)(R^n) \subseteq KJ(R)$, from which $\varepsilon(\omega - \varphi\varepsilon\omega)(R^n) \subseteq (J(R))^n$. Thus $\varepsilon(\omega - \varphi\varepsilon\omega) \in J(\mathrm{End}(R^n))$ [Anderson and Fuller 1992, Corollary 17.12]. Set $\beta = 1 - \varepsilon(\omega - \varphi\varepsilon\omega)$, and note that β is an invertible element of $\mathrm{End}(R^n)$. As $\omega\varepsilon\varphi = \varphi$, it is easy to see that $\beta\varepsilon\varphi = (\varepsilon\varphi)^2$. Observe that β induces the identity endomorphism on $(R/J(R))^n$ and also that $\varepsilon\varphi$ induces the idempotent endomorphism $\overline{\varepsilon\varphi}$ on $(R/J(R))^n$, whose image is $K/KJ(R) \cong Q$ and whose kernel is isomorphic to P. For any $m \in \mathbb{Z}$, let A_m be the matrix associated to the endomorphism $\beta^{-m-1}\varepsilon\varphi\beta^m : R^n_R \to R^n_R$. By Proposition 5.3(i), $(A_m)_{m\in\mathbb{Z}}$ is a net in the ring $S = S(n, R)$ of $n \times n$ matrices over R. Hence the left R-module ${}_RN = R^n / \bigcup_{m\in\mathbb{Z}} R^n(1 - A_m)$ is flat. By Remark 4.3, if we apply isomorphism (2) with $R' = R/J(R)$ and g the canonical projection, and using that $(\tilde{g}(A_m))_{m\in\mathbb{Z}}$ is the net in $S' = S(n, R/J(R))$ constantly equal to the matrix $\overline{A}$ of the endomorphism $\overline{\varepsilon\varphi}$ of $(R/J(R))^n$, we see that

$$(R/J(R))^n \otimes_{S'} \underline{\lim}_{S'}\, (\tilde{g}(A_m))_{m\in\mathbb{Z}} \cong (R/J(R))^n \otimes_{S'} \underline{\lim}_{S'}\, (\overline{A})_{m\in\mathbb{Z}}.$$

By Examples 2.4(1), $\underline{\lim}_{S'}\, (\overline{A})_{m\in\mathbb{Z}} \cong S'\overline{A}$. Thus

$$\begin{aligned} N/J(R)N &\cong R/J(R) \otimes_R N \cong (R/J(R))^n \otimes_{S'} \underline{\lim}_{S'}\, (\tilde{g}(A_m))_{m\in\mathbb{Z}} \cong \\ &\cong (R/J(R))^n \otimes_{S'} \underline{\lim}_{S'}\, (\overline{A})_{m\in\mathbb{Z}} \cong (R/J(R))^n \otimes_{S'} S'\overline{A} \cong (R/J(R))^n(\overline{A}). \end{aligned}$$

Since $Q \cong \overline{A}(R/J(R))^n$, we can conclude $N/J(R)N \cong \mathrm{Hom}_R(Q, R/J(R))$. As N is a finitely generated, countably presented, flat module, this shows that (iv) and (v) hold.

By symmetry, that is, applying (i) implies (iv) and (v) to the opposite ring R^{op}, we see that (iv) implies (i) and (ii). Hence (i), (ii), (iv) and (v) are equivalent statements.

By Proposition 6.3, (iii) implies (i). Assume that (ii) holds, so that there exist a finitely generated, countably presented, flat right R-module M and an isomorphism $\alpha : M/MJ(R) \to P$. The module M has projective dimension ≤ 1 [Lazard 1969, Théorème 3.2]. Let $\pi : (R/J(R))^n \to P$ be the projection with kernel Q. The onto mapping $\alpha^{-1}\pi : (R/J(R))^n \to M/MJ(R)$ can be lifted to a homomorphism of right R-modules $\psi : R^n_R \to M_R$, which is necessarily onto by Nakayama's Lemma. As the conditions of Proposition 6.3(i) are satisfied, we deduce the existence of a right projective module Q' such that $Q'/Q'J(R) \cong Q$. This proves that (ii) implies (iii), so that (ii) and (iii) are equivalent statements. By symmetry, (v) and (vi) are also equivalent. □

Recall that a projective module P is a direct sum of countably generated submodules, and that $P=0$ if and only if $P/PJ(R)=0$. Hence, if a projective module is finitely generated modulo the Jacobson radical, it must be countably generated. Thus the modules P' and Q' in the statement of Theorem 7.1 are necessarily countably generated.

It would be interesting to know whether the module M in Theorem 7.1(ii) is uniquely determined up to isomorphism. In Proposition 5.2 we saw an example of countably presented nonisomorphic cyclic flat modules that are isomorphic modulo the Jacobson radical, but the cyclic modules in that example are not projective modulo the Jacobson radical.

We conclude with two results related to this question, the first of which appears as [Lam 1999, p. 161, Exercise 20]. We give a proof for the sake of completeness.

Lemma 7.2. *Let M be a finitely generated flat right module over a ring R, and let P be a projective right R-module. If $\gamma : P \to M$ is a projective cover, then γ is an isomorphism.*

Proof. The module P is finitely generated because M is finitely generated and $\ker\gamma$ is small in P. Hence there exist n and a projective module Q such that $P \oplus Q \cong R^n$. As $\gamma \oplus 1_Q : P \oplus Q \to M \oplus Q$ is a projective cover, and γ is an isomorphism if and only if so is $\gamma \oplus 1_Q$, we may assume without loss of generality that P is R^n and that $\gamma : R^n \to M$ is a projective cover.

Let $x \in \ker\gamma$. By Proposition 4.2, there exists $A \in M_n(R)$ such that $Ax = x$ and $AR^n \subseteq \ker\gamma \subseteq R^n J(R)$. This implies that $(1-A)x = 0$ and that $A \in M_n(J(R))$, thus $x = 0$. This shows that $\ker\gamma = 0$, hence γ is an isomorphism. ☐

Proposition 7.3. *Let M be a finitely generated flat right module over a ring R, and let P be a projective module. If $M/MJ(R) \cong P/PJ(R)$, then $M \cong P$.*

Proof. As $M/MJ(R) \cong P/PJ(R)$, the module $P/PJ(R)$ is finitely generated. We will prove that P is, in fact, finitely generated.

By Theorem 7.1, there exists a finitely generated, countably presented, flat module M' such that $P/PJ(R) \cong M'/M'J(R)$. Let $\alpha : P/PJ(R) \to M'/M'J(R)$ be an isomorphism. Let $\pi : P \to P/PJ(R)$ and $\pi' : M' \to M'/M'J(R)$ denote the canonical projections. Since P is projective, there exists $\beta : P \to M'$ such that the diagram

$$\begin{tikzcd} P \arrow[r, "\beta"] \arrow[d, "\pi"'] & M' \arrow[d, "\pi'"] \\ P/PJ(R) \arrow[r, "\alpha"] & M'/M'J(R) \end{tikzcd}$$

is commutative. Since $\ker \pi'$ is small in M', β is onto. As M' has projective dimension 1, $\ker \beta$ is projective. Applying $- \otimes_R R/J(R)$ to the exact sequence

$$0 \longrightarrow \ker \beta \to P \xrightarrow{\beta} M' \longrightarrow 0,$$

we obtain the exact sequence

$$0 \longrightarrow \ker \beta \otimes_R R/J(R) \to P/PJ(R) \xrightarrow{\alpha} M'/M'J(R) \longrightarrow 0.$$

Since α is an isomorphism, we have $0 = \ker \beta \otimes_R R/J(R) \cong \ker \beta/(\ker \beta)J(R)$. But $\ker \beta$ is projective, hence β is an isomorphism. This proves that P is a finitely generated projective module.

Let $\rho : M \to M/MJ(R)$ denote the canonical projection. Since P is projective, there exists $\gamma : P \to M$ such that the diagram

$$\begin{tikzcd} P \arrow[r, "\gamma"] \arrow[d, "\pi"'] & M \arrow[d, "\rho"] \\ P/PJ(R) \arrow[r, "\cong"] & M/MJ(R), \end{tikzcd}$$

is commutative. Since $\ker \rho$ is small in M, γ is onto. Since P is finitely generated and $\ker \gamma \subseteq PJ(R)$, $\ker \gamma$ is small in P. Hence $\gamma : P \to M$ is a projective cover. By Lemma 7.2, γ is an isomorphism. □

Thus if a finitely generated projective right $R/J(R)$-module P satisfies condition (i) of Theorem 7.1 (that is, $P \cong M/MJ(R)$ for some finitely generated flat right R-module M) and the right/left symmetric of condition (vi) of Theorem 7.1 (that is, $P \cong P'/P'J(R)$ for some projective right R-module P'), then $M \cong P'$ is a projective cover of P.

Acknowledgement

We are grateful to G. Puninski for his comments on a previous version of this paper.

References

[Anderson and Fuller 1992] F. W. Anderson and K. R. Fuller, *Rings and categories of modules*, 2nd ed., Graduate Texts in Mathematics **13**, Springer, New York, 1992. MR 94i:16001 Zbl 0765.16001

[Azumaya 1987] G. Azumaya, "Finite splitness and finite projectivity", *J. Algebra* **106**:1 (1987), 114–134. MR 89a:16034 Zbl 0607.16017

[Bass 1960] H. Bass, "Finitistic dimension and a homological generalization of semi-primary rings", *Trans. Amer. Math. Soc.* **95** (1960), 466–488. MR 28 #1212 Zbl 0094.02201

[Jøndrup 1976] S. Jøndrup, "Projective modules", *Proc. Amer. Math. Soc.* **59**:2 (1976), 217–221. MR 54 #7546 Zbl 0338.16008

[Lam 1999] T. Y. Lam, *Lectures on modules and rings*, Graduate Texts in Mathematics **189**, Springer, New York, 1999. MR 99i:16001 Zbl 0911.16001

[Lazard 1969] D. Lazard, "Autour de la platitude", *Bull. Soc. Math. France* **97** (1969), 81–128. MR 40 #7310 Zbl 0174.33301

[Lazard 1974] D. Lazard, "Liberté des gros modules projectifs", *J. Algebra* **31** (1974), 437–451. MR 50 #2250 Zbl 0291.13004

[Puninski 2004] G. Puninski, "Projective modules over the endomorphism ring of a biuniform module", *J. Pure Appl. Algebra* **188**:1-3 (2004), 227–246. MR 2004k:16003

[Sakhaev 1987] I. I. Sakhaev, "On the group $K_0(A\Pi)$ for semilocal rings", *Math. Nachr.* **130** (1987), 157–175. MR 88e:18010 Zbl 0617.13009

[Sakhaev 1993] I. I. Sakhaev, "Finite generability of projective modules over rings with polynomial identities", *Izv. Vyssh. Uchebn. Zaved. Mat.* 8 (1993), 65–75. In Russian; translated in *Russian Math.* (*Iz. VUZ*) **37** (1993), 64–74. MR 95k:16004

[Sakhaev 1996] I. I. Sakhaev, "Finite generation of projective modules over some rings", *Izv. Vyssh. Uchebn. Zaved. Mat.* 10 (1996), 63–75. In Russian; translated in *Russian Math.* (*Iz. VUZ*) **40** (1997), 60–72. MR 98e:16004

[Vasconcelos 1969] W. V. Vasconcelos, "On projective modules of finite rank", *Proc. Amer. Math. Soc.* **22** (1969), 430–433. MR 39 #4134 Zbl 0176.31601

[Zöschinger 1981] H. Zöschinger, "Projektive Moduln mit endlich erzeugtem Radikalfaktormodul", *Math. Ann.* **255**:2 (1981), 199–206. MR 82e:16013 Zbl 0439.16015

Received July 11, 2003. Revised July 13, 2004.

ALBERTO FACCHINI
DIPARTIMENTO DI MATEMATICA PURA ED APPLICATA
UNIVERSITÀ DI PADOVA
35131 PADOVA
ITALY

facchini@math.unipd.it
http://www.math.unipd.it/~facchini

DOLORS HERBERA
DEPARTAMENT DE MATEMÀTIQUES
UNIVERSITAT AUTÒNOMA DE BARCELONA
08193 BELLATERRA (BARCELONA)
SPAIN

dolors@mat.uab.es

ISKHAK SAKHAJEV
DEPARTMENT OF MECHANICS AND MATHEMATICS
KAZAN'S UNIVERSITY OF TATARSTAN
420008 KAZAN
RUSSIA

ishak.sahajev@ksu.ru

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

MAXIMAL TORI DETERMINING THE ALGEBRAIC GROUPS

SHRIPAD M. GARGE

Let k be a finite field, a global field, or a local non-archimedean field, and let H_1 and H_2 be split, connected, semisimple algebraic groups over k. We prove that if H_1 and H_2 share the same set of maximal k-tori, up to k-isomorphism, then the Weyl groups $W(H_1)$ and $W(H_2)$ are isomorphic, and hence the algebraic groups modulo their centers are isomorphic except for a switch of a certain number of factors of type B_n and C_n.

(Due to a recent result of Philippe Gille, this result also holds for fields which admit arbitrary cyclic extensions.)

1. Introduction

Let H be a connected, semisimple algebraic group over a field k. It is natural to ask to what extent the group H is determined by the k-isomorphism classes of maximal k-tori contained in it. We study this question over finite fields, global fields and local non-archimedean fields, and prove the following theorem.

Theorem 1.1 (Theorem 4.1). *Let k be a finite field, a global field or a local non-archimedean field, and let H_1 and H_2 be split, connected, semisimple algebraic groups over k. Suppose that for every maximal k-torus $T_1 \subset H_1$ there exists a maximal k-torus $T_2 \subset H_2$ such that the tori T_1 and T_2 are k-isomorphic, and vice versa. Then the Weyl groups $W(H_1)$ and $W(H_2)$ are isomorphic.*

Moreover, if we write $W(H_1)$ and $W(H_2)$ as a direct product of Weyl groups of simple algebraic groups, $W(H_1) = \prod_{\Lambda_1} W_{1,\alpha}$, and $W(H_2) = \prod_{\Lambda_2} W_{2,\beta}$, then there exists a bijection $i : \Lambda_1 \to \Lambda_2$ such that $W_{1,\alpha}$ is isomorphic to $W_{2,i(\alpha)}$ for every $\alpha \in \Lambda_1$.

Since a split simple algebraic group with trivial center is determined by its Weyl group, except for the groups of the type B_n and C_n, we have following theorem.

Theorem 1.2. *Let k be as in Theorem 1.1, and let H_1 and H_2 be split, connected, semisimple algebraic groups over k, with trivial center. Write H_1 and H_2 as direct products of simple groups: $H_1 = \prod_{\Lambda_1} H_{1,\alpha}$, and $H_2 = \prod_{\Lambda_2} H_{2,\beta}$. If H_1 and H_2 satisfy the condition given in Theorem 1.1, then there is a bijection $i : \Lambda_1 \to \Lambda_2$*

MSC2000: 20G15.
Keywords: maximal tori, algebraic groups.

such that $H_{1,\alpha}$ is isomorphic to $H_{2,i(\alpha)}$, except for the case where $H_{1,\alpha}$ is a simple group of type B_n or C_n, in which case $H_{2,i(\alpha)}$ could be of type C_n or B_n.

From the explicit description of maximal k-tori in $\mathrm{SO}(2n+1)$ and $\mathrm{Sp}(2n)$ (see for instance [Kariyama 1989, Proposition 2]) one finds that these groups contain the same set of k-isomorphism classes of maximal k-tori.

We show by an example that the existence of split tori in the groups H_1 and H_2 is necessary. Note that if k is $\mathbb{Q}_p$, then the Brauer group of k is $\mathbb{Q}/\mathbb{Z}$. Consider the central division algebras of degree five, D_1 and D_2, corresponding to $1/5$ and $2/5$ in $\mathbb{Q}/\mathbb{Z}$ respectively, and let

$$H_1 = \mathrm{SL}_1(D_1) \quad \text{and} \quad H_2 = \mathrm{SL}_1(D_2).$$

The maximal tori in H_1 and H_2 correspond, respectively, to the maximal commutative subfields in D_1 and D_2. But over $\mathbb{Q}_p$ every division algebra of a fixed degree contains every field extension of that degree (see [Pierce 1982, Proposition 17.10 and Corollary 13.3]), so H_1 and H_2 share the same set of maximal tori over k. But they are not isomorphic, since it is known that $\mathrm{SL}_1(D) \cong \mathrm{SL}_1(D')$ if and only if $D \cong D'$ or $D \cong (D')^{\mathrm{op}}$ [Knus et al. 1998, 26.11].

This paper is arranged as follows. The description of the k-conjugacy classes of maximal k-tori in an algebraic group H defined over k can be given in terms of the Galois cohomology of the normalizer in H of a fixed maximal torus. Similarly, the k-isomorphism classes of n-dimensional tori defined over k can be described in terms of n-dimensional integral representations of the Galois group of $\bar{k}$ (the algebraic closure of k) over k. Using these two descriptions, in Section 2 we obtain a Galois cohomological description for the k-isomorphism classes of maximal k-tori in H. Since we are dealing with groups that are split over the base field k, the Galois action on the Weyl groups is trivial. This enables us to prove, in Section 4, that if split, connected, semisimple algebraic groups H_1 and H_2 of rank n share the same set of maximal k-tori up to k-isomorphism, then the Weyl groups $W(H_1)$ and $W(H_2)$, considered as subgroups of $\mathrm{GL}_n(\mathbb{Z})$, share the same set of elements up to conjugacy in $\mathrm{GL}_n(\mathbb{Z})$.

This then is the main question to be answered: if the Weyl groups of two split, connected, semisimple algebraic groups, W_1 and W_2, embedded in $\mathrm{GL}_n(\mathbb{Z})$ in the natural way, i.e., by their action on the character group of a fixed split maximal torus, have the property that every element of W_1 is $\mathrm{GL}_n(\mathbb{Z})$-conjugate to one in W_2 and vice versa, are the Weyl groups isomorphic? Much of the work in this paper seeks to prove this statement by using elaborate information available about the conjugacy classes in Weyl groups of simple algebraic groups together with their standard representations in $\mathrm{GL}_n(\mathbb{Z})$. Our analysis finally depends on the knowledge of characteristic polynomials of elements in the Weyl groups considered as subgroups of $\mathrm{GL}_n(\mathbb{Z})$. This information is summarized in Section 3. Using it we prove the main theorem in Section 4.

We emphasize that if we were proving Theorems 1.1 and 1.2 for simple algebraic groups, our proofs would be relatively very simple. However, for semisimple groups, we have to make a somewhat complicated inductive argument on the maximal rank among the simple factors of the semisimple groups H_1 and H_2.

We use the term "simple Weyl group of rank r" for the Weyl group of a simple algebraic group of rank r. Any Weyl group is a product of simple Weyl groups in a unique way (up to permutation). We say that two Weyl groups are *isomorphic* if and only if the simple factors and their multiplicities are the same.

The question studied in this paper seems relevant for the study of Mumford–Tate groups over number fields. The author was informed, after the completion of the paper, that Theorem 1.1 over a finite field is implicit in the work of Larsen and Pink [1992]. We would like to mention that although much of the paper could be said to be implicitly contained in [Larsen and Pink 1992], the theorems we state (and prove) are not explicitly stated there, and our proofs are also different.

2. Galois cohomological lemmas

We begin by fixing notation. Let k denote an arbitrary field and let $G(\bar{k}/k)$ be the Galois group of $\bar{k}$ (the algebraic closure of k) over k. Let H denote a split, connected, semisimple algebraic group defined over k and let T_0 be a fixed split maximal torus in H, of dimension n. Let W be the Weyl group of H with respect to T_0. Then we have an exact sequence of algebraic groups defined over k,

$$0 \longrightarrow T_0 \longrightarrow N(T_0) \longrightarrow W \longrightarrow 1$$

where $N(T_0)$ denotes the normalizer of T_0 in H.

The above exact sequence gives us a map $\psi : H^1(k, N(T_0)) \to H^1(k, W)$. It is well known that a certain subset of $H^1(k, N(T_0))$ classifies k-conjugacy classes of maximal k-tori in H. For the sake of completeness, we formulate this as a lemma in the case of split, connected, semisimple groups.

Lemma 2.1. *Let H be a split, connected, semisimple algebraic group defined over a field k and let T_0 be a fixed split maximal torus in H. The natural embedding $N(T_0) \hookrightarrow H$ induces a map $\Psi : H^1(k, N(T_0)) \to H^1(k, H)$. The set of k-conjugacy classes of maximal tori in H are in one-one correspondence with the subset of $H^1(k, N(T_0))$ which is mapped to the neutral element in $H^1(k, H)$ by the map Ψ.*

Proof. Let T be a maximal k-torus in H and let L be a splitting field of T, that is, assume that the torus T splits as a product of $\mathbb{G}_m$s over L. We assume that the field L is Galois over k. By the uniqueness of maximal split tori up to conjugacy, there exists an element $a \in H(L)$ such that $aT_0a^{-1} = T$. Then for any $\sigma \in G(L/k)$, we have $\sigma(a)T_0\,\sigma(a)^{-1} = T$, as both T_0 and T are defined over k. This implies that

$$\big(a^{-1}\sigma(a)\big)T_0\big(a^{-1}\sigma(a)\big)^{-1} = T_0.$$

Therefore $a^{-1}\sigma(a) \in N(T_0)$. This enables us to define a map $G(L/k) \to N(T_0)$ which sends σ to $a^{-1}\sigma(a)$, and by composing this map with the natural map $G(\bar{k}/k) \to G(L/k)$, we get a map $\phi_a : G(\bar{k}/k) \to N(T_0)$. We check that

$$\phi_a(\sigma\tau) = \phi_a(\sigma)\sigma\big(\phi_a(\tau)\big)$$

for all $\sigma, \tau \in G(\bar{k}/k)$, and hence that ϕ_a is a 1-cocycle. If $b \in H(L)$ is another element such that $bT_0\,b^{-1} = T$, we see that

$$\phi_a(\sigma) = (b^{-1}a)^{-1}\phi_b(\sigma)\sigma(b^{-1}a).$$

Thus the element $[\phi_a] \in H^1(k, N(T_0))$ is determined by the maximal torus T, and so we denote it by $\phi(T)$. It is clear that $\phi(T)$ is determined by the k-conjugacy class of T. Moreover, if $\phi(T) = \phi(S)$ for two maximal tori T and S in H, then one can check that these two tori are conjugate over k. Indeed, if $T = aT_0\,a^{-1}$ and $S = bT_0\,b^{-1}$ for $a, b \in H(\bar{k})$ then for any $\sigma \in G(\bar{k}/k)$,

$$a^{-1}\sigma(a) = c^{-1}\big(b^{-1}\sigma(b)\big)\sigma(c)$$

for some $c \in N(T_0)$. Then $\sigma(bca^{-1}) = bca^{-1}$ for all $\sigma \in G(\bar{k}/k)$, and hence $bca^{-1} \in H(k)$ and $(bca^{-1})T(bca^{-1})^{-1} = S$. Further, it is clear that the image of ϕ in $H^1(k, N(T_0))$ is mapped to the neutral element in $H^1(k, H)$ by Ψ.

Moreover, if $\phi_1 : G(\bar{k}/k) \to N(T_0)$ is a 1-cocycle such that $\Psi(\phi_1)$ is neutral in $H^1(k, H)$, then $\phi_1(\sigma) = a^{-1}\sigma(a)$ for some $a \in H(\bar{k})$. Then the cohomology class $[\phi_1] \in H^1(k, N(T_0))$ corresponds to the maximal torus $S_1 = aT_0a^{-1}$ in H. Since $a^{-1}\sigma(a) = \phi_1(\sigma) \in N(T_0)$, the torus S_1 is invariant under the Galois action, and so we conclude that it is defined over k. Thus the image of ϕ is the inverse image of the neutral element in $H^1(k, H)$ under the map Ψ. This is the complete description of the k-conjugacy classes of maximal k-tori in the group H.

Finally, we observe that the detailed proof we have given above amounts to looking at the exact sequence $1 \to N(T_0) \to H \to H/N(T_0) \to 1$ which gives an exact sequence

$$H/N(T_0)(k) \longrightarrow H^1(k, N(T_0)) \longrightarrow H^1(k, H)$$

of pointed sets. Therefore $H/N(T_0)(k)$, which is the variety of conjugacy classes of k-tori in H, is identified with the elements of $H^1(k, N(T_0))$ which become trivial in $H^1(k, H)$. □

We also recall the correspondence between k-isomorphism classes of n-dimensional k-tori and equivalence classes of n-dimensional integral representations of

$G(\bar{k}/k)$. Let $T_0 = \mathbb{G}_m^n$ be the split torus of dimension n, let T_1 be an n-dimensional torus defined over k, and let L_1 denote the splitting field of T_1. Since the torus T_1 is split over L_1, we have an L_1-isomorphism $f : T_0 \to T_1$. The Galois action on T_0 and T_1 gives us another isomorphism, $f^\sigma := \sigma f \sigma^{-1} : T_0 \to T_1$. Again one sees that the map $\varphi_f : G(\bar{k}/k) \to \mathrm{Aut}_{L_1}(T_0)$, given by $\sigma \mapsto f^{-1} f^\sigma$, is a 1-cocycle. Since the torus T_0 is already split over k, we have $\mathrm{Aut}_{L_1}(T_0) \cong \mathrm{Aut}_k(T_0)$, and hence the Galois group $G(\bar{k}/k)$ acts trivially on $\mathrm{Aut}_{L_1}(T_0)$, which is isomorphic to $\mathrm{GL}_n(\mathbb{Z})$. Therefore, φ_f is actually a homomorphism from the Galois group $G(\bar{k}/k)$ to $\mathrm{GL}_n(\mathbb{Z})$. This homomorphism gives an n-dimensional integral representation of the absolute Galois group, $G(\bar{k}/k)$. By changing the isomorphism f to any other L_1-isomorphism from T_0 to T_1, we get a conjugate of φ_f. Thus the element $[\varphi_f]$ in $H^1(k, \mathrm{GL}_n(\mathbb{Z}))$ is determined by T_1 and we denote it by $\varphi(T_1)$. Thus a k-isomorphism class of an n-dimensional torus gives us an equivalence class of n-dimensional integral representations of the Galois group, $G(\bar{k}/k)$. This correspondence is known to be bijective [Platonov and Rapinchuk 1994, 2.2].

Since the group H that we consider here is split over the base field k, the Weyl group W of H is defined over k, and $W(\bar{k}) = W(k)$. Therefore $G(\bar{k}/k)$ acts trivially on W, and hence $H^1(k, W)$ is the set of conjugacy classes of elements in $\mathrm{Hom}(G(\bar{k}/k), W)$. Since W acts faithfully on the character group of T_0, we can consider $W \hookrightarrow \mathrm{GL}_n(\mathbb{Z})$ and thus each element of $H^1(k, W)$ gives us an integral representation of the absolute Galois group. For a maximal torus T in H, we already have an n-dimensional integral representation of $G(\bar{k}/k)$, as described above. We prove that this representation is equivalent to a Galois representation given by an element of $H^1(k, W)$.

Lemma 2.2. *Let H be a split, connected, semisimple algebraic group defined over k. Fix a maximal split k-torus T_0 in H. Let T be a maximal k-torus in H, let $\phi(T) \in H^1(k, N(T_0))$ be the cohomology class corresponding to the k-conjugacy class of T in H, and let $\varphi(T) \in H^1(k, \mathrm{GL}_n(\mathbb{Z}))$ be the cohomology class corresponding to the k-isomorphism class of T. Then the integral representations given by $\varphi(T)$ and $i \circ \psi \circ \phi(T)$ are equivalent, where $\psi : H^1(k, N(T_0)) \to H^1(k, W)$ is induced by the natural map from $N(T_0)$ to W, and i is the natural map from $H^1(k, W)$ to $H^1(k, \mathrm{GL}_n(\mathbb{Z}))$.*

Proof. Let L be a splitting field of T, then an element $a \in H(L)$ such that $aT_0 a^{-1} = T$ enables us to define a 1-cocycle $\phi_a : G(\bar{k}/k) \to N(T_0)$ given by $\phi_a(\sigma) = a^{-1}\sigma(a)$. The element $\phi(T) \in H^1(k, N(T_0))$ is precisely the class $[\phi_a]$.

Further, we treat conjugation by a as an L-isomorphism $f : T_0 \to T$, and then it can be checked that the map $f^\sigma := \sigma f \sigma^{-1}$ is precisely conjugation by $\sigma(a)$. The element $\varphi(T) \in H^1(k, \mathrm{GL}_n(\mathbb{Z}))$ is equal to $[\varphi_f]$, where $\varphi_f(\sigma) = f^{-1} f^\sigma$. Now, the map $\psi : N(T_0) \to W$ is the natural map taking an element $\alpha \in N(T_0)$ to

$\bar{\alpha} := \alpha \cdot T_0 \in W = N(T_0)/T_0$. Hence we have

$$\psi\big(\phi_a(\sigma)\big) = \overline{a^{-1}\sigma(a)} = f^{-1}f^{\sigma} = \varphi_f(\sigma).$$

Since the action of W on T_0 is given by conjugation, it is clear that the integral representation of the Galois group $G(\bar{k}/k)$, given by $\psi(\phi(T))$, is equivalent to the one given by $\varphi(T)$. □

Thus, a k-isomorphism class of a maximal torus in H gives an element in $H^1(k, W)$. We note here that not every subgroup of the Weyl group W may appear as a Galois group of some finite extension K/k. For instance, if k is a local field of characteristic zero it is known that the Galois group of any finite extension over k is a solvable group [Serre 1979, IV].

If we assume that the base field k is either a finite field or a local non-archimedean field, we have the following result.

Lemma 2.3. *Let k be a finite field or a local non-archimedean field and let H be a split, connected, semisimple algebraic group defined over k. Fix a split maximal torus T_0 in H and let W denote the Weyl group of H with respect to T_0. An element in $H^1(k, W)$ which corresponds to a homomorphism $\rho : G(\bar{k}/k) \to W$ with cyclic image, corresponds to a k-isomorphism class of a maximal torus in H under the mapping $\psi : H^1(k, N(T_0)) \to H^1(k, W)$.*

Proof. Consider the map $\Psi : H^1(k, N(T_0)) \to H^1(k, H)$ induced by the inclusion $N(T_0) \hookrightarrow H$. If we denote the neutral element in $H^1(k, H)$ by ι, then by Lemma 2.1 the set

$$X := \big\{ f \in H^1(k, N(T_0)) : \Psi(f) = \iota \big\}$$

is in one-one correspondence with the k-conjugacy classes of maximal k-tori in H. By Lemma 2.2, it is enough to show that $[\rho] \in \psi(X)$, where $\psi : H^1(k, N(T_0)) \to H^1(k, W)$ is induced by the natural map from $N(T_0)$ to W.

By Tits' theorem [1966, 4.6], there exists a subgroup $\overline{W}$ of $N(T_0)(k)$ such that the sequence

$$0 \longrightarrow \mu_2^n \longrightarrow \overline{W} \longrightarrow W \longrightarrow 1$$

is exact. Let N denote the image of ρ in W. We know that N is a cyclic subgroup of W. Let w be a generator of N and $\overline{w}$ be a lifting of w to $\overline{W}$. Since the base field k admits cyclic extensions of any given degree, there exists a map ρ_1 from $G(\bar{k}/k)$ to $\overline{W}$ whose image is the cyclic subgroup generated by $\overline{w}$. Since the Galois action on $\overline{W}$ is trivial, as $\overline{W}$ is a subgroup of $N(T_0)(k)$, the map ρ_1 could be treated as a 1-cocycle from $G(\bar{k}/k)$ to $N(T_0)$. Consider $[\rho_1]$ as an element in $H^1(k, N(T_0))$, then $\psi[\rho_1] = [\rho] \in H^1(k, W)$. We now consider two cases.

Case 1: k is a finite field.

By Lang's Theorem [1956, Corollary to Theorem 1], $H^1(k, H)$ is trivial and so the set X coincides with $H^1(k, N(T_0))$. Therefore the element $[\rho_1] \in H^1(k, N(T_0))$ corresponds to a k-conjugacy class of maximal k-torus in H. Then, by Lemma 2.2, $[\rho] = \psi[\rho_1]$ corresponds to a k-isomorphism class of maximal k-tori in H.

Case 2: k is a local non-archimedean field.

By [Platonov and Rapinchuk 1994, Proposition 2.10] there exists a semisimple, simply connected algebraic group $\widetilde{H}$, which is defined over k, together with a k-isogeny $\pi : \widetilde{H} \to H$. We have already fixed a split maximal torus T_0 in H; let $\widetilde{T_0}$ be the split maximal torus in $\widetilde{H}$ which gets mapped to T_0 by the covering map π. It can be seen that by restriction we get a surjective map $\pi : N(\widetilde{T_0}) \to N(T_0)$, where the normalizers are taken in appropriate groups. Moreover, the induced map $\pi_1 : \widetilde{W} \to W$ is an isomorphism.

We define the maps

$$\tilde{\psi} : H^1(k, N(\widetilde{T_0})) \to H^1(k, \widetilde{W}) \quad \text{and} \quad \tilde{\Psi} : H^1(k, N(\widetilde{T_0})) \to H^1(k, \widetilde{H})$$

in the same way as the maps ψ and Ψ are defined for the group H.

Consider the following diagram, where the horizontal arrows represent natural maps.

$$\begin{tikzcd} \widetilde{H} \arrow[d, "\pi"'] & N(\widetilde{T_0}) \arrow[l] \arrow[r] \arrow[d, "\pi"'] & \widetilde{W} \arrow[d, "\pi_1"] \\ H & N(T_0) \arrow[l] \arrow[r] & W, \end{tikzcd}$$

It is clear that this diagram is commutative and hence so is the following one.

$$\begin{tikzcd} H^1(k, \widetilde{H}) \arrow[d, "\pi^*"'] & H^1(k, N(\widetilde{T_0})) \arrow[l, "\tilde{\Psi}"'] \arrow[r, "\tilde{\psi}"] \arrow[d, "\pi^*"'] & H^1(k, \widetilde{W}) \arrow[d, "\pi_1^*"] \\ H^1(k, H) & H^1(k, N(T_0)) \arrow[l, "\Psi"] \arrow[r, "\psi"'] & H^1(k, W). \end{tikzcd}$$

Since π_1 is an isomorphism, the map π_1^* is a bijection. Now consider an element $[\rho] \in H^1(k, W)$ such that the image of the 1-cocycle ρ is a cyclic subgroup of W, and let $[\tilde{\rho}]$ be its inverse image in $H^1(k, \widetilde{W})$ under the bijection π_1^*. Using Tits' theorem [1966] as above, we lift $[\tilde{\rho}]$ to an element $[\tilde{\rho}_1]$ in $H^1(k, N(\widetilde{T_0}))$. Since $\widetilde{H}$ is simply connected and k is a non-archimedean local field, $H^1(k, \widetilde{H})$ is trivial [Bruhat and Tits 1967; Kneser 1965a, 1965b]. Therefore, $\tilde{\Psi}[\tilde{\rho}_1]$ is neutral in $H^1(k, \widetilde{H})$ and so is $\pi^*(\tilde{\Psi}[\tilde{\rho}_1])$ in $H^1(k, H)$. By commutativity of the diagram, we have that the element $[\rho] \in H^1(k, W)$ has a lift $\pi^*[\tilde{\rho}_1]$ in $H^1(k, N(T_0))$ such that $\Psi(\pi^*[\tilde{\rho}_1])$ is neutral in $H^1(k, H)$. Thus the element $[\rho]$ corresponds to a k-isomorphism class of a maximal torus in H. $\square$

3. Characteristic polynomials

For a finite subgroup W of $\mathrm{GL}_n(\mathbb{Z})$, we define $\mathrm{ch}(W)$ to be the set of characteristic polynomials of elements of W, and $\mathrm{ch}^*(W)$ to be the set of irreducible factors of elements of $\mathrm{ch}(W)$. Since all the elements of W are of finite order, the irreducible factors (over $\mathbb{Q}$) of the characteristic polynomials are cyclotomic polynomials. We denote by ϕ_r the r-th cyclotomic polynomial, that is, the irreducible monic polynomial over $\mathbb{Z}$ satisfied by a primitive r-th root of unity. We define

$$\mathfrak{m}_i(W) = \max\left\{t : \phi_i^t \text{ divides } f \text{ for some } f \in \mathrm{ch}(W)\right\}$$

and

$$\mathfrak{m}'_i(W) = \min\left\{t : \phi_2^t \cdot \phi_i^{\mathfrak{m}_i(W)} \text{ divides } f \text{ for some } f \in \mathrm{ch}(W)\right\}.$$

For positive integers $i \neq j$, we define

$$\mathfrak{m}_{i,j}(W) = \max\left\{t+s : \phi_i^t \cdot \phi_j^s \text{ divides } f \text{ for some } f \in \mathrm{ch}(W)\right\}.$$

If U_1 is a subgroup of $\mathrm{GL}_n(\mathbb{Z})$ and U_2 is a subgroup of $\mathrm{GL}_m(\mathbb{Z})$, then $U_1 \times U_2$ can be treated as a subgroup of $\mathrm{GL}_{m+n}(\mathbb{Z})$. Then

$$\mathrm{ch}(U_1 \times U_2) = \left\{f_1 \cdot f_2 : f_1 \in \mathrm{ch}(U_1),\ f_2 \in \mathrm{ch}(U_2)\right\}.$$

Moreover, one can easily check that

$$\mathfrak{m}_i(U_1 \times U_2) = \mathfrak{m}_i(U_1) + \mathfrak{m}_i(U_2),$$
$$\mathfrak{m}'_i(U_1 \times U_2) = \mathfrak{m}'_i(U_1) + \mathfrak{m}'_i(U_2)$$

for all i, and

$$\mathfrak{m}_{i,j}(U_1 \times U_2) = \mathfrak{m}_{i,j}(U_1) + \mathfrak{m}_{i,j}(U_2)$$

for all i, j. A simple Weyl group W of rank n has a natural embedding in $\mathrm{GL}_n(\mathbb{Z})$. We obtain a description of the sets $\mathrm{ch}^*(W)$ with respect to this natural embedding. Here we use the following result due to T. A. Springer [1974, Theorem 3.4(i)] about the fundamental degrees of the Weyl group W. We recall that the degrees of the generators of the invariant algebra of the Weyl group are called as the fundamental degrees of the Weyl group.

Theorem 3.1 (Springer). *Let W be a complex reflection group with fundamental degrees $d_1, d_2, \dots, d_m$. An r-th root of unity occurs as an eigenvalue for some element of W if and only if r divides one of the fundamental degrees d_i of W.*

Equivalently, the irreducible polynomial ϕ_r is in $\mathrm{ch}^*(W)$ *if and only if r divides one of the fundamental degrees d_i of the reflection group W.*

Table 3.2 lists the fundamental degrees and the divisors of degrees for the simple Weyl groups (see [Humphreys 1990, 3.7]).

Type	Degrees	Divisors of degrees	
A_n	$2, 3, \ldots, n+1$	$1, 2, \ldots, n+1$	
B_n	$2, 4, \ldots, 2n$	$1, 2, \ldots, n, n+2, n+4, \ldots, 2n$	for n even
		$1, 2, \ldots, n, n+1, n+3, \ldots, 2n$	for n odd
D_n	$2, 4, \ldots, 2n-2, n$	$1, 2, \ldots, n, n+2, n+4, \ldots, 2n-2$	for n even
		$1, 2, \ldots, n, n+1, n+3, \ldots, 2n-2$	for n odd
G_2	$2, 6$	$1, 2, 3, 6$	
F_4	$2, 6, 8, 12$	$1, 2, 3, 4, 6, 8, 12$	
E_6	$2, 5, 6, 8, 9, 12$	$1, 2, 3, 4, 5, 6, 8, 9, 12$	
E_7	$2, 6, 8, 10, 12, 14, 18$	$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 18$	
E_8	$2, 8, 12, 14, 18, 20, 24, 30$	$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 18, 20, 24, 30$	

Table 3.2. Fundamental degrees and divisors of the simple Weyl groups

Using Theorem 3.1 and Table 3.2, we can now easily compute the set $\mathrm{ch}^*(W)$ for any simple Weyl group W. We summarize them below.

$$\begin{aligned}
&\mathrm{ch}^*(W(A_n)) = \{\phi_1, \phi_2, \ldots, \phi_{n+1}\}\\
&\mathrm{ch}^*(W(B_n)) = \{\phi_i, \phi_{2i} : i = 1, 2, \ldots, n\}\\
&\mathrm{ch}^*(W(D_n)) = \{\phi_i, \phi_{2j} : i = 1, 2, \ldots, n,\ \ j = 1, 2 \ldots, n-1\}\\
&\mathrm{ch}^*(W(G_2)) = \{\phi_1, \phi_2, \phi_3, \phi_6\}\\
&\mathrm{ch}^*(W(F_4)) = \{\phi_1, \phi_2, \phi_3, \phi_4, \phi_6, \phi_8, \phi_{12}\}\\
&\mathrm{ch}^*(W(E_6)) = \{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_8, \phi_9, \phi_{12}\}\\
&\mathrm{ch}^*(W(E_7)) = \{\phi_1, \phi_2, \ldots, \phi_{10}, \phi_{12}, \phi_{14}, \phi_{18}\}\\
&\mathrm{ch}^*(W(E_8)) = \{\phi_1, \phi_2, \ldots, \phi_{10}, \phi_{12}, \phi_{14}, \phi_{15}, \phi_{18}, \phi_{20}, \phi_{24}, \phi_{30}\}
\end{aligned}$$

4. Main result

In this section, k is either a finite field, a global field or a non-archimedean local field. We now restate the main result, Theorem 1.1.

Theorem 4.1. *Let H_1 and H_2 be split, connected, semisimple algebraic groups defined over k. Suppose that for every maximal k-torus $T_1 \subset H_1$ there exists a maximal k-torus $T_2 \subset H_2$ such that the torus T_2 is k-isomorphic to the torus T_1 and vice versa. Then, the Weyl groups $W(H_1)$ and $W(H_2)$ are isomorphic.*

Moreover, if we write $W(H_1)$ and $W(H_2)$ as a direct product of Weyl groups of simple algebraic groups, $W(H_1) = \prod_{\Lambda_1} W_{1,\alpha}$, and $W(H_2) = \prod_{\Lambda_2} W_{2,\beta}$, then there exists a bijection $i : \Lambda_1 \to \Lambda_2$ such that $W_{1,\alpha}$ is isomorphic to $W_{2,i(\alpha)}$ for every $\alpha \in \Lambda_1$.

The proof of this theorem occupies the rest of this section. Clearly the groups H_1 and H_2 are of the same rank, say n. Let W_1 and W_2 denote the Weyl groups of H_1 and H_2, respectively. We always treat W_1 and W_2 as subgroups of $\mathrm{GL}_n(\mathbb{Z})$. We first prove a lemma which transforms the information about k-isomorphism of maximal k-tori in the groups H_1 and H_2 into some information about the conjugacy classes of the elements of the corresponding Weyl groups W_1 and W_2.

Lemma 4.2. *Under the hypotheses of Theorem 4.1, for every element $w_1 \in W_1$, there exists an element $w_2 \in W_2$ such that w_2 is conjugate to w_1 in $\mathrm{GL}_n(\mathbb{Z})$ and vice versa.*

Proof. Let $w_1 \in W_1$ and let N_1 denote the subgroup of W_1 generated by w_1. Since the base field k admits any cyclic group as a Galois group, there is a map $\rho_1 : G(\bar{k}/k) \to W_1$ such that $\rho_1(G(\bar{k}/k)) = N_1$.

We first consider the case where k is a finite field or a local non-archimedean field. By Lemma 2.3, the element $[\rho_1] \in H^1(k, W_1)$ corresponds to a maximal k-torus in H_1, say T_1. By the hypothesis, there exists a torus $T_2 \subset H_2$ which is k-isomorphic to T_1. We know by Lemma 2.2 that there exists an integral Galois representation $\rho_2 : G(\bar{k}/k) \to \mathrm{GL}_n(\mathbb{Z})$ corresponding to the k-isomorphism class of T_2 which factors through W_2. Let $N_2 := \rho_2(G(\bar{k}/k)) \subseteq W_2$. Since T_1 and T_2 are k-isomorphic tori, the corresponding Galois representations, ρ_1 and ρ_2, are equivalent. This implies that there exists $g \in \mathrm{GL}_n(\mathbb{Z})$ such that $N_2 = gN_1g^{-1}$. Then $w_2 := gw_1g^{-1} \in N_2 \subseteq W_2$ is a conjugate of w_1 in $\mathrm{GL}_n(\mathbb{Z})$. We can start with an element $w_2 \in W_2$ and obtain its $\mathrm{GL}_n(\mathbb{Z})$-conjugate in W_1 in the same way.

Now we consider the case when k is a global field. Let v be a non-archimedean valuation of k and let k_v be the completion of k with respect to v. Clearly the groups H_1 and H_2 are defined over k_v. Let $T_{1,v}$ be a maximal k_v-torus in H_1. Then by Grothendieck's theorem [Borel and Springer 1968, 7.9, 7.11] and the weak approximation property [Platonov and Rapinchuk 1994, Proposition 7.3], there exists a k-torus in H, say T_1, such that $T_{1,v}$ is obtained from T_1 by the base change. By hypothesis, we have a k-torus T_2 in H_2 which is k-isomorphic to T_1. Then the torus $T_{2,v}$, obtained from T_2 by the base change, is k_v-isomorphic to $T_{1,v}$. Thus, every maximal k_v-torus in H_1 has a k_v-isomorphic torus in H_2. Similarly, we can show that every maximal k_v-torus in H_2 has a k_v-isomorphic torus in H_1. Then the proof follows by the previous case. □

Corollary 4.3. *Under the hypotheses of Theorem 4.1,* $\mathrm{ch}(W_1) = \mathrm{ch}(W_2)$ *and* $\mathrm{ch}^*(W_1) = \mathrm{ch}^*(W_2)$. *In particular,* $\mathfrak{m}_i(W_1) = \mathfrak{m}_i(W_2)$, $\mathfrak{m}'_i(W_1) = \mathfrak{m}'_i(W_2)$ *and* $\mathfrak{m}_{i,j}(W_1) = \mathfrak{m}_{i,j}(W_2)$ *for all i, j.*

Proof. Since the Weyl groups W_1 and W_2 share the same set of elements up to conjugacy in $\mathrm{GL}_n(\mathbb{Z})$, the sets $\mathrm{ch}(W_1)$ and $\mathrm{ch}(W_2)$ are the same, and hence the sets $\mathrm{ch}^*(W_1)$ and $\mathrm{ch}^*(W_2)$ are also the same. Further, for a fixed integer i, $\phi_i^{\mathfrak{m}_i(W_1)}$ divides an element $f_1 \in \mathrm{ch}(W_1)$. But since $\mathrm{ch}(W_1) = \mathrm{ch}(W_2)$, the polynomial $\phi_i^{\mathfrak{m}_i(W_1)}$ also divides an element $f_2 \in \mathrm{ch}(W_2)$. Therefore $\mathfrak{m}_i(W_1) \leq \mathfrak{m}_i(W_2)$. We obtain the inequality in the other direction in the same way and hence $\mathfrak{m}_i(W_1) = \mathfrak{m}_i(W_2)$. Similarly, we can prove that $\mathfrak{m}'_i(W_1) = \mathfrak{m}'_i(W_2)$, and also that, for integers $i \neq j$, the sets

$$\big\{(t_1, s_1) : \phi_i^{t_1} \cdot \phi_j^{s_1} \text{ divides some element } f_1 \in \mathrm{ch}(W_1)\big\},$$
$$\big\{(t_2, s_2) : \phi_i^{t_2} \cdot \phi_j^{s_2} \text{ divides some element } f_2 \in \mathrm{ch}(W_2)\big\}$$

are the same for $i = 1, 2$. It follows that $\mathfrak{m}_{i,j}(W_1) = \mathfrak{m}_{i,j}(W_2)$. □

We now prove the following result before going on to prove the main theorem.

Theorem 4.4. *Let H_1 and H_2 be split, connected, semisimple algebraic groups of rank n. Suppose that $\mathfrak{m}_i(W(H_1)) = \mathfrak{m}_i(W(H_2))$, that $\mathfrak{m}'_i(W(H_1)) = \mathfrak{m}'_i(W(H_2))$, and that $\mathfrak{m}_{i,j}(W(H_1)) = \mathfrak{m}_{i,j}(W(H_2))$ for all i, j. Let m be the maximum possible rank among the simple factors of H_1 and H_2. Let W'_1 and W'_2 denote the product of the Weyl groups of rank m simple factors of, respectively, H_1 and H_2. Then the groups W'_1 and W'_2 are isomorphic.*

Proof. We denote $W(H_1)$ by W_1 and $W(H_2)$ by W_2. We prove that if a simple Weyl group of rank m appears as a factor of W_1 with multiplicity p, then it appears as a factor of W_2, with the same multiplicity. We prove this lemma case by case, depending on the type of rank m simple factors of H_1 and H_2.

We prove this result by comparing the sets $\mathrm{ch}^*(W)$ for the simple Weyl groups of rank m. We observe from Table 3.2 that the maximal degree of the simple Weyl group of exceptional type, if any, is the largest among the maximal degrees of simple Weyl groups of rank m. The next largest maximal degree is that of $W(B_m)$, the next one is that of $W(D_m)$, and finally the Weyl group $W(A_m)$ has the smallest maximal degree. We use the relation between the elements of $\mathrm{ch}^*(W)$ and the degrees of the Weyl group W, given by Theorem 3.1. So, we begin the proof of the lemma with the case of exceptional groups of rank m, prove that it occurs with the same multiplicity for $i = 1, 2$. Then we prove the lemma for B_m, then for D_m and finally we prove the lemma for the group A_m.

Case 1: One of H_1 or H_2 contains a simple exceptional factor of rank m.

We first treat the case of the simple group E_8, that is, we assume that 8 is the maximum possible rank of the simple factors of the groups H_1 and H_2. We know that $\mathfrak{m}_{30}(W(E_8)) = 1$. Observe that ϕ_{30} is an irreducible polynomial of degree 8, and hence cannot occur in $\mathrm{ch}^*(W)$ for any simple Weyl group of rank at most 7. Moreover, from Theorem 3.1 and Table 3.2, it is clear that $\mathfrak{m}_{30}(W(A_8)) = \mathfrak{m}_{30}(W(B_8)) = \mathfrak{m}_{30}(W(D_8)) = 0$. Hence the multiplicity of E_8 in H_i is given by $\mathfrak{m}_{30}(W_i)$ which is the same for $i = 1, 2$.

Similarly for the simple algebraic group E_7, observe that $\mathfrak{m}_{18}(W(E_7)) = 1$ and $\mathfrak{m}_{18}(W) = 0$ for any simple Weyl group W of rank at most 7. Then the multiplicity of E_7 in H_i is given by $\mathfrak{m}_{18}(W_i)$ which is the same for $i = 1, 2$.

The case of E_6 is done by using $\mathfrak{m}_9$, since it is clear that $\mathfrak{m}_9(W) = 0$ for any simple Weyl group W of rank at most 6.

The cases of F_4 and G_2 are done similarly by using $\mathfrak{m}_{12}$ and $\mathfrak{m}_6$ respectively.

Case 2: One of H_1 or H_2 has B_m or C_m as a factor.

Since $W(B_m) \cong W(C_m)$, we treat the case of B_m only. By case 1, we can assume that the exceptional group of rank m, if any, occurs with the same multiplicities in both H_1 and H_2, and hence while counting the multiplicities $\mathfrak{m}_i$, $\mathfrak{m}'_i$ and $\mathfrak{m}_{i,j}$, we can (and will) ignore the exceptional groups of rank m.

Observe that $\mathfrak{m}_{2m}(W(B_m)) = 1$ and $\mathfrak{m}_{2m}(W) = 0$ for any other simple Weyl group W of classical type of rank at most m. However, it is possible that $\mathfrak{m}_{2m}(W) \neq 0$ for a simple Weyl group W of exceptional type of rank less than m. If $m \geq 16$ then this problem does not arise, therefore the multiplicity of B_m in H_i for $m \geq 16$ is given by $\mathfrak{m}_{2m}(W_i)$, which is the same for $i = 1, 2$. We do the cases of B_m for $m \leq 15$ separately.

For the group B_2, we observe that $\mathfrak{m}_4(W(B_2)) = 1$ and $\mathfrak{m}_4(W) = 0$ for any other simple Weyl group W of rank at most 2. Thus, the case of B_2 is done using $\mathfrak{m}_4(W_1) = \mathfrak{m}_4(W_2)$.

For the group B_3, we have $\mathfrak{m}_6(W(B_3)) = 1$, but then $\mathfrak{m}_6(W(G_2))$ is also 1. Observe that $\mathfrak{m}_4(W(B_3)) = 1$ and $\mathfrak{m}_4(W(G_2)) = 0$. We do this case by looking at the multiplicities of ϕ_4 and ϕ_6, so we do not worry about the simple Weyl groups W of rank at most 3 for which the multiplicities $\mathfrak{m}_4(W)$ and $\mathfrak{m}_6(W)$ are both zero. Now, let the multiplicities of B_3, G_2 and B_2 in the group H_i be, respectively, p_i, q_i and r_i, for $i = 1, 2$. Then, using $\mathfrak{m}_6(W_1) = \mathfrak{m}_6(W_2)$, we see that $p_1 + q_1 = p_2 + q_2$. Using $\mathfrak{m}_4$ we have $p_1 + r_1 = p_2 + r_2$ and using $\mathfrak{m}_{4,6}$ we see $p_1 + q_1 + r_1 = p_2 + q_2 + r_2$. Combining these equalities, we see that $p_1 = p_2$, that is, the group B_3 appears in both the groups H_1 and H_2 with the same multiplicity.

For the group B_4, we observe that $\mathfrak{m}_8(W(B_4)) = 1$. Since ϕ_8 has degree 4, it cannot occur in $\mathrm{ch}(W)$ for any simple Weyl group of rank at most 3 and $\mathfrak{m}_8(W(A_4)) = \mathfrak{m}_8(W(D_4)) = 0$. Since we are assuming by case 1 that the group F_4 occurs in both H_1 and H_2 with the same multiplicity, we are done in this case also.

For the group B_5, we have $\mathfrak{m}_{10}(W(B_5)) = 1$ and $\mathfrak{m}_{10}(W) = 0$ for any other simple Weyl group of classical type of rank at most 5. Since 5 does not divide the order of $W(G_2)$ or $W(F_4)$, it follows that $\mathfrak{m}_{10}(W(G_2)) = \mathfrak{m}_{10}(W(F_4)) = 0$ and so we are done.

The group B_6 is another group where the exceptional groups give problems. We have $\mathfrak{m}_{12}(W(B_6)) = 1$, but $\mathfrak{m}_{12}(W(F_4))$ is also 1. Observe that $\mathfrak{m}_{10}(W(B_6)) = 1$, but $\mathfrak{m}_{10}(W(F_4)) = 0$. Now, let the multiplicities of B_6, D_6, B_5 and F_4 in H_i be, respectively, p_i, q_i, r_i and s_i. Then $p_1 + s_1 = \mathfrak{m}_{12}(W_1) = \mathfrak{m}_{12}(W_2) = p_2 + s_2$. Similarly, comparing $\mathfrak{m}_{10}$, we see that

$$p_1 + q_1 + r_1 = p_2 + q_2 + r_2.$$

Then, we compare $\mathfrak{m}_{10,12}$ of the groups W_1 and W_2, to see that

$$p_1 + q_1 + r_1 + s_1 = p_2 + q_2 + r_2 + s_2.$$

Combining this equality with the one obtained by $\mathfrak{m}_{10}$, we get that $s_1 = s_2$ and hence $p_1 = p_2$. Thus the group B_6 occurs in both H_1 and H_2 with the same multiplicity.

We have $\mathfrak{m}_{14}(W(E_6)) = 0$, therefore the group B_7 is characterized by ϕ_{14} and hence it occurs in both H_1 and H_2 with the same multiplicity.

For the group B_8, we have $\mathfrak{m}_{16}(W(B_8)) = 1$. Since ϕ_{16} has degree 8, it cannot occur in $\mathrm{ch}^*(W)$ for any of the Weyl groups of G_2, F_4, E_6 or E_7. Thus, the group B_8 is characterized by ϕ_{16} and hence it occurs in both H_1 and H_2 with the same multiplicity.

The group B_9 has the property that $\mathfrak{m}_{18}(W(B_9)) = 1$. But $\mathfrak{m}_{18}(W(E_7)) = \mathfrak{m}_{18}(W(E_8)) = 1$, and so we conclude that the multiplicity of E_8 is the same for both W_1 and W_2 using $\mathfrak{m}_{30}$. Then we compare the multiplicities $\mathfrak{m}_{18}, \mathfrak{m}_{16}$ and $\mathfrak{m}_{16,18}$ to prove that the group B_9 occurs in both the groups H_1 and H_2 with the same multiplicity.

Now we examine the case B_{10}. Here $\mathfrak{m}_{20}(W(B_{10})) = 1$. Observe that $\mathfrak{m}_{20}(W) = 0$ for any other simple Weyl group W of rank at most 10, except E_8. Then the multiplicity of B_{10} in H_i is $\mathfrak{m}_{20}(W_i) - \mathfrak{m}_{30}(W_i)$ and hence it is the same for $i = 1, 2$.

The same method also works for B_{12}, that is, the multiplicity of B_{12} in H_i is $\mathfrak{m}_{24}(W_i) - \mathfrak{m}_{30}(W_i)$.

The multiplicities of B_{11}, B_{13} and B_{14} in H_i are given by $\mathfrak{m}_{22}(W_i)$, $\mathfrak{m}_{26}(W_i)$ and $\mathfrak{m}_{28}(W_i)$ and hence they are the same for $i = 1, 2$.

For B_{15}, we have $\mathfrak{m}_{30}(W(B_{15})) = \mathfrak{m}_{30}(W(E_8)) = 1$, and $\mathfrak{m}_{30}(W) = 0$ for any other simple Weyl group W of rank at most 15. Observe also that $\mathfrak{m}_{28}(W(B_{15})) = \mathfrak{m}_{28}(W(B_{14})) = 1$, and $\mathfrak{m}_{28}(W) = 0$ for any other simple Weyl group W of rank at most 15. Then by comparing $\mathfrak{m}_{30}$, $\mathfrak{m}_{28}$ and $\mathfrak{m}_{28,30}$ we get the desired result that B_{15} occurs in both H_1 and H_2 with the same multiplicity.

Case 3: One of H_1 or H_2 has D_m as a factor.

For this case, we assume that the exceptional group of rank m, if any, and the group B_m occur in both H_1 and H_2 with the same multiplicities.

We observe that $2m - 2$ is the largest integer r such that $\phi_r \in \mathrm{ch}^*(W(D_m))$, but $\mathfrak{m}_{2m-2}(W(B_{m-1})) = 1$. Hence we always have to compare the group D_m with the group B_{m-1}.

Let us assume that $m \geq 17$, so that $\phi_{2m-2} \notin \mathrm{ch}^*(W)$ for any simple Weyl group of exceptional type of rank less than m.

We know that $\mathfrak{m}_{2m-2}(W(D_m)) = \mathfrak{m}_{2m-2}(W(B_{m-1})) = 1$ and that $\mathfrak{m}_{2m-2}(W) = 0$ for any other simple Weyl group W of classical type of rank at most m. Further, $(X+1)(X^{m-1}+1)$ is the only element in $\mathrm{ch}(W(D_m))$ which has ϕ_{2m-2} as a factor. Similarly $X^{m-1}+1$ is the only element in $\mathrm{ch}(W(B_{m-1}))$ which has ϕ_{2m-2} as a factor. Observe that $\mathfrak{m}'_{2m-2}(W(D_m)) = \mathfrak{m}'_{2m-2}(W(B_{m-1}))+1$ and $m'_{2m-2}(W) = 0$ for any other simple Weyl group W of rank at most m. Let p_i and q_i be, respectively, the multiplicities of the groups D_m and B_{m-1} in H_i, for $i = 1, 2$. Then by considering $\mathfrak{m}_{2m-2}$, we have $p_1 + q_1 = p_2 + q_2$. Further if m is even, then by considering $\mathfrak{m}'_{2m-2}$ we have $2p_1 + q_1 = 2p_2 + q_2$. These two equalities imply that $p_1 = p_2$. If m is odd then $\mathfrak{m}'_{2m-2}$ itself gives $p_1 = p_2$. Thus the group D_m appears in both H_1 and H_2 with the same multiplicity.

Now we consider the groups D_m, for $m \leq 16$.

For D_4, we have to consider the simple algebraic groups B_3 and G_2. Comparing the multiplicities $\mathfrak{m}_6$, $\mathfrak{m}_4$ and $\mathfrak{m}_{4,6}$ we see that G_2 occurs in both H_1 and H_2 with the same multiplicity, and then we proceed as above to prove that D_4 also occurs with the same multiplicity in both the groups H_1 and H_2

For the group D_5, we first prove that the multiplicity of F_4 is the same for both H_1 and H_2 using $\mathfrak{m}_{12}$ and then prove the required result by considering $\mathfrak{m}_5$, $\mathfrak{m}_8$ and $\mathfrak{m}_{5,8}$. While dealing with the case D_6, we observe that $\mathfrak{m}_{10}(W(G_2)) = \mathfrak{m}_{10}(W(F_4)) = 0$, and so this case follows by an argument similar to that for $m \geq 17$. The case D_7 is proved by considering $\mathfrak{m}_7$, $\mathfrak{m}_{12}$ and $\mathfrak{m}_{7,12}$. For D_8, we first prove that the group E_7 occurs in both H_1 and H_2 with the same multiplicity by considering $\mathfrak{m}_{18}$ and then proceed as above. For D_9, we prove that E_8 occurs in both H_1 and H_2 with the same multiplicity by considering $\mathfrak{m}_{30}$ and proceed as for $m \geq 17$. For D_{10}, we prove that E_8 appears in both H_1 and H_2 with the same multiplicity by considering $\mathfrak{m}_{30}$, and the same follows for E_7 by considering $\mathfrak{m}_{18}$, $\mathfrak{m}_{16}$ and $\mathfrak{m}_{16,18}$.

For the groups D_m, where $m \geq 11$, the only simple Weyl group W of exceptional type such that $\phi_{2m-2} \in \mathrm{ch}^*(W)$ is $W(E_8)$, but for D_m, with $m \leq 15$, we can assume that E_8 occurs in both H_1 and H_2 with the same multiplicity by considering $\mathfrak{m}_{30}$ and hence we are done. For the group D_{16}, we take care of E_8 by considering $\mathfrak{m}_{30}$, $\mathfrak{m}_{28}$ and $\mathfrak{m}_{28,30}$. Other arguments are similar to the case $m \geq 17$.

Case 4: One of H_1 or H_2 has A_m as a factor.

We now consider the case of simple algebraic group of type A_m. Here, as usual, we assume that all other simple algebraic groups of rank m occur with the same multiplicities in both H_1 and H_2.

If m is even, then $m+1$ is odd and hence $\mathfrak{m}_{m+1}(W) = 0$ for any simple Weyl group W of classical type of rank less than m. If $m \geq 30$, then we do not have to bother about the exceptional simple groups of rank less than m. If m is odd and $m \geq 31$, then ϕ_{m+1} occurs in $\mathrm{ch}^*(W(B_r))$ and $\mathrm{ch}^*(W(D_{r+1}))$ for $r \geq (m+1)/2$. Then we compare the multiplicities $\mathfrak{m}_m$, $\mathfrak{m}_{m+1}$ and $\mathfrak{m}_{m,m+1}$ and find that the group A_m occurs in H_1 and H_2 with the same multiplicity. We must therefore consider the cases $m \leq 29$ separately.

The cases A_1 and A_2 are easy since there are no exceptional groups of rank 1. For A_3 we use $\mathfrak{m}_3$, $\mathfrak{m}_4$ and $\mathfrak{m}_{3,4}$ to get the result, and the case A_4 follows similarly by using $\mathfrak{m}_5$. The group A_5 is more problematic, since neither $\mathfrak{m}_6(W(B_3))$ nor $\mathfrak{m}_6(W(G_2))$ nor $\mathfrak{m}_6(W(F_4))$ vanish, but this is solved by first proving that F_4 appears with the same multiplicity using $\mathfrak{m}_{12}$ and then using the multiplicities $\mathfrak{m}_5$, $\mathfrak{m}_6$ and $\mathfrak{m}_{5,6}$. The case A_6 is solved by using $\mathfrak{m}_7$, and for A_7 we use $\mathfrak{m}_7$, $\mathfrak{m}_8$ and $\mathfrak{m}_{7,8}$.

With A_8, we can first assume that the multiplicity of E_7 is the same for both H_1 and H_2 by using $\mathfrak{m}_{18}$, and then use $\mathfrak{m}_7$, $\mathfrak{m}_9$ and $\mathfrak{m}_{7,9}$ to get the result. For A_9 we can again get rid of E_7 and E_8 using the multiplicities $\mathfrak{m}_{18}$ and $\mathfrak{m}_{30}$. Then we are left with the groups B_5 and E_6, and so here we use $\mathfrak{m}_7$, $\mathfrak{m}_{10}$ and $\mathfrak{m}_{7,10}$ to get the result.

Further, we note that for even $m \in \{10, 12, 16, \dots, 28\}$, we have $\mathfrak{m}_{m+1}(W) = 0$ for any simple Weyl group of rank less than m. Thus, the multiplicities of the groups A_m in H_i, for even $m \in \{10, 12, 16, \dots, 28\}$, are characterized by considering $\mathfrak{m}_{m+1}(W_i)$ and are hence the same for $i = 1, 2$. The case A_{14} follows by using $\mathfrak{m}_{13}$, $\mathfrak{m}_{15}$ and $\mathfrak{m}_{13,15}$.

Thus, the only remaining cases are A_m where m is odd and $11 \leq m \leq 29$. We observe that for odd $m \in \{11, 13, 17, \dots, 29\}$, the only simple Weyl group W of rank less than m, with $\mathfrak{m}_m(W) \neq 0$, is A_{m-1}. Moreover, $\mathfrak{m}_{m+1}(W(A_{m-1})) = 0$, so the cases of the groups A_m, for odd $m \in \{11, 13, 17, \dots, 29\}$, are solved by considering $\mathfrak{m}_m$, $\mathfrak{m}_{m+1}$ and $\mathfrak{m}_{m,m+1}$.

The only remaining case is A_{15}, which can be solved by considering $\mathfrak{m}_{13}$, $\mathfrak{m}_{16}$ and $\mathfrak{m}_{13,16}$. $\square$

We now prove the main theorem of this paper.

Proof of Theorem 4.1. Recall that W_1 and W_2 denote the Weyl groups of H_1 and H_2 respectively. Let m_0 be the maximum among the ranks of simple factors of the groups H_1 and H_2. It is clear from Corollary 4.3 that $\mathfrak{m}_i(W_1) = \mathfrak{m}_i(W_2)$, that $\mathfrak{m}'_i(W_1) = \mathfrak{m}'_i(W_2)$ and that $\mathfrak{m}_{i,j}(W_1) = \mathfrak{m}_{i,j}(W_2)$ for any i, j. Then we apply Theorem 4.4 to conclude that the products of rank m_0 simple factors in W_1 and W_2 are isomorphic.

Let m be a positive integer less than m_0. For $i = 1, 2$, let W'_i be the subgroup of W_i which is the product of the Weyl groups of simple factors of H_i of rank greater than m. We assume that the groups W'_1 and W'_2 are isomorphic and then we prove that the products of the Weyl groups of rank m simple factors of H_1 and H_2 are isomorphic. This will complete the proof of the theorem by an induction argument.

Let U_i be the subgroup of W_i such that $W_i = U_i \times W_i'$. Then, since $\mathfrak{m}_j(W_1') = \mathfrak{m}_j(W_2')$ and $\mathfrak{m}'_j(W_1') = \mathfrak{m}'_j(W_2')$, we have

$$\mathfrak{m}_j(U_1) = \mathfrak{m}_j(W_1) - \mathfrak{m}_j(W_1') = \mathfrak{m}_j(W_2) - \mathfrak{m}_j(W_2') = \mathfrak{m}_j(U_2),$$
$$\mathfrak{m}'_j(U_1) = \mathfrak{m}'_j(W_1) - \mathfrak{m}'_j(W_1') = \mathfrak{m}'_j(W_2) - \mathfrak{m}'_j(W_2') = \mathfrak{m}'_j(U_2)$$

and similarly

$$\mathfrak{m}_{i,j}(U_1) = \mathfrak{m}_{i,j}(U_2).$$

Now we use Theorem 4.4 to conclude that the subgroups of W_i which are products of the Weyl groups of simple factors of H_i of rank m are isomorphic, for $i = 1, 2$.

The proof of the theorem can now be completed by the downward induction on m. It also follows from the proof of Theorem 4.4, that the Weyl groups of simple factors of H_1 and H_2 are pairwise isomorphic. □

Remark 4.5. We remark here that the above proof is valid even if we assume that the Weyl groups $W(H_1)$ and $W(H_2)$ share the same set of elements up to conjugacy in $\mathrm{GL}_n(\mathbb{Q})$, not just in $\mathrm{GL}_n(\mathbb{Z})$. Thus Theorem 1.1 is true under the weaker assumption that the groups H_1 and H_2 share the same set of maximal k-tori up to k-isogeny, not just up to k-isomorphism.

We also remark that the above proof holds over the fields k which admit arbitrary cyclic extensions and which have cohomological dimension ≤ 1.

Remark 4.6. Philippe Gille [2004] has recently proved that the map ψ described in Lemma 2.2 is surjective for any quasisplit semisimple group H. Therefore our main result, Theorem 1.1, now holds for all fields k which admit cyclic extensions of arbitrary degree.

Acknowledgements

The author wishes to thank Prof. Dipendra Prasad for suggesting this question, for numerous fruitful discussions, and for spending a lot of time going through the paper and correcting mistakes. The author would also like to thank Prof. Gopal Prasad for pointing out a mistake, Prof. J.-P. Serre for several useful suggestions, Prof. M. S. Raghunathan, Prof. R. Parimala and Dr. Maneesh Thakur for encouraging comments, and Joost van Hamel for informing him of Gille's result [2004].

References

[Borel and Springer 1968] A. Borel and T. A. Springer, "Rationality properties of linear algebraic groups, II", *Tôhoku Math. J.* (2) **20** (1968), 443–497. MR 39 #5576 Zbl 0211.53302

[Bruhat and Tits 1967] F. Bruhat and J. Tits, "Groupes algébriques simples sur un corps local", pp. 23–36 in *Proc. Conf. Local Fields* (Driebergen, 1966), edited by T. A. Springer, Springer, Berlin, 1967. MR 37 #6396 Zbl 0155.00204

[Gille 2004] P. Gille, "Type des tores maximaux des groupes semisimples", *J. Ramanujan Math. Soc.* **19**:3 (2004), 1–18.

[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 0768.20016

[Kariyama 1989] K. Kariyama, "On conjugacy classes of maximal tori in classical groups", *J. Algebra* **125**:1 (1989), 133–149. MR 90j:20090 Zbl 0682.20031

[Kneser 1965a] M. Kneser, "Galois-Kohomologie halbeinfacher algebraischer Gruppen über $\mathfrak{p}$-adischen Körpern, I", *Math. Z.* **88** (1965), 40–47. MR 30 #4760 Zbl 0143.04702

[Kneser 1965b] M. Kneser, "Galois-Kohomologie halbeinfacher algebraischer Gruppen über $\mathfrak{p}$-adischen Körpern, II", *Math. Z.* **89** (1965), 250–272. MR 32 #5658 Zbl 0143.04702

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

[Lang 1956] S. Lang, "Algebraic groups over finite fields", *Amer. J. Math.* **78** (1956), 555–563. MR 19,174a Zbl 0073.37901

[Larsen and Pink 1992] M. Larsen and R. Pink, "On l-independence of algebraic monodromy groups in compatible systems of representations", *Invent. Math.* **107**:3 (1992), 603–636. MR 93h:22031 Zbl 0778.11036

[Pierce 1982] R. S. Pierce, *Associative algebras*, Graduate Texts in Mathematics **88**, Springer, New York, 1982. MR 84c:16001 Zbl 0497.16001

[Platonov and Rapinchuk 1994] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure Appl. Math. **139**, Academic Press, Boston, 1994. MR 95b:11039 Zbl 0841.20046

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 82e:12016 Zbl 0423.12016

[Springer 1974] T. A. Springer, "Regular elements of finite reflection groups", *Invent. Math.* **25** (1974), 159–198. MR 50 #7371 Zbl 0287.20043

[Tits 1966] J. Tits, "Normalisateurs de tores, I; Groupes de Coxeter étendus", *J. Algebra* **4** (1966), 96–116. MR 34 #5942 Zbl 0145.24703

Received 7 July 2003. Revised 14 January 2004.

SHRIPAD M. GARGE
SCHOOL OF MATHEMATICS
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
DR HOMI BHABHA ROAD
COLABA
MUMBAI 400 005
INDIA

shripad@math.tifr.res.in
http://www.math.tifr.res.in/~shripad/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

KNOT MUTATION: 4-GENUS OF KNOTS AND ALGEBRAIC CONCORDANCE

SE-GOO KIM AND CHARLES LIVINGSTON

Kearton observed that mutation can change the concordance class of a knot. A close examination of his example reveals that it is of 4-genus 1 and has a mutant of 4-genus 0. The first goal of this paper is to show by examples that for any pair of nonnegative integers *m* and *n* there is a knot of 4-genus *m* with a mutant of 4-genus *n*.

A second result is a crossing change formula for the algebraic concordance class of a knot, which is then applied to prove the invariance of the algebraic concordance class under mutation. We conclude with an application of crossing change formulas to give a short new proof of Long's theorem that strongly positive amphicheiral knots are algebraically slice.

1. Introduction

The main goal of this paper is to examine the effect of knot mutation on two concordance invariants of knots, the 4-ball genus and the algebraic concordance class. We completely describe the extent to which mutation can change the 4-genus, and show that the algebraic concordance class of a knot, as defined in [Levine 1969b], is invariant under mutation. In the course of our work we develop a crossing change formula for the algebraic concordance class of a knot. We apply such an approach to demonstrate that Long's theorem that strongly positive amphicheiral knots are algebraically slice is an immediate corollary of the Hartley–Kawauchi theorem that such knots have Alexander polynomials that are squares. Lastly, we show that the Hartley–Kawauchi theorem also follows from a similar crossing change approach.

Mutation and algebraic concordance. The construction of a mutant K^* of a knot K consists in removing a 3-ball B from S^3 that meets K in two proper arcs and gluing it back in via an involution τ of its boundary S, where τ is orientation-preserving and leaves the set $S \cap K$ invariant. This is among the subtlest constructions of knot theory in that it leaves a wide range of knot invariants unchanged [Adams 1989; Kawauchi 1994; 1996; Kirk 1989; Kirk and Klassen 1990; Meyerhoff and Ruberman 1990; Rong 1994; Ruberman 1987; 1999]. Most relevant to

MSC2000: 57M25.
Keywords: mutation, knot concordance, amphicheiral, 4-genus, knot genus.

the work here is the statement of [Cooper and Lickorish 1999] that the Tristram–Levine signatures, σ_ω, are invariant under mutation, since, for ω a prime power root of unity, these provide the strongest classical bounds on the 4-genus [Murasugi 1965; Tristram 1969]: $\frac{1}{2}|\sigma_\omega(K)| \leq g_4(K)$. We will prove a more general result involving Levine's homomorphism [1969b] from the knot concordance group $\mathscr{C}$ to the algebraic concordance group $\mathscr{G}$:

Theorem 1.1. *Mutation does not change the image of a knot under Levine's homomorphism.*

One proof, given in Section 7, is entirely self-contained and gives a previously unnoticed crossing change formula for the algebraic concordance class of a knot. (As a side note, in Section 9 we use this crossing change formula to give a quick derivation of a result of Long that strongly positive amphicheiral knots are algebraically slice.) Section 8 present an alternate proof of Theorem 1.1; this argument is briefer, but depends on the detailed analysis of Seifert forms given in [Cooper and Lickorish 1999].

Mutation and the 4-genus of a knot. The 4-genus of a knot, $g_4(K)$, is the least genus of an embedded surface bounded by K in the 4-ball. This can be defined in either the smooth or topological locally flat category; the results of this paper apply in either. It is an especially challenging invariant to compute; there remain knots of low crossing number for which it is uncomputed, though the smooth category has advanced considerably in recent years, most notably with the solution of the Milnor conjecture giving the 4-genus of torus knots [Kronheimer and Mrowka 1993].

Almost nothing has been known concerning the interplay between mutation and the 4-genus. Basically the only success in this realm consists of Kearton's observation [1989] that an example of [Livingston 1983] yields an example for which mutation changes the concordance class of a knot. A close examination of that example shows that it has 4-genus 1, but it has a mutant of 4-genus 0. Further such examples have since been developed in [Kirk and Livingston 1999; 2001]. Our main result regarding the 4-genus is:

Theorem 1.2. *For every pair of nonnegative integers m and n, there is a knot K with mutant K^* satisfying $g_4(K) = m$ and $g_4(K^*) = n$.*

It should be noted that the original argument of [Livingston 1983] was based on [Gilmer 1983], in which it is now known an error appears. To correct for that, one must base the argument of [Livingston 1983] on a 3-fold branched cover rather than the 2-fold cover. We do this here.

Strongly positive amphicheiral knots. A knot K is called strongly positive amphicheiral if, when viewed as a knot in $\mathbb{R}^3$, it has a representative that is invariant under the map $\tau(x, y, z) = (-x, -y, -z)$ of $\mathbb{R}^3$. We consider two theorems:

Theorem 1.3 [Long 1984]. *A strongly positive amphicheiral knot is algebraically slice.*

Theorem 1.4 [Hartley and Kawauchi 1979]. *If K is strongly positive amphicheiral, the Alexander polynomial Δ_K is the square of a symmetric polynomial.*

In Section 9 we use crossing change formulas developed earlier to prove that Long's theorem is an immediate corollary of the Hartley–Kawauchi result. In Section 10 we use a crossing change argument to give a new proof of the Hartley–Kawauchi theorem.

2. Background on Casson–Gordon invariants

A key tool in the proof of Theorem 1.2 is the main theorem from [Gilmer 1982] bounding Casson–Gordon invariants in terms of the 4-genus of a knot. Here is a simplified description of that result, based on the statement of the theorem and later remarks in [Gilmer 1982].

Theorem 2.1 (Gilmer). *Let K be an algebraically slice knot such that $g_4(K) = g$ and let M_q be the q-fold branched cover of S^3 branched over K, with q a prime power. Let β denote the linking form on $H_1(M_q, \mathbb{Z})$. Then β can be written as a direct sum $\beta_1 \oplus \beta_2$ such that*

(1) *β_1 has a presentation of rank $2(q-1)g$, and*

(2) *β_2 has a metabolizer D such that, for any character χ of prime power order on $H_1(M_q, \mathbb{Z})$ given by linking with an element in D, one has*

$$|\sigma(K, \chi)| \leq 2qg.$$

Here $\sigma(K, \chi)$ is the Casson–Gordon invariant, originally denoted $\sigma_1\tau(K, \chi)$ in [Casson and Gordon 1986; Gilmer 1982]. We will need to know that D can be taken to be equivariant with respect to the deck transformation of M_q. Details concerning this and other points will be given below, as they arise.

In our applications the group $H_1(M_q, \mathbb{Z})$ will also be a vector space over a finite field, in which case a metabolizer for β_2 will be half-dimensional. Hence:

Corollary 2.2. *In Theorem 2.1, if $H_1(M_q, \mathbb{Z})$ is isomorphic to $H_1(M_q, \mathbb{Z}_p)$, a $\mathbb{Z}_p$-vector space, conclusion (1) can be restated as*

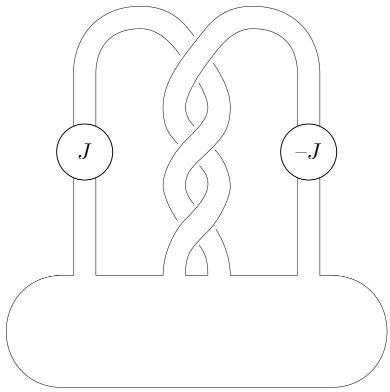
(1) $$\dim \beta_1 \leq 2(q-1)g$$

and in (2) the metabolizer D satisfies

$$\dim D \geq \tfrac{1}{2}\big(\dim H_1(M_q, \mathbb{Z}_p) - 2(q-1)g\big).$$

3. The building blocks

The figure illustrates a knot K_J of genus 1. The bands in the surface are tied in knots J and $-J$, for a knot J to be determined later. The twisting of the bands is such that the Seifert matrix for K_J is $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$.



Here $-J$ denotes the concordance inverse of J, formed from J by reversing the orientations of S^3 and the knot. A diagram for $-J$ is constructed by reflecting a diagram for J through a vertical line on the page and reversing the orientation of the knot. For K_J, the knot in the right band is the reflection through a vertical line of the knot in the left band. In all examples here, J can be taken to be reversible, so the details of the orientation issues for J are not critical.

Knots related to this one have been carefully analyzed elsewhere, for example [Gilmer and Livingston 1992; Livingston 1983; 2001], and the details of the following results can be found there. Here are the relevant facts.

(1) If M_3 denotes the 3-fold branched cover of S^3 branched over K_J, then

$$H_1(M_3, \mathbb{Z}) = \mathbb{Z}_7 \oplus \mathbb{Z}_7.$$

(2) As a $\mathbb{Z}_7$-vector space, $H_1(M_3, \mathbb{Z})$ splits as the direct sum of a 2-eigenspace, spanned by a vector e_2, and a 4-eigenspace, spanned by a vector e_4, with respect to the linear transformation induced by the deck transformation.

(3) Linking with e_i induces a character $\chi_i : H_1(M_3, \mathbb{Z}) \to \mathbb{Z}_7$. Results of Litherland [1984] (see also [Gilmer 1993; Gilmer and Livingston 1992]) give

$$\sigma(K, \chi_2) = \sigma_{1/7}(J) + \sigma_{2/7}(J) + \sigma_{3/7}(J),$$
$$\sigma(K, \chi_4) = -\sigma_{1/7}(J) - \sigma_{2/7}(J) - \sigma_{3/7}(J),$$

where $\sigma_{a/b}$ denotes the classical Levine–Tristram signature, also written as σ_ω with $\omega = e^{(a/b)2\pi i}$. To simplify notation we set, for any knot J,

$$s_7(J) = \sigma_{1/7}(J) + \sigma_{2/7}(J) + \sigma_{3/7}(J).$$

There are knots for which s_7 is arbitrarily large, for instance connected sums of trefoil knots, which are reversible.

4. The Basic Examples

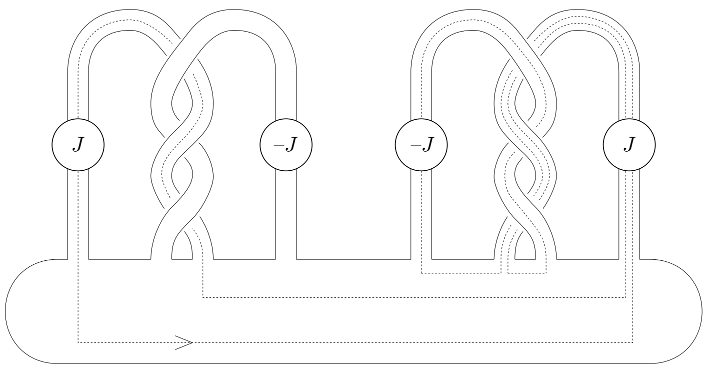
We denote by L_J the connected sum of K_J with the reverse of $-K_J$:

$$L_J = K_J \mathbin{\#} -K_J^r.$$

As observed by Kearton, L_J is a mutant of the slice knot $K_J \mathbin{\#} -K_J$.

Theorem 4.1. *For any choice of J, we have $g_4(L_J) \leq 1$ and thus $g_4(nL_J) \leq n$.*

Proof. Here is an illustration of L_J, showing also a simple closed curve on the genus-2 Seifert surface F. This curve has self-linking number 0 and represents the



slice knot $J \mathbin{\#} -J$. Thus F can be surgered in the 4-ball to reduce its genus to 1, showing that L_J bounds a surface of genus 1 in the 4-ball, as desired. □

The homology of the 3-fold branched cover of L_J, N_3, naturally splits as

$$(\mathbb{Z}_7 \oplus \mathbb{Z}_7) \oplus (\mathbb{Z}_7 \oplus \mathbb{Z}_7),$$

with a 2-eigenspace spanned by the vectors $e_2 \oplus 0$ and $0 \oplus e_2'$, which we abbreviate simply by e_2 and e_2'. Similarly for the 4-eigenspace. We denote the corresponding $\mathbb{Z}_7$-valued characters given by linking with e_2 and e_2' by χ_2 and χ_2', respectively.

Theorem 4.2. *The Casson–Gordon invariants of L_J are given by*

$$\sigma(L_J,\ a\chi_2 + b\chi_2') = \epsilon(a)s_7(J) + \epsilon(b)s_7(J),$$
$$\sigma(L_J,\ a\chi_4 + b\chi_4') = -\big(\epsilon(a)s_7(J) + \epsilon(b)s_7(J)\big),$$

where $\epsilon(x) = 0$ or 1 depending on whether $x = 0$ or $x \neq 0$ modulo 7.

Proof. This follows from the additivity of Casson–Gordon invariants; see [Litherland 1984] or [Gilmer 1983]. The only unexpected aspect of the formula is that, since we are dealing with $K_J \# -K_J^r$, it might have been anticipated that the difference $\epsilon(a)s_7(J) - \epsilon(b)s_7(J)$ would appear rather than the sum. This switch occurs because the connected sum involves the mirror image of the reverse, rather than simply the mirror image; thus the role of J and $-J$ are reversed in the second summand. □

5. Proof of Theorem 1.2

As observed by Kearton, for any knots L_1 and L_2, the connected sums $L_1 \# -L_2$ and $L_1 \# -L_2^r$ are mutants of each other. It follows immediately that for $m < n$, the knot nL_J is a mutant of $mL_J \# (n-m)(K_J \# -K_J)$. Since $K_J \# -K_J$ is slice, this second knot is concordant to, and hence of the same 4-genus as, mL_J. To prove Theorem 1.2 we show that for each positive integer n there exists a knot J such that $g_4(mL_J) = m$ for all $m \le n$.

Fix a positive integer n and select an arbitrary m with $1 \le m \le n$. The knot J will be chosen as its necessary properties become apparent.

Suppose that mL_J bounds a surface F in the 4-ball with genus $g(F) = k < m$. Let V_3 denote the 3-fold branched cover of B^4 branched over F having for boundary the m-fold connected sum mN_3. Also, abbreviate by D the image of Tor $H_2(V_3, mN_3, \mathbb{Z})$ in $H_1(mN_3, \mathbb{Z})$. An examination of the proof of Gilmer's theorem in [Gilmer 1982] reveals that this D is the metabolizer given in our statement of the result, Theorem 2.1. Thus $|\sigma(mL_J, \chi)| \le 6k$ for any χ corresponding to an element in D.

With $\mathbb{Z}_7$-coefficients, $H_1(mN_3, \mathbb{Z})$ has dimension $4m$, so by Gilmer's theorem we have $\dim H_1(mN_3, \mathbb{Z}) - 2\dim D \le 2(3-1)k = 4k$. Hence D is nontrivial, since $k < m$.

Observe that by its construction, D is equivariant with respect to the deck transformation and hence contains an eigenvector. Assume that it is a 2-eigenvector. If we write $H_1(mN_3, \mathbb{Z}) = \oplus_m H_1(N_3, \mathbb{Z})$, the 2-eigenvectors are naturally denoted $e_{2,i}$ and $e'_{2,i}$, with $1 \le i \le m$, where $e_{2,i}$ and $e'_{2,i}$ are the 2-eigenvectors in the i-th summand. A nontrivial 2-eigenvector in D will be of the form $\sum_i a_i e_{2,i} + \sum_i b_i e'_{2,i}$. Using additivity, the Casson–Gordon invariant corresponding to the dual character is given by:

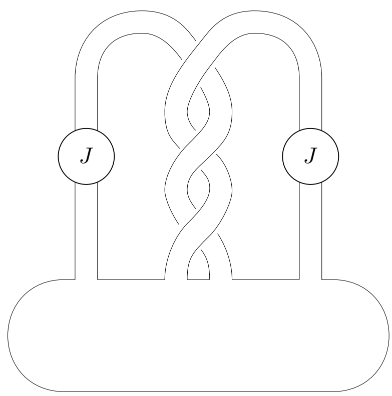
$$\Big(\sum_i \epsilon(a_i)\Big)\, s_7(J) + \Big(\sum_i \epsilon(b_i)\Big)\, s_7(J).$$

To complete the proof, observe that this sum is greater than or equal to $s_7(J)$, so that if J is chosen so that $s_7(J) > 6n$ a contradiction is achieved. Notice that the choice of J depends only on n and not m.

A similar argument applies if D contains only a 4-eigenvector.

6. The growth of $g_4(nK)$ for algebraically slice knots K

For a general knot K one has $g_4(nK) \leq ng_4(K)$ but one does not usually have an equality. In the case of a knot T, such as the trefoil, for which the 4-genus is detected by a classical (additive) invariant, such as the signature, one can sometimes demonstrate that $g_4(nT) = ng_4(T)$. But for algebraically slice knots with $g_4(K) \neq 0$ such arguments are not possible. In fact, it is unknown whether in the topological category there is such an algebraically slice knot for which the equality holds for all n. (In the smooth setting, Livingston [2003] has constructed an algebraically slice knot K for which $g_4(K) = \tau(K) = 1$, where τ is the invariant defined in [Ozsváth and Szabó 2003]. Since τ is additive and bounds g_4, it follows that $g_4(nK) = ng_4(K)$ for all n.) We will here observe that one can come quite close for the knot T_J, where T_J is the knot illustrated below, built as K_J is, only



with J tied in both bands rather than J in one band and $-J$ in the other. (Similar results hold for K_J and L_J but the proof would require the continued use of 3-fold covers rather than the 2-fold cover for which the estimates are simpler.)

Theorem 6.1. *For all ϵ with $0 < \epsilon < 1$, there is a knot J such that $g_4(nT_J) > (1-\epsilon)ng_4(T_J)$ for all $n > 0$.*

Proof. Our proof builds upon Gilmer's original argument [1982]. Observe first that $g_4(T_J) \leq 1$. For the 2-fold branched cover we have that $H_1(M_2, \mathbb{Z}) = \mathbb{Z}_3 \oplus \mathbb{Z}_3$ and the $\mathbb{Z}_3$-dimension satisfies $\dim H_1(nM_2, \mathbb{Z}_3) = 2n$.

If nT_J bounds a surface in the 4-ball of genus k at most $(1-\epsilon)n$, then by Gilmer's theorem there exists a self-annihilating summand D with

$$\dim H_1(nM_2, \mathbb{Z}_3) - 2\dim D \leq 2k$$

and such that $|\sigma(nK_J, \chi)| \leq 4k$ for all characters χ dual to elements in D.

One computes that $\dim D \geq n-k$. A linear algebra argument, basically Gauss–Jordan elimination, now implies that some element of D will be of the form $\bigoplus_i \chi_i$ with at least $n-k$ of the χ_i nontrivial, and for each of these χ_i the corresponding Casson–Gordon invariant is at least $2\sigma_{1/3}(J)$. Thus we have the equation

$$\left|(n-k)2\sigma_{1/3}(J)\right| \leq 4k.$$

Since $k \leq (1-\epsilon)n$, this reduces to $\left|\epsilon n 2\sigma_{1/3}(J)\right| \leq 4(1-\epsilon)n$, which is to say

$$\left|\sigma_{1/3}(J)\right| \leq \frac{2(1-\epsilon)}{\epsilon}.$$

The proof is completed by noting that for any ϵ one can select a J for which this inequality does not hold. □

7. Mutation and algebraic concordance

In this section we develop a crossing change formula for the algebraic concordance class of a knot in order to prove Theorem 1.1: mutation preserves the algebraic concordance class of a knot. Certain knot invariants, such as the Alexander polynomial and Tristram–Levine signatures, provide algebraic concordance invariants, and these have been shown to be mutation invariants (see for instance [Cooper and Lickorish 1999; Lickorish and Millett 1987]), but the general question of whether mutation can change the algebraic concordance class has remained open. We note that changing a knot to its orientation reverse is a very special case of mutation and reversal does not change the algebraic concordance class of a knot, as follows from [Long 1984]. (More directly, it can be shown that the complete set of algebraic concordance invariants defined by Levine [1969a] are unchanged by matrix transposition, the operation on Seifert matrices induced by reversal.)

We will first present a proof that the normalized Alexander polynomial is invariant under mutation; this argument is not new but must be presented to set up the needed notation for the analysis of algebraic concordance that follows. This is followed by a review of the theory and algebra of Levine's [1969a] algebraic concordance group $\mathcal{G}$. In the last part of the section we present a crossing change formula for the algebraic concordance class of a knot and use this to prove the mutation invariance of this class.

The Alexander and Conway polynomial. For an oriented link L, a choice of connected Seifert surface F for L, and a choice of basis for $H_1(F, \mathbb{Z})$ there is a Seifert matrix $V(L)$, say of dimension $r \times r$. The (normalized) Alexander polynomial $\Delta_L(t)$ of L can be defined by setting

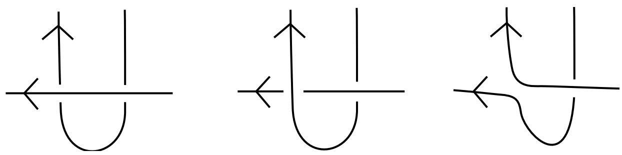
$$V_t(L) = (1-t)V + (1-\bar{t})V^t \qquad \text{and} \qquad \Delta_L(t) = \frac{1}{z^r}\det V_t(L),$$

where V^t denotes the transpose, $\bar{t} = t^{-1}$ and $z = t^{-1/2} - t^{1/2}$. (Recall that $\Delta_L(t)$ can be expressed as a polynomial in z, $\Delta_L(t) = C_L(z) \in \mathbb{Z}[z]$, and this defines the Conway polynomial [1970].) Notice that $z^2 = -(1-\bar{t})(1-t)$, so that if r is even (for instance, when L is connected, so r is twice the genus of F), we have $\Delta_L \in \mathbb{Z}[\bar{t}, t]$ and elementary algebraic manipulations lead to the usual normalized Alexander polynomial,

$$\Delta_L(t) = t^{-r/2} \det(V - tV^t).$$

(This polynomial is clearly independent of change of basis and an observation below will show that it is an S-equivalence invariant [Trotter 1973] and thus depends only on K.)

Here is a local picture of link diagrams for links L_-, L_+, and L_s, with the



L_- $\qquad$ L_+ $\qquad$ L_s

diagrams identical outside the local picture. Any crossing change and smoothing can be achieved using this local change. In the diagram for L_- a Reidemeister move eliminates the two crossings. If Seifert's algorithm is used to construct a Seifert surface F_0 for L_- using this simplified diagram, the corresponding Seifert matrix will be denoted A. The Seifert surfaces for the links L_- and L_+ that arise from Seifert's algorithm applied to the given diagrams are formed from F_0 by adding two twisted bands. From this we have that $V(L_\pm)$ is given by a $(r+2) \times (r+2)$ matrix of the form

$$V(L_\pm) = \begin{pmatrix} & & & a_1 & 0 \\ & A & & \vdots & \vdots \\ & & & a_r & 0 \\ a_1 & \cdots & a_r & b & 1 \\ 0 & \cdots & 0 & 0 & \epsilon_\pm \end{pmatrix},$$

where all entries are identical in these two matrices except that $\epsilon_- = 0$ and $\epsilon_+ = -1$. $V(L_s)$ is given by the same matrix, with the last row and column deleted.

A few consequences of these calculations follow quickly.

Theorem 7.1. *The normalized Alexander polynomial is an S-equivalence invariant and hence is a knot invariant.*

Proof. S-equivalence is generated by the operation on Seifert matrices that takes a matrix A and replaces it with the matrix denoted $V(L_-)$ above. That this doesn't change the Alexander polynomial is easily checked: expand the relevant determinant along the last column and then along the last row. □

Theorem 7.2 (The Conway skein relation). *The Alexander polynomial satisfies* $\Delta_{L_+} - \Delta_{L_-} = z\Delta_{L_s}$.

Proof. This again is a simple exercise in algebra, expanding the determinant along the last column and then last row. □

Theorem 7.3. *The Alexander polynomials of mutant knots are the same.*

Proof. In the construction of the mutant K^*, if the intersection of K with the ball B that is being taken out and replaced via an involution is invariant under the extension of that involution to the 3-ball, then $K^* = K$ and the polynomials are the same. In general, a series of crossing changes and smoothings converts $K \cap B$ into invariant tangles, so, via the Conway skein relation, the polynomial of K^* is the same as that for K. □

If K is a knot, the Alexander polynomial satisfies $\Delta_K(1) = 1$ and in particular $\Delta_K(t)$ is nontrivial. Hence, in the matrices above, working now with K instead of L, A_t is nonsingular. Thus, for $V_t(K_\pm)$ the same set of row and column operations can be used to eliminate the entries corresponding to the a_i in V. There results the following matrix $W_t(K_\pm)$, where the entries are rational functions in t and the matrix is hermitian with respect to the involution induced by the map $t \to \bar{t}$:

$$W_t(K_\pm) = \begin{pmatrix} & A_t & & 0 & 0 \\ & & & \vdots & \vdots \\ 0 & \cdots & 0 & c(t) & 1-t \\ 0 & \cdots & 0 & 1-\bar{t} & \epsilon_\pm(1-t)(1-\bar{t}) \end{pmatrix}.$$

Lemma 7.4. *The ratio* $\Delta_{K_+}/\Delta_{K_-}$ *is equal to* $c(t)+1$.

Proof. This follows from a calculation of the relevant determinants. □

Algebraic concordance. An algebraic Seifert matrix is a square integral matrix V satisfying $\det(V - V^t) = \pm 1$. Such a matrix is called metabolic if it is congruent to a matrix of the form

$$\begin{pmatrix} 0 & A \\ B & C \end{pmatrix},$$

with A, B, and C square. Levine defined the algebraic concordance group $\mathcal{G}$ to be the set of equivalence classes of algebraic Seifert matrices, with V_1 and V_2 equivalent if $V_1 \oplus -V_2$ is metabolic. The group operation is induced by direct sum.

A rational algebraic concordance group $\mathcal{G}^{\mathbb{Q}}$ can be similarly defined, where now it is required that $\det\big((V-V^t)(V+V^t)\big) \neq 0$. Levine [1969a] proved that the inclusion $\mathcal{G} \to \mathcal{G}^{\mathbb{Q}}$ is injective.

Consider next the set of nonsingular hermitian matrices with coefficients in the field $\mathbb{Q}(t)$, where $\mathbb{Q}(t)$ has the involution $t \to \bar{t}$. In this case the equivalence relation generated by congruence to metabolic matrices results in the Witt group of $\mathbb{Q}(t)$, denoted $W(\mathbb{Q}(t))$.

Theorem 7.5. *The map*

$$V \to V_t = (1-t)V + (1-\bar{t})V^t$$

induces an injection $\mathcal{G} \to W(\mathbb{Q}(t))$.

Proof. A proof is given in [Litherland 1984] for $\mathcal{G}^{\mathbb{Q}}$ (denoted there by $W_S(\mathbb{Q}, -)$), and the theorem follows from the injectivity of the inclusion $\mathcal{G} \to \mathcal{G}^{\mathbb{Q}}$. In defining $\mathcal{G}^{\mathbb{Q}}$, Litherland restricts to nonsingular matrices, but as he notes, Levine proved that every class in $\mathcal{G}$ has a nonsingular representative. To simplify notation, we will use $W_t(K)$ to denote both the matrix and the Witt class represented by the matrix when the meaning is clear in context. □

Crossing changes and algebraic concordance. From the calculations and notation above, if a crossing change is performed on a knot K, the difference of Witt classes associated to the Seifert forms is given by

$$W_t(K_+) - W_t(K_-) = (A_t \oplus C_+) \oplus -(A_t \oplus C_-),$$

where

$$C_\pm = \begin{pmatrix} c(t) & 1-t \\ 1-\bar{t} & \epsilon_\pm(1-t)(1-\bar{t}) \end{pmatrix}.$$

Since $A_t \oplus -A_t$ is Witt trivial, as is C_-, only C_+ contributes to the difference of Witt classes. Diagonalization, the identification of $c(t)+1$ with $\Delta_{L_+}/\Delta_{K_-}$, and a final multiplication of a basis element (by Δ_{K_-}) yields the following theorem.

Theorem 7.6. $W_t(K_+) - W_t(K_-)$ *is represented by the matrix*

$$\begin{pmatrix} \Delta_{K_+}(t)\Delta_{K_-}(t) & 0 \\ 0 & -1 \end{pmatrix},$$

and thus the difference is determined by the Alexander polynomials of the knots.

The special case of $\omega = -1$ in the following corollary is a result from [Murasugi 1965]. The proof of the corollary follows from Theorem 7.6 by setting $t = \omega$ and induction on the number of crossing changes needed to reduce K to an unknot. To avoid the matrix being nonsingular, we must restrict to prime power roots of unity.

Corollary 7.7. *For ω a prime power root of unity,* $\operatorname{sign}(\Delta_K(\omega)) = (-1)^{\sigma_\omega(K)/2}$.

We now have the main result of this section, the following corollary of Theorem 7.6, a restatement of Theorem 1.1.

Corollary 7.8. *The algebraic concordance class of a knot is invariant under mutation; that is,* $W_t(K) = W_t(K^*)$ *for any knot* K *and its mutant* K^*.

Proof. A sequence of crossing changes in the tangle in K that is being mutated converts it into a tangle that is invariant under mutation. Thus we have a sequence of knots

$$K = K_0,\ K_1,\ \dots,\ K_n = K_n^*,\ K_{n-1}^*,\ \dots,\ K_0^* = K^*,$$

where $K_n = K_n^*$. By the previous theorem and the mutation invariance of the Alexander polynomial, each pair of successive differences is equal:

$$W_t(K_i) - W_t(K_{i+1}) = W_t(K_i^*) - W_t(K_{i+1}^*).$$

Thus $W_t(K) - W_t(K_n) = W_t(K^*) - W_t(K_n^*)$. Since $K_n = K_n^*$, the proof is complete. □

8. Generalized Mutation

Cooper and Lickorish [1999] studied the effect of a generalization of mutation, called *genus-2 mutation*, on the Seifert form of a knot. Here we deduce from their result an alternative proof of Theorem 1.1. In fact, since they demonstrate that generalized mutation generates a finer relation than mutation, a stronger result than Theorem 1.1 is in fact achieved.

Genus-2 mutation consists of removing a solid handlebody of genus 2 that contains a knot K from S^3 and replacing it via an involution of the boundary. The involution is selected to extend to the solid handlebody so that it has three fixed arcs. The resulting knot is called K^*. According to [Cooper and Lickorish 1999] there are Seifert matrices for K and K^* of the form

$$V = \begin{pmatrix} A & B^t \\ B & C \end{pmatrix} \qquad \text{and} \qquad V^* = \begin{pmatrix} A & B^t \\ B & C^t \end{pmatrix},$$

respectively, where A and C are square and B is of the form $(0 \mid b)$ for some single column b. Since V is a Seifert matrix and $V - V^t = (A - A^t) \oplus (C - C^t)$, we see that A and C are also algebraic Seifert matrices. Note that

$$V_t = \begin{pmatrix} A_t & -z^2 B^t \\ -z^2 B & C_t \end{pmatrix} \qquad \text{and} \qquad V_t^* = \begin{pmatrix} A_t & -z^2 B^t \\ -z^2 B & (C^t)_t \end{pmatrix}$$

where $z = t^{-1/2} - t^{1/2}$ and $z^2 = -(1-t)(1-\bar{t}) = -(1-t) - (1-\bar{t})$.

Since A is a Seifert matrix, A_t is nonsingular and hermitian. Let

$$P = \begin{pmatrix} I & z^2 (A_t)^{-1} B^t \\ 0 & I \end{pmatrix}.$$

Then V_t and V_t^* are congruent to $\bar{P}^t V_t P$ and $\bar{P}^t V_t^* P$, respectively, which in turn are seen, after a simple computation, to equal

$$\begin{pmatrix} A_t & 0 \\ 0 & C_t - z^4 B(A_t)^{-1} B^t \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A_t & 0 \\ 0 & (C^t)_t - z^4 B(A_t)^{-1} B^t \end{pmatrix}.$$

Suppose that A is an $m \times m$ matrix. Let $\alpha(t) \in \mathbb{Q}(t)$ be the (m, m) entry of $(A_t)^{-1}$ and recall that $B = (0 \mid b)$ for some single column b with integral entries. It is easy to see that

$$B(A_t)^{-1} B^t = \alpha(t) b b^t.$$

In particular, it is symmetric. For simplicity, let $E = C_t - z^4 B(A_t)^{-1} B^t$. Then $E^t = (C^t)_t - z^4 B(A_t)^{-1} B^t$ and we have that V_t and V_t^* are congruent to $A_t \oplus E$ and $A_t \oplus E^t$, respectively. The difference of Witt classes of V_t and V_t^* is given by

$$(A_t \oplus E) \oplus -(A_t \oplus E^t).$$

Since $A_t \oplus -A_t$ is Witt trivial, only $E \oplus -E^t$ contributes to the difference of Witt classes. Observe that E is a nonsingular hermitian matrix since $A_t \oplus E$ and A_t are. There is a nonsingular matrix Q such that $F = \overline{Q}^t E Q$ is diagonal. This implies that $F = F^t = Q^t E^t \overline{Q}$. Using congruence by base change $Q \oplus \overline{Q}$, we see $E \oplus -E^t$ is congruent to $F \oplus -F$, which is Witt trivial. Thus, $V_t = V_t^*$ in $W(\mathbb{Q}(t))$ and K and K^* are algebraically concordant since $\mathcal{G} \to W(\mathbb{Q}(t))$ is injective.

9. Strongly positive amphicheiral knots

A knot K is called strongly positive amphicheiral if it is invariant under an orientation-reversing involution of S^3 that preserves the orientation of K. This is easily seen to be equivalent to the statement that K, when viewed as a knot in $\mathbb{R}^3 \subset S^3$, is isotopic to a knot, again denoted by K, that is invariant under the involution $\tau : \mathbb{R}^3 \to \mathbb{R}^3$ given by $\tau(x) = -x$, where $x \in \mathbb{R}^3$.

Hartley and Kawauchi [1979] proved that if K is strongly positive amphicheiral then $\Delta_K(t) = (F(t))^2$ for some Alexander polynomial F. Long [1984] proved that strongly positive amphicheiral knots are algebraically slice. Here we demonstrate that Long's theorem is in fact a corollary of the Hartley–Kawauchi theorem and the crossing change formula for the algebraic concordance class.

A bit of notation will be helpful: for a strongly amphicheiral knot that is invariant under the involution τ, τ defines a pairing of the crossing points in a diagram of K. A *paired crossing change* on such a K consists of changing both of a pair of crossings. Notice that since τ is orientation-reversing, the two crossings will be of opposite sign, so we denote the original knot K_{+-} and the knot formed by making the paired crossing changes K_{-+}.

Lemma 9.1. *A sequence of paired crossing changes converts a strongly positive amphicheiral knot into the unknot.*

Proof. Since an involution of S^1 cannot have one fixed point, K misses the origin in $\mathbb{R}^3$ and thus projects to a knot $\overline{K}$ in the quotient $\mathbb{R}^3 - \{0\}/\tau \equiv \mathbb{RP}^2 \times \mathbb{R}$. Since $\overline{K}$ lifts to a single component in the cover, it is homotopic to standard generator of $\pi_1(\mathbb{RP}^2 \times \mathbb{R})$, whose lift is an unknot in the cover. That homotopy can be carried out by a sequence of crossing changes, each of which lifts to a pair of crossing changes in the cover. □

Theorem 9.2 [Long 1984]. *A strongly positive amphicheiral knot is algebraically slice.*

Proof. Let K be the knot. By the previous lemma we need only show that $W_t(K_{+-}) - W_t(K_{-+})$ represents 0 in $W(\mathbb{Q}(t))$.

Working in the Witt group we can write

$$W_t(K_{+-}) - W_t(K_{-+}) = \big(W_t(K_{+-}) - W_t(K_{--})\big) - \big(W_t(K_{-+}) - W_t(K_{--})\big).$$

Applying Theorem 7.6, this is represented by the difference

$$\begin{pmatrix} \Delta_{K_{+-}}(t)\Delta_{K_{--}}(t) & 0 \\ 0 & -1 \end{pmatrix} \oplus -\begin{pmatrix} \Delta_{K_{-+}}(t)\Delta_{K_{--}}(t) & 0 \\ 0 & -1 \end{pmatrix}$$

Applying the Hartley–Kawauchi theorem, we write

$$\Delta_{K_{+-}}(t) = F(t)^2 \quad \text{and} \quad \Delta_{K_{-+}}(t) = G(t)^2,$$

and then cancel the (-1) summands to arrive at the difference

$$\begin{pmatrix} F(t)^2\Delta_{K_{--}}(t) & 0 \\ 0 & -G(t)^2\Delta_{K_{--}}(t) \end{pmatrix}.$$

This form has a metabolizer generated by the vector $(G(t), F(t)) \in \mathbb{Q}(t)^2$, and hence it is trivial in the Witt group, as desired. □

10. The Hartley–Kawauchi Theorem

Here we present a combinatorial proof of the theorem that for strongly positive amphicheiral knots the Alexander polynomial is a square of an Alexander polynomial. The proof also gives an alternative, though longer, route to Long's theorem than was given in the previous section. We begin by considering the existence of an equivariant Seifert surface for such a knot.

If Seifert's algorithm for constructing a Seifert surface is applied to a diagram for a strongly amphicheiral knot that is invariant under τ, the resulting surface will be invariant. In addition, τ restricted to this surface is orientation-preserving since τ preserves the orientation of the knot that is the boundary of the surface.

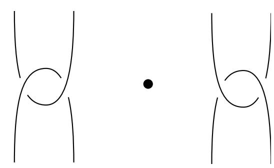
However τ reverses the positive normal direction since it reverses the orientation of R^3. Thus:

Lemma 10.1. *Let K be a strongly positive amphicheiral knot with involution τ. A Seifert surface F of K can be constructed so that F is invariant under τ and its Seifert form θ satisfies*

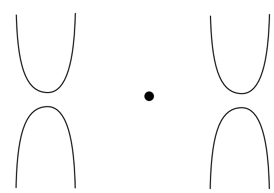
$$\theta(\tau u, \tau v) = -\theta(v, u)$$

for all $u, v \in H_1(F)$.

To understand the effect of crossing changes, we consider two figures. The first represents a portion of a symmetric diagram of a strongly amphicheiral knot, say K_{+-}:



The dot in center of the figure represents the origin in $\mathbb{R}^3$, the center of symmetry. For the knot K_{-+} the diagram will be the same, only a symmetric pair of crossing changes has been made. Thus, for K_{-+} the clasps pull apart, leaving a knot, denoted K', with diagram as follows:



Suppose that K' has an equivariant Seifert surface F_0 given by Seifert's algorithm and $H_1(F_0)$ has symplectic basis $w_1, \ldots, w_r$. Then an equivariant Seifert surface F for K_{+-} is given by adding four bands to F_0. The basis for $H_1(F_0)$ can be naturally extended to symplectic one for $H_1(F)$, $w_1, \ldots, w_r, x, y, \tau x, \tau y$, where y has trivial Seifert pairing with all elements other than x and itself, and x has trivial Seifert pairing with τy.

Let A be the Seifert matrix of F_0 with respect to $w_1, \ldots, w_r$ and let T denote the matrix representing the action of τ on $H_1(F_0)$. Then Lemma 10.1 applied to F_0 can be rewritten in terms of matrices: $T^t A T = -A^t$. After hermitianizing and taking inverses, we have

$$T(A_t)^{-1}T^t = -(A_t^t)^{-1} = \overline{-(A_t)^{-1}}.$$

To find the Seifert matrix for F with respect to the basis above, a couple of things have to be clarified. First, note that

$$\theta(x, \tau x) = -\theta(\tau\tau x, \tau x) = -\theta(x, \tau x),$$

and hence $\theta(x, \tau x) = 0$. Similarly, $\theta(\tau x, x) = 0$. Next, let

$$a = \begin{pmatrix} \theta(w_1, x) \\ \vdots \\ \theta(w_r, x) \end{pmatrix} \qquad \text{and} \qquad T = (t_{ij})_{1 \le i, j \le r}.$$

Then

$$\begin{pmatrix} \theta(w_1, \tau x) \\ \vdots \\ \theta(w_r, \tau x) \end{pmatrix} = \begin{pmatrix} -\theta(x, \tau w_1) \\ \vdots \\ -\theta(x, \tau w_r) \end{pmatrix} = \begin{pmatrix} -\sum_j t_{j1}\theta(x, w_j) \\ \vdots \\ -\sum_j t_{jr}\theta(x, w_j) \end{pmatrix}$$

$$= -T^t \begin{pmatrix} \theta(x, w_1) \\ \vdots \\ \theta(x, w_r) \end{pmatrix} = -T^t a.$$

It follows readily that the Seifert matrix for K_{+-} is the $(r+4) \times (r+4)$ matrix

$$V^\epsilon = \begin{pmatrix} A & a & 0 & -T^t a & 0 \\ a^t & b & 1 & 0 & 0 \\ 0 & 0 & \epsilon & 0 & 0 \\ -a^t T & 0 & 0 & -b & 0 \\ 0 & 0 & 0 & -1 & -\epsilon \end{pmatrix}, \qquad \text{where } \epsilon = -1.$$

Similarly, for K_{-+} the same matrix arise, only in this case $\epsilon = 0$. After hermitianizing we get

$$V_t^\epsilon = \begin{pmatrix} A_t & -z^2 a & 0 & z^2 T^t a & 0 \\ -z^2 a^t & -z^2 b & 1-t & 0 & 0 \\ 0 & 1-\bar{t} & -z^2\epsilon & 0 & 0 \\ z^2 a^t T & 0 & 0 & z^2 b & -(1-\bar{t}) \\ 0 & 0 & 0 & -(1-t) & z^2\epsilon \end{pmatrix},$$

where $z = t^{-1/2} - t^{1/2}$. Let

$$P = \begin{pmatrix} I & z^2(A_t)^{-1}a & 0 & -z^2(A_t)^{-1}T^t a & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Let $W_t^\epsilon = \bar{P}^t V_t^\epsilon P$. Then

$$W_t^\epsilon = \begin{pmatrix} A_t & 0 & 0 & 0 & 0 \\ 0 & -z^2 b - z^4 a^t (A_t)^{-1} a & 1-t & z^4 a^t (A_t)^{-1} T^t a & 0 \\ 0 & 1-\bar{t} & -z^2\epsilon & 0 & 0 \\ 0 & z^4 a^t T (A_t)^{-1} a & 0 & z^2 b - z^4 a^t T (A_t)^{-1} T^t a & -(1-\bar{t}) \\ 0 & 0 & 0 & -(1-t) & z^2\epsilon \end{pmatrix}.$$

Let $c(t) = -z^2 b - z^4 a^t (A_t)^{-1} a$. Since W_t^ϵ is hermitian, $c(t) = \overline{c(t)}$. The (1, 1)-entry of the lower right 2×2 submatrix of W_t^ϵ is

$$z^2 b - z^4 a^t \left(T (A_t)^{-1} T^t\right) a = \overline{z^2 b + z^4 a^t (A_t)^{-1} a} = \overline{-c(t)} = -c(t).$$

Let $d(t) = z^4 a^t (A_t)^{-1} T^t a$. Then the 1×1 matrix $d(t)$ is equal to its transpose

$$z^4 a^t T (A_t^t)^{-1} a = z^4 a^t T \left(-T (A_t)^{-1} T^t\right) a = -z^4 a^t (A_t)^{-1} T^t a = -d(t),$$

and hence $d(t) = 0$. Also, note that $z^4 a^t T (A_t)^{-1} a = \overline{d(t)} = 0$ since W_t^ϵ is hermitian.

Thus V_t^ϵ is congruent, by base change P, to

$$A_t \oplus C \oplus -C^t,$$

where

$$C = \begin{pmatrix} c(t) & 1-t \\ 1-\bar{t} & -z^2\epsilon \end{pmatrix}.$$

Since $\det P = 1$,

$$\Delta_{K_{+-}} = (c(t)+1)^2 \frac{1}{z^r} \det A_t = (c(t)+1)^2 \Delta_{K_{-+}},$$

where $c(t) = c(\bar{t})$. This proves the Hartley–Kawauchi theorem.

Next, to prove Long's theorem, we will show that $V_t(K_{+-})$, A_t, and $V_t(K_{-+})$ are all Witt-equivalent. It suffices to show that $C \oplus -C^t$ is Witt-trivial. Observe that C is nonsingular and hermitian since $A_t \oplus C \oplus -C^t$ and A_t are. There is a nonsingular matrix Q such that $D = \overline{Q}^t C Q$ is diagonal. This implies that

$$D = D^t = Q^t C^t \overline{Q}.$$

Using congruence by base change $Q \oplus \overline{Q}$, we see that $C \oplus -C^t$ is congruent to $D \oplus -D$, which is Witt trivial. Thus, K_{+-} and K_{-+} are algebraically concordant. This proves Long's theorem.

References

[Adams 1989] C. C. Adams, "Tangles and the Gromov invariant", *Proc. Amer. Math. Soc.* **106**:1 (1989), 269–271. MR 89k:57004 Zbl 0673.57009

[Casson and Gordon 1986] A. J. Casson and C. M. Gordon, “Cobordism of classical knots”, pp. 181–199 in *À la recherche de la topologie perdue*, edited by A. Marin and L. Guillou, Progr. Math. **62**, Birkhäuser, Boston, 1986. MR MR900252 Zbl 0597.57001

[Conway 1970] J. H. Conway, “An enumeration of knots and links, and some of their algebraic properties”, pp. 329–358 in *Computational problems in abstract algebra* (Oxford, 1967), Pergamon, Oxford, 1970. MR 41 #2661 Zbl 0202.54703

[Cooper and Lickorish 1999] D. Cooper and W. B. R. Lickorish, “Mutations of links in genus 2 handlebodies”, *Proc. Amer. Math. Soc.* **127**:1 (1999), 309–314. MR 99b:57008 Zbl 0905.57004

[Gilmer 1982] P. M. Gilmer, “On the slice genus of knots”, *Invent. Math.* **66**:2 (1982), 191–197. MR 83g:57003 Zbl 0495.57002

[Gilmer 1983] P. M. Gilmer, “Slice knots in S^3”, *Quart. J. Math. Oxford Ser.* (2) **34**:135 (1983), 305–322. MR 85d:57004 Zbl 0542.57007

[Gilmer 1993] P. Gilmer, “Classical knot and link concordance”, *Comment. Math. Helv.* **68**:1 (1993), 1–19. MR 94c:57007 Zbl 0805.57005

[Gilmer and Livingston 1992] P. Gilmer and C. Livingston, “The Casson–Gordon invariant and link concordance”, *Topology* **31**:3 (1992), 475–492. MR 93h:57037 Zbl 0797.57001

[Hartley and Kawauchi 1979] R. Hartley and A. Kawauchi, “Polynomials of amphicheiral knots”, *Math. Ann.* **243**:1 (1979), 63–70. MR 81c:57004 Zbl 0394.57009

[Kawauchi 1994] A. Kawauchi, “Topological imitation, mutation and the quantum SU(2) invariants”, *J. Knot Theory Ramifications* **3**:1 (1994), 25–39. MR 95a:57025 Zbl 0823.57012

[Kawauchi 1996] A. Kawauchi, “Mutative hyperbolic homology 3-spheres with the same Floer homology”, *Geom. Dedicata* **61**:2 (1996), 205–217. MR 97g:57020 Zbl 0857.57012

[Kearton 1989] C. Kearton, “Mutation of knots”, *Proc. Amer. Math. Soc.* **105**:1 (1989), 206–208. MR 89e:57001 Zbl 0675.57005

[Kirk 1989] P. A. Kirk, “Mutations of homology spheres and Casson’s invariant”, *Math. Proc. Cambridge Philos. Soc.* **105**:2 (1989), 313–318. MR 90a:57019 Zbl 0691.57003

[Kirk and Klassen 1990] P. A. Kirk and E. P. Klassen, “Chern–Simons invariants of 3-manifolds and representation spaces of knot groups”, *Math. Ann.* **287**:2 (1990), 343–367. MR 91d:57008 Zbl 0681.57006

[Kirk and Livingston 1999] P. Kirk and C. Livingston, “Twisted knot polynomials: inversion, mutation and concordance”, *Topology* **38**:3 (1999), 663–671. MR 2000c:57011 Zbl 0928.57006

[Kirk and Livingston 2001] P. Kirk and C. Livingston, “Concordance and mutation”, *Geom. Topol.* **5** (2001), 831–883. MR 2002j:57016 Zbl 1002.57007

[Kronheimer and Mrowka 1993] P. B. Kronheimer and T. S. Mrowka, “Gauge theory for embedded surfaces, I”, *Topology* **32**:4 (1993), 773–826. MR 94k:57048 Zbl 0799.57007

[Levine 1969a] J. Levine, “Invariants of knot cobordism”, *Invent. Math.* **8** (1969), 98–110. MR 40 #6563 Zbl 0179.52401

[Levine 1969b] J. Levine, “Knot cobordism groups in codimension two”, *Comment. Math. Helv.* **44** (1969), 229–244. MR 39 #7618 Zbl 0176.22101

[Lickorish and Millett 1987] W. B. R. Lickorish and K. C. Millett, “A polynomial invariant of oriented links”, *Topology* **26**:1 (1987), 107–141. MR 88b:57012 Zbl 0608.57009

[Litherland 1984] R. A. Litherland, "Cobordism of satellite knots", pp. 327–362 in *Four-manifold theory* (Durham, NH, 1982), edited by C. Gordon and R. Kirby, Contemp. Math. **35**, Amer. Math. Soc., Providence, RI, 1984. MR 86k:57003 Zbl 0563.57001

[Livingston 1983] C. Livingston, "Knots which are not concordant to their reverses", *Quart. J. Math. Oxford Ser.* (2) **34**:135 (1983), 323–328. MR 85d:57005 Zbl 0537.57003

[Livingston 2001] C. Livingston, "Infinite order amphicheiral knots", *Algebr. Geom. Topol.* **1** (2001), 231–241. MR 2002d:57009 Zbl 0997.57006

[Livingston 2003] C. Livingston, "Splitting the concordance group of algebraically slice knots", *Geom. Topol.* **7** (2003), 641–643. MR 2004j:57009 Zbl 02062472

[Long 1984] D. D. Long, "Strongly plus-amphicheiral knots are algebraically slice", *Math. Proc. Cambridge Philos. Soc.* **95**:2 (1984), 309–312. MR 85h:57007 Zbl 0547.57006

[Meyerhoff and Ruberman 1990] R. Meyerhoff and D. Ruberman, "Mutation and the η-invariant", *J. Differential Geom.* **31**:1 (1990), 101–130. MR 91j:57017 Zbl 0689.57012

[Murasugi 1965] K. Murasugi, "On a certain numerical invariant of link types", *Trans. Amer. Math. Soc.* **117** (1965), 387–422. MR 30 #1506 Zbl 0137.17903

[Ozsváth and Szabó 2003] P. Ozsváth and Z. Szabó, "Knot Floer homology and the four-ball genus", *Geom. Topol.* **7** (2003), 615–639. MR 2004i:57036 Zbl 1037.57027

[Rong 1994] Y. W. Rong, "Mutation and Witten invariants", *Topology* **33** (1994), 499–507. MR 95f:57021 Zbl 0823.57011

[Ruberman 1987] D. Ruberman, "Mutation and volumes of knots in S^3", *Invent. Math.* **90**:1 (1987), 189–215. MR 89d:57018 Zbl 0634.57005

[Ruberman 1999] D. Ruberman, "Mutation and gauge theory, I: Yang–Mills invariants", *Comment. Math. Helv.* **74**:4 (1999), 615–641. MR 2001c:57032 Zbl 0940.57033

[Tristram 1969] A. G. Tristram, "Some cobordism invariants for links", *Proc. Cambridge Philos. Soc.* **66** (1969), 251–264. MR 40 #2104 Zbl 0191.54703

[Trotter 1973] H. F. Trotter, "On S-equivalence of Seifert matrices", *Invent. Math.* **20** (1973), 173–207. MR 58 #31100 Zbl 0269.15009

Received November 26, 2003. Revised May 11, 2004.

SE-GOO KIM
DEPARTMENT OF MATHEMATICS
KYUNG HEE UNIVERSITY
HOEKI-DONG
DONGDAEMOON-KU
SEOUL 130-710
SOUTH KOREA
sgkim@khu.ac.kr

CHARLES LIVINGSTON
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, IN 47401
livingst@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

RATIONAL JET DEPENDENCE OF FORMAL EQUIVALENCES BETWEEN REAL-ANALYTIC HYPERSURFACES IN $\mathbb{C}^2$

R. TRAVIS KOWALSKI

Let (M, p) and $(\hat{M}, \hat{p})$ be the germs of real-analytic 1-infinite type hypersurfaces in $\mathbb{C}^2$. We prove that any formal equivalence sending (M, p) into $(\hat{M}, \hat{p})$ is formally parametrized (and hence uniquely determined by) its jet at p of a predetermined order depending only on (M, p). As an application, we use this to examine the local formal transformation groups of such hypersurfaces.

1. Introduction

A *formal* (*holomorphic*) *mapping* $H : (\mathbb{C}^2, p) \to (\mathbb{C}^2, \hat{p})$, with $p, \hat{p} \in \mathbb{C}^2$, is a $\mathbb{C}^2$-valued formal power series

$$H(Z) = \hat{p} + \sum_{|\alpha| \geq 1} c_\alpha (Z - p)^\alpha, \quad c_\alpha \in \mathbb{C}^2, \quad Z = (Z_1, Z_2).$$

The map H is *invertible* if there exists a formal map $H^{-1} : (\mathbb{C}^2, \hat{p}) \to (\mathbb{C}^2, p)$ such that $H(H^{-1}(Z)) \equiv H^{-1}(H(Z)) \equiv Z$ as formal power series; equivalently, if the Jacobian of H is nonvanishing at p. We denote by $J^k(\mathbb{C}^2, \mathbb{C}^2)_{p,\hat{p}}$ the *jet space* of order k of (formal) holomorphic mappings $(\mathbb{C}^2, p) \to (\mathbb{C}^2, \hat{p})$, and by $j_p^k(H) \in J^k(\mathbb{C}^2, \mathbb{C}^2)_{p,\hat{p}}$ the *k-jet* of H at p. (See Section 2 for further details.)

Suppose that (M, p) and $(\hat{M}, \hat{p})$ are (germs of) real-analytic hypersurfaces at p and $\hat{p}$ respectively, given by the real-analytic, real-valued local defining functions $\rho(Z, \bar{Z})$ and $\hat{\rho}(Z, \bar{Z})$. The formal map H is said to *take* (M, p) *into* $(\hat{M}, \hat{p})$ if

$$\hat{\rho}\big(H(Z), \overline{H(Z)}\big) \equiv c(Z, \bar{Z})\rho(Z, \bar{Z})$$

(in the sense of power series) for some formal power series $c(Z, \bar{Z})$; if in addition the formal map is invertible, it is called a *formal equivalence* between (M, p) and $(\hat{M}, \hat{p})$, and the germs themselves are called *formally equivalent*.

We wish to study the parametrization and finite determination of invertible formal holomorphic mappings of $\mathbb{C}^2$ taking one real-analytic hypersurface M into

MSC2003: 32H12, 32V20.
Keywords: real hypersurfaces, formal equivalence, jet determination.

another. There is a great deal of literature on this if M is assumed to be *minimal* at p, that is, if there is no complex hypersurface through p in $\mathbb{C}^2$ contained in M; see the remarks at the end of this introduction. In the present paper, however, we shall assume that M is not minimal at p, so that there exists a complex hypersurface $\Sigma \subset \mathbb{C}^2$ with $p \in \Sigma \subset M$. It is well known [Chern and Moser 1974; Baouendi et al. 1999b, Chapter IV] that for any real-analytic hypersurface $M \subset \mathbb{C}^2$ and point $p \in M$ (not necessarily minimal), there exist local holomorphic coordinates $(z, w) \in \mathbb{C} \times \mathbb{C}$, vanishing at p, such that M is defined locally by the equation

$$\operatorname{Im} w = \Theta(z, \bar{z}, \operatorname{Re} w),$$

where $\Theta(z, \bar{z}, s)$ is a real-valued, real-analytic function such that

$$\Theta(z, 0, s) \equiv \Theta(0, \bar{z}, s) \equiv 0.$$

Such coordinates are called *normal coordinates* for M at p, and are not unique. M is said to be of *finite type* at p if $\Theta(z, \bar{z}, 0) \not\equiv 0$; otherwise M is of *infinite type* at p. This definition is equivalent to being of finite type in the sense of [Kohn 1972] and [Bloom and Graham 1977]. For real-analytic hypersurfaces, it is also equivalent to minimality — indeed, if M is of infinite type at p, then (in normal coordinates) M contains the nontrivial complex hypersurface $\Sigma = \{w = 0\}$. (For details see [Baouendi et al. 1999b, Chapter I], for example.)

In this paper, we shall focus our attention on 1-infinite type points p of a real-analytic hypersurface $M \subset \mathbb{C}^2$, i.e., points at which the normal coordinates above satisfy the additional condition that $\Theta_s(z, \bar{z}, 0) \not\equiv 0$. (See Section 2 for precise definitions.) Our main result gives rational dependence of a formal equivalence between 1-infinite type hypersurfaces on its jet of a predetermined order.

Theorem 1.1. *Let $M \subset \mathbb{C}^2$ be a real-analytic hypersurface, and suppose $p \in M$ is of 1-infinite type. There exists an integer k such that, given any hypersurface $\hat{M} \subset \mathbb{C}^2$ with $(\hat{M}, \hat{p})$ formally equivalent to (M, p), there exists a formal power series of the form*

$$\Psi(Z; \Lambda) = \sum_{\alpha} \frac{p_\alpha(\Lambda)}{q(\Lambda)^{\ell_\alpha}} (Z - p)^\alpha, \tag{1}$$

where p_α, q are (respectively) $\mathbb{C}^2$- and $\mathbb{C}$-valued polynomials on the jet space $J^k(\mathbb{C}^2, \mathbb{C}^2)_{p,\hat{p}}$ and the ℓ_α are nonnegative integers, such that any formal equivalence $H : (M, p) \to (\hat{M}, \hat{p})$ satisfies

$$q\big(j_p^k(H)\big) = \det\Big(\frac{\partial H}{\partial Z}(p)\Big) \neq 0 \quad \textit{and} \quad H(Z) = \Psi\big(Z; j_p^k(H)\big).$$

Our proof (presented in Section 5) will actually give a constructive process for determining such an k.

Theorem 1.1 has a number of applications. The first states that any formal equivalence between two germs of 1-infinite type hypersurfaces (M, p) and $(\hat{M}, \hat{p})$ is determined by finitely many derivatives at p.

Theorem 1.2. *Let (M, p) and k be as in Theorem 1.1. If $H^1, H^2 : (M, p) \to (\hat{M}, \hat{p})$ are formal equivalences and*

$$\frac{\partial^{|\alpha|} H^1}{\partial Z^\alpha}(p) = \frac{\partial^{|\alpha|} H^2}{\partial Z^\alpha}(p) \quad \text{for all } |\alpha| \le k,$$

then $H^1 = H^2$ as power series.

Our second application deals with the structure of jets of formal equivalences in the jet space $J^k(\mathbb{C}^2, \mathbb{C}^2)_{p,\hat{p}}$, or rather in the submanifold $G^k(\mathbb{C}^2)_{p,\hat{p}}$ of jets of invertible maps taking $(\mathbb{C}^2, p)$ to $(\mathbb{C}^2, \hat{p})$. We shall denote by $\mathcal{F}(M, p; \hat{M}, \hat{p})$ the set of formal equivalences taking (M, p) into $(\hat{M}, \hat{p})$.

Theorem 1.3. *Let (M, p) and k be as in Theorem 1.1. Then for any (germ of a) real-analytic hypersurface $(\hat{M}, \hat{p})$ in $\mathbb{C}^2$, the mapping*

$$j_p^k : \mathcal{F}(M, p; \hat{M}, \hat{p}) \to G^k(\mathbb{C}^2)_{p,\hat{p}}$$

is an injection onto a real algebraic submanifold of $G^k(\mathbb{C}^2)_{p,\hat{p}}$.

Of special interest is the case $(\hat{M}, \hat{p}) = (M, p)$, since $\mathcal{F}(M, p; \hat{M}, \hat{p})$ becomes a group under composition, called the *formal stability group* of M at p and denoted by $\mathrm{Aut}(M, p)$. We shall denote by $G^k(\mathbb{C}^2)_p := G^k(\mathbb{C}^2)_{p,p}$ the *k-jet group* of $\mathbb{C}^2$ at p. The following result is then a corollary of Theorem 1.3.

Theorem 1.4. *Let (M, p) and k be as in Theorem 1.1. Then the mapping*

$$j_p^k : \mathrm{Aut}(M, p) \to G^k(\mathbb{C}^2)_p$$

defines an injective group homomorphism onto a real algebraic Lie subgroup of $G^k(\mathbb{C}^2)_p$.

The study of the (formal) transformation groups of hypersurfaces in $\mathbb{C}^N$ has a long history. Its roots can be traced back to E. Cartan, who studied the structure of the local transformation groups of smooth Levi nondegenerate hypersurfaces in $\mathbb{C}^2$ in [Cartan 1932a; 1932b]. These results were later extended to higher dimensions by Chern and Moser in [Chern and Moser 1974], who also proved the finite determination of such equivalences by their 2-jets.

Further results about the transformation groups of various classes of finite type generic submanifolds of $\mathbb{C}^N$ have been obtained more recently by a number of mathematicians. Regarding the parametrization of transformation groups, we mention the work of Zaitsev [1997], and Baouendi, Ebenfelt, and Rothschild [Baouendi et al. 1999a], which presents modified versions of Theorems 1.2–1.4 valid for smooth generic submanifolds M, $\hat{M}$ in $\mathbb{C}^N$ with M of finite type and $\hat{M}$ finitely nondegenerate. Moreover, there exist a number of results concerning the finite determination of local equivalences addressed in Theorem 1.2. We mention the work of Baouendi, Mir, and Rothschild [Baouendi et al. 2002], which gives the best finite determination results to date for the general case of finite type submanifolds in $\mathbb{C}^N$, and Ebenfelt, Lamel, and Zaitsev [Ebenfelt et al. 2003], which addresses the case $\mathbb{C}^2$ specifically, proving that the local equivalences between any two nonflat real-analytic hypersurface are determined by a finite jet. The reader interested in other recent work on these problems is directed to the excellent survey articles [Rothschild 2003] and [Zaitsev 2002].

For the proofs of the four theorems above, it is convenient to work with formal mappings between formal real hypersurfaces. Hence, the results presented here will be reformulated and proved in this more general context. The following section presents the necessary preliminaries and definitions. In what follows, the distinguished points p and $\hat{p}$ on M and $\hat{M}$, respectively, will, for convenience and without loss of generality, be assumed to be 0.

2. Preliminaries and basic definitions

Formal mappings and hypersurfaces. Let $X = (X_1, \dots, X_N)$ be a N-tuple of indeterminates, and let $\mathcal{R}$ denote a commutative ring with unity. We define

- $\mathcal{R}[\![X]\!]$:= the ring of formal power series in X with coefficients in $\mathcal{R}$;
- $\mathcal{R}[X]$:= the ring of polynomials in X with coefficients $\mathcal{R}$.

For $\mathcal{R} = \mathbb{C}$, we shall also define

- $\mathbb{C}\{X\}$:= the ring of convergent power series in X with coefficients in $\mathbb{C}$;
- $\mathcal{O}_\epsilon(X)$:= the ring of power series in X with coefficients in $\mathbb{C}$ that converge for $X_j \in \mathbb{C},\ |X_j| < \epsilon,\ 1 \le j \le N$.

We have canonical embeddings

$$\mathbb{C}[X] \subset \mathcal{O}_\epsilon(X) \subset \mathbb{C}\{X\} \subset \mathbb{C}[\![X]\!].$$

A power series $\rho \in \mathbb{C}[\![Z, \zeta]\!]$, where $Z = (Z_1, \dots, Z_N)$ and $\zeta = (\zeta_1, \dots, \zeta_N)$, is called *real* if $\rho(Z, \zeta) = \bar{\rho}(\zeta, Z)$, where $\bar{\rho}$ denotes the power series obtained by replacing the coefficients of ρ by their complex conjugates. If, in addition, the power series ρ satisfies the conditions

$$\rho(0) = 0, \quad d\rho(0) \neq 0, \tag{2}$$

we say that ρ defines a *formal real hypersurface* M of $\mathbb{C}^N$ through 0, and we write

$$M = \{\rho(Z, \bar{Z}) = 0\}$$

and say that the pair $(M, 0)$ is a *formal real hypersurface*. The function ρ is a *formal defining function* for M. The reader should observe that if M is a formal real hypersurface in $\mathbb{C}^N$ with formal defining function ρ, then in general there is no actual point set $M \subset \mathbb{C}^N$.

Suppose that $\hat{\rho}$ is another formal power series (not necessarily real) satisfying conditions (2). If there exists a power series $a(Z, \zeta)$ (necessarily invertible at 0) such that

$$\hat{\rho}(Z, \zeta) = a(Z, \zeta)\,\rho(Z, \zeta),$$

then we say that $\hat{\rho}$ also defines the formal real hypersurface M, and again we write $M = \{\hat{\rho}(Z, \bar{Z}) = 0\}$.

By a formal mapping $H : (\mathbb{C}^N, 0) \to (\mathbb{C}^N, 0)$, denoted $H \in E(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$, we shall mean an element $H \in \mathbb{C}[\![Z]\!]^N$ such that $H(0) = 0$. We say H is a *formal change of coordinates* if it is formally invertible, i.e., if there exists a formal map $H^{-1} : (\mathbb{C}^N, 0) \to (\mathbb{C}^N, 0)$ such that

$$H(H^{-1}(Z)) \equiv H^{-1}(H(Z)) \equiv Z$$

as formal power series. As noted in the introduction, H is a formal change of coordinates in $\mathbb{C}^N$ if and only if its Jacobian at 0 is nonzero.

Given a formal change of coordinates H in $\mathbb{C}^N$, we define its corresponding *formal holomorphic change of variable* by

$$Z = H(Z'), \quad \zeta = \overline{H}(\zeta').$$

If $M = \{\rho(Z, \bar{Z}) = 0\}$ is a formal real hypersurface of $\mathbb{C}^N$, we say M is expressed in the Z' coordinates by $\{\rho(H(Z'), \overline{H(Z')}) = 0\}$.

If $\hat{M} = \{\hat{\rho}(Z, \bar{Z}) = 0\}$ is another formal real hypersurface of $\mathbb{C}^N$, then a formal mapping $H \in E(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$ is said to *take M into $\hat{M}$* if there exists a power series $c(Z, \zeta)$ (not necessarily invertible at 0) such that

$$\hat{\rho}\big(H(Z), \overline{H}(\zeta)\big) = c(Z, \zeta)\,\rho(Z, \zeta).$$

In this situation we write as $H : (M, 0) \to (\hat{M}, 0)$. This definition is independent of the power series used to define M and $\hat{M}$.

If $H : (M, 0) \to (\hat{M}, 0)$ is as above and H is invertible, it follows that H^{-1} takes $\hat{M}$ into M. In this case we say that M and $\hat{M}$ are *formally equivalent*, and that H is a *formal equivalence* between them, denoted $H \in \mathcal{F}(M, 0; \hat{M}, 0)$.

The motivation behind these definitions is the following. If the formal series ρ defining the formal real hypersurface M is actually convergent, then the equation $\rho(Z, \bar{Z}) = 0$ defines a real-analytic hypersurface M of $\mathbb{C}^N$ passing through the origin. Moreover, if $H : \mathbb{C}^N \to \mathbb{C}^N$ is a holomorphic mapping with $H(0) = 0$, and M, $\hat{M}$ are both real-analytic hypersurfaces of $\mathbb{C}^N$, then $H(M) \subset \hat{M}$ if and only if the formal mapping H maps the formal real hypersurface M into the formal real hypersurface $\hat{M}$.

For each positive integer k, we denote by $J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$ the jet space of order k of (formal) holomorphic mappings $(\mathbb{C}^N, 0) \to (\mathbb{C}^N, 0)$, and by $j_0^k : E(\mathbb{C}^N, \mathbb{C}^N) \to J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$ the corresponding jet mapping taking a formal mapping H to its k-jet at 0, $j_0^k(H)$. We denote by $G^k(\mathbb{C}^N)_0 \subset J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$ the collection of k-jets of invertible formal mappings of $(\mathbb{C}^N, 0)$ to itself.

Given coordinates Z and $\hat{Z}$ on $\mathbb{C}^N$, we may identify the jet space $J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$ with the set of degree-k polynomial mappings of $(\mathbb{C}^N, 0) \to (\mathbb{C}^N, 0)$. The coordinates on $J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$, which we denote by Λ, can then be taken to be the coefficients of these polynomials. Formal changes of coordinates in $\mathbb{C}^N$ yield polynomial changes of coordinates in $J^k(\mathbb{C}^N, \mathbb{C}^N)_{0,0}$.

If M is a formal real hypersurface in $\mathbb{C}^N$, there is a formal change of coordinates $Z = (z, w) \in \mathbb{C}[\![z, w]\!]^N$ with $z = (z_1, \dots, z_{N-1})$, such that M, under the corresponding formal holomorphic change of variable $Z = Z(z, w)$, $\zeta = \bar{Z}(\chi, \tau)\big)$, is defined by

$$\rho(z, w, \chi, \tau) := \left(\frac{w-\tau}{2i}\right) - \Theta\left(z, \chi, \frac{w+\tau}{2}\right) \in \mathbb{C}[\![Z, \zeta]\!],$$

where $\Theta \in \mathbb{C}[\![z, \chi, s]\!]$ is real and satisfies $\Theta(z, 0, s) = \Theta(0, \chi, s) = 0$. Such coordinates are called *normal coordinates* for M; see [Baouendi et al. 1999b, Chapter IV].

Using the formal Implicit Function Theorem to solve for w above, we see that there exists a unique formal power series $Q \in \mathbb{C}[\![z, \chi, \tau]\!]$ with $Q(0, 0, 0) = 0$ such that $\rho\big(z, Q(z, \chi, \tau), \chi, \tau\big) \equiv 0$; moreover, Q is convergent whenever Θ is. This implies that there exists a power series $a(z, w, \chi, \tau)$, nonvanishing at 0, such that $\rho(z, w, \chi, \tau) = a(z, w, \chi, \tau)\big(w - Q(z, \chi, \tau)\big)$; whence we may write (abusing notation)

$$M = \left\{\left(\frac{w-\overline{w}}{2i}\right) = \Theta\left(z, \bar{z}, \frac{w+\overline{w}}{2}\right)\right\} = \big\{w = Q(z, \bar{z}, \overline{w})\big\}. \tag{3}$$

Observe that the normality of the coordinates implies $Q(z, 0, \tau) = Q(0, \chi, \tau) = \tau$.

Given normal coordinates $Z = (z, w)$ for M as above, define the numbers $m, r, L, K \in \{0, 1, 2, \dots\} \cup \{\infty\}$ as follows. Set

$$m := \sup\{q : \Theta_{s^j}(z, \chi, 0) \equiv 0 \text{ for all } j < q\}. \tag{4}$$

If $m = \infty$ (i.e., if $\Theta \equiv 0$), then set $r = L = K = \infty$. Otherwise, set

$$r := \sup\big\{q : \Theta_{z^\alpha \chi^\beta s^m}(0, 0, 0) = 0 \text{ for all } |\alpha| + |\beta| < q\big\}, \tag{5}$$

$$L := \sup\big\{q : \Theta_{\chi^\beta s^m}(z, 0, 0) \equiv 0 \text{ for all } |\beta| < q\big\}, \tag{6}$$

$$K := \sup\big\{q : \Theta_{z^\alpha \chi^\beta s^m}(0, 0, 0) = 0 \text{ for all } |\alpha| < q,\ |\beta| = L\big\}. \tag{7}$$

We shall show in Theorem 2.1 that this 4-tuple of numbers is independent of the normal coordinates used to define them.

We say that M is of *finite type* at 0 if $m = 0$; otherwise M is of *infinite type* at 0. If we wish to emphasize the number $m \geq 1$, we shall say that M is of *m-infinite type* at 0 if $m < \infty$, and is *flat* at 0 if $m = \infty$. We shall further say M is of *finite type r* at 0 if $m = 0$, and is of *m-infinite type r* at 0 if $1 \leq m < \infty$.

We conclude these definitions by stating a few known results concerning these numbers in the case when M is a real-analytic hypersurface in $\mathbb{C}^N$. In this case, it is known that the pair (m, r) is a biholomorphic invariant of M; see [Meylan 1995]. If M is of infinite type at 0, it contains a formal complex hypersurface Σ passing through 0. (In normal coordinates, we may take $\Sigma = \{w = 0\}$.) In fact, $m > 0$ is constant along the complex hypersurface $\Sigma \subset M$ through 0. And while r is not constant along Σ, there exists a proper, real-analytic subvariety $V \subset \Sigma$ outside of which all points are of m-infinite type 2. See [Ebenfelt 2002] for details.

Statement of results. Our first result shows that the 4-tuple (m, r, L, K) (and hence the notion of being m-infinite type r at a point) is in fact a *formal* invariant of a hypersurface.

Theorem 2.1. *Let $(M, 0)$ be a formal real hypersurface of $\mathbb{C}^N$. Then the numbers (m, r, L, K) are independent of the choice of normal coordinates used to define them. Moreover, if $(\hat{M}, 0)$ is formally equivalent to $(M, 0)$ and has the corresponding 4-tuple $(\hat{m}, \hat{r}, \hat{L}, \hat{K})$, then $(m, r, L, K) = (\hat{m}, \hat{r}, \hat{L}, \hat{K})$.*

We shall then focus on the case $N = 2$ and $m = 1$. We may now state the generalizations of Theorems 1.1 through 1.4 valid for formal real hypersurfaces. Our main result is the following.

Theorem 2.2. *Let $(M, 0)$ be a formal real hypersurface in $\mathbb{C}^2$ of 1-infinite type. There exists an integer k such that given any formal real hypersurface $(\hat{M}, 0)$ in $\mathbb{C}^2$ formally equivalent to $(M, 0)$, there exists a formal power series of the form*

$$\Psi(Z; \Lambda) = \sum_{\alpha} \frac{p_\alpha(\Lambda)}{q(\Lambda)^{\ell_\alpha}} Z^\alpha, \tag{8}$$

where p_α, q are (respectively) $\mathbb{C}^2$- and $\mathbb{C}$-valued polynomials on the jet space $J^k(\mathbb{C}^2, \mathbb{C}^2)_{0,0}$ and the ℓ_α are nonnegative integers, such that any formal equivalence $H \in \mathcal{F}(M, 0; \hat{M}, 0)$ satisfies

$$q\big(j_0^k(H)\big) = \det\Big(\frac{\partial H}{\partial Z}(0)\Big) \neq 0, \qquad H(Z) = \Psi\big(Z; j_0^k(H)\big).$$

It is clear from the remarks made in the previous section that Theorem 2.2 is a more general version of Theorem 1.1 from the introduction. As a consequence of this result, we have the following, from which Theorem 1.2 is derived.

Theorem 2.3. *Let $(M, 0)$ be a formal real hypersurface in $\mathbb{C}^2$ of 1-infinite type, and let k be the number described in Theorem 2.2. If $(\hat{M}, 0)$ is a formal hypersurface formally equivalent to $(M, 0)$, and $H^1, H^2 : (M, 0) \to (\hat{M}, 0)$ are formal equivalences such that*

$$\frac{\partial^{|\alpha|} H^1}{\partial Z^\alpha}(0) = \frac{\partial^{|\alpha|} H^2}{\partial Z^\alpha}(0) \quad \textit{for all } |\alpha| \le k,$$

then $H^1 = H^2$ as power series.

We shall then prove the following generalization of Theorem 1.4.

Theorem 2.4. *Let M and k be as in Theorem 2.2. The mapping*

$$j_0^k : \mathrm{Aut}(M, 0) \to G^k(\mathbb{C}^2)_0$$

defines an injective group homomorphism onto a real algebraic Lie subgroup of $G^k(\mathbb{C}^2)_0$.

The following generalization of Theorem 1.3 is a consequence of Theorem 2.4.

Theorem 2.5. *Let M and k be as in Theorem 2.2. For any formal real hypersurface $\hat{M}$ in $\mathbb{C}^2$, the mapping*

$$j_0^k : \mathcal{F}(M, 0; \hat{M}, 0) \to J^k(\mathbb{C}^2)_0$$

is an injection onto a real algebraic submanifold of $G^k(\mathbb{C}^2)_0$.

3. Formal invariance of type conditions

In this section, we shall prove Theorem 2.1, or rather a slightly sharper statement of which Theorem 2.1 is an immediate consequence:

Proposition 3.1. *Let $(M, 0)$ be a formal real hypersurface in $\mathbb{C}^N$, given in normal coordinates $Z = (z, w)$ by Equation (3). Let $(\hat{M}, 0)$ be a formal real hypersurface in $\mathbb{C}^N$, given in normal coordinates $\hat{Z} = (\hat{z}, \hat{w})$ by the corresponding "hatted" defining functions*:

$$\hat{M} = \left\{ \frac{\hat{w} - \bar{\hat{w}}}{2i} = \hat{\Theta}\Big(\hat{z}, \bar{\hat{z}}, \frac{\hat{w} + \bar{\hat{w}}}{2}\Big) \right\} = \left\{ \hat{w} = \hat{Q}(\hat{z}, \bar{\hat{z}}, \bar{\hat{w}}) \right\}.$$

Define as in Section 2 the 4-tuple (m, r, L, K) for M and the corresponding 4-tuple $(\hat{m}, \hat{r}, \hat{L}, \hat{K})$ for $\hat{M}$. If M and $\hat{M}$ are formally equivalent, then $(m, r, L, K) = (\hat{m}, \hat{r}, \hat{L}, \hat{K})$.

We begin with a useful lemma concerning the form of formal mappings in normal coordinates. It is proved in the same way as [Baouendi et al. 1999b, Lemma 9.4.4].

Lemma 3.2. *Let M, $\hat{M}$ be formal hypersurfaces in $\mathbb{C}^N$ through 0, expressed in normal coordinates as in Proposition 3.1. If $H = (F, G) : (M, 0) \to (\hat{M}, 0)$ is a formal mapping, then $G(z, w) = w\, g(z, w)$ for some $g \in \mathbb{C}[[z, w]]$. Moreover, if H is a formal equivalence, then $F(z, 0) \in \mathbb{C}[[z]]^{N-1}$ is a formal equivalence, and $g(0, 0) \neq 0$.*

As a consequence of this lemma, we shall henceforth write formal equivalences (in suitable normal coordinates) as

$$H(z, w) = \big(f(z, w), w\, g(z, w)\big), \tag{9}$$

with $f = (f^1, \dots, f^{N-1}) \in \mathbb{C}[[z, w]]^{N-1}$ satisfying $\det f_z(0, 0) \neq 0$ and $g \in \mathbb{C}[[z, w]]$ satisfying $g(0, 0) \neq 0$. Observe that the condition that H map M formally into $\hat{M}$ may be written as

$$Q(z, \chi, \tau)\, g\big(z, Q(z, \chi, \tau)\big) \equiv \hat{Q}\big(f\big(z, Q(z, \chi, \tau)\big), \bar{f}(z, \chi), \tau\, \bar{g}(\chi, \tau)\big). \tag{10}$$

Moreover, for convenience, we shall formally expand f and g as

$$f(z, w) = \sum_{n \geq 0} \frac{f_n(z)}{n!}\, w^n, \quad g(z, w) = \sum_{n \geq 0} \frac{g_n(z)}{n!}\, w^n. \tag{11}$$

The main technical lemma in the proof of Proposition 3.1 is the following.

Lemma 3.3. *Suppose M, $\hat{M}$ are formal hypersurfaces in $\mathbb{C}^N$ through 0, expressed in normal coordinates as in Proposition 3.1, and assume that $H : (M, 0) \to (\hat{M}, 0)$ is a formal equivalence. Then for every $j \geq 0$, if*

$$\hat{Q}(\hat{z}, \hat{\chi}, 0) \equiv \hat{Q}_{\hat{\tau}}(\hat{z}, \hat{\chi}, 0) - 1 \equiv \hat{Q}_{\hat{\tau}^2}(\hat{z}, \hat{\chi}, 0) \equiv \cdots \equiv \hat{Q}_{\hat{\tau}^j}(\hat{z}, \hat{\chi}, 0) \equiv 0,$$

then

$$Q(z, \chi, 0) \equiv Q_\tau(z, \chi, 0) - 1 \equiv Q_{\tau^2}(z, \chi, 0) \equiv \cdots \equiv Q_{\tau^j}(z, \chi, 0) \equiv 0. \tag{12}$$

Moreover, $g_0(z), g_1(z), \dots, g_j(z)$ are all real constants (with $g_0(z)$ nonzero), and

$$Q_{\tau^{j+1}}(z, \chi, 0) \equiv g(0)^j\, \hat{Q}_{\hat{\tau}^{j+1}}\big(f_0(z), \bar{f}_0(\chi), 0\big).$$

To prove Lemma 3.3, we make use of two results. The first is a generalization of the Chain Rule due to Faa de Bruno; see [Range 1986], for example:

Lemma 3.4 (Faa de Bruno's Formula). *Suppose that $f = \big(f_1, f_2, \dots, f_\ell\big) \in \mathbb{C}^\ell[[z]]$ with $z \in \mathbb{C}$ and $f(0) = 0$, and suppose $h(z_1, z_2, \dots, z_\ell) \in \mathbb{C}[[z_1, z_2, \dots, z_\ell]]$. Then*

$$\frac{\partial^v}{\partial z^v}\{h\big(f(z)\big)\} = \sum_{\substack{[\alpha^1]+[\alpha^2]+\cdots \\ +[\alpha^\ell]=v}} \frac{v!\, h_{z_1{}^{|\alpha^1|} z_2{}^{|\alpha^2|} \cdots z_\ell{}^{|\alpha^\ell|}}\big(f(z)\big)}{\alpha^1!\, \alpha^2! \cdots \alpha^\ell!} \prod_{\substack{1 \leq q \leq v \\ 1 \leq p \leq \ell}} \left(\frac{f_p{}^{(q)}(z)}{q!}\right)^{\alpha^p_q},$$

where each $\alpha^p = (\alpha_1^p, \dots, \alpha_v^p)$ *denotes an* v*-dimensional multi-index, and*

$$|\alpha^p| = \sum_{q=1}^{v} \alpha_q^p, \quad [\alpha^p] = \sum_{q=1}^{v} q\,\alpha_q^p, \quad \alpha^p! = \prod_{q=1}^{v} (\alpha_q^p)!$$

The proof is a routine induction, and is left to the reader. The other result we shall need gives a second characterization of the number m:

Proposition 3.5 [Baouendi and Rothschild 1991, Proposition 1.7]. *Let* M, m, Θ, *and* Q *be as above. Then*

$$m = \sup\Big\{q : \frac{\partial^j}{\partial \tau^j}\{Q(z, \chi, \tau) - \tau\}\Big|_{\tau=0} \equiv 0 \text{ for all } j < q\Big\}.$$

Furthermore,

$$Q_{\tau^m}(z, \chi, 0) = \begin{cases} \dfrac{1 + i\,\Theta_s(z, \chi, 0)}{1 - i\,\Theta_s(z, \chi, 0)} & \text{if } m = 1, \\ 2i\,\Theta_{s^m}(z, \chi, 0) & \text{if } 2 \le m < \infty. \end{cases}$$

Proof of Lemma 3.3. Differentiating identity (10) v times in τ, setting $\tau = 0$, and canceling $v!$ from both sides yields the identity

$$\text{(13)} \quad \sum_{k+[\xi]=v} \frac{g_{|\xi|}(z)\,Q_{\tau^k}(z, \chi, 0)}{k!\,\xi!} \prod_{p=1}^{v} \left(\frac{Q_{\tau^p}(z, \chi, 0)}{p!}\right)^{\xi_p}$$
$$\equiv \sum_{\substack{[\alpha^1]+\cdots+[\alpha^n]+[\beta^1]+\cdots \\ \cdots+[\beta^n]+[\gamma]=v}} \frac{\hat{Q}_{\hat{z}^{(|\alpha^1|,\dots,|\alpha^n|)}\hat{\chi}^{(|\beta^1|\cdots,|\beta^n|)}\hat{\tau}^{|\gamma|}}\big(f_0(z), \bar{f}_0(\chi), 0\big)}{\alpha^1!\cdots\alpha^n!\,\beta^1!\cdots\beta^n!\,\gamma!}$$
$$\times \prod_{\substack{1\le q\le v \\ 1\le u\le n}} \left(\sum_{[\eta]=q} \frac{f^u_{|\eta|}(z)}{\eta!} \prod_{r=1}^{q} \left(\frac{Q_{\tau^r}(z, \chi, 0)}{r!}\right)^{\eta_r}\right)^{\alpha_q^u} \left(\frac{\overline{f_q^u}(\chi)}{q!}\right)^{\beta_q^u} \left(\frac{\overline{g_{q-1}}(\chi)}{(q-1)!}\right)^{\gamma_q}.$$

We now proceed by induction. For $j = 0$, we assume only that $\hat{Q}(\hat{z}, \hat{\chi}, 0) \equiv 0$. Setting $\tau = 0$ in identity (10), we find

$$Q(z, \chi, 0)\; g\big(z, Q(z, \chi, 0)\big) \equiv \hat{Q}\big(f_0(z), \bar{f}_0(\chi), 0\big) = 0.$$

Since $g(z, Q(z, \chi, 0))$ does not vanish at $z = \chi = 0$, we conclude $Q(z, \chi, 0) \equiv 0$.

Applying the $v = 1$ case of identity (13), we find

$$Q_\tau(z, \chi, 0)\, g_0(z) \equiv \hat{Q}_{\hat{\tau}}\big(f_0(z), \bar{f}_0(\chi), 0\big)\,\bar{g}_0(\chi).$$

Setting $\chi = 0$ yields $g_0(z) \equiv \bar{g}_0(0) = \overline{g_0(0)}$, whence $g_0(z)$ is a real constant r, and since H is invertible, $r \ne 0$ necessarily. Dividing $g_n(z) = \bar{g}_0(\chi) = r \ne 0$ from both sides of the identity above yields

$$Q_\tau(z, \chi, 0) \equiv \hat{Q}_{\hat\tau}\big(f_0(z), \bar{f}_0(\chi), 0\big),$$

which proves the $j = 0$ case.

Now, assume that the lemma holds for some $j - 1 \ge 0$; we shall prove it for j. Suppose that (12) holds. By induction, we know that

$$Q(z, \chi, 0) \equiv Q_\tau(z, \chi, 0) - 1 \equiv Q_{\tau^2}(z, \chi, 0) \equiv \cdots \equiv Q_{\tau^{j-1}}(z, \chi, 0) \equiv 0,$$

that $g_0, g_1, \ldots, g_{j-1}$ are constant functions, and that

$$Q_{\tau^j}(z, \chi, 0) \equiv r^{j-1}\, \hat{Q}_{\hat\tau^j}\big(f_0(z), \bar{f}_0(\chi), 0\big).$$

In the $j = 1$ case, this implies $Q_\tau(z, \chi, 0) \equiv 1$; otherwise it implies $Q_{\tau^j}(z, \chi, 0) \equiv 0$, as desired.

Substituting these values into identity (13) (with $v = j + 1$), we obtain

$$r\, Q_{\tau^{j+1}}(z, \chi, 0) + (j+1)\, g_j(z) \equiv r^{j+1}\, \hat{Q}_{\hat\tau^{j+1}}\big(f_0(z), \bar{f}_0(\chi), 0\big) + (j+1)\, \bar{g}_j(\chi).$$

Setting $\chi = 0$ yields

$$(j+1)\, g_j(z) = (j+1)\, \bar{g}_j(0) = (j+1)\overline{g_j(0)},$$

so $g_j(z)$ is a real constant. Subtracting $(j+1)\, g_j(z)$ from both sides and dividing by $r \ne 0$ completes the induction. □

Corollary 3.6. *Let M, $\hat{M}$ be formal real submanifolds of $\mathbb{C}^N$ through 0, given in normal coordinates as in Proposition 3.1. Define m for M and the corresponding $\hat{m}$ for $\hat{M}$. If M and $\hat{M}$ are formally equivalent, then $m = \hat{m}$.*

Proof. Lemma 3.3 implies $m \ge \hat{m}$. Then reverse the roles of M and $\hat{M}$. □

We shall be primarily interested in formal real hypersurfaces which are of infinite type, but nonflat, at 0. That is, formal hypersurfaces of m-infinite type for some positive integer m. In this case, Corollary 3.6 may be strengthened as follows.

Proposition 3.7. *If M is of m-infinite type at 0 and $H \in \mathcal{F}(M, 0; \hat{M}, 0)$, then $\hat{M}$ is of m-infinite type at 0, $g_0, g_1, \ldots, g_{m-1}$ are constant, and*

$$0 \not\equiv \Theta_{s^m}(z, \chi, 0) \equiv g_0(0)^{m-1}\, \hat{\Theta}_{\hat{s}^m}\big(f_0(z), \bar{f}_0(\chi), 0\big).$$

Proof. Put together Lemma 3.3, Corollary 3.6, and Proposition 3.5. □

We now have the necessary ingredients to prove Proposition 3.1.

Proof of Proposition 3.1. We have seen that $m = \hat{m}$. If the hypersurfaces are of finite type, then it is well known that the triple (r, L, K) is a formal invariant. (An outline of the proof that r is a formal invariant, for example, may be found in

[Baouendi et al. 1999b, Chapter I].) Similarly, $r = \infty$ if and only if $m = \hat{m} = \infty$, which in turn holds if and only if $\hat{r} = \infty$; and likewise if $L = \infty$ or $K = \infty$. Hence, it suffices to assume that all the numbers in question are positive integers. By Proposition 3.7, we have

$$0 \not\equiv \Theta_{s^m}(z, \chi, 0) \equiv g_0(0)^{m-1}\, \hat{\Theta}_{\hat{s}^m}\big(f_0(z), \bar{f}_0(\chi), 0\big).$$

A straightforward induction using Faa de Bruno's formula implies that for any multi-indices α and β,

$$\Theta_{z^\alpha \chi^\beta s^m}(z, \chi, 0) = g_0(0)^{m-1} \sum_{\substack{|\mu|\le|\alpha| \\ |\nu|\le|\beta|}} \hat{\Theta}_{\hat{z}^\mu \hat{\chi}^\nu \hat{s}^m}\big(f_0(z), \bar{f}_0(\chi), 0\big) \times P^{\alpha\beta}_{\mu\nu}\Big(\big((f_0^u)_{z^\gamma}(z)\big)_{|\gamma|\le|\mu|}, \big((\bar{f}_0^u)_{\chi^\delta}(\chi)\big)_{|\delta|\le|\nu|}\Big),$$

where each $P^{\alpha\beta}_{\mu\nu}$ is a polynomial in its arguments.

This implies that $\Theta_{z^\alpha \chi^\beta s^m}(0, 0, 0) = 0$ whenever $|\alpha| + |\beta| < \hat{r}$, whence $r \ge \hat{r}$ necessarily. Reversing the roles of M and $\hat{M}$ yields $r = \hat{r}$. Similarly, the equality of r and $\hat{r}$ then implies that $\Theta_{\chi^\beta s^m}(z, 0, 0) \equiv 0$ whenever $|\beta| < \hat{L}$, whence $L \ge \hat{L}$; reversing the roles of the formal hypersurfaces establishes equality. The proof that $K = \hat{K}$ is similar, and is left to the reader. □

4. The 1-infinite type case in $\mathbb{C}^2$

Notation and results. From now on we deal only with formal real hypersurfaces of $\mathbb{C}^2$, and in particular those hypersurfaces that are of 1-infinite type at 0. Suppose that M is such a formal hypersurface. We shall write M in normal coordinates $Z = (z, w)$ as in (3). Since M is of 1-infinite type, this implies that we can write $Q(z, \chi, \tau) = \tau\, S(z, \chi, \tau)$ for some $S \in \mathbb{C}[[z, \chi, \tau]]$, so that

$$M = \left\{\Big(\frac{w-\overline{w}}{2i}\Big) = \Theta\Big(z, \bar{z}, \frac{w+\overline{w}}{2}\Big)\right\} = \big\{w = \overline{w}\, S(z, \bar{z}, \overline{w})\big\}. \tag{14}$$

For convenience, we shall write

$$\theta(z, \chi) = \sum_{j=0}^{\infty} \frac{\theta_j(z)}{j!}\, \chi^j := \Theta_s(z, \chi, 0) \not\equiv 0 \tag{15}$$

Observe that $\theta_j(z) \equiv 0$ if $j < L$ and $\theta_L^{(j)}(0) = 0$ if $j < K$,where L, K are defined by equations (6) and (7). It will be useful for later computations to observe that Proposition 3.5 implies

$$S(z, \chi, 0) = \frac{1 + i\theta(z, \chi)}{1 - i\theta(z, \chi)}, \tag{16}$$

whence repeated differentiation in χ yields

$$S_{\chi^j}(z,0,0) = \begin{cases} 1 & \text{if } j=0, \\ 0 & \text{if } 1 \le j \le L-1, \\ 2i\,\theta_L(z) & \text{if } j=L, \\ 2i\,\theta_{L+1}(z) - 4\theta_1(z)^2 & \text{if } j=L+1. \end{cases} \tag{17}$$

We now define a new, rather technical, invariant for 1-infinite type hypersurfaces. Letting δ_k^j denote the Kronecker delta function, we define the number $T \in \{0, 1\}$ by

$$T := \prod_{q=0}^{K-2} \delta^0_{\theta^{(q)}_{L+1}(0)}. \tag{18}$$

That is, $T = 1$ if and only if $\theta_{L+1}(z) = O(|z|^{K-1})$; by means similar to the proofs for the numbers r, L, and K, it can be shown that T is a formal invariant. Details are left to the reader.

Now assume that $\hat{M}$ is a formal real hypersurface of $\mathbb{C}^2$ that is formally equivalent to M, and write it in normal coordinates $\hat{Z} = (\hat{z}, \hat{w})$ as

$$\hat{M} = \left\{ \frac{\hat{w} - \bar{\hat{w}}}{2i} = \hat{\Theta}\left(\hat{z}, \bar{\hat{z}}, \frac{\hat{w} + \bar{\hat{w}}}{2}\right) \right\} = \left\{ \hat{w} = \bar{\hat{w}}\,\hat{S}(\hat{z}, \bar{\hat{z}}, \bar{\hat{w}}) \right\}, \tag{19}$$

We write $\hat{\theta}(\hat{z}, \hat{\chi}) := \hat{\Theta}_{\hat{s}}(\hat{z}, \hat{\chi}, 0)$ as above.

If $H : (M, 0) \to (\hat{M}, 0)$ is a formal equivalence, Lemma 3.2 implies that $H(z, w)$ is of the form given by (9), with $f, g \in \mathbb{C}[[z, w]]$ and $f_z(0,0)\,g(0,0) \ne 0$. Observe that identity (10) can be rewritten (after canceling an extra τ from both sides) as

$$\begin{aligned} &S(z,\chi,\tau)\,g\big(z, \tau\,S(z,\chi,\tau)\big) \\ &\qquad \equiv \bar{g}(\chi,\tau)\,\hat{S}\big(f\big(z, \tau\,S(z,\chi,\tau)\big), \bar{f}(z,\chi), \tau\,\bar{g}(\chi,\tau)\big). \end{aligned} \tag{20}$$

We shall continue to use the formal Taylor expansions of f and g in w given by equation (11), and shall write

$$f_n(z) := \sum_{k \ge 0} \frac{1}{k!}\,\overline{a_n^k}\,z^k, \qquad g_n(z) := \sum_{k \ge 0} \frac{1}{k!}\,\overline{b_n^k}\,z^k, \tag{21}$$

where the bar denotes complex conjugation. Note that, in particular, $a_0^0 = 0$, $a_0^1 \ne 0$, and $b_0^0 = \overline{b_0^0} \ne 0$.

Finally, for $n \geq 0$, define the formal rational mapping $\Upsilon^n : (\mathbb{C}^2, 0) \to (\mathbb{C}^4, 0)$ by

$$\Upsilon_1^n(z, \chi) := K \frac{\theta_L(z)}{\theta_L'(z)} \Big(\frac{1+i\theta(z, \chi)}{1-i\theta(z, \chi)}\Big)^n \theta_z(z, \chi) - L \frac{\bar{\theta}_L(\chi)}{\bar{\theta}_L'(\chi)} \theta_\chi(z, \chi),$$

$$\Upsilon_2^n(z, \chi) := (1 + \theta(z, \chi)^2)\bigg(\Big(\frac{1+i\theta(z, \chi)}{1-i\theta(z, \chi)}\Big)^n - 1\bigg) - 2in \frac{\bar{\theta}_L(\chi)}{\bar{\theta}_L'(\chi)} \theta_\chi(z, \chi),$$

$$\begin{aligned}\Upsilon_3^n(z, \chi) := \delta_L^1 \delta_T^1 \Bigg(& \delta_K^1 \theta_1^{(L)}(0) \frac{\theta_\chi(z, \chi, 0)}{\bar{\theta}_1'(\chi)} \\ & + \frac{\theta_1^{(K)}(0)\theta_2^{(K)}(0) - \theta_1^{(K+1)}(0)\theta_2^{(K-1)}(0)}{K\theta_1^{(K)}(0)^2} \frac{\bar{\theta}_1(\chi)}{\bar{\theta}_1'(\chi)} \theta_\chi(z, \chi) \\ & - \Big(\frac{1+i\theta(z, \chi)}{1-i\theta(z, \chi)}\Big)^n \bigg(\theta_1(z)\big(1 + \theta(z, \chi)^2\big) + \bigg(\frac{\theta_2(z)}{\theta_1'(z)} - 2in \frac{\theta_1(z)^2}{\theta_1'(z)}\bigg)\theta_z(z, \chi)\bigg) \\ & + \frac{\theta_2^{(K-1)}(0)}{\theta_1^{(K)}(0)} \bigg(\bar{\theta}_1(\chi)\big(1 + \theta(z, \chi)^2\big) + \bigg(\frac{\bar{\theta}_2(\chi)}{\bar{\theta}_1'(\chi)} + 2in \frac{\bar{\theta}_1(\chi)^2}{\bar{\theta}_1'(\chi)}\bigg)\theta_\chi(z, \chi)\bigg)\Bigg),\end{aligned}$$

$$\begin{aligned}\Upsilon_4^n(z, \chi) := \delta_K^1 \Bigg(& \frac{\bar{\theta}_1(\chi)}{\theta_1'(0)}\big(1 + \theta(z, \chi)^2\big) - \frac{\theta_z(z, \chi)}{\theta_1'(z)} \Big(\frac{1+i\theta(z, \chi)}{1-i\theta(z, \chi)}\Big)^n \\ & + \frac{\theta_\chi(z, \chi)}{\theta_1'(0)} \bigg(2in \frac{\bar{\theta}_1(\chi)^2}{\bar{\theta}_1'(\chi)} + \frac{\bar{\theta}_2(\chi)}{\bar{\theta}_1'(\chi)} - \frac{\theta_1''(0)}{\theta_1'(0)} \frac{\bar{\theta}_1(\chi)}{\bar{\theta}_1'(\chi)}\bigg)\Bigg),\end{aligned}$$

where the θ_j are defined by (15). We shall prove in the next section that these four equations actually define *formal power series* in (z, χ), rather than quotients of formal power series.

Observe that the formal mapping Υ^n depends on the choice of normal coordinates $Z = (z, w)$ for the formal hypersurface M.

We are now able to state the main technical result of the paper, which may be viewed as a sharper version of Theorem 2.2, but with conjugated derivatives.

Theorem 4.1. *Let* $(M, 0)$ *be a formal real hypersurface in* $\mathbb{C}^2$ *which is of* 1*-infinite type*, *given in normal coordinates* $Z = (z, w)$ *by equation* (14). *Define* $\Upsilon^n(z, \chi)$ *as immediately above*. *For each* $n \in \mathbb{N}$, *define the complex vector space*

$$\mathcal{V}^n(M) := \operatorname{span}_{\mathbb{C}} \big\{ \upsilon_{s,t}^n := \Upsilon^n_{z^s \chi^t}(0, 0) : s, t \in \mathbb{N} \big\} \subset \mathbb{C}^4. \tag{22}$$

Then the dimension of the vector space $\mathcal{V}^n(M)$ *is a formal invariant for each* n, *and the invariant set of integers*

$$\mathfrak{D}(M) := \big\{ n \in \mathbb{N} : \dim_{\mathbb{C}} \mathcal{V}^n(M) < 2 + \delta_K^1 + \delta_L^1 \delta_T^1 \big\} \tag{23}$$

is always finite.

Furthermore, given a formal real hypersurface $(\hat{M}, 0)$ *in* $\mathbb{C}^2$ *formally equivalent to* $(M, 0)$, *normal coordinates* $\hat{Z} = (\hat{z}, \hat{w})$ *for* $\hat{M}$, *and* $n \in \mathbb{N}$, *there exists a formal power series* $\mathcal{A}_n(z; \Delta, \Lambda) \in \mathbb{C}[\Delta, \Lambda][[z]]^2$, *with* $(z, \Delta, \Lambda) \in \mathbb{C} \times \mathbb{C} \times \mathbb{C}^{4|\mathcal{D}(M)|}$, *such that*

$$\big(f_n(z), g_n(z)\big) \equiv \mathcal{A}_n\bigg(z; \frac{1}{a_0^1 b_0^0}, \big(a_j^0, b_j^0, a_j^1, b_j^1\big)_{j \in \mathcal{D}(M)}\bigg).$$

for any $H \in \mathcal{F}(M, 0; \hat{M}, 0)$.

Moreover, if M *and* $\hat{M}$ *are convergent, there exists an* $\epsilon > 0$ *such that the map*

$$z \mapsto \mathcal{A}_n\bigg(z; \frac{1}{a_0^1 b_0^0}, \big(a_j^0, b_j^0, a_j^1, b_j^1\big)_{j \in \mathcal{D}(M)}\bigg)$$

lies in $\mathcal{O}_\epsilon(z)^2$ *for every* $H \in \mathcal{F}(M, 0; \hat{M}, 0)$ *and every* $n \in \mathbb{N}$.

Examples. We now use Theorem 4.1 and Proposition 3.7 to calculate the formal transformation groups of various 1-infinite type hypersurfaces.

Example 4.2. Consider the family of 1-infinite type hypersurfaces

$$M_c^j := \left\{(z, w) : \operatorname{Im} w = c \operatorname{Re} w\, |z|^{2j}\right\}, \quad c \in \mathbb{R} \setminus \{0\}, \quad j \geq 1.$$

Observe that $L = K = j$, $T = 1$, and $\theta(z, \chi) = cz\chi$. If $n > 0$, it can be shown that $\{\upsilon_{2j,2j}^n, \upsilon_{3j,3j}^n\}$ is a basis for $\mathcal{V}^n(M_c^j)$ if $j \geq 2$, and that adding the vectors $\{\upsilon_{2,3}^n, \upsilon_{3,2}^n\}$ extends this to a basis for $\mathcal{V}^n(M_c^1)$. Hence, in any case, we have $\mathcal{D}(M_c^j) = \{0\}$, so any formal equivalence with source M_c^j is determined by (a_0^1, b_0^0).

Applying Proposition 3.7 with $M = \hat{M} = M_c^j$ implies $f_0(z) = \varepsilon z$ for some $\varepsilon \in \mathbb{C}$ with $|\varepsilon| = 1$. It thus follows that

$$\operatorname{Aut}(M_c^j, 0) = \{(z, w) \mapsto (\varepsilon z, r w) : \varepsilon \in \mathbb{C},\ |\varepsilon| = 1,\ r \in \mathbb{R} \setminus \{0\}\}.$$

In particular, every formal automorphism converges.

Observe that for $j \neq k$, the hypersurfaces M_c^j and M_b^k are *not* formally equivalent (Theorem 2.1). On the other hand, M_c^j and M_b^j are formally equivalent if and only if $c/b > 0$. In this case, applying Proposition 3.7 implies that $f_0(z) = \alpha z$ for some $\alpha \in \mathbb{C}$ of modulus $(c/b)^{1/2j}$. It thus follows that

$$\mathcal{F}(M_c^j, 0; M_b^j, 0) = \left\{(z, w) \mapsto \left(\frac{c}{b}\right)^{1/2j} \big(\varepsilon z, r w\big) : \varepsilon \in \mathbb{C},\ |\varepsilon| = 1,\ r \in \mathbb{R} \setminus \{0\}\right\}.$$

Hence, the hypersurfaces M_c^j are formally equivalent if and only if they are biholomorphically equivalent if and only if b and c have the same sign.

Example 4.3. Consider the family of 1-infinite type hypersurfaces

$$N_b^j := \left\{(z, w) : \operatorname{Im} w = 2 \operatorname{Re} w \ \operatorname{Re}(bz\bar{z}^j)\right\}, \quad b \in \mathbb{C} \setminus \{0\},\ j \geq 2.$$

Note $L = 1$, $K = j$, and $\theta(z, \chi) = bz\chi^j + \bar{b}z^j\chi$. If $n > 0$, it can be shown that $\{\upsilon^n_{2,2}, \upsilon^n_{3,2}, \upsilon^n_{3,3}\}$ forms a basis for $\mathcal{V}^n(N_b^j)$, so we again conclude that $\mathfrak{D}(N_b^j) = \{0\}$. Hence, every formal equivalence H with source N_b^j is determined by the values a_0^1 and b_0^0.

Now, Proposition 3.7 applied to the case $M = \hat{M} = N_b^j$ implies that a_0^1 is a $(j-1)$-th root of unity and that $f_0(z) = z/a_0^1$. We conclude that

$$\operatorname{Aut}(N_b^j, 0) = \left\{(z, w) \mapsto (\varepsilon z, rw) : \varepsilon \in \mathbb{C},\ \varepsilon^{j-1} = 1,\ r \in \mathbb{R} \setminus \{0\}\right\}.$$

Note that every formal automorphism converges.

Example 4.4. Consider the hypersurface

$$M := \left\{(z, w) : \operatorname{Im} w = \frac{\operatorname{Re} w\, |z|^2}{1 + \sqrt{1 - |z|^4}}, \quad |z| < 1\right\}.$$

It is easy to check that $L = K = 1$ in this case and that $\mathfrak{D}(M) = \{0, 1, 2\}$. (In fact, $\Upsilon_4^1 \equiv 0$ and $2i\,\Upsilon_1^2 \equiv \Upsilon_2^2$.) A complete calculation of the stability group of this hypersurface is given in [Kowalski 2002b], and reveals it to be a real-analytic hypersurface whose stability group at the origin is determined by 3-jets *but not by 2-jets*.

In fact, this example can be generalized as follows. Define for $k = 2, 3, 4, \ldots$ the set

$$M_k := \left\{(z, w) : \overline{w} = w\bigl(i|z|^2 + \sqrt{1 - |z|^4}\bigr)^{2/k}\right\},$$

where the principal branch of $\zeta \mapsto \zeta^{2/k}$ is meant. A straightforward calculation shows that each M_k defines a real hypersurface and that $M_2 = M$ above. It can also be shown that $\mathfrak{D}(M_k) = \{0, k/2, k\} \cap \mathbb{Z}$, and that the stability group of M_k is determined by $(k + 1)$-jets, *but not by jets of any lesser order*; for details, see [Kowalski 2002a, Chapter 7]. Hence, even though Theorem 4.1 asserts that $\mathfrak{D}(M)$ is always finite, the integers themselves can be arbitrarily large and, consequently, the required jet-order can be as well.

5. Proofs of the main results

Proof of Theorem 4.1. A basic outline of the proof can be divided into four steps.

(1) Given a fixed set of normal coordinates $Z = (z, w)$, we prove that for each $n \in \mathbb{N}$ the power series $f_n(z)$ and $g_n(z)$ are rationally parametrized by the values (a_ℓ^j, b_ℓ^j) for $\ell = 0, 1$ and $0 \leq j \leq n$.

(2) We prove that under these conditions, if $n \notin \mathfrak{D}(M)$, the 4-tuple of complex numbers $(a_n^0, a_n^1, b_n^0, b_n^1)$ is itself a polynomial in $1/(a_0^1 b_0^0)$ and (a_ℓ^j, b_ℓ^j) for $\ell = 0, 1$ and $0 \le j \le n-1$.

(3) We prove that $\mathfrak{D}(M)$, defined by these normal coordinates, is always finite.

(4) We show that the dimension of $\mathcal{V}^n(M)$ (and hence the set $\mathfrak{D}(M)$) is independent of the normal coordinates used to define it.

To fix notation throughout the proof, we assume that M is always given in normal coordinates $Z = (z, w)$ by (14). We also set $\mathfrak{D} = \mathfrak{D}(M)$ and $\mathcal{V}^n = \mathcal{V}^n(M)$. Similarly, $\hat{M}$, whenever a target formal hypersurface is needed, will always be given in normal coordinates $\hat{Z} = (\hat{z}, \hat{w})$ by (19). If $H : (M, 0) \to (\hat{M}, 0)$ is a formal equivalence, we set

$$\begin{aligned}
\Delta(H) &:= \frac{1}{a_0^1 b_0^0} \in \mathbb{C} \setminus \{0\},\\
\lambda_2^n(H) &:= \left(a_n^1, b_n^0\right) \in \mathbb{C}^2,\\
\lambda_3^n(H) &:= \left(a_n^1, b_n^0, a_n^0\right) \in \mathbb{C}^3,\\
\lambda_4^n(H) &:= \left(a_n^1, b_n^0, a_n^0, b_n^1\right) \in \mathbb{C}^4,\\
\Lambda_j^n(H) &:= \left(\lambda_j^0(H), \lambda_j^1(H), \ldots, \lambda_j^n(H)\right) \in \mathbb{C}^{j(n+1)}.
\end{aligned}$$

We also use the following conventions for naming various types of polynomials and power series.

- $\mathfrak{Q}^d(X; \Lambda) \in \mathbb{C}[X, \Lambda] \equiv \mathbb{C}[\Lambda][X]$ denotes a polynomial in X of degree d whose coefficients are polynomial in Λ.
- $\mathcal{P}(\Lambda; X) \in \mathbb{C}[\![X, \Lambda]\!] \equiv \mathbb{C}[\![X]\!][\Lambda]$ denotes a polynomial in Λ whose coefficients are power series in X.
- $\mathcal{R}(X; \Lambda) \in \mathbb{C}[\![X, \Lambda]\!] \equiv \mathbb{C}[\Lambda][\![X]\!]$ denotes a power series in X whose coefficients are polynomial in Λ.

Assume the normal coordinates Z and $\hat{Z}$ for M and $\hat{M}$ are fixed. We now tackle the first step, the parametrizing of f_n and g_n. We begin with a lemma.

Lemma 5.1. *Let $(M, 0)$ and $(\hat{M}, 0)$ be formally equivalent formal 1-infinite type hypersurfaces as above. There exist unique formal power series $U, V \in \mathbb{C}[\![X, Y]\!]$, vanishing at 0, such that*

$$f_0(z) = U\left(z, \frac{z}{a_0^1}\right), \quad \bar{f}_0(\chi) = V\left(\chi, a_0^1 \chi\right)$$

for any $H \in \mathcal{F}(M, 0; \hat{M}, 0)$. If both M and $\hat{M}$ are convergent hypersurfaces, then $U, V \in \mathbb{C}\{X, Y\}$.

Proof. Proposition 3.7 implies that

$$\theta(z,\chi)\equiv\hat\theta\big(f_0(z),\bar f_0(\chi)\big). \tag{24}$$

Differentiating this L times in χ using Faa de Bruno's formula and setting $\chi=0$ yields the identity

$$\theta_L(z)\equiv(a_0^1)^L\hat\theta_L\big(f_0(z)\big). \tag{25}$$

Differentiating this K times in z and setting $z=0$ yields

$$\theta_L^{(K)}(0)=\big(\overline{a_0^1}\big)^K\big(a_0^1\big)^L\hat\theta_L^{(K)}(0). \tag{26}$$

In particular, we find that for any formal equivalence $H\in\mathcal{F}(M,0;\hat M,0)$,

$$\big|f_0'(0)\big|=\big|a_0^1\big|=\left|\frac{\theta_L^{(K)}(0)}{\hat\theta_L^{(K)}(0)}\right|^{1/(L+K)}=:\mu\in\mathbb{R}\setminus\{0\}. \tag{27}$$

We can write

$$\theta_L(z)=\frac{1}{K!}\theta_L^{(K)}(0)z^K t(z),$$

for some $t\in\mathbb{C}[[z]]$ with $t(0)=1$. Thus, there exists a unique power series $u(z)$ with $u(0)=1$ such that $u(z)^K=t(z)$. Similarly, write

$$\hat\theta_L(\hat z)=\frac{1}{K!}\hat\theta_L^{(K)}(0)\hat z^K\hat u(\hat z)^K,$$

with $\hat u(0)=1$. Define the formal power series

$$\iota(\hat z,X,Y):=\hat z\hat u(\hat z)-\mu^2Yu(X).$$

Observe that $\iota(0,0,0)=0$ and $\iota_{\hat z}(0,0,0)=1$, whence the formal Implicit Function Theorem implies the existence of a unique power series $U(X,Y)$, vanishing at $(0,0)$, such that $\iota\big(U(X,Y),X,Y\big)\equiv0$.

Now, suppose that $H\in\mathcal{F}(M,0;\hat M,0)$. Then identity (25) may be written as

$$\frac{1}{K!}\theta_L^{(K)}(0)\big(zu(z)\big)^K\equiv(a_0^1)^L\frac{1}{K!}\hat\theta_L^{(K)}(0)\big(f_0(z)\hat u(f_0(z))\big)^K.$$

Replacing ${\theta_L}^{(K)}(0)$ by equation (26) and canceling common terms yields the identity

$$\big(\overline{a_0^1}zu(z)\big)^K\equiv\big(f_0(z)\hat u(f_0(z))\big)^K.$$

Formally extracting K-th roots on both sides, we conclude that the two power series in the brackets differ only by some multiple $\varepsilon\in\mathbb{C}$ with $\varepsilon^K=1$. However, since

$$\frac{\partial}{\partial z}\big(\overline{a_0^1}zu(z)\big)\big|_{z=0}=\overline{a_0^1}=f_0'(0)=\frac{\partial}{\partial z}\big(f_0(z)\hat u(f_0(z))\big)\big|_{z=0},$$

we conclude that $\varepsilon = 1$ necessarily. Moreover, since $a_0^1 \overline{a_0^1} = \mu^2$, we have

$$\mu^2 \left(\frac{z}{a_0^1} \right) u(z) \equiv f_0(z)\, \hat{u}\big(f_0(z)\big).$$

Hence, $\iota\big(f_0(z), z, z/a_0^1\big) \equiv 0$, so by the uniqueness of U, we conclude $f_0(z) = U\big(z, z/a_0^1\big)$. Conjugating this result yields $\bar{f}_0(\chi) = V(\chi, a_0^1\chi)$, where V is defined by $V(X, Y) := \bar{U}(X, Y/\mu^2)$.

Finally, observe that if M and $\hat{M}$ are convergent, then the power series θ (hence also u) and $\hat{\theta}$ (hence $\hat{u}$) are convergent. Thus the holomorphic Implicit Function Theorem implies that U and V are necessarily convergent near $(0, 0) \in \mathbb{C}^2$. □

We can now extend this lemma to show that f_n and g_n are similarly parametrized for any $n \geq 0$.

Proposition 5.2. *Let* $(M, 0)$, $(\hat{M}, 0)$ *be formally equivalent formal* 1*-infinite type hypersurfaces as above. Then for every* $n \in \mathbb{N}$, *there exists a formal power series* $\mathcal{B}_n(z; \Delta, \Lambda) \in \mathbb{C}[\Delta, \Lambda][[z]]^2$ *such that*

$$\big(f_n(z), g_n(z)\big) = \mathcal{B}_n\big(z; \Delta(H), \Lambda^n_{2+\delta^1_K+\delta^1_T}(H)\big) \tag{28}$$

for any $H \in \mathcal{F}(M, 0; \hat{M}, 0)$. *In addition, if* $n \geq 1$, *then in fact*

$$\begin{aligned}\frac{f_n(z)}{f_0'(z)} &= T_n^1\big(z; \Delta(H), \Lambda^{n-1}_{2+\delta^1_K+\delta^1_T}(H)\big) - \frac{L}{a_0^1}\left(\frac{\theta_L(z)}{\theta_L'(z)}\right)a_n^1 + \frac{n}{b_0^0}\left(\frac{\theta_L(z)}{\theta_L'(z)}\right)b_n^0 \\ &\quad + \frac{i\,\delta^1_K}{2b_0^0}\left(\frac{1}{\theta_1'(z)}\right)b_n^1 + \frac{\delta^1_T}{a_0^1}\left(2i\,n\frac{\theta_1(z)^2}{\theta_L'(z)} - \frac{\theta_{L+1}(z)}{\theta_L'(z)} + \frac{L\,a_0^2}{a_0^1}\frac{\theta_L(z)}{\theta_L'(z)}\right)a_n^0,\end{aligned} \tag{29}$$

$$g_n(z) = T_n^2\big(z; \Delta(H), \Lambda^{n-1}_{2+\delta^1_K+\delta^1_T}(H)\big) + b_n^0 + \frac{2i\,b_0^0\,\delta^1_T}{a_0^1}\big(\theta_1(z)\big)a_n^0 \tag{30}$$

with $T(z; \Delta, \Lambda^{n-1}_{2+\delta^1_K+\delta^1_T}) \in \mathbb{C}^2[\Delta, \Lambda^{n-1}_{2+\delta^1_K+\delta^1_T}][[z]]$.

Moreover, if M *and* $\hat{M}$ *are convergent, there exists an* $\epsilon > 0$ *such that the map*

$$z \mapsto \mathcal{B}_n\big(z; \Delta(H), \Lambda^n_{2+\delta^1_K+\delta^1_T}(H)\big)$$

lies in $\mathbb{O}_\epsilon(z)^2$ *for every* $n \in \mathbb{N}$ *and every* $H \in \mathcal{F}(M, 0; \hat{M}, 0)$.

Proof. For convenience, we shall set $\gamma = 2 + \delta^1_K + \delta^1_T$. We proceed by induction. The $n = 0$ case follows immediately from Lemma 5.1 and the fact that $g_0(z) \equiv b_0^0$ (Proposition 3.7), so let us assume that the proposition is true up to some $n - 1 \geq 0$. To prove (28), it suffices to prove that equations (29) and (30) hold.

Suppose that $H : (M, 0) \to (\hat{M}, 0)$ is a formal equivalence.[1] Differentiating identity (20) n times in τ using Faa de Bruno's formula and setting $\tau = 0$ (or, equivalently, substituting $Q(z, \chi, \tau) = \tau S(z, \chi, \tau)$ and $v = n + 1$ into (13)) yields

$$\begin{aligned}(31)\quad -S(z,\chi,0)^{n+1}g_n(z) + b_0^0\hat{S}_{\hat{z}}\big(f_0(z),\bar{f}_0(\chi),0\big)S(z,\chi,0)^n f_n(z)\\ +b_0^0\hat{S}_{\hat{\chi}}\big(f_0(z),\bar{f}_0(\chi),0\big)\bar{f}_n(\chi) + \hat{S}\big(f_0(z),\bar{f}_0(\chi),0\big)\bar{g}_n(\chi)\\ \equiv \mathcal{P}_n\Big(b_0^0, \big(f_j(z), g_j(z), \bar{f}_j(\chi), \bar{g}_j(\chi)\big)_{j=1}^{n-1}; z, \chi, f_0(z), \bar{f}_0(\chi)\Big),\end{aligned}$$

where $\mathcal{P}_n(\Lambda; X)$, with $(\Lambda, X) \in \mathbb{C}^{4n-3} \times \mathbb{C}^4$, depends only on M and $\hat{M}$ and *not* the map H. (An explicit formula for $\mathcal{P}_n$ is given following the proof of Proposition 5.2.) Note that Lemma 3.3 implies $\hat{S}\big(f_0(z), \bar{f}_0(\chi), 0\big) = S(z, \chi, 0)$, whence

$$\hat{S}_{\hat{z}}\big(f_0(z),\bar{f}_0(\chi),0\big) = \frac{S_z(z,\chi,0)}{f_0'(z)}, \quad \hat{S}_{\hat{\chi}}\big(f_0(z),\bar{f}_0(\chi),0\big) = \frac{S_\chi(z,\chi,0)}{\bar{f}_0'(\chi)}.$$

If equation (28) holds for some $n \in \mathbb{N}$, then

$$\begin{aligned}(32)\quad \overline{\lambda_4^n(H)} &= \big((\mathcal{B}_n)_z^1, (\mathcal{B}_n)^2, (\mathcal{B}_n)^1, (\mathcal{B}_n)_z^2\big)(0; \Delta(H), \Lambda_\gamma^n(H))\\ &=: \beta_n(\Delta(H), \Lambda_\gamma^n(H)).\end{aligned}$$

Applying the inductive hypothesis to this and substituting this into equation (31) yields

$$(33)\quad \big(\bar{f}_j(\chi), \bar{g}_j(\chi)\big) = \overline{\mathcal{B}}_j\Big(\chi; \Big(\frac{a_0^1}{\mu}\Big)^2 \Delta(H), \big(\beta_\ell(\Delta(H), \Lambda_\gamma^\ell(H))\big)_{\ell=0}^j\Big)$$

for $j < n$, where μ is defined in equation (27). Substituting these values into (31) yields

$$\begin{aligned}(34)\quad -S(z,\chi,0)^{n+1}g_n(z) + S(z,\chi,0)\bar{g}_n(\chi) + b_0^0 S_z(z,\chi,0)S(z,\chi,0)^n\frac{f_n(z)}{f_0'(z)}\\ +b_0^0 S_\chi(z,\chi,0)\frac{\bar{f}_n(\chi)}{\bar{f}_0'(\chi)} \equiv \mathcal{R}_n(z,\chi; \Delta(H), \Lambda_\gamma^{n-1}(H)),\end{aligned}$$

with $\mathcal{R}_n(X; \Lambda)$ independent of the mapping H for each $n \geq 0$.

On one hand, substituting $\chi = 0$ and the identities from equations (16) and (17) into (34) yields

$$(35)\quad g_n(z) = \mathcal{R}_n(z, 0; \Delta(H), \Lambda_\gamma^{n-1}(H)) + b_n^0 + \frac{2i\, b_0^0}{a_0^1}\big(\theta_1(z)\big)a_n^0.$$

On the other hand, differentiating identity (34) L times in χ, setting $\chi = 0$, and using the identities from equations (16) and (17) yields (after rearranging terms)

[1] We remark that the construction given in this section can be carried out if *no* formal equivalence exists between M and $\hat{M}$.

the identity

$$
\begin{aligned}
&\theta_L'(z)\,\frac{f_n(z)}{f_0'(z)}\\
&\quad\equiv -\frac{i}{2b_0^0}(\mathcal{R}_n)_{\chi^j}\big(z,0;\,\Delta(H),\,\Lambda_\gamma^{n-1}(H)\big)+\frac{n+1}{b_0^0}\theta_L(z)\,g_n(z)+\frac{i}{2b_0^0}b_n^L\\
&\qquad -\frac{1}{b_0^0}\big(\theta_L(z)\big)b_n^0-\frac{L}{a_0^1}\big(\theta_L(z)\big)a_n^1-\frac{1}{a_0^1}\Big(\theta_{L+1}(z)+2i\,\theta_1(z)^2-\frac{L\,a_0^2}{a_0^1}\,\theta_L(z)\Big)\,a_n^0.
\end{aligned}
$$

Using the formula for $g_n(z)$ from equation (35) and observing that $(\theta_1)^2=\theta_1\theta_L$ for every $L\geq 1$, we can rewrite this identity as

$$
\begin{aligned}
(36)\quad &\theta_L'(z)\,\frac{f_n(z)}{f_0'(z)}\\
&\quad\equiv -\frac{i}{2b_0^0}(\mathcal{R}_n)_{\chi^j}\big(z,0;\,\Delta(H),\,\Lambda_\gamma^{n-1}(H)\big)-\frac{n}{b_0^0}\big(\theta_L(z)\big)b_n^0+\frac{i}{2b_0^0}b_n^L\\
&\qquad -\frac{L}{a_0^1}\big(\theta_L(z)\big)a_n^1+\frac{1}{a_0^1}\Big(-\theta_{L+1}(z)+2i\,n\,\theta_1(z)^2+\frac{L\,a_0^2}{a_0^1}\,\theta_L(z)\Big)\,a_n^0.
\end{aligned}
$$

We complete the proof by examining cases.

Case 1. $K=1$. In this case $L=T=1$ necessarily, so $\gamma=4$ and $\theta_L'(z)=\theta_1'(z)$ is a multiplicative unit. Dividing it on both sides of (36) yields (29); equation (30) follows from (35).

Case 2. $K>0$. In this case, setting $z=0$ in (36) yields

$$
0=-\frac{i}{2b_0^0}(\mathcal{R}_n)_{\chi^j}\big(z,0;\,\Delta(H),\,\Lambda_\gamma^{n-1}(H)\big)+\frac{i}{2b_0^0}b_n^L,
$$

whence we may replace b_n^L in identity (36) by $(\mathcal{R}_n)_{\chi^j}\big(z,0;\,\Delta(H),\,\Lambda_\gamma^{n-1}(H)\big)$. Thus, after rearranging the terms again, we may rewrite (36) as

$$
\begin{aligned}
(37)\quad &\theta_L'(z)\,\frac{f_n(z)}{f_0'(z)}\\
&\quad\equiv \sum_{j=0}^{K-2}\bigg(\frac{r_j^n\big(\Delta(H),\Lambda_\gamma^{n-1}(H)\big)}{j!}z^j+\mathcal{R}_n^1\big(z;\,\Delta(H),\,\Lambda_0^{n-1}(H)\big)\bigg)-\frac{n}{b_0^0}\big(\theta_L(z)\big)\,b_n^0\\
&\qquad -\frac{L}{a_0^1}\big(\theta_L(z)\big)\,a_n^1+\frac{1}{a_0^1}\bigg(-\theta_{L+1}(z)+2i\,n\,\theta_1(z)^2+\frac{L\,a_0^2}{a_0^1}\,\theta_L(z)\bigg)\,a_n^0,
\end{aligned}
$$

with the r_j^n polynomials and $\mathcal{R}_n^1(z;\,\Delta,\,\Lambda)$ of order at least $K-1$ in z.

Case 2A. $T = 1$. Note that $\gamma = 3$. Since $\theta_{L+1}^{(j)}(0) = 0$ for $j < K-1$, differentiating (37) in z (up to $K-2$ times) yields the relations

$$r_j^n\big(\Delta(H), \Lambda_3^{n-1}(H)\big) = 0, \quad 0 \leq j \leq K-2.$$

This does not imply that the polynomials $r_j^n(\Delta, \Lambda)$ are themselves identically zero; merely that they vanish whenever

$$(\Delta, \Lambda) = \big(\Delta(H), \Lambda_3^{n-1}(H)\big)$$

for some formal equivalence $H \in \mathcal{F}(M, 0; \hat{M}, 0)$.

Consequently, we may remove the first $K-1$ summands of the right-hand expression in identity (37). Observe that all the remaining summands are of order at least $K-1$ in z, and hence can be divided by $\theta_L'(z)$ to form another power series. This division yields (29); (30) follows from (35).

Case 2B. $T = 0$. Note that $\gamma = 2$. We know there exists some $j_0 \in \{1, 2, \ldots, K-2\}$ such that $\theta_{L+1}^{(j_0)}(0) \neq 0$. Differentiating the identity (37) j_0 times in z and setting $z = 0$, we obtain

$$0 = r_{j_0}^n\big(\Delta(H), \Lambda_2^{n-1}(H)\big) - \frac{\theta_{L+1}^{(j_0)}(0)}{a_0^1} a_n^0,$$

whence we may replace a_n^0 in (35) and (37) by $\dfrac{a_0^1 r_{j_0}^n(\Delta(H), \Lambda_2^{n-1}(H))}{\theta_{L+1}^{(j_0)}(0)}$ to obtain

$$\begin{aligned}\theta_L'(z)\frac{f_n(z)}{f_0'(z)} &\equiv \sum_{j=0}^{K-2}\left(\frac{\tilde{r}_j^n(\Delta(H), \Lambda_2^{n-1}(H))}{j!} z^j + \mathcal{R}_n^2(z; \Delta(H), \Lambda_2^{n-1}(H))\right) \\ &\qquad - \frac{n}{b_0^0}\big(\theta_L(z)\big)b_n^0 - \frac{L}{a_0^1}\big(\theta_L(z)\big)a_n^1, \\ g_n(z) &= \mathcal{R}_n^3\big(z, 0; \Delta(H), \Lambda_2^{n-1}(H)\big) + b_n^0.\end{aligned}$$

Thus, (30) holds; arguing as in the proof of Case 2A now yields (29).

The only thing missing from the proof is the convergence statement. Assume now that M and $\hat{M}$ define real-analytic hypersurfaces in $\mathbb{C}^2$ through 0. Hence, there exists a $\delta > 0$ such that

$$S(z, \chi, \tau) \in \mathbb{O}_\delta(z, \chi, \tau), \quad \hat{S}\big(\hat{z}, \hat{\chi}, \hat{\tau}\big) \in \mathbb{O}_\delta\big(\hat{z}, \hat{\chi}, \hat{\tau}\big).$$

Without loss of generality, we shall assume that δ is chosen small enough such that $\theta_L(z) \neq 0$ for $0 < |z| < \delta$, since the zeros of a nonconstant holomorphic function of one variable are isolated.

Similarly, since $U(X,Y) \in \mathbb{C}\{X,Y\}$ vanishes at 0 by Lemma 5.1, there exists an $\eta > 0$ such that $U(X,Y) \in \mathbb{O}_\eta(X,Y)$ and satisfies $\left|U(X,Y)\right| < \delta$ whenever $|X|, |Y| < \eta$.

Choose $\epsilon < \min\{\delta, \eta, \mu\eta\}$, where μ is defined by equation (27). We claim this is the desired $\epsilon > 0$; the proof is by induction. The case $n = 0$ follows from Lemma 5.1. Assuming this choice of ϵ holds up to some $n-1$, then observe that the mapping

$$(z,\chi) \mapsto \mathcal{R}_n\big(z,\chi;\Delta(H),\Lambda_\gamma^{n-1}(H)\big) \equiv \mathcal{P}_n\Big(b_0^0, \big(f_j(z), g_j(z), \bar{f}_j(\chi), \bar{g}_j(\chi)\big)_{j=1}^{n-1}; z, \chi, f_0(z), \bar{f}_0(\chi)\Big)$$

converges if $|z|, |\chi| < \delta$ for any $H \in \mathcal{F}(M,0;\hat{M},0)$. Fix such an H. By equation (35), we conclude $g_n(z)$ converges on the ball $B^1(0,\epsilon) = \{z \in \mathbb{C} : |z| < \epsilon\}$. On the other hand, we have shown that $\theta_L'(z) f_n(z)/f_0'(z) = z^{K-1} q\big(z;\Delta(H),\Lambda_\gamma^{n-1}(H)\big)$, with $q(\,\cdot\,;\Delta(H),\Lambda_\gamma^{n-1}(H))$ convergent on $B^1(0,\epsilon)$. Since $\theta_L'(z)$ converges for $|z| < \epsilon$ and in the ϵ-ball vanishes only at $z = 0$ (of order $K-1$), we conclude that $f_n(z)$ converges on $B^1(0,\epsilon)$ as well, which completes the proof. □

It is of interest to note that as a consequence of Proposition 5.2, we see that if M and $\hat{M}$ are real-analytic hypersurfaces in $\mathbb{C}^2$ and H is a formal equivalence between them, the formal mappings $z \mapsto H_{w^n}(z,0)$ are convergent for every $n \in \mathbb{N}$; moreover, they converge on some common ϵ-neighborhood of $0 \in \mathbb{C}$, with ϵ independent of n and H.

Because it is useful in doing calculations, we now give the explicit formula for $\mathcal{P}_n$. Using Faa de Bruno's formula, we have

$$\mathcal{P}_n\Big(\big(f_j, g_j, \bar{f}_j, \bar{g}_j\big)_{j=0}^{n-1}; z, \chi, \hat{z}, \hat{\chi}\Big) = p_n\Big(\big(f_j, g_j, \bar{f}_j, \bar{g}_j\big)_{0\le j\le n-1}, \big(S_{\tau^j}(z,\chi,0)\big)_{0\le j\le n}, \big(\hat{S}_{\hat{z}^j\hat{\chi}^k\hat{\tau}^\ell}(\hat{z},\hat{\chi},0)\big)_{0\le j+k+\ell\le n}\Big)$$

where p_n is the universal polynomial

$$p_n\Big(\big(f_j, g_j, \bar{f}_j, \bar{g}_j\big)_{0\le j\le n-1}, \big(S_j\big)_{0\le j\le n}, \big(\hat{S}_{(j,k,\ell)}\big)_{0\le j+k+\ell\le n}\Big) \equiv \sum_{\substack{\alpha\in\mathbb{N}^n\\ k+[\alpha]=n\\ |\alpha|<n}} \frac{n!\, g_{|\alpha|}\, S_k}{k!\,\alpha!} \prod_{p=1}^n \left(\frac{S_{p-1}}{(p-1)!}\right)^{\alpha_p} - \sum_{\substack{\alpha,\beta,\gamma\in\mathbb{N}^n\\ k+[\alpha]+[\beta]+[\gamma]=n\\ [\alpha],[\beta],k<n}} \frac{n!\,\bar{g}_k\,\hat{S}_{(|\alpha|,|\beta|,|\gamma|)}}{k!\,\alpha!\,\beta!\,\gamma!} \times \prod_{p=1}^n \left(\sum_{\substack{\xi\in\mathbb{N}^p\\ [\xi]=p}} \frac{f_{|\xi|}}{\xi!} \prod_{q=1}^n \left(\frac{S_{q-1}}{(q-1)!}\right)^{\xi_q}\right)^{\alpha_p} \left(\frac{\bar{f}_p}{p!}\right)^{\beta_p} \left(\frac{\bar{g}_{p-1}}{(p-1)!}\right)^{\gamma_p}.$$

In particular, observe that

(38)

$$\mathcal{P}_n\big((0,0,g_0,\bar{g}_0,0,0,\dots,0);z,\chi,\hat{z},\hat{\chi}\big)=-g_0S_{\tau^n}(z,\chi,0)+\bar{g}_0^n\hat{S}_{\hat{\tau}^n}(\hat{z},\hat{\chi},0).$$

This completes the first step of the proof. We move on to the second step, which involves parametrizing Λ^n.

Proposition 5.3. *Let $(M,0)$ and $(\hat{M},0)$ be formal hypersurfaces of 1-infinite type which are formally equivalent as above. Then for every $n\in\mathbb{N}$, there exists a power series*

$$\mathcal{A}_n(z;\Delta,\Lambda)\in\mathbb{C}[\Delta,\Lambda][\![z]\!]^2$$

such that

$$\big(f_n(z),g_n(z)\big)=\mathcal{A}_n\Big(z;\Delta(H),\big(\lambda^n_{2+\delta^1_K+\delta^1_L\delta^1_T}(H)\big)_{j\in\mathcal{D}(M),j\le n}\Big).$$

for any $H\in\mathcal{F}(M,0;\hat{M},0)$. Moreover, if M and $\hat{M}$ are convergent, there exists an $\epsilon>0$ such that the map

$$z\mapsto\mathcal{A}_n\Big(z;\Delta(H),\big(\lambda^n_{2+\delta^1_K+\delta^1_L\delta^1_T}(H)\big)_{j\in\mathcal{D}(M),j\le n}\Big)$$

lies in $\mathbb{O}_\epsilon(z)^2$ for every $n\in\mathbb{N}$ and every $H\in\mathcal{F}(M,0;\hat{M},0)$.

Proof. We continue with the notation from Proposition 5.2; in particular, we shall continue to let γ denote $2+\delta^1_K+\delta^1_T$. Observe that Proposition 5.3 follows immediately from Proposition 5.2 if it can be shown that for every $n\notin\mathcal{D}(M)$, there exists a $\mathbb{C}^\gamma$-valued polynomial $\omega^n(\Delta,\Lambda)$ such that

$$\lambda^n_\gamma(H)=\omega^n\big(\Delta(H),\Lambda^{n-1}_{2+\delta^1_K+\delta^1_L\delta^1_T}(H)\big)\quad\text{for all } H\in\mathcal{F}(M,0;\hat{M},0). \tag{39}$$

To see this, suppose equation (39) holds for every $n\notin\mathcal{D}(M)$. An easy induction shows that for *every* $n\in\mathbb{N}$, there exists a $\mathbb{C}^\gamma$-valued polynomial $\tilde{\omega}^n(\Delta,\Lambda)$ such that

$$\lambda^n_\gamma(H)=\tilde{\omega}^n\Big(\Delta(H),\big(\lambda^j_{2+\delta^1_K+\delta^1_L\delta^1_T}(H)\big)_{j\in\mathcal{D},j\le n}\Big).$$

Substituting this into the power series for $\mathcal{B}_n$ given by Proposition 5.2 completes the proof.

Hence, we must show that a relation of the form given in (39) holds for each $n\notin\mathcal{D}(M)$. To this end, define the power series

$$\tilde{\Upsilon}^n:(\mathbb{C}^2,0)\to(\mathbb{C}^4,0)$$

by $\tilde{\Upsilon}^n_j = \Upsilon^n_j$ for $j \neq 3$, and set

$$\tilde{\Upsilon}^n_3(z,\chi) := \delta^1_T \Bigg(\delta^1_K\, \theta_1^{(L)}(0)\, \frac{\theta_\chi(z,\chi)}{\bar{\theta}'_L(\chi)} + \frac{L\big(\theta_L^{(K)}(0)\theta_{L+1}^{(K)}(0) - \theta_L^{(K+1)}(0)\theta_{L_1}^{(K-1)}(0)\big)}{K\,\theta_1^{(K)}(0)^2}\, \frac{\bar{\theta}_L(\chi)}{\bar{\theta}'_L(\chi)}\theta_\chi(z,\chi)$$
$$- \Big(\frac{1+i\,\theta(z,\chi)}{1-i\,\theta(z,\chi)}\Big)^n \bigg(\theta_1(z)\big(1+\theta(z,\chi)^2\big) + \Big(\frac{\theta_{L+1}(z)}{\theta'_L(z)} - 2i\,n\,\frac{\theta_1(z)^2}{\theta'_L(z)}\Big)\theta_z(z,\chi)\bigg)$$
$$+ \frac{\theta_{L_1}^{(K-1)}(0)}{\theta_L^{(K)}(0)} \bigg(\bar{\theta}_1(\chi)\big(1+\theta(z,\chi)^2\big) + \Big(\frac{\bar{\theta}_{L+1}(\chi)}{\bar{\theta}'_L(\chi)} + 2i\,n\,\frac{\bar{\theta}_1(\chi)^2}{\bar{\theta}'_L(z)}\Big)\theta_\chi(z,\chi)\bigg)\Bigg).$$

Observe that

$$\delta^1_L\, \tilde{\Upsilon}^n_3 = \Upsilon^n_3.$$

Reconsider the identity (34). If we substitute into it the explicit formulas for $f_n(z)$ and $g_n(z)$ given in Proposition 5.2, as well as the corresponding formulas for $\bar{f}_n(\chi)$ and $\bar{g}_n(\chi)$ given by equation (33), we can rewrite this as

$$\tilde{\Upsilon}^n(z,\chi)^t\,\kappa^n\big(\Delta(H),\lambda_2^0(H)\big)\lambda_4^n(H) \equiv W^n\big(z,\chi;\Delta(H),\Lambda_\gamma^{n-1}(H)\big), \tag{40}$$

where the superscript ^{t} denotes the transpose operation, $\kappa^n(\Delta,\lambda)$ is the 4×4 matrix of polynomials defined by

$$\kappa^n(\Delta,\lambda_2^0) := \begin{pmatrix} (L/K)\,\Delta(b_0^0)^2 & -n/K & -\delta^1_T(L/K)\,a_0^2\,\Delta^2(b_0^0)^3 & 0 \\ 0 & -i/2 & 0 & 0 \\ 0 & 0 & -\delta^1_T\,\Delta(b_0^0)^2 & 0 \\ 0 & 0 & 0 & \delta^1_K i/2 \end{pmatrix}$$

(by Lemma 5.1, a_0^2 is a polynomial in a_0^1), and

$$W^n(z,\chi;\Delta,\Lambda) \in \mathbb{C}[\Delta,\Lambda][\![z,\chi]\!].$$

Denote by $\tilde{\kappa}^n$ the 4×4 matrix function

$$\tilde{\kappa}^n(\Delta,\lambda_2^0) := \begin{pmatrix} (K/L)\,\Delta(a_0^1)^2 & 2i\,n/L\Delta(a_0^1)^2 & -a_0^2\,\Delta\,a_0^1 & 0 \\ 0 & 2i & 0 & 0 \\ 0 & 0 & -\delta^1_T\,\Delta(a_0^1)^2 & 0 \\ 0 & 0 & 0 & -\delta^1_K\,2i \end{pmatrix}.$$

Observe that if $a_0^1 b_0^0 \neq 0$, then

$$\kappa^n\left(\frac{1}{a_0^1 b_0^0}, \lambda_2^0\right) \cdot \tilde{\kappa}^n\left(\frac{1}{a_0^1 b_0^0}, \lambda_2^0\right) = \begin{pmatrix} 1 & 0 & \dfrac{L a_0^2}{K a_0^1}(\delta_T^1 - 1) & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \delta_T^1 & 0 \\ 0 & 0 & 0 & \delta_K^1 \end{pmatrix}_j ,$$

For convenience, we denote by κ_j^n the upper-left $j \times j$ submatrix of κ^n for $1 \leq j \leq 4$; we define $\tilde{\kappa}_j^n$ similarly. We now complete the proof by examining cases.

Case 1. $K = 1$. Observe that $L = T = 1$ necessarily, so $\tilde{\Upsilon}^n = \Upsilon^n$ and $\kappa_4^n, \tilde{\kappa}_4^n$ are matrix inverses for all $n \in \mathbb{N}$. Suppose that $n \notin \mathfrak{D}(M)$, and choose a basis $\{\upsilon_{s_j,t_j}^n\}_{j=1}^4$ for $\mathcal{V}^n$. If Ξ is the 4×4 matrix whose j-th row is υ_{s_j,t_j}^n, then it follows that Ξ is invertible. Now, differentiating (40) s_j times in z, t_j times in χ, and setting $z = \chi = 0$ (for $j = 1, 2, 3, 4$), we obtain the 4×4 linear system of equations of the form

$$\Xi \kappa_4^n\big(\Delta(H), \lambda_0^2(H)\big) \lambda_4^n = w^n\big(\Delta(H), \Lambda_4^{n-1}(H)\big),$$

Thus, we may take

$$\omega^n(\Delta, \Lambda_4^{n-1}) := \tilde{\kappa}_4^n(\Delta, \lambda_0^2)\, \Xi^{-1}\, w^n(\Delta, \Lambda_4^{n-1})$$

to complete the proof.

Case 2. $K > L = 1 = T$. We have $\tilde{\Upsilon}^n = \Upsilon^n = (\Upsilon_1^n, \Upsilon_2^n, \Upsilon_3^n, 0)$ and $\kappa_3^n, \tilde{\kappa}_3^n$ are inverses for all $n \in \mathbb{N}$. Observe too that (40) reduces to

$$\big(\Upsilon_1^n(z, \chi), \Upsilon_2^n(z, \chi), \Upsilon_3^n(z, \chi)\big)^t \kappa_3^n\big(\Delta(H), \lambda_2^0(H)\big) \lambda_3^n(H) \equiv W^n\big(z, \chi; \Delta(H), \Lambda_3^{n-1}(H)\big).$$

The proof now follows the exact same lines as in the previous case.

Case 3. $T = 0$. Since this implies $K > 1$, it follows that $\tilde{\Upsilon}^n = \Upsilon^n = (\Upsilon_1^n, \Upsilon_2^n, 0, 0)$ and $\kappa_2^n, \tilde{\kappa}_2^n$ are inverses for all $n \in \mathbb{N}$. Here, the identity (40) reduces to

$$\big(\Upsilon_1^n(z, \chi), \Upsilon_2^n(z, \chi)\big)^t \kappa_2^n\big(\Delta(H), \lambda_2^0(H)\big) \lambda_2^n(H) \equiv W^n\big(z, \chi; \Delta(H), \Lambda_2^{n-1}(H)\big). \tag{41}$$

The proof now follows the exact same lines as in the previous two cases.

Case 4. $L > 1 = T$. Observe that identity (40) reduces to

$$\big(\Upsilon_1^n(z, \chi), \Upsilon_2^n(z, \chi), \tilde{\Upsilon}_3^n(z, \chi)\big)^t \kappa_3^n\big(\Delta(H), \lambda_2^0(H)\big) \lambda_3^n(H) \equiv W^n\big(z, \chi; \Delta(H), \Lambda_3^{n-1}(H)\big). \tag{42}$$

We claim that $a_n^0 = \sigma^n(\Delta(H), \Lambda_3^{n-1}(H))$ for every $n \in \mathbb{N}$, where σ^n is a polynomial. Hence, we can write

$$\big(f_n(z), g_n(z)\big) = \mathcal{B}_n\big(z; \Delta(H), \Lambda_3^n(H)\big) = \widetilde{\mathcal{B}}_n\big(z; \Delta(H), \Lambda_2^n(H)\big);$$

that is, $f_n(z)$ and $g_n(z)$ are given by expressions of the same form as in Proposition 5.2, *but without the a_n^0 term*. Hence, identity (40) reduces to identity (41), and the proof proceeds as in Case 3.

To prove the claim, we proceed by induction. For $n = 0$, this is trivial, as $a_0^0 = 0$. For the inductive step, we consider two cases.

Case 4A. $\theta_{L+1}^{(K-1)}(0) = 0$. Then equation (29) implies

$$\overline{a_n^0} = f_n(0) = \overline{a_0^1}\, T_n^1\big(0; \Delta(H), \Lambda_3^{n-1}(H)\big).$$

Conjugating this and applying equation (33) yields $a_n^0 = \tilde{T}(\Delta(H), \Lambda_3^{n-1}(H))$ for some polynomial $\tilde{T}(\Delta, \Lambda)$. But by the inductive hypothesis, $\Lambda_3^{n-1}(H)$ is itself a polynomial in $(\Delta(H), \Lambda_2^{n-1}(H))$, so the induction is complete in this case.

Case 4B. $\theta_{L+1}^{(K-1)}(0) \neq 0$. Differentiating (42) $L-1$ times in χ and setting $\chi = 0$ yields the identity

$$\frac{\big|\theta_{L+1}^{(K-1)}(0)\big|^2}{\big|\theta_L^{(K)}(0)\big|^2}\theta_L(z)\, a_n^0 = W_{\chi^{L-1}}\big(z, 0; \Delta(H), \Lambda_3^{n-1}(H)\big).$$

Differentiating this K times in z and setting $z = 0$ yields $a_n^0 = \tilde{T}(\Delta(H), \Lambda_3^{n-1}(H))$ for some polynomial $\tilde{T}(\Delta, \Lambda)$. But by the inductive hypothesis, $\Lambda_3^{n-1}(H)$ is itself a polynomial in $(\Delta(H), \Lambda_2^{n-1}(H))$, so the induction is complete in this case. $\square$

This completes the second step. We move on to the third step, counting the elements of $\mathcal{D}$.

Proposition 5.4. *Given a fixed set of normal coordinates Z on M, the set $\mathcal{D}(M)$ defined by equation* (23) *has at most* $2(2 + \delta_K^1 + \delta_L^1 \delta_T^1)$ *elements.*

Proof. Consider the power series $\Upsilon^n(z, \chi)$ defined on page 120; we must prove that for all but $2(2 + \delta_K^1 + \delta_L^1 \delta_T^1)$ integers $n \in \mathbb{N}$, the set $\mathcal{V}^n(M)$ has dimension $2 + \delta_K^1 + \delta_L^1 \delta_T^1$.

Consider the matrix

$$\xi(n) := \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \upsilon_{2K,2L}^n & \upsilon_{3K,3L}^n & \upsilon_{3K,2L}^n & \upsilon_{2K,3L}^n \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix}^t .$$

Our goal will be to show that for all but at most $2(2 + \delta_K^1 + \delta_L^1 \delta_T^1)$ integers $n \in \mathbb{N}$, the first $2 + \delta_K^1 + \delta_L^1 \delta_T^1$ rows are linearly independent, which implies that $n \notin \mathcal{D}(M)$.

Using Faa de Bruno's formula, we compute that

$$(\Upsilon_1^n)_{\chi^{2L}}(z,0) = 2i\frac{(2L)!}{(L!)^2}K\,\theta_L^{(K)}(z)^2 n + \mathfrak{Q}^0\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<3L+K+1}\big),$$

$$(\Upsilon_1^n)_{\chi^{3L}}(z,0) = -2\frac{(3L)!}{(L!)^3}K\,\theta_L^{(K)}(z)^3 n^2 + \mathfrak{Q}^1\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<4L+K+1}\big),$$

$$(\Upsilon_2^n)_{\chi^{2L}}(z,0) = -2\frac{(2L)!}{(L!)^2}\theta_L^{(K)}(z)^2 n^2 + \mathfrak{Q}^1\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<3L+K+1}\big),$$

$$(\Upsilon_2^n)_{\chi^{3L}}(z,0) = -\frac{4i}{3}\frac{(3L)!}{(L!)^3}\theta_L^{(K)}(z)^3 n^3 + \mathfrak{Q}^2\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<4L+K+1}\big),$$

$$(\Upsilon_3^n)_{\chi^{2}}(z,0) = \delta_L^1\delta_T^1\big(-4\theta_1^{(K)}(z)^3 n^2 + \mathfrak{Q}^1\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<K+4}\big)\big),$$

$$(\Upsilon_3^n)_{\chi^{3}}(z,0) = \delta_L^1\delta_T^1\big(-16i\,\theta_1^{(K)}(z)^4 n^3 + \mathfrak{Q}^2\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<K+5}\big)\big),$$

$$(\Upsilon_4^n)_{\chi^{2}}(z,0) = \delta_K^1\big(\mathfrak{Q}^0\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<5}\big)\big),$$

$$(\Upsilon_4^n)_{\chi^{3}}(z,0) = \delta_K^1\big(12\theta_1(z)^2 n^2 + \mathfrak{Q}^1\big(n;\,(\partial^\nu\theta(z,0))_{|\nu|<5}\big)\big).$$

Setting $\alpha := \theta_L^{(K)}(0)$ it follows, we may write $\xi(n) = C_1(n) + C_2(n)$, with $C_1(n)$ given by

$$\begin{pmatrix} \dfrac{2i\,K(2L)!\,(2K)!\,\alpha^2}{(L!\,K!)^2}n & \dfrac{-2(2L)!\,(2K)!\,\alpha^2}{(L!\,K!)^2}n^2 & 0 & 0 \\ \dfrac{-2K(3L)!\,(3K)!\,\alpha^3}{(L!\,K!)^3}n^2 & \dfrac{-4i(3L)!\,(3K)!\,\alpha^3}{3(L!\,K!)^3}n^3 & 0 & 0 \\ 0 & 0 & \delta_L^1\delta_T^1\dfrac{-4(3K)!\,\alpha^3}{(K!)^3}n^2 & 0 \\ 0 & 0 & 0 & \delta_K^1\,72\alpha^2 n^2 \end{pmatrix}$$

and $C_2(n)$ of the form

$$\begin{pmatrix} \mathfrak{Q}^0(n;\,j_0^{3L+3K+1}\theta) & \mathfrak{Q}^1(n;\,j_0^{3L+3K+1}\theta) & \delta_L^1\delta_T^1\mathfrak{Q}^1(n;\,j_0^{3K+4}\theta) & \delta_K^1\mathfrak{Q}^0(n;\,j_0^{7}\theta) \\ \mathfrak{Q}^1(n;\,j_0^{4L+4K+1}\theta) & \mathfrak{Q}^2(n;\,j_0^{4L+4K+1}\theta) & \delta_L^1\delta_T^1\mathfrak{Q}^2(n;\,j_0^{4K+5}\theta) & \delta_K^1\mathfrak{Q}^2(n;\,j_0^{9}\theta) \\ \mathfrak{Q}^1(n;\,j_0^{3L+4K+1}\theta) & \mathfrak{Q}^2(n;\,j_0^{3L+4K+1}\theta) & \delta_L^1\delta_T^1\mathfrak{Q}^1(n;\,j_0^{4K+4}\theta) & \delta_K^1\mathfrak{Q}^0(n;\,j_0^{8}\theta) \\ \mathfrak{Q}^1(n;\,j_0^{4L+3K+1}\theta) & \mathfrak{Q}^2(n;\,j_0^{4L+3K+1}\theta) & \delta_L^1\delta_T^1\mathfrak{Q}^2(n;\,j_0^{3K+5}\theta) & \delta_K^1\mathfrak{Q}^2(n;\,j_0^{8}\theta) \end{pmatrix}.$$

We shall denote by $\xi_j(n)$ the upper-left $j \times j$ submatrix of $\xi(n)$ for $j = 1, 2, 3, 4$. We complete the proof by examining cases.

Case 1. $K = 1$. In this case $L = T = 1$ as well, whence $2 + \delta_K^1 + \delta_L^1\delta_T^1 = 4$. By examining the matrix $\xi_4(n)$, and in particular the term of highest order in n in each of its entries, we find that

$$\det \xi_4(n) = 110592\alpha^{10} n^8 + \mathfrak{Q}^7(n;\, j_0^9\theta).$$

Since $\alpha \neq 0$, this is a nonzero, eighth degree polynomial in n, and hence has at most eight distinct zeros (in the complex plane). If $\det \xi_4(n_0) \neq 0$, then the four rows of $\xi(n_0)$ are linearly independent, which completes the claim.

Case 2. $K > L = T = 1$. In this case, we have $2 + \delta_K^1 + \delta_L^1 \delta_T^1 = 3$. By examining the highest order terms in n as above, we find that

$$\det \xi_3(n) = 64\, K \frac{(2K)!(3K)!^2}{(K!)^8} \alpha^8 n^6 + \mathfrak{Q}^5(n;\, j_0^{4K+5}\theta).$$

Arguing as above implies that for all but (at most) six integers n, the matrix $\xi_3(n)$ is invertible, whence the first three rows of $\xi(n)$ are linearly independent. This completes the claim.

Case 3. $L > 1$ or $T = 0$. Since either of these conditions necessarily implies $K > 1$, we conclude that $2 + \delta_K^1 + \delta_L^1 \delta_T^1 = 2$. Since

$$\det \xi_2(n) = -\frac{4}{3} K \frac{(2L)!\,(3L)!\,(2K)!\,(3K)!}{(L!\,K!)^5} \alpha^5 n^4 + \mathfrak{Q}^3(n;\, j_0^{4L+4K+1}\theta),$$

the proof is complete by arguments similar to the previous case. □

Note that while $\mathfrak{D}(M)$ is always finite, it is also never empty. Indeed, $0 \in \mathfrak{D}(M)$ for any 1-infinite type hypersurface M, since it is easy to check that $\Upsilon_2^0(z, \chi) \equiv 0$.

This completes the third step of the proof. We complete the proof by showing that $\mathfrak{D}(M)$ is independent of the choice of normal coordinates used to define it. In fact, we prove the following, which completes the proof of Theorem 4.1.

Proposition 5.5. *Suppose that M, $Z = (z, w)$, Υ^n, and $\mathcal{V}^n = \mathcal{V}^n(M)$ are as above. Let $(\hat{M}, 0)$ be formally equivalent to $(M, 0)$, with corresponding power series $\hat{\Upsilon}^n$ and subspaces $\hat{V}^n = \mathcal{V}^n(\hat{M})$ defined using the normal coordinates $\hat{Z} = (\hat{z}, \hat{w})$. Then for every $n \in \mathbb{N}$, the dimensions of $\mathcal{V}^n$ and $\hat{V}^n$ are equal. In particular, the dimension of subspace $\mathcal{V}^n(M) \subset \mathbb{C}^4$ is independent of the choice of normal coordinates used to define it.*

Proof. Let $H(z, w) = \big(f(z, w), w\, g(z, w)\big)$ be a formal equivalence between M and $\hat{M}$. Consider the formal power series

$$(z, \chi) \mapsto \hat{\Upsilon}^n\big(f_0(z), \bar{f}_0(\chi)\big) \in \mathbb{C}[[z, \chi]]^4,$$

which may be viewed as the power series $\hat{\Upsilon}^n$ given in the Z coordinates. Using Faa de Bruno's formula and the fact that $f_0 : (\mathbb{C}, 0) \to (\mathbb{C}, 0)$ is a formal change of coordinates, it is straightforward to verify that

$$\operatorname{span}_{\mathbb{C}}\left\{\hat{\upsilon}^n_{s,t} := \frac{\partial^{s+t}}{\partial z^s \partial \chi^t}\hat{\Upsilon}^n\big(f_0(z), \bar{f}_0(\chi)\big)\Big|_{\substack{z=0\\ \chi=0}} : s, t \in \mathbb{N}\right\} = \hat{V}^n.$$

From (24) we derive

$$\hat{\theta}_{\hat{z}}\big(f_0(z), \bar{f}_0(\chi)\big) = \frac{\theta_z(z,\chi)}{f_0'(z)}, \quad \hat{\theta}_{\hat{\chi}}\big(f_0(z), \bar{f}_0(\chi)\big) = \frac{\theta_\chi(z,\chi)}{\bar{f}_0'(\chi)},$$

whereas repeated differentiation of this in χ yields

$$\hat{p}_{L+1}\big(f_0(z)\big) = \frac{1}{2(a_0^1)^{L+2}}\big(2a_0^1 p_{L+1}(z) - (L+1)La_0^2 p_L(z)\big).$$

From this and identity (25), it follows by an elementary (albeit involved) calculation that

$$\begin{aligned}
\hat{\Upsilon}^n_1\big(f_0(z), \bar{f}_0(\chi)\big) &= \quad \Upsilon^n_1(z,\chi),\\
\hat{\Upsilon}^n_2\big(f_0(z), \bar{f}_0(\chi)\big) &= \quad \Upsilon^n_2(z,\chi),\\
\hat{\Upsilon}^n_3\big(f_0(z), \bar{f}_0(\chi)\big) &= \frac{1}{a_0^1}\Upsilon^n_3(z,\chi) + \frac{\delta^1_T a_0^2}{K(a_0^1)^2}\Upsilon^n_1(z,\chi),\\
\hat{\Upsilon}^n_4\big(f_0(z), \bar{f}_0(\chi)\big) &= a_0^1\,\Upsilon^n_4(z,\chi).
\end{aligned}$$

Now, suppose that $\{\hat{\upsilon}^n_{s_j,t_j}\}_{j=1}^{\ell_0}$ is any collection of vectors in $\hat{V}^n$; consider the corresponding vectors $\upsilon^n_{s_j,t_j} \in \mathcal{V}^n$. Observe that if $\hat{\Xi}$, Ξ denote the $4 \times \ell_0$ matrices whose columns are, respectively, the $\hat{\upsilon}^n_{s_j,t_j}$, $\upsilon^n_{s_j,t_j}$, then in view of the above identities, these matrices necessarily have the same rank. In particular, the columns of $\hat{\Xi}$ are linearly independent if and only if the columns of Ξ are. From this it follows that $\hat{V}^n$ and $\mathcal{V}^n$ have the same dimension. □

The main results. We use Theorem 4.1 to prove the main theorems stated at the end of Section 2. We begin with Theorem 2.2.

Proof. Let M be a formal real hypersurface of 1-infinite type at 0. Observe that the result of Theorem 2.2 is independent of the choice of coordinates Z, so without loss of generality let us take $Z = (z, w)$ to be normal coordinates for M, so that M is given by equation (14). Let $\mathcal{D} = \mathcal{D}(M)$ be as in Theorem 4.1, and set $k := 2 + \max \mathcal{D}$, which exists since $\mathcal{D}$ is a finite set.

To prove this k is sufficient, suppose $\hat{M}$ is a formally equivalent formal real hypersurface. Define the corresponding $\mathcal{A}^n$ as in Theorem 4.1. Fix a formal equivalence $H \in \mathcal{F}(M, 0; \hat{M}, 0)$. Conjugating the formula for (f_n, g_n) implies that

$$\big(\bar{f}_n(\chi), \bar{g}_n(\chi)\big) = \overline{\mathcal{A}}^n\left(\frac{1}{\overline{a_0^1}\,\overline{b_0^0}}, \big(\overline{a_j^0}, \overline{b_j^0}, \overline{a_j^1}, \overline{b_j^1}\big)_{j\in\mathcal{D}}\right),$$

whence

$$(a_n^0, b_n^0, a_n^1, b_n^1) = A_n\left(\frac{1}{\overline{a_0^1}\,\overline{b_0^0}}, \big(\overline{a_j^0}, \overline{b_j^0}, \overline{a_j^1}, \overline{b_j^1}\big)_{j\in\mathcal{D}}\right), \quad n = \mathbb{N},$$

with $A_n \in \mathbb{C}[\Delta, \Lambda]^4$. Substituting this into $\mathcal{A}_n$, and recalling that

$$\Delta(H) = \frac{1}{a_0^1 b_0^0} = \frac{\overline{a_0^1}}{\mu^2 \overline{b_0^0}},$$

where μ is defined by (27), we can write

$$\big(f_n(z), g_n(z)\big) = \Gamma^n\left(z; \frac{1}{\overline{a_0^1}\,\overline{b_0^0}}, \big(\overline{a_j^0}, \overline{b_j^0}, \overline{a_j^1}, \overline{b_j^1}\big)_{j\in\mathcal{D}}\right),$$

with $\Gamma^n(z; \Delta, \Lambda) \in \mathbb{C}[\Delta, \Lambda][\![z]\!]^2$. Write

$$\Gamma^n_{z^j}\left(0; \frac{1}{\overline{a_0^1}\,\overline{b_0^0}}, \big(\overline{a_j^0}, \overline{b_j^0}, \overline{a_j^1}, \overline{b_j^1}\big)_{j\in\mathcal{D}}\right) =: \frac{c_j^n\Big(\big(\overline{a_j^0}, \overline{b_j^0}, \overline{a_j^1}, \overline{b_j^1}\big)_{j\in\mathcal{D}}\Big)}{\big(\overline{a_0^1}\,\overline{b_0^0}\big)^{\ell_j^n}},$$

with $\ell_j^n \in \mathbb{N}$ and c_j^n a $\mathbb{C}^2$-valued polynomial.

Now, observe that

$$\frac{\partial^{\ell+j} H}{\partial z^\ell \partial w^j}(0,0) = \big(\overline{a_j^\ell}, j\,\overline{b_{j-1}^\ell}\big).$$

In particular, $\overline{a_j^0}$ is a term in (the coordinates of) $j_0^k(H)$, a_j^1 and b_j^0 are terms in $j_0^{k+1}(H)$, and b_j^1 is a term in $j_0^{j+2}(H)$. Hence, c_n^j is a polynomial in $j_0^{2+\max\mathcal{D}}(H) = j_0^k(H)$ and

$$0 \neq \overline{a_0^1}\,\overline{b_0^0} = \det\Big(\frac{\partial H}{\partial Z}(0,0)\Big) =: q\big(j_0^k(H)\big),$$

so the proof is complete in view of equation (11). □

By inspecting Propositions 5.2 through 5.5, we see that we can replace the k given in the proof by $k := 1 + \delta_K^1 + \max\mathcal{D}$ to get a better bound in the $K > 1$ case, and if $\mathcal{D} = \{0\}$, then we may take $k = 1$ since $b_0^1 = 0$ by Proposition 3.7.

We now use this result to prove Theorem 2.3.

Proof. Let M, k be as in Theorem 2.2. Suppose that $\hat{M}$ is formally equivalent to M, and let Ψ be the formal power series from Theorem 2.2. If $H^1, H^2 : (M, 0) \to (\hat{M}, 0)$ are two formal equivalences that satisfy

$$\frac{\partial^{|\alpha|} H^1}{\partial Z^\alpha}(0) = \frac{\partial^{|\alpha|} H^2}{\partial Z^\alpha}(0) \quad \text{for all } |\alpha| \leq k,$$

it follows that $j_0^k(H^1) = j_0^k(H^2)$. If we call this common jet Λ_0, it follows from Theorem 2.2 that $H^1(Z) \equiv \Psi(Z; \Lambda_0) \equiv H^2(Z)$, as desired. □

We now tackle the two applications of Theorem 2.2 mentioned in Section 2. First we prove Theorem 2.4.

Proof. Let M, k be as in Theorem 2.2, and let Ψ be the formal power series defined in accord with that theorem with $\hat{M} = M$. That the mapping

$$j_0^k : \mathrm{Aut}(M, 0) \to J_0^k(\mathbb{C}^2, \mathbb{C}^2)_{0,0}$$

is injective follows from Theorem 2.3. Observe that $\Lambda_0 \in J^k(\mathbb{C}^2, \mathbb{C}^2)_{0,0}$ is in the image of j_0^k if and only if $q(\Lambda_0) \neq 0$ — so that $\Lambda_0 \in G^k(\mathbb{C}^2)_0)$ — and

$$\Lambda_0 = j_0^k\big(\Psi(\,\cdot\,, \Lambda_0)\big), \tag{43}$$

$$\rho\big(\Psi(Z, \Lambda_0), \overline{\Psi}\zeta, \overline{\Lambda}0)\big) = a(Z, \zeta)\rho(Z, \zeta) \tag{44}$$

for some multiplicative unit $a(Z, \zeta) \in \mathbb{C}[\![Z, \zeta]\!]$, where ρ is a defining power series for M. In view of equation (8), (43) is a finite set of polynomial equations in Λ_0, whereas (44) is a (possibly countably infinite) set of polynomial equations in $(\Lambda_0, \overline{\Lambda}0)$. Hence, the image of the mapping j_0^k is a locally closed subgroup of the Lie group $G^k(\mathbb{C}^2)_0$, and so is a Lie subgroup. □

And as a corollary, we have Theorem 2.5.

Proof. Let M, k be as in Theorem 2.2, and let $(\hat{M}, 0)$ be formally equivalent to $(M, 0)$. Injectivity of the jet map again follows from Theorem 2.3. Now, fix a formal equivalence $H_0 : (M, 0) \to (\hat{M}, 0)$; then any other formal equivalence is of the form $H := H_0 \circ A$, where $A \in \mathrm{Aut}(M, 0)$. In particular,

$$\begin{aligned} j_0^k\big(\mathcal{F}(M, 0; \hat{M}, 0)\big) &= \big\{ j_0^k(H_0 \circ A) : A \in \mathrm{Aut}(M, 0) \big\} \\ &= \big\{ j_0^k(H_0) \cdot j_0^k(A) : A \in \mathrm{Aut}(M, 0) \big\} \\ &= j_0^k(H_0) \cdot j_0^k\big(\mathrm{Aut}(M, 0)\big). \end{aligned}$$

Hence, the image of $\mathcal{F}(M, 0; \hat{M}, 0)$ is merely a coset of the algebraic Lie subgroup $j_0^k\big(\mathrm{Aut}(M, 0)\big)$ in the Lie group $G^k(\mathbb{C}^2)_0$, and so is itself a real-algebraic submanifold of $G^k(\mathbb{C}^2)_0$. □

References

[Baouendi and Rothschild 1991] M. S. Baouendi and L. P. Rothschild, "A general reflection principle in $\mathbb{C}^2$", *J. Funct. Anal.* **99**:2 (1991), 409–442. MR 92j:32035 Zbl 0747.32014

[Baouendi et al. 1999a] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, "Rational dependence of smooth and analytic CR mappings on their jets", *Mathematische Annalen* **315**:2 (1999), 205–249. MR 2001b:32075 Zbl 0942.32027

[Baouendi et al. 1999b] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, *Real submanifolds in complex space and their mappings*, Princeton Mathematical Series **47**, Princeton University Press, Princeton, NJ, 1999. MR 2000b:32066 Zbl 0944.32040

[Baouendi et al. 2002] M. S. Baouendi, N. Mir, and L. P. Rothschild, "Reflection ideals and mappings between generic submanifolds in complex space", *J. Geom. Anal.* **12**:4 (2002), 543–580. MR 2003m:32035 Zbl 1039.32021

[Bloom and Graham 1977] T. Bloom and I. Graham, "On "type" conditions for generic real submanifolds of $\mathbb{C}^n$", *Invent. Math.* **40**:3 (1977), 217–243. MR 58 #28644 Zbl 0346.32013

[Cartan 1932a] E. Cartan, "Sur la géométrie pseudo-conforme des hypersurfaces de l'espace de deux variables complexes", *Ann. Mat. Pura Appl.* (4) **11** (1932), 17–90. Zbl 0005.37304

[Cartan 1932b] E. Cartan, "Sur la géométrie pseudo-conforme des hypersurfaces de l'espace de deux variables complexes, II", *Ann. Sc. Norm. Super. Pisa* (2) **1** (1932), 333–354. Zbl 0005.37401

[Chern and Moser 1974] S. S. Chern and J. K. Moser, "Real hypersurfaces in complex manifolds", *Acta Math.* **133** (1974), 219–271. MR 54 #13112 Zbl 0302.32015

[Ebenfelt 2002] P. Ebenfelt, "On the analyticity of CR mappings between nonminimal hypersurfaces", *Math. Ann.* **322**:3 (2002), 583–602. MR 2003a:32058 Zbl 1002.32028

[Ebenfelt et al. 2003] P. Ebenfelt, B. Lamel, and D. Zaitsev, "Finite jet determination of local analytic CR automorphisms and their parametrization by 2-jets in the finite type case", *Geom. Funct. Anal.* **13**:3 (2003), 546–573. MR 2004g:32033 Zbl 1032.32025

[Kohn 1972] J. J. Kohn, "Boundary behavior of δ on weakly pseudo-convex manifolds of dimension two", *J. Differential Geometry* **6** (1972), 523–542. MR 48 #727 Zbl 0256.35060

[Kowalski 2002a] R. T. Kowalski, *Formal equivalences between real-analytic hypersurfaces*, Ph.D. dissertation, University of California, San Diego, 2002.

[Kowalski 2002b] R. T. Kowalski, "A hypersurface in $\mathbb{C}^2$ whose stability group is not determined by 2-jets", *Proc. Amer. Math. Soc.* **130**:12 (2002), 3679–3686. MR 2003f:32018 Zbl 1007.32024

[Meylan 1995] F. Meylan, "A reflection principle in complex space for a class of hypersurfaces and mappings", *Pacific J. Math.* **169**:1 (1995), 135–160. MR 96f:32018 Zbl 0838.32004

[Range 1986] R. M. Range, *Holomorphic functions and integral representations in several complex variables*, Graduate Texts in Mathematics **108**, Springer, New York, 1986. MR 87i:32001 Zbl 0591.32002

[Rothschild 2003] L. P. Rothschild, "Mappings between real submanifolds in complex space", pp. 253–266 in *Explorations in complex and Riemannian geometry*, edited by J. Bland et al., Contemp. Math. **332**, Amer. Math. Soc., Providence, RI, 2003. MR 2004h:32039 Zbl 1042.32016

[Zaitsev 1997] D. Zaitsev, "Germs of local automorphisms of real-analytic CR structures and analytic dependence on k-jets", *Math. Res. Lett.* **4**:6 (1997), 823–842. MR 99a:32007 Zbl 0898.32006

[Zaitsev 2002] D. Zaitsev, "Unique determination of local CR-maps by their jets: a survey", *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **13**:3-4 (2002), 295–305. MR 2004i:32056

Received March 17, 2003. Revised March 27, 2004.

R. TRAVIS KOWALSKI
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
SOUTH DAKOTA SCHOOL OF MINES AND TECHNOLOGY
510 E. SAINT JOSEPH STREET
RAPID CITY, SD 57701
travis.kowalski@sdsmt.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

WEAKLY REGULAR EMBEDDINGS OF STEIN SPACES WITH ISOLATED SINGULARITIES

JASNA PREZELJ

We show that any n-dimensional Stein space X with isolated singular points admits a proper holomorphic injective map $X \to \mathbb{C}^{2n}$ which is regular on $\mathrm{Reg}(X)$. The proof is based on the fact that the Whitney cones $C_5(x, X)$ are at most $2n$-dimensional, which means that there exists a neighborhood of x in X having a weakly regular embedding into $\mathbb{C}^{2n}$. The homotopic principle then enables us to obtain a weakly regular embedding of X into $\mathbb{C}^{2n}$.

1. Introduction

The motivation for this paper was the following question: Let M be a smooth, compact, strongly pseudoconvex, integrable CR-manifold of dimension $2n-1 \geq 5$ and of CR-dimension $n-1 \geq 2$. Find the smallest integer $N = N(n)$ such that M admits a CR embedding into $\mathbb{C}^N$. By the results of Rossi [1965] and Ohsawa [1984a; 1984b], there exists a pure n-dimensional Stein space X with isolated singular points and a relatively compact domain $D \subset X$ such that $\partial D = M$ and $\partial D \cap \mathrm{Sing}(X) = \varnothing$. This leads to the following problem: Let X be an n-dimensional Stein space with isolated singular points. Find the smallest integer N such that there exists a proper holomorphic injective map $f : X \to \mathbb{C}^N$ which is regular on $\mathrm{Reg}(X)$. It turns out that the dimension N can be expressed in terms of the Whitney cones C_5 (for the definition see Section 2 or [Chirka 1989]).

Theorem 1.1. *Let X be an n-dimensional Stein space with isolated singular points. Let $N(X) = \max\{[n/2]+n+1, 3, \max\{\dim C_5(x, X) : x \in X\}\}$. Then there exists a proper holomorphic injective map $f : X \to \mathbb{C}^{N(X)}$, which is regular on* $\mathrm{Reg}(X)$.

Remark 1.2. Since we are not interested in regularity at singular points we may (and will), with no loss of generality, assume that the space is reduced. By [Acquistapace et al. 1975] there is a proper holomorphic injective map $f : X \to \mathbb{C}^N$,

MSC2000: 32C15, 32C22, 32E10, 32H02.
Keywords: Stein space, holomorphic map, weakly regular embedding, homotopic principle, Whitney cone.
The author was supported by the Ministry of Education, Science and Sports of the Republic of Slovenia.

where $N \geq 2n+1$, which is regular on $\operatorname{Reg}(X)$. The dimension $N(X)$ from Theorem 1.1, however, is at most $2n$, because $\dim C_5(x, X) \leq 2n$ for all n-dimensional Stein spaces X and all $x \in X$. If X is a Stein manifold, it was proved by Schürman that $N(X) = [n/2] + n + 1$ for $n > 1$.

Remark 1.3. In the case of normal Stein spaces any two weakly regular holomorphic embeddings are biholomorphically equivalent. Let us mention that this result does not necessarily give a minimal N for CR-embedding.

The paper is organized as follows: the second section contains the definition and some properties of Whitney cones and the third section consists of the proof of the main theorem.

Definitions and notation. For $y \in \mathbb{C}^n$ let $|y| := \sup\{|y_i| : 1 \leq i \leq n\}$ denote the sup norm and $\|y\|$ the euclidean norm. By $B_n(r)$ we denote the ball in $\mathbb{C}^n$ with radius r and center 0.

Let X be a complex space, $K \subset X$ a compact subset, and $f : X \to \mathbb{C}^n$ a continuous map. We will use the notation $|f|_K := \max\{|f(x)| : x \in K\}$ and $\|f\|_K := \max\{\|f(x)\| : x \in K\}$. By $\mathcal{O}(X)$ we denote the space of all holomorphic functions on a complex space X equipped with the standard topology of uniform convergence on compact sets. For an analytic set $Y \subset X$ let $\Gamma(X, \mathcal{J}(Y))$ denote the space of holomorphic functions on X which vanish on Y. By TX we denote the complex tangent space of X and by $T_x X$ the complex tangent space of X at the point x.

A holomorphic map $f : X \to Y$ is *almost proper* if for each compact set $K \subset Y$ the connected components of $f^{-1}(K)$ are compact. A *stratification of a complex space* X is a finite descending chain of analytic sets $A_m := X \supset A_{m-1} \supset \cdots \supset A_0$ such that $A_i \setminus A_{i-1}$ is a complex manifold, for $i = 1, \ldots, m$.

2. Some properties of tangent cones

The Whitney tangent cones C_3, C_4 and C_5 play an important role in our work.

Definition 2.1. Let $X \subset \mathbb{C}^m$ be an analytic set, $x \in X$. Then let

$$\begin{aligned}
C_3(x, X) := \{v \in \mathbb{C}^m :{} & \text{there exists a sequence } x_j \in X \text{ such that } x_j \to x, \\
& \text{and a sequence } \lambda_j \in \mathbb{C} \text{ such that } \lambda_j(x_j - x) \to v\}, \\
C_4(x, X) := \{v \in \mathbb{C}^m :{} & \text{there exists a sequence } z_j \in \operatorname{Reg}(X) \text{ such that } z_j \to x, \\
& \text{and a sequence } v_j \in T_{z_j} X \text{ such that } v_j \to v\}, \\
C_5(x, X) := \{v \in \mathbb{C}^m :{} & \text{there exist sequences } z_j, w_j \in X \text{ with } z_j, w_j \to x, \\
& \text{and a sequence } \lambda_j \in \mathbb{C} \text{ such that } \lambda_j(z_j - w_j) \to v\}.
\end{aligned}$$

Further, set $C_3(X) := \{(x, v) : x \in X, v \in C_3(x, X)\}$ and define $C_4(X)$ and $C_5(X)$ similarly. Clearly these are three subsets of TX. Using the fact that every analytic set is locally biholomorphic to an analytic set in some $\mathbb{C}^m$, we can extend the above definition of cones to an arbitrary complex space X. A more detailed discussion on this subject can be found in [Chirka 1989; Stutz 1972]. We state some simple properties of Whitney cones [Chirka 1989, Sections 8.4, 9.1 and 9.2]. If $n = \dim(x, X)$, then:

(i) The cones $C_4(x, X)$ and $C_5(x, X)$ are biholomorphically invariant, projective algebraic sets with $n \le \dim C_i(x, X) \le 2n$, and the cone $C_3(x, X)$ is an n-dimensional algebraic set.

(ii) $C_3(x, X) \subset C_4(x, X) \subset C_5(x, X)$.

(iii) $C_4(X)$ is the closure of $TX|_{\mathrm{Reg}(X)}$.

(iv) If $x \in \mathrm{Reg}(X)$ then $\dim C_4(x, X) = \dim C_5(x, X) = n$.

(v) If $\dim C_5(x, X) = n$, then $x \in \mathrm{Reg}(X)$.

Example 2.2. Let $X = (\mathbb{C}^n \times 0) \cup (0 \times \mathbb{C}^n) \subset \mathbb{C}^{2n}$. Then $C_3(0, X) = C_4(0, X) = X$ and $C_5(0, X) = \mathbb{C}^{2n}$.

Proposition 2.3 [Chirka 1989, Section 8.4]. *Let $X \subset \mathbb{C}^m$ be an analytic set containing* 0, *let $L = \mathbb{C}^{m-k} \times 0 \subset \mathbb{C}^m$, and suppose that $C_3(0, X) \cap L = \{0\}$. Then there exists an open set $U \subset \mathbb{C}^m$ containing* 0, *such that the orthogonal projection $\pi_L : U \cap X \to \mathbb{C}^k$ is proper.*

Remark 2.4. The condition $C_3(0, X) \cap L = \{0\}$ implies that the neighborhood of 0 lies in some cone. The condition is fulfilled for almost every $(m-k)$-dimensional linear subspace $L \subset \mathbb{C}^m$. Clearly, the projection along any L with $\dim L \le m - n$ and $C_3(0, X) \cap L = \{0\}$ is also proper.

Proposition 2.5 [Chirka 1989, Section 9.4]. *Let $X \subset \mathbb{C}^m$ be a pure n-dimensional analytic set containing* 0, *let $L = \mathbb{C}^{m-n} \times 0 \subset \mathbb{C}^m$, and let $\pi_L : U \cap X \to \mathbb{C}^n$ be the orthogonal projection. If $C_4(0, X) \cap L = \{0\}$ then there exists an open set $U \subset \mathbb{C}^m$ containing* 0, *such that* $\mathrm{br}(\pi_L, X \cap U) = (X \setminus \mathrm{Reg}(X)) \cap U$.

Remark 2.6. In the case of a general n-dimensional analytic set such projection is of course not a cover; it is, however, proper (because $C_3(0, X) \subset C_4(0, X)$) and regular on $\mathrm{Reg}(X) \cap U$.

Corollary 2.7. *Let X be a complex space, let $x \in X$ and suppose* $\dim C_4(x, X) = k$. *Then there exists an open neighborhood U of x, and a proper holomorphic map $f : U \to \mathbb{C}^k$ which is regular on* $\mathrm{Reg}(X) \cap U$. *Every holomorphic map $f : X \to \mathbb{C}^k$ with* $\mathrm{Ker}\, Df(x) \cap C_4(x, X) = \{0\}$ *is regular on* $\mathrm{Reg}(X) \cap U$ *for a suitable open neighborhood U of x.*

Proof. We may assume that $X \subset \mathbb{C}^m$, since the statement is local. For the first part, notice that the condition $\dim C_4(x, X) = k$ implies the existence of an $(m-k)$-dimensional linear subspace L with $L \cap C_4(x, X) = \{0\}$. The rest follows from Proposition 2.5 and the remark below.

As for the second part, if the statement is false, there exist sequences $x_j \in \operatorname{Reg}(X)$ with $x_j \to x$, and $v_j \in T_{x_j}X$ with $\|v_j\| = 1$, such that

$$Df(x_j)(v_j) = 0.$$

By passing to a subsequence we may assume that $v_j \to v$. But v is in $C_4(x, X)$ by definition; therefore $Df(x)(v) = 0$, which is a contradiction. $\square$

Proposition 2.8 [Chirka 1989, Section 9.4]. *Let $X \subset \mathbb{C}^m$ be a pure n-dimensional analytic set containing* 0, *and let $L = \mathbb{C}^{m-n-1} \times 0 \subset \mathbb{C}^m$. If $C_5(0, X) \cap L = \{0\}$ then there exists an open set $U \subset \mathbb{C}^m$ such that the orthogonal projection $\pi_L : U \to \mathbb{C}^{n+1}$ is an almost one-sheeted cover over some analytic subset of $\mathbb{C}^{n+1}$, that is, a homeomorphism of $X \cap U$ onto some hypersurface in $U \cap \mathbb{C}^{n+1}$.*

Remark 2.9. As before in the case of a general n-dimensional analytic set such a projection is not a cover; it is proper (because $C_3(0, X) \subset C_5(0, X)$), regular on $\operatorname{Reg}(X) \cap U$ and injective.

Corollary 2.10. *Let X be a complex space, let $x \in X$, and let* $\dim C_5(x, X) = k$. *Then there exists an open neighborhood U of x, and a proper, injective holomorphic map $f : U \to \mathbb{C}^k$, which is regular on* $\operatorname{Reg}(X) \cap U$. *Every holomorphic map $f : X \to \mathbb{C}^k$ satisfying* $\operatorname{Ker} Df(x) \cap C_5(x, X) = \{0\}$ *is injective, proper and regular on* $\operatorname{Reg}(X) \cap U$ *for some neighborhood U of x.*

Proof. The first part follows by a similar argument to that used in Corollary 2.7. We may assume that $X \subset \mathbb{C}^m$. Because $C_4(x, X) \subset C_5(x, X)$, the regularity of the map on $\operatorname{Reg}(X) \cap U$, for some small neighborhood U of x, follows from Corollary 2.7. If the map were not injective in any neighborhood of x then there would exist sequences $x_j, y_j \in X$ with $x_j, y_j \to x$ and $x_j \neq y_j$, such that $f(x_j) - f(y_j) = 0$. The Taylor series expansion gives us

$$f(x_j) - f(y_j) = Df(x_j)(x_j - y_j) + o(|x_j - y_j|) = 0,$$

which means that

$$Df(x_j)\Big(\frac{x_j - y_j}{|x_j - y_j|}\Big) \to 0.$$

By passing to a subsequence we may assume that $(x_j - y_j)/|x_j - y_j| \to v$, which lies in $C_5(x, X)$. But then $Df(x)(v) = 0$, which contradicts the assumption. $\square$

Definition 2.11. Let X be a complex space, let $x \in X$, and let $f : X \to \mathbb{C}^m$ be a holomorphic map. The map f is *weakly regular at x* if $C_5(x, X) \cap \operatorname{Ker} Df(x) = \{0\}$ and *weakly regular* if $C_5(X) \cap \operatorname{Ker} Df = 0$, where 0 is the zero section in TX.

On a complex manifold the notions of regular and weakly regular coincide. One of the key features of weakly regular maps that will be used in the sequel is local injectivity. Let us state two more lemmas describing properties of injective weakly regular maps.

Lemma 2.12. *Let X be a complex space, let $K \subset X$ be a compact set and let $f : X \to \mathbb{C}^m$ be a holomorphic map which is weakly regular and injective on K. Then there exists an open neighborhood $U \subset X$ of K such that f is injective and weakly regular on U.*

Proof. Weak regularity is obviously an open condition. Assume that the map is not injective. Then there are sequences $x_j, y_j \in X$, with $x_j \neq y_j$, such that $x_j \to x \in K$, $y_j \to y \in K$ and $f(x_j) = f(y_j)$. Injectivity of f on K implies that $x = y$ and since f is weakly regular on K it is injective in a neighborhood of x, which contradicts the existence of the sequences x_j and y_j. □

Lemma 2.13. *Let X be a complex space, let $K \subset X$ be a compact set, and let $f : X \to \mathbb{C}^m$ be a holomorphic map which is weakly regular and injective on K. Then there exists an $\varepsilon > 0$ such that any holomorphic map $g : X \to \mathbb{C}^m$ satisfying $|g - f|_K < \varepsilon$ is injective and weakly regular on K.*

Proof. Every map g close enough to f on K is weakly regular on K and therefore locally injective. The next step is to prove a local result:

Claim. Take $x \in X$ and assume that the map f is weakly regular (and therefore injective) in a small compact neighborhood U of x. Then there exists an $\varepsilon > 0$ such that if $|g - f|_U < \varepsilon$, then g is injective and weakly regular on U.

Proof of the claim. Note that any map close to f is weakly regular at x and therefore injective in some neighborhood of x. We need to prove that the map is injective on U. Assume the converse. Then there exists a sequence $\varepsilon_j \to 0$, a sequence $g_j : X \to \mathbb{C}^m$, and sequences $x_j, y_j \in U$, such that $|g_j - f| < \varepsilon_j$ and such that $g_j(x_j) - g_j(y_j) = 0$. We may assume that $x_j \to x$ and $y_j \to y$, and also that $(x_j - y_j)/|x_j - y_j| \to v$. Injectivity of f implies $x = y$. The Taylor series expansion gives

$$Dg_j(x_j)(x_j - y_j) = o_j(|x_j - y_j|).$$

Because of the Cauchy estimates there is $o(|x_j - y_j|)$ such that

$$\left|o_j(|x_j - y_j|)\right| < \left|o(|x_j - y_j|)\right|$$

for all j. Dividing the above equation by $|x_j - y_j|$ and passing to the limit we get $Df(x)(v) = 0$ which contradicts the fact that f is weakly regular. □

We have proved that there exists an open neighborhood V of the diagonal $\Delta \subset K \times K$ such that if g is close enough to f, the map $g(x) - g(y) : X \times X \to \mathbb{C}^m$

will have no zeroes in V except the diagonal Δ. Injectivity of f implies that $\min\{|f(x)-f(y)| : (x, y) \in K \times K \setminus V\} > 0$. The same holds for each map g close enough to f on K, which means that any such g is injective on K. $\square$

3. Proof of the main theorem

Let X be a n-dimensional Stein space with $S = \{s_j\} = X \setminus \operatorname{Reg}(X)$ discrete and let $N = N(X) = \max\big\{[n/2]+n+1,\ 3,\ \max\{\dim C_5(x, X) : x \in X\}\big\}$. We seek a proper, holomorphic, weakly regular, injective map $F = (H, G) : X \to \mathbb{C}^n \times \mathbb{C}^{N-n}$. We first construct an almost proper holomorphic map $H : X \to \mathbb{C}^n$ having certain additional properties (a generic almost proper map) and then construct the map $G : X \to \mathbb{C}^{N-n}$ such that $F = (H, G)$ has the desired properties.

By the definition of N there exist injective weakly regular holomorphic maps $\Phi_j : U_j \to \mathbb{C}^N$ defined on small neighborhoods U_j of s_j. For simplicity let us assume that $\Phi_{j,N}(s_j) = j$. Since X is Stein there is a holomorphic map $\Phi : X \to \mathbb{C}^N$ which coincides with the Φ_j to second order on S. Define $\varphi = (\Phi_1, \dots, \Phi_n)$. The following theorem gives a generic almost proper map H which coincides with φ on S to second order.

Proposition 3.1 (Generic almost proper maps) [Schürmann 1997; Prezelj 2003]. *Let X be a n-dimensional Stein space, let $Y \subset X$ be a discrete set, let* $\operatorname{Sing}(X) \subset Y$, *let $\varphi : X \to \mathbb{C}^n$ be a holomorphic map, and let $q' = \big[\frac{n+1}{2}\big]$. For each $y \in Y$ let a number $m_y \in \mathbb{N}$ be given. The set of all almost proper holomorphic maps $H : X \to \mathbb{C}^n$ satisfying*

(i) $(H-\varphi)_y \in \mathcal{J}(Y)^{m_j}$ *for each $y \in Y$, and*

(ii) $\dim\{x \in X \setminus Y : rank_x H \le n-i\} < 2(q'-i+1)$, *for $i = 1, \dots, n$*

is residual in the set $\mathcal{G}$ of all holomorphic maps G satisfying $(G-\varphi)_y \in \mathcal{J}(Y)^{m_j}$ for each $y \in Y$.

Remark 3.2. The first theorem of this sort was proved in [Schürmann 1997, Proposition 4.1]; see [Prezelj 2003, Proposition 2.4] for modifications. In our case $m_j = 2$ for each j. The maps H and φ will be fixed through the rest of the section.

The construction of the map G requires more work. We follow the proof of the embedding theorem in [Schürmann 1992]. The full proof is quite long and complicated, so we will only explain how to modify the theorems so that they hold for weakly regular maps. The main tool in the proof is the h-principle:

Definition 3.3 [Gromov 1989]. Let Z and X be complex spaces, let $h : Z \to X$ be a surjective submersion, and let $U \subset X$ be an open set. Then h admits a spray over U if, for some $m \in \mathbb{N}$, there exists a holomorphic map $s : h^{-1}(U) \times \mathbb{C}^m \to h^{-1}(U)$ such that

(i) $s(z, 0) = z$ for each $z \in h^{-1}(U)$,

(ii) $s(z, \mathbb{C}^m) \subset h^{-1}(h(z))$ for each $z \in h^{-1}(U)$, and

(iii) $(\partial/\partial t)\, s(z, t)|_{t=0} : \mathbb{C}^m \to \operatorname{Ker} D_z h$ is surjective.

Theorem 3.4 (The h-principle for Stein spaces) [Gromov 1989; Forstnerič and Prezelj 2001]. *Let X be a Stein space, Z a complex space and $h : Z \to X$ a holomorphic submersion* (*with constant corank*) *onto X. Assume that each $x \in X$ has a neighborhood $U \subset X$ such that h admits a spray over U. Let d be a metric on Z compatible with the complex space topology. Then*:

(i) *Each continuous section $f_0 : X \to Z$ can be deformed to a holomorphic section $f_1 : X \to Z$ through a continuous one-parameter family of continuous sections* (*a homotopy*) *$f_t : X \to Z$, for $t \in [0, 1]$.*

(ii) *If $K \subset X$ is a compact holomorphically convex set and the initial section f_0 is holomorphic in a neighborhood of K, then for each $\varepsilon > 0$ there exists a homotopy $f_t : X \to Z$, for $t \in [0, 1]$, such that $d(f_t(x), f_0(x)) < \varepsilon$ for each $x \in K$ and $t \in [0, 1]$, each f_t is holomorphic in a neighborhood of K and f_1 is holomorphic on X. In this case it suffices to assume that the submersion $h : Z \to X$ has a spray over small open subsets of $X \setminus K$.*

For $R > 0$, let X^R be an arbitrary union of finitely many connected components of the set $H^{-1}(B_n(R)) \subset X$, and let $Z^R = H(X^R) = B_n(R)$. Note that the map $H : X^R \to B(R)$ is proper. Let $\{X^k\}$ be a normal exhaustion of X and let the set U_k be an open Stein neighborhood of X^k contained in X^{k+1}, for each k. By the above definition the set X^k is Runge in X^{k+1}. We may assume that $S \cap (\partial X^k \cup \partial U_k) = \varnothing$. By $\psi(z) = \|z\|^2$ we denote the square of euclidean norm on $\mathbb{C}^n$.

We will construct a sequence of maps $G_k : U_k \to \mathbb{C}^{N-n}$ and a decreasing sequence $\varepsilon_j \to 0$ such that

(i) $(H, G_k) : U_k \to \mathbb{C}^N$ is weakly regular and injective,

(ii) $\|G_k - G_{k-1}\|_{X^{k-1}} < 2^{-k}\varepsilon_{k-1}$,

(iii) if $G' : U_k \to \mathbb{C}^{N-n}$ satisfies $\|G' - G_k\|_{X^k} < \varepsilon_k$ then $(H, G') : X \to \mathbb{C}^N$ is weakly regular and injective,

(iv) $\inf\{\|G_k(x)\| : x \in (H^{-1}(B_n(k-1)) \setminus X^{k-1}) \cap X^k\} > k - 1$,

(v) $DG_k(s_j) = D(\Phi_{n+1}, \ldots, \Phi_N)(s_j)$ for all j and k, whenever the expression makes sense.

Note that the sequence G_k converges uniformly on compact sets to a map G such that the map $(H, G) : X \to \mathbb{C}^N$ is weakly regular and injective by (i), (ii) and (iii), and proper by (iv). Condition (v) could be omitted since weak regularity is stable under small perturbations; in fact the construction ensures that we get (v) for free.

Before proceeding to the construction of maps the G_k, we define stratifications of X and $\mathbb{C}^n$ which we will need in the sequel.

Lemma 3.5 [Schürmann 1992; Prezelj 2003]. *There exist stratifications $X_n := X^R \supset X_{n-1} \supset \cdots \supset X_0 \supset X_{-1} = \varnothing$ and $Z_n := Z^R \supset Z_{n-1} \supset \cdots \supset Z_0 \supset Z_{-1} = \varnothing$, with X_0, $Z_0 \neq \varnothing$, such that*

(i) *$X_0 \supset X^R \cap S$ and $Z_0 \supset H(S \cap X^R)$,*

(ii) *$X_j = H^{-1}(Z_j) \cap X^R$,*

(iii) *the sets X_j and Z_j have dimension at most j and the sets $X_j^* := X_j \setminus X_{j-1}$ and $Z_j^* = Z_j \setminus Z_{j-1}$ are either complex j-dimensional manifolds or empty,*

(iv) *if X_j^* is not empty, the map $H : X_j^* \to Z_j^*$ is an immersion for $j \in \{0, \dots, n\}$,*

(v) *the rank of H is constant on each connected component of the set X_j^* for each $j \in \{0, \dots, n\}$.*

We quote some more results from [Schürmann 1992] (the almost proper map H is fixed). The original theorems deal with immersions and injective immersions; in our case the term "immersion" will be replaced with the term "weakly regular". Let $q = N - n$, fix some $R > 0$ and let $\{X_j\}$ and $\{Z_j\}$ be the stratifications from Lemma 3.5.

Theorem 3.6. *Choose $j \in \{1, \dots, n\}$ and $r, r_1, r_2 \in \mathbb{R} \setminus \{0\}$ such that $r_2 > 0$ and $r_1 < r_2 < r < R$. Let r_1^2 and r_2^2 be regular values for $\psi|_{Z_j^*}$ and suppose that $\big(X_{j-1} \cap \overline{X^{r_2}}\big) \cup \big(X_j \cap \overline{X^r}\big)$ is not empty. Let $f : X^r \to \mathbb{C}^q$ be a holomorphic map such that the map $(H, f) : X^r \to \mathbb{C}^N$ is weakly regular on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$. Then*:

(i) *If $r_1 < 0$, there is a holomorphic map $f' : X^{r_2} \to \mathbb{C}^q$ such that $f' - f|_{X^{r_2}}$ is in $\Gamma(X^{r_2}, \mathcal{J}(X_{j-1})^2 \cap \mathcal{J}(X_j))^q$ and (H, f') is weakly regular on $X_j \cap X^{r_2}$.*

(ii) *If $r_1 > 0$ then f can be approximated arbitrarily well on X^{r_1} by a holomorphic map $f' : X^{r_2} \to \mathbb{C}^q$ such that $f' - f|_{X^{r_2}}$ is in $\Gamma(X^{r_2}, \mathcal{J}(X_{j-1})^2 \cap \mathcal{J}(X_j))^q$ and (H, f') is weakly regular on $X_j \cap X^{r_2}$.*

Proof. Since the sheaf $\mathcal{J}(X_{j-1})^2 \cap \mathcal{J}(X_j)$ is coherent, there are finitely many holomorphic sections

$$f_1, \dots, f_M \in \Gamma(X, \mathcal{J}(X_{j-1})^2 \cap \mathcal{J}(X_j))^q$$

which generate $\Gamma(X^{r_2}, \mathcal{J}(X_{j-1})^2 \cap \mathcal{J}(X_j))^q$. We seek a map f' of the form

$$f' = f + \sum \alpha_j f_j,$$

where $\alpha = (\alpha_1, \dots, \alpha_m)$ is regarded as a section of the trivial bundle $X^r \times \mathbb{C}^M$. Denote by $K = \operatorname{Ker} DH \subset TX$ the kernel of DH and note that the definition of X_j^* implies that $K|_{X_j^*}$ is a vector bundle. Let $\Sigma \subset (X_j \cap X^r) \times \mathbb{C}^M =: V$ be the set of all

$(x, a_1, \dots, a_M)$ such that the map $(H, f + \sum a_j f_j)$ is not weakly regular in x. Let $p : V \to (X_j \cap X^r)$ be the trivial projection. Then because (H, f) is weakly regular on $X_{j-1} \cap X^r$ and the maps f_j vanish to second order on X_{j-1}, the set Σ is a subset of $X_j^* \times \mathbb{C}^M$. Since $K|_{X_j^*}$ is a vector bundle, Σ is the set of all $(x, a_1, \dots, a_M)$ in $X_j^* \times \mathbb{C}^M$ such that the map $K_x \to \mathbb{C}^{q'}$ given by $v \mapsto Df(x)v + \sum a_i Df_j(x)v$, is not injective, so Σ is analytic in $X_j^* \times \mathbb{C}^M$. But we want Σ to be analytic in V which means we have to prove that Σ is closed in V. Now, since (H, f) is weakly regular on $X_{j-1} \cap X^r$ any map of the form $\big(H, f + \sum \alpha_j f_j\big)$ is weakly regular on the same set and because weak regularity is an open condition, the map $\big(H, f + \sum \alpha_j f_j\big)$ is also weakly regular in some neighborhood of $X_{j-1} \cap X^r$ which means that Σ is a closed in V. As in [Schürmann 1992] we prove that the projection $p : V \setminus \Sigma \to (X_j^* \cap X^r)$ is a locally trivial fibration which admits a spray.

Our goal is to find a holomorphic section α of $(X^r \times \mathbb{C}^M) \setminus \Sigma \to X^r$. The zero section on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$ is a section of

$$((X^r \cap X_j) \times \mathbb{C}^M) \setminus \Sigma \to (X^r \cap X_j)$$

because the map (H, f) is weakly regular on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$. And since weak regularity is an open condition, the zero section defined in a neighborhood of the set $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$ is a section of $((X^r \cap X_j) \times \mathbb{C}^M) \setminus \Sigma \to (X^r \cap X_j)$ as well. As in [Schürmann 1992] this section can be extended to a continuous section of $((X^r \cap X_j) \times \mathbb{C}^M) \setminus \Sigma \to (X^r \cap X_j)$. Then the h-principle applies (if $r_1 < 0$ we use the existence version and if $r_1 > 0$ the approximation version) which yields a holomorphic section α' of $((X^r \cap X_j) \times \mathbb{C}^M) \setminus \Sigma \to (X^r \cap X_j)$. This section can be trivially extended to a holomorphic section α of $(X^r \times \mathbb{C}^M) \setminus \Sigma \to X^r$. □

Remark 3.7. The maps f_j used in Theorem 3.6 vanished to the first order on X_j, which means that if the initial map f is such that (H, f) is injective on $X_j \cap X^r$ then the map (H, f') is also injective on $X_j \cap X^r$.

Theorem 3.8. *Choose $j \in \{1, \dots, n\}$ and $r, r_1, r_2 \in \mathbb{R} \setminus \{0\}$ such that $r_2 > 0$ and $r_1 < r_2 < r < R$. Let r_1^2 and r_2^2 be regular values for $\psi|_{Z_j^*}$ and suppose that $\big(X_{j-1} \cap \overline{X^{r_2}}\big) \cup \big(X_j \cap \overline{X^r}\big)$ is not empty. Let $f : X^r \to \mathbb{C}^q$ be a holomorphic map such that the map $(H, f) : X^r \to \mathbb{C}^N$ is injective and weakly regular on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$. Then*:

(i) *If $r_1 < 0$ there is a holomorphic map $f' : X^{r_2} \to \mathbb{C}^q$ such that $f' - f|_{X^{r_2}}$ is in $\Gamma(X^{r_2}, \mathcal{J}(X_{j-1})^2)^q$ and such that (H, f') is injective and weakly regular on $X_j \cap X^{r_2}$.*

(ii) *If $r_1 > 0$ the map f can be approximated arbitrarily well on the set X^{r_1} by a holomorphic map $f' : X^{r_2} \to \mathbb{C}^q$ such that $f' - f|_{X^{r_2}} \in \Gamma(X^{r_2}, \mathcal{J}(X_{j-1})^2)^q$ and (H, f') is injective and weakly regular on $X_j \cap X^{r_2}$.*

Proof. Since the sheaves $\mathscr{F} = H_*\mathbb{O}(X^R)$ and $\mathscr{I}(Z_{j-1})^2\mathscr{F}$ are coherent, there exist finitely many holomorphic sections $\psi_1, \dots \psi_M \in \Gamma(Z, \mathscr{I}(Z_{j-1})^2\mathscr{F})^q$ generating $\Gamma(Z^{r_2}, \mathscr{I}(Z_{j-1})^2\mathscr{F})^q$. Let $f_j \in \Gamma(X^r, \mathbb{O}(X))^q$ be liftings of the sections ψ_j. We are looking for the map f' of the form

$$f' = f + \sum(\alpha_j \circ H) \cdot f_j,$$

where $\alpha = (\alpha_1, \dots, \alpha_m)$ is regarded as a section of the trivial bundle $B(r) \times \mathbb{C}^M$. Let $\Sigma \subset (Z_j \cap B(r)) \times \mathbb{C}^M =: V$ be the set of all $(z, a_1, \dots, a_M)$, such that the map $H^{-1}(z) \to \mathbb{C}^n$, given by $x \mapsto f(x) + \sum a_j f_j(x)$, is not injective. Let the map $p : V \to (Z_j \cap B(r))$ be the trivial projection. Then because (H, f) is injective on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$ and the maps f_j vanish to the second order on X_{j-1}, the set Σ is an analytic subset of $Z_j^* \times \mathbb{C}^M$. As above we want Σ to be an analytic subset of V, that is, closed in V. Since (H, f) is injective and weakly regular on $X_{j-1} \cap X^r$ and the maps f_j vanish to the second order on X_{j-1}, any map of the form $\big(H, f + \sum(\alpha_j \circ H) \cdot f_j\big)$ is injective and weakly regular on the same set and, because being injective and weakly regular is an open condition, such map is also injective and weakly regular in some neighborhood of $X_{j-1} \cap X^r$. This means that the set Σ is closed in V. As in [Schürmann 1992] we prove that $p : V \setminus \Sigma \to (Z_j^* \cap B(r))$ is a locally trivial fibration which admits a spray.

We seek a holomorphic section α of the submersion $(B(r) \times \mathbb{C}^M) \setminus \Sigma \to B(r)$. The zero section defined on $\big(Z_{j-1} \cap \bar{B}(r)\big) \cup \big(Z_j \cap \bar{B}(r_1)\big)$ is a section of the map $\big((B(r) \cap Z_j) \times \mathbb{C}^M\big) \setminus \Sigma \to (B(r) \cap Z_j)$ because the map (H, f) is weakly regular and injective on $\big(X_{j-1} \cap \overline{X^r}\big) \cup \big(X_j \cap \overline{X^{r_1}}\big)$. And since being injective and weakly regular is an open condition, the zero section defined in a neighborhood of the set $\big(Z_{j-1} \cap \bar{B}(r)\big) \cup \big(Z_j \cap \bar{B}(r_1)\big)$ is a section of $\big((B(r) \cap Z_j) \times \mathbb{C}^M\big) \setminus \Sigma \to (B(r) \cap Z_j)$ as well. As in [Schürmann 1992] this section can be extended to a continuous section of $\big((B(r) \cap Z_j) \times \mathbb{C}^M\big) \setminus \Sigma \to (B(r) \cap Z_j)$. Then the h-principle applies (if $r_1 < 0$ we use the existence version and if $r_1 > 0$ the approximation version) and yields a holomorphic section α' of $\big((B(r) \cap Z_j) \times \mathbb{C}^M\big) \setminus \Sigma \to (B(r) \cap Z_j)$, which can be trivially extended to a holomorphic section α of $\big(B(r) \times \mathbb{C}^M\big) \setminus \Sigma \to B(r)$. □

At the beginning of this section we defined the map Φ. Now we define

$$f = (\Phi_{n+1}, \dots, \Phi_N).$$

The map (H, f) clearly is weakly regular on S and injective in a small neighborhood of S, since $\Phi_N(s_j) = j$. Using in turn Theorem 3.8 and Theorem 3.6 we can proceed by induction over the strata starting with X_0 (as in [Schürmann 1992]), and using the fact that being weakly regular or injective and weakly regular is an open condition, to obtain the following results:

Theorem 3.9 (Existence). *Let $R > 0$ and let X^R be the union of a finite number of connected components of the set $H^{-1}(B_n(R))$. For $r \in (0, R)$ let $X^r := X^R \cap H^{-1}(B_n(r))$. There exists a holomorphic map $G : X^r \to \mathbb{C}^q$ satisfying the conditions*

$\alpha(r)$: *the map $(H, G) : X^r \to \mathbb{C}^N$ is injective and weakly regular*, *and*

$\beta(r)$: *(H, G) coincides with Φ to the second order $S \cap X^r$.*

Theorem 3.10 (Approximation). *Let $R, r > 0$, X^R and X^r be as in Theorem 3.9 Choose $r_1 \in (r, R)$ and set $X^{r_1} := X^R \cap H^{-1}(B_n(r_1))$. If a holomorphic map $G : X^r \to \mathbb{C}^q$ satisfies conditions $\alpha(r)$ and $\beta(r)$ from Theorem 3.9*, *it can be approximated arbitrarily well on the set X^r by a map $G' : X^{r_1} \to \mathbb{C}^q$ satisfying $\alpha(r_1)$ and $\beta(r_1)$ from Theorem 3.9.*

Remark 3.11. Note that the induction preserves the derivatives of Φ at the points in S since they are contained in X_0 and the maps f_j in Theorem 3.6 and Theorem 3.8 vanish to second order on X_0.

Proof of the main theorem. Now we can construct the maps G_j and the sequence $\varepsilon_j \to 0$ with the required properties (i)–(v). First we consider the case $k = 1$. By the existence Theorem 3.9 there is a map $G_1 : X^1 \to \mathbb{C}^q$ with properties (i) and (v). By Lemma 2.13 there is an $\varepsilon_1 > 0$ such that (iii) holds. Now we prove the induction step. Assume that $G_1, \dots, G_k$ and $\varepsilon_1, \dots, \varepsilon_k$ have already been constructed. Let $X^{k'} := X^{k+1} \cap (H^{-1}(B_k) \setminus X^k)$, that is, $X^{k'}$ is the union of those connected components of $H^{-1}(B(k))$ which lie in X^{k+1} but not in X^k. By Theorem 3.9 there is a map G'_k satisfying (i). By adding a sufficiently large positive constant we may assume that $\|G'\|_{X^{k'}} > 2k$ and that the map $G' : X^k \cup X^{k'} \to \mathbb{C}^q$, defined by $G'|_{X^k} = G_k$, $G'_{X^{k'}} = G'_k$ is such that $(H, G') : X^k \cup X^{k'} \to \mathbb{C}^N$ is injective. Now the assumptions of Theorem 3.9 are fulfilled so there exists a map $G_{k+1} : X^{k+1} \to \mathbb{C}^N$ satisfying (i), (ii), (iv) and (v). As above, there exists $\varepsilon_{k+1} \in (0, \varepsilon_k)$ such that (ii) holds as well. □

References

[Acquistapace et al. 1975] F. Acquistapace, F. Broglia, and A. Tognoli, “A relative embedding theorem for Stein spaces”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **2**:4 (1975), 507–522. MR 53 #870 Zbl 0313.32020

[Chirka 1989] E. M. Chirka, *Complex analytic sets*, Mathematics and its Applications **46**, Kluwer, Dordrecht, 1989. MR 92b:32016 Zbl 0683.32002

[Forstnerič and Prezelj 2001] F. Forstnerič and J. Prezelj, “Extending holomorphic sections from complex subvarieties”, *Math. Z.* **236**:1 (2001), 43–68. MR 2002b:32017 Zbl 0968.32005

[Gromov 1989] M. Gromov, “Oka’s principle for holomorphic sections of elliptic bundles”, *J. Amer. Math. Soc.* **2**:4 (1989), 851–897. MR 90g:32017 Zbl 0686.32012

[Ohsawa 1984a] T. Ohsawa, “Holomorphic embedding of compact s.p.c. manifolds into complex manifolds as real hypersurfaces”, pp. 64–76 in *Differential geometry of submanifolds* (Kyoto, 1984), edited by K. Kenmotsu, Lecture Notes in Math. **1090**, Springer, Berlin, 1984. MR 86j:32047 Zbl 0578.32032

[Ohsawa 1984b] T. Ohsawa, “Global realization of strongly pseudoconvex CR manifolds”, *Publ. Res. Inst. Math. Sci.* **20**:3 (1984), 599–605. MR 85m:32013 Zbl 0568.32014

[Prezelj 2003] J. Prezelj, “Interpolation of embeddings of Stein manifolds on discrete sets”, *Math. Ann.* **326**:2 (2003), 275–296. MR 2004b:32043 Zbl 1037.32024

[Rossi 1965] H. Rossi, “Attaching analytic spaces to an analytic space along a pseudoconcave boundary”, pp. 242–256 in *Proc. Conf. Complex Analysis* (Minneapolis, 1964), edited by A. Aeppli et al., Springer, Berlin, 1965. MR 31 #381 Zbl 0143.30301

[Schürmann 1992] J. Schürmann, *Einbettungen Steinscher Räume in affine Räume minimaler Dimension*, Schriftenreihe des Mathematischen Instituts (3. Serie) **7**, Universität Münster, Münster, 1992. MR 94g:32012 Zbl 0766.32014

[Schürmann 1997] J. Schürmann, “Embeddings of Stein spaces into affine spaces of minimal dimension”, *Math. Ann.* **307**:3 (1997), 381–399. MR 98a:32011 Zbl 0881.32007

[Stutz 1972] J. Stutz, “Analytic sets as branched coverings”, *Trans. Amer. Math. Soc.* **166** (1972), 241–259. MR 48 #2420 Zbl 0239.32006

Received November 21, 2001. Revised February 28, 2004.

JASNA PREZELJ
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LJUBLJANA
JADRANSKA 19
SI-1000 LJUBLJANA
SLOVENIA

jasna.prezelj@fmf.uni-lj.si

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

A DE RHAM THEOREM FOR SYMPLECTIC QUOTIENTS

REYER SJAMAAR

We introduce a de Rham model for stratified spaces arising from symplectic reduction. It turns out that the reduced symplectic form and its powers give rise to well-defined cohomology classes, even on a singular symplectic quotient.

1. Introduction

Let G be a compact Lie group and let M be a smooth G-manifold. Let $\Omega(M)$ be the de Rham complex of differential forms on M and $\Omega_{\mathrm{bas}}(M)$ the subcomplex of basic forms. It was proved by Koszul [1953] that the cohomology of $\Omega_{\mathrm{bas}}(M)$ is isomorphic to the cohomology with real coefficients of the orbit space M/G (which is usually not a manifold, unless G acts freely).

Now suppose that M is equipped with a symplectic form ω and that the G-action is Hamiltonian with equivariant moment map $\Phi\colon M \to \mathfrak{g}^*$, where $\mathfrak{g} = \mathrm{Lie}\, G$. The appropriate quotient in this category is the Marsden–Meyer–Weinstein symplectic quotient $X = \Phi^{-1}(0)/G$. It is usually not a manifold either, unless G acts freely on the fiber $\Phi^{-1}(0)$, but it always has a natural stratification into symplectic manifolds.

Much work has been done on the intersection cohomology of symplectic quotients; see, for example, [Kirwan 1985; Lerman and Tolman 2000]. The purpose of this note is rather more modest. We introduce a de Rham model for the ordinary cohomology of the symplectic quotient X, which is a straightforward adaptation of Koszul's complex of basic forms. It relies on a notion of a differential form on X that extends the concept of a smooth function developed in [Arms et al. 1991]. Relevant examples are the reduced symplectic form and its powers, which define cohomology classes of even degree. These classes are nonzero if the quotient is compact. Thus the symplectic quotient, even when singular, carries a suitable analogue of a symplectic form and a Liouville volume form.

MSC2003: 53D20, 58A12.
Keywords: symplectic reduction, de Rham theory.
The author was partially supported by NSF Grant DMS-0071625.

2. Review

Let (M, ω) be a connected symplectic manifold and let G be a compact Lie group acting on M in a Hamiltonian fashion with moment map $\Phi\colon M \to \mathfrak{g}^*$, where $\mathfrak{g} = \operatorname{Lie} G$. This means $d\Phi^\xi = i(\xi_M)\omega$, where ξ_M denotes the vector field on M induced by $\xi \in \mathfrak{g}$ and $\Phi^\xi = \langle \Phi, \xi \rangle$ denotes the component of the moment map along ξ. Also Φ is required to be equivariant with respect to the given action on M and the coadjoint action on $\mathfrak{g}^*$. The *symplectic quotient* of M by G is the space $X = Z/G$, where $Z = \Phi^{-1}(0)$ is the zero fiber of the moment map. It was proved in [Marsden and Weinstein 1974] that if G acts freely on Z, then Z and X are smooth manifolds and X carries a natural symplectic form. If G does not act freely on Z, often neither Z nor X are manifolds. In this case we proceed as in [Sjamaar and Lerman 1991], the relevant results of which we recall now. For any closed subgroup H of G let

$$M_{(H)} = \{m \in M \mid G_m \text{ is conjugate to } H\}$$

be the stratum of orbit type (H) in the G-manifold M. Here G_m denotes the stabilizer of m with respect to the G-action. Put

$$Z_{(H)} = Z \cap M_{(H)}.$$

Then $Z_{(H)}$ is a smooth G-stable submanifold of M. Let $\{Z_a \mid a \in A\}$ be the collection of connected components of all manifolds of the form $Z_{(H)}$, where (H) ranges over all conjugacy classes of subgroups of G. The decomposition

$$Z = \coprod_{a \in A} Z_a \tag{2.1}$$

is a Whitney stratification of the fiber Z. In particular the index set A has a partial order defined by $a \leq b$ if $Z_a \subseteq \bar{Z}_b$. There is a unique maximal element in A. The corresponding stratum, known as the *principal* or *top* stratum Z_{prin}, is open and dense in Z. Moreover the null foliation of the symplectic form ω restricted to any stratum Z_a is exactly given by the G-orbits. Hence there exists a unique symplectic form ω_a on the quotient manifold $X_a = Z_a/G$ satisfying $\pi_a^*\omega_a = \iota_a^*\omega$, where $\iota_a\colon Z_a \hookrightarrow M$ is the inclusion map and $\pi_a\colon Z_a \twoheadrightarrow X_a$ the orbit map. The decomposition

$$X = \coprod_{a \in A} X_a \tag{2.2}$$

is a locally normally trivial stratification of the quotient X into the symplectic manifolds X_a. The principal stratum $X_{\text{prin}} = Z_{\text{prin}}/G$ is open and dense in X.

3. Forms on a symplectic quotient

We use the same notation as in the previous section. We denote the de Rham complex of a manifold P by $\Omega(P)$. A *differential form* on the symplectic quotient X is a differential form α on the top stratum X_{prin} such that there exists a differential form $\tilde{\alpha}$ on M satisfying $\pi_{\mathrm{prin}}^{*}\alpha = \iota_{\mathrm{prin}}^{*}\tilde{\alpha}$. We say that $\tilde{\alpha}$ *induces* α. An easy averaging argument shows that we may assume $\tilde{\alpha}$ to be G-invariant on M. We denote the collection of differential forms on X by $\Omega(X)$.

If $X = X_{\mathrm{prin}}$, then X and Z are manifolds and the lift of any form on X to Z can be extended to M, so in this case our notion of a differential form on X reduces to the standard notion. Observe that $\Omega(X)$ is a subcomplex of $\Omega(X_{\mathrm{prin}})$, and it is closed under the wedge product.

Example 3.1. The symplectic form ω_{prin} on X_{prin} is induced by the symplectic form ω on M and so defines a closed 2-form on X.

Clearly not every invariant form on M induces a form on X. Indeed, if $\tilde{\alpha} \in \Omega(M)^G$ induces $\alpha \in \Omega(X)$, then $\iota_{\mathrm{prin}}^{*}\tilde{\alpha} = \pi_{\mathrm{prin}}^{*}\alpha$ is a G-horizontal form on the G-manifold Z_{prin}, so it is annihilated by all inner products $i(\xi_M)$ for $\xi \in \mathfrak{g}$. Recall that a form β on M is *basic* with respect to the G-action if it is G-invariant and G-horizontal. Adapting this notion to our context, we say that β is Φ-*basic* if it is G-invariant and if $\iota_{\mathrm{prin}}^{*}\beta \in \Omega(Z_{\mathrm{prin}})$ is horizontal. Let $\Omega_\Phi(M)$ denote the set of Φ-basic forms. This is a subcomplex of $\Omega(M)$ and the kernel of the natural surjection $\Omega_\Phi(M) \to \Omega(X)$ is the ideal

$$I_\Phi(M) = \{\beta \in \Omega(M)^G \mid \iota_{\mathrm{prin}}^{*}\beta = 0\}.$$

Thus the de Rham complex of X is isomorphic to

$$\Omega(X) \cong \Omega_\Phi(M)/I_\Phi(M), \tag{3.2}$$

a subquotient of the de Rham complex of M. In degree 0 we have the smooth functions on X as defined in [Arms et al. 1991],

$$C^\infty(X) \cong C^\infty(M)^G \big/ \{f \in C^\infty(M)^G \mid f = 0 \text{ on } Z\}.$$

If O is a G-invariant open neighborhood of Z, then O is a Hamiltonian G-manifold in its own right, so we can define $\Omega_\Phi(O)$ and $I_\Phi(O)$. Plainly (3.2) remains valid if we replace M with O. Thus $\Omega(X)$ depends only on the G-germ of M at Z.

It is true, though not completely obvious from the definition, that every form on X restricts to a form on each stratum of X.

Lemma 3.3. (i) *Let $\beta \in \Omega_\Phi(M)$. Then $\iota_a^{*}\beta$ is a horizontal form on Z_a for all a.*

(ii) *Let $\beta \in I_\Phi(M)$. Then $\iota_a^{*}\beta = 0$ for all a.*

(iii) *There is a well-defined restriction map $\Omega(X) \to \Omega(X_a)$ for each stratum X_a.*

Proof. Let $\beta \in \Omega_\Phi(M)$ and $z \in Z_a$. Choose a sequence $\{z_n\}$ in Z_{prin} converging to z. By compactness of the Grassmannian we may assume that the sequence of tangent spaces $T_{z_n} Z_{\text{prin}}$ converges to a subspace T of $T_z M$. By definition $i(\xi_M)\beta_{z_n} = 0$ on $T_{z_n} Z_{\text{prin}}$ for all $\xi \in \mathfrak{g}$, so by continuity $i(\xi_M)\beta_z = 0$ on T for all ξ. By Whitney's Condition A we have $T_z Z_a \subseteq T$. Hence $i(\xi_M)\beta_z = 0$ on $T_z Z_a$ for all ξ. This proves (i).

Similarly, if $\beta \in I_\Phi(M)$ then $\beta_{z_n} = 0$ on $T_{z_n} Z_{\text{prin}}$, so by continuity $\beta_z = 0$ on T and hence $\beta_z = 0$ on $T_z Z_a$, which proves (ii).

It follows from (i) that if $\beta \in \Omega_\Phi(M)$ then $\iota_a^* \beta$ descends to a form β_a on X_a. The assignment $\beta \mapsto \beta_a$ defines a homomorphism $\Omega_\Phi(M) \to \Omega(X_a)$ for each a. It follows from (ii) that this map is 0 on the ideal $I_\Phi(M)$. Using the isomorphism (3.2) we obtain the desired restriction map $\Omega(X) \to \Omega(X_a)$. □

4. Symplectic induction

A shortcoming of the de Rham complex $\Omega(X)$ is that it appears to depend on the way in which X is written as a quotient. But in certain interesting situations this defect turns out to be illusory. For instance, let H be a closed subgroup of G and let (N, ω_N) be a Hamiltonian H-manifold with equivariant moment map $\Psi \colon N \to \mathfrak{h}^*$. Consider the Hamiltonian $G \times H$-space

$$P = T^*G \times N,$$

where the action of G on P is given by left multiplication on T^*G and the action of H by right multiplication on T^*G and the given action on N. Let M be the symplectic quotient of P with respect to the H-action. This is called the G-space *induced* by the H-space N. Since H acts freely on T^*G, M is a smooth manifold and from P it inherits a symplectic form ω and a Hamiltonian G-action with moment map Φ. Let Y be the symplectic quotient of N by the H-action and X the symplectic quotient of M by the G-action. The principle of reduction in stages implies that X and Y are isomorphic in the sense that there is a stratification-preserving homeomorphism $Y \to X$ that restricts to a symplectomorphism on each stratum. We can represent the situation symbolically by a commutative diagram

$$\begin{tikzcd} P \arrow[r, dotted, "G"] \arrow[d, dotted, "H"'] & N \arrow[d, dotted, "H"] \\ M \arrow[r, dotted, "G"'] & Y \cong X, \end{tikzcd}$$

where the dotted arrows indicate symplectic reduction with respect to the relevant group. We assert that the de Rham complexes of X and Y are likewise isomorphic.

To prove this we need to recall from [Sjamaar and Lerman 1991, §2] the definition of the isomorphism $Y \to X$. Choose an H-invariant subspace $\mathfrak{m}$ of $\mathfrak{g}$ complementary to the subalgebra $\mathfrak{h}$. Then we have H-invariant decompositions $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$ and $\mathfrak{g}^* = \mathfrak{h}^* \oplus \mathfrak{m}^*$. Define a map

$$G \times \mathfrak{m}^* \times N \to P \cong G \times \mathfrak{g}^* \times N$$

by sending (g, α, p) to $(g, \alpha - \Psi(p), p)$. This is an H-equivariant diffeomorphism from $G \times \mathfrak{m}^* \times N$ onto the zero fiber of the H-moment map on P. Taking quotients by H we obtain a G-equivariant diffeomorphism

$$M \cong (G \times \mathfrak{m}^* \times N)/H$$

from M to the homogeneous vector bundle over G/H with fiber $\mathfrak{m}^* \times N$. We identify M with this bundle and write a typical point in it as $[g, \alpha, p]$, with $g \in G$, $\alpha \in \mathfrak{m}^*$ and $p \in N$. The G-action on M is given by $k[g, \alpha, p] = [kg, \alpha, p]$ for $k \in G$ and the moment map by

$$\Phi([g, \alpha, p]) = \mathrm{Ad}^*(g)(\alpha + \Psi(p)). \tag{4.1}$$

Let $f : N \to M$ be the embedding defined by $f(p) = [1, 0, p]$. Then f is H-equivariant and (4.1) shows that $\Phi \circ f = \mathrm{pr}^* \circ \Psi$, where $\mathrm{pr}^* : \mathfrak{h}^* \to \mathfrak{g}^*$ is the transpose of the projection map $\mathfrak{g} \to \mathfrak{h}$. Hence f maps the zero fiber $Z_N = \Psi^{-1}(0)$ into Z and descends to a map $Y \to X$, which is the required isomorphism. In particular f maps the principal stratum $(Z_N)_{\mathrm{prin}}$ into the principal stratum of Z. In fact Z and Z_{prin} are homogeneous bundles over G/H,

$$Z = (G \times Z_N)/H \qquad \text{and} \qquad Z_{\mathrm{prin}} = (G \times (Z_N)_{\mathrm{prin}})/H.$$

This implies that the restriction map $f^* : \Omega(M) \to \Omega(N)$ sends $\Omega_\Phi(M)$ to $\Omega_\Psi(N)$ and $I_\Phi(M)$ to $I_\Psi(N)$. Therefore, because of the isomorphism (3.2), it descends to a map $r : \Omega(X) \to \Omega(Y)$.

Proposition 4.2. *The map $r : \Omega(X) \to \Omega(Y)$ is an isomorphism.*

Proof. This relies on material developed in the Appendix. Let $\iota_{\mathrm{prin}} : Z_{\mathrm{prin}} \to M$ be the inclusion map. This is a bundle map of fiber bundles over the base G/H. Its restriction to a fiber is the inclusion map $(\iota_N)_{\mathrm{prin}} : (Z_N)_{\mathrm{prin}} \to N$. Let

$$\begin{aligned} e_M &: \Omega(N)^H \to \Omega(M)^G, \\ e_Z &: \Omega((Z_N)_{\mathrm{prin}})^H \to \Omega(Z_{\mathrm{prin}})^G \end{aligned}$$

be the extension homomorphisms for the homogeneous bundles M and Z_{prin} as defined in the Appendix. Then

$$e_Z \circ (\iota_N)^*_{\mathrm{prin}} = \iota^*_{\mathrm{prin}} \circ e_M \tag{4.3}$$

by Lemma A.2.

Now we show that r is surjective. In fact we must show that $f^*\Omega_\Phi(M) = \Omega_\Psi(N)$. Let $\gamma \in \Omega_\Psi(N)$. Then by definition $(\iota_N)^*_{\mathrm{prin}}\gamma$ is H-basic, so $e_Z((\iota_N)^*_{\mathrm{prin}}\gamma)$ is G-basic by Lemma A.1(ii). From (4.3) we get that $\iota^*_{\mathrm{prin}}e_M(\gamma)$ is G-basic, i.e. $e_M(\gamma) \in \Omega_\Phi(M)$. Using Lemma A.1(i) we find that $\gamma = f^*\beta$ with $\beta = e_M(\gamma) \in \Omega_\Phi(M)$. Hence $f^*\Omega_\Phi(M) = \Omega_\Psi(N)$.

Next we prove that r is injective. Suppose that $\beta \in \Omega_\Phi(M)$ satisfies $f^*\beta \in I_\Psi(N)$. We need to show that $\beta \in I_\Phi(M)$. The assumptions on β mean that $\iota^*_{\mathrm{prin}}\beta$ is G-basic and that $(\iota_N)^*_{\mathrm{prin}}f^*\beta = 0$. Using Lemma A.1(iii) we get

$$\iota^*_{\mathrm{prin}}\beta = e_Z(f^*\iota^*_{\mathrm{prin}}\beta) = e_Z((\iota_N)^*_{\mathrm{prin}}f^*\beta) = e_Z(0) = 0,$$

that is, $\beta \in I_\Phi(M)$. □

5. The de Rham sheaf

To prove a de Rham theorem we need to sheafify the de Rham complex. Let U be an open subset of the symplectic quotient X. The stratification of X induces one on U, so we can talk about the principal stratum of U etc. A *differential form* on U is a differential form α on U_{prin} such that for all $x \in U$ there exist $\alpha' \in \Omega(X)$ and an open neighborhood U' of x in U such that $\alpha = \alpha'$ on U'_{prin}. The set of differential forms on U is denoted by $\Omega(U)$. It is easy to check that the presheaf of differential graded algebras $\Omega\colon U \mapsto \Omega(U)$ is a sheaf. Its space of global sections is the previously defined de Rham complex $\Omega(X)$.

Lemma 5.1. Ω *is an acyclic sheaf, i.e.* $H^i(X, \Omega^j) = 0$ *for all* $i \geq 1$ *and* $j \geq 0$.

Proof. The space X possesses smooth partitions of unity subordinate to arbitrary open covers $\mathcal{U}$. Indeed, for each $U \in \mathcal{U}$ choose a G-invariant open $\tilde{U}$ in M such that $U = (\tilde{U} \cap Z)/G$ and let O be the union of the $\tilde{U}$'s. Choose a smooth G-invariant partition of unity on the G-manifold O subordinate to the cover defined by the $\tilde{U}$'s; this induces a smooth partition of unity on X subordinate to $\mathcal{U}$. Thus the sheaf of smooth functions Ω^0 is fine in the sense of [Godement 1973, §3.7]. A standard result in sheaf theory (see [Godement 1973, Théorème 4.4.3], for example) now implies that Ω^0 is acyclic. Since Ω is a module over Ω^0, it is fine, and therefore acyclic, as well. □

There is an alternative characterization of forms on open subsets of X. The proof is an easy exercise involving partitions of unity.

Lemma 5.2. *Let* U *be an open subset of* X *and let* $\alpha \in \Omega(U_{\mathrm{prin}})$. *Then* $\alpha \in \Omega(U)$ *if and only if there exist a* G*-invariant open subset* $\tilde{U}$ *of* M *and a form* $\tilde{\alpha} \in \Omega(\tilde{U})$ *such that* $U = (\tilde{U} \cap Z)/G$ *and* $\iota^*_{\mathrm{prin}}\tilde{\alpha} = \pi^*_{\mathrm{prin}}\alpha$.

Now let $\underline{\mathbb{R}}$ be the sheaf of locally constant real-valued functions on X and consider the sequence

$$0 \to \underline{\mathbb{R}} \xrightarrow{i} \Omega^0 \xrightarrow{d} \Omega^1 \xrightarrow{d} \cdots, \tag{5.3}$$

where $i: \underline{\mathbb{R}} \to \Omega^0$ is the natural inclusion. The following assertion is proved in the next section.

Lemma 5.4. *The sequence* (5.3) *is exact.*

Thus the de Rham complex is an acyclic resolution of the constant sheaf, which by standard sheaf theory (see [Godement 1973, Théorèmes 4.7.1, 6.2.1], for example) implies the following de Rham theorem.

Theorem 5.5. *The de Rham cohomology ring $H(\Omega(X))$ is naturally isomorphic to the* (*Čech or singular*) *cohomology ring of X with real coefficients $H(X, \mathbb{R})$.*

6. The Poincaré lemma

In this section we prove the following (marginally stronger) version of Lemma 5.4: every $x \in X$ has a basis of open neighborhoods U such that the sequence

$$0 \to \mathbb{R} \xrightarrow{i} \Omega^0(U) \xrightarrow{d} \Omega^1(U) \xrightarrow{d} \cdots \tag{6.1}$$

is exact. The proof is a variation on a familiar homotopy argument in de Rham theory, which requires a brief look into the functorial properties of $\Omega(X)$.

Let (M', ω', Φ') be a second Hamiltonian G-manifold with zero fiber $Z' = (\Phi')^{-1}(0)$ and symplectic quotient $X' = Z'/G$. Then we have stratifications $Z' = \coprod_{a \in A'} Z'_a$ and $X' = \coprod_{a \in A'} X'_a$ analogous to those for Z and X. Call a map $f: M \to M'$ *allowable* if

(i) f is smooth and G-equivariant;

(ii) $f(Z) \subseteq Z'$;

(iii) $df(T_z Z_{\text{prin}}) \subseteq T_{f(z)} Z'_{a(z)}$ for all $z \in Z_{\text{prin}}$, where $Z'_{a(z)} \subseteq Z'$ is the stratum of $f(z)$.

For instance, if f is smooth and equivariant and maps Z_{prin} into a single stratum of Z', then f is allowable.

Example 6.2. Let (V, ω) be a symplectic vector space on which G acts linearly and symplectically. A moment map is given by $\Phi_V^\xi(v) = \frac{1}{2}\omega(\xi v, v)$, where $\xi \in \mathfrak{g}$ acts on V via the infinitesimal representation $\mathfrak{g} \to \mathfrak{sp}(V)$. Let $t \in \mathbb{R}$ and let $f: V \to V$ be the dilation $f(v) = tv$. Clearly f preserves Z. Furthermore, if $t \neq 0$, then $f(v)$ has the same stabilizer as v, so f maps Z_{prin} to itself. If $t = 0$, then f maps Z_{prin} to 0. In either case f maps Z_{prin} into a single stratum of Z and it is obviously smooth and equivariant, so it is allowable. Similarly, if $|t| \leq 1$ and B is a G-invariant open ball about the origin, the restriction of f is an allowable map from B to itself.

The following result is easy to deduce from Lemma 3.3.

Lemma 6.3. *Let $f\colon M \to M'$ allowable. Then the pullback homomorphism $f^*\colon \Omega(M') \to \Omega(M)$ sends $\Omega_{\Phi'}(M')$ to $\Omega_\Phi(M)$ and $I_{\Phi'}(M')$ to $I_\Phi(M)$, and therefore induces a homomorphism $f^*\colon \Omega(X') \to \Omega(X)$.*

Homotopies induce chain homotopies on the de Rham complex in a standard way. Let $F\colon M \times [0,1] \to M'$ be a smooth homotopy and put $F_t = F|_{M\times\{t\}}$. Let t be the coordinate on $[0,1]$ and for $\gamma \in \Omega(M')$ put $\kappa_F\gamma = \int_0^1 i(\partial/\partial t)F^*\gamma\, dt$. Then κ_F lowers the degree by 1 and

$$F_1^* - F_0^* = \kappa_F d + d\kappa_F.$$

Assume that F is equivariant with respect to the given G-actions on M and M' and the trivial action on $[0,1]$. It is straightforward to check that

$$\kappa_F \circ g^* = g^* \circ \kappa_F \qquad \text{for all } g \in G, \tag{6.4}$$

$$\kappa_F \circ i(\xi_{M'}) = -i(\xi_M) \circ \kappa_F \qquad \text{for all } \xi \in \mathfrak{g}. \tag{6.5}$$

Call the homotopy F *allowable* if

(i) F is smooth and G-equivariant;

(ii) $F_t\colon M \to M'$ is allowable for almost all $t \in [0,1]$;

(iii) $dF_{(z,t)}(\partial/\partial t) \in T_{F(z,t)}Z'_{a(z,t)}$ for almost all $t \in [0,1]$ and for all $z \in Z_{\mathrm{prin}}$, where $Z'_{a(z,t)} \subseteq Z'$ is the stratum of $F(z,t)$.

For instance, if F is smooth and equivariant and if there exists a single stratum Z'_a of Z' such that $F_t(Z_{\mathrm{prin}}) \subseteq Z'_a$ for almost all t, then F is allowable.

Example 6.6. Let (V,ω) be a symplectic representation space for G as in Example 6.2. The radial contraction $F\colon V \times [0,1] \to V$ given by $F(v,t) = tv$ is smooth and equivariant and satisfies $F_t(Z_{\mathrm{prin}}) \subseteq Z_{\mathrm{prin}}$ for $t \neq 0$. Hence it is allowable. Likewise, F defines an allowable homotopy $B \times [0,1] \to B$ for any G-invariant open ball B about the origin.

Lemma 6.7. *Let $F\colon M \times [0,1] \to M'$ be an allowable homotopy. Then the homotopy operator $\kappa_F\colon \Omega(M') \to \Omega(M)$ sends $\Omega_{\Phi'}(M')$ to $\Omega_\Phi(M)$ and $I_{\Phi'}(M')$ to $I_\Phi(M)$, and therefore induces a homotopy $\kappa_F\colon \Omega(X') \to \Omega(X)$.*

Proof. Let $\gamma \in \Omega^k_{\Phi'}(M')$. Then γ is invariant, so $\kappa_F\gamma$ is invariant by (6.4). Let $z \in Z_{\mathrm{prin}}$. Using (6.5) we find that for any multivector $v \in \Lambda^{k-1}(T_z Z_{\mathrm{prin}})$

$$i(\xi_M)(\kappa_F\gamma)_z(v) = \int_0^1 \phi(t)\, dt, \tag{6.8}$$

where $\phi(t) = -\gamma_{F(z,t)}(\xi_{M'}, F_*\partial/\partial t, (F_t)_*v)$. Let $Z'_{a(z,t)}$ be the stratum of Z' containing $F(z,t)$. Since F is allowable,

$$F_*\partial/\partial t \in T_{F(z,t)}Z'_{a(z,t)} \qquad \text{and} \qquad (F_t)_*v \in \Lambda^{k-1}\big(T_{F(z,t)}Z'_{a(z,t)}\big)$$

for most t. Moreover, by Lemma 3.3(i) the restriction of γ to $Z'_{a(z,t)}$ is horizontal. Hence $\phi(t) = 0$ for almost all t. From (6.8) we get $i(\xi_M)(\kappa_F\gamma)_z(v) = 0$; in other words $\kappa_F\gamma \in \Omega^{k-1}_\Phi(M)$. The inclusion $\kappa_F I_{\Phi'}(M') \subseteq I_\Phi(M)$ is proved in a similar way, and the last assertion now follows from the isomorphism (3.2). □

Example 6.9. Applying Lemma 6.7 to the radial contraction of Example 6.6 we find that the de Rham complex of the symplectic quotient of a vector space V is homotopically trivial. More generally, if $Y = (B \cap Z)/G$ is the symplectic quotient of any G-invariant open ball B about the origin, then the de Rham complex of Y is homotopically trivial.

Example 6.10. Let H be a closed subgroup of G and let V be a symplectic H-module. Let B be an H-invariant open ball about the origin and let O be the Hamiltonian G-manifold induced by B. Let Y be the symplectic quotient of B by the H-action and U the symplectic quotient of O by the G-action. Then $\Omega(U) \cong \Omega(Y)$ by Proposition 4.2, so $\Omega(U)$ is homotopically trivial by Example 6.9.

This example generalizes to arbitrary Hamiltonian G-manifolds by means of a slice argument. Let $z \in Z$ and let $H = G_z$ be the stabilizer of z. Consider the symplectic H-module $V = (T_zGz)^\omega / T_zGz$ known as the *symplectic slice* at z. Choose an H-invariant open ball B in V and let O be the G-space induced by B. The symplectic slice theorem due to Marle and to Guillemin and Sternberg (see [Sjamaar and Lerman 1991, §2], for instance) says that, for sufficiently small B, z has a G-invariant open neighborhood that is isomorphic to O as a Hamiltonian G-manifold. Hence the point $x \in X$ determined by z has an open neighborhood U for which $\Omega(U)$ is homotopically trivial. By letting B shrink to a point we obtain a collection of such neighborhoods, which is a basis of the topology at x. This finishes the proof of (6.1).

7. Integration and the symplectic class

In this section we show that top-degree forms on a compact symplectic quotient are always integrable and establish a version of Stokes' theorem. We conclude that the cohomology class of the symplectic form and its powers are nonzero.

For technical reasons we do not assume at the outset that X is compact. We start by introducing a metric on X_{prin} and demonstrating that X has "locally finite" volume. Choose a G-invariant compatible almost complex structure J on the Hamiltonian G-manifold M. The volume element determined by the Riemannian metric $\sigma = \omega(\cdot\,, J\,\cdot)$ is identical to the Liouville volume form $\omega^d/d!$ (where

$2d = \dim M$). The almost complex structure and Riemannian metric descend in a natural way to each stratum of X. Let $2n = \dim X$ and write $\mu = \omega_{\mathrm{prin}}^n/n!$ for the volume element of the principal stratum X_{prin}.

Lemma 7.1. *Every $x \in X$ has an open neighborhood U such that* $\operatorname{vol} U_{\mathrm{prin}}$ *is finite. Hence X_{prin} has finite volume if X is compact.*

Proof. Choose $z \in Z$ mapping to x and let $H = G_z$. By the symplectic slice theorem we may take U to be the symplectic quotient of an H-invariant neighborhood B of the origin in the symplectic slice V at z. The almost complex structure on M induces one on V, turning V into a unitary H-module. The metric on U_{prin} induced by the flat metric σ_V on V is quasi-isometric to the metric induced by σ. Therefore it is enough to show that U has finite volume with respect to the former. Let W be the orthogonal complement in V of the subspace of invariants V^H. The quadratic moment map Φ_V is constant along V^H, so $Z_V = V^H \times Z_W$, where $Z_V = \Phi_V^{-1}(0)$ and $Z_W = \Phi_V^{-1}(0) \cap W$. Let $B = B_1 \times B_2$, where B_1 is an open ball about the origin in V^H and B_2 an H-invariant open ball about the origin in W. Then B has a product metric and so do $(Z_V)_{\mathrm{prin}} = V^H \times (Z_W)_{\mathrm{prin}}$ and the quotient

$$U_{\mathrm{prin}} = B_1 \times (B_2 \cap (Z_W)_{\mathrm{prin}})/H. \tag{7.2}$$

Recall that the *metric cone* over a Riemannian manifold (Y, σ_Y) is the product $Y \times (0, 1)$ with metric $t^2\sigma_Y + dt \otimes dt$, where t is the coordinate on $(0, 1)$. The metric cone over Y has finite volume if Y does. For instance, the ball B_2 in W is the metric cone over the sphere $S = \partial B_2$. Similarly, with respect to the metric induced by σ_W, $B_2 \cap (Z_W)_{\mathrm{prin}}$ is a metric cone over $S \cap (Z_W)_{\mathrm{prin}}$. Upon taking quotients we see that $(B_2 \cap (Z_W)_{\mathrm{prin}})/H$ is a metric cone over $(S \cap (Z_W)_{\mathrm{prin}})/H$. The link $S \cap Z_W$ is the zero fiber of the moment map $v \mapsto \big(\Phi_W(v), \frac{1}{2}(1 - |v|^2)\big)$ for the $H \times \mathbf{U}(1)$-action on W, where $\mathbf{U}(1)$ acts by complex scalar multiplication. By induction on the depth of the stratification, the principal stratum of the symplectic quotient $(S \cap (Z_W))/H$ has finite volume. Hence $(B_2 \cap (Z_W)_{\mathrm{prin}})/H$ has finite volume and therefore, because of the product decomposition (7.2), so does U_{prin}. ◻

The Riemannian metric on M determines metrics on $\Lambda^k(TM)$ for all k. Let $|\beta| \in C^0(M)$ denote the pointwise norm of a form β on M. Similarly, for $\alpha \in \Omega(X)$ let $|\alpha| \in C^0(X_{\mathrm{prin}})$ denote the pointwise norm over the principal stratum. If α is induced by $\tilde{\alpha} \in \Omega_\Phi(M)$, then $|\tilde{\alpha}|$ is a G-invariant continuous function on M and

$$\pi_{\mathrm{prin}}^*|\alpha| \leq \iota_{\mathrm{prin}}^*|\tilde{\alpha}|. \tag{7.3}$$

The *support* of a form $\alpha \in \Omega(X)$ is its support as a section of the sheaf Ω. This is the same as the closure in X of the support of α considered as a form on X_{prin}. The estimate (7.3) implies that for $\alpha \in \Omega(X)$ with compact support the pointwise norm $|\alpha|$ is a bounded function on X_{prin} and therefore by Lemma 7.1 the global norm $\int_{X_{\mathrm{prin}}} |\alpha|\mu$ is finite. In particular, for α of top degree $2n$ the integral $\int_{X_{\mathrm{prin}}} \alpha$ is absolutely convergent.

We can now prove Stokes' theorem. The proof is based on the fact that the singular strata of X have codimension ≥ 2, which makes the boundary terms in the integral vanish.

Proposition 7.4. *$\int_{X_{\mathrm{prin}}} d\gamma = 0$ if $\gamma \in \Omega^{2n-1}(X)$ has compact support.*

Proof. We use the notation of the proof of Lemma 7.1. By using partitions of unity we can reduce the general case to the case where γ has compact support in an open subset U of the form $B_1 \times (B_2 \cap Z_W)/H$. Let $2m = \dim Z_W/H$. If $m = 0$ then U is nonsingular and the result follows from the usual version of Stokes' theorem, so we may assume $m \geq 1$. Let $\chi\colon [0, \infty) \to [0, 1]$ be a smooth function satisfying $\chi(t) = 0$ for t near 0 and $\chi(t) = 1$ for $t \geq 1$. Define a sequence of H-invariant functions $\tilde{\chi}_k\colon V \to [0, 1]$ for $k \geq 1$ by $\tilde{\chi}_k(v) = \chi(k|\mathrm{pr}_W v|)$, where $\mathrm{pr}_W\colon V \to W$ is the orthogonal projection. These functions descend to smooth functions $\chi_k\colon U \to [0, 1]$. The functions $1 - \chi_k$ are bump functions supported near the singularities of U. In fact the sets $S_k = \mathrm{supp}(1 - \chi_k)$ form a decreasing sequence satisfying

$$\bigcap_k S_k = B_1 \times \{0 \bmod H\}, \tag{7.5}$$

the most singular stratum of U. Therefore $\bigcap_k (S_k)_{\mathrm{prin}}$ is empty and

$$\left|\int_{X_{\mathrm{prin}}} d\gamma - \int_{X_{\mathrm{prin}}} \chi_k d\gamma\right| = \left|\int_{(S_k)_{\mathrm{prin}}} (1 - \chi_k) d\gamma\right| \leq C \operatorname{vol}(S_k)_{\mathrm{prin}} \to 0$$

as $k \to \infty$. (Here C is an upper bound for $|(1 - \chi_k)d\gamma|$.) This shows that

$$\int_{X_{\mathrm{prin}}} d\gamma = \lim_{k\to\infty} \int_{X_{\mathrm{prin}}} \chi_k d\gamma.$$

To see that this limit is 0 we use

$$\int_{X_{\mathrm{prin}}} \chi_k d\gamma = \int_{X_{\mathrm{prin}}} d(\chi_k \gamma) - \int_{X_{\mathrm{prin}}} d\chi_k \wedge \gamma.$$

Since $d(\chi_k\gamma)$ is supported away from the most singular stratum (7.5), we can assume by induction on the depth of the stratification that $\int_{X_{\mathrm{prin}}} d(\chi_k\gamma) = 0$. Moreover,

$$\left|\int_{X_{\mathrm{prin}}} d\chi_k \wedge \gamma\right| \leq \int_{X_{\mathrm{prin}}} |d\chi_k||\gamma|\mu \leq C \int_{(S_k)_{\mathrm{prin}}} |d\chi_k|\mu,$$

where C is an upper bound for $|\gamma|$. Let $\tilde{\rho}_k\colon W \to W$ be the dilation $v \mapsto kv$ and ρ_k the induced map on Z_V/H. Then $\chi_k = \chi_1 \circ \rho_k$ and $S_k = \rho_k^{-1}(S_1)$. It follows that $d\chi_k(x) = kd\chi_1(\rho_k(x))$. By (7.2), U_{prin} is the product of a ball and a metric cone, so $\mathrm{vol}(S_k)_{\mathrm{prin}} = k^{-2m}\,\mathrm{vol}(S_1)_{\mathrm{prin}}$, where $2m = \dim Z_W/H \geq 2$. Hence

$$\left|\int_{X_{\mathrm{prin}}} d\chi_k \wedge \gamma\right| \leq Ck^{1-2m}\int_{(S_1)_{\mathrm{prin}}} |d\chi_1|\mu \to 0$$

as $k \to \infty$. Therefore $\lim_{k\to\infty}\int_{X_{\mathrm{prin}}} \chi_k d\gamma = 0$. □

Stokes' theorem implies that the volume form of a compact quotient is not exact.

Corollary 7.6. *Suppose that X is compact. Then the class of ω^k_{prin} in $H^{2k}(\Omega(X))$ is nonzero for $0 \leq k \leq n$, where $2n = \dim X$.*

8. Generalizations

The results above can be generalized in two obvious ways. First we consider symplectic quotients at nonzero levels. Let $\mathbb{O}$ be a coadjoint orbit in $\mathfrak{g}^*$. The *symplectic quotient* at $\mathbb{O}$ is $X_{\mathbb{O}} = Z_{\mathbb{O}}/G$, where $Z_{\mathbb{O}}$ is the fiber $\Phi^{-1}(\mathbb{O})$. The spaces $Z_{\mathbb{O}}$ and $X_{\mathbb{O}}$ stratify in exactly the same way as when $\mathbb{O} = \{0\}$ and the strata of $X_{\mathbb{O}}$ again carry natural symplectic forms. Differential forms on $X_{\mathbb{O}}$ can now be defined as before. There is a symplectic slice theorem for orbits in $Z_{\mathbb{O}}$, so all our results generalize to this situation with virtually unchanged proofs.

Next we consider actions of a noncompact group G. The symplectic slice theorem remains valid, provided that G acts properly on M. For locally closed coadjoint orbits $\mathbb{O}$ stratifications of $Z_{\mathbb{O}}$ and $X_{\mathbb{O}}$ were obtained in [Bates and Lerman 1997]. However, our definition of forms on X is valid as it stands only when $\mathbb{O}$ is closed, because forms on a nonclosed subset may not extend to the ambient manifold. If $\mathbb{O}$ is locally closed we define $\Omega(X_{\mathbb{O}}) = \Omega_\Phi(N)/I_\Phi(N)$. Here $N = \Phi^{-1}(D)$ is the preimage of any G-invariant open neighborhood D of $\mathbb{O}$ in $\mathfrak{g}^*$ such that $\mathbb{O}$ is closed in D, $\Omega_\Phi(N)$ is the set of G-invariant forms on N that restrict to basic forms on $(Z_{\mathbb{O}})_{\mathrm{prin}}$, and $I_\Phi(N)$ is the set of G-invariant forms on N that restrict to 0 on $(Z_{\mathbb{O}})_{\mathrm{prin}}$. With this minor modification our results carry over to symplectic quotients by proper actions at locally closed coadjoint orbits. (For general orbits one might try to apply the methods developed in [Cushman and Śniatycki 2001], but we have not attempted this.)

Appendix: Forms on homogeneous bundles

Let G be a compact Lie group and H a closed subgroup. For any H-manifold F we can form the homogeneous fiber bundle with fiber F over G/H,

$$E = (G \times F)/H.$$

The map $f\colon F \to E$ defined by $f(p) = [1, p]$ identifies F with the fiber over the coset $0 \bmod H$. (Here $[g, p]$ denotes the coset $(g, p) \bmod H$ of $(g, p) \in G \times F$.)

Restriction to the fiber is a homomorphism

$$f^*\colon \Omega(E)^G \to \Omega(F)^H.$$

It is not hard to see that G-basic forms on E restrict to H-basic forms on F and that $f^*\colon \Omega_{\mathrm{bas}}(E) \to \Omega_{\mathrm{bas}}(F)$ is an isomorphism. We require a slight generalization of this elementary fact.

Choose an H-equivariant projection $\mathfrak{g} \to \mathfrak{h}$; this determines a G-invariant connection 1-form $\theta \in \Omega^1(G, \mathfrak{h})^{G\times H}$ on the principal H-bundle $G \to G/H$. Let VE be the vertical tangent bundle of E and let $\theta_E \in \Omega^1(E, VE)^G$ be the G-invariant connection 1-form on E associated to θ. Let $\gamma \in \Omega(F)^H$ be any invariant form on the fiber. Define a form $e(\gamma) \in \Omega(E)$ by putting

$$e(\gamma)_{[g,p]}(v) = \gamma_p\big((g^{-1})_*\theta_E(v)\big)$$

for $[g, p] \in E$ and $v \in \Lambda(T_{[g,p]}E)$. (For simplicity we write θ_E for the extension of the connection $\theta_E\colon TE \to VE$ to a multiplicative map $\Lambda(TE) \to \Lambda(VE)$.) The H-invariance of γ implies that $e(\gamma)_{[g,p]}(v)$ does not depend on the choice of the representative (g, p) of the coset $[g, p]$. The G-invariance of θ_E implies that $e(\gamma)$ is G-invariant. Thus we have defined a map

$$e\colon \Omega(F)^H \to \Omega(E)^G,$$

which we call the *extension homomorphism* determined by θ. (An alternative definition runs as follows. Let $\mathcal{V} = \mathrm{pr}^* TF$, where $\mathrm{pr}\colon G \times F \to F$ is the Cartesian projection. The vertical bundle of E is then the quotient $VE \cong \mathcal{V}/H$. A form $\gamma \in \Omega(F)^H$ is a section of $\Lambda(TF)$ and as such extends uniquely to a section $\tilde{\gamma}$ of $\mathcal{V}$ that is constant along G. Then $\tilde{\gamma}$ is $G \times H$-invariant and so descends to a G-invariant section $\bar{\gamma}$ of VE. Thus $e(\gamma) = \theta_E^*\bar{\gamma}$ is a G-invariant section of $\Lambda(TE)$. This argument also shows that $e(\gamma)$ is smooth.) The following result is immediate from the definition.

Lemma A.1. (i) *$f^*e(\gamma) = \gamma$ for $\gamma \in \Omega(F)^H$;*

(ii) *e maps $\Omega(F)_{\mathrm{bas}}$ to $\Omega(E)_{\mathrm{bas}}$;*

(iii) *$e(f^*\beta) = \beta$ for $\beta \in \Omega(E)_{\mathrm{bas}}$.*

It follows from (i) that $f^*\colon \Omega(E)^G \to \Omega(F)^H$ is surjective and from (ii)–(iii) that $f^*\colon \Omega_{\mathrm{bas}}(E) \to \Omega_{\mathrm{bas}}(F)$ is an isomorphism, as noted above. Now let F' be a second H-manifold and let $j\colon F \to F'$ be an H-equivariant map. Then j extends naturally to an G-equivariant bundle map $\bar{j}\colon E \to E' = (G \times F')/H$. Moreover θ_E is the pullback of the associated connection $\theta_{E'}$ on E'. This implies that the extension homomorphism is functorial in the following sense.

Lemma A.2. *$e \circ j^* = \bar{j}^* \circ e'$, where $e'\colon \Omega(F')^H \to \Omega(E')^G$ is the extension homomorphism for E'.*

References

[Arms et al. 1991] J. M. Arms, R. H. Cushman, and M. J. Gotay, "A universal reduction procedure for Hamiltonian group actions", pp. 33–51 in *The geometry of Hamiltonian systems* (Berkeley, 1989), edited by T. Ratiu, Math. Sci. Res. Inst. Publ. **22**, Springer, New York, 1991. MR 92h:58059 Zbl 0742.58016

[Bates and Lerman 1997] L. Bates and E. Lerman, "Proper group actions and symplectic stratified spaces", *Pacific J. Math.* **181**:2 (1997), 201–229. MR 98i:58085 Zbl 0902.58008

[Cushman and Śniatycki 2001] R. Cushman and J. Śniatycki, "Differential structure of orbit spaces", *Canad. J. Math.* **53**:4 (2001), 715–755. MR 2002j:53109 Zbl 01688933

[Godement 1973] R. Godement, *Topologie algébrique et théorie des faisceaux*, 3ème ed., Actualités scientifiques et industrielles **1252**, Hermann, Paris, 1973. MR 49 #9831 Zbl 0275.55010

[Kirwan 1985] F. C. Kirwan, "Partial desingularisations of quotients of nonsingular varieties and their Betti numbers", *Ann. of Math.* (2) **122**:1 (1985), 41–85. MR 87a:14010 Zbl 0592.14011

[Koszul 1953] J. L. Koszul, "Sur certains groupes de transformations de Lie", pp. 137–141 in *Géométrie différentielle* (Strasbourg, 1953), Colloques Internationaux du CNRS, Centre National de la Recherche Scientifique, Paris, 1953. MR 15,600g Zbl 0101.16201

[Lerman and Tolman 2000] E. Lerman and S. Tolman, "Intersection cohomology of S^1 symplectic quotients and small resolutions", *Duke Math. J.* **103**:1 (2000), 79–99. MR 2001k:53163 Zbl 0985.53041

[Marsden and Weinstein 1974] J. Marsden and A. Weinstein, "Reduction of symplectic manifolds with symmetry", *Rep. Mathematical Phys.* **5**:1 (1974), 121–130. MR 53 #6633 Zbl 0327.58005

[Sjamaar and Lerman 1991] R. Sjamaar and E. Lerman, "Stratified symplectic spaces and reduction", *Ann. of Math.* (2) **134**:2 (1991), 375–422. MR 92g:58036 Zbl 0759.58019

Received July 11, 2002.

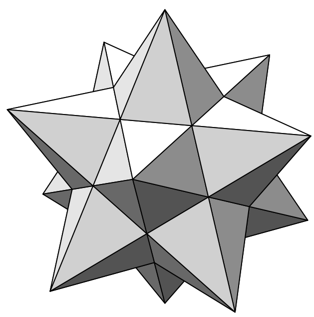
REYER SJAMAAR
DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NY 14853-4201
sjamaar@math.cornell.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

KEPLER'S SMALL STELLATED DODECAHEDRON AS A RIEMANN SURFACE

MATTHIAS WEBER

We provide a new geometric computation for the Jacobian of the Riemann surface of genus 4 associated to the small stellated dodecahedron. Starting with Threlfall's description, we introduce other flat conformal geometries on this surface which are related to holomorphic 1-forms. They allow us to show that the Jacobian is isogenous to a fourfold product of a single elliptic curve whose lattice constant can be determined in two essentially different ways, yielding an unexpected relation between hypergeometric integrals. We also obtain a new platonic tessellation of the surface.



1. Introduction

In his *Harmonice Mundi*, Kepler [1619] considers regular shapes in 2 and 3 dimensions. Besides the classical convex regular polygons he describes regular star polygons, so it is natural to allow also polyhedra that have such star polygons as faces. He comes up with several examples, among them the small stellated dodecahedron. It is therefore plausible that he didn't consider the 60 triangles of the stellated dodecahedron as its natural faces but the 12 star pentagons. This given, the polyhedron has 12 vertices and 30 edges, so the Euler formula gives

$$V - E + F = 12 - 30 + 12 = -6$$

MSC2003: 30F30.
Keywords: Jacobians, flat structures, small stellated dodecahedron.

which is not the Euler characteristic of the sphere but of a Riemann surface of genus 4. This was first observed by Poinsot and started some confusion about the validity of Euler's formula; see [Lakatos 1976].

All this can be resolved by viewing each star pentagon as a Riemann surface with a branch point in the center: The same way a regular pentagon is composed of 5 isosceles triangles with angle $2\pi/5$, the regular pentagram is composed by 5 isosceles triangles with angle $4\pi/5$. In fact, one can try to imagine the stellated dodecahedron as an immersed surface where each star pentagon is realized as a branched pentagon whose center branch point is hidden by a stellating pyramid. In this way, the stellated dodecahedron inherits from its singular euclidean metric a conformal structure and becomes a compact Riemann surface Σ of genus 4 whose automorphism group contains at least the icosahedral group.

This possibility was probably first observed by Klein [1877], who showed that the Riemann surface defined in $\mathbb{P}^4$ as the complete intersection

$$\sum_{i=1}^{5} z_i = 0, \quad \sum_{i=1}^{5} z_i^2 = 0, \quad \sum_{i=1}^{5} z_i^3 = 0$$

is biholomorphic to Kepler's small stellated dodecahedron. We will briefly discuss this in Section 4.

Threlfall [1932] gives a detailed description of the pentagon tessellation of this genus 4 surface Σ in terms of hyperbolic geometry. In particular, he finds another tessellation of the same surface by quadrilaterals such that 10 meet in one vertex. Because he is working in hyperbolic geometry, it is clear a priori that these two tessellations live on the same Riemann surface. Though Threlfall mentions the term Riemann surface frequently, he is interested neither in the properties of this particular surface as an algebraic curve nor in its automorphism group.

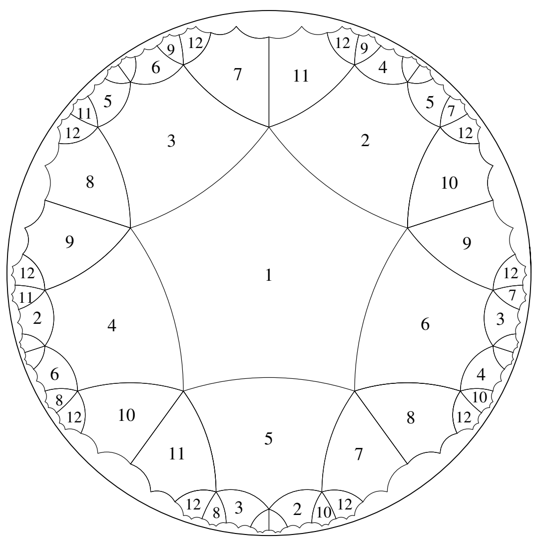
We will conformally replace the quadrilaterals in Threlfalls's description by other euclidean quadrilaterals to obtain new locally flat structures on the surface. These lead directly to a basis of holomorphic 1-forms by taking the exterior derivative of the developing maps of the flat structures. As the periods of the 1-forms are determined by the geometric data of the new metrics, we obtain easily a period matrix for the surface. In particular:

Theorem 1.1. *The Jacobian of Σ is isogenous to a* 4*-fold product of a rhombic torus. Its lattice constant can be computed either using the Schwarz–Christoffel formula for the new quadrilaterals or via the modular invariant of this torus.*

Remark. G. Riera and R. E. Rodríguez [1992] follow quite a different approach to compute the Jacobian of Σ: They first show that some 1-parameter family of polarized abelian varieties of dimension 4 is stabilized under the only 4-dimensional symplectic irreducible representation of S_5. Then they determine the parameter (implicitly) using an algebraic characterization of the quotient tori $\Sigma/\langle\phi\rangle$ and $\Sigma/(\mathbb{Z}/2\mathbb{Z})^2$ that differs from our description in Section 6.

2. A hyperbolic metric on the stellated dodecahedron

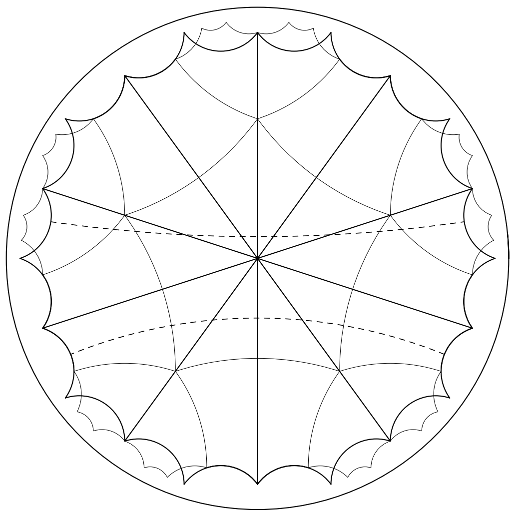
We now view the small stellated dodecahedron as a surface of genus 4, which comes with a natural tessellation by 12 star pentagons. Each star pentagon can be obtained by gluing together 5 isosceles euclidean triangles with obtuse angle $4\pi/5$. Map such a triangle conformally to a hyperbolic $(2\pi/5, 2\pi/10, 2\pi/10)$-triangle and continue this map by reflection first to the star pentagon. We obtain a conformal map from the star pentagon to a regular hyperbolic $2\pi/5$-pentagon. Continuing again by reflection to the whole surface yields a nonsingular conformal hyperbolic metric on the surface which is now tessellated by these hyperbolic pentagons. Here is the lift of this tessellation to the hyperbolic plane; the numbers designate the 12 faces:



Our next goal is to derive Threlfall's tessellation of the surface by hyperbolic quadrilaterals. The key for this is the rotation ρ of order-5 of the stellated dodecahedron around the axes through two opposite vertices. These vertices are two fixed points, but there are two more, namely the branch points of the dodecahedron faces which are intersected by the rotation axes. Hence the quotient $\Sigma/\langle\rho\rangle$ is a four-punctured sphere. More precisely:

Lemma 2.1. Σ *is a fivefold cyclic branched covering over the four-punctured sphere whose conformal structure is obtained by doubling a square. Using four branch slits* γ_i *from the center of one of the squares to the corners, the covering is* *given by gluing together five copies of the sphere thus slit, so that the left edge of slit γ_i of copy j is glued to the right edge of slit γ_i of copy $j+d_i$, where $d_i = 1, 2, 4, 3$.*

Sketch of proof. This statement can be proved by analyzing the next figure, where we have added to the 72° pentagon tessellation 10 fat hyperbolic $2\pi/10$-squares.



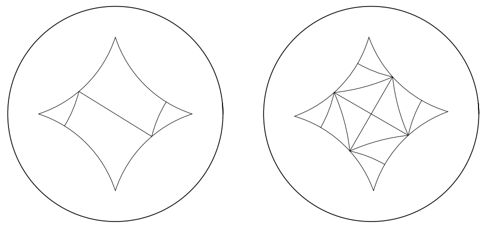
Using the figure on the previous page, one checks that these 10 squares constitute a fundamental domain for the surface. The edges are identified according to the two dashed geodesics and the order-5 rotational symmetry around the center of the figure. Now it is clear that two adjacent squares constitute a fundamental domain of the group $\langle\rho\rangle$ on Σ. The faces of these two squares have to be glued together by "flipping over", i.e., the quotient has the conformal structure claimed.

To see that the description of the covering in the lemma gives the same fundamental domain is straightforward; see [Threlfall 1932]. $\square$

We digress a bit to discuss also the other natural automorphisms of the surface:

The order-3 rotation around an axes through two opposite vertices of the *unstellated* dodecahedron defines a fixed point free automorphism τ of Σ which can be seen in the hyperbolic picture as a translation along the lower identification geodesic by 1/3 of its length. The quotient surface $\Sigma/\langle\tau\rangle$ is a nonsingular surface of genus 2 which comes with a tessellation by 4 hyperbolic 72°-pentagons; it is discussed in detail in [Threlfall 1932].

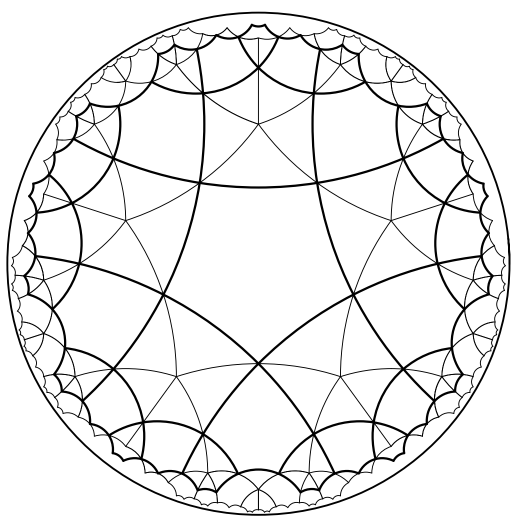
One can also obtain an order-2 rotation around the midpoints of the dodecahedron edges. But it turns out that this automorphism is actually the square of an order-4 rotation ϕ which is (of course) not an automorphism of the euclidean polyhedral structure on Σ but a conformal automorphism. That this rotation is really well defined on Σ becomes clear if we convince ourselves that the midpoints of some pentagon edges are also the centers of the quadrilaterals:



The left picture shows one of the quadrilaterals moved to a central position with the pentagon geodesics inside. Comparing the angles of the (congruent) triangles in the right picture with the two triangles in the left one shows easily the claimed symmetry.

To actually define this automorphism ϕ one can check that an order-4 rotation of one square is compatible with the identifications. One also finds a second fixed point, so that by the Riemann–Hurwitz formula, the quotient surface $\Sigma/\langle\phi\rangle$ is a torus. Because there are many different such automorphisms, this observation is the first indication that the Jacobian of Σ might be quite interesting. The investigation of this torus will be one of our primary goals.

Another way to see this automorphism is by looking at a new platonic tessellation of Σ by 24 right-angled regular pentagons:

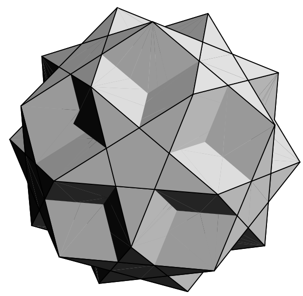


The figure shows the previous pentagon tessellation and the new one with thick lines. The order-4 rotation becomes a rotation around a vertex of this (preserved) tessellation. From this picture one can also deduce that ϕ has two fixed points.

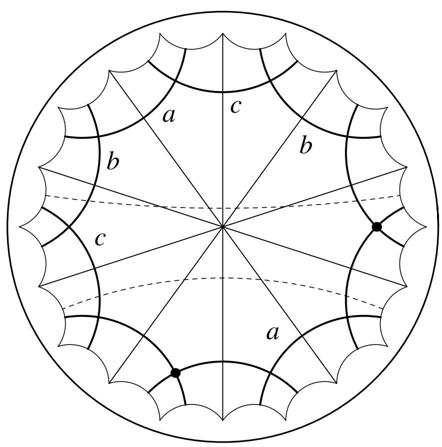
Furthermore, the thick lines are defined as geodesics connecting midpoints of adjacent pentagon edges: The sequence of edges hit by such a geodesic constitutes a *Petri polygon*; see [Coxeter and Moser 1972] for details.

The vertices of the 90° pentagons are either centers of the quadrilaterals or midpoints of the 72° pentagon edges.

This tessellation has also a euclidean realization as a euclidean uniform polyhedron, the so-called *dodecadodecahedron*, which is thus recognized as another (new) conformal version of Kepler's dodecahedron. This polyhedron has both regular pentagons and star-pentagons as faces:



The central right-angled regular decagon in the next figure shows a fundamental domain for the rotation ϕ on Σ. The fixed points are marked by a dot, and the nonadjacent edges are to be identified according to the labels.



This fundamental domain allows us to construct a degree-5 map from the quotient torus $T = \Sigma/\langle\phi\rangle$ to the sphere which is branched only over 3 points, as follows. Decompose the regular decagon into ten $(45°, 45°, 36°)$-triangles with vertices at the decagon vertices and its center. Map one of these triangles to the upper half-plane and continue by reflection. In principle, such a map pins down the conformal structure of the torus, but in general it is very hard to determine (say) the modular invariant of the torus from this map.

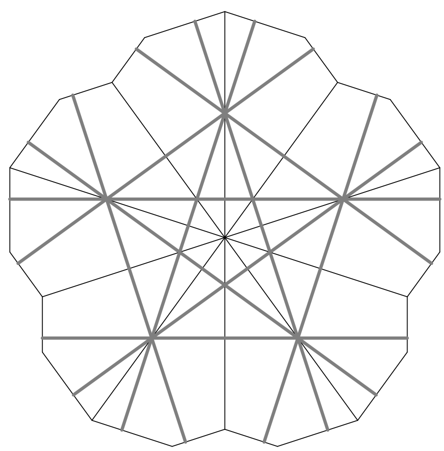
Proposition 2.2. *The automorphism group of* Σ *is* S_5, *the symmetric group of* 5 *elements.*

Proof. We know that $\operatorname{Aut}\Sigma$ contains the icosahedral group A_5 and has order at least 120. Assume that the automorphism group is strictly larger, that is, at least of order 240. Now the standard proof of Hurwitz's theorem about the order of the automorphism group of a compact Riemann surface forces $\operatorname{Aut}\Sigma$ to be a (2, 3, 7)-triangle group. But S_5 contains no element of order 7, so $\operatorname{Aut}\Sigma$ had to have at least $7 \cdot 120$ elements which contradicts the conclusion of Hurwitz's theorem. □

3. Σ as an algebraic curve

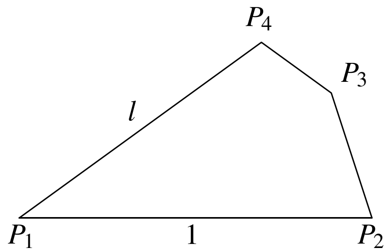
In this section, we construct a base of holomorphic 1-forms on Σ and derive an algebraic equation.

The first holomorphic 1-form ω_1 can be visualized by the following figure:



This is another fundamental domain of Σ, using euclidean quadrilaterals instead of hyperbolic $2\pi/10$-squares as in the figure on page 170. The identifications (which are indicated by the shaded lines) are realized by euclidean parallel translations. This is because we have chosen the quadrilateral with angles $\pi/5$, $2\pi/5$, $4\pi/5$, $3\pi/5$. Hence this description gives a singular flat metric on Σ with trivial linear holonomy. This means that the exterior derivative of the locally defined developing map of this flat metric is a globally well-defined holomorphic 1-form on Σ. Its zeros coincide with the singular points of this metric: Whenever the angles at a point add up to $k \cdot 2\pi$, the holomorphic 1-form will have a zero of order $k-1$.

Hence the 1-form ω_1 defined by the preceding figure has divisor $P_2+3P_3+2P_4$, where the points are located as follows:



Unfortunately, up to now we haven't proved that the fundamental domain above defines the correct conformal structure on Σ. In fact, this is impossible, because we haven't really specified *which* quadrilateral we are going to use for this construction. To guarantee that the resulting surface is biholomorphic to Σ, it is sufficient to ensure that the chosen quadrilateral is biholomorphic to any square, or, by the Riemann mapping theorem, to the upper half-plane with vertices at -1, 0, 1, ∞.

We do not know how explicitly it is possible to find such a quadrilateral, but at least we know these data in terms of Schwarz–Christoffel integrals. Denote by e_i the edge P_iP_{i+1}. Then

$$\begin{aligned}
e_1 &= \int_{-1}^{0} (t-1)^{-1/5} t^{-3/5} (t+1)^{-4/5} dt, \\
e_2 &= \int_{0}^{1} (t-1)^{-1/5} t^{-3/5} (t+1)^{-4/5} dt, \\
e_3 &= \int_{1}^{\infty} (t-1)^{-1/5} t^{-3/5} (t+1)^{-4/5} dt, \\
e_4 &= \int_{-\infty}^{-1} (t-1)^{-1/5} t^{-3/5} (t+1)^{-4/5} dt.
\end{aligned}$$

Denote by $l_i = |e_i/e_1|$ the corresponding normalized edge lengths, with $l = l_4$. By trigonometry,

$$l_1 = 1, \qquad l_2 = -1 + l\frac{\sqrt{5}+1}{2} \approx 0.373129,$$
$$l_3 = \frac{\sqrt{5}+1}{2}(1-l) \approx 0.244905, \quad l_4 = l \approx 0.848641.$$

Now three more holomorphic 1-forms ω_i can be defined using *the same* quadrilateral: Because it is conformally a square, we can permute the vertices cyclically.

This results in cyclically permuted divisors:

	P_1	P_2	P_3	P_4
ω_1	0	1	3	2
ω_2	1	3	2	0
ω_3	3	2	0	1
ω_4	2	0	1	3

Using this, we can derive an algebraic equation for Σ:

Proposition 3.1. Σ *is biholomorphic to the algebraic curve defined by the affine equation*

$$y^5 = (x+1)x^2(x-1)^{-1}.$$

Proof. Denote by $x : \Sigma \to \mathbb{P}^1$ the branched quotient map $\Sigma \to \Sigma/\rho$, where we choose the images of the branch points to be $-1, 0, 1, \infty$, which is possible by symmetry. Hence

$$\big((x+1)x^2(x-1)^{-1}\big) = P_1^5 + P_2^{10} + P_3^{-5} - P_4^{-10}.$$

Now put $y = \omega_2/\omega_1$ and obtain the same divisor for y^5. After scaling y appropriately, the equation follows. □

The function y will be explained geometrically in the next section.

4. Excursion: Bring's curve

In this section we show why the small stellated dodecahedron is biholomorphic to Bring's curve B, which is the complete intersection in $\mathbb{P}^4$ of the three hypersurfaces

$$\sum_{i=1}^{5} z_i = 0, \quad \sum_{i=1}^{5} z_i^2 = 0, \quad \sum_{i=1}^{5} z_i^3 = 0,$$

This was first shown by Klein [1877; 1884]. Bring's curve B occurs naturally as the locus of solutions of the reduced quintic equation

$$z^5 + pz + q = 0$$

because the vanishing of the coefficients of z^2, z^3, z^4 is equivalent to the equations above.

For projective properties of B, see [Edge 1978].

Following Klein, we first construct a threefold branched covering

$$\pi_1 : \Sigma \to \mathbb{P}^1$$

which is branched twice at all 72°-pentagon vertices. This is done by mapping the hyperbolic $(2\pi/5, 2\pi/10, 2\pi/10)$-triangle that constitutes one fifth of the tesselating 72°-pentagon onto a spherical $(2\pi/5, 2\pi/5, 2\pi/5)$-triangle, and continuing this map by reflection. The image of all the triangles will form the icosahedral tessellation of the sphere. Each vertex has two preimages: one is a branched pentagon vertex, the other an unbranched pentagon midpoint.

There is also a second such map π_2, using the dual 72°-pentagon tessellation instead. Both of these maps can be given explicitly in terms of the 1-forms ω_i: By considering divisors we see easily that (up to normalization)

$$\omega_1\omega_3 = \omega_2\omega_4,$$

so that we have an explicit equation of the quadric Q on which the canonical curve of Σ lies. Now the projections on the respective factors of $Q \simeq \mathbb{P}^1 \times \mathbb{P}^1$ are given by the meromorphic functions

$$z \mapsto \omega_2/\omega_1 \quad \text{and} \quad z \mapsto \omega_4/\omega_1,$$

which have precisely the same branching behavior as the functions π_i above. This shows also that π_1 is proportional to the function y from the last section. We leave to the reader the transformation of the ω_i to the z_j and the proof that the latter then satisfy the cubic equation as well. See also [Edge 1978; Klein 1884, 1877, Slodowy 1986].

5. The Jacobian of Σ

In this section, we compute the Jacobian of Σ in terms of tenth roots of unity and the constant l of Section 3, which is the ratio of two hypergeometric functions. This also allows us to compute the lattice of the quotient tori.

To compute the Jacobian, we first choose an appropriate base for the homology of Σ. This base will not be canonical but adapted to our representation of Σ as a branched covering over a 4-punctured sphere. Denote by c_k the curve on Y that winds k times around P_1, then once around P_2 and finally as often around P_1 as is necessary to lift to a closed curve on Σ. Similarly, denote by $\tilde{c}_k$ the curve on Y that winds k times around P_2, then once around P_3 and finally as often around P_2 as is necessary to lift to a closed curve on Σ.

For the holomorphic 1-forms, we take the ω_j of Section 3. Here we are still free to choose a normalization. Because we intend to compute also the lattice of the quotient torus of Σ by the order-4 rotation subgroup $\langle\phi\rangle$, we will eventually need a nonzero holomorphic 1-form that is invariant under this rotation ϕ and whose periods we can compute. If we normalize the ω_i in such a way that $\phi^*\omega_i = \omega_{i+1}$, the 1-form $\omega = \omega_1 + \omega_2 + \omega_3 + \omega_4$ will do. This normalization can be achieved by

(1) taking the same sized quadrilateral for the different 1-forms and just relabeling the vertices, and

(2) fixing the developing map for all of them simultaneously.

Using these two normalizations, we obtain

Lemma 5.1. *Denote by* $\zeta = e^{2\pi i/10}$ *and by* $\Phi = \frac{1}{2}(\sqrt{5}+1)$. *For* $i = 1, 2, 3, 4$, *set* $\alpha_i = 2^i\pi/5$ *reduced modulo* 2π. *Indices are to be taken cyclically. Then*

$$\int_{c_k} \omega_j = e^{ki\alpha_j} e_j (1 - e^{i\alpha_{j+1}}),$$
$$\int_{\tilde{c}_k} \omega_j = e^{ki\alpha_{j+1}} e_{j+1} (1 - e^{i\alpha_{j+2}})$$

Hence the period matrix of the Jacobian with respect to the ω_j *and the cycles* $c_0, \dots, c_3, \tilde{c}_0, \dots, \tilde{c}_3$ *is given by*

$$\Omega = \begin{pmatrix} \left.\begin{matrix} \zeta^{2k}(1-\zeta^4) \\ \zeta^{4k+7}(1-\zeta^8)(-1+l\Phi) \\ \zeta^{8k+6}(1-\zeta^6)\Phi(1-l) \\ \zeta^{6k+4}(1-\zeta^2)l \end{matrix}\right|_{k=0}^{3} & \left.\begin{matrix} \zeta^{4k+7}(1-\zeta^8)(-1+l\Phi) \\ \zeta^{8k+6}(1-\zeta^6)\Phi(1-l) \\ \zeta^{6k+4}(1-\zeta^2)l \\ \zeta^{2k}(1-\zeta^4) \end{matrix}\right|_{k=0}^{3} \end{pmatrix}$$

Proof. To compute the period of an ω_k, we use the definition of ω_k by a flat metric on the 4-punctured sphere which is given by doubling the quadrilateral of figure 9. Because the developing map of the flat metric is the integral of the corresponding 1-form, the period can be read off from the picture: Winding around a vertex P_j changes the direction into which we develop by the cone angle at P_j, and the loop from P_j to P_{j+1}, around this point and back to P_j contributes the factor $e_j(1 - e^{i\alpha_{j+1}})$. The rest is straightforward computation. □

For a similar computation, see [Karcher and Weber 1999].

This construction also shows that ρ acts on the 1-forms by multiplication with roots of unity:

$$\omega_1 \mapsto \zeta^2\omega_1, \quad \omega_2 \mapsto \zeta^4\omega_2, \quad \omega_3 \mapsto \zeta^8\omega_3, \quad \omega_4 \mapsto \zeta^6\omega_4.$$

This is because ρ changes the direction of the developing map by a rotation of order 5 if we choose the base point for the development in one of the fixed points, and the amount depends on the respective cone angle in this point.

Because we haven't normalized our homology base, the polarization of the Jacobian still has to be computed. We do this by giving the intersection matrix of the cycles:

Lemma 5.2. *The intersection matrix of the cycles* $c_0, \dots, c_3, \tilde{c}_0, \dots, \tilde{c}_3$ *is given by*

$$I = \begin{pmatrix} 0 & 1 & 1 & -1 & -1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 1 & 0 & 1 & 0 & -1 \\ -1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 & -1 & -1 & 0 \end{pmatrix}.$$

The proof is straightforward but tedious and we omit it.

The claims may be checked by verifying the Riemann period conditions

$$\Omega I^{-1}\Omega^t = 0 \quad \text{and} \quad -i\Omega I^{-1}\overline{\Omega}^t > 0.$$

In fact,

$$\begin{aligned} -i\Omega I^{-1}\overline{\Omega}^t &= \left(-5\,\zeta^2 - 5\,\zeta^3 + 10\,l\,\Phi\,\left(\zeta^2+\zeta^3\right) - 5\,l^2\,(1+\Phi)\,\left(\zeta+\zeta^4\right)\right) Id \\ &\approx 5.52531 Id. \end{aligned}$$

Corollary 5.3. *The lattice of the quotient torus* $\Sigma/\langle\phi\rangle$ *is spanned by*

$$\begin{aligned} \tau_1 &= (1+\zeta)^2\left(-1+l+\zeta-\zeta^2\right) \approx 1.79303 - 0.321884i, \\ \tau_2 &= (1+\zeta)\left(-1+2l-l\zeta+\zeta^2+l\zeta^2-l\zeta^3\right) \approx 1.26139 + 1.31433i, \\ \tau_2/\tau_1 &= \frac{-1+\zeta^2+(1+\zeta^{-1})l}{-1+\zeta^{-2}+(1+\zeta)l} = \bar{\zeta}\cdot\frac{l-\zeta(1-\zeta)}{l-\bar{\zeta}(1-\bar{\zeta})} \approx 0.554051 + 0.832482i. \end{aligned}$$

Proof. We have to show that the periods $\pi_j, \tilde{\pi}_j$ of $\omega = \omega_1+\omega_2+\omega_3+\omega_4$ constitute this lattice. By Lemma 5.1 we have $\tilde{\pi}_j = \pi_j$ and $\pi_0 = \tau_1,\ \pi_1 = \tau_2,\ \pi_2 = -2\tau_1+\tau_2,\ \pi_3 = 0,\ \pi_4 = \tau_1 - 2\tau_2.$ □

Remark. The specific value of l is only defined by the condition that our euclidean quadrilateral has to be a square. This also means that the formulas above do not make sense for any other surface.

We have computed the Jacobian of Σ and found at least three different quotient maps from Σ to tori. The relationship between all these tori will now be clarified.

Lemma 5.4. *Let* Γ *be a lattice in* $\mathbb{C}^n$ *and* $\alpha_1, \dots, \alpha_n$ *be* n *linearly independent linear functionals on* $\mathbb{C}^n$ *such that* $\Gamma_i = \alpha_i(\Gamma)$ *is a lattice in* $\mathbb{C}$. *Then* $\mathbb{C}^n/\Gamma$ *is isogenous to the product* $\mathbb{C}/\Gamma_1 \times \cdots \times \mathbb{C}/\Gamma_n$.

Proof. The regular linear map $\alpha_1 \times \cdots \times \alpha_n : \mathbb{C}^n \to \mathbb{C}^n$ induces a holomorphic Lie group homomorphism $\mathbb{C}^n/\Gamma \to \mathbb{C}/\Gamma_1 \times \cdots \times \mathbb{C}/\Gamma_n$. If this map had a nondiscrete kernel, there would be a $v \in \mathbb{C}^n - \{0\}$ such that $\alpha_i(v) = 0$ for all i, contradicting the linear independence of the α_i. □

Corollary 5.5. Jac Σ *is isogenous to the product* $T \times T \times T \times T$.

Proof. The idea is to conjugate the map ϕ by ρ to obtain enough different quotient maps to the same torus. In our base of the lattice, the functional $z \mapsto z_1+z_2+z_3+z_4$ describes the map to the quotient torus induced by the quotient map $\Sigma \to \Sigma/\langle\phi\rangle$. Now we can as well consider the quotient maps associated to the conjugate maps $\rho^{-k}\phi\rho^k$ which are different quotient maps to the same torus. By the definition of the ω_i, ρ acts on them by multiplication as

$$\omega_i \mapsto \zeta^{2^i}\omega_i.$$

Thus $\rho^{-1}\phi\rho$ acts as

$$\omega_1 \mapsto \zeta^8\omega_2, \quad \omega_2 \mapsto \zeta^6\omega_3, \quad \omega_3 \mapsto \zeta^2\omega_4, \quad \omega_4 \mapsto \zeta^4\omega_1$$

and hence the induced map from Jac $\Sigma \to$ Jac T is described by the functional $z \mapsto \zeta^4 z_1+\zeta^8 z_2+\zeta^6 z_3+\zeta^2 z_4$. Similarly, the functionals $z \mapsto \zeta^8 z_1+\zeta^6 z_2+\zeta^2 z_3+\zeta^4 z_4$ and $z \mapsto \zeta^2 z_1 + \zeta^4 z_2 + \zeta^8 z_3 + \zeta^6 z_4$ describe the maps induced by $\rho^{-2}\phi\rho^2$ and $\rho^{-3}\phi\rho^3$. These 4 functionals are clearly independent, and the claim follows from the previous lemma. □

Corollary 5.6. *All holomorphic image tori of* Σ *are isogenous.*

Proof. Any holomorphic surjective map $f : \Sigma \to E$ to an elliptic curve induces a group homomorphism f : Jac $\Sigma \to$ Jac $E = E$. This map cannot be trivial on all factors of Jac Σ; hence there is a nontrivial restriction $f_1 : T \to E$ that is necessarily a covering. □

6. An algebraic equation for the quotient torus

In this section we derive an algebraic equation for the quotient torus $T = \Sigma/\langle\phi\rangle$ and compute its modular invariant. The arithmetic nature of this torus has been investigated by Serre [1980], and an equation is given (without proof) in [Slodowy 1986].

Our strategy for producing such an equation is as follows: Using the representation of Σ as a branched covering over the four-punctured sphere, we construct a degree-3 function y and a degree-4 function w on Σ having poles of order at most 2 and 3, respectively, and only at the branch points of the covering $\pi : \Sigma \to \Sigma/\langle\rho\rangle$. Averaging this function over the action of ϕ yields functions of degrees 2 and 3 on the quotient torus T. To determine an equation, we investigate these functions at their poles.

To start, we need to understand the action of ϕ in terms of the equation

$$y^5 = (x+1)x^2(x-1)^{-1}$$

(see Section 3). Recall that y represents a function on Σ with divisor $P_1 + 2P_2 - P_3 - 2P_4$ and x has branch points of order 5 with values $-1, 0, 1, \infty$ at the P_i. This implies that the new function

$$z = y^2/x$$

has divisor $2P_1 - P_2 - 2P_3 + P_4$ and is therefore proportional to the function π_2 from Section 5. From the two equations above one easily obtains

$$(*) \qquad yz^2 = \frac{y^2+z}{y^2-z}$$

and this equation reflects the order-4 automorphism ϕ as the map

$$y \mapsto z \quad z \mapsto -1/y.$$

Hence the average

$$Y = y + z - 1/y - 1/z$$

of Y will descend to T as a function with one double-order pole at the image of the P_i. Similarly, the function

$$w = y/z$$

on Σ has divisor $-P_1 + 3P_2 + P_3 - 3P_4$ and the average

$$W = \frac{y}{z} - \frac{1}{yz} + \frac{z}{y} - yz$$

descends to T as a function with one triple-order pole at the image of the P_i. We keep the names Y and W for the functions on T.

This means that there are constants $a, b, c, d, e, f \in \mathbb{C}$ such that

$$(**) \qquad (W - aY)^2 - bY^3 - cY^2 - dW - eY - f \equiv 0.$$

To determine them, we compute this expression on Σ in a neighborhood of P_1, using y as a local coordinate. Note that

$$z = -y^2 + O(y^7)$$

because $x = z/y^2$ has a branch point of order-5 with value -1 at P_1. This leads to

$$\frac{1-b}{y^6} + \frac{-2a+3b}{y^5} + \frac{-2+2a+a^2-3b-c}{y^4} + \frac{2a-2a^2-2b+2c-d}{y^3} \\ + \frac{-1-4a+a^2+9b-c-e}{y^2} + O(y^{-1}) = 0,$$

which determines the first 5 constants as

$$a = \tfrac{3}{2}, \quad b = 1, \quad c = \tfrac{1}{4}, \quad d = -3, \quad e = 4.$$

Putting this back into $(**)$ gives

$$h = (y^3 + y\,z + y^5\,z + y^2\,z^2 + y^3\,z^2 - y^4\,z^2 + z^3 + y^2\,z^3 - 4\,y^3\,z^3 - y^4\,z^3 - y^6\,z^3 \\ - y^2\,z^4 - y^3\,z^4 + y^4\,z^4 + y\,z^5 + y^5\,z^5 - y^3\,z^6)/(y^3\,z^3),$$

which reduces to -4 using $(*)$.

Hence we obtain the desired equation in Y and W:

$$4 - 4\,Y - \frac{Y^2}{4} - Y^3 + 3\,W + \Big(\frac{-3\,Y}{2} + W\Big)^2 = 0.$$

In new variables this equation can be brought into the form

$$y^2 = 4x^3 - 75x - 1475.$$

These equations allow to compute the modular invariant λ of T as the cross ratio of ∞ and the three algebraic numbers

$$\tfrac{1}{8}\big(\big(\tfrac{11}{5}\big)^{1/3}(59 - 24\sqrt{6})^{1/3} + 5^{2/3}(59 + 24\sqrt{6})^{1/3} - 13\big),$$
$$(5^{2/3}/16)\,\big((-1 + i\sqrt{3})(59 - 24\sqrt{6})^{1/3} - (1 + i\sqrt{3})(59 + 24\sqrt{6})^{1/3} - 26\big),$$
$$(5^{2/3}/16)\,\big(-(1 + i\sqrt{3})(59 - 24\sqrt{6})^{1/3} - (1 - i\sqrt{3})(59 + 24\sqrt{6})^{1/3} - 26\big),$$

which gives roughly

$$\lambda \approx 0.660609 - 0.75073i.$$

This modular invariant can be used to compute the periods of the quotient torus in a different way. One obtains the period quotient τ_2/τ_1 of T as a quotient of two hypergeometric integrals, but this time as

$$\frac{\tau_2}{\tau_1} = \frac{\int_1^\infty u^{-1/2}(u-1)^{-1/2}(u-\lambda)^{-1/2}du}{\int_0^1 u^{-1/2}(u-1)^{-1/2}(u-\lambda)^{-1/2}du}.$$

Combining this expression with Corollary 5.3 gives an unexpected identity between hypergeometric integrals.

Acknowledgment

The author thanks Hermann Karcher for his interest.

References

[Coxeter and Moser 1972] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, Third ed., Ergebnisse der Math. und ihrer Grenzgebiete **14**, Springer, New York, 1972. MR 50 #2313 Zbl 0239.20040

[Edge 1978] W. L. Edge, "Bring's curve", *J. London Math. Soc.* (2) **18**:3 (1978), 539–545. MR 80a:14014 Zbl 0397.51013

[Karcher and Weber 1999] H. Karcher and M. Weber, "The geometry of Klein's Riemann surface", pp. 9–49 in *The eightfold way*, Math. Sci. Res. Inst. Publ. **35**, Cambridge Univ. Press, Cambridge, 1999. MR 2001a:14027 Zbl 0964.14029

[Kepler 1619] J. Kepler, *Harmonices mundi libri V*, Linz, 1619. Translated as *The harmony of the world* by E. J. Aiton et al., Philadelphia, Amer. Philos. Society, 1997.

[Klein 1877] F. Klein, "Weitere Untersuchungen über das Ikosaeder", *Math. Annalen* **12** (1877), 321–384.

[Klein 1884] F. Klein, *Vorlesungen über das Ikosaeder und die Auflösung der Gleichungen vom fünften Grade*, Teubner, Leipzig, 1884. Reprinted Birkhäuser, Basel, 1993 (edited by P. Slodowy); translated as *Lectures on the icosahedron and the solution of equations of the fifth degree*, Kegan Paul, London, 1913 (2nd edition); reprinted by Dover, 1953. MR 96g:01046

[Lakatos 1976] I. Lakatos, *Proofs and refutations: the logic of mathematical discovery*, Cambridge University Press, Cambridge, 1976. MR 58 #122 Zbl 0334.00022

[Riera and Rodríguez 1992] G. Riera and R. E. Rodríguez, "The period matrix of Bring's curve", *Pacific J. Math.* **154**:1 (1992), 179–200. MR 93e:14033 Zbl 0734.30036

[Serre 1980] J.-P. Serre, "Extensions icosaedriques", in *Séminaire de Théorie des Nombres de Bordeaux, Année* 1979–1980, exposé no. 19, 1980. Reprinted as pp. 550–554 in *Œuvres*, vol. III, Springer, Berlin, 1985. Zbl 0474.12011

[Slodowy 1986] P. Slodowy, "Das Ikosaeder und die Gleichungen fünften Grades", pp. 71–113 in *Arithmetik und Geometrie*, Math. Miniaturen **3**, Birkhäuser, Basel, 1986. MR 88i:01056

[Threlfall 1932] W. Threlfall, "Gruppenbilder", *Abh. Sächs. Akad. Wiss. Leipzig Math.-Natur. Kl.* **41**:6 (1932), 1–59. Zbl 0004.02202 JFM 58.0132.01

Received June 3, 1999. Revised January 27, 2004.

MATTHIAS WEBER
DEPARTMENT OF MATHEMATICS
DEPARTMENT OF MATHEMATICS
RAWLES HALL
INDIANA UNIVERSITY
BLOOMINGTON, IN 47405
matweber@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

SHARP ISOPERIMETRIC INEQUALITIES AND SPHERE THEOREMS

SHIHSHU WALTER WEI AND MEIJUN ZHU

Various relations between sharp isoperimetric inequalities and volumes of manifolds are studied. In particular, we introduce and estimate sharp isoperimetric constants τ^* and γ^* corresponding to two types of isoperimetric inequalities. We show that for a complete n-dimensional manifold M with Ricci curvature $\mathrm{Ric}(M) \geq n-1$, the volume of M is close to that of S^n if and only if τ^* is close to $n(n-1)/\big(2(n+2)\omega_n^{2/n}\big)$ and M is simply connected (for $n=2$ or 3), or γ^* is close to 1 (for any $n \geq 2$).

1. Introduction

A sharp Sobolev inequality of Aubin and Li [1999] states that on an n-dimensional smooth, compact, connected Riemannian manifold M, for $p \in (1, n)$ if $n \geq 4$, or for $p \in (1, \sqrt{n}\,) \cup (2, n)$ if $n = 2$ or 3, and for $r > r^* = np/(n+2-p)$, there exists a constant $A(p, r) > 0$ depending only on n, the bound on the injectivity radius, and the bound on the curvature tensor and its covariant derivatives on M such that, for all $\varphi \in W^{1,p}(M)$,

$$\left(\int_M |\varphi|^{p^*}\, dv\right)^{p/p^*} \leq K(n, p)^p \int_M |\nabla\varphi|^p dv + A(p, r) \left(\int_M |\varphi|^r\, dv\right)^{p/r}, \tag{1–1}$$

where $p^* = np/(n-p)$ and

$$K(n, p) = \frac{1}{n}\left(\frac{n(p-1)}{n-p}\right)^{(p-1)/p} \left(\frac{\Gamma(n+1)}{\Gamma\left(\frac{n}{p}\right)\Gamma\left(n+1-\frac{n}{p}\right) n\omega_n}\right)^{1/n},$$

for Γ the gamma function, ω_n the volume of the unit ball in $\mathbb{R}^n$, and dv the volume element of M. This inequality solves a conjecture raised by Aubin in the late 1970's; similar results were obtained independently in [Druet 1998]. It is natural

MSC2000: 58E35, 53C20, 53A99.
Keywords: isoperimetric inequality, Ricci curvature, sectional curvature, Sobolev inequality.

to ask whether (1–1) holds for $p = 1$ and $r = r^*$. Equivalently, does there exist, for every domain $\Omega \subset M$, a constant $C(M)$ depending on M such that

$$P^n \geq n^n \omega_n V^{n-1} \bigl(1 - C(M) V^{2/n}\bigr), \tag{1–2}$$

where $P = \operatorname{vol}_{n-1} \partial\Omega$ and $V = \operatorname{vol}_n \Omega$? It is well known that (1–2) does hold for a geodesic ball with small volume; see (2–2), for example.

The case $p = 1$ in (1–1) is not addressed in [Aubin and Li 1999]. On the other hand, an elegant local inequality due to Morgan and Johnson [2000] implies:

Theorem A. *If the sectional curvature K of M is less than K_0, then an enclosure of small volume V has perimeter P satisfying*

$$P \geq \bigl(1 - C K_0 V^{2/n}\bigr) P^*, \tag{1–3}$$

where C is a constant and P^* is the perimeter of the Euclidean ball of volume V.

This local result was previously only known for small geodesic balls — see (1–2) and (2–2). Equation (1–3) improved on the bound $P \geq \bigl(1 - C' V^{2/(n(n+3))}\bigr) P^*$ found in [Bérard and Meyer 1982] and valid for small volume V.

As a consequence of (1–3) we can make the following statement, valid even when V is not small, extending the Aubin–Li inequality (1–1) to the case $p = 1$ and $r = r^*$, and initiating the study of the isoperimetric inequality (1–4):

Theorem 1.1 (An isoperimetric inequality). *For every domain $\Omega \subset M$, there exists a constant $C(M)$ depending on M such that*

$$P^n \geq n^n \omega_n V^{n-1} \bigl(1 - C(M) V^{2/n}\bigr), \tag{1–4}$$

where $P = \operatorname{vol}_{n-1} \partial\Omega$, $V = \operatorname{vol}_n \Omega$, and ω_n is the volume of the unit ball in $\mathbb{R}^n$. (One can take, for example, $C(M) = \max\bigl\{nCK_0, \epsilon_0(M)^{-2/n}\bigr\}$, where CK_0 is as in (1–3) and $\epsilon_0(M) > 0$ is a constant depending on M so that (1–3) holds for small $V \leq \epsilon_0$.)

Remark. After we completed our work, we learned that Druet [2002] had given another proof of (1–4) by a different approach.

By a standard technique involving the coarea formula and Cavalieri's principle, we see that (1–4) is equivalent to the following:

Theorem 1.2 (A Sobolev inequality). *There exists a constant $A = A(M)$ such that for all $\varphi \in W^{1,1}(M)$,*

$$\left(\int_M |\varphi|^{n/(n-1)}\, dv\right)^{(n-1)/n} \leq K(n,1) \int_M |\nabla\varphi|\, dv + A(M) \left(\int_M |\varphi|^{n/(n+1)}\, dv\right)^{(n+1)/n},$$

where $K(n,1) = \lim\limits_{p\to 1} K(n,p) = (n\omega_n)^{-1/n}$.

The isoperimetric inequality (1–4) has its roots in global analysis and partial differential equations (see, for example, [Aubin and Li 1999]). The optimal constants in (1–4), too, will have geometric and even topological applications. An immediate example is that a sharp estimate on $C(M)$ in (1–4) in two dimensions will recapture the Bernstein isoperimetric inequality [1905] on S^2,

$$L^2 \geq 4\pi A\Big(1 - \frac{1}{4\pi} A\Big), \tag{1–5}$$

with equality if and only if the domain in question is a disk; see Theorem 1.3(I).

Now introduce, for an n-dimensional, smooth, compact, connected Riemannian manifold M, the *isoperimetric constant* $\tau^* = \tau^*(M)$, defined as the constant $C(M)$ that makes (1–4) sharp:

$$\tau^* := \inf \big\{ C(M) : C(M) \text{ is a constant such that (1–4) holds} \big\}. \tag{1–6}$$

The constant τ^* depends deeply on the geometric properties of the underlying manifold M. In turn, it may even completely determine the metric of M:

Theorem 1.3. *Let M be a complete, simply connected Riemannian manifold with* $\mathrm{Ric}(M) \geq n-1$.

(I) *The isoperimetric constant τ^* satisfies*

$$\tau^* \geq \tau_0 := \frac{n(n-1)}{2(n+2)\omega_n^{2/n}}. \tag{1–7}$$

For $n = 2$ or 3, we have $\tau^* = \tau_0$ if and only if M is isometric to S^n with the standard metric.

(II) *For $n = 2$ or 3, if the isoperimetric constant τ^* is close to τ_0, then* $\mathrm{vol}\, M$ *is close to* $\mathrm{vol}\, S^n$.

This theorem is sharp, and generalizes the Bernstein inequality (1–5). Also, the assumption of simple connectedness is necessary for the last sentence of (I), as can be seen from the example of three-dimensional real projective space, which is complete, not simply connected, and satisfies $\tau^* = \tau_0$.)

Open Problem. For M of dimension $n \geq 4$, complete and simply connected, with $\mathrm{Ric}(M) \geq n-1$, does $\tau^* = \tau_0$ still imply that M is isometric to the standard unit sphere S^n?

In Section 5 we prove that $\tau^* = \tau_0$ also for $M = S^4$ and S^5:

Theorem 1.4. *For any domain Ω of volume V and perimeter P in S^n, where $n = 2, 3, 4$ or 5, and with τ_0 as in* (1–7), *we have*

$$P^n \geq n^n \omega_n V^{n-1}\big(1 - \tau_0 V^{2/n}\big). \tag{1–8}$$

Open Problem. Is the isoperimetric constant still τ_0 on the standard unit sphere S^n, for all $n \geq 6$? That is, does (1–8) (or equivalently (5–3) below) hold for $n \geq 6$?

Remark. For a complete manifold M with $\mathrm{Ric}(M) \geq n-1$, the equality $\tau^* = \tau_0$ implies that M is (positive) Einstein (see the proof of Theorem 1.3). This opens up the perspective of studying positive Einstein metrics via isoperimetric constants.

In high dimensions, we have an analog of Toponogov's version of S. Y. Cheng's Maximum Diameter Theorem, in the setting of the sharp isoperimetric inequality Theorem 1.3 being realized on the sphere:

Theorem 1.5. *If M is a complete, simply connected n-manifold of sectional curvature* $\mathrm{Sec}(M) \geq 1$ *and such that $\tau^*(M^n)$ is close to τ_0, then* $\mathrm{vol}\, M$ *is close to* $\mathrm{vol}\, S^n$ *for all $n \geq 2$.*

Open Problem. Does Theorem 1.5 remain true in dimensions $n \geq 4$ if one weakens the assumption that $\mathrm{Sec}(M) \geq 1$ to the assumption that $\mathrm{Ric}(M) \geq n-1$?

One may also investigate the converse of Theorem 1.3(II) on the estimates of τ^* under some assumptions on the Ricci curvature and volume of the manifold. This is related to the study of the second constant of sharp Sobolev inequalities (see, for example, [Hebey 1999]). However, we will show by an example that τ^* might not be close to τ_0 even if $C \geq \mathrm{Ric}(M) \geq n-1$ and $\mathrm{vol}\, M$ is close to $\mathrm{vol}\, S^n$. Therefore, under the assumption that M has bounded Ricci curvature, saying that $\mathrm{vol}\, M$ is close to $\mathrm{vol}\, S^n$ is not equivalent to saying that τ^* is close to τ_0. In an attempt to solve this problem of searching for a new equivalence, we turn to the isoperimetric inequality of Gromov [1980] (see also [Chavel 1993, Theorem 6.6]):

Theorem B (Gromov's isoperimetric inequality). *Given an n-dimensional compact manifold M with* $\mathrm{Ric}\,(M) \geq n-1$ *and a domain $\Omega \subset M$ with smooth boundary $\partial\Omega$, let $\Omega_0 \subset S^n$ be a spherical cap such that*

$$\frac{\mathrm{vol}\,\Omega_0}{\mathrm{vol}\, S^n} = \frac{\mathrm{vol}\,\Omega}{\mathrm{vol}\, M}. \tag{1–9}$$

Then

$$\mathrm{vol}\,\partial\Omega \geq \frac{\mathrm{vol}\, M}{\mathrm{vol}\, S^n} \cdot \mathrm{vol}\,\partial\Omega_0.$$

Thus it makes sense to consider, for a complete manifold M with $\mathrm{Ric}(M) \geq n-1$, Gromov's isoperimetric constant $\gamma^* = \gamma^*(M)$, defined by

$$\gamma^* := \sup\{\gamma(M) : \mathrm{vol}\,\partial\Omega \geq \gamma(M)\,\mathrm{vol}\,\partial\Omega_0 \text{ for any domain } \Omega \subset M\}, \tag{1–10}$$

where $\partial\Omega$ is smooth and $\Omega_0 \subset S^n$ is a spherical cap satisfying (1–9).

The isoperimetric constants $\tau^*(M)$ and $\gamma^*(M)$ open up a new perspective on complete manifolds M with $\mathrm{Ric}(M) \geq n-1$. In particular, there are a variety of equivalent ways of stating that γ^* is close to 1, such as the following:

Theorem 1.6. *Assume that M is complete with* $\operatorname{Ric}(M) \geq n-1$. *Then γ^* is close to* 1 *if and only if* $\operatorname{vol} M$ *is close to* $\operatorname{vol} S^n$ *for all $n \geq 2$.*

This provides a new approach to the relation $\operatorname{vol} M \sim \operatorname{vol} S^n$. Other equivalent relations [Colding 1996a; 1996b; Petersen 1999] involve the Gromov–Hausdorff distance, the radius, and the $(n+1)$-st eigenvalue. As a consequence of this work, Theorem 1.5, and work of Cheeger and Colding [1997], one can conclude:

Theorem 1.7. *Let M be complete with* $\operatorname{Ric}(M) \geq n-1$. *For all $n \geq 2$, the following properties* (1)–(5) *are equivalent and each of them implies property* (6):

(1) *γ^* is close to* 1.
(2) $\operatorname{vol} M$ *is close to* $\operatorname{vol} S^n$.
(3) *M is Gromov–Hausdorff close to S^n.*
(4) *M has radius close to S^n, where the radius of M is that of the smallest closed metric ball that covers M.*
(5) *The $(n+1)$-st eigenvalue is close to n.*
(6) *M is diffeomorphic to S^n.*

Corollary 1.8. *Let M be complete and simply connected with* $\operatorname{Sec}(M) \geq 1$ *if $n \geq 2$, or* $\operatorname{Ric}(M) \geq n-1$ *if $n = 2$ or* 3. *Then the properties* (2)–(6) *below are equivalent, each of them is implied by property* (1), *and each implies properties* (7)–(9):

(1) *τ^* is close to the constant* $\tau_0 := \dfrac{n(n-1)}{2(n+2)\omega_n^{2/n}}$.
(2) *γ^* is close to* 1.
(3) $\operatorname{vol} M$ *is close to* $\operatorname{vol} S^n$.
(4) *M is Gromov–Hausdorff close to S^n.*
(5) *M has radius close to S^n.*
(6) *The $(n+1)$-st eigenvalue is close to n.*
(7) *M is diffeomorphic to S^n.*
(8) *M has diameter close to S^n.*
(9) *The first eigenvalue is close to n.*

2. Proof of Theorem 1.3

We begin with the asymptotic formulas for the perimeter P and volume V of a geodesic ball $B_r(\bar{x})$ of scalar curvature $\operatorname{Scal}_{\bar{x}}(M)$ about a point $\bar{x}$ (see, for example, [Gallot et al. 1987, Theorem 3.98]):

$$\frac{P^n}{V^{n-1}} = n^n \omega_n \left(1 - \frac{\operatorname{Scal}_{\bar{x}}(M)}{2(n+2)} r^2 + O(r^4)\right). \tag{2–1}$$

Thus, for a domain that is a geodesic ball $B_r(\bar{x})$ with small volume,

$$P^n = n^n \omega_n V^{n-1} \left(1 - \frac{\mathrm{Scal}_{\bar{x}}(M)}{2(n+2)\omega_n^{2/n}} V^{2/n} + o(1) V^{3/n}\right), \tag{2–2}$$

where $o(1)$ is small and tends to 0 as $V \to 0$.

Since $\mathrm{Ric}(M) \geq n-1$, we have $\mathrm{Scal}_{\bar{x}}(M) \geq n(n-1)$ at any point of M; thus by (1–4) and (1–6),

$$\tau^* \geq \frac{\mathrm{Scal}_{\bar{x}}(M)}{2(n+2)\omega_n^{2/n}} \geq \frac{n(n-1)}{2(n+2)\omega_n^{2/n}} = \tau_0. \tag{2–3}$$

For $n=2$ or 3, if $\tau^* = \tau_0$ we know from (2–3) that $\mathrm{Scal}_x(M) \leq n(n-1)$ at any point x in M; thus $\mathrm{Ric}(M) = n-1$. This in turn implies that M has constant sectional curvature $K = 1$, and is therefore isometric to the standard unit sphere. On the other hand, for S^2, due to Gromov's isoperimetric inequality (Theorem B), we need only prove that for any spherical cap domain Ω, the equality in (1–7) holds. This is obvious, since in terms of the spherical coordinate θ (which measures down from the north pole) we have $L = 2\pi \sin\theta$ and $A = 2\pi \int_0^\theta \sin\alpha \, d\alpha$ for $0 \leq \theta \leq \pi$. It follows that $\tau^* = 1/4\pi$, and we recapture the standard Bernstein isoperimetric inequality (1–5). In the case of S^3, due to Theorem B, it suffices to prove that, for any spherical cap domain Ω,

$$P^3 \geq 36\pi V^2 \left(1 - \frac{3}{5}\left(\frac{4\pi}{3}\right)^{-2/3} V^{2/3}\right). \tag{2–4}$$

In terms of the spherical coordinate function, for $0 \leq \theta \leq \pi$,

$$P = 4\pi \sin^2\theta \quad \text{and} \quad V = 4\pi \int_0^\theta \sin^2\alpha \, d\alpha.$$

Viewing P as a function of V, we define

$$f(V) = P^3 - 36\pi V^2 \left(1 - \frac{3}{5}\left(\frac{4\pi}{3}\right)^{-2/3} V^{2/3}\right).$$

Direct computation yields

$$\frac{dP}{dV} = \frac{2\cos\theta}{\sin\theta}, \qquad \frac{d^2P}{dV^2} = \frac{-8\pi}{(4\pi \sin^2\theta)^2} = \frac{-8\pi}{P^2},$$

so that

$$\frac{df(V)}{dV} = 24\pi^2 \left(4\sin^3\theta\cos\theta - 3(2\theta - \sin 2\theta) + \frac{12}{5}\left(\frac{4}{3}\right)^{-2/3} (2\theta - \sin 2\theta)^{5/3}\right)$$

and

$$\frac{d^2 f(V)}{dV^2} = -96\pi \left(\sin^2\theta - \left(\frac{4}{3}\right)^{-2/3} (2\theta - \sin 2\theta)^{2/3}\right).$$

Specializing for $\theta = 0$ (so $V = 0$ and $P = 0$) we have

$$f(0) = \frac{df}{dV}(0) = \frac{d^2 f}{dV^2}(0) = 0,$$

Note that $d^2 f(V)/dV^2$ has the same sign as

$$\mu(\theta) = 2\theta - \sin 2\theta - \tfrac{4}{3}\sin^3\theta.$$

One easily checks that $\mu(0) = 0$, and $\mu'(\theta) = 4\sin^2\theta - 4\sin^2\theta\cos\theta > 0$ for $\theta \in (0, \pi)$. It follows that $f(V) \geq 0$. This completes the proof of part (I).

To prove part (II), first observe that vol $M \leq$ vol S^n by the Bishop volume comparison theorem [Bishop and Crittenden 1964]. For $n = 2$, if the statement were not true, there would exist $\delta > 0$ such that vol $M^2 \leq 4\pi - \delta$ for some manifold M^2 with

$$\tau^*(M^2) - \frac{1}{4\pi} \leq \frac{\delta}{8\pi(4\pi - \delta)}.$$

We then choose $\Omega = M \setminus B_\epsilon$, where B_ϵ is a small geodesic ball of radius ϵ in M. For such a domain Ω,

$$P^2 \geq 4\pi V \left(1 - \left(\frac{1}{4\pi} + \frac{\delta}{8\pi(4\pi - \delta)}\right) V\right)$$

which would imply that $0 \geq \frac{1}{2}\delta V > 0$ as $\epsilon \to 0$, a contradiction.

For $n = 3$, we need the following lemma, which is a slight variation on a convergence theorem due to Petersen [1998, 10.5.4, Theorem 5.10]:

Lemma 2.1. *Given $n \geq 2$ and $\lambda > 0$, there is an $\epsilon = \epsilon(n, \lambda) > 0$ such that any closed, simply connected Riemannian n-manifold (M, g) with $|\mathrm{Sec}(M) - \lambda| \leq \epsilon$ is $C^{1,\alpha}$-close to a metric of constant curvature λ.*

Proof of Lemma 2.1. By the Bonnet Theorem [1855], $\mathrm{Sec}(M) \geq \lambda - \epsilon$ implies $\mathrm{diam}(M) \leq \pi/\sqrt{\lambda - \epsilon}$. Then, for $n \geq 3$, replacing Cheeger's lemma (see [Petersen 1998, pages 300–301]) by Klingenberg's Theorem [1959], which implies that the injectivity radius is at least $\pi/\sqrt{\lambda + \varepsilon}$, one may readily modify [Petersen 1998, proof on page 312] to deduce the conclusion. For $n = 2$ instead of Klingenberg's Theorem one can use Synge's Theorem and [Carmo 1992, Proposition 3.4, p. 281]. $\square$

From (2–3) we can see that $\mathrm{Ric}(M) \to 2$ as $\tau^* \to \frac{3}{5}\left(\frac{3}{4\pi}\right)^{2/3}$. This implies that $\mathrm{Sec}(M) \to 1$, since on a 3-manifold M and for some constant K_0, there is equivalence between $\mathrm{Ric}(M) \equiv 2K_0$ and $\mathrm{Sec}(M) \equiv K_0$. Then from Lemma 2.1 we know that the metric of M converges to the standard metric of S^3 in the $C^{1,\alpha}$ topology as $\tau^* \to \frac{3}{5}\left(\frac{3}{4\pi}\right)^{2/3}$. This implies that vol $M \to$ vol $S^3 = 2\pi^2$, completing the proof of part (II) and of Theorem 1.3.

Remark. Conceivably, estimates on τ^* may yield estimates on the first eigenvalue. For instance, assuming that M is complete and simply connected, and that $\mathrm{Ric}(M) \geq n-1$, then λ_1 is close to n if τ^* is close to τ_0 for $n=2$ or 3. This can be proved as follows. According to Theorem 1.3 we know that $\operatorname{vol} M$ is close to $\operatorname{vol} S^n$, thus $\mathrm{rad}(M)$ is close to π (see, for example, [Petersen 1999]). This of course yields that $\mathrm{diam}(M)$ is close to π. Then due to a theorem of Cheng [1975] we know that λ_1 is close to n (see Corollary 1.8).

3. A small manifold with large isoperimetric constant

We show that the converse of Theorem 1.3(II) is not true. Assume that $\mathrm{Ric}(M) \geq n-1$. For a geodesic ball $B_r(\bar{x})$ with small volume, we recall (2–2)

$$P^n = n^n \omega_n V^{n-1}\left(1 - \frac{\mathrm{Scal}_{\bar{x}}(M)}{2(n+2)\omega_n^{2/n}} V^{2/n} + o(1) V^{3/n}\right),$$

where $\mathrm{Scal}_{\bar{x}}(M)$ is the scalar curvature at point $\bar{x}$ and $o(1) \to 0$ as $V \to 0$.

One can check that, for $n=2$,

$$\frac{n(n-1)}{2(n+2)\omega_n^{2/n}} = ((n+1)\omega_{n+1})^{-2/n}.$$

If $\operatorname{vol} M \to \operatorname{vol} S^2$ implied that $\tau^* \to 1/4\pi$, then $\mathrm{Scal}_{\bar{x}}(M)$ would be less than $2+\delta$ as $\operatorname{vol} M \to \operatorname{vol} S^2$, for any $\delta > 0$. But the following example shows that this is impossible.

Example 3.1. For any small positive ϵ (less than $\frac{1}{100}$, say), define a C^2-smooth function by

$$f_\epsilon(x) = \begin{cases} \sqrt{1-(x-\epsilon)^2} & \text{if } -1+\epsilon \leq x \leq -\epsilon, \\ h_\epsilon(x) & \text{if } \quad -\epsilon \leq x \leq \epsilon, \\ \sqrt{1-(x+\epsilon)^2} & \text{if } \quad \epsilon \leq x \leq 1-\epsilon, \end{cases}$$

where h_ϵ is a symmetric function to be determined. Direct computation shows that

$$\begin{aligned} f'_\epsilon(-\epsilon) &= -f'_\epsilon(\epsilon) = 2\epsilon + o_\epsilon(1)\epsilon^2, \\ f''_\epsilon(\epsilon) &= f''_\epsilon(-\epsilon) = -1 + o_\epsilon(1)\epsilon, \end{aligned} \tag{3–1}$$

where $o_\epsilon(1)$ is small and tends to 0 as $\epsilon \to 0$. For a small $\epsilon > 0$, we choose a negative continuous symmetric function g_ϵ satisfying $g_\epsilon(\pm\epsilon) = f''_\epsilon(\epsilon) = f''_\epsilon(-\epsilon)$, $g'_\epsilon(x) < 0$ for $-\epsilon \leq x < 0$,

$$-5 \leq \min_{-\epsilon \leq x \leq \epsilon} g_\epsilon(x) \leq -2 \quad \text{and} \quad \int_{-\epsilon}^{\epsilon} g_\epsilon \, dx = 2 f'_\epsilon(\epsilon).$$

The existence of such a function is guaranteed by (3–1). We then define h_ϵ to be a symmetric function such that $h_\epsilon(-\epsilon) = \sqrt{1-4\epsilon^2}$ and

$$h'_\epsilon(x) = \int_{-\epsilon}^{x} g_\epsilon(s)\, ds + f'_\epsilon(-\epsilon) \quad \text{for } -\epsilon < x \le 0.$$

Let M_ϵ be the surface obtaining by rotating $y = f_\epsilon(x)$ around the x-axis. Recall that the Gaussian curvature K_ϵ is given by

$$K_\epsilon = -\frac{f''_\epsilon}{f_\epsilon(1+(f'_\epsilon)^2)^2},$$

where differentiation is with respect to x. It is easy to check that $K_\epsilon \ge 1 + o_\epsilon(1)$ and $\operatorname{vol} M_\epsilon = \operatorname{vol} S^2 + o_\epsilon(1)$, but K_ϵ is greater than $\frac{3}{2} + o_\epsilon(1)$ at the equator of M_ϵ, so the scalar curvature $\operatorname{Scal}_x(M_\epsilon)$ is at least $3 + o_\epsilon(1)$ at the equator of M_ϵ. By rescaling, one easily obtains a sequence of manifolds M_ϵ with Gaussian curvatures $K_\epsilon \ge 1$ and volumes $\operatorname{vol} M_\epsilon \to \operatorname{vol} S^2$, but with scalar curvatures $\operatorname{Scal}_x(M_\epsilon) > \frac{5}{2}$ at some points.

4. Proof of proximity results

Proof of Theorem 1.5. In view of (2–3), $\operatorname{Sec}(M) \to 1$ as $\tau^* \to \tau_0$. It then follows from Lemma 2.1 that the metric of M converges to the standard metric of S^n in the $C^{1,\alpha}$ topology as $\tau^* \to \tau_0$. This implies that $\operatorname{vol} M \to \operatorname{vol} S^n$. □

Proof of Theorem 1.6. Let M be complete with $\operatorname{Ric}(M) \ge n-1$. We claim that

$$\gamma^* \le 1. \tag{4–1}$$

If not, there is $\delta > 0$ such that

$$\operatorname{vol} \partial\Omega \ge (1+\delta) \operatorname{vol} \partial\Omega_0 \tag{4–2}$$

for any smooth domain $\Omega \subset M$. Now, [Morgan and Johnson 2000, Theorem 3.4] says that given V, the manifold M has regions of volume V and perimeter at most equal to the perimeter $P_0(V)$ of a ball of volume V in S^n. Choose $V = \frac{1}{2} \operatorname{vol} M$; since $\operatorname{vol} M \le \operatorname{vol} S^n$, we know that

$$P(V) \le P_0(V) \le \operatorname{vol} S^{n-1}.$$

However, from (4–2) we have

$$P(V) \ge (1+\delta) \operatorname{vol} S^{n-1},$$

which is a contradiction. This proves (4–1).

If $\operatorname{vol} M$ is close to $\operatorname{vol} S^n$, we know from Theorem B that

$$\gamma^* \ge \frac{\operatorname{vol} M}{\operatorname{vol} S^n} \to 1.$$

Combining this with (4–1) we get $\gamma^* \to 1$.

Conversely, if $\gamma^* \to 1$, we claim $\operatorname{vol} M \to \operatorname{vol} S^n$. Otherwise, there is $\delta > 0$ such that $\operatorname{vol} M \le \operatorname{vol} S^n - \delta$. Choose $V = \frac{1}{2} \operatorname{vol} M$ in [Morgan and Johnson 2000, Theorem 3.4] and let R be the region whose perimeter is $P(V)$; then

$$P(V) \le P_0(V) \le (1-\epsilon) \operatorname{vol} S^{n-1}$$

for some fixed $\epsilon = \epsilon(\delta) > 0$, since $\operatorname{vol} M \le \operatorname{vol} S^n - \delta$. Thus

$$\operatorname{vol} \partial R \le (1-\epsilon) \operatorname{vol} \partial R_0,$$

which contradicts the fact that $\gamma^* \to 1$. $\square$

5. Spheres in dimensions up to 5: Proof of Theorem 1.4

Thanks to Gromov's isoperimetric inequality, to prove Theorem 1.4 we need only show that (1–8) holds for any spherical cap domain Ω in S^n for $n = 4$ or 5 (the cases $n = 2, 3$ being covered by Theorem 1.3.

For $n = 4$, we must prove $P^4 \ge 4^4 \omega_4 V^3 \big(1 - \omega_4^{-1/2} V^{1/2}\big)$, where $\omega_4 = \pi^2/2$. Using the spherical coordinate θ that measures angles down from the north pole, we know that, for $0 \le \theta \le \pi$,

$$P = 2\pi^2 \sin^3\theta \quad \text{and} \quad V = 2\pi^2 \int_0^\theta \sin^3\alpha \, d\alpha.$$

Viewing P as a function of V, we define

$$f(V) = P^4 - 4^4 \omega_4 V^3 \left(1 - (\omega_4)^{-1/2} V^{1/2}\right).$$

Direct computation yields

$$\frac{df(V)}{dV} = 32\pi^6 \left(3 \sin^8\theta \cos\theta - 48A^2 + 112A^{5/2}\right),$$

where $A = A(\theta) = \int_0^\theta \sin^3\alpha \, d\alpha$. Since $f(0) = 0$ and $dV/d\theta \ge 0$, it suffices to show that $f_1(\theta) := df(V)/dV \ge 0$ for any $\theta \in (0, \pi)$.

Again, since $f_1(0) = 0$, it is enough to show that $df_1(\theta)/d\theta \ge 0$ for any $\theta \in (0, \pi)$. Equivalently, it suffices to show, for any $\theta \in (0, \pi)$, that

$$f_2(\theta) := 24 \sin^4\theta \cos^2\theta - 3 \sin^6\theta - 96A + 280A^{3/2} \ge 0.$$

Note again that since $f_2(0) = 0$, it is enough to show that $df_2(\theta)/d\theta \ge 0$ for any $\theta \in (0, \pi)$. Equivalently, we only need to show

$$f_3(\theta) := 162 \cos^3\theta - 66 \cos\theta - 96 + 420A^{1/2} \ge 0 \tag{5–1}$$

for $\theta \in (0, \pi)$. Since A is an increasing function of θ, we can check, for $\theta \geq \pi/2$, that

$$f_3(\theta) \geq 420A^{1/2}\pi/2 - 258 \geq 0.$$

To prove (5–1) for $\theta \leq \pi/2$, it is sufficient to show that

$$g_1(\theta) = 420^2 A - (162\cos^3\theta - 66\cos\theta - 96)^2 \geq 0,$$

for $\theta \in (0, \pi/2)$. Again, since $g_1(0) = 0$, it is enough to prove that $dg_1/d\theta \geq 0$ for $\theta \in (0, \pi/2)$. Equivalently, we need only show, for $\theta \in (0, \pi/2)$, that

$$g_2(\theta) := 420^2 \sin^2\theta - 2(162\cos^3\theta - 66\cos\theta - 96)(66 - 486\cos^2\theta) \geq 0. \tag{5–2}$$

To check this, we have, for $\theta \in (0, \pi/2)$, and setting $s := \sin\theta$, $c := \cos\theta$,

$$\begin{aligned}
&420^2 s^2 - 2(162c^3 - 66c - 96)(66 - 486c^2) \\
&\qquad = 420^2 s^2 - 2\left(-162cs^2 + 96(c-1)\right)(66 - 486c^2) \\
&\qquad = 420^2 s^2 + 324 \cdot 66cs^2 - 324 \cdot 486c^3 s^2 - 192(1-c)(486c^2 - 66) \\
&\qquad \geq 420^2 s^2 + 324 \cdot 66cs^2 - 324 \cdot 486cs^2 - 192(1-c)(486c^2 - 66) \\
&\qquad = 420^2 s^2 - 324 \cdot 420cs^2 - 192(1-c)(486c^2 - 66) \\
&\qquad \geq 420^2 s^2 - 324 \cdot 420s^2 - 192(1-c)(486c^2 - 66c^2) \\
&\qquad = 420 \cdot 96 \cdot \left(s^2 - 2(1-c)c\right) \\
&\qquad = 420 \cdot 96 \cdot 4\sin^2(\theta/2) \cdot \left(\cos^2(\theta/2) - c^2\right) \geq 0,
\end{aligned}$$

proving the case $n = 4$.

For general n, we note that (1–8) is equivalent to the integral inequality

$$\sin^{n(n-1)}\theta \geq n^{n-1}A^{n-1} - \frac{n^{n+(2/n)}(n-1)}{2(n+2)}A^{n-1+(2/n)} \tag{5–3}$$

for $\theta \in [0, \pi]$, where

$$A = \int_0^\theta \sin^{n-1}\alpha\, d\alpha.$$

For $n = 5$, we can follow the same argument used for $n = 4$ and find that it is enough to show that

$$f_4(\theta) = 128\sin^4\theta - 187\sin^2\theta + 5^{2/5} \cdot 11 \cdot 17 \cdot A^{2/5} \geq 0$$

for $\theta \in [0, \pi]$. Notice that $128\sin^4\theta - 187\sin^2\theta \leq 0$, and so it suffices to prove

$$g(\theta) = 5^2 \cdot 11^5 \cdot 17^5 \cdot A^2 + (128\sin^4\theta - 187\sin^2\theta)^5 \geq 0 \quad \text{for } \theta \in [0, \pi].$$

Since $g(0) = 0$, it is enough to show that, for $\theta \in [0, \pi]$, and with s, c as before,

$$g'(\theta) = 2 \cdot 5^2 \cdot 11^5 \cdot 17^5 \cdot A \cdot s^4 + 5(128s^4 - 187s^2)^4 \cdot (4 \cdot 128s^3c - 2 \cdot 187sc) \geq 0.$$

Let $g_1(\theta) = 5 \cdot 11^5 \cdot 17^5 \cdot A + (128s^3 - 187s)^4 \cdot (2 \cdot 128s^3c - 187sc)$. Note that $g_1(\theta)$ has the same sign as $g'(\theta)$ and that $g_1(0) = 0$, so we need only show that $g_1'(\theta) \geq 0$. Let

$$\begin{aligned} g_2(\theta) = 5 \cdot 11^5 \cdot 17^5 + 4 \cdot (128s^2 - 187)^3 (3 \cdot 128s^2c - 187c)(2 \cdot 128s^2c - 187c) \\ + (128s^2 - 187)^4 (6 \cdot 128s^2c^2 - 2 \cdot 128s^4 - 187c^2 + 187s^2). \end{aligned}$$

Note that $g_2(\theta)$ has the same sign as $g_1'(\theta)$ for $\theta \in [0, \pi]$. Then, by means of some delicate computations, we can check that $g_2(\theta) \geq 0$ for $\theta \in [0, \pi]$. It should be pointed out that $g_2'(\theta)$ is no longer nonnegative for $\theta \in [0, \pi]$.

Acknowledgment

The authors thank the referee for valuable comments and remarks, and Silvio Levy and Nicholas Jackson for a superb job in editing the article.

References

[Aubin and Li 1999] T. Aubin and Y. Y. Li, "On the best Sobolev inequality", *J. Math. Pures Appl.* (9) **78**:4 (1999), 353–387. MR 2000e:46041 Zbl 0944.46027

[Bérard and Meyer 1982] P. Bérard and D. Meyer, "Inégalités isopérimétriques et applications", *Ann. Sci. École Norm. Sup.* (4) **15**:3 (1982), 513–541. MR 84h:58147 Zbl 0489.35058

[Bernstein 1905] F. Bernstein, "Über die isoperimetrische Eigenschaft des Kreises auf der Kugeloberfläche und in der Ebene", *Math. Ann.* **60** (1905), 117–136.

[Bishop and Crittenden 1964] R. L. Bishop and R. J. Crittenden, *Geometry of manifolds*, Pure and Applied Mathematics **15**, Academic Press, New York, 1964. MR 29 #6401 Zbl 0132.16003

[Bonnet 1855] O. Bonnet, "Sur quelques propriétés des lignes géodésiques", *C. R. Acad. Sci. Paris* **40** (1855), 1311–1313.

[Carmo 1992] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, Boston, 1992. MR 92i:53001 Zbl 0752.53001

[Chavel 1993] I. Chavel, *Riemannian geometry—a modern introduction*, Cambridge Tracts in Mathematics **108**, Cambridge University Press, Cambridge, 1993. MR 95j:53001 Zbl 0810.53001

[Cheeger and Colding 1997] J. Cheeger and T. H. Colding, "On the structure of spaces with Ricci curvature bounded below, I", *J. Diff. Geom.* **46**:3 (1997), 406–480. MR 98k:53044 Zbl 0902.53034

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

[Colding 1996a] T. H. Colding, "Shape of manifolds with positive Ricci curvature", *Invent. Math.* **124**:1-3 (1996), 175–191. MR 96k:53067 Zbl 0871.53027

[Colding 1996b] T. H. Colding, "Large manifolds with positive Ricci curvature", *Invent. Math.* **124**:1-3 (1996), 193–214. MR 96k:53068 Zbl 0871.53028

[Druet 1998] O. Druet, "Optimal Sobolev inequalities of arbitrary order on compact Riemannian manifolds", *J. Funct. Anal.* **159**:1 (1998), 217–242. MR 99m:53076 Zbl 0923.46035

[Druet 2002] O. Druet, "Isoperimetric inequalities on compact manifolds", *Geom. Dedicata* **90** (2002), 217–236. MR 2003a:53045 Zbl 1025.58014

[Gallot et al. 1987] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, Universitext, Springer, Berlin, 1987. MR 88k:53001 Zbl 0636.53001

[Gromov 1980] M. Gromov, "Paul Lévy's isoperimetric inequality", preprint M/80/320, Inst. Haut Étud. Sci., 1980.

[Hebey 1999] E. Hebey, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*, Courant Lecture Notes in Mathematics **5**, New York University Courant Institute of Mathematical Sciences, New York, 1999. MR 2000e:58011 Zbl 0981.58006

[Klingenberg 1959] W. Klingenberg, "Contributions to Riemannian geometry in the large", *Ann. of Math.* (2) **69** (1959), 654–666. MR 21 #4445 Zbl 0133.15003

[Morgan and Johnson 2000] F. Morgan and D. L. Johnson, "Some sharp isoperimetric theorems for Riemannian manifolds", *Indiana Univ. Math. J.* **49**:3 (2000), 1017–1041. MR 2002e:53043 Zbl 1021.53020

[Petersen 1998] P. Petersen, *Riemannian geometry*, Graduate Texts in Mathematics **171**, Springer–Verlag, New York, 1998. MR 98m:53001 Zbl 0914.53001

[Petersen 1999] P. Petersen, "On eigenvalue pinching in positive Ricci curvature", *Invent. Math.* **138**:1 (1999), 1–21. MR 2001e:53032 Zbl 0988.53011

Received February 5, 2004. Revised August 3, 2004.

SHIHSHU WALTER WEI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73019
UNITED STATES

wwei@math.ou.edu

MEIJUN ZHU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73019
UNITED STATES

mzhu@math.ou.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

CORRECTION TO: EGGERT'S CONJECTURE ON THE DIMENSIONS OF NILPOTENT ALGEBRAS

LAKHDAR HAMMOUDI

Volume **202**:1 (2002), 93–97

The author acknowledges that the Theorem in the paper in question does not solve Eggert's conjecture completely, because it uses the fact that $R\cap(A\oplus\mathbb{K})=\{0\}$. Indeed, in the proof of the Theorem (page 96, lines 19 to 24), we assume that the unions in lines 17 and 18 are disjoint. If we add this hypothesis to the Theorem, which is fulfilled by graded algebras for example, the result is correct.

Therefore, the proof of the Theorem yields only a particular case of Eggert's conjecture. Eggert's conjecture in general remains open.

The author thanks Professors B. Amberg and L. Kazarin for pointing out the mistake and for the valuable discussions.

Received September 9, 2004.

LAKHDAR HAMMOUDI
DEPARTMENT OF MATHEMATICS
OHIO UNIVERSITY-CHILLICOTHE
101 UNIVERSITY DRIVE
CHILLICOTHE, OH 45601
UNITED STATES
hammoudi@ohio.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 220, No. 1, 2005

CORRECTION TO: MODULAR DIOPHANTINE INEQUALITIES AND NUMERICAL SEMIGROUPS

J. C. ROSALES, P. A. GARCÍA-SÁNCHEZ AND J. M. URBANO-BLANCO

Volume **218**:2 (2005), 379–398

The modular numerical semigroup $\mathrm{S}(2, 6) = \langle 3, 4, 5\rangle$ is pseudo-symmetric. Thus Corollary 60 of the paper is false, since it asserts that $\mathrm{S}(a, ab)$ is not pseudo-symmetric for any positive integers $a, b > 1$. The mistake comes from part (ii) of Proposition 58, which should read

$$S \text{ is pseudo-symmetric if and only if } (a-1, b) + (a-1) \bmod b = b - 1.$$

(The sign on right-hand side of this equality was incorrect.)

Received March 21, 2005.

J. C. ROSALES
DEPARTAMENTO DE ÁLGEBRA
UNIVERSIDAD DE GRANADA
E-18071 GRANADA
SPAIN
jrosales@ugr.es

P. A. GARCÍA-SÁNCHEZ
DEPARTAMENTO DE ÁLGEBRA
UNIVERSIDAD DE GRANADA
E-18071 GRANADA
SPAIN
pedro@ugr.es

J. M. URBANO-BLANCO
DEPARTAMENTO DE ÁLGEBRA
UNIVERSIDAD DE GRANADA
E-18071 GRANADA
SPAIN
jurbano@ugr.es

Guidelines for Authors

Authors may submit manuscripts for publication to any of the editors. Submission of a manuscript acknowledges that the paper is original and has not been submitted elsewhere. Information regarding the preparation of manuscripts is provided below. For further information, 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 submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Electronic submissions are strongly encouraged; the editors' email addresses are listed inside the front cover. Papers submitted in hard copy should be sent in triplicate and authors should keep a copy.

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 Web site 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 220 No. 1 May 2005

0030-8730(200505)220:1;1-N