Pacific Journal of Mathematics

Volume 163 No. 1 March 1994

CONTENTS

Vol. 163, No. 1 March 1994

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

WEIGHTED MAXIMAL FUNCTIONS AND DERIVATIVES OF INVARIANT POISSON INTEGRALS OF POTENTIALS

PATRICK AHERN AND WILLIAM COHN

In this paper we prove L^p estimates for weighted maximal functions of invariant Poisson integrals of potentials. From this it follows that the exceptional sets of the Poisson integrals of potentials are sets of zero Hausdorff capacity.

Let S denote the boundary of B_n, the unit ball in C^n, and let $d\sigma$ be the unusual rotation invariant measure defined on S. If g is a function belonging to the usual Lebesgue space $L^1(d\sigma)$ of functions defined on the sphere then by $P[g]$ we will mean the invariant Poisson integral of g defined by the equation

$$P[g](z) = \int_S g(\eta)\frac{(1-|z|^2)^n}{|1-\langle z, \eta\rangle|^{2n}}\, d\sigma(\eta),$$

where $z \in B_n$.

In this paper we will continue the work of Ahern and Cascante [**ACa**] and study invariant Poisson integrals of potentials of distributions in the atomic Hardy spaces H^p_{at} where $0 < p \le 1$. Precisely, if v denotes a distribution in the space H^p_{at} defined by Garnett and Latter and if $0 < \beta < n$ and $\zeta \in S$ define the (non-isotopic) potential of v by

$$I_\beta v(\zeta) = \int_S v(\eta)\frac{d\sigma(\eta)}{|1-\langle \zeta, \eta\rangle|^{n-\beta}}.$$

Let $f(z) = P[I_\beta v](z)$ and denote by f^*_α the admissible maximal function of f defined on the sphere S associated with the admissible approach region of aperture α. Thus, for each fixed $\alpha > 1$

$$f^*_\alpha(\zeta) = \sup_{w\in\Gamma_\alpha(\zeta)} |f(w)|,$$

where $\Gamma_a(\zeta)$ is the admissible approach region

$$\Gamma_\alpha(\zeta) = \{w \in B_n : |1-\langle w, \zeta\rangle| < \tfrac{\alpha}{2}(1-|w|^2)\}.$$

Suppose that μ is a positive measure on S satisfying the condition

$$\mu(B(\zeta\,;\delta)) \le C\delta^{n-\beta p} \tag{1}$$

for every Koranyi ball

$$B(\zeta\,;\delta) = \{\eta \in S : |1 - \langle \eta\,, \zeta\rangle| \le \delta\}$$

centered at ζ of radius δ contained in S. In **[ACa]** the following result is proved.

THEOREM 1. *Suppose that β is an integer between* 0 *and* $n-1$. *Let* μ *be a positive measure satisfying condition* (1). *Then with* v *and* f *related as above, there is a positive constant* C, *depending on* α *but independent of* v, *such that*

$$\int (f_\alpha^*)^p\, d\mu \le C\|v\|^p_{H^p_{at}}\,.$$

In this paper we will remove from Theorem 1 the restriction that β be an integer. In order to explain the method we pursue we first recall the basic idea used to establish Theorem 1.

For $z \in B_n$ let R be the operator given by

$$Rf(z) = \sum_{j=1}^{n} z_j D_j f(z)\,,$$

where $D_j = \frac{\partial}{\partial z_j}$ and let $\overline{R}$ be the operator given by

$$\overline{R}f(z) = \sum_{j=1}^{n} \overline{z}_j \overline{D}_j f(z)\,,$$

where $\overline{D}_j = \frac{\partial}{\partial \overline{z}_j}$. If $z = r\zeta$ where $\zeta \in S$ then it is easily verified that

$$\frac{\partial}{\partial r}(rf(r\zeta)) = (R + \overline{R} + \mathrm{id})f(z)\,.$$

From this it follows that

$$(2) \qquad (k-1)!f(z) = \int_0^1 \log^{k-1}\left(\frac{1}{t}\right)(R + \overline{R} + \mathrm{id})^k f(tz)\,dt\,.$$

In **[ACa]** it is shown that if $v \in H^p_{at}$ and $f = P[I_k v]$, then the admissible maximal function of $(R + \overline{R} + \mathrm{id})^k f(z)$ belongs to L^p. The argument used in **[A]** then can be applied to derive the conclusion of Theorem 1. For the case we are considering, that is, $f = P[I_\beta v]$ where β is not an integer, in order to use an argument patterned on the one above, we must find a suitable replacement for equation (2). The difficulty we face is that if we tailor the definition of $(R + \overline{R} + \mathrm{id})^k f(z)$ for non-integral k in such a way that equation (2) still holds then the methods of [ACa] are no longer sufficient by themselves to show the other fact that is needed, namely that the admissible maximal function of $(R+\overline{R}+\mathrm{id})^k f(z)$ is in L^p. (This problem does not occur if v belongs to the Hardy space H^p of holomorphic functions; see [A].) We circumvent this obstacle in the following fashion. With $f(z)=P[I_\beta](z)$ let

$$u(z)=(1-|z|)^{k-\beta}(R+\overline{R}+\mathrm{id})^k f(z),$$

where k is an integer greater than β but less than n. It can be verified that

$$(3)\qquad (k-1)!|z|f(z)=\int_0^{|z|}\left(\log\left(\frac{|z|}{t}\right)\right)^{k-1}(1-t)^{\beta-k}u(t\zeta)\,dt,$$

where $z=r\zeta$, and $\zeta\in S$. The main result of this paper will be the following theorem.

THEOREM 2. *Let* $v\in H^p_{at}$, $0<\beta<n-1$, *and* $f=P[I_\beta v]$. *If* k *is an integer greater than* β *but less than* n, *then the function* $u(z)=(1-|z|)^{k-\beta}(R+\overline{R}+\mathrm{id})^k f(z)$ *has admissible maximal function in* L^p.

Theorem 2 and the representation given by equation (3) can be used to apply the method of [A] to estimate f^*_α; the idea is that the factor $(\log(\frac{|z|}{t}))^{k-1}(1-t)^{\beta-k}$ will serve just as well as the factor $(\log(\frac{1}{t}))^{\beta-1}$ appearing in (2). We thus obtain the following corollary.

COROLLARY 1. *Theorem* 1 *remains true for all values of* β *between* 0 *and* $n-1$.

We will need to make use of the following objects. Let $\zeta\in S$ and for $1\le j,k\le n$ define the complex tangential vector field

$$T_{j,k}=\overline{\zeta}_j\frac{\partial}{\partial\zeta_k}-\overline{\zeta}_k\frac{\partial}{\partial\zeta_j}$$

and let $\overline{T}_{j,k}$ be the conjugate of $T_{j,k}$. Furthermore, let

$$L=\sum_{j<k}\overline{T}_{j,k}T_{j,k}$$

and

$$\overline{L}=\sum_{j<k}T_{j,k}\overline{T}_{j,k}.$$

If f is a function defined on B_n then for $z \in B_n$ with $z = r\zeta$ and $\zeta \in S$ we define

$$T_{j,k}f(z) = T_{j,k}f(r\zeta),$$

where the right-hand side is computed by holding r fixed and interpreting $f(r\zeta)$ as a function defined on the sphere. Then other operators above are also extended to act on functions on the ball in a similar fashion. We will need the following observations. Suppose that g is a smooth function of one complex variable. Let ζ and η range over the sphere S. Then

$$L_\zeta g(\langle\zeta, \eta\rangle) = \overline{L}_\eta g(\langle\zeta, \eta\rangle), \tag{4}$$

where the subscripts on the operators denote which variable the derivatives are taken with respect to. Furthermore, there is a second function, h, of one complex variable, such that

$$L_\zeta g(\langle\zeta, \eta\rangle) = h(\langle\zeta, \eta\rangle). \tag{5}$$

In fact, direct calculation shows that formula (4) is valid and that both expressions are equal to

$$(1 - |\langle\zeta, \eta\rangle|^2)D\overline{D}g(\langle\zeta, \eta\rangle) - (n-1)\langle\zeta, \eta\rangle Dg(\langle\zeta, \eta\rangle),$$

where D and $\overline{D}$ denote the usual operators $D = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y})$ and $\overline{D} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$. This proves the second assertion as well.

The following variants of the Poisson kernel used by Geller in [G] will also be of importance. For integers j and l let $P_{j,l}$ be the kernel

$$P_{j,l}(z, \eta) = \frac{(1-|z|^2)^{n+j+l}}{(1-\langle z, \eta\rangle)^{n+j}(1-\langle\eta, z\rangle)^{n+l}}.$$

These kernels will concern us when j and l are non-positive integers whose sum is greater than $-n$. Notice that $P_{0,0}$ is the usual Poisson kernel.

Before proceeding to the proof of Theorem 2 we will need some preliminary results. We remark that in what follows we will follow the custom of using the letter C to stand for a positive constant which changes its value from one appearance to another while remaining independent of the important variables.

LEMMA 1. *Let g and h be bounded functions defined on the unit ball in C^1 and suppose ζ and η are points on the sphere in C^n. Then*

$$\int_S g(\langle\zeta, \omega\rangle)h(\langle\omega, \eta\rangle)\, d\sigma(\omega) = \int_S h(\langle\zeta, \omega\rangle)g(\langle\omega, \eta\rangle)\, d\sigma(\omega).$$

Proof. Denote by $L(\zeta, \eta)$ the left-hand integral in the statement of the theorem and by $R(\zeta, \eta)$ the right-hand integral. For each ζ both expressions are continuous functions in the variable η. The desired conclusion will therefore follow if we show that for any smooth function ϕ defined on the sphere we have the equality

$$\int_S \phi(\eta)L(\zeta, \eta)\, d\sigma(\eta) = \int_S \phi(\eta)R(\zeta, \eta)\, d\sigma(\eta).$$

This in turn will follow if we show it to be true for all functions ϕ belonging to the space $H(p, q)$ of restrictions to S to homogeneous harmonic polynomials of bidegree (p, q) for all p and q. Now,

$$\begin{aligned}\int_S \phi(\eta)L(\zeta, \eta)\, d\sigma(\eta)\\ = \int_S g(\langle\zeta, \omega\rangle) \int_S h(\langle\omega, \eta\rangle)\phi(\eta)\, d\sigma(\eta)\, d\sigma(\omega).\end{aligned}$$

Let the inside integral of the right-hand side of the last equation define the operator

$$T(\phi)(\omega) = \int_S h(\langle\omega, \eta\rangle)\phi(\eta)\, d\sigma(\eta).$$

It is easily checked that T commutes with the usual action of the group of unitary operators on S. By Theorem 12.3.8 in **[R]** it follows that for all $\phi \in H(p, q)$

$$T(\phi) = C_h\phi,$$

where C_h is a constant depending only on h, p and q. Therefore

$$\int_S \phi(\eta)L(\zeta, \eta)\, d\sigma(\eta) = C_h \int_S g(\langle\zeta, \omega\rangle)\phi(\omega)\, d\sigma(\omega)$$

and by the same reasoning it follows that

$$\int_S \phi(\eta)L(\zeta, \eta)\, d\sigma(\eta) = C_h C_g \phi(\zeta).$$

An identical argument gives the formula

$$\int_S \phi(\eta)R(\zeta, \eta)\, d\sigma(\eta) = C_g C_h \phi(\zeta)$$

for all $\phi \in H(p, q)$, where the constants C_g and C_h are the same as before. This completes the proof. □

REMARK. The hypothesis that h and g be bounded is clearly not the weakest on h and g which allows some version of the conclusion of Lemma 1 to hold. If, for example, we assume only that the functions used in the proof are integrable on the sphere and therefore permit the application of Fubini's theorem, the argument will show that the equality of Lemma 1 holds almost everywhere $d\sigma(\zeta)\,d\sigma(\eta)$. In what follows, we will use this version of Lemma 1 whenever the hypotheses on g and h satisfy these less restrictive conditions.

While there is no natural group structure that allows us to define convolution, the Hermitian inner product provides a well-known substitute. If g is a function defined on the unit ball in C^1 and $\zeta \in S$ for a function F defined on the sphere let $F * g$ be given by

$$F * g(\zeta) = \int_S F(\eta) g(\langle \zeta, \eta \rangle)\, d\sigma(\eta).$$

The integral will be well-defined whenever $F \in L^1(d\sigma)$ and $g(\zeta_1) \in L^1(d\sigma)$. Here, of course, by ζ_1 we mean the first coordinate of the variable $\zeta \in C^n$. As a corollary of Lemma 1 we have the following result.

COROLLARY 2. *Let g and h be functions defined on the unit ball in C^1 such that both $g(\zeta_1)$ and $h(\zeta_1)$ are in $L^1(d\sigma)$. Let $F \in L^1(d\sigma)$. Then*

$$(F * g) * h = (F * h) * g.$$

Proof. The proof is accomplished through Fubini's theorem and the remark following Lemma 1. □

We will also need to notice that "convolution" commutes with the operators L and $\overline{L}$.

LEMMA 2. *Let F be a smooth function on S and g a smooth function of one complex variable. Let X be either L or $\overline{L}$. Then*

$$X(F * g)(\zeta) = (XF * g)(\zeta).$$

Proof. Use integration by parts and formula (4) to compute that

$$\begin{aligned} X(F * g)(\zeta) &= X_\zeta \int_S F(\eta) g(\langle \zeta, \eta \rangle)\, d\sigma(\eta) \\ &= \int_S F(\eta) X_\zeta g(\langle \zeta, \eta \rangle)\, d\sigma(\eta) \\ &= \int_S F(\eta) \overline{X}_\eta g(\langle \zeta, \eta \rangle)\, d\sigma(\eta) \\ &= \int_S XF(\eta) g(\langle \zeta, \eta \rangle)\, d\sigma(\eta) \\ &= (XF * g)(\zeta), \end{aligned}$$

as claimed. □

We will also need pointwise estimates on the derivatives of an invariant harmonic functions. See Theorem 1.2 of **[G]** for the analogous estimates associated with the Heisenberg group. Let $a \in B_n$ and for $\varepsilon > 0$ define

$$Q(a; \varepsilon) = \{w \in B_n : |1 - \langle w, a\rangle| < \varepsilon\}.$$

LEMMA 3. *Let U be an invariant harmonic function defined on B_n. If $a \in B_n$ let*

$$U^+(a) = \sup\left\{|U(w)| : w \in Q\left(a; \frac{1-|a|}{2}\right)\right\}.$$

Then for each pair of non-negative integers j and l there is a constant $C = C(j, l)$ independent of a or U such that

$$|\overline{R}^j R^l U(a)| \le C(1 - |a|)^{-j-l} U^*(a).$$

Proof. The proof is based on the same idea as the proof of Theorem 1.2 in **[G]**. For each $a \in B_n$ let ϕ_a be the automorphism of the ball given on page 25 of **[R]**. Let ψ be a smooth nonnegative function of a real variable supported on the interval $[0, s]$. We may choose s so small that for all a ϕ_a maps the ball in C^n centered at the origin of radius s into $Q(a; \frac{1-|a|}{2})$. Next, let $\Psi(w) = \psi(|w|)$ for $w \in B_n$. The argument used in **[G**, p. 130] (see also **[ACa**, equation 1.2]) shows that there is a constant C independent of U or a such that

$$U(a) = C\int_{B_n} U(w)\Psi(\phi_a(w))\, d\nu(w),$$

where $d\nu$ is the invariant measure

$$d\nu(w) = \frac{dV(w)}{(1 - |w|^2)^{n+1}},$$

and dV is Lebesgue measure on C^n. The desired estimate follows now by first differentiating under the integral sign, then using the fact that $\Psi(\phi_a(w))$ is supported on the set $Q(a; \frac{1-|a|}{2})$ together with the formula for $\phi_a(w)$ to bound the resulting expressions by $C(1 - |a|)^{-j-l}$, and finally observing that the invariant measure of $Q(a; \frac{1-|a|}{2})$ is bounded by a constant independent of a. This completes the proof. □

We are now ready to give the proof of Theorem 2.

Proof of Theorem 2. At times we will simplify the notation by suppressing the dependence of the admissible maximal function on the parameter α determining its aperture. The atomic decomposition of Garnett and Latter shows that Theorem 2 is a consequence of the following assertion.

Claim. Let a be a (p, ∞) atom in H^p_{at}. Suppose $f = P[I_\beta a]$ and $u(z) = (1 - |z|)^{k-\beta}(\overline{R} + R)^k f(z)$. Then there is a constant C depending only on α, k, and p, but not a such that

$$\int_S (u^*_\alpha)^p \, d\sigma \leq C.$$

We first give a detailed proof of the claim for the case where $0 < \beta < 1$ and $k = 1$. Since all the ideas necessary to establish the claim in full are present in this situation we will only sketch how the argument goes in general. Assume then that $0 < \beta < 1$, $k = 1$, and a is a (p, ∞) atom in H^p_{at}. We may assume that a is an atom centered at e_1 supported in the Koranyi ball

$$B(e_1; \delta) = \{\eta \in S : |1 - \eta_1| \leq \delta\},$$

where $e_1 = (1, 0, \ldots, 0)$. Recall that

$$|a| \leq \delta^{-n/p},$$

and that a has vanishing moments up to a certain order depending on p; see [GL] for details. We note for later use that the construction of the atomic decomposition given in [GL] shows that this order may actually be taken to be arbitrarily large. Let

$$u(z) = (1 - |z|)^{1-\beta}(\overline{R} + R)P[I_\beta a](z).$$

For λ a complex number in the unit disk and $r < 1$ define

$$P_r(\lambda) = \frac{(1 - r^2)^n}{|1 - r\lambda|^{2n}}.$$

If $z = r\zeta$ with $\zeta \in S$ and $F \in L^1(d\sigma)$ then we may write

$$P[F](z) = F * P_r(\zeta).$$

By Corollary 2 it follows that

$$P[I_\beta a](z) = (a * I_\beta) * P_r(\zeta) = (a * P_r) * I_\beta(\zeta).$$

Let V be the invariant harmonic function given by $V = P[a]$. Since $\overline{R} + R = r\frac{\partial}{\partial r}$ it follows that

$$(\overline{R} + R)P[I_\beta a](z) = \left(r\frac{\partial}{\partial r}a * P_r\right) * I_\beta(\zeta). \tag{6}$$

The right-hand side may be rewritten as

$$\int_S \frac{1}{|1-\langle\zeta, \eta\rangle|^{n-\beta}}(\overline{R}+R)V(r\zeta)\,d\sigma(\eta).$$

Notice that the operator $\overline{R}+R$ now acts on the variable $r\eta$. We may therefore write u as the sum of

$$u_1(z) = (1-|z|)^{1-\beta}\int_S \frac{1}{|1-\langle\zeta, \eta\rangle|^{n-\beta}} RV(r\eta)\,d\sigma(\eta) \tag{7}$$

and a similar expression, $u_2(z)$, which is obtained from the formula for u_1 by replacing R by $\overline{R}$. We proceed to show that there is a constant C independent of a such that

$$\int_S (u_1^*)^p\,d\sigma \leq C.$$

The same argument will establish the same inequality for u_2, and complete the proof for the case we are considering.

We first split u_1 into two parts. Let ψ be a non-negative $\mathscr{C}^\infty$ function supported on the disk in the complex plane centered at the origin of radius $\frac{1}{2}$ which is identically 1 on the disk centered at the origin of radius $\frac{1}{4}$. For $0 \leq r < 1$ and ζ and η in S let

$$\psi_r(\zeta, \eta) = \psi\left(\frac{1-\langle\zeta, \eta\rangle}{1-r}\right).$$

Then

$$u_1(z) = J_1(z) + J_2(z),$$

where

$$J_1(z) = (1-|z|)^{1-\beta}\int_S \frac{\psi_r(\zeta, \eta)}{|1-\langle\zeta, \eta\rangle|^{n-\beta}} RV(r\eta)\,d\sigma(\eta) \tag{8}$$

and

$$J_2(z) = (1-|z|)^{1-\beta}\int_S \frac{1-\psi_r(\zeta, \eta)}{|1-\langle\zeta, \eta\rangle|^{n-\beta}} RV(r\eta)\,d\sigma(\eta). \tag{9}$$

Consider first J_1. Let $\xi \in S$ and suppose that $z = r\zeta \in \Gamma_\alpha(\xi)$ for some aperture α. Since the integrand in J_1 vanishes for $|1-\langle\zeta, \eta\rangle| > \frac{1}{2}(1-r)$ it is easy to see that on the support of $\psi_r(\zeta, \eta)$ we may apply Lemma 3 to $RV(r\eta)$ to get the estimate

$$|RV(r\eta)| \leq C(1-r)^{-1}V^*(\xi), \tag{10}$$

provided that the maximal function V^* is taken with respect to an aperture equal to a fixed constant c times α, where c is independent of the atom a. It follows that

$$|J_1(z)| \leq C(1-r)^{-\beta} V^*_{c\alpha}(\xi) \int_S \frac{\psi_r(\zeta, \eta)}{|1 - \langle \zeta, \eta \rangle|^{n-\beta}} \, d\sigma .$$

If we use again the fact that $\psi_r(\zeta, \eta)$ is supported on a Koranyi ball centered at ζ of radius $1-r$ then the integral in the last inequality may be estimated as in [**ACo**, p. 427] to yield the conclusion that

$$|J_1(z)| \leq CV^*(\xi) .$$

From this it follows easily from the fact that a is a (p, ∞) atom that

$$\int_S (J_1^*)^p \, d\sigma \leq C \int_S (V^*)^p \, d\sigma \leq C ;$$

see [**GL**].

The analysis necessary to handle J_2 will be more complicated. We first make use of Theorem 1 and Lemma 1.4 from [**ACa**] together with Lemma 2 above to write

$$(11) \qquad -(n-1)RP[a](r\eta) = P_{0,-1}[La](r\eta) = LP_{0,-1}[a](r\eta) .$$

We remark that the equality of the first and last terms above may be verified directly by showing that

$$-(n-1)R_z P_{0,0}(z, \eta) = L_\zeta P_{0,-1}(z, \eta) .$$

In any event, let

$$G_r(\lambda) = \frac{(1-r^2)^{n-1}}{(1-r\lambda)^n (1-r\overline{\lambda})^{n-1}} ,$$

so

$$-(n-1)RV(r\eta) = RP[a](r\eta) = LP_{0,-1}[a](r\eta) = L(a * G_r)(\eta) .$$

Next let

$$-(n-1)K_r(\lambda) = \left(1 - \psi\left(\frac{1-\lambda}{1-r}\right)\right) |1-\lambda|^{\beta-n} .$$

Then we may write

$$J_2(z) = (1-r)^{1-\beta} (L(a * G_r)) * K_r(\zeta) .$$

Integration by parts shows that

$$\begin{aligned}(L(a * G_r)) * K_r(\zeta) &= \int_S L(a * G_r(\eta)) K_r(\langle \zeta, \eta \rangle) \, d\sigma(\eta) \\ &= \int_S (a * G_r(\eta)) \overline{L} K_r(\langle \zeta, \eta \rangle) \, d\sigma(\eta) .\end{aligned}$$

By formula (5) $\overline{L}K_r(\langle\zeta, \eta\rangle)$ is, for fixed r, a function of $\langle\zeta, \eta\rangle$. In fact, a calculation shows that $\overline{L}K_r(\langle\zeta, \eta\rangle)$ is the sum of

$$\left(1-\psi\left(\frac{1-\langle\zeta, \eta\rangle}{1-r}\right)\right)\overline{L}|1-\langle\zeta, \eta\rangle|^{\beta-n}$$

and three other terms each of which has a factor which is a derivative of ψ. Now, differentiating ψ yields a function which is supported on the region

$$\left\{\eta \in S : \frac{1-r}{4} < |1-\langle\zeta, \eta\rangle| < \frac{1-r}{2}\right\}.$$

These terms can then be handled in the same fashion as J_1 above. We therefore are left with the final task of estimating the admissible maximal function of

$$J_3(z) = (1-|z|)^{1-\beta}\int_S\left(1-\psi\left(\frac{1-\langle\zeta, \eta\rangle}{1-r}\right)\right) \cdot\overline{L}|1-\langle\zeta, \eta\rangle|^{\beta-n}(a * G_r(\eta))\,d\sigma.$$

To simplify the notation, let

$$Q_r(\langle\zeta, \eta\rangle) = (1-r)^{1-\beta}\left(1-\psi\left(\frac{1-\langle\zeta, \eta\rangle}{1-r}\right)\right)\overline{L}|1-\langle\zeta, \eta\rangle|^{\beta-n},$$

where Q_r is a function of one complex variable; equation (5) shows that this is possible. We therefore obtain the formula

$$J_3(z) = (a * G_r) * Q_r(\zeta).$$

Recall that the atom a is supported on the Koranyi ball

$$B(e_1; \delta) = \{\eta \in S : |1-\eta_1| \le \delta\},$$

where $e_1 = (1, 0, \dots, 0)$. We will need to partition unity in a manner that lets us take advantage of the support of a. It is possible to find smooth functions ϕ_0 and ϕ defined on the complex plane such that ϕ_0 is supported on the unit disk, ϕ is supported on the annulus $\{\lambda \in C^1 : 1/2 \le |\lambda| \le 2\}$ and

$$1 = \phi_0(\lambda) + \sum_{j=0}^{\infty}\phi\left(\frac{\lambda}{2^j}\right).$$

For η and $\tau \in S$ and $r\eta \in B_n$ let

$$\Phi_0(r\eta, \tau) = \phi_0\left(\frac{1-\langle r\eta, \tau\rangle}{8\delta}\right)$$

and for $j = 1, 2, \ldots, N$ let

$$\Phi_j(r\eta, \tau) = \phi\left(\frac{1 - \langle r\eta, \tau\rangle}{2^{j-1}8\delta}\right).$$

It follows that

$$1 = \sum_{j=0}^{N} \Phi_j(r\eta, \tau),$$

where N is a sufficiently large integer which depends only on δ.

We now write

$$a * G_r(\eta) = P_{0,-1}[a](r\eta) = \sum_{j=0}^{N} A_j(r\eta),$$

where

$$A_j(r\eta) = \int_S P_{0,-1}(r\eta, \tau)\Phi_j(r\eta, \tau)a(\tau)\, d\sigma(\tau).$$

We claim that there is an integer m that we may choose to be arbitrarily large (and whose choice will depend on p) such that

(i) $A_0(r\eta)$ is supported on the set

$$\{r\eta : |1 - r\eta_1| < 32\delta\};$$

(ii) $|A_0(r\eta)| \le C\delta^{-n/p}$;

(iii) For $j = 1, \ldots, N$, $A_j(r\eta)$ is supported on the set

$$\{r\eta : 2^{j-1}\delta < |1 - r\eta_1| < 32 \cdot 2^j g\}d;$$

(iv) For $j = 1, \ldots, N$

$$|A_j(r\eta)| \le C(2^j)^{-n-m+n/p}(2^j\delta)^{-n/p}.$$

Properties (i), (ii) and (iii) follow immediately from the definition of Φ_j, the support and size of the atom a, and the triangle inequality for the pseudometric $d(z, w) = |1 - \langle z, w\rangle|^{1/2}$ proved in **[R]**, Proposition 5.1.2. To verify property (iv) we must use the cancellation properties of the atom a in the usual way. For $2^{j-1}\delta < |1 - r\eta_1|$ estimate that

$$|A_j(r\eta)| \le \left|\int a(\tau)[P_{0,-1}(r\eta, \tau)\Phi_j(r\eta, \tau) - T_m(\tau, e_1)]\, d\sigma(\tau)\right|,$$

where for each fixed $r\eta$, $T_m(\tau, e_1)$ is the non-isotropic Taylor polynomial for $P_{0,-1}(r\eta, \tau)\Phi_j(r\eta, \tau)$ expanded about e_1 of degree m; see **[GL]** for the precise details. Since

$$2^{j-1}\delta < |1 - r\eta_1| < 32 \cdot 2^j\delta$$

we may estimate that

$$|P_{0,-1}(r\eta, \tau)\Phi_j(r\eta, \tau) - T_m(\tau, e_1)| \leq C\frac{\delta^m}{(2^j\delta)^{n+m}}.$$

It follows that

$$|A_j(r\eta)| \leq C\delta^{n-n/p}\frac{\delta^m}{(2^j\delta)^{n+m}},$$

as claimed.

We now write

$$J_3(z) = \sum_{j=0}^{N} F_j(r\zeta),$$

where

$$F_j(r\zeta) = \int_S A_j(r\eta)Q_r(\zeta, \eta)\, d\sigma(\eta)$$

and proceed to estimate F_j^* for each j. From the formula for $Q_r(\zeta, \eta)$ it is not hard to see that

$$\int_S |Q_r(\zeta, \eta)|\, d\sigma(\eta) \leq C$$

for a constant C independent of r; we have used the fact that $Q_r(\zeta, \eta)$ vanishes identically on the Koranyi ball centered at ζ of radius $1-r$ as well as the estimates found in **[R]** Proposition 1.4.10. It follows therefore that for each j

$$\|F_j\|_\infty \leq C\|A_j\|_\infty$$

and therefore

$$\|F_j^*\|_\infty \leq C(2^j)^{-n-m+n/p}(2^j\delta)^{-n/p}. \tag{12}$$

Recall that the admissible maximal region depends on the parameter α which controls its aperture. Set $M = 1000\alpha$. We will use inequality (12) above to estimate F_j^* on the set $\{\xi \in S : |1-\xi_1| \leq M2^j\delta\}$.

Assume then that $|1-\xi_1| > M2^j\delta$, and let $r\zeta \in \Gamma_\alpha(\xi)$. From properties (i) and (iii) it follows that $F_j(r\zeta)$ vanishes unless $1-r < 32\cdot 2^j\delta$ so we may as well assume that $1-r < 32\cdot 2^j\delta$. Let

$$U_{j,r}(\langle \eta, \tau\rangle) = P_{0,-1}(r\eta, \tau)\Phi_j(r\eta, \tau),$$

where, for each fixed r, $U_{j,r}$ is a function of one complex variable; notice that the definition of $\Phi_j(r\eta, \tau)$ makes this possible. Then by Corollary 2

$$\begin{aligned} F_j(r\zeta) &= (a * U_{j,r}) * Q_r(\zeta) \\ &= (a * Q_r) * U_{j,r}(\zeta). \end{aligned}$$

Thus

$$F_j(r\zeta) = \int_S a * Q_r(\eta) U_{j,r}(\langle \zeta, \eta \rangle)\, d\sigma(\eta).$$

Notice that, since $U_{j,r}(\langle \zeta, \eta \rangle)$ vanishes if $|1 - r\langle \zeta, \eta \rangle| > 16 \cdot 2^j \delta$, and since $r\zeta \in \Gamma_\alpha(\xi)$ with $1 - r < 32 \cdot 2^j \delta$, it follows from the triangle inequality of [**R**], Proposition 5.1.2 that $U_{j,r}(\langle \zeta, \eta \rangle) = 0$ unless $\eta \in B(\xi, 128\alpha 2^j \delta)$. For each such η use the cancellation properties of a to write

$$\begin{aligned} a * Q_r(\eta) &= \int_S a(\tau) Q_r(\langle \eta, \tau \rangle)\, d\sigma(\tau) \\ &= \int_S a(\tau)[Q_r(\langle \eta, \tau \rangle) - T_m(\tau; e_1)]\, d\sigma(\tau), \end{aligned}$$

where for each fixed η, $T_m(\tau, e_1)$ is the non-isotropic Taylor polynomial for $Q_r(\langle \eta, \tau \rangle)$ expanded about e_1 of degree m. From the formula $Q_r(\langle \eta, \tau \rangle)$ and the facts that $\tau \in B(e_1; \delta)$, $\eta \in B(\xi; 128\alpha 2^j \delta)$ and $|1 - \xi_1| > M2^j\delta$ it can be seen that

$$|Q_r(\langle \eta, \tau \rangle) - T_m(\tau; e_1)| \le C \frac{\delta^m (1-r)^{1-\beta}}{|1 - \xi_1|^{n+1-\beta+m}}.$$

Therefore with η as above

$$|a * Q_r(\eta)| \le C \frac{\delta^{m+n-n/p} (2^j \delta)^{1-\beta}}{|1 - \xi_1|^{n+1-\beta+m}}.$$

From this it follows that

$$|F_j(r\zeta)| \le C \frac{\delta^{m+n-n/p}}{|1 - \xi_1|^{n+m}},$$

and therefore, if $|1 - \xi_1| > M2^j\delta$, then

$$F_j^*(\xi) \le C(2^j)^{-n-m+n/p} \frac{(2^j\delta)^{n+m-n/p}}{|1 - \xi_1|^{n+m}}. \tag{13}$$

We now specify that $m > n/p - n$. Then the estimates in (12) and (13) show that

$$\int_S (F_j^*)^p\, d\sigma = \int_{B(e_1; M2^j\delta)} (F_j^*)^p\, d\sigma + \int_{S - B(e_1; M2^j\delta)} (F_j^*)^p\, d\sigma,$$

where the first integral on the right-hand side is dominated by

$$C(2^j\delta)^n \|F_j\|_\infty^p \le C(2^j\delta)^n (2^j)^{-np-mp+n} (2^j\delta)^{-n} = C(2^j)^{-np-mp+n},$$

and the second integral is less than

$$C(2^j)^{-np-mp+n} \int_{S - B(e_1; M2^j\delta)} \frac{(2^j\delta)^{mp+np-n}}{|1 - \xi_1|^{np+mp}}\, d\sigma \le C(2^j)^{-np-mp+n}.$$

Since $0 < p \leq 1$ we may use the triangle inequality to conclude that

$$\int_S (J_3^*)^p \, d\sigma \leq \sum_{j=0}^{N} \int_S (F_j^*)^p \, d\sigma$$
$$\leq C \sum_{j=0}^{\infty} (2^j)^{-np-mp+n}.$$

This completes the proof of the claim for the special case where $k = 1$.

The proof of the claim for the arbitrary case where $0 < \beta < k \leq n-1$ proceeds in an analogous fashion; we point out some of the minor differences. Since equation (6) will be replaced by

$$(\overline{R} + R)^k P[I_\beta a](z) = \left(\left(r\frac{\partial}{\partial r} \right)^k a * P_r \right) (I_\beta(\zeta),$$

instead of u_1 as given by equation (7), we will have to consider a sum of terms of the form

$$(1-|z|)^{k-\beta} \int_S \frac{1}{|1 - \langle \zeta , \eta \rangle|^{n-\beta}} \overline{R}^j R^{k-j} V(r\eta) \, d\sigma(\eta).$$

We split each such item into two pieces J_1 and J_2 as given by equations (8) and (9) with $1-\beta$ replaced by $k-\beta$ and RV replaced by $\overline{R}^j R^{k-j} V$. To handle J_1 we use the pointwise estimates of Lemma 3 in place of inequality (10). To handle J_2, in place of equation (11) we use Theorem 1 from **[ACa]** and Lemma 2 to get the fact that

$$\overline{R}^j R^{k-j} P[a] = \sum_{|l|+|m| \leq k} Q_{l,m}(L, \overline{L}) P_{l,m}[a],$$

where $Q_{l,m}$ is a polynomial in two variables of degree no greater than k. This lets us write J_2 as a sum of terms of the form

$$(1-r)^{k-\beta} (\overline{L}^j L^l a * G_r) * K_r(\zeta),$$

where

$$G_r(\lambda) = \frac{(1-r^2)^{n-t-s}}{(1-r\lambda)^{n-t}(1-r\overline{\lambda})^{n-s}},$$

where $j + l \leq n-1$ and t and s are non-positive integers such that $|t| + |s| \leq n-1$. The remainder of the argument proceeds without difficulty. This completes the proof of Theorem 2. □

References

[A] P. Ahern, *Exceptional sets for holomorphic Sobolev functions*, Michigan Math. J., **35** (1988), 29–41.

[ACa] P. Ahern and C. Cascante, *Exceptional sets for Poisson integrals of potentials on the unit sphere in* C^n, $p \leq 1$, Pacific. J. Math., **153**, no. 1 (1992), 1–13.

[ACo] P. Ahern and W. Cohn, *Exceptional sets for Hardy Sobolev functions,* $p > 1$, Indiana Univ. Math. J., **38**, no. 2 (1989), 417–453.

[GL] J. Garnett and R. Latter, *The atomic decomposition for Hardy spaces in several complex variables*, Duke Math. J., **45** (1978), 814–845.

[G] D. Geller, *Some results on* H^p *theory for the Heisenberg group*, Duke Math. J., **47**, no. 2 (1980), 365–390.

[R.] W. Rudin, *Function Theory in the Unit Ball of* C^n, Springer, New York, 1980.

Received December 1, 1991.

Wayne State University
Detroit, MI 48202
E-mail address: cohn@math.wayne.edu

and

University of Wisconsin
Madison, WI 53706
E-mail address: ahern@math.wisc.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

FANO BUNDLES AND SPLITTING THEOREMS ON PROJECTIVE SPACES AND QUADRICS

VINCENZO ANCONA, THOMAS PETERNELL AND JAROSLAW A. WIŚNIEWSKI

The aim of this paper is to describe the structure of Fano bundles in dimension ≥ 4.

Introduction. In this paper rank 2 vector bundles E on projective spaces $\mathbb{P}_n$ and quadrics Q_n are investigated which enjoy the additional property that their projectized bundles $\mathbb{P}(E)$ are Fano manifolds, i.e. have negative canonical bundles. Such bundles are shortly called Fano bundles. Up to dimension 3 Fano bundles are completely classified by [SW], [SW′], [SW″], [SSW]. The aim of this paper is to describe the structure of Fano bundles in dimension ≥ 4. Namely we prove the following

MAIN THEOREM. *Let E be a rank 2 Fano bundle on $\mathbb{P}_n$ or Q_n, $n \geq 4$. Then up to some explicit exceptions on Q_4 and Q_5 (see ex. (2.1), (2.2), (2.3)), E splits into a direct sum of line bundles.*

A rank 2 bundle E on $\mathbb{P}_n$ is Fano if and only if the "$\mathbb{Q}$-vector bundle" $E \otimes (\det E^*)/2 \otimes \mathscr{O}(\frac{n+1}{2})$ is ample, i.e.

$$\mathscr{O}_{\mathbb{P}(E)}(2) \otimes \pi^* \left(\det E^* \otimes \mathscr{O}\left(\frac{n+1}{2}\right)\right) \quad \text{is ample}.$$

If we normalize E in the following sense: $E_0 = E \otimes (\det E^*)/2$, so that $c_1(E_0) = 0$; then E is Fano iff $E_0(\frac{n+1}{2})$ is ample. Similarly on quadrics. In other words, we show that bundles with $E_0(\frac{n+1}{2})$ ample must split (on $\mathbb{P}_n$, $n \geq 4$). In other words: ample bundles with $c_1(E) \leq n+1$ split.

We prove even more:

THEOREM (9.1). *Let F be an ample 2-bundle on $\mathbb{P}_n$. Then F splits if one of the following assumptions hold:*

(1) $n = 4$, $c_1(F) \leq 6$,
(2) $n = 5$, $c_1(F) \leq 8$,

(3) $n = 6$ *or* 7, $c_1(F) \leq \frac{4n+2}{3}$,
(4) $n \geq 8$, $c_1(F) \leq \frac{5n-1}{3}$.

For testing the well-known conjecture of Hartshorne, that every 2-bundle on $\mathbb{P}_n$ ($n \geq 5$, or 6, or 7) should split, it would certainly be interesting to prove better bounds than in (9.1).

It is equally interesting to prove splitting theorems assuming only information of E on the lines in $\mathbb{P}_n$. The archaeopteryx of these theorems is the uniform splitting theorem. In the last section we prove among other things:

THEOREM (10.11). *Let E be a 2-bundle on $\mathbb{P}_n$. Assume for every line $L \subset \mathbb{P}_n$:*

$$E|L = \mathcal{O}(a_1(L)) \oplus \mathcal{O}(a_2(L))$$

with $|a_1(L) - a_2(L)| < \frac{n}{2} - 1$.

Then E splits.

1. Preliminaries. In this section we fix notations, give basic definitions and some elementary propositions which will be frequently used in the later sections.

(1.1) We will consider vector bundles only on the projective space $\mathbb{P}_n$ and on the n-dimensional quadric Q_n. If E is a vector bundle, we let $\mathbb{P}(E)$ be its associated projective bundle—taking hyperplanes in the fibers of E. We always let

$$\xi = c_1(\mathcal{O}_{\mathbb{P}(E)}(1)) \quad \text{and} \quad \eta = \pi^*(c_1(\mathcal{O}_X(1))),$$

where $\pi: \mathbb{P}(E) \to X$ is the projection and $X = \mathbb{P}_n$ or Q_n. If E is a 2-bundle on X, we denote by $c_i(E)$ its Chern classes, $i = 1, 2$ and consider them as numbers. Since we work only in dimension at least 4, we have

$$H^2(X, \mathbb{Z}) \simeq \mathbb{Z}$$

with the possible exception of Q_4; in this case

$$H^4(Q_4, \mathbb{Z}) \simeq \mathbb{Z}^2,$$

and we fix generators H_1, H_2 and identify $c_2(E) = aH_1 + bH_2$ with the pair (a, b).

DEFINITION 1.2. Let E be a vector bundle on a projective manifold X.

(1) E is said to be a Fano bundle if $\mathbb{P}(E)$ is a Fano manifold, i.e. $-K_{\mathbb{P}(E)}$ is ample.

(2) E is said to be nef ("numerically effective") if $\mathscr{O}_{\mathbb{P}(E)}(1)$ is a nef line bundle, i.e.:

$$c_1(\mathscr{O}_{\mathbb{P}(E)}(1)) \cdot C \geq 0$$

for all curves $C \subset \mathbb{P}(E)$.

REMARK 1.3. Let $r = \operatorname{rk} E$. Since

$$-K_{\mathbb{P}(E)} = \mathscr{O}_{\mathbb{P}(E)}(r) \otimes \pi^*(-K_X \otimes \det E^*),$$

E is Fano iff the "$\mathbb{Q}$-vector bundle"

$$E \otimes \frac{\det E^*}{r} \otimes \frac{-K_X}{r}$$

is ample, i.e. the $\mathbb{Q}$-Cartier divisor

$$O_{\mathbb{P}\left(E \otimes \frac{\det E^*}{r} \otimes \frac{-K_X}{r}\right)}(1)$$

is ample. We will often abbreviate $E \otimes (\det E^*)/r$ by E_0; we have $c_1(E_0) = 0$.

PROPOSITION 1.4. (A) *Let F be a nef 2-bundle on an n-dimensional projective manifold X with $b_2(X) = 1$, $b_4(X) = 1$ where the square of a generator of $H^2(X, \mathbb{Z})$ generates $H^4(X, \mathbb{Z})$. Let $c_i = c_i(F)$ (as numbers). Then:*

(1) $c_2 \geq 0$,
(2) $c_1^2 \geq 2c_2$ *if* $n \geq 3$,
(3) $c_1^2 \geq 3c_2$ *if* $n \geq 5$,
(4) $c_1^2 \geq (2+\sqrt{2})c_2$ *if* $n \geq 7$,
(5) $c_1^2 \geq (\frac{5}{2}+\frac{\sqrt{5}}{2})c_2$ *if* $n \geq 9$,
(6) $c_1^2 \geq (2+\sqrt{3})c_2$ *if* $n \geq 11$.

If F is ample, all inequalities are strict.

(B) *Let F be a nef 2-bundle on Q_4. Write $c_1 = c_1(F)$, $c_2 = c_2(F) = (a, b)$. Then:*

(1) $a \geq 0$, $b \geq 0$,
(2) $c_1^2 \geq a$, $c_1^2 \geq b$,
(3) $a^2 + b^2 - 3c_1^2(a+b) + 2c_1^4 \geq 0$.

Again the inequalities are strict for F ample.

Proof. (1) in (A) or (B) is well known. The other inequalities follow from positivity of the Segre classes for ample or nef bundles (**[FL]**, [**Fu**]) and the following computations for the Segre classes $s_i = s_i(F)$:

$$\begin{aligned}
s_1 &= c_1\,,\\
s_2 &= c_1^2 - c_2\,,\\
s_5 &= c_1(c_1^2 - c_2)(c_1^2 - 3c_2)\,,\\
s_7 &= c_1(c_1^2 - 2c_2)(c_1^2 - (2 - \sqrt{2}c_2))(c_1^2 - (2 + \sqrt{2})c_2)\,,\\
s_9 &= c_1\left(c_1^2 - \left(\frac{3}{2} + \frac{\sqrt{5}}{2}\right)c_2\right)\left(c_1^2 - \left(\frac{3}{2} - \frac{\sqrt{5}}{2}\right)c_2\right)\\
&\quad\times\left(c_1^2 - \left(\frac{5}{2} + \frac{\sqrt{5}}{2}\right)c_2\right)\left(c_1^2 - \left(\frac{5}{2} - \frac{\sqrt{5}}{2}\right)c_2\right)\,,\\
s_{11} &= c_1(c_1^2 - c_2)(c_1^2 - 2c_2)(c_1^2 - 3c_2)(c_1^2 - (2 + \sqrt{3}c_2))\\
&\quad\times(c_1^2 - (2 - \sqrt{3})c_2)\,.
\end{aligned}$$

(3) is just the semi-positivity of s_4.

Later we will also use s_6 on Q_6:

$$s_6 = 2c_1^6 - 10c_1^4c_2 + 6c_1^2c_2^2 - c_2^3\,.$$

An important tool will be le Potier's vanishing theorem [SS]:

THEOREM 1.5. *Let X be a projective manifold, E an ample vector bundle of rank r. Then:*

$$H^i(X\,, E \otimes K_X) = 0 \quad \textit{for } i \geq r\,.$$

We will also use

PROPOSITION 1.6. *Let F be a 2-bundle on $X = \mathbb{P}_n$ or on $X = Q_n$ with $n \geq 5$.*

Assume

$$H^0(X\,, F) \neq 0\,, \qquad H^0(X\,, F(-1)) = 0\,,$$

and that $c_2(F) \leq 1$. Then F splits.

The same holds for $X = Q_4$, provided $a < 0$ or $b < 0$ or $a = b = 0$ where $c_2(F) = (a\,, b)$.

Proof. Take $s \in H^0(F)$, $s \neq 0$, and let $Z = \{s = 0\}$. If $Z = \varnothing$, clearly F splits. If $Z \neq \varnothing$, then Z is of pure codimension 2, and

$$\deg Z = c_2(F)\,.$$

By our assumption on $c_2(F)$, we obtain a contradiction.

(1.7) A rank 2-bundle E on $\mathbb{P}_n$ or Q_n $(n \geq 3)$ is called (semi-) stable if for every line bundle $\mathcal{L} \subset E$:

$$c_1(\mathcal{L}) < \frac{c_1(E)}{2} \left(c_1(\mathcal{L}) \leq \frac{c_1(E)}{2} \right).$$

If E is stable on $\mathbb{P}_n$, it is well known that

$$(*) \qquad c_1^2(E) < 4c_2(E)$$

([**Ba**]). This is also true for quadrics Q_n; observe that for $n \neq 4$, $(*)$ is just an inequality of numbers, for $n = 4$ $(*)$ means

$$\int_{Q_4} c_1^2(E) \wedge \omega^2 < 4 \int_{Q_4} c_1^2(E) \wedge \omega^2$$

for every Kähler form ω on Q_4.

In order to see $(*)$ for quadrics, one can proceed as follows. If E is semistable then E carries an "approximate" Hermite-Einstein connection and hence

$$c_1^2(E) \leq 4c_2(E);$$

see [**Ko**].

Now assume $c_1^2 = 4c_2$. Since we may also assume $c_1(E) = 0$, we have $c_2(E) = 0$. But it is obvious that such an E cannot be stable. Thus a stable bundle satisfies $c_1^2 < 4c_2$.

(1.8) Some further notations: $h^i(X, \mathcal{F})$ will always be the dimension of $H^i(X, \mathcal{F})$; K_X will denote canonical line bundle of the complex manifold X; $[x]$ denotes the integral part of x.

2. Statement of the main result. Before stating our main result we shortly review some facts on special rank 2 vector bundles on quadrics.

EXAMPLE 2.1. We denote by S' and S'' the two "spinor bundles" on the 4-dimensional quadric Q_4. These are bundles of rank 2 with Chern classes $c_1(S') = c_1(S'') = -1$ and $c_2(S') = (1, 0)$, $c_2(S'') = (0, 1)$. Since $S'(1)$ and $S''(1)$ are globally generated, they are Fano bundles, i.e. $\mathbb{P}(S')$ and $\mathbb{P}(S'')$ are Fano manifolds (see [**Ot1**]). We will need in the sequel the following fact due to Ottaviani ([**Ot1**, Remark 3.4]): Every stable 2-bundle on Q_4 with Chern classes $c_1 = -1$ and $c_2 = (1, 0)$ (resp. $(0, 1)$) is isomorphic to S' (resp. S'').

EXAMPLE 2.2. Applying the Serre correspondence (see e.g. [**OSS**]) to the union of two disjoint planes in Q_4 we can construct a family of stable rank 2-bundles F with

$$c_1(F) = -1, \quad c_2(F) = (1, 1).$$

Their moduli space can be identified with $\mathbb{P}_7 \backslash Q_6$ (**[Ot2**, Remark 3.4]). By **[OT2]**, $F(2)$ is generated by global sections and thus F is a Fano bundle. Moreover: every stable 2-bundle on Q_4 with Chern classes $c_1 = -1$ and $c_2 = (1, 1)$ is isomorphic to some F described above.

EXAMPLE 2.3. On Q_5 there is a family of stable 2-bundles C with $c_1(C) = -1$, $c_2(C) = 1$. These were introduced in **[Ot2]**, where they are called Cayley bundles. Again $C(2)$ is globally generated; hence Cayley bundles are Fano. Moreover we have by **[Ot2**, main theorem and Theorem 3.2]: Every stable rank 2-bundle on Q_5 with Chern classes $c_1 = -1$, $c_2 = 1$ is isomorphic to a Cayley bundle. No Cayley bundle extends to Q_6.

We are now able to state the main result of this paper.

MAIN THEOREM 2.4. (1) *Let E be a Fano bundle of rank 2 on $\mathbb{P}_n$, $n \geq 4$. Then E splits as a direct sum of two line bundles.*

(2) *Let E be a Fano bundle of rank 2 on Q_n, $n \geq 4$. Then either E splits or:*

(a) *$n = 4$ and E is—up to a twist—a spinor bundle or one of the bundles described in* (2.2);

(b) *$n = 5$ and E is—up to a twist—a Cayley bundle (Example* (2.3)).

Fano 2-bundles on $\mathbb{P}_n$ or Q_n with $n \leq 3$ are classified in **[SW]** and **[SSW]**. Let E be a 2-bundle on $X = \mathbb{P}_n$ or Q_n. Since E is Fano if and only if $E \otimes (\det E^*)/2 \otimes -K_X/2$ is ample, we can restate Theorem 2.4 as follows.

COROLLARY 2.5. (1) *Let E be a normalized 2-bundle on $\mathbb{P}_n$, $n \geq 4$. If $c_1(E) = 0$ assume that $E(\frac{n+1}{2})$ is ample. If $c_1(E) = -1$, assume that $E(\frac{n+2}{2})$ is ample. Then E splits.*

(2) *Let E be a normalized 2-bundle on Q_n, $n \geq 4$. If $c_1(E) = 0$, assume that $E(\frac{n}{2})$ is ample. If $c_1(E) = -1$, assume that $E(\frac{n+1}{2})$ is ample. Then either E splits or E is as in* 2.4 (2)(a), (b).

The rest of this section is devoted to the proof of the following important technical result.

PROPOSITION 2.6. *Let E be a normalized Fano bundle of rank 2 on $\mathbb{P}_n$, $n \geq 4$. Then:*

(1) *If $c_1(E) = -1$ and n is odd, then $E([\frac{n}{2}] + 3)$ is generated by global sections and $E([\frac{n}{2}] + 2)$ is ample.*

(2) *In the other cases,* $E([\frac{n}{2}]+2)$ *is generated by global sections and* $E([\frac{n}{2}]+1)$ *is ample.*

In particular, in all cases $E(n)$ *is generated by global sections and ample.*

Proof. The ampleness statements are just translations of (1.3).

(1) The le Potier vanishing Theorem (1.5) gives

$$H^i(\mathbb{P}_n, E(t)) = 0$$

for $i \geq 2$ and $t \geq k+2-(n+1) = -k$ with $k = [\frac{n}{2}]$.

In particular:

$$H^i(\mathbb{P}_n, E(k+3-i)) = 0 \quad \text{for } i \geq 2.$$

Now we claim that this holds also for $i = 1$.

Consider on $\mathbb{P}(E)$ the divisor

$$D = 3\xi + (3k+5)\eta.$$

D is clearly ample; hence by Kodaira vanishing

$$H^1(\mathbb{P}(E), D + K_{\mathbb{P}(E)}) = 0,$$

i.e. $0 = H^1(\mathbb{P}(E), \xi + (k+2)\eta) \cong H^1(\mathbb{P}_n, E(k+2))$, whence our claim.

Now $E(k+3)$ is globally generated by the Castelnuovo-Mumford lemma.

(2) We treat shortly the case $n = 2k$ and $c_1(E) = 0$ leaving the remaining cases to the reader.

The le Potier vanishing theorem gives now

$$H^i(\mathbb{P}_n, E(k+2-i)) = 0, \qquad i \geq 2,$$

while the Kodaira vanishing theorem applied to the ample divisor $3\xi + (3k+2)\eta$ yields

$$H^1(\mathbb{P}_n, E(k+1)) = 0.$$

Thus $E(k+2)$ is globally generated.

The corresponding result for Q_n reads

PROPOSITION 2.7. *Let* E *be a normalized Fano bundle of rank* 2 *on* Q_n, $n \geq 4$. *Then:*

(1) *If* $c_1(E) = 0$ *and* n *is even,* $E(\frac{n}{2})$ *is ample and* $E(\frac{n}{2}+1)$ *is globally generated.*

(2) *In the other cases,* $E([\frac{n}{2}]+1)$ *is ample and* $E([\frac{n}{2}]+2)$ *is globally generated. In particular,* $E(n-1)$ *is ample and generated in all cases.*

The proof of (2.7) is just an adaptation of (2.6) and will be omitted.

The proof of Main Theorem 2.4 will be given in the subsequent sections; several cases have to be treated separately.

3. The case $\mathbb{P}_n$, $c_1^2(E) \geq 4c_2(E)$. In this section we shall prove

PROPOSITION 3.1. *Let* E *be a Fano 2-bundle on* $\mathbb{P}_n$, $n \geq 4$. *Assume* $c_1^2(E) \geq 4c_2(E)$. *Then* E *splits.*

The proof rests on the following result due to Holme and Schneider [**HS**, Theorem 4.2].

PROPOSITION 3.2. *Let* F *be a 2-bundle on* $\mathbb{P}_n$ *admitting a section whose zero locus is of pure codimension* 2. *If* F *is not stable and if moreover*

$$c_2(F) < (n-1)(c_1(F) - n + 2), \tag{3.2.1}$$

then F *splits.*

COROLLARY 3.3. *Let* F *be a globally generated 2-bundle on* $\mathbb{P}_n$. *If* F *is not stable and if* (3.2.1) *holds, then* F *splits.*

Proof. Let $s \in H^0(F)$ be a general section. Then $Z = \{s = 0\}$ is either empty (hence F splits) or Z is smooth of codimension 2. In this second case now apply (3.2).

Proof of (3.1). We may assume E to be normalized. E is unstable by [**Ba**], because of the inequality $c_1^2(E) \geq 4c_2(E)$ (which is invariant under twists). Put $F = E(n)$. Then by (2.6) F is globally generated. Since $c_2(E) \leq 0$ and $c_1(F) = c_1(E) + 2n$, we have

$$c_2(F) = c_2(E) + c_1(E)n + n^2 \leq c_1(E)n + n^2;$$

hence (3.2.1) holds as is easily verified. Thus F—as well as E—splits by (3.3).

4. The case Q_n, $n \geq 5$, **and** $c_1^2(E) \geq 4c_2(E)$. We now treat the analogous case to §3 for quadrics Q_n, $n \geq 5$. The case Q_4 will be done later.

PROPOSITION 4.1. *Let E be a Fano 2-bundle on Q_n, $n \geq 5$. Assume $c_1^2(E) \geq 4c_2(E)$. Then E splits. In order to prove* (4.1) *we need the following analogy to* (3.2):

PROPOSITION 4.2. *Let F be a 2-bundle on Q_n, $n \geq 5$, admitting a section whose zero locus is of pure codimension 2. If F is unstable and if moreover*

(4.2.1) $$c_2(F) \leq (n-2)(c_1(F) - n + 2) + n - 3,$$

then F splits.

Postponing the proof of (4.2) for a moment we have as in §3 the immediate

COROLLARY 4.3. *Let F be a globally generated 2-bundle on Q_n, $n \geq 5$. If F is unstable and if* (4.2.1) *holds, then F splits.*

Proof of 4.1. Let E be normalized. Since $c_1^2(E) \geq 4c_2(E)$, E is unstable (1.7). By (2.7), $F = E(n-1)$ is generated by global sections. Now

$$c_2(F) = c_2(E) - c_1(E)(n-1) + (n-1)^2 \leq c_1(E)(n-1) + (n-1)^2,$$

so (4.2.1) holds. Hence F (and E) splits by (4.3).

Proof of 4.2. The proof of (4.2) follows the same lines as that one of (3.2), so we give only a sketch, following **[Ra]** and **[HS]**. We may assume that our section vanishes in codimension 2, so we have a sequence

$$0 \to \mathcal{O}_{Q_n} \to F \to J_X(c_1) \to 0,$$

where $c_1 = c_1(F)$ and $X = \{s = 0\}$ is a locally complete intersection of codimension 2 in Q_n and of degree $d = c_2 = c_2(F)$. For $t \in \mathbb{Z}$ let

$$e(t) = c_2 - c_1 t + t^2 = c_2(F(-t)).$$

For a fixed point $p \in Q_n$ let S_p be the set of lines $l \subset Q_n$ with $p \in l$, and let

$$\Sigma_k = \Sigma_{k,p} = \{l \in S_p \,|\, \text{length}(l \cap X) \geq k\}$$

be the set of k-secant lines through p contained in Q_n.

Then we have (compare **[Ra]**).

PROPOSITION 4.3.1. *Assume* $k \le n-3$ *and* $e(0)\cdot e(1)\cdot\dots\cdot e(k) \neq 0$. *Then* $\dim \Sigma_{k+1} \ge n-k-2$. *In particular,* $\Sigma_{k+1} \neq \varnothing$.

Proof of 4.3.1. Since $\Sigma_0 = S_p \simeq Q_{n-2}$ it suffices by induction on k to show the following. If $C \subset \Sigma_k$ is an irreducible curve with $C \cap \Sigma_{k+1} = \varnothing$ and with $\min\{\text{length}(l \cap X) | l \in C\} = k$, then $e(k) = 0$. But this is proved by easily adapting the proof of the proposition in **[Ra]** to our situation.

Arguing as in **[Ra]** we obtain

LEMMA 4.3.2. *If* $c_1(F) \ge c_2(F)/(n-3) + n - 3$ *or if* $c_2(F) \le n-3$, *then* F *splits.*

Finally, the proof of Theorem (4.2) in **[HS]** can be copied almost word for word to give a proof of (4.2) (note that the inequality (4.2.1) is equivalent to $e(n-2) \le n-3$).

5. The case $\mathbb{P}_n$, $n \ge 6$, **and** $c_1^2(E) < 4c_2(E)$.

PROPOSITION 5.1. *There is no Fano 2-bundle* E *on* $\mathbb{P}_n$, $n \ge 6$, *with* $c_1^2(E) < 4c_2(E)$.

The proof of (5.1) will be based on the following result of **[HS]** (Corollary 3.4 and Proposition 6.1).

PROPOSITION 5.2. *Let* F *be a 2-bundle on* $\mathbb{P}_n$ *admitting a section whose zero locus is smooth and of pure codimension* 2. *Assume* $c_1^2(F) < 4c_2(F)$. *Then*:
(1) $c_1(F) \ge 2n+3$ *for* $n \ge 6$,
(2) $c_1(F) \ge 3n$ *for* $n \ge 8$.

Actually only (1) is used at this place but (2) will be needed later.

Proof of 5.1. Assume E to be a normalized Fano bundle of rank 2 on $\mathbb{P}_n$, $n \ge 6$, with $c_1^2(E) < 4c_2(E)$. By (2.6) $E(n)$ is globally generated. Now take a general section of $E(n)$ which vanishes along a smooth 2-codimension subvariety (of course the zero locus is non-empty). Hence $c_1(F) \ge 2n+3$ by 5.2(i); hence $c_1(E) \ge 3$, contradicting the fact that E is normalized.

6. The case: Q_n, $n \geq 12$, and $c_1^2(E) < 4c_2(E)$.

PROPOSITION 6.1. *There are no Fano 2-bundles on Q_n, $n \geq 12$, with $c_1^2(E) < 4c_2(E)$.*

In order to prove (6.1) we must have a substitute of (5.2) which is given by

PROPOSITION 6.2. *Let F be a 2-bundle on Q_n, $n \geq 12$, admitting a section whose zero locus is smooth and of pure codimension 2. Assume $c_1^2(F) < 4c_2(F)$. Then*

$$c_2(F) > \frac{71}{4}\left(\sin\frac{\pi}{n-1}\right)^{-2}.$$

First we show how (6.1) is derived from (6.2).

Proof of 6.1. Suppose again E to be normalized and let $F = E(\frac{n}{2}+1)$ if $c_1(E) = 0$ and n even, $F = E([\frac{n}{2}] + 2)$ otherwise. In any case $c_1(F) \leq n+3$, E being normalized. By (1.4), we have $c_1^2(F) > 3c_2(F)$; thus

$$c_2(F) \leq \frac{1}{3}c_1(F)^2 \leq \frac{1}{3}(n+3)^2,$$

F being globally generated (2.7), (6.2) applies to F. Hence (6.2.1) leads to a contradiction, since for $n \geq 12$ we have an inequality

$$\frac{1}{3}(n+3)^2 < \frac{71}{4}\left(\sin\left(\frac{\pi}{n-1}\right)\right)^{-2},$$

for $n \geq 12$.

Proof of 6.2. We mimic step by step the proof of the corresponding Theorem 2.2 of **[Sch]** on $\mathbb{P}_n$. Note that the Segre class $s_k(E)$ can be written as

$$s_k(E) = \tilde{s}_k(E)h^k$$

with $\tilde{s}_k(E) \in \mathbb{Z}$ and h the class of a hyperplane section of Q_n. According to the fact that the normal bundle of a submanifold of Q_n is always globally generated, we find as in **[Sch**, Corollary 1.2] that

$$\tilde{s}_k(E) \geq 0 \quad \text{for } k \leq n-2.$$

Now write

$$c_1(E) = \delta + \overline{\delta}, \quad c_2(E) = |\delta|^2$$

with $\delta = re^{i\varphi}$, $r \geq 0$, $-\pi \leq \varphi < \pi$.

Then $\tilde{s}_k(E) = \sum_{\nu=0}^{k} \delta^{k-\nu}\overline{\delta}^{-\nu}$.

Repeating the proof of Proposition 2.1 of **[Sch]** we get

LEMMA 6.2.1.
$$|\varphi| < \frac{\pi}{n-1}.$$

Now put $F(Q_n) = \min\{m \in \mathbb{N} | m = 4c_2(G) - c_1^2(G)$, with G a topological 2-bundle on $Q_n\}$, analogously $F(\mathbb{P}_n)$.

As in **[Sch]** we obtain from (6.2.2)
$$c_2(F) \geq \frac{1}{4} F(Q_n) \sin^2\left(\frac{\pi}{n-1}\right).$$

Since $n \geq 12$, we find a linearly embedded $\mathbb{P}_6$ in Q_n; hence
$$F(Q_n) \geq F(\mathbb{P}_6).$$

By **[Sch]**: $F(\mathbb{P}_6) \geq 71$, hence
$$F(Q_n) \geq 71,$$

finishing the proof of (6.2).

7. The case $\mathbb{P}_n$, $n = 4, 5$, and $c_1^2(E) < 4c_2(E)$. This is the last case to finish the main theorem for projective spaces.

PROPOSITION 7.1. *There are no Fano 2-bundles on $\mathbb{P}_n$, $n = 4, 5$ with $c_1^2(E) < 4c_2(E)$.*

Proof. Assume E is such a Fano 2-bundle. We may assume E to be normalized. Let $c_i = c_i(E)$ and introduce the $\mathbb{Q}$-vector bundle
$$E_0 = E\left(-\frac{c_1}{2}\right).$$

The fact that E is Fano can be expressed as $E_0(\frac{5}{2})$ (if $n = 4$) resp. $E_0(3)$ (if $n = 5$) to be ample.

As usual let ξ be the class of $\mathcal{O}_{\mathbb{P}(E)}(1)$, η the class of the pull-back of the hyperplane divisor.

(1) $n = 4$. Applying (1.4) to $E_0(\frac{5}{2})$ gives (by $c_1^2 > 2c_2$)
$$c_2(E_0) < \frac{25}{4};$$

hence $c_2 = c_2(E) \leq 6$ (and of course we also have $c_2 > 0$). Moreover
$$0 < s_4\left(E_0\left(\frac{5}{2}\right)\right) = (c_1^4 - 3c_1^2c_2 + c_2^2)(E_0).$$

Both inequalities easily imply

$$c_2(E) \leq 3.$$

By the Schwarzenberger conditions

$$c_2(c_2+1) \equiv 0 \ (12) \quad (\text{if } c_1 = 0),$$
$$c_2(c_2+2) \equiv 0 \ (12) \quad (\text{if } c_1 = -1),$$

we conclude that only the case $c_1(E) = 0$, $c_2(E) = 3$ is left.

By Riemann-Roch we obtain

$$\chi(\mathbb{P}_4, E(-2)) = 0.$$

The le Potier vanishing theorem applied to $E(3)$ yields

$$H^i(\mathbb{P}_4, E(-2)) = 0, \qquad i \geq 2.$$

Hence

$$H^0(\mathbb{P}_4, E(-2)) \neq 0,$$

and consequently $\zeta - 2\eta$ is an effective divisor on $\mathbb{P}(E)$. Thus

$$(\zeta - 2\eta)(2\zeta + 5\eta)^4 \geq 0.$$

On the other hand one computes easily

$$(\zeta - 2\eta)(2\zeta + 5\eta)^4 < 0,$$

a contradiction. Thus also the case $c_1 = 0$, $c_2 = 3$ is excluded.

(2) $n = 5$. This case is even simpler. The ampleness of $E_0(3)$ gives by $c_1^2 > 3c_2$ (1.4):

$$c_2(E) < 3 \quad \text{if } c_1 = 0,$$
$$c_2(E) \leq 3 \quad \text{if } c_1 = -1.$$

The claim follows again by using the Schwarzenberger conditions on $\mathbb{P}_4$.

8. The case: Q_n, $4 \leq n \leq 11$. We now treat the final case of low-dimensional quadrics in order to finish the proof of the main theorem.

PROPOSITION 8.1. *Let E be a Fano 2-bundle on Q_4. Then either E splits or is—up to a twist—a spinor bundle or one of the bundles of Example* (2.2).

Proof. As usual we assume E to be normalized and let $c_i = c_i(E)$, $E_0 = E(-\frac{c_1}{2})$, $\zeta = \mathcal{O}_{\mathbb{P}(E)}(1)$. Moreover let $c = c_2(E_0)$, $\zeta_0 = \zeta - c_1\eta/2$.

We write $c_2 = (a, b)$, $c = (a_0, b_0)$, so that $(a, b) = (a_0 + c_1^2/4, b_0 + c_1^2/4)$. We will use Riemann-Roch on Q_4:

$$\chi(E(-2)) = \frac{1}{12}(a(a+1) + b(b+1)) \quad \text{if } c_1(E) = 0;$$
$$\chi(E(-1)) = \frac{1}{12}(a(a-1) + b(b-1)) \quad \text{if } c_1(E) = -1.$$

The Fano condition says that $E_0(2)$ is ample. Hence by (1.4):

$$c_2(E_0(2)) > 0, \quad c_1^2(E_0(2)) > 2c_2(E_0(2)),$$

and consequently we obtain the bounds

$$-3 \le a \le 3, \quad -3 \le b \le 3 \qquad (\text{if } c_1 = 0),$$
$$-3 \le a \le 4, \quad -3 \le b \le 4 \qquad (\text{if } c_1 = -1).$$

LEMMA 8.1.1. *Suppose* $\chi(Q_4; E) > 0$ *and moreover*:
(1) *if* $c_1 = 0$: $a + 1 < 0$ *or* $b + 1 < 0$ *or* $a = b = -1$
(2) *if* $c_1 = -1$: $a + 2 < 0$ *or* $b + 2 < 0$ *or* $a = b = -2$.
Then E *splits.*

Proof. (1) Assume $c_1 = 0$. So $E(2)$ is ample. By le Potier vanishing:

$$H^i(Q_4, E(t-2)) = 0 \quad \text{for } i \ge 2, \ t \ge 0.$$

Hence

$$\chi(Q_4, E) = h^0(E) - h^1(E).$$

Since $\chi(E) > 0$, we conclude $h^0(E) \neq 0$.

By duality: $H^0(E(-2)) \simeq H^4(E(-2)) = 0$.

If now $H^0(E(-1)) \neq 0$, then by (1.6) E splits, since $c_2(E(-1)) = (a+1, b+1)$. If $H^0(E(-1)) = 0$, use (1.6) for E instead of $E(-1)$.

(2) The case $c_1 = -1$ is done in the same way starting with the ample bundle $E(3)$. We omit the details.

Since the condition $\chi(E) > 0$ is always satisfied if $a + b \le 0$ (by Riemann-Roch), the following cases are settled by Lemma 8.1.1:

$c_1 = 0$: $a < -1$, $b \le 2$ and $a \le 2$, $b < -1$ and $(a, b) = (-1, -1)$.

$c_1 = -1$: $a < -2$, $b \le 3$, and $a \le 3$, $b < -2$, and $(a, b) = (-2, -2)$.

(a) Suppose now $c_1 = 0$.

Riemann-Roch on Q_3 and Q_4 for the bundle E gives the following congruences:

$$a+b \equiv 0\ (2), \qquad -23(a+b)+a^2+b^1 \equiv 0\ (12).$$

Hence (a, b) must be one of the following: $(-1, 3)$, $(3, -1)$, $(0, 0)$, $(2, 2)$, $(3, 3)$.

(a_1) In the first two cases we have

$$\zeta(\zeta_0+2\eta)^4 = -6 < 0;$$

hence ζ cannot be effective, i.e.

$$H^0(E) = 0.$$

In particular E is stable. This is a contradiction because for $(a, b) = (-1, 3)$ or $(3, -1)$ the discriminant $c_1^2 - 4c_2(E) \geq 0$.

(a_2) Let now $(a, b) = (0, 0)$. Then $\chi(E) = 2$. By le Potier vanishing we get $h^0(E) \neq 0$. On the other hand $h^0(E(-1)) = 0$, since

$$(\zeta-\eta)(\zeta_0+4\eta)^4 = -32 < 0.$$

Now apply (1.6) to obtain the splitting of E.

(a_3) If $(a, b) = (2, 2)$ then $\chi(E(-2)) = 1$ by Riemann-Roch; hence le Potier vanishing gives $H^0(E(-2)) \neq 0$, contradicting

$$(\zeta-2\eta)(\zeta_0+2\eta)^4 = -120 < 0.$$

(a_4) If $(a, b) = (3, 3)$, we have

$$(\zeta_0+2\eta)^5 = a_0^2+b_0^2-40(a_0+b_0)+160 < 0$$

(observe $(a, b) = (a_0, b_0)$), which is in contradiction to the ampleness of $E(2)$.

(b) We consider now the case $c_1 = -1$. We have a congruence

$$-13(a+b)+a^2+b^2 \equiv 0\ (12).$$

Thus the only possible values for (a, b) are: $(4, -3)$, $(-3, 4)$, $(3, -2)$, $(-2, 3)$, $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$, $(0, 4)$, $(4, 0)$, $(1, 4)$, $(4, 1)$, $(3, 3)$, $(4, 4)$. The first two cases are settled by (8.1.1) since then $\chi(E) = 2$. In the next three ones, we have $\chi(E) > 0$, hence $h^0(E)$ by le Potier vanishing; moreover $h^0(E(-1)) = 0$ since

$$(\zeta-\eta)(\zeta_0+2\eta)^4 < 0.$$

So E splits by (1.6).

In the case $(0, 0)$, $(1, 0)$ or $(0, 1)$, E must be stable because $H^0(E) = 0$ by

$$\zeta(\zeta_0 + 2\eta)^4 < 0.$$

So E is as in (2.1) or (2.2).

The case $(0, 4)$, $(4, 0)$ are ruled out as in a_1. The remaining cases finally contradict

$$(\zeta_0 + 2\eta)^5 > 0.$$

PROPOSITION 8.2. *Let E be a Fano 2-bundle on Q_n, $5 \leq n \leq 11$, with $c_1^2(E) < 4c_2(E)$. Then $n = 5$ and E is—up to a twist—a Cayley bundle* (2.3).

Proof. Again suppose E to be normalized. Let $E_0 = E(-c_1/2)$. E being Fano, $E_0(\frac{n}{2})$ is an ample $\mathbb{Q}$-vector bundle. By positivity of the Segre classes of $E_0(\frac{n}{2})$ we obtain

$$c_1^2\left(E_0\left(\frac{n}{2}\right)\right) > \alpha c_2\left(E_0\left(\frac{n}{2}\right)\right)$$

with

$$\alpha = \begin{cases} 3 & \text{for } n = 5, 6, \\ 2 + \sqrt{2} & n = 7, 8, \\ \frac{5}{2} + \frac{\sqrt{5}}{2} & n = 9, 10, \\ 2 + \sqrt{3} & n = 11 \end{cases}$$

(see **[FL]**, **[FU]**).

Hence: $n^2 > \alpha(c_2(E_0) + n^2/4)$.

Since $c_2(E_0) > 0$, we easily obtain

$$c_2(E) \leq 2,$$

with the exception $n = 6$, $c_1(E) = -1$.

Let us first consider this exceptional case. Then we have only $c_2(E) \leq 3$. Assume $c_2 = 3$. Then we compute

$$s_6\left(E\left(\frac{7}{2}\right)\right) = (2c_1^6 - 10c_1^4c_2 + 6c_1^2c_2^2 - c_2^3)\left(E\left(\frac{7}{2}\right)\right) < 0,$$

contradicting ampleness of $E(\frac{7}{2})$.

So we may assume $c_2(E) \leq 2$ in all cases. The case $c_2 = 2$ is ruled out as follows: if $c_1(E) = 0$ (resp. $c_1(E) = -1$) take a $Q_5 \subset Q_n$ (resp. $Q_4 \subset Q_n$) and Riemann-Roch gives $\chi(Q_5, E|Q_5) \notin \mathbb{Z}$ (resp. $\chi(Q_4, E|Q_4) \notin \mathbb{Z}$).

By observing $\chi(Q_4, E|Q_4) \notin \mathbb{Z}$, also the case $c_1(E) = 0$, $c_2(E) = 1$ is impossible. It remains to consider the case $c_1(E) = -1$, $c_2(E) = 1$.

If E is unstable, apply (4.2) to the bundle $F = E(n-1)$ which is globally generated by (2.7) (the condition (4.2.1) is immediately verified). So E splits.

If E is stable, the restriction $E|Q_5$ to a generic linear $Q_5 \subset Q_n$ is stable again with $c_1 = -1$, $c_2 = 1$. Hence by (2.3), $E|Q_5$ is a Cayley bundle.

Since no Cayley extends to Q_6 (Ottaviani, see 2.3), we must have $n = 5$. The proof of (8.2) is now complete.

Combining all results of §§3–8 gives a proof of the Main Theorem.

9. Generalizations. The Main Theorem for projective spaces can be improved considerably (we will not consider the case of quadrics here):

THEOREM 9.1. *Let F be an ample 2-bundle on $\mathbb{P}_n$. Then F splits under one of the following assumptions.*

(1) $n = 4$, $c_1(F) \leq 6$,

(2) $n = 5$, $c_1(F) \leq 8$,

(3) $n = 6$ *or* 7, $c_1(F) \leq \frac{4n+2}{3}$,

(4) $n \geq 8$, $c_1(F) \leq \frac{5n-1}{3}$.

REMARK. (9.1) can be reformulated as follows. Assume that F is a $\mathbb{Q}$-vector bundle with $c_1(F) = 0$. Then e.g. (1) says that in case $n = 4$, F splits if $F(3)$ is ample. We should also mention the Horrocks-Mumford bundle H in this context. It has $c_1(H) = -1$ and $c_2(H) = 4$; moreover $H(4)$ is generated by global sections. So the statement (1) or (9.1) is almost sharp, see also (9.2) below.

Part (1) of Theorem 9.1 will follow from the more general statement:

PROPOSITION 9.2. *Let E be a 2-bundle on $\mathbb{P}_4$. Let $E_0 = E \otimes (\det E^*/2)$. If $E_0(3)$ is nef, then E splits.*

For the proof of (9.2) we will need

LEMMA 9.3. *Let E be a normalized 2-bundle on $\mathbb{P}_n$ such that $E(m)$ is nef for some $m \in \mathbb{Q}$. Let $r \in \mathbb{Z}$ be the maximal number such that*

$$H^0(\mathbb{P}_n, E(-r)) \neq 0.$$

Then either E splits or

(a) $r \leq m - 2$ (*if* $c_1(E) = 0$) *or*

(b) $r \leq m - 3$ (*if* $c_1(E) = -1$).

Proof. We treat only the case $c_1(E)=0$, the other case being similar. Let $s\in H^0(E(-r))$, $s\neq 0$, and let $Z=\{s=0\}$. If $Z=\varnothing$, E splits. So assume $Z\neq\varnothing$. By our assumptions, Z is locally a complete intersection of codimension 2. If $\deg Z=1$, Z is a complete intersection and E splits. So let $\deg Z\geq 2$. Then take a 2-secant line L of Z with $L\not\subset Z$.

Then $E(-r)|L$ has a section with at least two zeros; hence

$$E(-r)|L=\mathscr{O}_L(2+k)\oplus\mathscr{O}_L(-2r-2-k)$$

for some $k\geq 0$. Hence

$$E(m)|L=\mathscr{O}_L(2+k+r+m)\oplus\mathscr{O}_L(m-r-2-k)$$

and by nefness of $E(m)|L$ we conclude.

Proof of 9.2. We may assume E to be normalized.

(a) First let $c_2(E)\leq 0$. So E is unstable. Let r be the biggest positive integer such that

$$H^0(\mathbb{P}_4, E(-r))\neq 0.$$

Assume that E does not split. Then we deduce from (9.3): $r\leq l$ in case $c_1(E)=0$; $r\leq\frac{1}{2}$ if $c_1(E)=-1$. In the second case $r\leq 0$; we must have $r=1$, and thus $E(-1)$ has a section whose zero locus Z is either empty or of codimension 2 with $\deg Z=c_2(E)+1\leq 1$. But then clearly E splits.

(b) Now we consider the case $c_2(E)>0$. Let $c=c_2(E_0)$. By nefness of $E_0(3)$ we obtain

$$0\leq c_1(E_0(3))^2-2c_2(E_0(3))=36-2(c+9);$$

hence $c\leq 9$. On the other hand, the highest Segre class $s_4(E_0(3))\geq 0$; hence $c^2-90c+405\geq 0$, which together with $c\geq 9$, proves $c\leq 45-\sqrt{1620}<5$.

Hence $c_2(E)\leq 4$ if $c_1(E)=0$ and $c_2(E)\leq 5$ if $c_1(E)=-1$. By the Schwarzenberger conditions we find:

$$c_1(E)=0,\quad c_2(E)=3\quad\text{or}\quad c_1(E)=-1,\quad c_2(E)=4.$$

In both cases a short computation shows

$$\zeta(\zeta_0+3\eta)^4<0;$$

hence $H^0(E)=0$ and E is thus stable.

By **[BE]** there is no stable 2-bundle on $\mathbb{P}_4$ with $c_1 = 0$, $c_2 = 3$; by **[DS]** the only stable 2-bundle on $\mathbb{P}_4$ with $c_1 = -1$, $c_2 = 4$ is the Horrocks-Mumford bundle for which it is easy to see that $E_0(3)$ is not nef (restrict to jumping lines). This completes the proof of (9.2).

REMARK. (1) In the proof of (9.2) one shows also the following stronger statement: let E be an unstable 2-bundle on $\mathbb{P}_4$, assume $E_0(3)$ to be nef on every line $L \subset \mathbb{P}_4$. Then E splits.

(2) It would be interesting to do the next step in (9.2): assume only $E_0(4)$ to be nef. This leads to some interesting problems. Let e.g. E be a (semi-stable) 2-bundle on $\mathbb{P}_4$ with $c_1 = -1$, $c_2 = 6$ and assume $E_0(4)$ even to be generated by global sections. Take a general section with smooth zero locus X. Then

$$K_X = \mathcal{O}_X(2),$$

i.e. X is a "half-canonical" surface in $\mathbb{P}_4$ with $\deg X = 18$. Half-canonical surfaces are investigated in **[DPPS]** and it is shown that they cannot exist (or are complete intersections) with possible exceptions in degree 18 and 22 (and some other restrictions). "Of course" one expects that half-canonical surfaces are complete intersections in these degrees, too.

Part (2) of (9.1) will be a consequence of

PROPOSITION 9.4. *Let E be a 2-bundle on $\mathbb{P}_5$ such that $E_0(4)$ is nef. Then E splits.*

Proof. As usual we suppose E normalized.

(a) Assume $c_2(E) \leq 0$; so E is not stable. Let r be the maximal positive integer such that

$$H^0(E(-r)) \neq 0.$$

By (9.3): $r \leq 2$ if $c_1(E) = 0$; $r \leq \frac{3}{2}$ if $c_1(E) = -1$. If $c_1(E) = -1$ we have $r = 1$, so by $c_2(E(-1)) = c_2(E) \leq 0$, E splits (1.6). This argument settles also $c_1(E) = 0$ and $r = 1$. Finally let $c_1(E) = 0$ and $r = 2$. Then (1.6) settles the case $c_2(E) \leq -3$. Take a linear $\mathbb{P}_4 \subset \mathbb{P}_5$ and use the Schwarzenberger condition for $E|\mathbb{P}_4$ to obtain $c_2(E) = 0$ or -1. But in both cases:

$$(\zeta - 2\eta)(\zeta + 4\eta)^5 < 0,$$

contradicting the nefness of $\zeta + 4\eta$.

(b) Assume now $c_2(E) > 0$. From $c_1^2(E_0(4)) > 3c_2(E_0(4))$ we deduce $c_2(E) \leq 5$. Now the Schwarzenberger condition for $E|\mathbb{P}_4$ implies:

$$c_1(E) = 0, \quad c_2(E) = 3, \quad \text{or} \quad c_1(E) = -1, \quad c_2(E) = 4.$$

In both cases:

$$\zeta(\zeta_0 + 4\eta)^5 < 0;$$

hence $H^0(E) = 0$, and consequently E is stable.

Now for a general linear $\mathbb{P}_4 \subset \mathbb{P}_5$, $E|\mathbb{P}_4$ is again stable with the same Chern classes, so $c_1(E) = 0$, $c_2(E) = 3$ is ruled out by **[BE]** and the other case by **[DS]**, since the Horrocks-Mumford bundle does not extend to $\mathbb{P}_5$.

Proof of 9.1, *parts* (3) *and* (4). Let $c_i = c_i(F)$. If $c_1 \leq n+1$, F is a Fano bundle and hence F splits by the Main Theorem. So assume now $c_1 \geq n+2$.

First let us show that

(1) $F(c_1 - n)$ is generated by global sections. In fact, $H^i(\mathbb{P}_n F(t)) = 0$ for $i \geq 2$, $t \geq -n-1$ by le Potier vanishing; moreover by Kodaira vanishing for the divisor ζ_F:

$$0 = H^1(\mathbb{P}(F), 3\zeta_F + K_{\mathbb{P}(F)}) \simeq H^1(\mathbb{P}_n, F(c_1 - n - 1)).$$

So (1) follows from the Castelnuovo-Mumford lemma. As a consequence we obtain

(2) $c_1^2(F) \geq 4c_2(F)$.

In fact, if $c_1^2(F) < 4c_2(F)$ we can apply—using (1)—Proposition 5.2 for $F(c_1 - n)$ conflicting our assumptions.

We suppose c_1 to be even, the odd case being treated similarly. Let $E = F(-c_1/2)$. Let r be the maximal integer such that

$$H^0(E(-r)) \neq 0.$$

Since E is unstable by (2), r must be positive.

Since $H^n(\mathbb{P}_n, F(-n-1)) = 0$ we have by duality $H^0(\mathbb{P}_n, E(-c_1/2)) = 0$; hence $r < c_1/2$.

Since $c_2(E) \leq 0$, we have moreover $c_2(E(-r)) \leq r^2 \leq c_1^2/4$, so our assumption yields

$$c_2(E(-r)) < (n-1)(n+5).$$

Since any section of $E(-r)$ vanishes nowhere or in codimension 2, E splits by [**HS**, 4.7].

10. Numerical splitting of rank 2-bundles on $\mathbb{P}_n$. In the previous sections we considered Fano bundles E on $\mathbb{P}_n$, i.e. $E_0(\frac{n+1}{2})$ is ample, E_0 denoting the normalization $E(c_1(E^*)/2)$. Here we want only to make an assumption on the behaviour of E on the lines and try to get some information.

Let E always denote a vector bundle of rank 2 on $\mathbb{P}_n$.

10.1. DEFINITION. (1) For a line $L \subset \mathbb{P}_n$ put

$$\delta_L(E) = \delta_L = a_2 - a_1,$$

if $E|L = \mathcal{O}(a_1) \oplus \mathcal{O}(a_2)$ with $a_1 \leq a_2$.

(2) For $x \in \mathbb{P}_n$ define

$$\delta_x^{\max} = \max\{\delta_L | L \text{ a line through } x\} \text{ and}$$
$$\delta_x^{\min} = \min\{\delta_L | L \text{ a line through } x\}.$$

10.2. DEFINITION. For $x \in \mathbb{P}_n$ let $\mathbb{P}_x$ be the variety of lines through x. Write $\delta_x^{\min} = \delta_0 < \delta_1 < \cdots < \delta_r = \delta_x^{\max}$, where δ_i are the "splitting types" realized by E on some line passing through x.

Define $V_{\delta_i} = \{L \in \mathbb{P}_x | \delta_L = \delta_i\}$.

10.3. REMARK. We have clearly:

(a) $\overline{V}_{\delta_0} = \mathbb{P}_x$,

(b) $\overline{V}_{\delta_i} = V_{\delta_i} \cup \bigcup_{j>i}(\overline{V}_{\delta_i} \cap V_{\delta_j})$.

10.4. DEFINITION. If $\delta_L > 0$, the ruled surface $\mathbb{P}(E|L)$ has a unique exceptional section C_L (i.e. $C_L^2 < 0$). We define a map (for fixed $x \in \mathbb{P}_n$)

$$\Phi_{\delta_i} : V_{\delta_i} \to \mathbb{P}(E_x) \simeq \mathbb{P}_1 \qquad (i > 0)$$

by setting

$$\Phi_{\delta_i}(L) = C_L \cap \mathbb{P}(E_x).$$

It is easy to check that Φ_{δ_i} is holomorphic.

The key to this section is

10.5. THEOREM. *Assume that for some δ_i the map ϕ_{δ_i} has a fiber containing a compact curve. Then E splits numerically:*

$$c_1(E) = a + b, \quad c_2(E) = ab, \quad \textit{where } E|L_{\delta_i} = \mathcal{O}(a) \oplus \mathcal{O}(b).$$

In other words E has the same Chern classes as a decomposable bundle.

REMARK. The assumption means that there is a "compact" family $(L_t)_{t\in T}$ of lines through x with T compact, such that $C_{L_t} \cap \mathbb{P}(E_x)$ does not depend on t.

Proof. After normalizing T we obtain a compact curve C and a geometrically ruled surface $p: S \to C$ with a map $\psi: S \to \mathbb{P}(E)$ such that $\psi(p^{-1}(c)) = C_{L_t}$ where c is a point in C over t. By our assumption the ruled surface contains a section, say C_0, such that $\psi(C_0) = \phi_{\delta_t}(t)$ for any $t \in T$.

Now consider the relative Euler sequence

$$0 \to \omega_{\mathbb{P}(E)/\mathbb{P}_n}(1) \to \pi^*(E) \to \mathscr{O}_{\mathbb{P}(E)}(1) \to 0$$

where $\mathbb{P}(E)$ is the projective bundle taking hyperplanes and $\omega_{\mathbb{P}(E)/\mathbb{P}_n}$ is the relative dualizing sheaf.

Since $\omega_{\mathbb{P}(E)/\mathbb{P}_n}(1) \simeq \mathscr{O}_{\mathbb{P}(E)}(-1) \otimes \pi^*(\mathscr{O}(a_1 + a_2))$ we obtain by tensoring with $\pi^*(\mathscr{O}(-a_1))$:

$$\begin{aligned} 0 \to \mathscr{O}_{\mathbb{P}(E)}(-1) \otimes \pi^*(\mathscr{O}(a_2)) \to (\pi^* E)(-a_1) \\ \to \mathscr{O}_{\mathbb{P}(E)}(1) \otimes \pi^*(\mathscr{O}(-a_1)) \to 0. \end{aligned}$$

Now we have

$$(*) \qquad \psi^*((\mathscr{O}_{\mathbb{P}(E)}(1)) \otimes \pi^*(\mathscr{O}(-a_1))) \simeq \mathscr{O}_s:$$

this has only to be checked on C_0 (obvious!) and on a fiber $p^{-1}(c)$. But for this it is sufficient to see

$$\mathscr{O}_{\mathbb{P}(E)}(1) \otimes \pi^*(\mathscr{O}(-a_1))|C_L \simeq \mathscr{O}$$

which is clear since

$$\mathscr{O}_{\mathbb{P}(E)}(1) \otimes \pi^*(\mathscr{O}(-a_1))|C_L = \mathscr{O}_{\mathbb{P}(\mathscr{O}_1(a_2-a_1)\otimes\mathscr{O}_L)}(1)|C_L$$

and since C_L is the exceptional section (see [**Ha**, Chap. V.2]).

Since $\pi \circ \psi$ is generically finite, $(*)$ implies

$$c_2(E(-a_1)) = c_2(\psi^*(\pi^*(E(-a_1)))) = 0.$$

Hence $c_2(E) = a_1 a_2$.

An obvious consequence of 6.5 is

10.6. COROLLARY. *Assume that E does not split numerically. Then*
(1) $\dim V_{\delta_x^{\max}} \leq 1$ *and*
(2) $\dim \overline{V}_{\delta_i} - \dim(\overline{V}_{\delta_i} \cap \bigcup_{j>i} V_{\delta_j}) \leq 2$, *for* $\delta_i < \delta_x^{\max}$.

10.7. THEOREM. *Assume that there is some $x \in \mathbb{P}_n$ such that*

$$\delta_x^{\max} - \delta_x^{\min} < n - 2 .$$

Then E splits numerically.

Proof. First observe that in general

$$\#\{V_{\delta_i}\} \leq \frac{1}{2}(\delta_x^{\max} - \delta_x^{\min}) + 1 < \frac{n}{2}$$

by our assumption.

On the other hand (10.6) implies: $2 \cdot \#\{V_{\delta_i}\} \geq n$, if E does not split numerically. Both inequalities being incompatible, E has to split numerically.

For $n = 3$ Theorem 10.7 says that every uniform (w.r.t. lines through x) 2-bundle E numerically splits. Of course it is well known that E in fact splits. But already for $n = 5$, the assumption of 10.7 is less restrictive than uniformity.

Another immediate consequence of 10.5 is

10.8. COROLLARY. *If there is some $x \in \mathbb{P}_n$ and some i such that V_{δ_i} contains a compact surface, then E splits numerically.*

10.9. COROLLARY. *Assume that E is a semi-stable 2-bundle on $\mathbb{P}_n$, $n \geq 4$, with $c_1(E) = 0$. Assume that there is some $x \in \mathbb{P}_n$ and some $a > 0$ such that for all $L \in \mathbb{P}_x$ either $E|L = \mathcal{O} \oplus \mathcal{O}$ or $E|L = \mathcal{O}(a) \oplus \mathcal{O}(-a)$.*

Then E splits numerically.

Proof. Of course we may assume that $E|L = \mathcal{O}(a) \oplus \mathcal{O}(-a)$ for some line. Since $c_1(E) = 0$, the jumping lines of E form a divisor D in $G(1, n)$ (=lines in $\mathbb{P}_n$, see e.g. **[OSS]**). Hence $D \cap \mathbb{P}_x$—which is the set of jumping lines through x—is a divisor in $\mathbb{P}_x$. Therefore we obtain a compact surface in $V_{\delta_1} = \overline{V}_{\delta_1} = \overline{V}_{2a}$, since $n \geq 4$.

In order to prove splitting criteria rather than merely criteria for numerically splitting we prove

10.10. PROPOSITION. *Let E be a 2-bundle on $\mathbb{P}_n$. Choose $x \in \mathbb{P}_n$ generic such that $\delta_x^{\max}$ is minimal. If* $\dim V_{\delta_x^{\max}} > \frac{n}{2}$, *then there is* $s \in H^0(E((-c_1 + \delta_x^{\max})/2))$ *with* $s(x) \neq 0$.

Proof. (a) We claim that $\phi := \phi_{\delta_x^{\max}}: V := V_{\delta_x^{\max}} \to \mathbb{P}(E_x)$ is constant. In fact, otherwise by our assumption ϕ would have fibers of dimension $> \frac{n}{2} - 1$. On the other hand we have $V \subset \mathbb{P}_x \simeq \mathbb{P}_{n-1}$; so algebraic sets in V of dimension $> \frac{n}{2} - 1$ must meet, contradiction.

(b) Let $D = \bigcup\{C_L | X \in \mathbb{P}_n,\ L \in \mathbb{P}_x \text{ and } \delta_L = \delta_x^{\max}\} \subset \mathbb{P}(E)$. If x is general, then $D \cap \mathbb{P}(E_x)$ consists of one point by (a). Hence there is an irreducible component $D_0 \subset D$ and some $d \in \mathbb{Z}$ such that

$$D_0 \in |\mathcal{O}_{\mathbb{P}(E)}(1) \otimes \pi^*(\mathcal{O}(d))|.$$

Taking a line L through our general x and observing

$$D_0 \cap \pi^{-1}(L) = C_L,$$

we conclude

$$d = \frac{-c(E) + \delta_x^{\max}}{2}.$$

10.11. THEOREM. *Let E be a 2-bundle on $\mathbb{P}_n$. Assume $\delta_L < \frac{n}{2} - 1$ for every line $L \subset \mathbb{P}_n$. Then E splits.*

Proof. This follows from (the proof of) Proposition 10.10 since our condition implies

$$\dim V_{\delta_x^{\max}} > \frac{n}{2}$$

for every x; hence the section constructed in the proof of 10.10 does not vanish at any point and consequently E splits.

Again 10.11 can be viewed as a generalization of the fact that uniform 2-bundles on $\mathbb{P}_n$, $n \geq 3$, split.

10.12. REMARK. Most of the above can be applied to manifolds containing "plenty of lines" if we only can control their cohomology. For example, if X is a Fano manifold of index $r > \frac{1}{2}\dim X + 1$ (recall that the index r is the largest integer dividing $-K_X$ in $\mathrm{Pic}(X)$ and that for $r > \frac{1}{2}\dim X + 1$, we have $\mathrm{Pic}(X) = \mathbb{Z}$ by **[Wi]**) then through every point of X there passes a line (i.e. a rational curve whose intersection with the ample generator H of Pic(X) is 1). For a 2-bundle E on X we can define (via normalization) the splitting type of E on any such line. Similarly we can define $\delta_x^{\max}$, $\delta_x^{\min}$. Then we obtain an equivalent of 10.7.

10.13. THEOREM. *Let X and E be as above, let moreover $b_4(X) = 1$. If $\delta_x^{\max} - \delta_x^{\min} < r - 3$ for some $x \in X$, then E splits numerically, i.e. $c_1(E) = (a+b)c_1(H)$, $c_2(E) = (ab)c_1(H)^2$ for some $a, b \in \mathbb{Z}$.*

REFERENCES

[Ba] W. Barth, *Some properties of stable rank 2 vector bundles on $\mathbb{P}_n$*, Math. Ann., **226** (1977), 125–150.

[BE] W. Barth and G. Elencwajg, *Concernant la Cohomologie des Fibrés Algébriques Stables sur $\mathbb{P}_n(\mathbb{C})$*. In: *Variétés Analytiques Compactes*, Colloque Nice 1977, Lecture Notes in Math., vol. 683, Springer-Verlag, Berlin and New York, 1977, pp.1–24.

[DS] W. Decker and F. O. Schreyer, *On the uniqueness of the Horrochs-Mumford bundle*, Math. Ann., **273** (1986), 415–443.

[DPPS] W. Decker, Th. Peternell, J. le Potier, and M. Schneider, *Half-canonical Surfaces in $\mathbb{P}_4$*, Lecture Notes in Math., vol. 1417, Springer-Verlag, Berlin and New York, 1990, pp. 91–110.

[Fr] K. Fritzsche, *Linear-uniforme Bündel auf Quadriken*, Annali Sci. Norm. Sup. Pisa, (Serie 4), **10** (1983), 313–339.

[FL] W. Fulton and R. Lazarsfeld, *Positive polynomials for ample vector bundles*, Ann. Math., **118** (1983), 35–60.

[Fu] W. Fulton, *Intersection Theory*. Erg. d. Math. 3. Folge, Band 2. Springer 1984.

[HS] A. Holme and M. Schneider, *A computer aided approach to codimension 2 subvarieties of $\mathbb{P}_n$, $n \geq 6$*. Crelles J., **357** (1985), 205–220.

[OSS] C. Okonek, M. Schneider and H. Spindler, *Vector bundles on complex projective spaces*, Progress in Math., vol. 3, Birkhäuser 1980.

[Ot1] G. Ottaviani, *Spinor bundles on quadrics*, Trans. Amer. Math. Soc., **307** (1988), 301–316.

[Ot2] ——, *On Cayley bundles on the five-dimensional quadric*, Boll. UMI7 (4-A), (1990), 87–100.

[Ra] Z. Ran, *On projective varieties of codimension 2*, Invent. Math., **73** (1983), 333–336.

[Sch] M. Schneider, *Submanifolds of projective space with semistable normal bundle*, In: *Several variables*, Proc. of the 1981 Hangzhon Conference 151–160, Boston 1984.

[SS] B. Shiffman and A. J. Sommese, *Vanishing Theorems*, Birkhäuser 1985.

[SSW] I. Sols, M. Szurek, and J. A. Wisniewski, *Rank 2 Fano bundles over a smooth quadric Q_3*, Pacific J. Math., **148** (1991), 153–159.

[SW] M. Szurek and J. A. Wiśniewski, *Fano bundles over $\mathbb{P}_3$ and Q_3*, Pacific J. Math., **141** (1990), 197–208.

[SW′] ——, *On Fano manifolds which are $\mathbb{P}_k$-bundles over $\mathbb{P}_2$*, Nagoya Math. J., **120** (1990), 89–101.

[SW″] ——, *Fano bundles of rank 2 on surfaces*, Compositio Math., **76** (1990), 295–305.
[Wi] J. Wiśniewski, *On a conjecture of Mukai*, Manuscripta Math., **68** (1990), 135–141.

Received February 27, 1992.

"U. DINI"
UNIV. DI FIRENZE
VIALE MORAGAGNI 67A
50139 FIRENZE, ITALY

UNIV. BAYREUTH
95440 BAYREUTH, GERMANY

AND

UNIWERSYTET WARSZAWSKI
UL. BANACHA 2
02-097 WARSZAWSKI, POLAND

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

A TRANSVERSE STRUCTURE FOR THE LIE-POISSON BRACKET ON THE DUAL OF THE VIRASORO ALGEBRA

GLORIA MARÍ BEFFA

KdV equations can be described as Hamiltonian systems on the dual of the Virasoro algebra with the canonical Lie-Poisson (also called Berezin-Kirillov-Kostant) bracket. In this paper we give an explicit transverse structure for this Poisson manifold along a finite dimensional submanifold. The structure is linearizable and equivalent to the Lie-Poisson structure on $\mathrm{sl}(2, \mathbf{R})^*$. This problem is closely related to the classification of Hill's equations.

1. Introduction and main definitions. It was known since Lie's time that if a manifold has a Poisson structure and the rank of the Poisson tensor is constant around a point (that is, the point is *regular*), then the manifold can be locally described at such a point as foliated into leaves of maximum rank or *symplectic* leaves. If the Poisson manifold is the dual of a Lie algebra with its *Lie-Poisson bracket*, then the symplectic leaves coincide with the orbits under the coadjoint action of the group. If the point is singular the local description can be achieved by finding a section which is transversal to the orbit of the point and which is endowed with a Poisson structure induced by the global Poisson bracket. This induced bracket, or *transverse structure*, was initially introduced by A. Weinstein for finite dimensional Poisson manifolds (see [**20**]) and it describes the relation between the symplectic structures on the different leaves as we cross them transversally to the orbit of a singular point. Weinstein also proved that transverse structures were unique in the following sense: if we have two sections transversal to the orbit of a singular point with Poisson brackets induced on them and with dimensions equal to the codimension of the orbit, then there exists a Poisson isomorphism of the manifold, defined between two neighbourhoods of the intersections with the orbit, which will clearly preserve the two transverse structures.

The aim of this paper is to show the geometrical description of the coadjoint orbits on the dual of the Virasoro algebra as we move transversally through them and to use this description to find an explicit transverse structure for its Lie-Poisson bracket. Descriptions and classifications of the coadjoint orbits have been given by different authors (see **[8]**, **[9]**, **[17]**, or **[21]**). The problem is closely related to finding normal forms for Hill's equations as we will see later.

In §2 we try to find a suitable transversal section in which we will define our structure. This direction might have a role in this work. In the case of the Virasoro algebra (a Frechèt manifold) there are no results on uniqueness of transverse structures available to us, so that, in principle, the transverse structure we get might not have been canonically chosen. We will show that it is enough to describe a transverse structure for constant potentials of the form $p = \frac{n^2}{2}$, for all integers n (these are analogous to *singular* points in finite dimensions). An orbit that does not contain such a potential will automatically possess a trivial transverse structure (potentials on these orbits are analogous to finite dimensional *regular* points). A direction transversal to an orbit which goes through a potential of the form $p = \frac{n^2}{2}$ is given by a 3-dimensional submanifold which is isomorphic to $\mathrm{sl}(2, \mathbf{R})$. In §3 we find a transverse structure along that section and we show how, although it is nonlinear, it can be linearized along the submanifold and therefore it is equivalent to the standard Lie-Poisson structure on $\mathrm{sl}(2, \mathbf{R})^*$. We also discuss how this fact does not imply a uniqueness result. The definition of transverse structure is also revised, to make it easier to adapt to the infinite dimensional case.

In the last section we provide an expression for the Taylor expansion of the transverse structure in terms of the even moments corresponding to a certain moment functional. This linear functional is defined as follows: the symplectic structure on the intersection of the coadjoint orbits with the transverse section can be, in some sense, represented by a Jacobi matrix. There exists a Jacobi fraction (continued fraction) associated to such a matrix and its corresponding partial denominators can be described as orthogonal polynomials with respect to certain discrete measure. The linear functional we are looking for is given by integrating against that measure.

Finally we show how the transverse structure can also be expressed in terms of the Fourier coefficients of a periodic solution of a nonhomogeneous equation whose homogeneous part is given by the coadjoint action of the algebra on its dual.

This paper is part of the author's Ph.D. thesis at the University of Minnesota. The author wishes to express her gratitude to J. F. Conn for introducing her to Poisson Geometry and infinite dimensional Lie algebras and for sharing with her his ideas about the subject. She would also like to thank D. Stanton, S. Angenent and especially Peter Olver. This work benefited very much from her talks with them.

Central extensions: The Virasoro algebra and the Lie-Poisson structure on its dual.

Let G be a Lie group, g its associated Lie algebra and g^* the dual space of g. We define the *Chevalley-Eilenberg complex* associated to a representation (V, ρ) for g, as the chain complex given by

$$\cdots \longrightarrow V \otimes \Lambda^1 g^* \xrightarrow{\delta_1} V \otimes \Lambda^2 g^* \xrightarrow{\delta_2} V \otimes \Lambda^3 g^* \longrightarrow \cdots$$

with the coboundaries defined as

$$\delta_1(\alpha)(\xi_1 \wedge \xi_2) = \rho(\xi_1)\alpha(\xi_2) - \rho(\xi_2)\alpha(\xi_1) - \alpha([\xi_1, \xi_2]),$$
$$\delta_2(\beta)(\xi_1 \wedge \xi_2 \wedge \xi_3) = \sum_{\sigma \in A_3} \rho(\xi_{\sigma_1})\beta(\xi_{\sigma_2} \wedge \xi_{\sigma_3}) + \sum_{\sigma \in A_3} \beta(\xi_{\sigma_1} \wedge [\xi_{\sigma_2}, \xi_{\sigma_3}]),$$

and where $\xi_1, \xi_2, \xi_3 \in g$, $[\ ,\ ]$ is the Lie bracket in the algebra, and A_3 is the space of cyclic permutations of $\{1, 2, 3\}$.

In particular, if $(V, \rho) = (\mathbf{R}, 0)$, the conditions above become

$$\delta_1(\alpha)(\xi_1 \wedge \xi_2) = -\alpha([\xi_1, \xi_2]),$$
$$\delta_2(\beta)(\xi_1 \wedge \xi_2 \wedge \xi_3) = \sum_{\sigma \in A_3} \beta(\xi_{\sigma_1} \wedge [\tau_{\sigma_2}, \tau_{\sigma_3}]).$$

We will denote by $H^2(g, (V, \rho)) \equiv H^2(g, V)$ the second cohomology group associated to the Chevalley-Eilenberg complex.

Given a nontrivial 2-cocycle $c \in H^2(g, \mathbf{R})$, define the Lie algebra $g_0 = g \oplus \mathbf{R}$ with Lie bracket

$$[(\xi, t), (\mu, s)]_0 = ([\xi, \mu], c(\xi, \mu)).$$

g_0 is called a *central extension* for g.

Let S^1 be the unit circle and G be the group of diffeomorphisms of S^1, $\mathrm{diff}(S^1)$, with the composition $\circ$ as the operation of the group. We can naturally identify g with the space of vector fields of the circle, $\mathrm{vect}(S^1)$ (for more information about infinite dimensional Lie algebras see **[14]** and **[16]**). The Lie bracket on g is given by the usual commutator

$$\left[\xi(\theta)\frac{\partial}{\partial\theta}, \eta(\theta)\frac{\partial}{\partial\theta}\right] = (\xi\eta' - \xi'\eta)\frac{\partial}{\partial\theta},$$

and the adjoint action of the group is carried out through a simple change of variables in the vector field, $\mathrm{Ad}(\phi)(\xi(\theta)\frac{\partial}{\partial\theta}) = (\phi'\xi)\circ\phi^{-1}\frac{\partial}{\partial\theta}$.

On the other hand, g^* can be identified with the space of 2-tensors on S^1 acting as

$$\left\langle p(\theta)\, d\theta \otimes d\theta\, , \xi(\theta)\frac{\partial}{\partial\theta}\right\rangle = \int_0^{2\pi} p(\theta)\xi(\theta)\, d\theta\, ,$$

and the coadjoint action of the group is then given by $\mathrm{Ad}(\phi)(p(\theta)\, d\theta^2) = \frac{p}{\phi'^2} \circ \phi^{-1}\, d\theta^2$, which is the usual change of variable for 2-tensors.

It is known that $H^2(\mathrm{vect}(S^1)\, , \mathbf{R}) \cong \mathbf{R}$ and a generator is given by

$$c\left(\xi\frac{\partial}{\partial\theta}\, , \mu\frac{\partial}{\partial\theta}\right) = \int_0^{2\pi} \xi'\mu''\, d\theta = \int_0^{2\pi} \xi'''\mu\, d\theta\, .$$

c is the so-called *Gelfand-Fuks cocycle.*

In the case when $g = \mathrm{vect}(S^1)$ and c is the Gelfand-Fuks cocycle, the central extension g_0 is called the *Virasoro algebra.*

c can be integrated to a cocycle in the group

$$B(\varphi\, , \phi) = \int_0^{2\pi} [\ln(\varphi \circ \psi)]'\, d(\ln \psi')\, ,$$

called *Bott's cocycle.* The group $G_0 = \mathrm{diff}(S^1) \times \mathbf{R}$ with operation

$$(\varphi\, , s) * (\psi\, , t) = (\varphi \circ \psi\, , t + s + B(\varphi\, , \psi))$$

is the Lie group that has g_0 as its corresponding Lie algebra. It is called the *Virasoro group.* Finally, g_0^* can be viewed as

$$g_0^* = \{(p(\theta)\, d\theta^2\, , s)\, , p(\theta)\ 2\pi\text{-periodic function, } s \in \mathbf{R}\} = g^* \oplus \mathbf{R}\, ,$$

acting on g_0 as

$$\langle (p\, , s)\, , (\xi\, , t)\rangle = \int_0^{2\pi} p(\theta)\xi(\theta)\, d\theta + ts\, ,$$

where, for convenience, we have denoted $p(\theta)\, d\theta \otimes d\theta$ and $\xi(\theta)\frac{\partial}{\partial\theta}$ by p and ξ, as we will often do from now on.

Let $\mathscr{H}$ be an element of $C^\infty(g^*)$. Define the *gradient* of $\mathscr{H}$ to be the element of g given by $\delta_p\mathscr{H}(\theta)\frac{\partial}{\partial\theta} \in g$, where $\delta_p\mathscr{H}(\theta)$ is a 2π-periodic function such that

$$\left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} \mathscr{H}(p + \varepsilon h) = \int_0^{2\pi} h(\theta)\delta_p\mathscr{H}(\theta)\, d\theta\, ,$$

for any 2π-periodic function h.

This definition establishes a correspondence between elements of $C^\infty(g^*)$ and elements of the Lie algebra. We can define the classical

Lie-Poisson structure on g^* as the one induced on g^* by the Lie bracket on g through the correspondence above, i.e.

$$\{\mathscr{H}, \mathscr{P}\}(\alpha) = \left\langle \alpha, \left[\delta_p\mathscr{H}(\theta)\frac{\partial}{\partial\theta}, \delta_p\mathscr{P}(\theta)\frac{\partial}{\partial\theta}\right]\right\rangle,$$

for any $\mathscr{H}, \mathscr{P} \in C^\infty(g^*)$, and any $\alpha \in g^*$.

If we denote by $\mathscr{H}$ and $\mathscr{P}$ two elements of $C^\infty(g_0^*)$, their gradients will have two partial components $(\delta_p\mathscr{H}, \delta_t\mathscr{H})$. By definition, the Lie-Poisson bracket on g_0^* is given by

$$\begin{aligned}\{\mathscr{H}, \mathscr{P}\}_0(p, s) &= \langle (p, s), [(\delta_p\mathscr{H}, \delta_t\mathscr{H}), (\delta_p\mathscr{P}, \delta_t\mathscr{P})]_0\rangle \\ &= \int_0^{2\pi} [\delta_p\mathscr{H}, \delta_p\mathscr{P}]\, p(\theta)\, d\theta + sc(\delta_p\mathscr{H}, \delta_p\mathscr{P}),\end{aligned}$$

for all $\mathscr{H}, \mathscr{P} \in C^\infty(g_0^*)$. Since the expression above does not depend on the value of $\mathscr{H}$ and $\mathscr{P}$ in the central direction, we can rewrite it in the usual way

$$\{\mathscr{H}, \mathscr{P}\}_0(p, s) = \{\mathscr{H}, \mathscr{P}\}(p) + sc(\delta\mathscr{H}, \delta\mathscr{P}). \tag{1.1}$$

The KdV equation $u_t = 3uu_x - u_{xxx}$ can be interpreted as a Hamiltonian system with respect to $\{\ ,\ \}_0$ in the following sense:

Consider the *evaluation operator* $\mathscr{D}$ defined as $\mathscr{D}(p) = p(\theta)$. That is, $\mathscr{D}$ has Dirac's delta function as gradient (Dirac's delta function does not give rise to a differentiable operator but it can be expressed as a series of differential kernels, so we view it in such an approximate way). If we consider the Hamiltonian operator $\mathscr{H}$ defined as

$$\mathscr{H}(p) = \frac{1}{2}\int_0^{2\pi} p^2(\theta)\, d\theta,$$

it is straightforward to check that the KdV equation is equal to the Hamiltonian system $u_t = \{\mathscr{H}, \mathscr{D}\}_0(u)$, with central charge $s = -1$ (for more information see **[1]**, **[2]**, **[6]** or **[7]**).

2. A transverse section to the orbits: Classification of Hill's equations. An explicit expression for the coadjoint action of the Virasoro group on the dual of the Virasoro algebra can be found in Kirillov's paper **[8]** and it is given by

$$K^*(\varphi)(p, s) = \left(\frac{p + sS(\varphi)}{\varphi'^2} \circ \varphi^{-1}, s\right), \tag{2.1}$$

where $S(\varphi)$ denotes the *Schwartz derivative* of φ, $S(\varphi) = (\varphi'''\varphi' - \frac{3}{2}\varphi''^2)/\varphi'^2$. One can obtain the coadjoint action of the Virasoro algebra on its dual by differentiating the expression (2.1)

$$k^*(\xi)(p, s) = (s\xi''' - 2p\xi' - p'\xi, 0). \tag{2.2}$$

The central charge s remains invariant under the action, that is, g_0^* stratifies into a family of Poisson submanifolds with constant central parameter. Each one of them is isomorphic to g^* with Poisson structure given as in (1.1) and they are all geometrically equivalent, except for the case $s = 0$. This is the usual change in the Poisson Geometry of the dual of a Lie algebra, produced by a central extension. Let's fix once and for all an adequate hyperplane inside g_0^*, namely $s = -1$.

Define *the stabilizer of a point* p to be the set of diffeomorphisms of the circle that fix the point p under the coadjoint action, i.e.,

$$\mathrm{Stab}(p) = \{\varphi \in \mathrm{diff}(S^1) \text{ such that } K^*(\varphi)(p, -1) = (p, -1)\}.$$

From (2.2) we deduce that the tangent of the stabilizer of p at the identity element is given by the vector space

$$T_{\mathrm{id}}(\mathrm{Stab}(p)) = \left\{\xi\frac{\partial}{\partial\theta} \in \mathrm{vect}(S^1) \text{ such that } \xi''' + 2p\xi' + p'\xi = 0\right\}. \tag{2.3}$$

A classification of the stabilizers of potentials was given in [**8**]. It was shown there that the set of solutions of (2.3) has a structure of Lie algebra which is isomorphic to $\mathrm{sl}(2, \mathbf{R})$, and that, furthermore, the number of periodic solutions of (2.3) is either 1 or 3, i.e., $\mathrm{Stab}(p)$ is either 1 or 3 dimensional. This dimension coincides with the codimension of the coadjoint orbit.

Let g be a Lie algebra with Lie bracket $[\ ,\ ]$. If $\mathrm{ad}(\xi)(\mu) = [\xi, \mu]$ is the usual adjoint action of the algebra, we define the *Killing form of* g to be the bilinear form $B(\xi, \mu) = \mathrm{tr}(\mathrm{ad}\,\xi(\mathrm{ad}\,\mu))$, for any $\xi, \mu \in g$.

PROPOSITION 2.1. *In the case of codimension* 3, *the coadjoint orbit contains a point of the form* $\frac{n^2}{2}\,d\theta^2$ *for some integer* n.

Proof. Assume that the codimension is three and let us consider $T_{\mathrm{id}}(\mathrm{Stab}(p))$ with its $\mathrm{sl}(2, \mathbf{R})$-structure. An expression for its Killing form was given in [**8**] and it is equal to

$$B(\xi) = -\frac{1}{4}\left(\xi\xi'' + p\xi^2 - \frac{1}{2}\xi'^2\right).$$

We know that the Killing form of $\mathrm{sl}(2, \mathbf{R})$ takes positive, negative and zero values; so does the Killing form of $T_{\mathrm{id}}(\mathrm{Stab}(p))$. If ξ is a periodic function with simple zeros we obtain

$$I(\xi) = \xi\xi'' + p\xi^2 - \frac{1}{2}\xi'^2 = -\frac{1}{2}\xi'^2 < 0.$$

If ξ has double zeros then $I(\xi) \equiv 0$, so that $I(\xi) > 0$ implies that ξ never vanishes.

For such a nonvanishing vector field $\xi \in T_{\text{id}}(\text{Stab}(p))$ choose φ to be equal to

$$\varphi(\theta) = \int_0^\theta \frac{A}{\xi(\theta)}\, d\theta ,$$

$A = \text{constant}$ chosen so that $\varphi(2\pi) = 2\pi$.

It is immediate that $\text{Ad}(\varphi)(\xi) = (\varphi'\xi)\circ\varphi^{-1} = A$ (remember that we denote by $\text{Ad}(\varphi)$ the coadjoint action of $\text{diff}(S^1)$). It is straightforward to check that the constant A should be a solution of the equation $\mu''' + 2K^*(\varphi)(p)\mu' + [K^*(\varphi)(p)]'\mu = 0$, and therefore $K^*(\varphi)(p) = p_1$ is also constant. Since the number of periodic solutions of (2.3) is preserved along the orbit, the equation $\mu''' + 2p_1\mu' = 0$ must have three independent periodic solutions. The only choice is $p_1 = \frac{n^2}{2}$ for some integer n and we are done. □

This last result entitles us to restrict the problem of finding a transverse structure to the case $p = \frac{n^2}{2}$: if the orbit does not go through $\frac{n^2}{2}$ for some n, the transverse section would be 1-dimensional and the transverse structure trivial. In fact, the codimension of the orbit is constant around a point different from $p = \frac{n^2}{2}$, and therefore we can refer to them as *regular* potentials. Furthermore, if $\text{Or}(p)$ goes through $\frac{n^2}{2}$ for some n we can immediately obtain a transverse structure at p translating from $\frac{n^2}{2}$ to p using the coadjoint action.

When $p = \frac{n^2}{2}$, three independent solutions for equation (2.3) are $\xi_1 = \cos(n\theta)$, $\xi_2 = \sin(n\theta)$, $\xi_3 = \textit{constant}$. Consider the linear section

$$(2.4)\qquad Q_n = \left\{ \left(\frac{n^2}{2} + a\cos(n\theta) + b\sin(n\theta) + c \right) d\theta^2 ,\quad a, b, c \in \mathbf{R},\ |c|, |a|, |b| < \delta \right\}$$

for some fixed integer n and some small δ that we will fix later on.

Proposition 2.2. *Q_n is transversal to the orbit of $\frac{n^2}{2}$ at $\frac{n^2}{2}$.*

Proof. Denote the orbit through p by $\text{Or}(p)$ and define the *annihilator of $T_{\text{id}}(\text{Or}(p))$* as the subset of g_0 given by

$$T_{\text{id}}(\text{Or}(p))^\perp = \{\xi \in g_0 \text{ such that } \langle \xi , k^*(\nu)(p)\rangle = 0 \text{ for all } \nu \in g_0\} .$$

It is easy to see that $T_{\mathrm{id}}(\mathrm{Or}(p))^{\perp} =$ kernel of $k^*(p)$, since

$$\begin{aligned}\langle \xi, k^*(\nu)p\rangle &= \int_{S^1} -\xi(\nu''' + 2p\nu' + p'\nu)\,d\theta \\ &= \int_{S^1} \nu(\xi''' + 2p\xi' + p'\xi)\,d\theta = -\langle k^*(\xi)(p), \nu\rangle .\end{aligned}$$

Besides, the pairing

$$\begin{gathered} T_{\mathrm{id}}\left(\mathrm{Or}\left(\frac{n^2}{2}\right)\right)^{\perp} \times Q_n \to \mathbf{R}, \\ (\xi, p) \to \langle p, \xi\rangle \end{gathered}$$

is nondegenerate, so that Q_n has to necessarily be transversal to $\mathrm{Or}(\frac{n^2}{2})$. □

Since the Virasoro algebra is a Frechèt manifold, there are no general inverse function theorems we could apply at this point to deduce straightforwardly that Q_n intersects all nearby orbits. This is an important condition on Q_n if we wish to describe the Poisson structure around $\frac{n^2}{2}$. To avoid this problem we need a description of the invariants of the coadjoint orbits to later check that they are all locally reached along Q_n. The classification of the orbits has been studied by several authors. Kirillov gave a classification of the stabilizers in his paper **[8]**. Lazutkin and Pankratova provided a partial description in **[9]**. Later on, Segal **[17]** pointed out a discrete invariant that was missing in **[9]** and gave the complete set of invariants which we are going to describe next.

First of all, we can identify g_0^* with the space of Hill's operators associating to a tensor $p\,d\theta^2$ the equation

$$\xi'' + \frac{p}{2}\xi = 0. \tag{2.5}$$

If ξ is a solution of (2.5), it is straightforward to prove that its *Liouville-Green transform*, $\mu = [(\varphi')^{1/2}\xi]\circ\varphi^{-1}$, is a solution of

$$\xi'' + \frac{K^*(\varphi)(p)}{2}\xi = 0. \tag{2.6}$$

Moreover, this is the only transform which preserves Hill's equations. In that sense we will view our manifold as the manifold of Hill's operators and the coadjoint action as a change of variable in the corresponding equation. Using this interpretation it is immediate to check that, if F_p is the *Floquet matrix* or *monodromy* associated to (2.5), then RF_pR^{-1} is the monodromy associated to (2.6) where

$$R = \begin{pmatrix} \varphi'^{1/2} & 0 \\ \frac{1}{2}\varphi'^{-3/2} & \varphi'^{-1/2} \end{pmatrix} (2\pi) \in \mathrm{SL}(2, \mathbf{R}).$$

That is, the $\mathrm{SL}(2, \mathbf{R})$-conjugation class of the monodromy matrix is preserved along the coadjoint orbit. This is one of the invariants of the orbit, in fact the only one that changes continuously. There exists a second invariant which we can describe in the following way:

Consider $u: \mathbf{R} \to \mathbf{R}^2 \backslash \{0\}$ to be an immersion given by two independent solutions of (2.5); we can assume that $u(0) = (1, 0)$ and $u'(0) = (0, 1)$. Let $\hat{u}: \mathbf{R} \to S^1$ be its radial projection and let n_p be the number of complete turns that $\hat{u}$ makes in a period. n_p is another invariant of the orbit, called a *discrete invariant* since it does not change continuously (again we can easily check that n_p is invariant using a Liouville-Green's transformation).

THEOREM 2.1. *Let u_1 and u_2 be two orientation-preserving immersions given as above by the solutions of two equations $\xi'' + \frac{p_1}{2}\xi = 0$ and $\xi'' + \frac{p_2}{2}\xi = 0$, respectively. If $n_{p_1} = n_{p_2} = m$ and $F_{p_1} = RF_{p_2}R^{-1}$ for some $R \in \mathrm{SL}(2, \mathbf{R})$, then p_1 is in the same orbit as p_2. That is, up to a Liouville-Green transformation, each Hill's equation corresponds to a different conjugacy class of the universal covering space of* $\mathrm{SL}(2, \mathbf{R})$ *under the* $\mathrm{SL}(2, \mathbf{R})$*-action.*

Proof. 1*st case*. Assume that $F_{p_1} = F_{p_2} = F$.

Then $\hat{u}_1$ and $\hat{u}_2$ make the same number of turns in a period and $\hat{u}_1(2\pi) = \hat{u}_2(2\pi)$. Divide the interval $[0, 2\pi]$ into several subintervals $[0, \theta_1], \dots, [\theta_i, \theta_{i+1}], \dots, [\theta_{m-1}, \theta_m], [\theta_m, 2\pi]$, such that $\hat{u}_1$ covers a complete turn on S^1 at each subinterval, except for $[\theta_m, 2\pi]$. Repeat the subdivision for $\hat{u}_2$. Then $\varphi = \hat{u}_2^{-1} \circ \hat{u}_1$ is smooth and well defined if we map each one of the $\hat{u}_1$-subintervals diffeomorphically into the corresponding $\hat{u}_2$-subinterval, that is, following in the mapping a natural order.

Because of the condition $F_{p_1} = F_{p_2} = F$, we get that $\varphi(\theta + 2\pi) = \varphi(\theta) + 2\pi$, and therefore φ is a diffeomorphism of the circle with $\hat{u}_2 \circ \varphi = \hat{u}_1$.

Finally, since $\hat{u}_s(\theta)$ is the radial projection of $u_s(\theta)$, $s = 1, 2$, we obtain that $u_1(\theta) = f(\theta)(u_2 \circ \varphi)$, for some differentiable and real-valued function f. Both u_1 and u_2 were given by solutions of Hill's equations. Through a straight substitution in the equations one can check that this condition imposes a unique possible choice, $f = \varphi'^{-1/2}$. Therefore, $p_1 = K^*(\varphi)p_2$ and this case is proved.

2nd case.

$$F_{p_2} = RF_{p_1}R^{-1}.$$

We know that

$$\hat{u}_i(\theta + 2\pi) = (F_{p_i}u_i)^\frown(\theta), \qquad i = 1, 2. \tag{2.7}$$

Therefore

$$(R^{-1}u_2)^\frown(\theta + 2\pi) = (F_{p_1}R^{-1}u_2)^\frown(\theta).$$

Denote the image of u as subset of $\mathbf{R}^2$ by $\mathrm{Im}(u)$. It is not hard to check that R can be chosen so that $\mathrm{Im}(\hat{u}_1)$ and $\mathrm{Im}(R^{-1}u_2)^\frown$ intersect at some point. In fact, we could use above $-R$ instead of R if they do not intersect (the sets $\mathrm{Im}(\hat{u}_1) \cap \mathrm{Im}(R^{-1}u_2)^\frown$ and $\mathrm{Im}(\hat{u}_1) \cap \mathrm{Im}(-R^{-1}u_2)^\frown$ cannot be simultaneously void). By translation in the argument, we can make the initial values coincide. $R^{-1}u_2$ and any translation of it is given by solutions of the same Hill's equation as u_2. We can now obtain this case as a corollary of the previous one. □

PROPOSITION 2.3. *The space of Hill's equations, up to Green-Liouville's transformations, is in one-to-one correspondence with the space of* $\mathrm{SL}(2, \mathbf{R})$ *conjugation classes of the Universal covering space of* $\mathrm{SL}(2, \mathbf{R})$, *with the point (Identity,* $n = 0$) *removed.*

Proof. First of all notice that if a matrix $M \in \mathrm{SL}(2, \mathbf{R})$ has two different eigenvalues (that is, $|\mathrm{trace}(M)| > 2$), then its $\mathrm{GL}(2, \mathbf{R})$ and $\mathrm{SL}(2, \mathbf{R})$ conjugation classes coincide. If $\mathrm{trace}(M) = \pm 2$, then the two different $\mathrm{SL}(2, \mathbf{R})$-Jordan forms are $\pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\pm\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$, and if $|\mathrm{trace}(M)| < 2$, then both eigenvalues are imaginary and there are also two different $\mathrm{SL}(2, \mathbf{R})$-Jordan forms, namely $\pm\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$ and $\pm\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$, $a, b > 0$.

Next, consider the potential

$$p_{\alpha,\beta}(\theta) = \begin{cases} \alpha^2 & \text{if } 0 \le \theta < \theta_0, \\ -\beta^2 & \text{if } \theta_0 \le \theta \le 2\pi \end{cases}$$

and consider its associated Hill's equation $\xi'' + p_{\alpha,\beta}\xi = 0$. A funda-

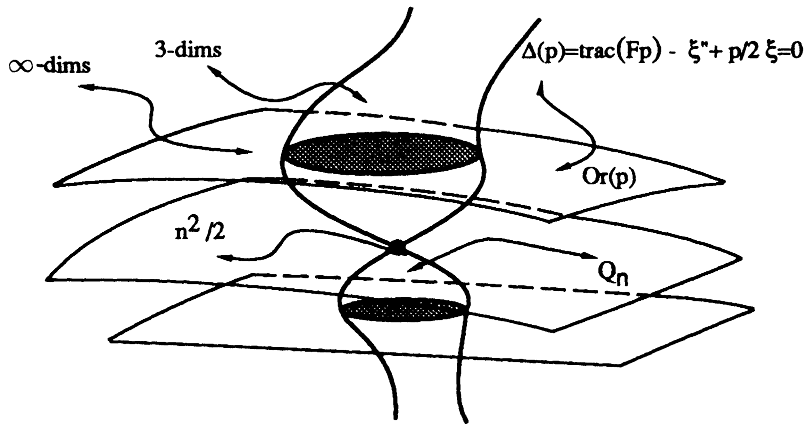

FIGURE 1

mental matrix of solutions for it is given by $X = X_2X_1$ where

$$X_1 = \begin{pmatrix} \cos(\alpha\theta_0) & \frac{1}{\alpha}\sin(\alpha\theta_0) \\ -\alpha\sin(\alpha\theta_0) & \cos(\alpha\theta_0) \end{pmatrix},$$

$$X_2 = \begin{pmatrix} \cosh(\beta(2\pi-\theta_0)) & \frac{1}{\beta}\sinh(\beta(2\pi-\theta_0)) \\ \beta\sinh(\beta(2\pi-\theta_0)) & \cosh(\beta(2\pi-\theta_0)) \end{pmatrix} \quad \text{if } \beta \neq 0, \text{ or}$$

$$X_2 = \begin{pmatrix} 1 & 2\pi-\theta_0 \\ 0 & 1 \end{pmatrix} \quad \text{if } \beta = 0.$$

The rotation number (in the above sense) of this equation is $\alpha\theta_0/2\pi$ plus an angle ω_0 with $\tan(\omega_0) < \frac{1}{\beta}$ and which is, in any case, less than $\frac{\pi}{2}$. Next, we will show that, for different values of α, β and θ_0 we obtain all possible SL(2, **R**)-Jordan forms, with all possible rotation numbers, except for the case of no complete turns (rotation number 0) and monodromy equals the identity.

If $\theta_0 = 2\pi$, then

$$X = \begin{pmatrix} \cos(2\pi\alpha) & \frac{1}{\alpha}\sin(2\pi\alpha) \\ -\alpha\sin(2\pi\alpha) & \cos(2\pi\alpha) \end{pmatrix}.$$

We can therefore cover the four possible Jordan forms corresponding to complex conjugated eigenvalues by choosing different values of α from the intervals $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$ and $[\frac{3}{4}, 2]$, respectively. Considering values $m\alpha$ with m being an integer $m \geq 1$, we would obtain the same Jordan forms but the rotation number would be $m-1$. The identity matrix is reached here whenever α is a nonzero whole number. It is never reached for rotation number equals zero, since the solution curve is in this case periodic and it should, at least, give a complete turn around S^1.

If $\theta_0 = \pi$ and $\alpha = 2, 4, \dots$, then $\operatorname{trace}(X) = e^{2\pi\beta} + e^{-2\pi\beta} > 2$. Changing the values of β and α we will reach all possible values of the trace and all rotation numbers. If $\theta_0 = \pi$ and $\alpha = 1, 3, 5, \dots$, then $\operatorname{trace}(X) = -(e^{2\pi\beta} + e^{-2\pi\beta}) < -2$ and the same result holds.

Therefore, we are left with the classes which have Jordan forms $\pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\pm\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$. If $\theta_0 = \alpha$, $\alpha = 0, 2, 4, \dots$ and $\beta = 0$, we obtain the classes with Jordan form $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ (and $-\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ for the choices $\alpha = 1, 3, 5, \dots$) and all different rotation numbers.

Finally, consider $\theta_0 = \frac{\pi}{2}$ and $\alpha = 3, 7, \dots$. Then,

$$X = \begin{pmatrix} \cosh(\frac{3\pi\beta}{2}) & \frac{1}{\beta}\sinh(\frac{3\pi\beta}{2}) \\ \beta\sinh(\frac{3\pi\beta}{2}) & \cosh(\frac{3\pi\beta}{2}) \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{\alpha} \\ \alpha & 0 \end{pmatrix}$$

has a double eigenvalue whenever $(\frac{\alpha}{\beta} - \frac{\beta}{\alpha})^2 \sinh^2(\frac{3\pi\beta}{2}) = 4$. Its eigenvalues would be $\pm 1 = \pm\frac{1}{2}(\frac{\alpha}{\beta} - \frac{\beta}{\alpha})\sinh(\frac{3\pi\beta}{2})$ depending on the sign of $(\frac{\alpha}{\beta} - \frac{\beta}{\alpha})\sinh(\frac{3\pi\beta}{2})$. In this case, X will have Jordan form $\pm\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$. On the other hand,

$$\lim_{\beta\to 0+} \left(\frac{\alpha}{\beta} - \frac{\beta}{\alpha}\right) \sinh\left(\frac{3\pi\beta}{2}\right) = \frac{3\pi\alpha}{2} > 2$$

and

$$\lim_{\beta\to +\infty} \left(\frac{\alpha}{\beta} - \frac{\beta}{\alpha}\right) \sinh\left(\frac{3\pi\beta}{2}\right) = -\infty.$$

From here it is obvious that this last case is also covered.

If we approximate $p_{\alpha,\beta}$ by C^∞ periodic functions we will immediately obtain the claim of the proposition. □

Using this geometrical description it is easy to prove the following theorem.

THEOREM 2.2. *The transverse section* Q_n *(see* (2.4)) *intersects all orbits nearby the one going through the potential* $\frac{n^2}{2}\, d\theta^2$.

Proof. It suffices to prove that the map $Q_n \to \mathrm{SL}(2, \mathbf{R})$, which associates to each potential the monodromy of equation (2.5), is locally surjective. If we expand the monodromy as a function of (a, b, c) we obtain

$$F_p = \text{Identity } + \pi\left\{\begin{pmatrix} 0 & -1 \\ \frac{4}{n^2} & 0 \end{pmatrix} c + \begin{pmatrix} 0 & -\frac{1}{2} \\ -\frac{2}{n^2} & 0 \end{pmatrix} a + \begin{pmatrix} -\frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{pmatrix} b\right\}$$
$$+ \text{ higher order terms},$$

which has maximal rank. □

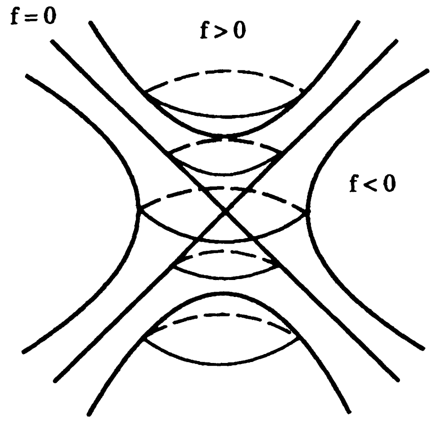

FIGURE 2

Assume that $p = \frac{n^2}{2} + c + a\cos(n\theta) + b\sin(n\theta) \in Q_n$. We can also calculate the Taylor expansion of $\Delta(p)$ in a, b and c up to second order (see **[10]**). We are required to solve two ordinary second order differential equations for each Taylor coefficient we want to find.

After some long calculations one gets

$$\Delta(p) = (-1)^n \left\{ 2 + \frac{2\pi^2}{n^2} \left[\left(\frac{a}{2}\right)^2 + \left(\frac{b}{2}\right)^2 - c^2 \right] \right\} \\ + \text{ higher order terms in } (a, b, c).$$

Let us have a closer look to the real function $f(x, y, z) = x^2+y^2-z^2$. Its level sets are given as in Figure 2. Recall that f is a nontrivial Casimir element for the Lie-Poisson bracket of $\mathrm{sl}(2, \mathbf{R})^*$.

Recall also that $\Delta(p)$ is constant along each orbit, in particular along each intersection of the orbit with Q_n. If there exists a transverse structure for $(g_0, \{\ ,\ \}_0)$ on Q_n, say $\{\ ,\ \}_Q$, one expects the Kirillov leaves of $\{\ ,\ \}_Q$ to be such an intersection. Since a function that is constant along the symplectic leaves is a Casimir function, $\Delta(p)$ would be a Casimir for $\{\ ,\ \}_Q$. Therefore, we can make a guess and claim that $(Q_n, \{\ ,\ \}_Q)$ is locally isomorphic to $\mathrm{sl}(2, \mathbf{R})^*$ with its canonical Lie-Poisson structure. This is actually one of the main results in the next section.

Theorem 2.2 partially proves a claim by Lazutkin and Pankratova about normal forms of Hill's equations. In their paper (**[9]**) they claim that any Hill's equation has normal form $\xi''+(d+e\cos(n\theta))\xi = 0$, for some real numbers d and e. This normal form can be achieved under a Liouville-Green transformation. From Theorem 2.2 any potential p can be taken to the intersection of the leaf with Q_n using the coadjoint action, as far as it belongs to an orbit close enough to $\mathrm{Or}(\frac{n^2}{2})$, for some integer n. Besides $c + a\cos(n\theta) + b\sin(n\theta) = c + \beta\cos(n\theta + \alpha)$ for some $\alpha, \beta \in \mathbf{R}$. The result follows. The methods used on ([**9**]) are different from the ones in this paper.

3. A transverse structure for g_0^*, $\{\ ,\ \}_0$.

3.1. *Induced Poisson structures*: *Transverse structures and Dirac formalism. Transverse structures in infinite dimensions.* In the finite dimensional case, transverse structures were introduced by A. Weinstein [**20**] and they were proved to be unique. Some results have already been proved in the infinite dimensional case, whenever the manifold is modelled by a Hilbert or Banach space (see [**11**]). That is not our case either since $\mathrm{vect}(S^1)$ is a Frechèt manifold (it is not only that the Fourier series of an element has to converge, but all the series of its derivatives). Therefore, we now encounter one of the obstacles in this work: it is not clear how to induce a Poisson structure in this kind of space.

The idea we will follow is to imitate the finite dimensional procedure, covering any gap we find in some appropriate way. In particular, we will find the analogue of Dirac's formula for transverse structures in finite dimensions and we will check that it actually defines a Poisson structure on Q_n which is induced by the Lie-Poisson structure of the Virasoro algebra (for more information about induced Poisson structures see for example [**12**], [**18**], or [**13**], or [**20**]).

DEFINITION. Let $L_p = Q_n \cap \mathrm{Or}(p)$. Assume (1) $\{\ ,\ \}_0$ induces a nondegenerate (symplectic) structure on L_p, for all $p \in Q_n$.

(2) There exists a smooth (resp. analytic) Poisson structure on Q_n, $\{\ ,\ \}_Q$, that induces the same structure as $\{\ ,\ \}_0$ on L_p.

$\{\ ,\ \}_Q$ is called a *smooth* (*resp. analytic*) *transverse structure for* $(g_0^*, \{\ ,\ \}_0)$ in the direction of Q_n.

THEOREM 3.1 [**20**] (*Induced Poisson structures in finite dimensions*). *Let M be a finite dimensional Poisson manifold with Poisson tensor P. Let Q be an immersed submanifold of M. Assume that, for all $x \in Q$,*

(a) $$P(x)(T_x(Q)^\perp) \cap T_x(Q) = \{0\},$$

(b) $$\mathrm{Ker}(P(x)) \cap T_x(Q)^\perp = 0.$$

Then Q canonically inherits a Poisson structure from M, which we will denote by P_Q.

We will make some comments on how the induced Poisson structure is found.

Using (a) and (b) it is not hard to show that

$$T_x(M) = T_x(Q) \oplus P(x)(T_x(Q)^{\perp})$$

which provides itself a smooth projection $\pi\colon T_x(M) \to T_x(Q)$ whenever $x \in Q$. The induced Poisson structure is then defined as

$$P_Q = \pi \circ P \circ \pi^*\colon T_x(Q)^* \to T_x(Q).$$

In other words, given a Hamiltonian function $\alpha \in T_x(Q)^*$ we can find an extension of it, $\pi^*(\alpha) = \hat{\alpha} \in T_x(M)^*$. The vector field $P(x)(\hat{\alpha}) \in T_x(M)$ has a component on $T_x(Q)$. Such a component is the value of $P_Q(x)(\alpha)$, and it is found taking away from $P(x)(\hat{\alpha})$ a linear combination of elements in $P(x)(T_x(Q)^{\perp})$.

In local coordinates the idea is as follows:

Let $\{z_1, \dots, z_{2s}\}$ be independent defining functions for Q near x. That is, $Q = \{x \in M\colon z_1(x) = z_2(x) = \cdots = z_{2s}(x) = 0\}$. Denote by $C(y) = (C_{ij}(y))$ the matrix $C_{ij}(y) = \{z_i, z_j\}(y)$, with $i, j = 1, \dots, 2s$ and $y \in Q$. This matrix has smooth (resp. analytic) entries and it is nonsingular. Let $C^{-1}(y) = (C^{ij}(y))$ be its inverse matrix, which also has smooth (resp. analytic) entries. Let f be a smooth function on Q and $\hat{f}$ be any extension of f to M. Due to the invertibility of C one can easily show that there exist unique smooth functions $\{g_i(y)\}_{i=1}^{2s}$ defined on a neighbourhood of x such that, if

$$P_Q(f)(y) = P(\hat{f})(y) + \sum_{i=1}^{2s} g_i(y) P(z_i)(y)$$

then $P_Q(f)(y) \in T_y(Q)$ for all $y \in Q$ in a neighbourhood of x. Imposing the tangency condition on $P_Q(f)$ we can uniquely solve for g_i in terms of the entries of C^{-1}.

The final expression for P_Q is

$$\{f, g\}_Q(y) = \{\hat{f}, \hat{g}\}(y) + \sum_{i=1}^{2s} \{\hat{f}, z_i\}(y) C^{ij}(y) \{z_j, \hat{g}\}(y), \tag{3.1}$$

for all $y \in Q$ around x. This formula is referred as *Dirac's formula for induced structures.*

P_Q immediately induces a structure on Q whose symplectic leaves coincide with the intersection of Q and the symplectic leaves of P. Notice that, in order to find an expression for P_Q, we need not only nondegeneracy of the bracket along the coadjoint orbits but also to invert it locally.

PROPOSITION 3.1. *For δ small enough, $\{\ ,\ \}_0(p)$ is nondegenerate (symplectic) on $T_p(Q_n)^\perp$, for all $p \in Q_n$, and locally invertible when considered as a linear operator from l^2 to l^2 (δ as in (2.4)). Coordinates can be chosen such that $\{\ ,\ \}_0$ and its inverse are represented by infinite matrices with analytic entries.*

Proof. Consider Fourier coefficients as coordinates for the dual of the Virasoro algebra, $\varepsilon_m\colon g_0^* \to \mathbf{R}$ defined as

$$\varepsilon_m(p) = \frac{1}{2\pi}\int_0^{2\pi} e^{im\theta} p(\theta)\, d\theta, \quad \text{for any integer } m,$$

Q_n is locally defined as the zero set of $\{\varepsilon_m\}_{m\neq \pm n, 0}$.

In order to prove the proposition, we need to check that the linear operator $C\colon l^2 \to l^2$, represented by the infinite matrix $C(p) = (\{\varepsilon_m, \varepsilon_k\}_0(p))_{m,k\neq \pm n, 0}$, is invertible for any $p \in Q_n$.

Assume that $p = \frac{n^2}{2} + c + a\cos(n\theta) + b\sin(n\theta) \in Q_n$, so that $\varepsilon_0(p) = \frac{n^2}{2} + c$, $\varepsilon_n(p) = \frac{1}{2}(a+bi)$, $\varepsilon_{-n}(p) = \frac{1}{2}(a-bi)$ $(i^2 = -1)$. Straightforwardly, one can show that

$$\tag{3.2} \begin{aligned} \{\varepsilon_{-k}, \varepsilon_k\}_0(p) &= \frac{k}{2\pi i}(2\varepsilon_0(p) - k^2), \\ \{\varepsilon_{-(n+k)}, \varepsilon_k\}_0(p) &= \frac{(2k+n)}{2\pi i}\varepsilon_{-n}(p), \\ \{\varepsilon_{n-k}, \varepsilon_k\}_0(p) &= \frac{(2k-n)}{2\pi i}\varepsilon_n(p), \\ \{\varepsilon_m, \varepsilon_k\}_0(p) &= 0, \quad \text{otherwise}. \end{aligned}$$

To invert the matrix above is equivalent to solve for $\{\gamma_m\}_{m=-\infty}^{+\infty}$ in the system

$$\tag{3.3} \sum_m \gamma_m \{\varepsilon_m, \varepsilon_k\}_0(p) = b_k,$$

for all k, and for some $B = \{b_k\}$ given. Let us assume that $\{\gamma_k\}$ and $\{b_k\}$ are elements of l^2. We can rescale so that system (3.3) becomes

$$\tag{3.4} \begin{aligned} \gamma_{-k}k(2\varepsilon_0(p) - k^2) + \gamma_{-(n+k)}(2k+n)\varepsilon_{-n}(p) \\ + \gamma_{n-k}(2k-n)\varepsilon_n(p) = b_k, \end{aligned}$$

for any integer k.

Observe that the system (3.4) can be divided into a finite number of autonomous subsystems. Each one of them involves only the γ's whose subindices belong either to the set $\{k = sn + r,\ s = 1, 2, \dots\}$ for a fixed integer $r \neq 0$, $-n < r < n$ or to the set $\{k = 2n,\ s = 2, 3, \dots\}$. There are a finite number of subsystems so that it is enough to prove that each one can be solved in l^2. The reasoning is the same for any one of them. We will prove it here only for the system given by the subindices $\{k = sn,\ s = 2, 3, \dots\}$. For simplicity call $\gamma_{-sn} = \gamma_s$ since no confusion is possible from now on.

We can rewrite the system to solve as $A(p)\gamma = b$, where $A(p)$ is given by the infinite tridiagonal matrix (Jacobi matrix)

$$\begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & (s-1)(2\varepsilon_0(p) - [[s-1]n]^2) & [2s-1]\varepsilon_{-n}(p) & 0 & \cdots \\ \cdots & [2s-1]\varepsilon_n(p) & s(2\varepsilon_0(p) - [sn]^2) & [2s+1]\varepsilon_{-n}(p) & \cdots \\ \cdots & 0 & [2s+1]\varepsilon_n(p) & (s+1)(2e_0(p) - [[s+1]n]^2) & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}.$$

Observe that $A(\frac{n^2}{2})$ is a diagonal matrix with nonvanishing diagonal entries. Therefore, $A(\frac{n^2}{2})$ is an invertible matrix and its inverse is a diagonal matrix with diagonal entries $\frac{1}{s(n^2-(sn)^2)}$, $s = 2, 3, \dots$. Although $A(\frac{n^2}{2})$ does not take l^2 into l^2, its inverse does.

Observe also that, if we define the matrix $D(p)$ through the relation $A(p) = A(\frac{n^2}{2}) + D(p)$, to solve the system $A(p)\gamma = b$ is equivalent to solve for γ in

$$A\left(\frac{n^2}{2}\right)^{-1} b = \left[I + A\left(\frac{n^2}{2}\right)^{-1} D(p)\right] \gamma,$$

where $A(\frac{n^2}{2})^{-1}$ is the inverse matrix of $A(\frac{n^2}{2})$. The matrix $A(\frac{n^2}{2})^{-1}D(p)$ is given by the tridiagonal matrix

$$\begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & \frac{(s-1)(2\varepsilon_0(p)-[[s-1]n]^2)}{(1-(s-1)^2)(s-1)n^2} & \frac{[2s-1]\varepsilon_{-n}(p)}{(1-(s-1)^2)(s-1)n^2} & 0 & \cdots \\ \cdots & \frac{[2s-1]\varepsilon_n(p)}{s(1-s^2)n^2} & \frac{s(2\varepsilon_0(p)-[sn]^2)}{s(1-s^2)n^2} & \frac{[2s+1]\varepsilon_{-n}(p)}{s(1-s^2)n^2} & \cdots \\ \cdots & 0 & \frac{[2s+1]\varepsilon_n(p)}{[1-(s+1)^2](s+1)n^2} & \frac{[s+1](2\varepsilon_0(p)-[[s+1]n]^2)}{[1-(s+1)^2](s+1)n^2} & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}.$$

The infinite dimensional operator $A(\frac{n^2}{2})^{-1}D(p)\colon l^2 \to l^2$ represented by this matrix is clearly bounded with norm bounded by $|\delta|$

(δ as in (2.4)). We can apply standard theorems on invertibility of linear operators on Hilbert spaces that are perturbations of the identity to obtain that, for δ small enough, the matrix $I + A(\frac{n^2}{2})^{-1}D(p)$ is invertible, for any $p \in Q_n$.

The matrix $I + A(\frac{n^2}{2})D(p)$ has analytic entries in ε_0, ε_n and ε_{-n}. Therefore, $A(p)^{-1} = [I + A(\frac{n^2}{2})^{-1}D(p)]^{-1}A(\frac{n^2}{2})^{-1}$ also has analytic entries and the result of the proposition is now proved. □

Denote by $(C^{ij}(p)) = C^{-1}$ the inverse matrix of C as in Proposition 3.1.

LEMMA 3.1. *Let $p = (a, b, c) = (\varepsilon_0(p), \varepsilon_n(p), \varepsilon_{-n}(p)) \in Q_n$, and let $C^{-2n2n}(p)$ be the entry in place $(-2n, 2n)$ of C^{-1}.*

Then, $\frac{1}{2\pi i}C^{-2n2n}$ is a real analytic function of $(\varepsilon_0(p), (\varepsilon_n(p)\varepsilon_{-n}(p)))$. That is, it depends only on c and the ratio $a^2 + b^2$.

Proof. This lemma is a corollary of §3.2, Theorem 3.7, in which we give an explicit Taylor expansion for it. A shorter proof can be given but we will avoid it. □

THEOREM 3.2. *A transverse structure for the dual of the Virasoro algebra at the point $\frac{n^2}{2}\,d\theta^2$ is given locally by an antisymmetric tensor, $\{\ ,\ \}_Q$, defined as*

$$\{\varepsilon_0, \varepsilon_n\}_Q(p) = \frac{n}{2\pi i}\varepsilon_n(p),$$
$$\{\varepsilon_0, \varepsilon_{-n}\}_Q(p) = \frac{-n}{2\pi i}\varepsilon_{-n}(p),$$
$$\{\varepsilon_n, \varepsilon_{-n}\}_Q(p) = \frac{1}{2\pi i}\left[-2n\varepsilon_0(p) + n^3 - 9n^2\varepsilon_n(p)\varepsilon_{-n}(p)\frac{C^{-2n2n}(p)}{2\pi i}\right].$$

The structure is analytic in $\{\varepsilon_0, \varepsilon_n, \varepsilon_{-n}\}$, linearizable and equivalent to the Lie-Poisson structure on $\mathrm{sl}(2, \mathbf{R})^*$.

Proof. Notice that we are actually copying the formula in coordinates given by Theorem 3.1. Define $\{\ ,\ \}_Q$ as

$$\{\varepsilon_i, \varepsilon_j\}_Q(p) = \{\varepsilon_i, \varepsilon_j\}_0(p) + \sum_{k,l\neq\pm n,0} \{\varepsilon_i, \varepsilon_k\}_0(p)C^{kl}(p)\{\varepsilon_l, \varepsilon_j\}_0(p).$$

Applying commutation relations (3.2), the formula above gives the expression in the statement of the theorem. This expression is found following formally the finite dimensional reasoning in Theorem 3.1.

On the other hand, since this is only a formal approach we will have to check straightforwardly that P_Q defines a Poisson structure on Q_n and that the intersections of Q_n with the symplectic leaves are symplectic with respect to both structures. Notice also that (a) and (b) on Theorem 3.1 are also true here due to the nondegeneracy of $\{\ ,\ \}_0$ along the leaves. Nevertheless, one cannot get the splitting of the tangent space into a direct sum as it happened in the finite dimensional case.

Leibniz's rule is obvious from the definition. To check Jacobi's identity for $\{\ ,\ \}_Q$ reduces to prove that

$$(3.5) \qquad \{\varepsilon_0, \{\varepsilon_n, \varepsilon_{-n}\}_Q\} + \{\varepsilon_n, \{\varepsilon_{-n}, \varepsilon_0\}_Q\} + \{\varepsilon_{-n}, \{\varepsilon_0, \varepsilon_n\}_Q\}_Q = 0,$$

on Q_n. Substituting we reduce (3.5) to

$$(3.6) \qquad \{\varepsilon_0, (\varepsilon_n\varepsilon_{-n})C^{-2n2n}\}_Q(p) = 0.$$

As a result of Lemma 3.1, $(\varepsilon_n\varepsilon_{-n})C^{-2n2n}$ restricted to Q_n is actually a function of the ratio $(\varepsilon_n(p)\varepsilon_{-n}(p))$. Applying Leibniz's rule and the definition of $\{\ ,\ \}_Q$ one gets that ε_0 commutes with the ratio along Q, and therefore (3.6) holds.

The last part is to check that P is symplectic on the intersections of Q_n with the symplectic leaves, L_p. As it happened in the finite dimensional case, that is a consequence of property (a) in Theorem 3.1, since the intersection $P(p)(T_p(Q_n)^{\perp}) \cap T_p(Q_n)$ is equal to the kernel of P along L_p, and in this case it vanishes.

Finally, we apply the following result by J. Conn **[4]** (see **[5]** for the smooth case): *if a Lie algebra* g *is semisimple* (*as* $\mathrm{sl}(2, \mathbf{R})$ *is*), *then any analytic Poisson structure on* g^*, *which is a perturbation of the Lie-Poisson structure by a tensor of order at least* 2 *that vanishes at the origin, is linearizable.*

It is now obvious that P_Q is linearizable and equivalent to the Lie-Poisson structure on $\mathrm{sl}(2, \mathbf{R})^*$. □

One comment on the linearization. Notice that by being linearizable we mean linearizable as structure on Q_n, not as a structure induced by g_0^*. That is, this result does not imply that a canonical transverse structure for the Lie-Poisson structure on the dual of the Virasoro algebra is the Lie-Poisson structure on $\mathrm{sl}(2, \mathbf{R})^*$, since no uniqueness result has been proved yet. What the result really means is that we can find coordinates (only) on Q_n such that $\{\ ,\ \}_Q$ on those coordinates is linear. In order to prove uniqueness we would need to extend that change of coordinates to g_0^* obtaining in this way an automorphism of the Lie-Poisson structure on g_0^*. We will comment more about uniqueness at the end of the section.

3.2. *The explicit expression for* $\{\ ,\ \}_Q$. In this section we will give an explicit expression for the Taylor expansion of the function $C^{-2n2n}(p)$.

Recall that our function is the entry in place $(-2n, 2n)$ of the matrix $(C^{km}(p))$, inverse of $(C_{km}(p)) = (\{\varepsilon_k, \varepsilon_m\}_0(p))_{k,m\neq\pm n,0}$. If again we set the system of equations $C\gamma = e_{2n}$, where $\gamma = \{\gamma_k\}$ and e_{2n} is a vector that has all its components equal to zero except for 1 in place $2n$, then $\gamma_{-2n} = C^{-2n2n}(p)$. Again, if for simplicity we write γ_k instead of γ_{-nk}, we get the recurrence relation

$$k(2\varepsilon_0 - (kn)^2)\gamma_k + (2k-1)\varepsilon_n\gamma_{k-1} + (2k+1)\varepsilon_{-n}\gamma_{k+1} = 0, \tag{3.7}$$

for any $k > 2$, and

$$2(2\varepsilon_0 - (2n)^2)\gamma_2 + 5\varepsilon_{-n}\gamma_3 = \beta,$$

where $\beta = \frac{2\pi i}{n}$.

Proposition 3.2. *For any* $k \geq 2$,

$$F_k(\varepsilon_n\varepsilon_{-n}, \varepsilon_0)\gamma_2 = H_k(\varepsilon_n\varepsilon_{-n}, \varepsilon_0)\beta - \delta_k\gamma_k$$

where F_k, H_k *satisfy the recurrence relation*

$$G_{k+1} = XG_k - \frac{(2k-1)^2}{k(k-1)((kn)^2 - Y)(((k-1)n)^2 - Y)}G_{k-1}, \tag{1}$$

with

$$X = \frac{1}{(\varepsilon_n\varepsilon_{-n})^{\frac{1}{2}}}, \quad Y = 2\varepsilon_0, \quad \delta_k = -\frac{(2k-1)\varepsilon_{-n}X}{(k-1)(Y - ((k-1)n)^2)}\delta_{k-1}$$

and initial conditions

$$F_1 = 0, \quad F_2 = 1, \quad H_2 = 0, \quad H_3 = \frac{1}{2(Y - (2n)^2)},$$
$$\delta_3 = -\frac{5\varepsilon_{-n}X}{2(Y - (2n)^2)}.$$

Proof. The proof of this proposition is by induction on k. □

The solutions of the recurrence relation (1) can be interpreted as orthogonal polynomials in X as we will show next.

A. *Jacobi fractions and orthogonal polynomials*: *Definitions and some results.*

DEFINITIONS. (a) Let $\{a_n\}$ and $\{b_n\}$ be arbitrary sequences of complex numbers and write

$$C_0 = b_0\,, \quad C_1 = b_0 + \frac{a_1}{b_1}\,, \quad C_2 = b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2}}\,, \ldots$$

$$C_n = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \ddots\, (a_n/b_n)}}\,.$$

C_n is called the nth *approximant of the continued fraction associated to the sequences* $\{a_n\}$, $\{b_n\}$. We will denote C_n as

$$C_n = b_0 + \frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + \cdots + \frac{a_n|}{|b_n}\,.$$

(b) A continued fraction of the form

$$\frac{\lambda_1|}{|x - c_1} - \frac{\lambda_2|}{|x - c_2} - \frac{\lambda_3|}{|x - c_3} - \cdots$$

is called a *Jacobi type continued fraction* (J-*fraction*).

(c) If C is a continued fraction and $C_n = A_n/B_n$, then A_n and B_n are called nth *partial numerator and* nth *partial denominator*, respectively.

Note. If A_n and B_n are the partial numerators and denominators for a J-fraction

$$\frac{\lambda_1|}{|x - c_1} - \cdots - \frac{\lambda_n|}{|x - c_n} - \cdots$$

it is very simple to prove that they satisfy the recurrence relations

$$B_n(x) = (x - c_n)B_{n-1}(x) - \lambda_n B_{n-2}(x)\,, \qquad n = 1, 2, 3, \ldots,$$
$$B_{-1}(x) = 0\,, \quad B_0(x) = 1\,,$$
$$A_n(x) = (x - c_n)A_{n-1}(x) - \lambda_n A_{n-2}(x)\,, \qquad n = 1, 2, \ldots,$$
$$A_{-1}(x) = 1\,, \quad A_0(x) = 0\,.$$

Notice the similarities between these expressions and the recurrence problem (1).

(d) Let $\{\mu_n\}$ be a sequence of complex numbers and let $\mathscr{L}$ be a complex-valued linear function defined on the vector space of all polynomials by the rule

$$\mathscr{L}(x^n) = \mu_n\,, \qquad n = 0, 1, 2, \ldots.$$

$\mathscr{L}$ is called *moment functional* determined by the *formal moment sequence* $\{\mu_n\}$. μ_n is called the *moment of order* n.

(e) A moment functional $\mathscr{L}$ is called *positive-definite* if $\mathscr{L}(\pi(x)) > 0$ for every polynomial $\pi(x)$ that is not identically zero and is nonnegative for all real x.

(f) Let $\mathscr{L}$ be a moment functional with moment sequence $\{\mu_n\}$. Define

$$\Delta_n = \det(\mu_{i+j})_{i,j=0}^n .$$

We will say that $\mathscr{L}$ is *quasi-definite* whenever $\Delta_n \neq 0$ for all $0 \leq n$.

(g) A sequence $\{P_n(x)\}$ is called an *Orthogonal Polynomial Sequence* (OPS) with respect to a moment functional $\mathscr{L}$ provided that, for all nonnegative integers m and n

(i) $P_n(x)$ is a polynomial of degree n,
(ii) $\mathscr{L}(P_m(x)P_n(x)) = 0$ for all $m \neq n$,
(iii) $\mathscr{L}(P_n^2(x)) \neq 0$.

It is not hard to notice that OPS are uniquely determined up to the product by a nonvanishing constant. The next theorem shows how partial denominators for a J-fraction can be interpreted as OPS with respect to a certain moment functional.

FAVARD'S THEOREM. *Let $\{c_n\}$ and $\{\lambda_n\}$ be arbitrary sequences of complex numbers and let $\{P_n(x)\}$ be defined by the recurrence formula*

$$(3.8) \qquad P_n(x) = (x - c_n)P_{n-1}(x) - \lambda_n P_{n-2}(x), \quad n = 1, 2, 3, \dots,$$
$$P_{-1}(x) = 0, \quad P_0(x) = 1.$$

Then, there is a unique moment functional $\mathscr{L}$ such that

$$\mathscr{L}(1) = \lambda_1, \quad \mathscr{L}(P_m(x)P_n(x)) = 0$$

for $m \neq n$, $m, n = 0, 1, 2, \dots$.

$\mathscr{L}$ is quasi-definite and $\{P_n(x)\}$ are the corresponding monic OPS if and only if $\lambda_n \neq 0$, while $\mathscr{L}$ is positive-definite if and only if c_n are real and $\lambda_n > 0$ $(n \geq 1)$.

Consider the OPS $\{P_n(x)\}$ with recurrence formula as in Favard's theorem, and define $P_n^{(1)}(x)$ to be a monic polynomial of degree n which satisfies the recurrence

$$P_n^{(1)}(x) = (x - c_{n+1})P_{n-1}^{(1)}(x) - \lambda_{n+1}P_{n-2}^{(1)}(x), \qquad n = 1, 2, 3, \dots,$$
$$P_{-1}^{(1)}(x) = 0, \quad P_0^{(1)}(x) = 1.$$

The polynomials $P_n^{(1)}(x)$ are called the *monic numerator polynomials* (or *associated polynomials*) corresponding to $P_n(x)$.

It is now clear to us that partial denominators and numerators of a J-fraction, B_k and A_{k+1}, are respectively OPS and associated polynomials with respect to a certain moment functional that is a positive-definite if and only if the partial fractions have all real coefficients and the numerators λ_k are all positive. Observe that A_k are not actually monic unless $\lambda_1 = 1$. To be correct, the associated polynomials are $\lambda_1^{-1} A_{k+1}$, $0 \leq k$.

DEFINITION. A moment functional is called *symmetric* if all of its moments of odd order are zero. This is equivalent to $c_n = 0$, $n \geq 1$, in the corresponding recurrence formula.

We can easily recognize the recurrence in problem (1) as corresponding to a symmetric problem, a fact that will be crucial for our final result.

Next we will give some definitions and quote without proof some of the results in the theory of OPS, Jacobi fractions and representation theory that will be more relevant in the resolution of our problem.

THEOREM 3.3. *Let $\mathscr{L}$ be a positive-definite moment functional and let $\mu_0 = \mathscr{L}(1)$. Let ψ_n be defined as*

$$\psi_n = \begin{cases} 0 & \text{if } x < x_{n1}, \\ A_{n1} + \cdots + A_{np} & \text{if } x_{np} \leq x < x_{n,p+1} \ (1 \leq p < n), \\ \mu_0 & \text{if } x \geq x_{nn}, \end{cases}$$

where $x_{n1} < x_{n2} < \cdots < x_{nn}$ are the zeros of $P_n(x)$ (OPS corresponding to $\mathscr{L}$), and $A_{n1}, \ldots, A_{nn}$ are positive numbers given by the Gauss quadrature formula

$$\mathscr{L}(x^k) = \mu_k = \sum_{i=1}^{n} A_{ni} x_{ni}^k, \qquad k = 0, 1, \ldots, 2n-1.$$

Then there is a subsequence in $\{\psi_n\}$ that converges on $(-\infty, +\infty)$ to a distribution function ψ which has an infinite spectrum and such that

$$\mathscr{L}(x^k) = \int_{-\infty}^{+\infty} x^k \, d\psi(x).$$

ψ is called a natural representative of $\mathscr{L}$.

From now on we will consider $\mathscr{L}$ to be positive-definite, and the associated data $\{x_{nm}\}$, $\{A_{nm}\}$, $\{\mu_n\}$, μ, μ_n defined as in the Theorem 3.3.

THEOREM 3.4. *Let $P_n(x)$ and $\lambda_1 P_n^{(1)}(x)$ be the partial denominators and numerators of a J-fraction as above, with c_n real numbers and $\lambda_n > 0$, $n \geq 1$. Let $\mathscr{L}$ be their associated moment functional. Then we have that*

$$\frac{\lambda_1 P_{n-1}^{(1)}(x)}{P_n(x)} = \sum_{k=1}^{n} \frac{A_{nk}}{x - x_{nk}} = \int_{-\infty}^{+\infty} \frac{d\psi_n(t)}{x-t}.$$

Moreover, A_{nk} can be expressed as

$$A_{nk} = \frac{\lambda_1 P_{n-1}^{(1)}(x_{nk})}{P_n'(x_{nk})}.$$

From Theorems 3.3 and 3.4 one can deduce the main result we will use later on, namely

COROLLARY 3.1. *In the conditions and notations of Theorems 5.2 and 5.3, there exists a subsequence $\{\psi_{n_k}\}$ in $\{\psi_n\}$ such that*

$$\lim_{k \to +\infty} \frac{\lambda_1 P_{n_k-1}^{(1)}(x)}{P_{n_k}(x)} = \int_{-\infty}^{+\infty} \frac{d\psi(t)}{x-t},$$

whenever x is not in the closure of the spectrum of ψ.

Next we will give a result describing the spectrum of distributions corresponding to symmetric problems. For broader information see Chihara **[3]** or Szegö **[18]**. Our notation and most of the results are stated as in Chihara's book.

THEOREM 3.5. *If a system is symmetric and $\lim_{n \to +\infty} \lambda_n = 0$ the set of limit points of the spectrum of ψ reduces to 0, and therefore the measure associated to $\mathscr{L}$ is discrete with 0 as the only possible accumulation point.*

Finally, we will quote a theorem that will be useful to actually compute the coefficients of a Taylor expansion for $C^{-2n2n}(\varepsilon)$.

THEOREM 3.6. *With reference to the recurrence formula (3.9) the following are valid for $n \geq 1$:*

(a) $\mathscr{L}(P_n^2(x)) = \lambda_1 \lambda_2 \cdots \lambda_{n+1}$, *provided that we define $\lambda_1 = \mu_0$.*

(b) $\mathscr{L}(\pi(x) P_n(x)) = 0$ *for any polynomial $\pi(x)$ of degree $m < n$, while $\mathscr{L}(\pi(x) P_n(x)) \neq 0$ if $m = n$.*

(c) $\mathscr{L}(x^n P_n(x)) = \mathscr{L}(P_n^2(x))$.

Now we are in condition to find a Taylor expansion for $C^{-2n2n}(\varepsilon)$.

B. *Taylor expansion for* $C^{-2n2n}(\varepsilon)$. Recall the recurrence problem (1)

$$G_{k+1} = XG_k - \frac{(2k-1)^2}{k(k-1)((kn)^2 - Y)(((k-1)n)^2 - Y)} G_{k-1}$$

with initial conditions

$$F_1 = 0, \quad F_2 = 1, \quad H_2 = 0, \quad H_3 = \frac{1}{2(Y - (2n)^2)}.$$

We can now assert that $F_{k+2}(X)$, $k \geq -1$, as in Proposition 3.2, are the set of monic orthogonal polynomials with respect to certain measure $d\psi_Y(X)$ and $\lambda_3^{-1} H_{k+2}$ the associated polynomials.

We also know that the associated moment functional is symmetric (since $c_n = 0$ in the recurrence). On the other hand

$$\lambda_{k+1} = \frac{(2k-1)^2}{k(k-1)((kn)^2 - Y)(((k-1)n)^2 - Y)} \to 0$$
$$\text{whenever } k \to +\infty,$$

so we can apply Theorem 3.5 to deduce that the measure associated to these orthogonal polynomials is absolutely discrete with zero as the only limit point of the spectrum of the natural representative ψ. Summarizing, one gets that, if we denote by $\mathscr{S}(\psi)$ the spectrum of ψ,

$$S(\psi) = \{z_k, -z_k, k \geq 0 | z_k \to 0 \text{ as } k \to +\infty\}$$

and $\{a_m\}$ are the weights of the corresponding measure, then $\mathscr{L}$ is defined as

$$\mathscr{L}(X^m) = \mu_m = 2\sum_{k=0}^{\infty} z_k^m a_m, \qquad m \geq 0.$$

Next, notice that Favard's theorem actually obtains a whole family of moment functionals associated to a fixed set of polynomials, one for each choice of λ_1, ($\{P_n(x)\}$ are independent of λ_1 given the initial condition $P_{-1}(x) = 0$). Due to the shift in the indices that we have, fix the value

$$\lambda_3 = \frac{1}{2(Y - (2n)^2)},$$

so that the pair (H_k, F_k) can be viewed as the kth partial numerator and denominator of the continued J-fraction

$$\frac{\lambda_3|}{|X} - \frac{\lambda_4|}{|X} - \frac{\lambda_5|}{|X} - \cdots .$$

Therefore, $\lambda_3^{-1} H_{k+2}$ are the associated polynomials with respect to problem (1). Now we can easily obtain a first expression for $C^{-2n2n}(\varepsilon)$.

If we apply Corollary 3.1 we obtain

$$\lim_{k\to+\infty} \frac{H_k(X)}{F_k(X)} = \frac{C^{-2n2n}(\varepsilon)}{\beta} = \int_{-\infty}^{+\infty} \frac{d\psi(t)}{x-t}$$
$$= \sum_{m=0}^{+\infty} \frac{a_m}{X - z_m} + \sum_{m=0}^{+\infty} \frac{a_m}{X + z_m}, \qquad X = (\varepsilon_n \varepsilon_{-n})^{\frac{-1}{2}}.$$

Observe that a_m and z_m depend on $Y = 2\varepsilon_0$ for all m.

We do not have much information about either the weights of the measure or the zeros of the polynomials. Even though this expression does not seem to be easy to compute we will give another expansion with coefficients that can be found following an easy algorithm.

Main Theorem 3.7.

$$C^{-2n2n}(\varepsilon) = \frac{2\pi i}{n} \sum_{k=1}^{\infty} \mu_{2k}(\varepsilon_0)(\varepsilon_n \varepsilon_{-n})^k,$$

where μ_k are the moments corresponding to $\mathscr{L}$. Moreover, there exists an algorithm to obtain the moments up to any desired order.

Proof. Applying the result of Theorem 3.4 one gets

$$\frac{H_k(X)}{F_k(X)} = \sum_{m=3}^{k} \frac{A_{km}}{X - x_{km}} = \sum_{m=3}^{k} \frac{H_k(x_{km})}{F'(x_{km})(X - x_{km})}, \qquad k = 3, 4, \ldots,$$

where A_{km} and x_{km} are analogous to the ones in Theorem 3.4. If we Taylor-expand the expressions as a function of $\frac{1}{X}$ we obtain

$$\frac{A_{km}}{X - x_{km}} = \sum_{l=0}^{p} A_{km} \left(\frac{x_{km}^l}{X^{l+1}}\right) + o\left(\frac{1}{X^{p+2}}\right).$$

Substituting above

$$\frac{H_k(X)}{F_k(X)} = \sum_{m=3}^{k} \frac{A_{km}}{X - x_{km}} = \sum_{m=3}^{k} \left[\sum_{l=0}^{p} A_{km} \left(\frac{x_{km}^l}{X^{l+1}}\right) + o\left(\frac{1}{X^{p+2}}\right)\right].$$

On the other hand, if ψ_k is given as in Theorem 3.3, $\mathscr{L}_k$ is the associated functional and μ_l^k is the corresponding 1-moment, then

$$\mu_l^k = \mathscr{L}_k(x^l) = \sum_{m=3}^{k} A_{km} x_{km}^l ,$$

and therefore

$$\frac{H_k(X)}{F_k(X)} = \sum_{l=0}^{p} \frac{\mu_l^k}{X^{l+1}} + o\left(\frac{1}{X^{p+2}}\right) .$$

A priori we know that the sequence converges to an analytic function on ε_n, ε_{-n}, ε_0; therefore, we can take limits without any problem and deduce the result of the theorem. Notice that $\mathscr{L}$ is symmetric and that property implies $\mu_{2k+1} = 0$ for $k \geq 0$. That is the reason to have only even powers of $X = 1/(\varepsilon_n \varepsilon_{-n})^{1/2}$ in the series above.

To finish with the proof of the theorem, we will give the algorithm to find the moments, avoiding the inconvenience of not having information about the explicit form of $d\psi_Y$.

From Theorem 3.6(a), we can deduce

$$\mu_0 = \int_{-\infty}^{+\infty} F_2^2(X)\, d\psi_Y(X) = \lambda_3 = \frac{1}{2(Y - (2n^2)^2)} ,$$

$$\mu_2 = \int_{-\infty}^{+\infty} F_3^2(X)\, d\psi_Y(X) = \lambda_3\lambda_4 = \frac{5^2}{12(Y - (2n^2)^2)(Y - (3n^2)^2)} .$$

In order to find μ_4, notice that $F_4 = XF_3 - \lambda_4 F_2 = X^2 - \lambda_4$, so $X^2 = F_4 + \lambda_4$ and therefore $X^4 = F_4^2 + \lambda_4^2 + 2\lambda_4 F_4$. Applying Theorem 3.6(b), we get zero when integrating the last term of the sum, so that

$$\mu_4 = \int_{-\infty}^{+\infty} X^4\, d\psi_Y(X) = \int_{-\infty}^{+\infty} F_4^2(X)\, d\psi_Y(X) + \int_{-\infty}^{+\infty} \lambda_4^2\, d\psi_Y(X)$$
$$= \lambda_3\lambda_4\lambda_5 + \lambda_4^2 \mu_0 .$$

In this way we can always obtain μ_{2k} in terms of μ_{2l}, $l < k$, and the integral of the square of $P_k(X)$ which value we know from Theorem 3.6(a). Repeating this process we can give the expression for moments up to any order we wish. This algorithm is not very fast since it requires us to solve for the orthogonal polynomials in the first place. For example, we obtain

$$\mu_6 = \lambda_3\lambda_4\lambda_5\lambda_6 + (\lambda_4 + \lambda_5)^2 \lambda_2 ,$$

$$\mu_8 = \lambda_3\lambda_4\lambda_5\lambda_6\lambda_7 + (\lambda_4 + \lambda_5 + \lambda_6)^2 \mu_4$$
$$- 2\lambda_4\lambda_6(\lambda_4 + \lambda_5 + \lambda_6)\mu_2 + \lambda_4^2\lambda_6^2 \mu_0 ,$$

$$\ldots$$

□

The author has a faster algorithm and a short computer program to calculate the moments. It involves Favard's path's theory (see Viennot's notes [**19**]), but we will not give further details in this paper.

Notice that, with very few adjustments, we can follow the exact same reasoning to find a Taylor expansion for any entry of the inverse matrix C^{-1}. That is, this is a general technique to find entries for the inverse of an infinite Jacobi matrix.

C. *Another interpretation for a transverse structure.* Let us look at the function $C^{-2n2n}(p)$ from another point of view. The next theorem will show us how to express transverse structures in terms of the solutions of some nonhomogeneous ordinary differential equations. The corresponding homogeneous equation is always given by the coadjoint action along Q_n.

THEOREM 3.8. *Consider the differential equation*

$$\xi''' + 2p\xi' + p'\xi = 2\cos(2n\theta), \tag{3.9}$$

with $p \in Q_n$.

There exists a periodic solution of (3.9), ξ, *whose Taylor expansion is given by* $\xi = \sum_{k=-\infty}^{+\infty} \gamma_k e^{-ik\theta}$, *with* $\gamma_{2n} = C^{-2n2n}(p)$.

Proof. Assume $\xi = \sum_{k=-\infty}^{+\infty} \gamma_k e^{-ik\theta}$. If we make a simple substitution we can observe that the action of the differential operator $-(\frac{d}{d\theta^3} + 2p\frac{\partial}{\partial\theta} + \frac{dp}{d\theta})$ on ξ is equivalent to the one of the matrix C on γ, where $\gamma = \{\gamma_k\} \in l^2$. This is true since

$$\begin{aligned}\xi''' &+ 2p\xi' + p'\xi \\ &= \sum_{k=-\infty}^{+\infty} [\gamma_k k(2\varepsilon_0(p) - k^2) + \gamma_{n+k}(2k+n)\varepsilon_{-n}(p) \\ &\qquad\qquad + \gamma_{k-n}(2k-n)\varepsilon_n(p)]e^{-ik\theta}.\end{aligned}$$

Notice at this point that the matrix C is antisymmetric. Therefore, we can solve the equation $C\gamma = b$, with b having entries all 0's except for the entry in place $-2n$, and obtain that $\gamma_{2n} = C^{-2n2n}(p)$.

But, on the other hand, to solve $C\gamma = b$ is equivalent to solving the differential equation (3.9), in the sense that the solution of $C\gamma = b$ would correspond to the Fourier coefficients of a solution of (3.9). We are done with the proof. □

Notice that we can follow the same strategy in order to find any entry of the inverse matrix for C. That is, C^{kl} *would be given by the*

lth Fourier coefficient of a periodic solution of the equation

$$\xi''' + 2p\xi' + p'\xi = 2\cos(k\theta).$$

From Proposition 3.1 we know that such a solution exists.

Finally one comment on the uniqueness problem. If we try to prove uniqueness in the same way that it is done in the finite dimensional case, we would have to try to connect two different transverse sections using the flow of a time-dependent Hamiltonian vector field. This flow would be defined on a neighbourhood of the intersection with the symplectic leaves and would automatically preserve the induced transverse structures. The existence of such a flow would automatically imply uniqueness.

In finite dimensions such a Hamiltonian vector field can always be found. In infinite dimensions we can connect two transverse sections Q_1 and Q_2 with a family of transverse sections Q_t with $1 \leq t \leq 2$. We can possibly fix the variation on the time so that the equations for the Hamiltonian operator are involutive. Nevertheless, that fact would not imply its integrability. This kind of integrability problem in infinite dimensions is quite complicated and not many results are available.

References

[1] M. Adler, *On a trace functional for formal pseudo-differential operators and the symplectic structure of the KdV*, Invent. Math., **50** (1979), 219–248.

[2] M. Antonowicz and A. P. Fordy, *Hamiltonian structure of nonlinear evolution equations*, in *Soliton Theory: A Survey of Results*, Manchester University Press, New York (1990).

[3] T. S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon & Breach, (1978).

[4] J. F. Conn, *Normal forms for analytic Poisson structures*, Ann. of Math. (2), **119** (1984), 577–601.

[5] ——, *Normal forms for smooth Poisson structures*, Ann. of Math., **221** (1985), 565–593.

[6] B. A. Dubrovin and S. P. Novikov, *Hydrodynamics of weakly deformed soliton lattices. Differential Geometry and Hamiltonian Theory*, Russian Math. Surveys, vol. 44, no. 6 (1989), 35–124.

[7] L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons*, Springer-Verlag (1987).

[8] A. A. Kirillov, *Infinite dimensional groups: their representation, orbits, invariants*, Proc. of International Congress of Math., (Helsinki, 1978).

[9] V. F. Lazutkin and T. F. Pankratova, *Normal forms and versal deformations for Hill's equations*, Funktsional. Anal. i Prilozhen., **9**, no. 4 (1976), 41–48.

[10] W. Magnus and S. Winklez, *Hill's equation*, Interscience Tracts in Pure and Applied Math., vol. 20 (1966).

[11] J. Marsden, *Applications of Global Analysis in Mathematical Physics*, Mathematics Lecture Series, 2. Publish or Perish Inc. (1974).
[12] J. Marsden and T. Ratiu, *Reduction of Poisson manifolds*, Letters in Math. Phys., **11** (1986), 161–169.
[13] J. Marsden, T. Ratiu and A. Weinstein, *Reduction and Hamiltonian structures on duals of semidirect product Lie algebras*, in *Fluids and Plasmas: Geometry and Dynamics*, Contemp. Math., vol. 28, Amer. Math. Soc., Providence, RI, 1984.
[14] J. Milnor, *Remarks on Infinite Dimensional Lie Groups. Relativity, Groups and Topology* II, Elsevier Science Publishers, (1984).
[15] V. Yu. Ovsienko and B. A. Khesin, *Symplectic leaves of the Gelfand-Dikii brackets and homotopy classes of nondegenerate curves*, Funktsional. Anal. i Prilozhen., **24**, no. 1 (1990), 38–47 or Functional Anal. Appl., **24** (1990).
[16] A. Pressley and G. Segal, *Loop Groups*, Clarendon Press, Oxford, (1986).
[17] G. Segal, *Unitary representations of some infinite dimensional groups*, Comm. Math. Phys., **80** (1981), 301–342.
[18] G. Szegö, *Orthogonal polynomials*, Amer. Math. Soc. Colloq. Publ., vol. 23, Amer. Math. Soc., Providene, RI (1967).
[19] G. Viennot, *Une Théorie Combinatoire des Polynomes Orthogonaux Généraux*, Notes de conférences données au Département de Mathématiques et d'Informatique, Université du Québec à Montréal, September 1983.
[20] A. Weinstein, *The local structure of Poisson manifolds*, J. Differential Geom., **18** (1983), 523–557.
[21] E. Witten, *Coadjoint orbits of the Virasoro group*, Comm. Math. Phys., **114**, (1988), 1–53.

Received March 30, 1992.

UNIVERSITY OF WISCONSIN
MADISON, WI 53706

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

ON AMBIENTAL BORDISM

CARLOS BIASI

Let M^m be a closed and oriented submanifold of a closed or oriented manifold N^n, such that $[M, i] = 0 \in \Omega_m(N)$, where $i: M \to N$ is the inclusion and $\Omega_m(N)$ is the mth oriented bordism group of N. If $n = m+2$ or $m \leq 3$ or $m \leq 4$ and $n \neq 7$ then M bounds in N.

Introduction. Let us consider M^m a closed submanifold of N^n. In this paper, we study the possibility that there exists submanifold $W^{m+1} \subset N^n$ such that $\partial W = M$. If $M = S^m$ and $N = S^{m+2}$, such that a submanifold W is called a Seifert surface knot S^m. In **[5]**, Sato showed that every connected closed and oriented submanifold M^m of S^{m+2} is a boundary of an oriented surface of S^{m+2}.

In **[4]**, Hirsch studies the following problem: If a compact connected and oriented manifold M^m bounds, does there exist embedding from M^m into $\mathbb{R}^n$ which is a boundary in $\mathbb{R}^n$?

The answer is yes, if $n \geq 2m$.

The difference between the two problems is that, in our case, the embedding from M into N is fixed.

There is an obvious necessary condition for the existence of W, when M and N are oriented manifolds.

Let $\Omega_m(N)$ be the mth oriented bordism group of N **[2]**. If $i: M \to N$ is the inclusion map, we can define an element $[M, i]$ in $\Omega_m(N)$ and see that $[M, i] = 0$ if M bounds in N.

Generally, the converse in not true, but sometimes the vanishing of $[M, i]$ guarantees the existence of W, for example if the codimension $n - m$ is large.

We prove the following theorem.

THEOREM 5.2. *Let us suppose that $M^m \subset N^n$, $n > m+1$, is such that $[M, i] = 0$ in $\Omega_m(N)$. Then M bounds in N if one of the following conditions occurs:*

(a) $n = m + 2$,
(b) $m \leq 3$,
(c) $m \leq 4$ *and* $n \neq 7$.

In his Doctoral thesis [1] the author proved that, when $n = 2m + 1$, and M and N are closed and oriented, a submanifold $M \subset N$ bounds in N if, and only if, $[M, i] = 0 \in \Omega_m(N)$.

1. A more general problem of ambiental bordism. Let

$$G \subset O(n-m-1), \qquad n > m+1,$$

be a closed transformation group and let $\gamma_G \to BG$ be the classifying fiber bundle of $(n-m-1)$-vector bundles which have a G-structure.

Let us consider MG the Thom space of γ_G. We have:

$$\pi_i(MG) = \begin{cases} 0, & i < n-m-1, \\ \mathbb{Z}, & i = n-m-1 \text{ and } G \subset \mathrm{SO}(n-m-1), \\ \mathbb{Z}_2, & i = n-m-1 \text{ and } G \not\subset \mathrm{SO}(n-m-1). \end{cases}$$

Let us consider now N^n to be a closed connected manifold which we assume to be oriented if $G \subset \mathrm{SO}(n-m-1)$. (If $G \not\subset \mathrm{SO}(n-m-1)$ we drop the orientability hypothesis.)

Let $M^m \subset N^n$ be a closed submanifold and let us suppose that the normal fiber bundle ν_M of M in N has a cross section s, nowhere zero, such that $\nu_M = \{s\} \oplus \xi$, where $\{s\}$ is a subbundle generated by s and ξ is a $(n-m-1)$-vector bundle endowed with a G-structure.

We shall say that a submanifold $W \subset N$ satisfies condition $(*)$ if it has the properties:

(i) $\partial W = M$ and s is the inward-pointing vector field on ∂W.

(ii) the normal fiber bundle ν_W has a G-structure which agrees with the given G-structure of ξ over M. (Observe that $\xi = \nu_W|M$.)

2. Primary obstruction to the existence of W. Let V be a closed tabular neighborhood of M in N, $A = \partial W$ and $X = N - \overset{\circ}{V}$. We can think s a function $s: M \to A$. Then $s(M)$ is a submanifold of A, whose normal fiber bundle is isomorphic to ξ. By the Thom construction there exists a function $f: A \to MG$ such that, if ∞ is the point at infinity to MG, then f is differentiable on $A - f^{-1}(\infty)$, f is transversal to BG and $f^{-1}(BG) = (M)$ [6].

We shall take $\pi_{m-n-1}(MG)$ as the cohomology coefficient group. Let $e \in H^{n-m-1}(MG)$ be the fundamental class of the space MG. We know that $f^*(e) = \alpha$, where α is the dual class of $s_*(\mu_M)$ and μ_M is the fundamental class of M.

If $f: A \to MG$ extends to a map $\overline{f}: X \to MG$, then we can suppose, up to homotopy, that $\overline{f}$ is differentiable in $X - \overline{f}^{-1}(\infty)$ and that $\overline{f}$ is transversal to BG. Taking $W_1 = \overline{f}^{-1}(BG)$ we obtain a submanifold of X whose boundary is $s(M)$.

Let us observe that this submanifold can be extended to a submanifold W which satisfies condition $(*)$.

We conclude then that there exists W, satisfying $(*)$, if and only if f extends to X.

The class $\delta f^*(e)$ is the obstruction to the extension of f to the $(n-m)$-skeleton of X, where $\delta\colon H^{n-m-1}(A) \to H^{n-m}(X, A)$ is the coboundary operator.

Consider the commutative diagram:

$$\begin{array}{ccc} H^{n-m-1}(A) & \xrightarrow{\delta} & H^{n-m}(X, A) \\ \downarrow D & & \downarrow D \\ H_m(A) & \xrightarrow{s_*} & H_m(X) \cong H_m(N-M). \end{array}$$

We conclude that the primary obstruction to the extension of f, up to duality, is the element $s_*(\mu_M) \in H_m(N-M)$ (regarding s as function from M into $N-M$).

Hence, we have:

PROPOSITION 2.1. *f extended to the $(n-m)$-skeleton of X if, and only if, $s_*(\mu_M) = 0$ in $H_m(N-M)$.*

Assuming that $s_*(\mu_M) = 0$, let us consider two cases:

1. $G = O(n-m-1)$.

Here, f extends up to the $(n-m+1)$-skeleton of X, because $\pi_{n-m}(MG) = 0$ and, if $n-m=2$, then f extends to all of X since $MO(1)$ is a $K(\mathbb{Z}_2, 1)$ space.

2. $G = \mathrm{SO}(n-m-1)$.

Since $\pi_{n-m+i}(MG) = 0$, $i = 0, 1, 2$, f extends up to the $(n-m+3)$-skeleton of X. Hence, if $\dim M \leq 3$, f extends.

On the other hand, if $n-m = 2$ or 3 then MG is a $K(\mathbb{Z}, 1)$ or $K(\mathbb{Z}, 2)$, respectively. In any case, f extends globally.

3. Oriented ambiental bordism. From now on, all manifolds and submanifolds will be considered to be oriented.

THEOREM 3.1. *Let us suppose that:*

(a) $H_j(X) = 0$, $0 < j < m-3$.

(b) *The canonical homomorphism* $\pi_{n-1}(\mathrm{MSO}(n-m-1)) \xrightarrow{\varphi} \Omega_m$ *is injective.*

There exists W satisfying $(*)$ if, and only if, $s_*(\mu_M) = 0 \in H_m(X)$ and M is a boundary.

Proof. Let us use the notation $\pi_i = \pi_i(\mathrm{MSO}(n-m-1))$. If $s_*(\mu_M) = 0$, then f extends to the $(n-m)$-skeleton of X.

From hypothesis (a) and Lefschetz duality, we conclude that

$$H^j(X, A, \pi_{j-1}) = 0, \qquad n-m < j < n.$$

Let D be an open disk of $X - A$. Since X is orientable, $H^j(X-D, A, \pi_{j-1}) \cong H^j(X, A, \pi_{j-1}) = 0$, $n-m < j < n$. Hence, there exists an extension $\overline{f}\colon X - D \to Y$ of $f\colon A \to Y$, where $Y = \mathrm{MSO}(n-m-1)$.

Let us consider $S = \partial D$ and $h = \overline{f}|\partial D\colon S \to Y$. We may suppose that h is transversal to $\mathrm{BSO}(n-m-1)$ and let

$$M^m = h^{-1}(\mathrm{BSO}(n-m-1)).$$

Consider $\overline{W} = \overline{f}^{-1}(\mathrm{BSO}(n-m-1))$, a bordism between M_1 and $s(M)$. Since $s(M)$ is a boundary, M_1 also is.

We have also that $\psi([h]) = [M_1] = 0$ and since ψ is a monomorphism, h is homotopic to a constant map and so h extends over D.

The converse is straightforward. □

4. On the existence of normal vector fields homologous to zero in $N - M$. In the next section, we show that in certain situations it is possible to obtain a cross-section $s\colon M \to S(\nu_M)$ such that $s_*(\mu_M) = 0 \in H_m(N-M)$, where $S(\nu_M) \to M$ is the normal sphere bundle of M in N.

PROPOSITION 4.1. *The Euler class of the normal bundle of M^m in N^n is zero if and only if $i_*(\mu_M) \subset \operatorname{im} j_*$, where μ_M is the fundamental class of M and $i\colon M \to N$, $j\colon N - M \to N$ are inclusion maps.*

Proof. Let us consider $e \in H^{n-m}(M, \mathbb{Z})$, the Euler class of the normal bundle ν_M, and let $D_A\colon H^{n-m}(M : \mathbb{Z}) \to H_m(N, N-M; \mathbb{Z})$ be the Alexander duality. We have that $D_A(e) = \alpha_*(\mu_M)$ where α_* is induced by the inclusion map $\alpha\colon (N, N-M)$.

Using the exact sequence of pair $(N, N-M)$ it follows that $\alpha_*(\mu_M) = 0$ if, and only if, $i_*(\mu_M) \subset \operatorname{im} j_*$. □

COROLLARY 4.2. *If $M^m \subset N^n$ is homologous to zero, $n - m = 2$ or $n \geq 2m$, then M has a normal vector field that is nowhere zero.*

Proof. By Proposition 4.1 the Euler class of ν_M is zero. Then there is a nowhere zero normal vector field on the $(n-m)$-skeleton of M, which can be extended to all M, because $n-m \geq m$ or $\pi_i(R^2-0)=0$, $i>1$ in the case $n-m=2$. □

Let $\pi\colon E \to M^m$ be a differentiable SO$(n+1)$-bundle with fiber S^n and base M^m (and oriented manifold).

If $s\colon M \to E$ is a cross-section, let θ_s be the Poincaré dual to $\bar{s}_*(\mu_M)$, where $\bar{s}=-s$ is the opposite cross-section to s.

Having fixed a cross-section $s_0\colon M \to E$, the following diagrams are commutative:

$$\begin{array}{ccccc}
[M, E] & & & & \\
\big\downarrow \xi & \searrow^{\varphi} & & & \\
H^n(M) & \xrightarrow{\pi^*} & H^n(E) & & \\
\big\downarrow D & & \big\downarrow D & & \\
H_{m-n}(M) & \xrightarrow{\Delta} & H_m(E) & \xrightarrow{\pi_*} & H_m(M)
\end{array}$$

where $[M, E]$ is the set of homotopy classes of cross-sections, $\xi([s])=\bar{s}^*(\theta_{\bar{s}_0})$; $\varphi([s])=\theta_{\bar{s}_0}-\theta_{\bar{s}}$, is Poincaré duality and last line is a portion of the generalized Gysin sequence.

We define $\psi\colon [M, E] \to H_m(E)$ by $\psi([s])=s_{s_*}(\mu_M)-s_*(\mu_M)$ and observe that $\psi = D \circ \psi$.

If $m \leq n+1$ or $n=1$, then the function ξ is onto and so the image of ψ is the kernel of π_*.

This fact will be applied in the proof of Proposition 4.3 below, where the fiber bundle to be considered is $S(\nu_M) \to M$.

PROPOSITION 4.3. *Let $M^m \subset N^n$, $n=m+2$ or $n \geq 2m$, be an oriented submanifold homologous to zero in an oriented manifold N. Then there exists a cross-section $r\colon M \to S(\nu_M)$ such that its image is homologous to zero in $H_m(N-m)$.*

Proof. Let $s_0\colon M \to S(\nu_M)$ be a cross-section that is nowhere zero (Corollary 4.2) and let us consider the commutative diagrams:

$$\begin{array}{ccccc}
 & \nearrow^{s_{0_*}} & H_m(S(\nu_M)) & \xrightarrow{\pi_*} & H_m(M) \\
H_m(M) & & \big\downarrow l_* & & \big\downarrow i_* \\
 & \searrow_{s_*} & H_m(N-M) & \xrightarrow{j_*} & H_m(N)
\end{array}$$

where $s_* = l_*(s_{0_*})$ and l_* is induced by the inclusion $S(\nu_M) \to (N-M)$.

We have $j_* s_*(\mu_M) = i_* \pi_* s_{0_*}(\mu_M) = 0$ implying that $s_*(\mu_M)$ belongs to the kernel of j_* which is the image of $\partial : H_{m+1}(N, N-M) \to H_m(N-M)$.

Let us consider the following commutative diagram:

$$\begin{array}{ccc} H_{m+1}(D(\nu_M), S(\nu_M)) & \xrightarrow{\partial} & H_m(S(\nu_M)) \\ \big\downarrow exc & & \big\downarrow j_* \\ H_{m=1}(N, N-M) & \xrightarrow{\partial} & H_m(N-M). \end{array}$$

It follows that there exists an element $\mu \in H_m(S(\nu_M))$ such that $\mu \in \operatorname{Ker} \pi_*$ and $j_* = s_*(\mu_M)$.

Since $\operatorname{im} \psi = \ker \pi_*$, there exists a cross-section $r: M \to S(\nu_M)$ such that $\psi([r]) = \mu$.

But $\psi([r]) = s_{0_*}(\mu_M) \to r_*(\mu_M)$ so $j_* r_*(\mu_M) = 0$ in $H_m(N-M)$. Hence, the image of $r: M \to S(\nu_M)$ is homologous to zero in $N-M$.

5. A theorem on ambiental bordism. Let us consider $\Omega_j(N)$ to be the jth bordism group of N.

If $H_j(N) = 0$, $0 < j < m-3$, it is possible using the bordism spectual sequence [2] to show that the function $\Omega_m(N) \to H_m(N) \oplus \Omega_m$, which associates to each pair $[M, f]$ the element $\mu([M, f]) + [M]$, is an isomorphism, where μ is the canonical homomorphism.

In the proof of Theorem 5.2, we are going to use the following lemma, which has been proved in [1] (the proof, if $q > m$, is due to Thom [6]).

LEMMA 5.1. *The homomorphism* $\varphi: \pi_{q+m}(\mathrm{MSO}(q)) \to \Omega_m$, $q \geq m$, *is an isomorphism.*

THEOREM 5.2. *Let us suppose* $M^m \subset N^n$, $n > m+1$, *is such that* $[M, i] = 0$ *in* $\Omega_m(N)$. *Then* M *bounds in* N *if one of the following conditions occurs:*

(a) $n = m+2$,

(b) $m \leq 3$,

(c) $m \leq 4$ *and* $n \neq 7$.

Proof. Any one of the conditions (a), (b) and (c), based on previous results, imply that normal bundle ν_M has a cross-section nowhere zero such that, considering s as a function from M into $N-M$, $s_*(\mu_M) = 0 \in H_m(N-M)$.

If (a) or (b) occurs, the theorem follows from case 2, already discussed in §2

If $n = 4$ and $n \geq 8$, we apply Theorem 3.1.

REMARK 1. If $n = m + 2$ or $m \leq 3$, then $[M, i] = 0 \in \Omega_m(N)$ if, and only if, M is homologous to zero in N.

REMARK 2. When $m = 4$ and $n \neq 7$, although we shall prove that $[M, i] = 0$ implies the existence of a normal section nowhere zero (Th. 5.3) we are not able to prove that there exists a normal vector field homologous to zero in $N - M$, which in this case would be sufficient to prove the conclusion of Theorem 5.2.

THEOREM 5.3. *Let us suppose* $M^4 \subset N^7$. *If* $[M, i] = 0$ *in* $\Omega_4(N)$ *then* ν_M *has a cross-section which is nowhere zero.*

Proof. There exists $W \subset N \times I$ such that $\partial W = M \times 0 \subset N \times I$ [**1**].

Let ν_W and ν_M be the normal fiber bundles of W in $N \times I$ and of M in N, respectively. We can also suppose that $\nu_W | M \times 0 = \nu_M$.

Let us consider $\overline{W} \subset N \times \mathbb{R}$ to be the double of W and let $i: \overline{W} \to N \times \mathbb{R}$ and $j: N \times \mathbb{R} \to \overline{W} \to N \times \mathbb{R}$ be inclusion maps.

Since $i_*(\mu_{\overline{W}}) \subset \operatorname{im} j_*$, then $\overline{W}$ has a normal vector field which is nowhere zero in $N \times \mathbb{R}$ up to the 3-skeleton of $\overline{W}$.

Hence, there exists a 2-dimensional oriented vector bundle ξ over M such that $\nu_M | M^{(3)} = \xi \otimes \mathscr{E}^1$.

Let us consider e to be the Euler class of ξ in $H^2(M^{(3)})$ and let $\overline{e} \in H^2(M)$ be such that $io^*(\overline{e}) = e$, where $i: M^{(3)} \to M$ is the inclusion map.

Let $\overline{\xi}$ be a 2-dimensional vector bundle over M such that its Euler class is $\overline{e}$. Let us observe that $\overline{\xi} | M^{(3)} = \xi$.

Let f, $g: M \to \mathrm{BSO}(3)$ be classifying maps $\overline{\xi} \oplus \mathscr{E}^1$ and ν_M, respectively.

Since the Euler classes of $\overline{\xi} \oplus \mathscr{E}^1$ and of ν_M are equal, then their second Stiefel-Whitney classes are equal.

Let $\tilde{p}_1$ be the Pontryagin class of the classifying fiber bundle $\tilde{\gamma} \to \mathrm{BSO}(3)$ and let $\tilde{e}$ be the Euler class of $\tilde{\gamma}$. Since $f^*(\tilde{p}_1) = g^*(\tilde{p}_1)$. Hence, the vector bundles $\xi \oplus \mathscr{E}^1$ and ν_M are equivalent [**3**]. □

References

[1] C. Biasi, *L-equivalent and ambiental bordism of submanifolds*, (Doctoral Thesis), ICMSC-USP, São Carlos (1981).

[2] P. E. Conner and E. E. Floyd, *Differentiable Periodic Maps*, Springer-Verlag (1964).

[3] A. Dold and H. Whitney, *Classification of oriented sphere bundles over a 4-complex*, Ann. of Math., **69** (1959), 667–677.

[4] M. Hirsch, *On embedding of bounding manifolds in euclidean space*, Ann. of Math., (2), **74** (1961), 494–497.

[5] N. Sato, *Cobordism of semiboundary links*, Topology Appl., **18** (1984), 225–234.

[6] R. Thom, *Quelques propriétés globales de variétés différentiable*, Comment. Math. Helv., **28** (1954), 17–86.

Received November 13, 1991 and in revised form November 19, 1992.

ICMSC-USP
Departamento de Matemática
Rua Dr. Carlos Botelho, 1465
Caixa Postal 668
13560-970-São Carlos-SP-Brasil

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

NONRIGID CONSTRUCTIONS IN GALOIS THEORY

PIERRE DEBES AND MICHAEL D. FRIED

The context for this paper is the *Inverse Galois Problem*. First we give an if and only if condition that a finite group is the group of a Galois regular extension of $\mathbb{R}(X)$ with only real branch points. It is that the group is generated by elements of order 2 (Theorem 1.1 (a)). We use previous work on the action of the complex conjugation on covers of $\mathbb{P}^1$. We also show each finite group is the Galois group of a Galois regular extension of $\mathbb{Q}^{tr}(X)$. Here $\mathbb{Q}^{tr}$ is the field of all totally real algebraic numbers (Theorem 5.7). Sections 1, 2, and 3 discuss consequences, generalizations, and related questions.

The second part of the paper, §4 and §5, concerns descent of fields of definition from $\mathbb{R}$ to $\mathbb{Q}$. Use of Hurwitz families reduces the problem to finding $\mathbb{Q}$-rational point on a special algebraic variety. Our first application considers realizing the symmetric group S_m as the group of a Galois extension of $\mathbb{Q}(X)$, regular over $\mathbb{Q}$, satisfying two further conditions. These are that the extension has four branch points, and it also has some totally real residue class field specializations. Such extensions exist for $m = 4, 5, 6, 7, 10$ (Theorem 4.11).

Suppose that m is a prime larger than 7. Theorem 5.1 shows that the dihedral group D_m of order $2m$ is not the group of a Galois regular extension of $\mathbb{Q}(X)$ with fewer than 6 branch points. The proof interprets realization of certain dihedral group covers as corresponding to rational points on *modular curves*. We then apply Mazur's Theorem. New results of Kamienny and Mazur suggest that no bound on the number of branch points will allow realization of all D_m s.

0.1. Description of Theorem 1.1. Throughout, $\mathbb{C}$ denotes the complex number field, X an indeterminate, and $\overline{\mathbb{C}(X)}$ a fixed algebraic closure of $\mathbb{C}(X)$. Let k be a subfield of $\mathbb{C}$. We say a finite extension $Y/k(X)$ with $\overline{\mathbb{C}(X)} \supset Y$ is regular over k if $\overline{k} \cap Y = k$. Equivalently $[Y: k(X)] = [Y\mathbb{C}: \mathbb{C}(X)]$. Denote this degree by n. Regard the degree n field extension $Y\mathbb{C}/\mathbb{C}(X)$ as the function field extension of a degree n cover $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$. Here $\mathbb{P}^1$ is the complex projective line and $Y_{\mathbb{C}}$ is an irreducible non-singular curve.

The map φ is ramified over a finite number of points $x_1, \dots, x_r$. We call these the *branch points* of the cover (or of the extension $Y/k(X)$). Our first result (Theorem 1.1(a)) shows exactly when a finite group G is the group of a Galois regular extension of $\mathbb{R}(X)$ with only real branch points.

This happens if and only if G is generated by involutions.

Theorem 1.1 uses formulas for the action of complex conjugation on the fundamental group of $\mathbb{P}^1\setminus\{x_1, \dots, x_r\}$ (cf. §2.3). Hurwitz **[Hur]** knew these. Krull and Neukirch investigated them further **[KN]**. Still, no one has exploited this simple statement about groups generated by involutions.

0.2. *Relations with the inverse Galois problem.* Here is a weak version of the Inverse Galois problem. Does each group occur as the Galois group of a field extension of $\mathbb{Q}$? As do others, we approach this through its geometric analog. That is, we consider it over $\mathbb{Q}(X)$ rather than $\mathbb{Q}$. This is a descent problem. Suppose we are given a group G, a suitably large integer r, and r points $x_1, \dots, x_r \in \mathbb{P}^1(\mathbb{C})$. Topology then constructs a Galois extension of $\mathbb{C}(X)$ with Galois group G and branch points $x_1, \dots, x_r$. One must then restrict the scalars from $\mathbb{C}$ to $\mathbb{Q}$. Theorem 1.1 gives a form of descent from $\mathbb{C}$ to $\mathbb{R}$. Proposition 2.3 and Comment 3 of §3.5 refine these for specific applications (see §0.4).

We stress the condition on the branch points. Theorem 1.1 (a) shows that Galois groups occur over $\mathbb{Q}$ (or even $\mathbb{R}$) using r branch points in $\mathbb{P}^1(\mathbb{R})$ only if r elements of order 2 generate G. Therefore, in practice, classical "rigidity" **[Se3**; Theorem 9.1] realizes only groups over $\mathbb{Q}(X)$ that are generated by 3 elements of order 2.

Corollary 1.2 is another consequence of Theorem 1.1 (a). Each finite group has a *totally nonsplit cover* (cf. §1.2) that is not the Galois group of a regular extension of $\mathbb{R}(X)$ with only real branch points. Nevertheless, every finite group is the Galois group of a regular extension of $\mathbb{R}(X)$, with branch points consisting of complex conjugate pairs ([**Se3**; Ex. p. 107] or Theorem 3.1). Theorem 5.7 notes that each finite group is the Galois group of a regular extension of $\mathbb{Q}^{\mathrm{tr}}(X)$. Here $\mathbb{Q}^{\mathrm{tr}}$ is the field of all totally real algebraic numbers.

0.3. *Extension of Theorem* 1.1. Theorem 1.1 (b) applies to not necessarily Galois extensions. Finite group G is the monodromy group of a cover $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ defined over $\mathbb{R}$ with only real branch points if and only if

(*) G has an automorphism h and a system of generators $\alpha_1, \dots, \alpha_s$ such that $h(\alpha_i) = \alpha_i^{-1}$ for $i = 1, \dots, s$.

Of course, (*) holds if G is generated by elements of order 2. Sections 1.2–1.5 have a more complete discussion on (*) and related conditions. In particular, we discuss the *persistence* of property $(*)$. Given a group G satisfying $(*)$, when does there exist a totally *nonsplit* cover of G that does not satisfy $(*)$ (§1.5).

Notation and basic tools appear in §2. Classical identifications in the theory of covers appear in §2.1 and §2.2. Skip these on a first reading. Sections 3.1–§3.4 give the proof of Theorem 1.1. The final descent argument for the constructive part ($\Leftarrow$) uses Weil's general criterion. This says that the *field of moduli* (§2.4) K of a cover is a field of definition if a certain cocycle condition holds. We add an observation to a result of Coombes and Harbater **[CoH]** for Galois covers (Theorem 2.4 (ii)). Thus, K is also a field of definition for the G-cover; the cover and its automorphisms can be defined over K.

This method is natural, but perhaps intricate. Serre suggested simplifying this using the algebraic fundamental group rather than the classical topological fundamental group. Section 3.6 gives a second proof of Theorem 1.1 (a) following Serre's viewpoint. This is constructive. Assume we have a group G and generators of G with property $(*)$. We give an explicit description, in terms of "branch cycles," of a cover $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ that has the properties stated in Theorem 1.1 (b). Furthermore, we can force this cover to have some fibers of only *real points*.

0.4. *Enhanced applications.* The topological action of complex conjugation c induces its arithmetic action. (Section 3.7 has a precise formulation.) We note that no naive p-adic analog of this representation of complex conjugation holds for the Frobenius $F_p \in G(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ (§3.7).

Comment 3 in §3.5 answers a question of E. Dew in his thesis **[D]**. In so doing it refines the technique of descending from $\mathbb{C}$ to $\mathbb{R}$. Consider the field of moduli K of a G-cover when K is a number field. How can we effectively decide if each completion of K is a field of definition of the G-cover? We give iff conditions for the field of moduli, on one hand, *and* the field of definition, on the other, of a G-cover to be (in) $\mathbb{R}$. Dew has started an investigation of a local-global question for the field of moduli being a field of definition. Knowing the answer over each local place (including infinite places) does not answer the global question.

Descent to $\mathbb{Q}$ appears in §4. We consider $G = S_m$ and specific choices of 3 generators of order 2. Then, we investigate if certain 4 branch point covers $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ derived from Theorem 1.1 can be defined over $\mathbb{Q}$. "Rigidity assumptions" from **[Se3**; Ch. 8, 9**]** do not apply. They rarely do when there are 4 (or more) branch points.

In §4.1 and §4.2, we recall from **[Fr1]** how to handle nonrigid cases. When $r = 4$, Hurwitz family ideas reduce the problem to finding a rational point on a certain curve $C(\mathbb{C})$. Section 4.3 gives a formula for the genus of $C(\mathbb{C})$. We can answer our original question about S_m when the curve $C(\mathbb{C})$ has genus 0. Our computation shows this happens exactly when $m = 4, 5, 6, 7, 10$. So, for these values of m, we realize the symmetric group S_m as the Galois group of a regular extension of $\mathbb{Q}(X)$ with 4 branch points and with some totally real residue class specializations (Theorem 4.11). Serre noted, with 3 branch points instead of 4, only one centerless group, $G = S_3$, had the same property **[Se2]**.

We do not know how to improve on our sporadic 3 generator cases to draw the conclusion of Theorem 4.11 for an infinite number of groups. Descent from $\mathbb{R}$ to $\mathbb{Q}$ is the difficulty because we must find rational points on low dimensional Hurwitz spaces. Even with easy groups this is a difficult obstruction. For example, the dihedral group D_m of order $2m$ is generated by 2 elements of order 2.

Consider a prime $m > 7$. Theorem 5.1 shows that D_m requires covers with at least 6 branch points to be realized as the Galois group of a regular extension of $\mathbb{Q}(X)$. Mazur has formulated conjectures that imply that realization of all D_m s will require an unbounded number of branch points **[KM]**. We borrow some of his formulation from an e-mail discussion with him.

0.5. **Acknowledgments**. David Harbater made expositional simplifications in our proof on Comment 3—Dew's question—in §3. In addition, much of the proof of Theorem 2.4 (§2.4) is implicit in the result in **[CH]**. Our concern is with Property (ii) which was not stated there.

1. First results and consequences. Let $Y/K(X)$ be a regular extension of degree n and $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ the associated cover. That is, $Y_{\mathbb{C}}$ is the set of places of the field $Y\mathbb{C}$ and φ is the natural restriction of places—points of $\mathbb{P}^1$—to $\mathbb{C}(X)$. Branch points $x_1, \dots, x_r$ are the places ramified in the extension $Y\mathbb{C}/\mathbb{C}(X)$.

1.1. *Statement of Theorem* 1.1. Let x_0 be a point in $\mathbb{P}^1(\mathbb{R})\setminus \{x_1, \dots, x_r\}$. Denote the fundamental group

$$\pi - 1(\mathbb{P}^1\setminus\{x_1, \dots, x_r\}, x_0)$$

for short by π_1. There is a natural action T of π_1 called the *monodromy action* on the points of the fiber $\varphi^{-1}(x_0)$. For its description, start with $[\gamma]$, the homotopy class of a closed path based at x_0. Then $T([\gamma])$ permutes $\varphi^{-1}(x_0)$; it maps $y \in \varphi^{-1}(x_0)$ to $T([\gamma])(y)$, the terminal point of the unique lift of γ through φ with initial point y.

The permutation $T([\gamma])$ is independent of the representative of $[\gamma]$. Fix a labeling $y_1, \dots, y_n$ of the points of the fiber $\varphi^{-1}(x_0)$. Regard T as an action $T\colon \pi_1 \to S_n$ of π_1 on the integers $1, \dots, n$. Up to conjugation by an element of S_n, this action does not depend on labeling the y_is or on the base point x_0. Call the group $T(\pi_1)$ *the monodromy group* of the cover. This defines a subgroup of S_n up to conjugation by elements of S_n.

THEOREM 1.1. (a) *Consider a finite group G. It is the group of a regular Galois extension of $\mathbb{R}(X)$ with only real branch points exactly when*

(1.1) $\qquad$ *G is generated involutions.*

(b) *Furthermore, G is the monodromy group of a cover $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ defined over $\mathbb{R}$ with only real branch points if and only if*

$$(1.2) \qquad \exists h \in \mathrm{Aut}(G),\ \exists \alpha_1, \dots, \alpha_s \in G| \\ \langle \alpha_1, \dots, \alpha_s \rangle = G,\ h(\alpha_i) = \alpha_i^{-1}, \qquad i = 1, \dots, s.$$

Addition to Theorem 1.1 (a). We can take the number of generating involutions of G equal to the number of branch points of the regular Galois extension of $\mathbb{R}(X)$ in the statement.

Addition to Theorem 1.1 (b). The cover $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ defined over $\mathbb{R}$ produced in §3.3 for the only if part of (b) has *branch cycles*

$$(\alpha_1, \alpha_1^{-1}\alpha_2, \dots, \alpha_{s-1}^{-1}\alpha_s, \alpha_s^{-1})$$

(cf. §2.3). It is Galois over $\mathbb{C}$. Indeed, it is Galois over $\mathbb{R}$ if h is induced by conjugation by $h' \in G$ with h' of order 2.

1.2. *Group theoretical conditions.* As noted, (1.1) $\Rightarrow$ (1.2). The converse is false: abelian groups distinct from $(\mathbb{Z}/2)^m$ satisfy (1.2) but not (1.1). For example, the cyclic group $\mathbb{Z}/m$ is the monodromy group of the Galois cover $\varphi\colon \mathbb{P}^1 \to \mathbb{P}^1$ given by $\varphi(y) = y^m$. For $m \neq 2$, it is defined over $\mathbb{R}$ with only real branch points. Yet, the corresponding function field extension $\mathbb{R}(y)/R(y^m)$ is not Galois.

Consider two further conditions.

(1.3) G is a subgroup of G' with $[G'\colon G] = 2$, and G' is generated by involutions in $G' \backslash G$.

Further: If $h' \in G$ of order 2 induces h, then G is generated by involutions.

(1.4) $G = \mathbb{Z}/2$ or $\mathrm{Aut}(G)$ is of even order.

We now show $(1.1) \Rightarrow (1.2) \Leftrightarrow (1.3) \Rightarrow (1.4)$.

$(1.2) \Rightarrow (1.3)$. Define G' to be $G' = G \times {}^s\langle h\rangle$ the semi-direct product of G and the group generated by the automorphism h. The elements (α_i, h), $i = 1, \dots, s$, and h generate G' and they are of order 2:

$$(\alpha_i, h)(\alpha_i, h) = (\alpha_i h(\alpha_i), h^2) = (\alpha_i \alpha_i^{-1}, 1) = 1.$$

Also, $(\alpha_i, h) \in G' \backslash G$. Suppose h is represented by inner automorphism by an element $h' \in G$ with h' of order 2. Then G is generated by involutions; include h' with $\alpha_i h'$, $i = 1, \dots, s$.

$(1.2) \Leftarrow (1.3)$. Consider the situation where $g_0, g_1, \dots, g_r$ are involutions in $G' \backslash G$ that generate G'. Then, $\beta_i = g_0 g_i$, $i = 1, \dots, r$, are in G. Clearly, g_0 conjugates them to their inverses: $g_0(g_0 g_i)g_0 = g_i g_0 = (g_0 g_i)^{-1}$. We have only to check if they generate G.

Take H to be the subgroup that the β_i s generate. We show G is the union of the cosets of H and $g_0 H$ to conclude the proof. Do an induction on elements of G presented as words $g_{i_1} \cdots g_{i_t}$ in the g_i s. Assume words of length at most $t-1$ are in one of the cosets H or $g_0 H$. Now do cases for $g_{i_2} \cdots g_{i_t} = \sigma$ in H or $g_0 H$. If $\sigma \in H$, then $g_0 g_{i_1} \sigma$ is also in H. Multiply by g_0 to see $g_{i_1} \sigma \in g_0 H$. On the other hand, if $\sigma \in g_0 H$, then multiply by $(g_{i_1} g_0) g_0$ to get $g_{i_1} \sigma$ in H. We're done.

$(1.3) \Rightarrow (1.4)$ Suppose G' contains τ of order 2 not in the centralizer $\mathrm{Cen}_{G'}(G)$ in G'. Then, conjugation by τ is an automorphism of G of order 2. Thus, $|\mathrm{Aut}(G)|$ is even. Assume all elements of G' of order 2 are in $\mathrm{Cen}_{G'}(G)$. Pick an element a of order 2 from $G' \backslash G$. Then $a \in \mathrm{Cen}_{G'}(G)$. Therefore, G' is the direct product $G \times \langle a\rangle$ and involutions—au with u running over involutions of $G' \backslash G$—generate G. Since those generators of G are also in $\mathrm{Cen}_{G'}(G)$, the group G is abelian. Conclude: $|\mathrm{Aut}(G)|$ is even unless $G = \mathbb{Z}/2$. □

So, groups distinct from $\mathbb{Z}/2$, with odd order automorphism group, are not monodromy groups of a cover over $\mathbb{R}$ with only real branch points. Here is how to get such a group. Consider a p-group P with p odd. Then, $\mathrm{Aut}(P)$ acts on the *frattini quotient module* $P/[P, P]P^p$ with kernel a p-group [**Hu**; Satz 3.17, p. 274]. There exists P with any desired nontrivial representation occurs in the frattini quotient [**BK**; Th. 1]. In particular, choose P so that its automorphism group is odd.

1.3. *A corollary of Theorem* 1.1. Recall that a *cover* of a group G is a surjective homomorphism $\psi: F \to G$. The cover is finite if F is a finite group. It is *totally nonsplit* if F has no proper subgroup that maps surjectively to G. This is equivalent to the condition for a *frattini cover* as after Lemma 1.3 below. The *frattini subgroup* of a group H is the intersection of all the maximal proper open subgroups of H.

COROLLARY 1.2. *Let G be any finite group. Then there is a totally nonsplit finite cover $\psi: F \to G$ of G where F is not the group of a regular Galois extension of $\mathbb{R}(X)$ with only real branch points.*

Corollary 1.2 follows from Theorem 1.1 (a) and this lemma.

LEMMA 1.3. *Let G be a finite group. There is a totally nonsplit finite cover $\psi: F \to G$ of G where F is not generated by elements of order 2.*

Consider a homomorphism $\psi: H \to K$ of profinite groups: projective limits of finite groups. Call it a *frattini cover* if the equivalent conditions (i) or (ii) hold.

(i) ψ is surjective and $\ker(\psi)$ is contained in the frattini group of H.
(ii) Subset S of H generates H if and only if $\psi(S)$ generates K.

The main result for frattini covers is the existence of a *universal* frattini cover for any profinite group. This is the cover $\tilde{G}$ in the following statement.

PROPOSITION 1.4 ([**FrJ**; *Proposition* 20.33]). *Each profinite group G has a cover $\tilde{\psi}: \tilde{G} \to G$, unique up to isomorphism, satisfying this condition. If $\psi: H \to G$ is any frattini cover of G, there exists a cover $\gamma: \tilde{G} \to H$ such that $\psi \circ \gamma = \tilde{\psi}$. Furthermore, $\tilde{g}$ is a profinite projective group.*

1.4. *Proof of Lemma* 1.3. We may assume $G \neq \{1\}$. Consider the universal frattini cover, $\tilde{\psi}: \tilde{G} \to G$, of G. Let $\mathcal{N} = \{N_i | i \in I\}$ be the collection of all normal subgroups of finite index of $\tilde{G}$. Let $F_i = \tilde{G}/N_i$, $i \in I$, and for 2 indices $i, j \in I$ such that $N_j \supseteq N_i$, let

$\pi_{ij}\colon F_i \to F_j$ be the natural homomorphism. The system $\langle F_i, \pi_{ij}\rangle$ is projective. From compactness of $\widetilde{G}$, $\varprojlim F_i = \widetilde{G}$. Take $n = |G|$. For each $i \in I$, let $\text{gen}_2(F_i)$ be the subset of F_i^n consisting of all n-tuples $\alpha = (\alpha_1, \dots, \alpha_n)$ such that $\langle \alpha_1, \dots, \alpha_n\rangle = F_i$ and $\alpha_i^2 = 1$, $i = 1, \dots, n$. For $i, j \in I$ with $N_j \supseteq N_i$, denote the restriction to $\text{gen}_2(F_i)$ of the natural map induced by π_{ij} on F_i^n by $\pi_{ij}\colon \text{gen}_2(F_i) \to \text{gen}_2(F_j)$.

The system $\{\text{gen}_2(F_i), \pi_{ij}\}$ is projective and an element of $\varprojlim \text{gen}_2(F_i)$ is an n-tuple $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_n)$ such that $\langle \tilde{\alpha}_1, \dots, \tilde{\alpha}_n\rangle = \widetilde{G}$ and $\tilde{\alpha}_i^2 = 1$ for $i = 1, \dots, n$. Yet, such an n-tuple cannot exist. Indeed, from Proposition 1.4, $\widetilde{G}$ is projective. Therefore, it has no nontrivial element of finite order **[FrJ**; Cor. 20.14]. Conclude that $\varprojlim \text{gen}_2(F_i)$ is empty. For all $i \in I$, $\text{gen}_2(F_i)$ is finite, hence compact. Thus, $\text{gen}_2(F_i)$ is empty for some $i \in I$. That is, elements of order 2 in F_i do not generate F_i.

Next, set $F = \widetilde{G}/(\ker \tilde{\psi} \cap N_i)$. We easily see that the natural map $\psi\colon F \to G$ is a frattini cover. From Axiom (ii) for frattini covers, the elements of order 2 in F do not generate F. The finite cover $\psi\colon F \to G$ is the required cover. □

1.5. *Persistence of condition* (1.2) *to frattini covers.* The collection of finite groups has no practical topology on it. Therefore, a statement about a property being *general* for finite groups has traditionally been applied by restricting consideration to natural sequences of finite groups. For example, a statement that indexes the subscript n among the alternating groups A_n is typical.

On the other hand, suppose a property P can be interpreted for all finite groups. Assume that G has property P. As above, consider those frattini covers of G that also have property P. For one, Proposition 1.4 shows these groups—as a collection—intrinsically attach to G. Therefore, *persistence* of property P to hold for frattini covers is intrinsic to the immediate seed group G. In addition, the kernel of the universal frattini cover $\widetilde{G}$ of G is pro-nilpotent. Thus, there are *measures* of the persistence of property P. The following question introduces an analog of Lemma 1.3 that fits the above discussion.

Question 1.5. Consider a group G that satisfies condition (1.2). Does its universal frattini cover satisfy (1.2)?

If "Yes" is the answer to Question 1.5, then a cofinal family of finite frattini covers of G satisfies (1.2).

If G is a p-group, then the universal frattini cover $\tilde{G}$ of G is a free pro-p-group. In addition, in all cases, $\tilde{G}$ has the same *rank*—minimal number of generators—as G [**FrJ**; §20.8].

Observation 1.6. Question 1.5 has a positive answer when G is a p-group satisfying (1.2).

Proof. A characteristic subgroup of $\tilde{G}$ gives the quotient G. Since $\tilde{G}$ is a free group, there is an automorphism of $\tilde{G}$ satisfying (1.2) that extends condition (1.2) for G. □

Let $\mathscr{C}$ be a nontrivial family of finite groups. We say $\mathscr{C}$ is *full* [**FrJ**; p. 189] if $\mathscr{C}$ is closed under taking subgroups, quotients, and middle terms of short exact sequences with end terms in $\mathscr{C}$. If $\mathscr{C}$ is full, there is a unique free pro-$\mathscr{C}$-group of any given rank [**FrJ**; Prop. 15.17]. For the case of rank s, denote this by $\hat{F}_s(\mathscr{C})$. In fact, the free pro-$\mathscr{C}$-group on s generators clearly has an automorphism h that satisfies (1.2).

If G is not a p-group, then we do not know the answer to Question 1.5. We conclude this section by showing that the universal frattini cover $\tilde{G}$ of G is not of the form $\hat{F}_s(\mathscr{C})$. Here $\mathscr{C}$ can be any full family of finite groups. In particular, this suggests a negative answer to Question 1.5 for such a G.

Suppose, on the contrary that $\hat{F}_s(\mathscr{C}) = \tilde{G}$. Let p' and p'' be distinct primes that divide $|G|$. Then, the kernel of $\tilde{G} \to G$ is pro-nilpotent with at least two sylow subgroups, $P_{p'}$ and $P_{p''}$ corresponding to these primes. These are nontrivial free pro-p-groups of finite rank. Since $\ker(\tilde{G} \to G)$ is a subgroup of finite index of $\hat{F}_s(\mathscr{C})$, it is of the form $\hat{F}_{s'}(\mathscr{C})$ for some finite number $s' > s$ [**FrJ**; Prop. 15.27]. The next result gives a contradiction by showing that $\hat{F}_{s'}(\mathscr{C})$ has a non-nilpotent quotient. For this, denote the primes p' and p'' as p and q. Let $\mathbb{Z}/p$ act on $A = (\mathbb{Z}/q)^p$ as cyclic permutations of the coordinates. Consider the semi-direct product $B = A \times {}^s\mathbb{Z}/p$ generated by this action.

PROPOSITION 1.7. *The group B is a non-nilpotent group of rank 2. Assume that pq divides $|G|$. Then, $\tilde{G}$ is not of the form $\hat{F}_s(\mathscr{C})$ for some full family $\mathscr{C}$.*

Proof. Assume we have shown B to have the properties of the proposition. From above, we are done if the non-nilpotent group B

is a quotient of $\widetilde{F}_{s'}(\mathscr{C})$. We know that $\mathscr{C}$ is full family, containing groups whose orders are divisible by p and q. Thus, $\mathscr{C}$ contains B. Since $s' \geq 2$, there is a surjection of $\widehat{F}_{s'}(\mathscr{C})$ on B. It remains to show the properties of B.

Here are two generators of B: $\alpha = (1, 0, \dots, 0) \in a$ and $\tau = 1 \in \mathbb{Z}/p$. Indeed, the $\mathbb{Z}/p$ orbit of α gives a basis for A. Finally, $\mathbb{Z}/p$ is a p-sylow for B. It is not, however, normal: $\alpha\tau\alpha^{-1}$ is $(1, -1, 0, \dots, 0) \times \tau$. Thus, B is not nilpotent. □

2. Basic tools.

2.1. *Identification of Galois and monodromy actions.* Let y_1 be a primitive element of the regular extension $Y/K(X)$. Take $P \in K[X, Y]$ to be an irreducible polynomial such that $P(X, y_1) = 0$ and $\deg_Y P = n$. Identify the curve $Y_{\mathbb{C}}$ with projective normalization of the affine plane curve $P(x, y) = 0$. Here $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ is projection: $(x, y) \to x$. Take x_0 to be distinct from the branch points of the cover.

Let $\widehat{Y\mathbb{C}}$ be the Galois closure of $Y\mathbb{C}/\mathbb{C}(X)$. The Galois group $G(\widehat{Y\mathbb{C}}/\mathbb{C}(X))$ is the *geometric Galois group* of the extension $Y/K(X)$. Embed it in S_n through its action on the n conjugates $y_1, \dots, y_n$ of y_1. Since we assume $Y/K(X)$ is regular, it is a transitive action.

Identify the points $p_1, \dots, p_n$ in the fiber $\varphi^{-1}(x_0)$ and the conjugates $y_1, \dots, y_n$ of y_1 as follows. Each embedding $Y\mathbb{C} \to \mathbb{C}((X-x_0))$ in the Laurent series around x_0 determines a point $p_i \in Y$ above x_0. Since x_0 is not a branch point, there are n such embeddings. Each corresponds to one of the y_is.

From now on, fix an embedding $\widehat{Y\mathbb{C}} \to \mathbb{C}((X - x_0))$. That is, regard $\widehat{Y\mathbb{C}}$ as a subfield of $\mathbb{C}((X-x_0))$ and label the points $p_1, \dots, p_n$ so that p_i corresponds to the power series y_i in $\mathbb{C}((X - x_0))$, $i = 1, \dots, n$. From classical analytic continuation theory, for this labeling, the images in S_n of both $T(\pi_1)$ and the geometric Galois group $G(\widehat{Y\mathbb{C}}/\mathbb{C}(X))$ are the same. Denote this common group by Γ_Y (or simply Γ). Furthermore, denote the image in Γ of an element $s \in T(\pi_1)$ by $\bar{s}$, and the image in Γ of an element $\sigma \in G(\widehat{Y\mathbb{C}}/\mathbb{C}(X))$ by $\bar{\sigma}$. Even in the case where $Y_{\mathbb{C}} \to \mathbb{P}^1$ is Galois, automorphisms of this cover do not naturally identify with automorphisms of $\widehat{Y\mathbb{C}}/\mathbb{C}(X)$. In particular, restriction of the former automorphisms to the fiber over x_0 do not correspond to automorphisms of $\widehat{Y\mathbb{C}}/\mathbb{C}(X)$.

We make an assumption a little stronger than saying that x_0 is not a branch point. We ask that $\frac{\partial}{\partial Y}(P(x_0, Y))$ has no repeated zeros.

Then, the first term $y_i(x_0)$ determines each of the power series y_i. Thus, for the labeling above, identify p_i with the geometric point $(x_0, y_i(x_0))$ on the affine plane curve $P(x, y) = 0$.

2.2. *The arithmetic Galois group.* From here on, assume the base point x_0 is in $\mathbb{P}^1(\mathbb{Q})$. Consider the automorphism group $\mathrm{Aut}(\mathbb{C})$. An automorphism $\tau \in \mathrm{Aut}(\mathbb{C})$ acts coordinatewise on the geometric points of any affine variety defined over $\mathbb{C}$. This action transforms the affine curve with equation $P(x, y) = 0$ into the affine curve of equation $P^\tau(x, y) = 0$. Denote the projective normalization of the curve $P^\tau(x, y) = 0$ by $Y^\tau_{\mathbb{C}}$ and the associated cover by $\varphi^\tau_{\mathbb{C}}\colon Y^\tau_{\mathbb{C}} \to \mathbb{P}^1$.

On the other hand, there is a natural extension of τ to $\mathbb{C}((X-x_0))$. Apply τ to the coefficients of a power series y to get y^τ. Indicate the transform of a subfield F of $\mathbb{C}((X-x_0))$ by F^τ. This action maps the power series $y_1, \dots, y_n$ onto the n roots $y_1^\tau, \dots, y_n^\tau$ in $\mathbb{C}((X-x_0))$ of the polynomial P^τ. Also, the field extension $(Y\mathbb{C})^\tau/\mathbb{C}(X)$ is the function field extension of the cover $\varphi^\tau_{\mathbb{C}}\colon Y^\tau_{\mathbb{C}} \to \mathbb{P}^1$.

Points on $Y^\tau_{\mathbb{C}}$ above x_0 correspond to the power series $y_1^\tau, \dots, y_n^\tau$. Label these, respectively, $p_1^\tau, \dots, p_n^\tau$. As in §2.1, p_i^τ corresponds to the point $(x_0, y_i(x_0)^\tau)$ on the affine curve of equation $P^\tau(x, y) = 0$. Conclude that the effect of τ on $p_1, \dots, p_n$ agrees with the action on the power series and with coordinatewise action on the geometric points.

Denote the subgroup of $\mathrm{Aut}(\mathbb{C})$ consisting of all automorphisms that fix K by $\mathrm{Aut}_K(\mathbb{C})$. Assume, in addition, that $\tau \in \mathrm{Aut}_K(\mathbb{C})$. Then $P = P^\tau$, $\widehat{Y\mathbb{C}} = \widehat{Y\mathbb{C}}^\tau$ and τ permutes the points $p_1, \dots, p_n$ in the fiber $\varphi^{-1}(x_0)$. Thus, τ induces a permutation $\overline{\tau} \in S_n$. Now consider $\widehat{Y}$, the Galois closure over $K(X)$ of the extension $Y/K(X)$. Call the Galois group $G(\widehat{Y}/K(X))$ the *arithmetic Galois group* of the extension. Label the image of $y \in \widehat{Y}$ under the automorphism $\sigma \in G(\widehat{Y}/K(X))$ by $\sigma(y)$. Also, denote the permutation of $\{1, \dots, n\}$ induced by σ on $\{y_1, \dots, y_n\}$ by $\overline{\sigma}$. Use $\widehat{\Gamma}$ for the group $\{\overline{\sigma} | \sigma \in G(\widehat{Y}/K(X))\}$. Note that $\overline{\tau} \in \widehat{\Gamma}$, for all $\tau \in \mathrm{Aut}_K(\mathbb{C})$.

PROPOSITION 2.1. *The group Γ is normal in $\widehat{\Gamma}$. The quotient group $\widehat{\Gamma}/\Gamma$ consists of the cosets modulo Γ of the elements $\overline{\tau}$, with $\tau \in \mathrm{Aut}_K(\mathbb{C})$.*

Proof. Let $\widehat{K}$ be the constant field of the extension $\widehat{Y}/K(X)$: $\widehat{K} = \widehat{Y} \cap \overline{K}$. Clearly, $\widehat{Y\mathbb{C}} = \widehat{Y}\mathbb{C}$; restriction $G(\widehat{Y\mathbb{C}}/\mathbb{C}(X)) \to G(\widehat{Y}/\widehat{K}(X))$ is an isomorphism. In particular, Γ is the image of $G(\widehat{Y}/\widehat{K}(X))$ in

S_n. It is a normal subgroup of $\widehat{\Gamma}$ because $\widehat{K}/K$ is Galois. The map $\mathrm{Aut}_K(\mathbb{C})$ to $G(\widehat{K}/K)$ is onto. Therefore, $\overline{\tau}$, with $\tau \in \mathrm{Aut}_K(\mathbb{C})$, form a full set of representatives (perhaps not distinct) for the quotient $\widehat{\Gamma}/\Gamma$. The result follows. □

2.3. *Complex conjugation and monodromy.* Retain §2.1–§2.2 notation. We know generators for the fundamental group $\pi_1 = \pi_1(\mathbb{P}^1\setminus\{x_1, \ldots, x_r\}, x_0)$. These are homotopy classes $[\gamma_i]$ of suitably chosen loops starting from x_0 around the branch points x_i, $i = 1, \ldots, r$. These freely generate except for one relation, $[\gamma_1][\gamma_2]\cdots[\gamma_r] = 1$. For $i = 1, \ldots, r$, set $s_i = T([\gamma_i])$; the s_is generate the monodromy group of the cover and satisfy $s_1 s_2 \cdots s_r = 1$.

Call the τ-tuple $(s_1, \ldots, s_r)$ *the branch cycle description* of the cover associated with the data (or *bouquet*) $(\gamma_1, \ldots, \gamma_r)$. It is an element of S_n^r when we label the points $p_1, \ldots, p_n$ in the fiber $\varphi^{-1}(x_0)$. Another labeling of the fiber $\varphi^{-1}(x_0)$ defines an element of S_n^r that is coordinatewise conjugate by an element of S_n to the first branch cycle description of the cover coming from the bouquet $(\gamma_1, \ldots, \gamma_r)$. This produces a one-one correspondence between the following sets:

- degree n covers $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ (up to equivalence of covers) ramified over the points $x_1, \ldots, x_r$; and
- r-tuples $(s_1, \ldots, s_r) \in S_n^r$ (modulo coordinatewise conjugation by S_n) with $s_1 s_2 \cdots s_r = 1$ and $\langle s_1, \ldots, s_r\rangle$ transitive on $1, \ldots, n$.

Unless otherwise specified, assume from here the following.

(2.1) Branch points $x_1, \ldots, x_r$, $r \geq 3$,

are in $\mathbb{P}^1(\mathbb{R})$ and $x_1 < x_2 < \cdots < x_r \leq \infty$.

Fix the base point $x_0 \in \mathbb{P}^1(\mathbb{Q})\setminus\{\infty\}$ on the arc between x_1 and x_r not containing x_2 on the real projective line. Denote complex conjugation on $\mathbb{C}$ by c. It maps the homotopy class $[\gamma] \in \pi_1$ of a closed path γ based at x_0 to the homotopy class $[\gamma^c]$ of the conjugate path γ^c. With suitable loops around the x_is, we write this action explicitly. For the rest of §2 and §3 use the specific bouquet $(\gamma_1, \ldots, \gamma_r)$ from **[FrD**; §2.1]. For this we have the following.

PROPOSITION 2.2. *The paths $\gamma_1^c, \ldots, \gamma_r^c$ are respectively homotopic to*

$$(\gamma_2\cdots\gamma_r)^{-1}\gamma_1^{-1}(\gamma_2\cdots\gamma_r),\ (\gamma_3\cdots\gamma_r)^{-1}\gamma_2^{-1}(\gamma_3\cdots\gamma_r),\ \ldots,\ (\gamma_r)^{-1}\gamma_{r-1}^{-1}\gamma_r,\ \gamma_r^{-1}.$$

Hurwitz knew these formulas [**Hur**; p. 357]. Krull and Neukirch [**KN**] investigated them further. We consider them deriving from the action of a general operator. Suppose we have a group U and an integer $r > 0$. Define $\mathscr{C}_r\colon U^r \to U^r$ to send $\mathbf{u} = (u_1, \dots, u_r) \in U^r$ to $\mathscr{C}_r(\mathbf{u}) \stackrel{\text{def}}{=} (u_1^{\mathscr{C}}, \dots, u_r^{\mathscr{C}})$ with $u_r^{\mathscr{C}} = u_r^{-1}$ and

$$(2.2)\qquad u_i^{\mathscr{C}} = (u_{i+1}\cdots u_r)^{-1}u_i^{-1}(u_{i+1}\cdots u_r), \qquad i = 1, \dots, r-1.$$

We also have

$$(2.3)\qquad u_i^{\mathscr{C}}\cdots u_r^{\mathscr{C}} = (u_i\cdots u_r)^{-1}, \qquad i = 1, \dots, r-1.$$

Consider a cover $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ and its conjugate $\varphi^c\colon Y_{\mathbb{C}}^c \to \mathbb{P}^1$. The fiber $(\varphi^c)^{-1}(x_0)$ consists of the points $p_1^c, \dots, p_n^c$. Let T^c denote the monodromy action on the fiber $(\varphi^c)^{-1}(x_0)$. For any closed path γ based at x_0, we have $T^c([\gamma^c])(p_i^c) = [T([\gamma])(p_i)]^c$. Replace γ by γ^c and apply c to both sides. This gives the equivalent expression:

$$(2.4)\qquad T^c([\gamma])(p_i^c)^c = T([\gamma^c])(p_i).$$

From (2.4):

(2.5) the r-tuple $(T([\gamma_1^c]), \dots, T([\gamma_r^c]))$ is the branch cycle description of the cover $\varphi^c\colon Y_{\mathbb{C}}^c \to \mathbb{P}^1$ associated with the bouquet $(\gamma_1, \dots, \gamma_r)$.

The (a) part of the next proposition rephrases (2.4) and (2.5). The (b) part follows because the assumptions imply $Y_{\mathbb{C}}^c = Y_{\mathbb{C}}$.

PROPOSITION 2.3. (a) *Suppose* $\mathbf{s} = (s_1, \dots, s_r)$ *is the branch cycle description of the cover* $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ *associated with the bouquet* $(\gamma_1, \dots, \gamma_r)$. *Then,* $\mathscr{C}_r(\mathbf{s}) = (s_1^{\mathscr{C}}, \dots, s_r^{\mathscr{C}})$ *is the branch cycle description of the cover* $\varphi^c\colon Y_{\mathbb{C}}^c \to \mathbb{P}^1$ *associated with the bouquet* $(\gamma_1, \dots, \gamma_r)$.
(b) *If* $\mathbb{R} \supset K$ *then* $\mathscr{C}_r(\bar{\mathbf{s}}) = \bar{c}\,\bar{\mathbf{s}}\,\bar{c}$. *That is* $\bar{s}_i^{\mathscr{C}} = \bar{c}\,\bar{s}_i\bar{c}$, $i = 1, \dots, r$.

2.4. *Descending the base field—Weil's method.* We now descend the base field in the second part of the proof of Theorem 1.1. Without condition (ii) below, it results from Prop. 2.5 of [**CoH**]. Here is the framework. Let $\Psi\colon E \to \mathbb{P}^1$ be a Galois cover, and let H be the subgroup of $\mathrm{Aut}(\mathbb{C})$ given as

$$\{\tau \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q}) \mid \Psi\colon E \to \mathbb{P}^1 \text{ and } \Psi^\tau\colon E^\tau \to \mathbb{P}^1 \text{ are equivalent covers}\}.$$

Take $K = \mathbb{C}^H$, the fixed field of H in $\mathbb{C}$. Then, K is the *field of moduli* of the cover. Choose x_0, a point in $\mathbb{Q}$ distinct from the branch points of the cover.

THEOREM 2.4. *Assume the conditions of the paragraph above. There exists an extension* $Y/K(X)$, *regular over* K, *such that*

(i) *the cover* $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ *is equivalent to the cover* $\Psi\colon E \to \mathbb{P}^1$, *and*

(ii) $K((x-x_0))$ *contains* Y.

Condition (ii) is equivalent to the following.

(ii′) Permutations $\overline{\tau}$ acting on the Galois closure of $Y/K(X)$ have a common fixed point for all $\tau \in H$ (notation as in §2.2).

Danger: $Y/K(X)$ need not be Galois. It is Galois if and only if $\overline{\tau} = 1$, for all $\tau \in H$. That is, one point of the cover over x_0 is defined over K. Thus, if the cover is Galois, all points over x_0 must be defined over K. In the other direction, let $\widehat{K}$ be the constants of the Galois closure of the extension $Y/K(X)$. Then $\widehat{K} = K$ if and only if $Y/K(X)$ is Galois. We know the field generated by coordinates of the collection of points above x_0 contains $\widehat{K}$. Therefore, if these points are defined over K, then $\widehat{K} = K$.

Proof. By definition, for each $\tau \in H$, there is an isomorphism $\delta_\tau\colon E \to E^\tau$ such that $\Psi^\tau \circ \delta_\tau = \Psi$. The automorphism δ_τ sends the fiber $\Psi^{-1}(x_0) = \{e_1, \ldots, e_n\}$ to the fiber $(\Psi^\tau)^{-1}(x_0) = \{e_1^\tau, \ldots, e_n^\tau\}$. The cover $\Psi^\tau\colon E^\tau \to \mathbb{P}^1$ is Galois. Thus, there exists an automorphism $\chi_\tau\colon E^\tau \to E^\tau$ such that $\chi_\tau \circ \delta_\tau$ sends e_1 to e_1^τ. Denote the isomorphism $\chi_\tau \circ \delta_\tau$ by c_τ. The collection $\{c_\tau\}_{\tau\in H}$ satisfies the cocycle condition: $c_{\tau_1}^{\tau_2} \circ c_{\tau_2} = c_{\tau_1\tau_2}$ for all $\tau_1, \tau_2 \in H$. Indeed:

$$c_{\tau_1}^{\tau_2} \circ c_{\tau_2}(e_1) = c_{\tau_1}^{\tau_2}(e_1^{\tau_2}) = c_{\tau_1}(e_1)^{\tau_2} = e_1^{\tau_1\tau_2} = c_{\tau_1\tau_2}(e_1).$$

Weil's cocycle criterion now reduces the field of definition **[We]**. There exists a cover $\varphi_K\colon E_k \to \mathbb{P}^1$, defined over K with the following properties. There is an isomorphism $\Theta\colon E_K \to E$ (defined over $\mathbb{C}$) such that

$$\text{(2.6)} \qquad \begin{aligned} &\text{(a)} \quad \Psi \circ \Theta = \varphi_K, \quad \text{and} \\ &\text{(b)} \quad \Theta^\tau \circ \Theta^{-1} = c_\tau, \quad \text{for all } \tau \in H. \end{aligned}$$

Define Y to be the function field over K of E_K. The extension $Y/K(X)$ is regular and satisfies condition (i). In fact, $\varphi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ is the cover $\varphi_K\colon E_K \to \mathbb{P}^1$.

Finally, consider the point $p_1 = \Theta^{-1}(e_1)$ on E_K. From (2.6) (b), $p_1^\tau = p_1$, for all $\tau \in H$. That is, $p_1 \in E_K$ is K-rational. As before, let y_1 be the power series corresponding to p_1. Then $y_1 \in K((X-x_0))$. Since $Y = K(X, y_1)$, $K((X-x_0)) \supset Y$. □

3. Proof of Theorem 1.1.

3.1. *Proof of Theorem* 1.1 (b) $\Rightarrow$. Let $Y/\mathbb{R}(X)$ be a degree n regular extension whose associated cover $\varphi : Y_{\mathbb{C}} \to \mathbb{P}^1$ has monodromy group $\Gamma_Y = G$. Let $\mathbf{s} = (s_1, \dots, s_r)$ be the branch cycle description of φ: $Y_{\mathbb{C}} \to \mathbb{P}^1$ associated to the bouquet $(\gamma_1, \dots, \gamma_r)$ of Proposition 2.2. From Proposition 2.3 (b), we have $s_i^{\mathscr{C}} = \bar{c}\,\bar{s}_i\bar{c}$, $i = 1, \dots, r$. Apply (2.3). Then, $(\bar{s}_i \cdots \bar{s}_r)^{-1} = \bar{c}(\bar{s}_i \cdots \bar{s}_r)\bar{c}$, $i = 1, \dots, r$. Set $\alpha_i = \bar{s}_{i+1} \cdots \bar{s}_r$, $i = 1, \dots, r-1$. Thus

$$\bar{c}\alpha_i\bar{c} = \alpha_i^{-1}, \qquad i = 1, \dots, r-1. \tag{3.1}$$

Conjugating G by $\bar{c} \in S_n$ gives the h that Theorem 1.1 (b) requires. □

3.2. *Proof of Theorem* 1.1 (a) $\Rightarrow$. Here, $Y/\mathbb{R}(X)$ is a degree n Galois regular extension with group $\Gamma_Y = G$. So (3.1) of §3.1 still holds. In addition, since $\widehat{\Gamma}_Y = \Gamma_Y$, we have $\bar{c} \in G$ (statement prior to Proposition 2.1). Thus, $\bar{c}, \overline{c\alpha}_1, \dots, \overline{c\alpha}_{r_1}$ are of order ≤ 2 and they generate G. □

3.3. *Proof of Theorem* 1.1 (b) $\Leftarrow$. Let G be a group with property (1.2). Let $r = s + 1$ and $n = |G|$. Regard G as a subgroup of S_n through its regular representation. Consider the r-tuple $\mathbf{s} = (s_1, \dots, s_r) \in S_n^r$ defined by

$$\mathbf{s} = (\alpha_1, \alpha_1^{-1}\alpha_2, \alpha_2^{-1}\alpha_3, \dots, \alpha_{r-2}^{-1}\alpha_{r-1}, \alpha_{r-1}^{-1}). \tag{3.2}$$

The s_is generate G. They also satisfy $s_1 \cdots s_r = 1$. Fix $r+1$ points $x_0, x_1, \dots, x_r$ in $\mathbb{P}^1(\mathbb{R})$ and a bouquet $(\gamma_1, \dots, \gamma_r)$ as in §2.3. From Riemann's Existence Theorem (§2.3), there exists a cover Ψ: $E \to \mathbb{P}^1$, unique up to equivalence of covers, with the following properties. Its branch points are $x_1, \dots, x_r$, and $\mathbf{s} = (s_1, \dots, s_r)$ is the branch cycle description of the cover associated to the bouquet $(\gamma_1, \dots, \gamma_r)$. Furthermore, since $G \to S_n$ is the regular representation, Ψ: $E \to \mathbb{P}^1$ is a Galois cover with automorphism group G.

From Proposition 2.3 (a), $\mathscr{C}_r(\mathbf{s}) = (s_1^{\mathscr{C}}, \dots, s_r^{\mathscr{C}})$ is the branch cycle description of the cover Ψ^c: $E^c \to \mathbb{P}^1$ associated to the bouquet $(\gamma_1, \dots, \gamma_r)$. From the definition of $\mathscr{C}_r$ and (1.2) check easily that $s_i^{\mathscr{C}} = h(s_i)$, $i = 1, \dots, r$. Suppose that conjugation by $\kappa \in S_n$ coincides with the automorphism h on G. Thus:

$$s_i^{\mathscr{C}} = \kappa s_i \kappa^{-1} \quad \text{for } i = 1, \dots, r. \tag{3.3}$$

From Riemann's Existence Theorem (§2.3), the covers $\Psi: E \to \mathbb{P}^1$ and $\Psi^c: E^c \to \mathbb{P}^1$ are equivalent covers. Apply Theorem 2.4 to conclude there exists a regular extension $Y/\mathbb{R}(X)$ with these properties.

(i) $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ is equivalent to the cover $\Psi^c: E^c \to \mathbb{P}^1$.

(ii) $\mathbb{R}((X - x_0))$ contains Y.

The cover $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ is defined over $\mathbb{R}$. It is the desired cover. □

3.4. *Proof of Theorem* 1.1 (a) $\Leftarrow$. Let G be a group generated by involutions $\alpha_1, \dots, \alpha_s$. In particular, G has property (1.2) with $h = 1$. Thus, the construction around (3.3) holds, with $h = 1$, $\kappa = 1$. Consider the regular extension $Y/\mathbb{R}(X)$ produced in §3.3. It is Galois over $\mathbb{C}(x)$ with (geometric) Galois group G. Also, $\mathbb{R}((X-x_0))$ contains Y. The branch cycle description $\mathbf{s} = (s_1, \dots, s_r)$ of the cover $\varphi: Y_{\mathbb{C}} \to \mathbb{P}^1$ associated with the bouquet $(\gamma_1, \dots, \gamma_r)$ has this property:

$$s_i^{\mathscr{C}} = s_i \quad \text{for } i = 1, \dots, r. \tag{3.4}$$

From Proposition 2.3 (b), we also have $s_i^{\mathscr{C}} = \overline{c} s_i \overline{c}$, $i = 1, \dots, r$. Therefore, $\overline{c} \in \mathrm{Cen}_{S_n}(G)$. Since $\mathbb{R}((X - x_0))$ contains Y, $\overline{c}$ has a fixed point. Conclude that $\overline{c} = 1$. Therefore, from Proposition 2.1, $\widehat{\Gamma}_Y = \Gamma_Y$: $Y/\mathbb{R}(X)$ is a Galois regular extension with Galois group $\Gamma_Y = G$. □

REMARK. In the above argument, $\overline{c} = 1$. That is, $\varphi^{-1}(x_0)$ has only real points. Equivalently, $\mathbb{R}$ contains the residue class algebra Y_{x_0}. □

3.5. *Comments.* This section consists of elaborate comments. Each uses the proof of Theorem 1.1 for further exploration. These are the topics.

- Branch points need not be real.
- The cover need not be Galois.
- You can decide when the field of moduli of a cover is $\mathbb{R}$.

Comment 1. *Dropping the assumption "the branch points are real."* The "real branch point situation" of Theorem 1.1 allowed special generators $[\gamma_1], \dots, [\gamma_r]$ of the fundamental group π_1 from §2.3. Explicit formulas gave $[\gamma_1^c], \dots, [\gamma_r^c]$ as words in $[\gamma_1], \dots, [\gamma_r]$ (cf. Proposition 2.2). We can work with the general cover defined over $\mathbb{R}$ similarly.

Here, the branch points consist of r_1 real points and r_2 complex conjugate pairs, where $r = r_1 + 2r_2$. Use the paths of [**FrD**; §2.2] for which we know the complex conjugation action explicitly. Slight adjustments to the proof above lead to this more general result.

THEOREM 3.1. *Finite group* G *is the group of a regular extension* $Y/\mathbb{R}(X)$ *with* r *branch points,* r_1 *of these real, exactly when* G *has special generators. Specifically,* $(r+r_1)/2$ *elements generate* G *with at least* r_1 *of them involutions.*

More precisely, the following statements are equivalent.

(a) There exists a Galois regular extension $Y/\mathbb{R}(X)$ of group G, with r branch points $t_1, \dots, t_{r_1}, \bar{z}_{r_2}, \dots, \bar{z}_1, z_1, \dots, z_{r_2}$, where $t_i \in \mathbb{R}$, $i = 1, \dots, r_1$, and $z_i \notin \mathbb{R}$, $i = 1, \dots, r_2$.

(b) There exists $(g'_1, \dots, g'_r) \in G^r$ which satisfy these conditions:

(i) $g'_1 \cdots g'_r = 1$,

(ii) $\langle g'_1, \dots, g'_r \rangle = G$,

(iii) $\exists g'_0 \in G$ such that $(g'_0 \cdots g'_i)^2 = 1$, $i = 0, \dots, r_1 - 1$,

$$g'_{r-i} = g'_0 (g'_{r_1+1+i})^{-1} g'_0, \quad i = 0, \dots, r_2 - 1.$$

The special case $r = r_1$ corresponds to Theorem 1.1 (a). For $r_1 = 0$, we get a result from the introduction. Namely, every finite group G is the Galois group of a Galois regular extension of $\mathbb{R}(X)$.

Comment 2. *Nonregular representations.* Here, suppose G has an embedding in S_n (not necessarily the regular representation). Assume $\alpha_1, \dots, \alpha_s$ are generators for which (1.2) holds. Denote the r-tuple of (3.2) by $\mathbf{s}(\alpha)$. Let $x_0, x_1, \dots, x_r$ be $r+1$ points in $\mathbb{P}^1(\mathbb{R})$. Take $(\gamma_1, \dots, \gamma_r)$ to be a bouquet as in §2.3 with $\mathbf{s}(\alpha)$ the associated branch cycle description of the cover with $x_1, \dots, x_r$ as branch points. Denote the degree n (not necessarily Galois) cover from §3.3 by $\Psi_{\mathbf{s}(\alpha),\mathbf{x}}: E \to \mathbf{P}^1$. We ask if we can define this cover over $\mathbb{R}$.

We showed the answer to be positive in the Galois case, thanks to Theorem 2.4. In greater generality, the answer is yes whenever you can construct a collection $\{c_\tau\}_{\tau \in G(\mathbb{C}/\mathbb{R})}$ as in Theorem 2.4. It must satisfy the cocycle condition $c_{\tau_1}^{\tau_2} \circ c_{\tau_2} = c_{\tau_1\tau_2}$, for all $\tau_1, \tau_2 \in G(\mathbb{C}/\mathbb{R})$. For example, you can do this when the cover $\Psi: E \to \mathbb{P}^1$ has no nontrivial automorphism. This is the same as the condition $\mathrm{Cen}_{S_n}(G) = \{1\}$.

Comment 3—*from E. Dew* **[D]**. *When the field of moduli is* $\mathbb{R}$. Suppose $\psi: E \to \mathbb{P}^1$ is a Galois cover and complex conjugation gives an equivalent cover $\Psi^e: E^e \to \mathbb{P}^1$. We say $\mathbb{R}$ contains the *field of moduli.* Suppose also that the covers have real branch points. Let

$\gamma = (\gamma_1, \ldots, \gamma_r)$ be a bouquet as in §2.3 and let $(s_1, \ldots, s_r)$ be the branch cycle description associated to the bouquet γ. With $\alpha_i = s_1 \cdots s_i$, $i = 1, \ldots, r-1$, Proposition 2.3 gives this:

(∗) The set $N_\alpha = \{\kappa \in S_n : \kappa\alpha_i\kappa^{-1} = \alpha_i^{-1},\ i = 1, \ldots, r-1\}$ is nonempty.

Thus, (∗) is a necessary condition. We want to know what to add to this for an if and only if condition for the following:

(∗∗) There is a cover equivalent to $\psi : E \to \mathbb{P}^1$ defined and Galois over $\mathbb{R}$.

It is tempting to answer: $N_\alpha \cap G$ is nonempty. Here G denotes the monodromy group of the cover. Yet, this condition may not be sufficient in general. The correct answer is this:

(∗∗∗) $$\exists \kappa \in N_\alpha \cap G \text{ with } \kappa^2 = 1.$$

Note. In the *addition* following Theorem 1.1 (b) we selected the s_is so $\kappa = 1$ lies in N_α. Also, (∗∗∗) is equivalent to asking that κ^2 be the square of an element of the center $Z(G)$; divide κ by this element.

Proof of the equivalence of (∗∗) *and* (∗∗∗). Assume that the cover $\psi : E \to \mathbb{P}^1$ is defined and Galois over $\mathbb{R}$. Then the element $\overline{c}$ (see §2.2 for the definition of $\overline{c}$) is in $N_\alpha \cap G$ and it satisfies $\overline{c}^2 = 1$.

In the other direction, assume (∗∗∗). Following the proof of Theorem 2.4 we use Weil's criterion. Here, however, we choose a different cocycle. Let $H = \{1, c\}$ denote the Galois group of $\mathbb{C}/\mathbb{R}$. Recall the dictionary between covers and branch cycle descriptions (for the bouquet γ). An isomorphism $\delta : E \to E^c$ such that $\psi^c \circ \delta = \psi$ comes from an element κ in N_α.

To use (∗∗∗), label points $\mathbf{p}$ on E above the base point x_0. Apply c to $\mathbf{p}$; then permute the naming of the image points $\mathbf{p}^c$ by κ. The new points $\kappa(\mathbf{p}^c)$ give us points above x_0 in E^c. These produce exactly the same branch cycle description (relative to γ) for E^c as do the points $\mathbf{p}$ for E. Thus, these respective namings of the points give a unique isomorphism $\delta_c : E \to E^c$ that sends points $\mathbf{p}$ to the respective points $\kappa(\mathbf{p}^c)$. In addition to $\psi^c \circ \delta = \psi$, δ_c satisfies these two conditions:

(†) $\delta_c^c \circ \delta_c = 1$; and

(††) δ_c commutes with the action of c that takes automorphisms of $E \to \mathbb{P}^1$ to automorphisms of $E^c \to \mathbb{P}^1$.

Indeed, (†) follows because the effect of the left side of (†) on $\mathbf{p}$ is given by κ^2. As for (††), automorphisms of the covers commute with a renaming of the points of $\mathbf{p}$.

For convenience take δ_1 to be the identity. Condition (†) guarantees that the collection $\{\delta_\tau\}$ satisfies the cocycle condition

$$\delta_{\tau_r}^{\tau_2} \circ \delta_{\tau_2} = \delta_{\tau_1 \tau_2}.$$

Therefore, one can descend the field of definition of the cover to $\mathbb{R}$. Condition (††) assures the automorphisms are also defined over $\mathbb{R}$.

Section 3.6 gives a more algebraic approach to the above. In particular, the equivalence of (**) and (***) follows immediately from Lemma 3.3.

3.6. *Serre's approach.* Serre suggested that the algebraic fundamental group, rather than the topological fundamental group, would be more convenient for proving Theorem 1.1 (a). We follow Serre's exposition [**Se3**; cf. Ch. 7, 8, 9].

Assume K has characteristic 0. Let $x_1, \ldots, x_r$ be r distinct points in $\mathbb{P}^1(\overline{K})$. Denote the maximal algebraic extension of $\overline{K}(X)$ unramified outside $x_1, \ldots, x_r$ by Ω. The extension $\Omega/\overline{K}(X)$ is Galois. Its group is the *algebraic fundamental group* of $\mathbb{P}^1(\overline{K})\backslash\{x_1, \ldots, x_r\}$. Denote this profinite group by π^{alg}.

When $\overline{K} = \mathbb{C}$, π^{alg} is the profinite completion $\hat{\pi}$ of the topological fundamental group π [**Se3**; Theorem 7.5, p. 69]. By analogy with the complex case, denote the free group on r generators $\Gamma_1, \ldots, \Gamma_r$ with the single relation $\Gamma_1 \cdots \Gamma_r = 1$ by π. There is a map $i\colon \pi \to \pi^{\text{alg}}$ with the following properties.

(i) $i(\Gamma_i) \stackrel{\text{def}}{=} \Gamma_i$ is a generator of an inertia group of the extension $\Omega/\overline{K}(X)$ above x_i, $i = 1, \ldots, r$.

(ii) The map i extends to an isomorphism $\hat{i}\colon \hat{\pi} \to \pi^{\text{alg}}$.

If the divisor $(x_1) + (x_2) + \cdots + (x_r)$ of $\mathbb{P}^1$ is K-rational, the extension $\Omega/K(X)$ is Galois. Let π_K denote the Galois group of this extension. We have this exact sequence:

$$(3.5) \qquad 1 \to \pi^{\text{alg}} \to \pi_K \to \Lambda_K \to 1.$$

Here Λ_K denotes the Galois group of the extension $\overline{K}/K$. Note: the map $\pi_K \to \Lambda_K$ has many sections. Indeed, for each $x_0 \in \mathbb{P}^1(K)\backslash\{x_1, \ldots, x_r\}$, we can embed Ω in $\overline{K}((X - x_0))$ where the elements of Λ_K act naturally (cf. §2.2).

Given a finite group G, a surjective homomorphism $\psi \in \text{Hom}(\pi^{\text{alg}}, G)$ produces a Galois extension $E/\overline{K}(X)$ with group G.

We say E descends to K if there exists a Galois regular extension $E_K/K(X)$ with $\mathbb{C}E_K = E$. This happens if and only if the homomorphism ψ extends to π_K.

In our context, $K = \mathbb{R}$ and the branch points $x_1, \dots, x_r$ are real. Section 2.3 gives generators $\Gamma_1, \dots, \Gamma_r$ of π^{alg} so that complex conjugation $c \in \Lambda_{\mathbb{R}}$ acts on them by the formulas (2.2). Recall from §2.3 the operator $\mathscr{C}$ in our next result.

PROPOSITION 3.2. *Assume the branch points $x_1, \dots, x_r$ are real. Then, π_K is isomorphic to the semi-direct product $\pi^{\text{alg}} \times {}^s\mathbb{Z}/2$ where $c = 1 \in \mathbb{Z}/2$ maps $\Gamma \in \pi^{\text{alg}}$ to Γ^c as follows:*

$$(3.6) \qquad \Gamma_i^c = \Gamma_i^{\mathscr{C}}, \qquad i = 1, \dots, r.$$

The group theoretical observation that supports Theorem 1.1 (a) now appears clearly.

LEMMA 3.3. *Let $\psi \in \operatorname{Hom}(\pi^{\text{alg}}, G)$ and $g_i = \psi(\Gamma_1)\cdots\psi(\Gamma_i)$, $i = 1, \dots, r$. Then, ψ extends to $\tilde{\psi} \in \operatorname{Hom}(\pi^{\text{alg}} \times {}^s\mathbb{Z}/2, G)$ if and only if there exists an involution $\kappa \in G$ with all of $\kappa g_1, \dots, \kappa g_r$ involutions.*

Proof. Assume $\tilde{\psi} \in \operatorname{Hom}(\pi^{\text{alg}} \times {}^s\mathbb{Z}/2, G)$ extends ψ. Set $\kappa = \tilde{\psi}(c)$; $|\kappa| = 2$ and

$$(3.7) \qquad \psi(\Gamma^c) = \kappa\psi(\Gamma)\kappa$$

for each $\Gamma \in \pi^{\text{alg}}$. Substitute Γ_i for Γ and use (2.3) to get $g_i^{-1} = \kappa g_i \kappa$, $i = 1, \dots, r$.

For the converse, define $\tilde{\psi} \in \operatorname{Hom}(\pi^{\text{alg}} \times {}^s\mathbb{Z}/2, G)$ by $\tilde{\psi}(\Gamma, \varepsilon) = \psi(\Gamma)\kappa^\varepsilon$ for each $\Gamma \in \pi^{\text{alg}}$ and $\varepsilon = 0, 1$. Use (3.6) to check that (3.7) holds for $\Gamma = \Gamma_i$, $i = 1, \dots, r$, and so for all $\Gamma \in \pi^{\text{alg}}$. This guarantees that $\tilde{\psi}$ is a homomorphism of groups. □

3.7. *p-adic analogs.* Proposition 3.2 gives the effect of complex conjugation c:

(3.8) Γ_i^c is conjugate in π to Γ_i^{-1}, $i = 1, \dots, r$.

The exponent -1 comes from the "branch cycle argument" ([**Fr1**; p. 62] or [**DFr**; §1.4 Proposition 1.9]). We explain. Consider the cyclotomic character $\chi\colon \Lambda_K \to \prod_N G(K(\mu_N)/K)$, $i = 1, \dots, r$. Here μ_N denotes the group of Nth roots of 1. The action of each $\tau \in \Lambda_K$ on the group π^{alg} looks like this:

(3.9) Γ_i^τ is conjugate in π^{alg} to $\Gamma_j^{\chi(\tau)}$ where $x_j = x_i^\tau$.

Now take $K = \mathbb{Q}_p$. It is natural to ask if the Frobenius $F_p \in \Lambda_{\mathbb{Q}_r}$ satisfies an analog of (3.8). One cannot just replace the exponent -1 in (3.8) by the exponent p. Indeed, if this were true, conjugates of Γ_i^p, $i = 1, \ldots, r$ would generate π. This, however, would imply that a group generated by elements of order p would be trivial, a contradiction.

We are not tempted to use the exponent p when we recognize a simple property of the Frobenius F_p. It acts on μ_N as pth powers only when p does not divide N. Question 3.4 below is subtler. Say that a finite extension $L/\overline{\mathbb{Q}}_p(X)$ is p'-ramified if p does not divide any of the orders e_i of the inertia groups above x_i, $i = 1, \ldots, r$. For such extensions, p is relatively prime to $N = \text{lcm}(e_1, \ldots, e_r)$. In this case, the value in $G(K(\mu_N)/K)$ of the cyclotomic character at F_p is p. Define π_p^{alg} to be the projective limit $\varprojlim \pi^{\text{alg}}/D$. Here D ranges over normal subgroups of π of finite index where the field extension corresponding to D is p'-ramified.

Question 3.4. Is the action of the Frobenius F_p on $\tilde{\pi} \cong \pi_p^{\text{alg}}$ induced by an action on π such that $\Gamma_i^{F_p}$ is conjugate in π to Γ_j^p where $x_j = x_i^{F_p}$, $i = 1, \ldots, r$?

We believe the answer is still "No!" Here is an outline in this direction in the case of covers with branch points in $\mathbb{Q}_p$. Such a "frobenius" action would give a formula like this:

$$(3.10) \qquad F_p \sigma_i F_p^{-1} = \omega_i(\boldsymbol{\sigma}) \sigma_i^p \omega_i^{-1}(\boldsymbol{\sigma}), \qquad i = 1, \ldots, r.$$

Here $\omega_i(\boldsymbol{\sigma})$ is a word in the entries of $\boldsymbol{\sigma}$. To regard the formula as similar to that over $\mathbb{R}$ requires some conditions on the words $\boldsymbol{\omega}(\boldsymbol{\sigma})$. At the minimum, they should be independent of considerable data describing the cover.

Suppose we ask that $\boldsymbol{\omega}(\boldsymbol{\sigma})$ be independent of the branch points and the choice of elements in the conjugacy classes given by the entries of $\boldsymbol{\sigma}$. Then, such a formula implies the existence of a correspondence—much like a Hecke correspondence—on the naturally attached *Hurwitz space*. We conclude by showing how this gives a contradiction.

When $r = 4$, consider the observation of [**Fr**, 2; §4.2]. This relates all Hurwitz spaces to curves defined by the action of a subgroup of finite index in $\mathbf{SL}_2(\mathbb{Z})$ on the upper half plane. Our assumptions on $\boldsymbol{\omega}(\boldsymbol{\sigma})$ would imply the existence of an actual nontrivial Hecke theory on these curves. Some of these curves are modular curves, and they have a well known Hecke theory. Still, most are not. For these, this contradicts a result of Atkin [A]: noncongruence subgroup curves have only trivial Hecke correspondences.

REMARK. The existence of a Galois regular extension of $\mathbb{Q}_p(X)$ with group any given group G was proved by Harbater [H]. In this subsection we wanted more. An analog of Lemma 3.3 would be a practical criterion for defining a given cover over $\mathbb{Q}_p$. □

4. Hurwitz spaces and rationality over $\mathbb{Q}$.

4.1. *Reduction of the problem.* Suppose G is a group with an embedding $G \to S_n$. This need not be the regular representation. Let $\alpha_1, \dots, \alpha_s$ be generators for which condition (1.2) holds. Denote a specific cover produced by Comment 2 of §3.5 by $\Psi_{\mathbf{s}(\alpha),\mathbf{x}}: E \to \mathbb{P}^1$. Finally, we assume either

(4.1) $G \to S_n$ is the regular representation or $\mathrm{Cen}_{S_n}(G) = \{1\}$.

From Comment 2 of §3.5, we can define $\Psi_{\mathbf{s}(\alpha),\mathbf{x}}: E \to \mathbb{P}^1$ over $\mathbb{R}$. In this section, we try to descend to $\mathbb{Q}$.

Question 4.1. Is there some choice of branch points $x_1, \dots, x_4$ in $\mathbb{P}^1(\mathbb{R})$ that gives a cover $\Psi_{\mathbf{s}(\alpha),x}: E \to \mathbb{P}^1$ produced by Comment 2 of §3.5 and defined over the rational number field $\mathbb{Q}$.

We use *Nielsen classes* and *Hurwitz families* to investigate this. Branch cycle descriptions provide much information (cf. §2.3 and **[DFr]** §1.1). Still, they depend on many choices: a base point x_0, a labeling of the points in the fiber $\Phi^{-1}(x_0)$, an ordering of the branch points $x_1, \dots, x_r$, and a sample bouquet $\gamma_1, \dots, \gamma_r$. There is an intrinsic notion.

Consider the data attached to any branch cycle description $(s_1, \dots, s_r)$ of a cover. Most importantly, there is the group $\langle \mathbf{s} \rangle$ generated by the s_is. Up to conjugation by S_n, this is the monodromy group of the cover. Secondly, there is the collection $\{C_1, \dots, C_r\}$ of conjugacy classes of $s_1, \dots, s_r$ in the group $\langle \mathbf{s} \rangle$. From Lemma 1 of **[Fr1]**, up to conjugation by S_n, this data is an invariant of the cover. This observation gives the definition of the Nielsen class of a cover.

Let G be a subgroup of S_n and let $\mathbf{C} = (\mathrm{C}_1, \dots, \mathrm{C}_r)$ be an r-tuple of nontrivial (not necessarily distinct) conjugacy classes of G.

DEFINITION 4.2. To the data $(G, \mathbf{C})$ we associate its *Nielsen class*:

$$\mathrm{ni}(\mathbf{C}) = \{\mathbf{s} \in G^r | \langle \mathbf{s} \rangle = G,\ s_1 \cdots s_r = 1 \text{ and there exists } \omega \in S_r,\ s_{(i)\omega} \in C_i,\ i = 1, \dots, r\}.$$

Suppose a cover $\Psi: E \to \mathbb{P}^1$ has any branch cycle description $\mathbf{s}$, up to conjugation by elements of S_n, in $\mathrm{ni}(\mathbf{C})$. We say the cover is in $\mathrm{ni}(\mathbf{C})$. Alternatively, $\mathrm{ni}(\mathbf{C})$ is the Nielsen class of the cover. The order in which we list the conjugacy classes does not matter. The *straight Nielsen class* of $(\mathbf{C}, G)$ is

$$\mathrm{sni}(\mathbf{C}) = \{\mathbf{s} \in \mathrm{ni}(\mathbf{C}) | s_i \in C_i, \ i = 1, \ldots, r\}.$$

We speak of a cover $\Psi: E \to \mathbb{P}^1$ with an ordering of its branch points being in $\mathrm{sni}(\mathbf{C})$. This means, up to conjugation by elements of S_n, that any branch cycle description of the cover with this ordering is in $\mathrm{sni}(\mathbf{C})$. The normalizer (resp., the straight normalizer) of the Nielsen class is

$$N(\mathbf{C}) = \{\kappa \in S_n | \text{conjugation by } \kappa \text{ permutes } \mathrm{C}_1, \ldots, \mathrm{C}_r\},$$
$$SN(\mathbf{C}) = \{\kappa \in S_n | \text{conjugation by } \kappa \text{ fixes } \mathrm{C}_1, \ldots, \mathrm{C}_r\}.$$

Note that $N(\mathbf{C})$ acts on the Nielsen class $\mathrm{ni}(\mathbf{C})$ by conjugation: $\kappa \in N(\mathbf{C})$ maps $s \in \mathrm{ni}(\mathbf{C})$ to $\kappa s \kappa^{-1} \in \mathrm{ni}(\mathbf{C})$. Similarly, $SN(\mathbf{C})$ acts on the straight Nielsen class $\mathrm{sni}(\mathbf{C})$. Denote the quotients of these actions by $\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$, $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$, the *absolute Nielsen classes.*

Under certain assumptions, there is a space representing a solution to a natural *moduli problem.* This is the problem of parametrizing equivalence classes of covers in a given Nielsen class. *Hurwitz monodromy action* interprets properties of this moduli space. We explain the monodromy action.

Consider the free group on r generators, Q_i, $i = 1, \ldots, r-1$, with these relations:

(4.2) (a) $Q_iQ_{i+1}Q_i = Q_{i+1}Q_iQ_{i+1}$, $i = 1, \ldots, r-2$;
(b) $Q_iQ_j = Q_jQ_i$, $|i-j| > 1$; and
(c) $Q_1Q_2 \cdots Q_{r-1}Q_{r-1} \cdots Q_1 = 1.$

This group, a quotient of the Artin braid group **[Bo]**, is called the *Hurwitz monodromy group* of degree r. We denote it by H_r. The Q_is act on $\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$ by this formula: for $\mathbf{s} \in \mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$

(4.3) $$(\mathbf{s})Q_i = (s_1, \ldots, s_{i-1}, s_is_{i+1}s_i^{-1}, s_i, s_{i+2}, \ldots, s_r), \quad i = 1, \ldots, r-1.$$

Thus they induce a permutation representation of H_r on $\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$: the Hurwitz monodromy action on the Nielsen class $\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$.

Denote the kernel of the natural permutation representation $H_r \to S_r$ sending Q_i to the 2-cycle $(i\,i{+}1)$ by SH_r. This is the *straight Hurwitz monodromy group.* The group SH_r acts on the straight Nielsen class $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. The next statement summarizes the basic moduli space properties in the special case that all of the conjugacy classes are *rational* ([**Fr1**; §4 and 5] or [**DFr**; §1]). (A conjugacy class is rational if it is closed under putting elements to powers relatively prime to the order of elements in the class.)

THEOREM 4.3. *Assume that* (4.1) *holds, that* G *has no center, that* SH_r *acts transitively on* $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$, *and that* $\mathrm{C}_1, \ldots, \mathrm{C}_r$ *are rational conjugacy classes. Then there is an algebraic family* $\mathscr{F}(\mathbf{C})$ *of covers of* $\mathbb{P}^1$ *(a priori over* $\mathbb{C}$*)*

$$\mathscr{F}(\mathbf{C})\colon \mathscr{T}(\mathbf{C}) \to \mathscr{H}(\mathbf{C}) \times \mathbb{P}^1.$$

This universal Hurwitz family associated to $\mathrm{ni}(\mathbf{C})$ *satisfies* (4.4)–(4.7).

(4.4) $\mathscr{F}(\mathbf{C})$ is a finite morphism of quasiprojective varieties, $\mathscr{H}(\mathbf{C})$ is irreducible and the generic fiber of $\mathrm{pr}_1 \circ \mathscr{F}(\mathbf{C})\colon \mathscr{T}(\mathbf{C}) \to \mathscr{H}(\mathbf{C})$ is irreducible.

(4.5) The family $\mathscr{F}(\mathbf{C})$ is defined over $\mathbb{Q}$.

(4.6) Each cover $\Psi\colon E \to \mathbb{P}^1$ in the Nielsen class $\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}$ is equivalent to a unique fiber cover $\mathscr{F}(\mathbf{C})_{\mathbf{h}}\colon \mathscr{T}(\mathbf{C})_{\mathbf{h}} \to \mathbb{P}^1$ (with $\mathbf{h} \in \mathscr{H}(\mathbf{C})$) of the family $\mathscr{F}(\mathbf{C})$. Also, $\mathscr{F}(\mathbf{C})_{\mathbf{h}}\colon \mathscr{T}(\mathbf{C})_{\mathbf{h}} \to \mathbb{P}^1$ is defined over $\mathbb{Q}(\mathbf{h})$, the field of definition of the point $\mathbf{h}$ on the algebraic variety $\mathscr{H}(\mathbf{C})$; $\mathbb{Q}(\mathbf{h})$ is the smallest field of definition for a cover that is equivalent to the cover $\Psi\colon E \to \mathbb{P}^1$.

(4.7) Denote the subvariety of $(\mathbb{P}^1)^r$ consisting of r-tuples with distinct coordinates by U^r. Then, consider the algebraic variety $U^r/S_r = U_r$ given by the quotient action of S_r. The "branch point reference map" $\Psi(\mathbf{C})\colon \mathscr{H}(\mathbf{C}) \to U_r$ sends each $\mathbf{h} \in \mathscr{H}(\mathbf{C})$ to the branch point set of the fiber cover $\mathscr{F}(\mathbf{C})_{\mathbf{h}}\colon \mathscr{T}(\mathbf{C})_{\mathbf{h}} \to \mathbb{P}^1$. This is an étale morphism of degree $|\mathrm{ni}(\mathbf{C})^{\mathrm{ab}}|$ defined over $\mathbb{Q}$.

The original conjugacy classes, $\mathrm{C}_1, \ldots, \mathrm{C}_r$, are the conjugacy classes in G of the entries of the r-tuple $\mathbf{s}(\alpha)$. Theorem 4.3 has this consequence.

PROPOSITION 4.4. *Assume the hypotheses of Theorem* 4.3. *The answer to Question* 4.1 *is yes if and only if there are branch points,* $x_1, \dots, x_r \in \mathbb{P}^1(\mathbb{Q})$, *so that the point* $\mathbf{h} \in \mathscr{H}(\mathbf{C})$ *that corresponds to the cover* $\Psi_{\mathbf{s}(\alpha), x}: E \to \mathbb{P}^1$ *is a* $\mathbb{Q}$*-rational point on* $\mathscr{H}(\mathbf{C})$.

4.2. *Description of* $\mathscr{H}(\mathbf{C})$ *for* $r = 4$. See [**BFr**; §1, Lemma 1.6)], [**Fr2**; §4.1]. Both our examples will be 4 branch point situations. In this case, $\mathscr{H}(\mathbf{C})$ has a more explicit description. Consider natural map $U^r \to U_r$. Let $\mathscr{H}(\mathbf{C})'$ be an irreducible component of the fiber product $\mathscr{H}(\mathbf{C}) \times_{U_r} U^r$ and $p: \mathscr{H}(\mathbf{C})' \to U^r$ the natural projection. Theorem 4.5 uses the permutations of $\mathrm{sni}(\mathbf{c})^{\mathrm{ab}}$ induced by these elements of SH_r: Q_1^2; $Q_1^{-1}Q_2^2Q_1$; $Q_1^{-1}Q_2^{-1}Q_3^2Q_2Q_1$. Denote these by a_{12}, a_{13}, a_{14}, respectively. These act on $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. The transitivity hypothesis of Theorem 4.3 implies that the a_{1j}s are transitive on $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$.

THEOREM 4.5. *For each* $(x_2, x_3, x_4) \in U^3$, *denote the inverse image* $p^{-1}(\mathbb{P}^1 \times (x_2, x_3, x_4))$ *by* $\mathscr{H}(\mathbf{C})'(x_2, x_3, x_4)$. *Composition of* p *with projection* $U^r \to \mathbb{P}^1$ *on the first factor gives an unramified cover*

$$\mathscr{H}(\mathbf{C})'(x_2, x_3, x_4) \to \mathbb{P}^1 \backslash \{x_2, x_3, x_4\}.$$

Complete this to a (ramified) cover $C(\mathbf{C}) \to \mathbb{P}^1$ *of projective nonsingular curves. This will have the following properties.*

(4.8) x_2, x_3, x_4 *are the* 3 *branch points of the cover.*

(4.9) (a_{12}, a_{13}, a_{14}) *(acting on* $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$*) is a branch cycle description of the cover.*

(4.10) *The cover is defined over* $\mathbb{Q}$.

COROLLARY 4.6. *The variety* $\mathscr{H}(\mathbf{C})'$ *is birational to* $C(\mathbf{C}) \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$.

Proof. For (x_2, x_3, x_4) take the generic point of U^3 in the above. The birational equivalence $\mathscr{H}(\mathbf{C})'(x_2, x_3, x_4) \cong C(\mathbf{C})$ induces a birational map $\mathscr{H}(\mathbf{C})' \to C(\mathbf{C}) \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. □

Section 4.3 has examples where $C(\mathbb{C})$ is $\mathbb{P}^1$ (over $\mathbb{Q}$). Consequently, the space $\mathscr{H}(\mathbf{C})'$ is a $\mathbb{Q}$-rational variety. In particular, the $\mathbb{Q}$-rational points on $\mathscr{H}(\mathbf{C})'$ form a dense subset of $\mathscr{H}(\mathbf{C})'(\mathbb{R})$ (for the complex topology) and Question 4.1 has an affirmative answer.

4.3. *A formula for the genus of the curve* $C(\mathbb{C})$. The Riemann-Hurwitz formula gives the genus $g(\mathbf{C})$ of the curve $C(\mathbf{C})$ (cf. Theorem 4.5):

$$\text{(4.11)} \quad \operatorname{ind}(a_{12}) + \operatorname{ind}(a_{13}) + \operatorname{ind}(a_{14}) = 2(N + g(C) - 1) \quad \text{with } N = |\operatorname{sni}(\mathbf{C})^{\mathrm{ab}}|.$$

Here is how we compute $\operatorname{ind}(a_{1j})$. Denote the length of the orbit of $\mathbf{s} \in \operatorname{sni}(\mathbf{C})^{\mathrm{ab}}$ under a_{1j} by $i_{1j}(\mathbf{s})$, $j = 1, 2, 3$. Then

$$\text{(4.12)} \qquad \operatorname{ind}(a_{1j}) = \sum_{\mathbf{s} \in \operatorname{sni}(\mathbf{C})^{\mathrm{ab}}} \frac{i_{1j}(\mathbf{s}) - 1}{i_{1j}(\mathbf{s})}.$$

Check easily that

$$\text{(4.13)} \quad \begin{aligned} (\mathbf{s})a_{12} &= ((s_1s_2)s_1(s_1s_2)^{-1}, s_1s_2s_1^{-1}, s_3, s_4) \\ &= (s_1, s_2, (s_1s_2)^{-1}s_3(s_1s_2), (s_1s_2)^{-1}s_4(s_1s_2)) \\ &\qquad (\text{in } \operatorname{sni}(\mathbf{C})^{\mathrm{ab}}). \end{aligned}$$

Thus, a_{12} acts by conjugation by s_1s_2 on the third and fourth components and leaves the others unchanged. It follows that $(\mathbf{s})(a_{12})^q = \mathbf{s}$ in $\operatorname{sni}(\mathbf{C})^{\mathrm{ab}}$ if and only if

$$\text{(4.14)} \quad \begin{aligned} &(s_1, s_2, (s_1s_2)^{-q}s_3(s_1s_2)^q, (s_1s_2)^{-q}s_4(s_1s_2)^q) \\ &\quad = \kappa(s_1, s_2, s_3, s_4)\kappa^{-1} \end{aligned}$$

for some $\kappa \in SN(\mathbf{C})$. For any subset A of $G = \langle \mathbf{s} \rangle$, denote the centralizer of A in $SN(\mathbf{C})$ by $Z(A)$. Then, condition (4.14) is equivalent to this:

(4.15) There exists $\gamma \in Z(s_1, s_2)$ such that $\gamma(s_1s_2)^{-q} \in Z(s_3)$.

Hence, $i_{12}(\mathbf{s})$ is the smallest integer $q > 0$ with $(s_1s_2)^{-q} \in Z(s_1,s_2)Z(s_3)$. Therefore, the factor group $\langle s_1s_2\rangle/\langle s_1s_2\rangle \cap Z(s_1,s_2)Z(s_3)$ has order $i_{12}(\mathbf{s})$. Similarly, check that

$$\begin{aligned} (\mathbf{s})a_{12} &= ((s_2s_4)^{-1}s_1(s_2s_4), s_2, (s_4s_2)^{-1}s_3(s_4s_2), s_4), \quad \text{and} \\ (\mathbf{s})a_{14} &= (s_1, (s_4s_1)^{-1}s_2(s_4s_1), (s_4s_1)^{-1}s_3(s_4s_1), s_4) \quad (\text{in } \operatorname{sni}(\mathbf{C})^{\mathrm{ab}}). \end{aligned}$$

Thus, the integer $i_{13}(\mathbf{s})$ (resp. $i_{14}(\mathbf{s})$) is the smallest integer $q > 0$ such that $(s_4s_2)^q \in Z(s_2, s_4)Z(s_3)$ (resp., $(s_4s_1)^q \in Z(s_1, s_4)Z(s_3)$). Finally, we get

$$\text{(4.16)} \quad \begin{aligned} i_{12}(\mathbf{s}) &= |\langle s_1s_2\rangle / \langle s_1s_2\rangle \cap Z(s_1, s_2)Z(s_3)|, \\ i_{13}(\mathbf{s}) &= |\langle s_4s_2\rangle / \langle s_4s_2\rangle \cap Z(s_4, s_2)Z(s_3)|, \\ i_{14}(\mathbf{s}) &= |\langle s_4s_1\rangle / \langle s_4s_1\rangle \cap Z(s_4, s_1)Z(s_3)|. \end{aligned}$$

THEOREM 4.7. *Assume the hypotheses of Theorem* 4.3 *and Theorem* 4.5. *Then,* (4.11) *gives the genus* $g(\mathbf{C})$, *where* (4.12) *and* (4.16) *give* $\mathrm{ind}(a_{12})$, $\mathrm{ind}(a_{13})$ *and* $\mathrm{ind}(a_{14})$.

4.4. *Symmetric groups.* In this section, $n = 2p+1$ is an odd prime and the group G is the symmetric group S_n embedded in itself. Condition (4.1) holds. Consider the following involutions of S_n:

$$\alpha_1 = (2\,n-1)(3\,n-2)\cdots(p-1\,p+3)(p\,p+2);$$
$$\alpha_2 = (1\,n)(2\,n-1)(3\,n-2)\cdots(p-1\,p+3)(p\,p+2);$$
$$\alpha_3 = (1\,n-1)(2\,n-2)(3\,n-3)\cdots(p-1\,p+2)(p\,p+1).$$

Since these generate a transitive subgroup of S_n, it is easy to see that they generate all of S_n. Indeed, as n is a prime, the representation is primitive. It is well known that a primitive subgroup of S_n containing a 2-cycle is all of S_n. As $\alpha_1\alpha_2$ is a 2-cycle, we are done. Therefore, condition (1.2) is satisfied.

Here is the 4-tuple $\mathbf{s}(\alpha) = (s_1, s_2, s_3, s_4)$ of (3.2):

$$s_1 = \alpha_1 = (2\,n-1)(3\,n-2)\cdots(p-1\,p+3)(p\,p+2);$$
$$s_2 = \alpha_1\alpha_2 = (1\,n);$$
$$s_3 = \alpha_2\alpha_3 = (n\,n-1\cdots 2\,1);$$
$$s_4 = \alpha_3 = (1\,n-1)(2\,n-2)(3\,n-3)\cdots(p-1\,p+2)(p\,p+1).$$

Order C_1, C_2, C_3, C_4 so they respectively denote the conjugacy classes of s_4, s_1, s_2, s_3. Thus $(s_1, s_2, s_3, s_4) \in \mathrm{ni}(\mathbf{c})^{\mathrm{ab}}$ and $(s_4, s_1, s_2, s_3) \in \mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. Specifically, we have: $C_1 = \{$products of p disjoint 2-cycles$\}$; $\mathrm{C}_2 = \{$products of $p-1$ disjoint 2-cycles$\}$; $\mathrm{C}_3 = \{$2-cycles$\}$; $\mathrm{C}_4 = \{n$-cycles$\}$. Any conjugacy class in S_n is rational. In particular, these are.

We now investigate the Hurwitz monodromy action on $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. First, a lemma helps us list the elements in $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. In the following, for $s, \omega \in S_n$, we let s^ω denote the conjugate of s under ω (i.e., $s^\omega = \omega^{-1}s\omega$). For $i \in \{1, \ldots, n\}$, i^ω is the integer $(i)\omega$.

LEMMA 4.8. *Let* $a, b \in S_n$ *be involutions. Let* O *be a disjoint cycle in* ab *that contains an integer* ρ_0 *fixed by* b. *There are two possibilities.*

(i) $O = (\rho_0\rho_1\cdots\rho_t\rho_t^b\cdots\rho_1^b)$ *with* $t \geq 0$ *and none of the integers* ρ_i, $i > 0$, *fixed by* b; ρ_t *is then fixed by* a *and* O *is a cycle of odd length.*

(ii) $O = (\rho_0\rho_1 \cdots \rho_t\rho_0^*\rho_t^b \cdots \rho_1^b)$ *with* $t \geq 0$ *and none of the integers* ρ_i, $i > 0$, *fixed by* b; ρ_0^* *is then fixed by* b *and* O *is a cycle of even length.*

Conversely, we have these partial products from Ob.

(i′) $(\rho_0\rho_1 \cdots \rho_t\rho_t^b \cdots \rho_1^b)(\rho_1\rho_1^b)\cdots(\rho_t\rho_t^b)$ *is a product of* t *disjoint 2-cycles.*

(ii″) $(\rho_0\rho_1 \cdots \rho_t\rho_0^*\rho_t^b \cdots \rho_1^b)(\rho_1\rho_1^b)\cdots(\rho_t\rho_t^b)$ *is a product of* $t+1$ *disjoint 2-cycles.*

Proof. Conjugation by b turns ab into $(ab)^{-1}$. Therefore, $(O^b)^{-1}$ is a disjoint cycle in ab. Since O and $(O^b)^{-1}$ have an integer in common, namely ρ_0, we obtain $O = (O^b)^{-1}$. The only cycles with that property are those described in statement (i) and (ii) of Lemma 4.8. The converse statements (i′) and (ii′) are immediate. □

We now show there is a one-to-one correspondence between the elements of $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ and the subset S of $\mathbf{N}^3$ of triples $[\mu, \beta, \gamma]$ satisfying

$$1 \leq \mu \leq p; \quad 1 \leq \beta \leq 2\mu - 1; \quad p + \mu + 1 \leq \gamma \leq n.$$

Start with this observation. Every element of the absolute straight Nielsen class $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ has a unique representative $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ with $\sigma_4 = (n\, n-1 \cdots 1)$ and $\sigma_3 = (1\, 2\mu)$, $\mu \in \{1, \ldots, p\}$. Existence is easy. Lemma 4.9 below (and $\mathrm{Cen}(S_n) = \{1\}$) gives uniqueness.

LEMMA 4.9. *The group* S_n *is generated by* σ_3 *and* σ_4.

Proof. Consider a partition I of $\{1, \ldots, n\}$. We say that I is a set of *imprimitivity* for a subgroup H of S_n, if H permutes the elements of I. Sets of imprimitivity of the n-cycle $(n\, n-1 \cdots 1)$ are the cosets modulo a nontrivial divisor of n. Since n is prime, $\langle \sigma_3, \sigma_4 \rangle$ is a primitive subgroup of S_n, which contains a 2-cycle. Therefore, it is all of S_n. □

For the representative $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ above, we obtain

$$\sigma_1\sigma_2 = (\sigma_3\sigma_4)^{-1} = (1\, 2 \cdots 2\mu - 1)(2\mu\, 2\mu + 1 \cdots n).$$

Both σ_1 and σ_2 are of order 2 and σ_2 fixes 3 integers. Lemma 4.8 shows that only one of these integers, say β, occurs in the odd length cycle $(1\, 2 \cdots 2\mu - 1)$ of $\sigma_1\sigma_2$. The two other integers fixed by σ_2 appear in the even length cycle $(2\mu\, 2\mu + 1 \cdots n)$. Denote the integer fixed by σ_2 that is in the second half of $\{2\mu, 2\mu + 1, \ldots, n\}$ by γ.

That is, γ is in the set $\{p+\mu+1, \dots, n\}$. This defines a triple $[\mu, \beta, \gamma]$ which lies in the set S. The next proposition gives us the genus of covers with branch cycles coming from our previous lemmas. In §4.5 we draw conclusions from this about Question 4.1. Since the result is not terribly positive, §4.6 makes further comment on what we can expect from variations of this technique.

PROPOSITION 4.10. *The map* $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}} \to S$ *that assigns to each element of* $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ *the triple* $[\mu, \beta, \gamma]$ *defined above is one-one and onto. In particular,*

$$|\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}| = \sum_{1\le\mu\le p} (2\mu-1)(p-\mu+1) = \frac{p(p+1)(2p+1)}{6}.$$

Proof. Let $[\mu, \beta, \gamma]$ be a triple in S. Set $\sigma_4 = (n\, n-1 \cdots 1)$ and $\sigma_3 = (1\, 2\mu)$. We need to show that there is a unique pair (σ_1, σ_2) with these properties:

(4.16) $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4) \in \mathrm{sni}(\mathbf{C})$ and σ_2 fixes β and γ.

Existence. One has $(\sigma_3\sigma_4)^{-1} = (1\, 2 \cdots 2\mu{-}1)(2\mu\, 2\mu{+}1 \cdots n)$. Using Lemma 4.8 (i′) and (ii′), write $(1\, 2 \cdots 2\mu-1) = a'b'$ with a' and b' products of $(\mu-1)$ 2-cycles with support in $\{1, 2, \dots, 2\mu-1\}$ and β fixed by b'. Also, $(2\mu\, 2\mu+1 \cdots n) = a''b''$ with a'' and b'' products of respectively $(n{-}2\mu{+}1)/2$ and $(n{-}2\mu{-}1)/2$ 2-cycles with support in $\{2\mu, 2\mu+1, \dots, n\}$ and γ fixed by b''. Take $\sigma_1 = a'a''$ and $\sigma_2 = b'b''$. The 4-tuple $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ has the required properties (4.16).

Uniqueness. σ_1 and σ_2 satisfy

$$\sigma_1\sigma_2 = (1\, 2 \cdots 2\mu-1)(2\mu\, 2\mu+1 \cdots n).$$

From Lemma 4.8 (i) and (ii), $(1\, 2 \cdots 2\mu-1)$ is of the form $(\rho_0\rho_1 \cdots \rho_t\rho_t^{\sigma_2} \cdots \rho_1^{\sigma_2})$ with $\rho_0 = \beta$, and $(2\mu\, 2\mu{+}1 \cdots n)$ is of the form $(\tau_0\tau_1 \cdots \tau_t\tau_0^*\tau_t^{\sigma_2} \cdots \tau_1^{\sigma_2})$ with $\tau_0 = \gamma$. This determines σ_2 on $\{1, 2, \dots, 2\mu-1\}$ and on $\{2\mu, 2\mu+1, \dots, n\}$ (i.e., on all of $\{1, \dots, n\}$). □

In the rest of this section identify each element of $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ with its image in S. The next step consists in computing indices of a_{12}, a_{13}, a_{14} acting on $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$.

Index of a_{12}. Let $\mathbf{s} = [\mu, \beta, \gamma] \in \mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$; the centralizer $Z(s_1, s_2)$ is the subgroup of S_n generated by $(2\mu\, 2\mu+1 \cdots n)^{p-\mu+1}$. Indeed, let

$t \in Z(s_1, s_2)$. Then t commutes with

$$s_1 s_2 = (1\,2 \cdots 2\mu - 1)(2\mu\, 2\mu + 1 \cdots n).$$

Therefore, t is of the form $(1\,2 \cdots 2\mu - 1)^h (2\mu\, 2\mu + 1 \cdots n)^k$. Since t fixes β and permutes the 2 other fixed points of s_2, $h = 0$ and $k = \lambda(p - \mu + 1)$, for some integer λ. Recall from §4.3, the integer $i_{12}(\mathbf{s})$ is the smallest integer $q > 0$ such that, for some integer λ,

$$(1\,2 \cdots 2\mu - 1)^q (2\mu\, 2\mu + 1 \cdots n)^{q - \lambda(p+\mu-1)}$$

commutes with $s_3 = (1\,2\mu)$ (i.e., fixes the pair $\{1, 2\mu\}$).

The two disjoint cycles $(1\,2 \cdots 2\mu - 1)$ and $(2\mu\, 2\mu + 1 \cdots n)$ of $s_1 s_2$ are of relatively prime order. Thus, $i_{12}(\mathbf{s}) = (2\mu - 1)(p - \mu + 1)$. Formulas (4.16) gives this:

$$\begin{aligned}\operatorname{ind}(a_{12}) &= \sum_{1 \le \mu \le p} (2\mu - 1)(p - \mu + 1)\left(1 - \frac{1}{(2\mu - 1)(p - 2\mu + 1)}\right) \\ &= N - p = \frac{p(p-1)(2p+5)}{6}.\end{aligned}$$

Index of a_{13}. Let $\mathbf{s} = [\mu, \beta, \gamma] \in \mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$. We easily see the centralizer $Z(s_2, s_4)$ is trivial. The integer $i_{13}(\mathbf{s})$ is the smallest integer $q > 0$ such that $(s_4 s_2)^q$ fixes the pair $\{1, 2\mu\}$. Let a and b in $\{1, \ldots, n\}$. These observations are helpful:

(4.17)

(i) if $a^{s_4} = b \in \{2\mu, \ldots, n\}$ and $b^{s_2} \ne 2\mu$, then $a(s_4 s_2)^2 = a^{s_3}$;

(ii) if $a^{s_4} = b \in \{1, \ldots, 2\mu - 1\}$ and $b^{s_2} \ne 1$, then $a(s_4 s_2)^2 = a^{s_3}$.

We prove (i)–(ii) is similar. From Lemma 4.8, s_2 fixes the set $\{2\mu, \ldots, n\}$. Thus, $b^{s_2} \in \{2\mu, \ldots, n\}$ and $b^{s_2} \ne 2\mu$. Therefore, $(b^{s_2}) s_3 = b^{s_2}$ and

$$(a)(s_4 s_2)^2 = (b^{s_2}) s_3 s_2 s_1 s_2 = (b) s_1 s_2 = (a) s_3 s_2 s_1 s_1 s_2 = (a) s_3.$$

Let $a = 1$. We have $1^{s_4} = n \in \{2\mu, \ldots, n\}$ and $n^{s_2} \ne 2\mu$. Indeed, from Lemma 4.8 (ii), no two consecutive integers in the even length orbit of $s_1 s_2$ can be images of one another by s_2. The even length orbit of $s_1 s_2$ is $(2\mu\, 2\mu + 1 \cdots n)$. From (4.17)(i), $(1)(s_4 s_2)^2 = 2\mu$. Let $a = 2\mu$. We have $(2\mu)^{s_4} = 2\mu - 1 \in \{1, \ldots, 2\mu - 1\}$. Lemma 4.8 (i) implies $(2\mu - 1)^{s_2} = 1$ if and only if $2\mu - 1 = \rho_t$. Distinguish two cases.

- If $2\mu - 1 \ne \rho_t$, (4.17)(ii) gives $(2\mu)(s_4 s_2)^2 = 1$ and $i_{13}(\mathbf{s}) = 2$. (Note: $i_{13}(\mathbf{s}) \ne 1$ because $(1)(s_4 s_2) = n^{s_2} \ne 1, 2\mu$.)

- •• If $2\mu - 1 = \rho_t$, we obtain $(2\mu)(s_4 s_2)^3 = 2\mu$ and $(1)(s_4 s_2)^3 = 1$. Hence $i_{13}(\mathbf{s}) = 3$.

The number of occurrences of •• is

$$\sum_{1 \le \mu \le p} (p - \mu + 1) = \frac{p(p+1)}{2},$$

$p - \mu + 1$ for each value of μ. Therefore, $\operatorname{ind}(a_{13}) = p(p+1)^2/6$.

Index of a_{14}. Let $\mathbf{s} = [\mu, \beta, \gamma] \in \operatorname{sni}(\mathbf{C})^{\mathrm{ab}}$. Again, the centralizer $Z(s_1, s_4)$ is trivial. The integer $i_{14}(\mathbf{s})$ is the smallest integer $q > 0$ such that $(s_4 s_1)^q = (s_2 s_3)^{-q}$ fixes the pair $\{1, 2\mu\}$. The calculation depends on the intersection set $\{1, 2\mu\} \cap \{1^{s_2}, (2\mu)^{s_2}\}$. Note: By construction of $[\mu, \beta, \gamma]$, $1 \le 1^{s_2} \le 2\mu - 1$ and $2\mu \le (2\mu)^{s_2} \le n$. So we only have 4 cases to consider.

1st case. $1^{s_2} = 1$ and $(2\mu)^{s_2} = 2\mu$. That is, $\mathbf{s} = [\mu, 1, \mu + p + 1]$. Here, $i_{14}(\mathbf{s}) = 1$.

2st case. $1^{s_2} \ne 1$ and $(2\mu)^{s_2} = 2\mu$. That is, $\mathbf{s} = [\mu, \beta, \mu + p + 1]$ with $\beta \ne 1$. Here, $(2\mu)(s_2 s_3)^3 = 2\mu$ and therefore, $(1)(s_2 s_3)^3 = 1$. Thus $i_{14}(\mathbf{s}) = 3$. (Note that $i_{14}(\mathbf{s}) \ne 1$ since $1^{s_2} \ne 1, 2\mu$.)

3rd case. $1^{s_2} = 1$ and $(2\mu)^{s_2} \ne 2\mu$. That is, $\mathbf{s} = [\mu, 1, \gamma]$ with $\gamma \ne \mu + p + 1$. This is exactly as in the 2nd case.

4th case. $1^{s_2} \ne 1$ and $(2\mu)^{s_2} \ne 2\mu$. Here, $(1)(s_2 s_3)^2 = 2\mu$ and thus, $(2\mu)(s_2 s_3)^2 = 1$. Therefore, $i_{14}(\mathbf{s}) = 2$.

We have only to count the possibilities for $\mathbf{s}$ in each case: p for the first case, $(2\mu - 2)$ for each μ of the second case, $(p - \mu)$ for each μ for the third case, and the rest for the fourth case. The result:

$$\operatorname{ind}(a_{14}) = \sum_{1 \le \mu \le p} \frac{2}{3}(\mu + p - 2) + \frac{1}{2}\left[N - p - \sum_{1 \le \mu \le p} (\mu + p - 2)\right].$$

Finally, $\operatorname{ind}(a_{14}) = \frac{p(p-1)(p+4)}{6}$.

Now (4.11) gives the genus $g(\mathbf{C})$ of $C(\mathbf{C})$ (cf. Theorem 4.5):

$$g(\mathbf{C}) = \frac{(p-2)(p-3)}{6}.$$

Thus, Question 4.1 has a positive answer for $p = 2, 3$ (i.e., $n = 5, 7$). There is one condition, however, in Theorem 4.3 we have not checked yet: transitivity of SH_r on $\operatorname{sni}(\mathbf{C})^{\mathrm{ab}}$. We use two steps.

From (4.13), $a_{12} = Q_1^2$ conjugates by $s_1 s_2$ on the first two components of the 4-tuple $\mathbf{s}$ and leaves the others unchanged. So for

$\mathbf{s} = [\mu, \beta, \gamma]$, we obtain:

$$[\mu, \beta, \gamma]a_{12} = [\mu, (\beta)(s_1s_2)^{-1}, (\gamma)(s_1s_2)^{-1}].$$

Still, the two disjoint cycles $(1\,2\cdots 2\mu-1)$ and $(2\mu\,2\mu+1\cdots n)$ of s_1s_2 are of relatively prime order. Therefore, the group generated by s_1s_2 acts transitively on the ordered pairs (β, γ) with $1 \le \beta \le 2\mu-1$ and $p+\mu+1 \le \gamma \le n$. Conclude that the orbits of a_{12} are the p subsets of $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ corresponding to each value of μ.

Now consider a_{13}. We are done if we show that for any $\mu = 1, \ldots, p$, a_{13} sends some element $[1, 1, \gamma]$ to some element $[\mu, \beta', \gamma']$. For $\mathbf{s} = [1, 1, \gamma]$, a_{13} leaves s_2 and s_4 unchanged and turns $s_3 = (1\,2)$ into

$$(s_4s_2)^{-1}s_3(s_4s_2) = (1^{s_4s_2}2^{s_4s_2}) = (n^{s_2}\,1).$$

That is, a_{13} sends some element $[1, 1, \gamma]$ on some element $[\mu, \beta', \gamma']$ with $(1\,2\mu) = (1\,n^{s_2})$, up to conjugation by a power of s_4.

We have $s_1s_2 = (1)(2\,3\cdots n)$. Lemma 4.8 (ii) implies the cycle $(2\,3\cdots n)$ has form $(\rho_0\rho_1\cdots\rho_t\rho_0^*\rho_t^b\cdots\rho_1^b)$ with $\rho_0 = \gamma$. Check: when γ ranges over $\{p+2, \ldots, n\}$, n^{s_2} takes on all values in $\{3, 5, \ldots, n\}$. That is, 2μ takes on all values in $\{2, 4, \ldots, n-1\}$. □

4.5. *Conclusions from* §4.4 *Example.*

THEOREM 4.11. *For $n = 5, 7$, S_n is the Galois group of a regular extension $E/\mathbb{Q}(T)$ with these properties:*

(i) *$E/\mathbb{Q}(T)$ is ramified over 4 rational points*; *and*
(ii) *for all t in a nonempty interval of the real line, the residue class extension $E_t/\mathbb{Q}$ is a totally real extension.*

End of proof. For $n = 5, 7$, (4.18) yields $g(\mathbf{C}) = 0$. Hence, the curve $C(\mathbf{C})$ is $\mathbb{P}^1$ if it has a $\mathbb{Q}$-rational point. The disjoint cycles in the permutation a_{12} of $\mathrm{sni}(\mathbf{C})^{\mathrm{ab}}$ are in 1-1 correspondence with the points over the branch point $x_2 \in \mathbb{P}^1$ in the cover $C(\mathbf{C}) \to \mathbb{P}^1$ of Theorem 4.5. The previous study of a_{12} shows, for $n = 5$ (resp., $n = 7$), there are 2 ramified points (resp., 3 ramified points) over x_2 of ramification indices 1, 3 (resp., 3, 5, 6). Each of these points has a unique ramification index. Thus, these points are rational over $\mathbb{Q}(x_2, x_3, x_4)$.

Consider (y, x_2, x_3, x_4) on $C(\mathbf{C}) \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ with y not lying over one of x_2, x_3, x_4. From Proposition 4.4 and Corollary 4.6, each such $\mathbb{Q}$-rational point corresponds to a cover $\psi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ defined over $\mathbb{Q}$. Equivalently, such a point corresponds to a regular extension $Y/\mathbb{Q}(T)$, with 4 rational branch points and monodromy group S_5 (resp., S_7).

Pick a $\mathbb{Q}$-rational point (y, x_2, x_3, x_4) that corresponds to a cover having the 4-tuple $(\alpha_1, \alpha_1\alpha_2, \alpha_2\alpha_3, \alpha_3)$ as a branch cycle description. (This is with respect to a bouquet as in §2.3.) The $\mathbb{Q}$-points are dense in the space $\mathscr{H}(\mathbf{C})'(\mathbb{R})$. Thus, such a choice of point is possible. Choose x_0 between x_1 and x_4 on the real projective line. The remark in §3.4 shows that the action of complex conjugation is trivial on the fiber $\psi^{-1}(x_0)$. That is, in the notation of §3.4, $\bar{c} = 1$. Let $E/\mathbb{Q}(T)$ be the Galois closure of the extension $Y/\mathbb{Q}(T)$. It is a regular extension with properties (i) and (ii). □

4.6. *Additions to Theorem* 4.11.

Comment (1). The §4.4 method applies to any 3-tuple $(\alpha_1, \alpha_2, \alpha_3)$ of generators of S_n of order 2. For example, we have computed with $n = 2p$ where p is an odd prime and

$$\begin{aligned}\alpha_1 &= (1\,n),\\ \alpha_2 &= (2\,n)(3\,n-1)(4\,n-2)\cdots(p-1\,p+3)(p\,p+2),\\ \alpha_3 &= (1\,n)(2\,n-1)(3\,n-2)\cdots(p-1\,p+2)(p\,p+1).\end{aligned}$$

The associated curve $C(\mathbf{C})$ has genus $g(\mathbf{C}) = \frac{1}{8}(p-3)(p-5)$. That is, the conclusion of Theorem 4.11 holds for $n = 6$ and $n = 10$. It also holds for the special case $n = 4$. Here, take $\alpha_1 = (2\,3)$, $\alpha_2 = (1\,4)(2\,3)$, and $\alpha_3 = (1\,3)$.

Comment (2). There is only one centerless group G for which Theorem 4.11 is true with 3 branch points instead of 4 branch points: $G = S_3$ (**[Se2]**, **[FrD]**). If we allow a center, there are other candidates: the groups $\mathbb{Z}/m \times^s \mathbb{Z}/2$, for $m = 2, 4, 6$. Moreover, the group $\mathbb{Z}/2 \times \mathbb{Z}/2$ does satisfy the conclusions of Theorem 4.11 for 3 branch points.

5. Two further applications. The dihedral group D_m is the easiest non-abelian finite group. The reader must be surprised to hear that there are serious questions about realizing it as a Galois group of a regular extension $L/\mathbb{Q}(x)$. The problem is not realizing the group, it is realizing it with extensions having few branch points. The problem is similar to that of §4: finding rational points on variants of Hurwitz spaces defined over $\mathbb{Q}$ as in §4.5. There we could only proceed when we knew that a certain curve $C(\mathbb{C})$ was of genus 0.

Suppose, however, that the curve is of genus greater than 0. It could still have rational points on it. One rational point was all we needed to conclude realization of the groups with the properties of §4.5. With dihedral groups we can interpret existence of rational points even when the number of branch points is large. We owe this to identifications of the particular Hurwitz spaces with variants on classical modular curves. Section 5.1 gives a definitive result when the number of branch points is less than 6. Section 5.2 considers larger values of r based on generalizations of Mazur's theorem.

Finally, we illustrate a new large field over which we know that all groups are Galois groups of regular extensions. For each prime p, there is a field $\mathbb{Q}^{\mathrm{tp}}$, the *totally p-adic* algebraic numbers. An algebraic number α is in $\mathbb{Q}^{\mathrm{tp}}$ if each conjugate of α is in $\mathbb{Q}_p$, the p-adic numbers. Section 5.3 considers the case of the real valuation.

5.1. *Dihedral groups with r small.* In this section, m is an odd prime. Consider the dihedral group $D_m = \mathbb{Z}/m \times {}^s\mathbb{Z}/2$ in its regular representation. The order of D_m is $n = 2m$. Two involutions generate it.

THEOREM 5.1. *For $m > 7$ a prime, D_m is not the Galois group of a regular extension of $\mathbb{Q}(X)$ with 5 or fewer branch points.*

Proof. Assume that $G = D_m$ is the Galois group of a regular extension $Y/\mathbb{Q}(X)$. Let $\Phi\colon Y_{\mathbb{C}} \to \mathbb{P}^1$ be the associated cover. Take $x_1, \ldots, x_r$ to be an ordering of the branch points. Identify G with the monodromy group of the cover. For $i = 1, \ldots, r$, let C_i be the conjugacy class of the branch cycles associated with x_i. That is, the cover is in $\mathrm{sni}(\mathrm{C}_i, \ldots, \mathrm{C}_r)$. We divide the proof into 2 cases. Let C be the conjugacy class of all involutions in G: $\mathrm{C} = \{(a, 1) | a \in \mathbb{Z}/m\}$.

1*st case. One of* $\mathrm{C}_1, \ldots, \mathrm{C}_r$, *say* C_i, *is different from* C. Let $(a, 0) \in C_i$. This is an element of order m and its nontrivial powers lie in $(m-1)/2$ distinct conjugacy classes of G. We show that $r \geq (m-1)/2 \geq 5$. Indeed, this follows from the rationality properties that the inertia groups inherit from the rationality of the cover. Specifically, apply the branch cycle argument §3.7, expression (3.9) in the following form. The order of C_i is the order of the elements in C_i.

(5.1) For each $i \in \{1, \ldots, r\}$, for all α relatively prime to the order of C_i, $c_i^{\alpha} = C_j$ for some $j \in \{1, \ldots, r\}$.

To complete the first case we show $r \neq 5$. For $r = 5$, $G = D_{11}$ and $C_1, \dots, C_5$ are conjugacy classes of 11-cycles. These classes, however, do not generate D_{11}, a contradiction.

2nd case. $C_1 = \cdots = C_r = C$. Observe that $r \neq 2$ when g is not a cyclic group. Also, that $r \neq 3, 5$; the relation $s_1 \cdots s_r = 1$ implies that r is even. Assume $r = 4$. The Riemann-Hurwitz formula yields the genus g of the cover $\Phi: Y_{\mathbb{C}} \to \mathbb{P}^1$:

$$m + m + m + m = 2(n + g - 1).$$

That is, $g = 1$. In addition, the elliptic curve $Y_{\mathbb{C}}$ has an automorphism χ of order m, for example $(1, 0)$.

Assume first that $Y_{\mathbb{C}}(\mathbb{Q}) \neq \varnothing$: $Y_{\mathbb{C}}$ is an elliptic curve over $\mathbb{Q}$. Translation by a point $\mathbf{p}$ of order m on $Y_{\mathbb{C}}$ gives χ. Since $Y/\mathbb{Q}(X)$ is regular, χ is defined over $\mathbb{Q}$ and $\mathbf{p}$ is a rational point in $Y_{\mathbb{C}}$.

Thus, we have produced an elliptic curve $Y_{\mathbb{C}}$ and a point $\mathbf{p}$ of order m . Both are defined over $\mathbb{Q}$. It is classical that the data $(Y_{\mathbb{C}}, \mathbf{p})$ corresponds to a rational point on the modular curve $X_1(m) \setminus \{\text{cusps}\}$. As $m > 7$, this contradicts Mazur's theorem **[Se1**; Theorem 3] (or **[M]**, **[MS]**).

If $Y_{\mathbb{C}}(\mathbb{Q}) = \varnothing$, the same argument works on the Jacobian $\text{Pic}^0(Y_{\mathbb{C}})$ of $Y_{\mathbb{C}}$. Recall: $\text{Pic}^0(Y_{\mathbb{C}})$ consists of divisor classes of degree 0 on $Y_{\mathbb{C}}$. The automorphism group of $Y_{\mathbb{C}}$ naturally embeds as automorphisms of $\text{Pic}^0(Y_{\mathbb{C}})$. Thus, this is an elliptic curve over $\mathbb{Q}$. And, it has an automorphism of order m defined over $\mathbb{Q}$. Therefore, $r \neq 4$. □

5.2. *Bounding r with dihedral groups.* This subsection discusses Conjecture 5.2.

Conjecture 5.2. Let m run over odd primes. There is no finite r_0 such that each D_m is the group of a regular Galois extension $L/\mathbb{Q}(x)$ with at most r_0 branch points.

Kamienny and Mazur have recent results that approach what we need to show this conjecture [M]. Suppose that such a bound r_0 as in the conjecture exists. The proof of Theorem 5.1 shows we can realize only a finite number of the D_ms under the following conditions. At least one inertia group generator is an m-cycle and there are no more than r_0 branch points. We restate the conjecture as follows.

Conjecture 5.2$'$. Realization of $L/\mathbb{Q}(x)$ with group D_m and all inertia group generators involutions requires more than r_0 branch points if m is suitably large.

We call a Galois realization of D_m over $\mathbb{Q}$ satisfying the condition that all inertia group generators are involutions an *involution realization of* D_m. Consider such an involution realization.

The fixed field T of an automorphism of order m is a degree 2 extension of $\mathbb{Q}(x)$ ramified over r (even) points. Also, L/T is a cyclic unramified extension of degree m. That is, T is the function field of a hyperelliptic curve of genus $\frac{r-2}{2}$.

We want $\varphi: \widehat{X} \to \mathbb{P}^1$ of degree $2m$ with a description of the branch cycles of form $(\sigma_1, \dots, \sigma_r)$. Here, each σ_i is in the conjugacy class C (§5.1) of involutions. A complete combinatorial count of these is easy. At least two of these are not equal (to generate D_m). Write $\sigma_i = (a_i, 1)$. Then, the product of the σs is 1 reduces to $a_1 - a_2 + \cdots - a_r = 0$. Calculations are sufficiently easy to compute elements a_{1j}, $j = 2, \dots, r$ that generalize those in §4.2. Their action on sni($\mathbf{C}$) is transitive. Formula (4.11), with r replacing 4, gives the genus of the analog of $C(\mathbf{C})$. The computation shows this grows quadratically with r when m is fixed. The 1st complex cohomology group of a projective algebraic variety is a birational invariant. Consider the analog for general r of Theorem 4.5. Conclude that the variety $\mathscr{H}(\mathbf{C})'$ for this Nielsen class cannot be *unirational* if r is large. (See Problem 5.6.)

The variety $\mathscr{H}(\mathbf{C})'$ covers the actual variety $\mathscr{H}(\mathbf{C}) = \mathscr{H}(r, m)$ that parametrizes the equivalence classes of covers that we want. Consider $\mathscr{H}(\mathbf{C})'$ as the parameter space for these covers with some ordering on the branch points of the covers. From **[FrV2]** there is a variety $\mathscr{H}(\mathbf{C})^{\text{in}} = \mathscr{H}(r, m)^{\text{in}}$, defined over $\mathbb{Q}$, whose rational points give us the desired extensions. Rational points exactly correspond to regular extensions $L/\mathbb{Q}(x)$ that give involution realizations of D_m. Below we use cover notation. These field extensions correspond to Galois covers $\varphi: \widehat{X} \to \mathbb{P}^1$ defined over $\mathbb{Q}$ with group D_m. Our problem is to decide if $\mathscr{H}(r, m)^{\text{in}}$ has $\mathbb{Q}$ points. We relate $\mathscr{H}(r, m)^{\text{in}}$ to more classical looking objects.

Take $\alpha \in D_m$ of order m. Form $\widehat{X}/\langle\alpha\rangle = Y$, the quotient of $\widehat{X}$ by the group generated by α. The degree 2 cover $Y \to \mathbb{P}^1$ presents Y as a hyperelliptic curve of genus $\frac{r-2}{2}$. Also, $\widehat{X}$ is a cyclic degree m unramified cover of Y. Lemma 5.3 interprets existence of $\widehat{X}$ as a property of $\text{Pic}^0(Y)$, the Picard group of divisor classes of degree 0 on Y. Denote the points of order m on $\text{Pic}^0(Y)$ by $T_m = T_m(Y)$. Then, $G(\overline{\mathbb{Q}}/\mathbb{Q}) = G_{\mathbb{Q}}$ acts on T_m. If $\mathbf{p} \in T_m \setminus \{0\}$ is a point defined over $\mathbb{Q}$, then $G(\overline{\mathbb{Q}}/\mathbb{Q})$ has trivial action $\langle\mathbf{p}\rangle$. When a point has this property, denote the group it generates by $\mathbb{Z}/m$. This says $G_{\mathbb{Q}}$ has trivial action on it.

Similarly, $G_{\mathbb{Q}}$ acts on the mth roots of 1. This is another copy of $\mathbb{Z}/m$, but to show $G_{\mathbb{Q}}$ has a particular nontrivial action on it, denote it by μ_m. Consider the set $G_m(d)$, $d = \frac{r-2}{2}$, of involution realizations of D_m, as above with r branch points, defined over $\mathbb{Q}$. Let $\text{Pic}^1(Y)$ be the Picard space of divisor classes of degree 1 on Y.

LEMMA 5.3. *Continue the notation above. The set of involution realizations of D_m associated to a fixed Y as above naturally inject into the set of $G_{\mathbb{Q}}$ equivariant injections from μ_m into $T_m(Y)$. The image of this map includes all $G_{\mathbb{Q}}$ equivariant injections $\mu_m \to T_m(Y)$ when* $\text{Pic}^1(Y)$ *has a* $\mathbb{Q}$ *point.*

Proof. Consider multiplication by m on $\text{Pic}^0(Y)$. Denote this endomorphism by ψ_m. The kernel is exactly T_m. Since Y consists of positive divisors of degree 1, Y naturally embeds in $\text{Pic}^1(Y)$ (assuming $g(Y) > 0$—that is, $r \geq 4$). Suppose we have an involution realization of D_m attached to Y as above. Universal properties of $\text{Pic}^0(Y)$ produce a natural surjective $G_{\mathbb{Q}}$ equivariant map $T_m(Y) \to \mathbb{Z}/m$. Here $\mathbb{Z}/m$ represents the Galois group of the cover $\widehat{X} \to Y$ as above. Below we show how this gives an injection from μ_m into $T_m(Y)$.

Suppose $\mathbf{q} \in \text{Pic}^1(Y)$ is defined over $\mathbb{Q}$. Define translation $\lambda_{\mathbf{q}}$: $\text{Pic}^1(Y) \to \text{Pic}^0(Y)$ as the map that takes a divisor class $[\mathbf{D}]$ of degree 1 to $[D - \mathbf{q}]$. Denote the image of Y under $\lambda_{\mathbf{q}}$ by $Y_{\mathbf{q}}$. This curve in $\text{Pic}^0(Y)$ is isomorphic to Y over $\mathbb{Q}$. The preimage $\psi_m^{-1}(Y) = Y_{m,\mathbf{q}}$ is the maximal exponent m abelian unramified geometric cover of Y. At least that is correct over $\overline{\mathbb{Q}}$. We cannot expect the automorphisms to be defined over $\mathbb{Q}$.

We want a $G_{\mathbb{Q}}$ invariant hyperplane V in T_m such that the quotient T_m/V is a copy of $\mathbb{Z}/m$. That is, $G_{\mathbb{Q}}$ acts trivially on the quotient. In more homological terms, we want a surjective element $\beta \in \text{Hom}_{G_{\mathbb{Q}}}(T_m, \mathbb{Z}/m) \stackrel{\text{def}}{=} M$. Then, V is the kernel of β.

Conclusion. The quotient $Y_{m,\mathbf{q}}/V \to Y_{m,\mathbf{q}}/T_m = Y_{\mathbf{q}}$ is the cyclic unramified cover we seek. We have identified its automorphism group with $\mathbb{Z}/m$ with trivial $G_{\mathbb{Q}}$ action. That is, the automorphisms are defined over $\mathbb{Q}$. The lemma is complete—from the first paragraph of proof—when we have shown how to go from an injective map β': $\mu_m \to T_m$ to a β above.

The abelian variety $\text{Pic}^0(Y)$ is principally polarized. That means it is isomorphic to its dual abelian variety. This is the abelian variety of linear equivalence classes of divisors on $\mathrm{Pic}^0(Y)$ that are algebraically equivalent to 0. In particular, the Weil pairing produces a nondegenerate symplectic form $w: T_m \times T_m \to \mu_m$ [L]. Thus, $\mathrm{Hom}_{G_\mathbb{Q}}(T_m, \mu_m)$ is isomorphic to T_m as a $G_\mathbb{Q}$ module.

Apply $\mathrm{Hom}_{G_\mathbb{Q}}(\cdot, \mu_m)$ to the map $\beta: \mu_m \to T_m$. This gives

$$\beta: \mathrm{Hom}_{G_\mathbb{Q}}(T_m, \mu_m) \to \mathrm{Hom}_{G_\mathbb{Q}}(\mu_m, \mu_m).$$

The first term identifies to T_m. Check easily that the second term is just $\mathbb{Z}/m$ acting as multiplications. □

REMARK 5.4. When $\mathrm{Pic}^1(Y)(\mathbb{Q})$ is empty. The proof of Lemma 5.3 used a $\mathbb{Q}$ point in $\mathrm{Pic}^1(Y)$ to construct the cover $Y_m \to Y$ canonically. We have not shown that a μ_p point on $\mathrm{Pic}^0(Y)$ produces the Galois sequence of an involution realization of D_m. This is a subtler problem.

We can interpret this as a question on the fibers of a map of the Hurwitz space $\mathscr{H}(\mathbf{C})^{\mathrm{in}} = \mathscr{H}(r, m)^{\mathrm{in}}$ to the space of cyclic order m subgroups of m division points on hyperelliptic jacobians. These fibers are homogeneous spaces for the action of PGL(2). If the image of a fiber is a μ_m point, when does the fiber have a rational point?

We list some *boundedness assertions.* Then, we comment on how these effect Conjecture 5.2.

(1) Let $S(d)$ be primes that are orders of rational points on the elliptic curve defined over some number field K with $[K: \mathbb{Q}] \le d$.

(2) Let $T(d)$ be primes that are orders of rational points on some abelian variety of dimension d over $\mathbb{Q}$.

(3) Let $V(d)$ be primes m that are orders of $G_\mathbb{Q}$ modules isomorphic to μ_m in abelian varieties over $\mathbb{Q}$ of dimension d.

(4) Let $W(d)$ be elements of $V(d)$ from jacobians of hyperelliptic curves of genus d.

The results of [M] include this: $S(d)$ is finite for $d < 9$. In addition, $S(d)$ is of density zero for all d. According to Lemma 5.3, a density 0 result for $V(d)$ would be a satisfactory contribution to Conjecture 5.2. Mazur communicated the following observations.

PROPOSITION 5.5. *We have* $S(d) \subset T(d)$. *Also, if* $m \in V(d)$, *then* $m \in T((m-1)d)$.

Proof. Suppose E is an elliptic curve over K with $[K: \mathbb{Q}] \le d$. Denote the Galois closure of $K/\mathbb{Q}$ by $\widehat{K}$. It is common to call the following formalism, "taking the *Weil trace*" of the elliptic curve over the number field down to $\mathbb{Q}$. Choose a primitive element $\alpha = \alpha_1$ for $K/\mathbb{Q}$. Let $\alpha_1, \ldots, \alpha_d$ be the complete list of conjugates of α_1.

Each conjugate α_i gives a conjugate elliptic curve E_i, defined over $\mathbb{Q}(\alpha_i)$. Let $G = G(\widehat{K}/\mathbb{Q})$ act on $A = E_1 \times E_2 \times \cdots \times E_d$ by permutation of the coordinates. For $\sigma \in G$ indicate this action by $T(\sigma)(A)$. In addition, regard σ as giving a conjugate of A by its action on the coefficients of the equations for A. Call the conjugate A^σ. Thus, for each $\sigma \in G$, the sets $T(\sigma^{-1})(A^\sigma)$ and A are identical. Now apply Weil's cocycle condition to assert that we can define A over $\mathbb{Q}$. To draw the strongest conclusions, we note this construction is universal in the following sense [**FrJ**; Proposition 9.34].

Consider $\mathbb{A}^n$ defined over K. There is a linear map $L\colon \mathbb{A}^{nd} \to \mathbb{A}^n$ defined over K with the following general property. For any subvariety $V \subset \mathbb{A}^n$ defined K, there is a subvariety $W \subset \mathbb{A}^{nd}$ defined over $\mathbb{Q}$ such that $(L_1, L_2, \ldots, L_d)\colon \mathbb{A}^{nd} \to (\mathbb{A}^n)^d$ maps W isomorphically to $V_1 \times \cdots \times V_d$. Here the L_is are the conjugates of L and the V_is are the conjugates of V. This means that we also can apply this to the K subvarieties in V. This produces a $\mathbb{Q}$ rational subvariety of W from the product of their conjugates. Thus, conjugates of a K point $\mathbf{p} \in E$ of order m produce a $\mathbb{Q}$ point of order m on the $\mathbb{Q}$ form of A. From this conclude $S(d) \subset T(d)$.

Now suppose $m \in V(d)$. Apply the Weil trace to $K = \mathbb{Q}(\zeta_m)$ as above to conclude that $m \in T((m-1)d)$. □

Problem 5.6. For each prime m consider the spaces $\mathscr{H}(r, m)^{\text{in}}$ at the beginning of this subsection. Is there a value r_0 such that $\mathscr{H}(r, m)^{\text{in}}$ is unirational over $\mathbb{C}$ for $r > r_0$?

A variety W is unirational if there is a map $\varphi\colon \mathbb{P}^t \to W$ defined on an open subset of $\mathbb{P}^t$ with image a zariski open subset of W. If W and φ are defined over $\mathbb{Q}$, we say W is unirational over $\mathbb{Q}$. Since $\mathbb{P}^t$ has so many rational points, this would imply W has a dense set of rational points. Thus, if Problem 5.6 has an affirmative answer for a given prime m, there are many involution realizations of D_m for an arbitrary prime m. (Although it is not hard to realize D_m as a Galois group of a regular extension of $\mathbb{Q}(X)$.)

5.3. *Descent to the totally real algebraic number field.* Denote the field of all totally real algebraic numbers by $\mathbb{Q}^{\text{tr}}$. These are the algebraic numbers whose complete set of conjugates are real. In this section we prove the following result.

THEOREM 5.7. *Each finite group G is the Galois group of a regular extension of $\mathbb{Q}^{\mathrm{tr}}(X)$.*

Section 4.1 recalls the theory of Hurwitz spaces of covers. **[FrV2]** develops a similar theory, but for G-covers—Galois covers given with their automorphisms. Consider a centerless group G and an r-tuple of conjugacy classes of G. The Hurwitz space $\mathscr{H}^{\mathrm{in}}(G, \mathbf{C})$ is a (reducible) algebraic variety defined over an explicitly computable field $K(\mathbf{C})$. Here is the key property of this space. Let K be a field containing $K(\mathbf{C})$. Then, G-covers in the Nielsen class $\mathrm{ni}(\mathbf{C})$, defined over K, correspond to K-rational points on $\mathscr{H}^{\mathrm{in}}(G, \mathbf{C})$.

Proof of Theorem 5.7. Consider a finite group G. Lemma 2 of **[FrV2]** constructs a cover $G' \to G$ with these properties.

(5.2) The center of G' is trivial and commutators generate the *Schur multiplier* of G'.

We do not explain the commutator statement in (5.2). It appears as a condition in the main theorem of **[FrV2]** which carefully explains it. Suppose we realize G' as a Galois group of a regular extension of $\mathbb{Q}^{\mathrm{tr}}(X)$. Then we automatically realize the quotient G as such a Galois group. Therefore, without loss, assume G satisfies (5.2).

Let b be an integer. Let $\mathrm{C}_1, \dots, \mathrm{C}_s$ be an ordering of nontrivial conjugacy classes of G. Assume each conjugacy class of G appears in this list with the same multiplicity, say m. It is automatic that if we pick g_i out of the conjugacy class C_i, then

(5.3) $$\mathbf{g} = (g_1, \dots, g_s) \text{ generate } G.$$

With $r = 2sb$, consider the r-tuple $\mathbf{C}$

$$(\mathrm{C}_s^{-1}, \dots, \mathrm{C}_1^{-1}, \dots, \mathrm{C}_s^{-1}, \dots, \mathrm{C}_1^{-1}, \mathrm{C}_1, \dots, \mathrm{C}_s, \dots, \mathrm{C}_1, \dots, \mathrm{C}_s).$$

Here, the first sb components are the conjugacy classes $\mathrm{C}_s^{-1}, \dots, \mathrm{C}_1^{-1}$ repeated in this order b times. The last sb components are the conjugacy classes $\mathrm{C}_1, \dots, \mathrm{C}_s$ repeated in this order b times. The Nielsen class $\mathrm{ni}(\mathbf{C})$ is not empty. With $\mathbf{g}$ from (5.3), the r-tuple

$$(g_s^{-1}, \dots, g_1^{-1}, \dots, g_s^{-1}, \dots, g_1^{-1}, g_1, \dots, g_s, \dots, g_1, \dots, g_s)$$

lies in the Nielsen class $\mathrm{ni}(\mathbf{C})$. Observe that all conjugacy classes appear the same number of times, namely $2bm$, in the r-tuple $\mathbf{C}$.

The main theorem of **[FrV2**; Appendix] shows that, if b is suitably large, then $\mathscr{H} = \mathscr{H}^{\text{in}}(G, \mathbf{C})$ is defined over $\mathbb{Q}$ and irreducible over $\overline{\mathbb{Q}}$. This uses (5.2) to apply a theorem of Conway and Parker **[FrV2**; Appendix].

We are left with finding $\mathbb{Q}^{\text{tr}}$-points on the absolutely irreducible variety $\mathscr{H}$. Pop **[P]** proved that every absolutely irreducible variety defined over $\mathbb{Q}^{\text{tr}}$ has $\mathbb{Q}^{\text{tr}}$-points provided it has $\mathbb{R}$-points. This reduces the problem to finding $\mathbb{R}$-points on $\mathscr{H}^{\text{in}}(G, \mathbf{C})$. And their existence follows from Theorem 3.1. Indeed, take $g_0' = 1$ and $r_1 = 1$ in (iii) of condition (b) of Theorem 3.1. This shows that $(g_1, \dots, g_r)$ satisfies the hypotheses of that theorem. □

REMARK. **[FrV]** consists of applications of **[FrV2]**. In particular, this observes that each finite complex extension L of $\mathbb{Q}^{\text{tr}}$ is

P(seudo)A(lgebraically)C(losed)

and Hilbertian. A field P has the PAC property if each absolutely irreducible variety over P has a P-point. The main theorem of **[FrV]** applies to show that the absolute Galois group $G(\overline{\mathbb{Q}}/L)$ is a free profinite group.

On the other hand, $\mathbb{Q}^{\text{tr}}$ is not even Hilbertian. In fact, involutions—conjugates of complex conjugation—generate the absolute Galois group of $\mathbb{Q}^{\text{tr}}$. Thus, Galois extensions of $\mathbb{Q}^{\text{tr}}$ have only groups that are generated by involutions. □

REFERENCES

[A] R. Atkin, *There is no nontrivial Hecke operator theory for noncongruence subgroups, Thompson et al. on non-congruence subgroups*: in *Group Theory*, Cheng et al. eds., Proc. of Singapore Group Theory Conference 1987.

[BK] Bryant and Kovacs, *Lie representations and groups of finite order*, J. London Math. Soc., **17** (1978), 415–221.

[BFr] R. Biggers and M. Fried, *Moduli spaces of covers and the Hurwitz monodromy group*, Crelles J., **335** (1982), 87–121.

[Bo] F. Bohnenblust, *The algebraical braid group*, Ann. of Math., (2) **48** (1947), 127–136.

[CoH] K. Coombes and D. Harbater, *Hurwitz families and arithmetic Galois groups*, Duke Math. J., **52** (1985), 821–839.

[DFr] P. Dèbes and M. Fried, *Arithmetic variation of fibers*, Crelles J., **409** (1990), 106–137.

[D] E. Dew, *Fields of moduli of arithmetic covers*, Thesis (1991).

[Fr1] M. Fried, *Fields of definition of function fields and Hurwitz families-groups as Galois groups*, Comm. Algebra, **5(1)** (1977), 17–82.

[Fr2] ——, *Arithmetic of* 3 *and* 4 *Branch Point Covers*, Sèminaire de Théorie des Nombres, Delange-Pisot-Poitou, Birkhauser (1987/88).

[FrD] M. Fried and P. Dèbes, *Rigidity and real residue class fields*, Acta Arith., **56**, no. 4, (1990), 13–45.
[FrJ] M. Fried and M. Jarden, *Field Arithmetic*, Springer, Ergebnisse, **11** (1986).
[FrV] M. Fried and H. Völklein, *The embedding problem over a Hilbertian PAC-field*, Ann. of Math., **135** (1992), 469–481.
[FrV2] ——, *The inverse Galois problem and rational points on moduli spaces*, Math. Ann., **290** (1992), 771–800.
[FrV3] ——, *Unramified abelian extensions of Galois covers*, Proc. Sympos. Pure Math., **49** (1989), 675–693.
[Gr] A. Grothendieck, *Géométrie formelle et géométrie algébrique*, Seminaire Bourbaki, t. 11 **182** (1958/59).
[Ha] D. Harbater, *Galois Covering of the Arithmetic Line*, Proc. of the NY Number Theory. Conf., LNM 1240, Springer (1985).
[Hup] B. Huppert, *Endliche Gruppen* I, Springer-Verlag, **134** (1967).
[Hur] A. Hurwitz, *Riemann'sche Flächen mit gegebenen Veizweigungspunkten*, Mathematische Werke I, 321–383.
[KM] B. Mazur and S. Kamienny, *Rational torsion of prime order in elliptic curves over number fields*, to appear in Asterisque, Columbia University Number Theory Seminar 1992.
[KN] A. Krull and J. Neukirch, *Die Struktur der absoluten Galois gruppe über dem Körper* $\mathbb{R}(T)$, Math. Ann., **193** (1971), 197–209.
[L] S. Lang, *Abelian Varieties*, Inter. Science Tracts, New York, **7** (1959).
[M] B. Mazur, *Rational Points on Modular Curves*, Lecture Notes in Math., Springer-Verlag, **601** (1977), 107–148.
[MS] B. Mazur and J.-P. Serre, *Points rationnels des courbes modulaires* $X_0(N)$, Séminaire Bourbaki, 27e année (1974/75), Exposé **469**, Lecture Notes in Math., Springer-Verlag, **514** (1976), 238–255.
[P] F. Pop, *Fields of totally* Σ*-adic numbers*, preprint 1991.
[Se1] J.-P. Serre, *Points rationnels des courbes modulaires*, Séminaire Bourbaki, 30ème année, no. 511 (1977/78).
[Se2] ——, *Groupes de Galois sur* $\mathbb{Q}$, Séminarie Bourbaki, no. 589 (1987/88).
[Se3] ——, *Topics in Galois Theory*, Research Notes in Mathematics, Jones and Bartlett, (1992).
[W] A. Weil, *The Field of Definition of a Variety*, Oeuvres Complètes (Collected papers) II, Springer-Verlag, 291–306.

Received April 1, 1992. The first author received support from the University of California (Irvine) and NSF grant DMS-8702150. The second author was supported by BSF grant #87-00038, NSA grant MDA 904-910-H-0057, the Institute for Advanced Studies in Jerusalem and IFR Grant #90/91-15.

University of California, Irvine
Irvine, CA 92717
E-mail address: mfried@math.uci.edu

and

Institue Henri Poincaré
11 Rue P. et M. Curie
75231 Paris, France
E-mail address: pde@ccr.jussieu.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

INTERPOLATED FREE GROUP FACTORS

KEN DYKEMA

The interpolated free group factors $L(\mathbf{F}_r)$ for $1 < r \leq \infty$ (also defined by F. Rădulescu) are given another (but equivalent) definition as well as proofs of their properties with respect to compression by projections and free products. In order to prove the addition formula for free products, algebraic techniques are developed which allow us to show $R * R \cong L(\mathbf{F}_2)$ where R is the hyperfinite II_1-factor.

Introduction. The free group factors $L(\mathbf{F}_n)$ for $n = 2, 3, \dots, \infty$ (introduced in [4]) have recently been extensively studied [**11**, **2**, **5**, **6**, 7] using Voiculescu's theory of freeness in noncommutative probability spaces (see [**8**, **9**, **10**, **11**, **12**, **13**], especially the latter for an overview). One hopes to eventually be able to solve the isomorphism question, first raised by R. V. Kadison of whether $L(\mathbf{F}_n) \cong L(\mathbf{F}_m)$ for $n \neq m$. In [7], F. Rădulescu introduced II_1-factors $L(\mathbf{F}_r)$ for $1 < r \leq \infty$, equalling the free group factor $L(\mathbf{F}_n)$ when $r = n \in \mathbf{N}\backslash\{0, 1\}$ and satisfying

$$(1) \qquad L(\mathbf{F}_r) * L(\mathbf{F}_{r'}) = L(\mathbf{F}_{r+r'}) \qquad (1 < r, r' \leq \infty)$$

and

$$(2) \qquad L(\mathbf{F}_r)_\gamma = L\left(\mathbf{F}\left(1 + \frac{r-1}{\gamma^2}\right)\right) \qquad (1 < r \leq \infty,\ 0 < \gamma < \infty),$$

where for a II_1-factor $\mathscr{M}$, $\mathscr{M}_\gamma$ means the algebra [4] defined as follows: for $0 < \gamma \leq 1$, $\mathscr{M}_\gamma = p\mathscr{M}p$, where $p \in \mathscr{M}$ is a selfadjoint projection of trace γ; for $\gamma = n = 2, 3, \dots$ one has $\mathscr{M}_\gamma = \mathscr{M} \otimes M_n(\mathbf{C})$; for $0 < \gamma_1, \gamma_2 < \infty$ one has

$$\mathscr{M}_{\gamma_1\gamma_2} = (\mathscr{M}_{\gamma_1})_{\gamma_2}.$$

We had independently found the interpolated free group factors $L(\mathbf{F}_r)$ $(1 < r \leq \infty)$ and the formulas (1) and (2), defining them differently and using different techniques. In this paper we give our definition and proofs. This picture of $L(\mathbf{F}_r)$ is sometimes more convenient, e.g. §4 of [**3**]. It is a natural extension of the result [**2**] that

$$(3) \qquad \mathrm{L}(\mathbf{Z}) * R \cong \mathrm{L}(\mathbf{F}_2),$$

where R is the hyperfinite II_1-factor. We also introduce some elementary algebraic techniques for freeness which have further application in [3]. One consequence of them that we prove here is that $R * R \cong L(\mathbf{F}_2)$.

This paper has four sections. In §1 we state a random matrix result (from [2], [7]) and some consequences; in §2 we define the interpolated free group factors and prove the formula (2); in §3 we develop the algebraic techniques; in §4 we prove the addition formula (1) and also make an observation from (1) and (2) (also observed in [7]) that, as regards the isomorphism question, we must have one of two extremes. Our original proof of the addition formula (1) was a fairly messy application of the algebraic techniques developed in §3. The proof of Theorem 4.1 that appears here, while still using the algebraic techniques in an essential way, benefits significantly from ideas found in the proof of F. Rădulescu [7].

1. The matrix model. Voiculescu, as well as developing the whole notion of freeness in noncommutative probability spaces, had the fundamental idea of using Gaussian random matrices to model freeness, which he developed in [12]. In [2], we extended this matrix model to the non-Gaussian case and also to be able to handle semicircular families together with a free finite dimensional algebra. As Rădulescu observed in [7], the matrix model necessary to be able to handle the free finite dimensional algebra can be easily proved in the Gaussian case directly using Voiculescu's methods (cf. the appendix of [2]). In any case, we shall use this matrix model in this paper, and quote it here, as well as some results of it. Our notation for random matrices will be as in [2]. A trivial reformulation of Theorem 2.1 of [2] gives

THEOREM 1.1. *Let $Y(s, n) \in M_n(L)$ for $s \in S$ be selfadjoint independently distributed $n \times n$ random matrices as in Theorem 2.1 of [2]. For*

$$c = \begin{pmatrix} c_{11} & \dots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{N1} & \dots & c_{NN} \end{pmatrix} \in M_N(\mathbf{C})$$

and for n a multiple of N let

$$c(n) = \begin{pmatrix} c_{11}I_{n/N} & \dots & c_{1N}I_{n/N} \\ \vdots & \ddots & \vdots \\ c_{N1}I_{n/N} & \dots & c_{NN}I_{n/N} \end{pmatrix}$$

be a constant matrix in $M_n(L)$. *Then*

$$\{(\{Y(s, n)\})_{s\in S}, \{c(n)|c \in M_N(\mathbf{C})\}\}$$

is an asymptotically free family as $n \to \infty$, *and each* $Y(s, n)$ *has for limit distribution a semicircle law.*

An immediate result of the above is (3.2 of **[2]**):

THEOREM 1.2. *In a noncommutative probability space* $(\mathscr{M}, \phi)$ *with* ϕ *a trace, let* $\nu_1 = \{X^s|s \in S\}$ *be a semicircular family and let* $\nu_2 = \{e_{ij}|1 \le i, j \le n\}$ *be a system of matrix units such that* $\{\nu_1, \nu_2\}$ *is free. Then in* $(e_{11}\mathscr{M}e_{11}, n\phi|_{e_{11}\mathscr{M}e_{11}})$, $\omega_1 = \{e_{1i}X^s e_{i1}|1 \le i \le n, s \in S\}$ *is a semicircular family and* $\omega_2 = \{e_{1i}X^s e_{j1}|1 \le i < j \le n, s \in S\}$ *is a circular family such that* $\{\omega_1, \omega_2\}$ *is free.*

The following is analogous to Theorem 2.4 of **[11]**.

THEOREM 1.3. *In a noncommutative probability space* $(\mathscr{M}, \phi)$ *with* ϕ *a trace, let* $\nu = \{X^s|s \in S\}$ *be a semicircular family and let* R *be a copy of the hyperfinite* II_1*-factor such that* $\{\nu, R\}$ *is free. Let* $p \in R$ *be a nonzero selfadjoint projection. Then in* $(p\mathscr{M}p, \phi(p)^{-1}\phi|_{p\mathscr{M}p})$, $\omega = \{pX^s p|s \in S\}$ *is a semicircular family and* $\{pRp, \omega\}$ *is free.* (*Note from* **[4]** *that* pRp *is also a copy of the hyperfinite* II_1*-factor.*)

Proof. Suppose first that $\phi(p) = m/2^k$, a dyadic rational number. Since for $U \in R$ a unitary, $\{R, U\nu U^*\}$ is free, we may let p be any projection in R of the given trace. Writing $R = M_{2^k} \otimes M_2 \otimes M_2 \otimes \cdots$, we use Theorem 1.1 in order to model ν as the limit of selfadjoint independently distributed random matrices of size $n = 2^k, 2^{k+1}, 2^{k+2}, \ldots,$ and model a dense subalgebra of R (equal to the tensor product of matrix algebras) by constant random matrices. Choosing p to correspond to a diagonal element of M_{2^k}, we may apply Theorem 1.1 again to see that ω is a semicircular family, $pRp \cong M_m \otimes M_2 \otimes M_2 \otimes \cdots$, and $\{pRp, \omega\}$ is free.

Now for general p, let $(p_l)_{l=1}^{\infty}$ be a decreasing sequence of projections in R which converge to p and such that each $\phi(p_l)$ is a dyadic rational number. Then

$$\{p_l R p_l = \{p_l y p_l | y \in R\}, \{p_l X^s p_l | s \in S\}\}$$

has limit distribution equal to $\{pRp, \omega\}$ as $l \to \infty$. For each l we have freeness and semicircularity, hence also in the limit. □

In addition, modeling R and a semicircular family as in the above proof, we can easily prove

THEOREM 1.4. *In a noncommutative probability space* $(\mathscr{M}, \phi)$ *with* ϕ *a trace, let* $\nu = \{X^s | s \in S\}$ *be a semicircular family, and let* R *be a hyperfinite* II_1*-factor containing a system of matrix units* $\{e_{ij} | 1 \leq i, j \leq n\}$, *such that* $\{\nu, R\}$ *is free. Then in* $(e_{11}\mathscr{M}e_{11}, n\phi|_{e_{11}\mathscr{M}e_{11}})$, $\omega_1 = \{e_{1i}X^s e_{i1} | 1 \leq i \leq n, \ s \in S\}$ *is a semicircular family and* $\omega_2 = \{e_{1i}X^s e_{j1} | 1 \leq i < j \leq n, \ s \in S\}$ *is a circular family such that* $\{\omega_1, \omega_2, e_{11}Re_{11}\}$ *is* $*$*-free.*

2. Definition and compressions of $L(\mathbf{F}_r)$.

DEFINITION 2.1. In a W*-probability space $(\mathscr{M}, \tau)$, where τ is a faithful trace, let R be a copy of the hyperfinite II_1-factor and $\omega = \{X^t | t \in T\}$ be a semicircular family such that R and ω are free. Then $L(\mathbf{F}_r)$ for $1 < r \leq \infty$ will denote any factor isomorphic to $(R \cup \{p_t X^t p_t | t \in T\})''$, where $p_t \in R$ are selfadjoint projections and $r = 1 + \sum_{t \in T} \tau(p_l)^2$.

PROPOSITION 2.2. $L(\mathbf{F}_r)$ *is well-defined, i.e. if*

$$\mathscr{A} = (R \cup \{p_t X^t p_t | t \in T\})'' \quad \textit{and} \quad \mathscr{B} = (R \cup \{q_t X^t q_t | t \in T\})'',$$

where $1 + \sum \tau(p_t)^2 = r = 1 + \sum \tau(q_t)^2$, *then* $\mathscr{A} \cong \mathscr{B}$.

Proof. We show that $\mathscr{A}$ (and thus also $\mathscr{B}$) is isomorphic to an algebra of a certain "standard form." Let $(f_k)_{k=1}^{\infty}$ be an orthogonal family of projections in R such that $\tau(f_k) = 2^{-k}$, and let $f_0 = 1$. If $r < \infty$ let N_l $(l \geq 0)$ be nonnegative integers corresponding to the base 4 expansion for r, i.e. $r = \sum_{l=0}^{\infty} N_l 4^{-l}$, $N_l \leq 3$ if $l \geq 1$ and $\sum_{l > l'} N_l 4^{-l} < 4^{-l'}$ $\forall l' \geq 0$. If $r = \infty$ we let $N_0 = \infty$ and $N_l = 0$ $\forall l \geq 1$. Let $S \subseteq T$, $k_s \in \mathbf{N} = \{0, 1, 2, \ldots\}$ for $s \in S$ be such that $|\{s \in S \,|\, k_s = l\}| = N_l$ $\forall l \geq 0$. The algebra of standard form is then $\mathscr{C} = (R \cup \{f_{k_s} X^s f_{k_s} \,|\, s \in S\})''$. Showing $\mathscr{A} \cong \mathscr{C}$ will prove the proposition.

Proving $\mathscr{A} \cong \mathscr{C}$ is an exercise in cutting and pasting. Note that if U_t are unitaries in R $(t \in T)$, then $\{R, (\{U_t X^t U_t^*\})_{t \in T}\}$ is free in $(\mathscr{M}, \tau)$. Moreover, each projection $p \in R$ is conjugate by a unitary in R to a projection that is a (possibly infinite) sum of projections in $\{f_k | k \geq 1\}$. Hence letting $T' = \{t \in T | p_t \neq 0\}$, we may assume without loss of generality that each p_t for $t \in T'$ is equal to such a sum, and we write $p_t = \sum_{k \in K_t} f_k$, for $K_t \subseteq \mathbf{N} \backslash \{0\}$ whenever $t \in T'$ and $p_t \neq 1$, and we set $K_t = \{0\}$ if $p_t = 1$. Then

$$\mathscr{A} = (R \cup \{f_k X^t f_{k'} | k, k' \in K_t, \ k' \leq k, \ t \in T'\})''.$$

Now we may appeal to the matrix model (§1) to see that (enlarging T if necessary),

$$\mathscr{A} = (R \cup \{f_k X^{\alpha(k,k',t)} f_{k'} | k, k' \in K_t, \ k' \le k, \ t \in T'\})'',$$

where α is a 1-1 map from $\{k, k' \in K_t, \ k' \le k, \ t \in T'\}$ onto a subset T'' of T. (The truth of the above assertion is most easily demonstrated when T' and each K_t are finite; the general case then follows by taking inductive limits.)

Consider for a moment $f_k X^t f_{k'}$ for $k' < k$, $t \in T$. Note that $f_{k'}$ is the sum of $2^{k-k'}$ orthogonal projections, each of which is equivalent in R to f_k. Using the matrix model shows that

$$\text{(4)} \quad (R \cup \{f_k X^t f_{k'}\})'' \cong (R \cup \{f_k X^{t_j} f_k | 1 \le j \le 2^{k-k'}\} \cup \{f_k X^{t'_j} f_k | 1 \le j \le 2^{k-k'}\})'',$$

where $t_1, \dots, t_{2^{k-k'}}, t'_1, \dots, t'_{2^{k-k'}}$ are distinct elements of T, and the isomorphism in (4) maps R identically into itself. Using inductive limits, one obtains

$$\text{(5)} \qquad \mathscr{A} \cong \widetilde{\mathscr{C}} = (R \cup \{f_{k_s} X^s f_{k_s} | s \in S'\})'',$$

for S' some subset of T, $k_s \in \mathbf{N}$ for each $s \in S'$. Moreover, checking the arithmetic of the above moves shows that $1 + \sum_{s \in S'} \tau(f_{k_s})^2 = r$.

Now for the pasting. Note that by the matrix model,

$$\text{(6)} \qquad (R \cup \{f_k X^{t_i} f_k | 1 \le i \le 4\})'' \cong (R \cup \{f_{k-1} X^t f_{k-1}\})''$$

by an isomorphism mapping R identically to itself, whenever $k \ge 1$, $t_1, \dots, t_4$ are distinct elements of T and $t \in T$. Suppose $r < \infty$. If r is not a dyadic rational then for each $l \ge 0$ let $\mathscr{C}(l) = (R \cup \{f_{k_s} X^s f_{k_s} \,|\, s \in S, \ k_s \le l\})'' \subseteq \mathscr{C}$. There is an increasing sequence $S'(l)$ of finite subsets of S' such that $\sum_{s \in S'(l)} 4^{-k_s} = \sum_{0 \le k \le l} N_k 4^{-k}$ and $\bigcup_{l \ge 1} S'(l) = S'$. Let

$$\text{(7)} \qquad \widetilde{\mathscr{C}}(l) = (R \cup \{f_{k_s} X^s f_{k_s} \,|\, s \in S'(l)\})'' \subseteq \widetilde{\mathscr{C}}.$$

Using (6) repeatedly we can find a compatible family of isomorphisms $\phi_l : \widetilde{\mathscr{C}}(l) \to \mathscr{C}(l)$, and taking inductive limits yields $\widetilde{\mathscr{C}} \cong \mathscr{C}$. If r equals a dyadic rational and S' is finite then a finite number of applications of (6) yields $\widetilde{\mathscr{C}} \cong \mathscr{C}$. If S' is infinite, let l be largest such that $N_l \ne 0$, let $\sigma \in S$ be such that $k_\sigma = l$ and let $f_l \ge g_{l+1} \ge g_{l+2} \ge \cdots$ be projections in R where $\tau(g_m) = 2^{-m}$. For $m > l$ let $\mathscr{C}(m) = (R \cup \{f_{k_s} X^s f_{k_s} \,|\, s \in S \backslash \{\sigma\}\} \cup \{f_{k_\sigma} X^\sigma f_{k_\sigma} - g_m X^\sigma g_m\})'' \subseteq \mathscr{C}$.

Then as before we can use the matrix model to find an increasing family $\widetilde{\mathscr{C}}(m)$ of subalgebras of $\widetilde{\mathscr{C}}$ whose union generates $\widetilde{\mathscr{C}}$ and compatible isomorphisms $\widetilde{\mathscr{C}}(m) \to \mathscr{C}(m)$. Taking inductive limits yields $\widetilde{\mathscr{C}} \cong \mathscr{C}$.

If $r = \infty$, then considering S' from (5) and letting $S'_k = \{s \in S' | k_s = k\}$, we have that $\sum_{k=0}^{\infty} |S'_k| 4^{-k} = \infty$. Now by repeated application of (6), we can transform the situation (by isomorphisms mapping R identically to itself) so that first some $|S'_k| = \infty$, then all $|S'_k| = \infty$, then $|S'_0| = \infty$ and $|S'_k| = 0$ for all $k \geq 1$. Thus $\widetilde{\mathscr{C}} = L(\mathbf{F}_\infty)$ by (3). □

REMARK 2.3. Formula (2), together with the fact that $L(\mathbf{F}_r)$ for $r \in \mathbf{N}$ is the free group factor on r generators, shows that Definition 2.1 is equivalent to Rădulescu's definitions 4.1 and 5.3 of [7]. However, for $r \geq 2$ (i.e. Rădulescu's 4.1), this equivalence can be seen directly using the "standard form" of $L(\mathbf{F}_r)$ as defined in Proposition 1.3, and by noting that the isomorphism

$$R * L(\mathbf{Z}) \xrightarrow{\sim} L(\mathbf{Z}) * L(\mathbf{Z}) \tag{8}$$

in [2] sends the set of projections $\{f_k | k \geq 1\} \subset R$ into one of the copies of $L(\mathbf{Z})$ on the right-hand side of (8).

The formula in the following theorem for the compression of an interpolated free group factor $L(\mathbf{F}_r)$ by a projection of trace γ was first proved by Voiculescu [11] for the cases $r = 2, 3, \dots$, $\gamma = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ and $r = \infty$, $\gamma \in \mathbf{Q}_+$. It was then extended by F. Rădulescu in [5] for $r = \infty$ and $\gamma \in \mathbf{R}_+$, and in [6] for $r = 2, 3, \dots$ and $\gamma = \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \dots$. Of course, Rădulescu also proved this theorem in the generality stated here in [7].

THEOREM 2.4.

$$L(\mathbf{F}_r)_\gamma = L\left(\mathbf{F}\left(1 + \frac{r-1}{\gamma^2}\right)\right) \tag{9}$$

for $1 < r \leq \infty$ *and* $0 < \gamma < \infty$.

Proof. It suffices to show the case $0 < \gamma < 1$. Let $L(\mathbf{F}_r) = \mathscr{A} = (R \cup \{p_t X^t p_t | t \in T\})''$ be as in Definition 2.1, so $1 + \sum_{t\in T} \tau(p_t)^2 = r$. Let $p \in R$ be a projection having trace γ. Without loss of generality, we may assume that each $p_t \leq p$. Then

$$p\mathscr{A}p = (pRp \cup \{p_t X^t p_t | t \in T\})'',$$

which by Theorem 1.3 is an interpolation free group factor. Counting gives the formula (9). □

3. Algebraic techniques. A crucial ingredient of our proof of the addition formula for free products (1) will be showing that $R * R \cong R * L(\mathbf{Z})$, with the isomorphism being the identity map on the first copy of R. In order to show this, we will introduce some elementary techniques (Definition 3.4, proof of Theorem 3.5) that are algebraic in nature. These techniques have extensive further applications to free products, as will be seen in [**3**].

REMARK 3.1. In this section, all von Neumann algebras will be finite and have fixed normalized faithful traces associated to them, and all isomorphisms and inclusions of von Neumann algebras will be assumed to be trace preserving. Von Neumann algebras that we obtain from others by certain operations will have associated traces given by the following conventions:

(1) group von Neumann algebras $L(G)$ for G a discrete group will have their canonical traces (equal to the vector-state for the vector $\delta_e \in l^2(G)$);

(2) factors, such as matrix algebras $M_n = M_n(\mathbf{C})$ or the hyperfinite II_1-factor R, will have (of course) their unique normalized traces;

(3) a tensor product $A \otimes B$ of algebras will have the tensor product trace $\tau_A \otimes \tau_B$ of the given traces on A and B;

(4) a free product $A * B$ of algebras will have the free product trace $\tau_A * \tau_B$ of the given traces on A and B;

(5) if $\mathscr{M}$ is a von Neumann algebra with faithful trace τ, and p is a projection in $\mathscr{M}$, then $p\mathscr{M}p$ will have trace $\tau(p)^{-1}\tau|_{p\mathscr{M}p}$.

Also, if A is an algebra with specific trace, $\overset{\circ}{A}$ will denote the ensemble of elements of A whose trace is zero.

First we examine $L(\mathbf{Z}_2) * L(\mathbf{Z}_2)$ (where $\mathbf{Z}_2$ is the two element group). The fact that $\mathscr{M} = L(\mathbf{Z}_2 * \mathbf{Z}_2) \cong L(\mathbf{Z}) \otimes M_2$ is well known, but we will need the following picture of $\mathscr{M}$.

PROPOSITION 3.2. *Consider $\mathscr{M} = L(\mathbf{Z}_2) * L(\mathbf{Z}_2)$ with trace τ, and let p and q be projections of trace $\frac{1}{2}$ generating the first and respectively the second copy of $L(\mathbf{Z}_2)$. Then*

$$\mathscr{M} \cong L^\infty([0, \tfrac{\pi}{2}], \nu) \otimes M_2, \tag{10}$$

where ν is a probability measure on $[0, \frac{\pi}{2}]$ without atoms and τ is given by integration with respect to ν tensored with the normalized trace on $M_2 = M_2(\mathbf{C})$. Moreover, in the setup of (10), *we have that*

$$(11)\qquad p=\begin{pmatrix}1&0\\0&0\end{pmatrix}\quad and\quad q=\begin{pmatrix}\cos^2\theta&\cos\theta\sin\theta\\\cos\theta\sin\theta&\sin^2\theta\end{pmatrix},$$

where $\theta\in[0,\frac{\pi}{2}]$.

Proof. It is well known that the universal unital C^*-algebra generated by two projections p and q is $A=\{f\colon[0,\frac{\pi}{2}]\to M_2(\mathbf{C})\mid f(0)$ and $f(\frac{\pi}{2})$ diagonal$\}$, with p and q as in (11). $\mathscr{M}$ thus has a dense subalgebra equal to a quotient of A, and τ gives a trace on A. One can easily see that a trace on A must be of the following form. Let $f(t)_1$ and $f(t)_2$ be the diagonal values of $f(t)$ for $t=0$ or $\frac{\pi}{2}$. Then

$$\tau(f)=a_1f(0)_1+a_2f(0)_2+\int_0^{\pi/2}\tau_2(f(t))\,d\nu(t)\\+b_1f\left(\frac{\pi}{2}\right)_1+b_2f\left(\frac{\pi}{2}\right)_2,$$

where τ_2 is the normalized trace on $M_2(\mathbf{C})$, ν is a positive measure on $[0,\frac{\pi}{2}]$, $a_1, a_2, b_1, b_2\geq 0$ and $|\nu|+a_1+a_2+b_1+b_2=1$. By Example 2.8 of **[9]**, the distribution of pqp in $p\mathscr{M}p$ has no atoms, which implies that $|\nu|=1$ and ν has no atoms. □

REMARK 3.3. In the right-hand side of (10), let

$$x=\begin{pmatrix}0&0\\1&0\end{pmatrix}=\mathrm{pol}((1-p)qp),$$

where "pol" means "polar part of." Then x is a partial isometry from p to $1-p$ and $\mathscr{M}$ is generated by pqp together with x. Let $y=\mathrm{pol}((1-q)pq)$. Then y is a partial isometry from q to $1-q$. Let

$$w=\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}.$$

Then w is unitary and $wpw^*=q$, $wxw^*=y$.

DEFINITION 3.4. Let $(S_\iota)_{\iota\in I}$ be subsets of a unital algebra $A\ni 1$. A nontrivial *traveling product* in $(S_\iota)_{\iota\in I}$ is a product $a_1a_2\cdots a_n$ such that $a_j\in S_{\iota_j}$ $(1\leq j\leq n)$ and $\iota_1\neq\iota_2\neq\iota_3\neq\cdots\neq\iota_n$. The *trivial traveling product* is the identity element 1. $\Lambda((S_\iota)_{\iota\in I})$ denotes the set of all traveling products in $(S_\iota)_{\iota\in I}$, including the trivial one. If $|I|=2$, we will often call traveling products *alternating products*.

THEOREM 3.5. *Let A and B be finite von Neumann algebras* (*with specified faithful traces—see Remark* 3.1). *Then*

(i) $(A\otimes L(\mathbf{Z}_2))*(B\otimes L(\mathbf{Z}_2))\cong(A*A*B*B*L(\mathbf{Z}))\otimes M_2$,
(ii) $(A\otimes M_2)*(B\otimes L(\mathbf{Z}_2))\cong(A*B*B*L(\mathbf{F}_2))\otimes M_2$,
(iii) $(A\otimes M_2)*(B\otimes M_2)\cong(A*B*L(\mathbf{F}_3))\otimes M_2$.

Proof. Let $\mathscr{M}$ be the von Neumann algebra on the left-hand side of (i) with trace τ. It will be notationally convenient to identify A with $A \otimes 1 \subseteq \mathscr{M}$ and B with $B \otimes 1 \subseteq \mathscr{M}$. Let p and q be projections of trace $\frac{1}{2}$ contained in the copy of $1 \otimes L(\mathbf{Z}_2)$ that commute with A and respectively B. Let $\mathscr{N}_0 = \{p, q\}'' \cong L(\mathbf{Z}_2) * L(\mathbf{Z}_2)$, and let $x, y, w \in \mathscr{N}_0$ be as in Remark 3.3. Then

$$p\mathscr{M}p = (\{pqp\} \cup pA \cup x^*Ax \cup w^*qBw \cup w^*y^*Byw)''.$$

We claim moreover that $\{\{pqp\}, pA, x^*Ax, w^*qBw, w^*y^*Byw\}$ is a free family in $p\mathscr{M}p$, which then clearly implies (i).

Let us first show that $\{\{pqp\}, pA, x^*Ax\}$ is free in $p\mathscr{M}p$. Let $g_k = (pqp)^k - 2\tau((pqp)^k)p$ $(k \geq 1)$. To show freeness means to show that a nontrivial traveling product in $\{g_k \mid k \geq 1\}$, $p\overset{\circ}{A}$ and $x^*\overset{\circ}{A}x$ has trace zero. Regrouping gives a traveling product in $\Omega_0 = \{x, x^*\} \cup \{g_k, xg_k, g_kx^*, xg_kx^* \mid k \geq 1\}$ and $\overset{\circ}{A}$. Let $a = p - \frac{1}{2}$, $b = q - \frac{1}{2}$. Then $\mathscr{N}_0 = \{a, b\}''$, and $\operatorname{span}\Lambda(\{a\}, \{b\})$ is a dense $*$-subalgebra of $\mathscr{N}_0$. Note that $\Omega_\circ \subset \mathscr{N}_0$, so that by the Kaplansky Density Theorem, any $z \in \Omega_0$ is the s.o.-limit of a bounded sequence in $\operatorname{span}\Lambda(\{a\}, \{b\})$. Note also that since a and b are free and each has trace zero, the trace of an element of $\operatorname{span}\Lambda(\{a\}, \{b\})$ is equal to the coefficient of 1. Since $\tau(z) = 0$, we may choose that approximating sequence in $\operatorname{span}\Lambda(\{a\}, \{b\})$ so that each coefficient of 1 equals zero. Moreover, since also $\tau(pz) = 0$, we may also insist that each coefficient of a be zero, i.e. we have a bounded approximating sequence for z of elements of $\operatorname{span}(\Lambda(\{a\}, \{b\})\backslash\{1, a\})$. We must now only show that a nontrivial alternating product in $\Lambda(\{a\}, \{b\})\backslash\{1, a\}$ and $\overset{\circ}{A}$ has trace zero. Regrouping gives a nontrivial alternating product in $\{a\} \cup \overset{\circ}{A} \cup a\overset{\circ}{A}$ and $\{b\}$, which by freeness has trace zero.

Let $\mathscr{N}_1 = (A \cup \mathscr{N}_0)''$, and let us show that $\{qw\mathscr{N}_1w^*, qB, y^*By\}$ is free in $q\mathscr{M}q$, which will complete the proof of (i). We show that a nontrivial traveling product in $w\mathscr{N}_1w^*$, $q\overset{\circ}{B}$ and $y^*\overset{\circ}{B}y$ has trace zero. Regrouping gives a traveling product in $\Omega_1 = \{y, y^*\} \cup qw\overset{\circ}{\mathscr{N}}_1w^* \cup yw\overset{\circ}{\mathscr{N}}_1w^* \cup w\overset{\circ}{\mathscr{N}}_1w^*y^* \cup yw\overset{\circ}{\mathscr{N}}_1w^*y^*$ and $\overset{\circ}{B}$. Now $\Omega_1 \subset \mathscr{N}_1$, $\operatorname{span}\Lambda(\{a\} \cup \overset{\circ}{A} \cup a\overset{\circ}{A}, \{b\})$ is a dense $*$-subalgebra of $\mathscr{N}_1$ and $\tau(z) = \tau(qz) = 0$ $\forall z \in \Omega_1$, so that as above, each $z \in \Omega_1$ is the s.o.-limit of a bounded sequence in

$$\operatorname{span}(\Lambda(\{a\} \cup \overset{\circ}{A} \cup a\overset{\circ}{A}, \{b\})\backslash\{1, b\}).$$

So it suffices to show that a nontrivial alternating product in $\operatorname{span}(\Lambda(\{a\} \cup \overset{\circ}{A} \cup a\overset{\circ}{A}, \{b\}) \backslash \{1, b\})$ and $\overset{\circ}{B}$ has trace zero. Regrouping gives a nontrivial alternating product in $\{a\} \cup \overset{\circ}{A} \cup a\overset{\circ}{A}$ and $\{b\} \cup \overset{\circ}{B} \cup b\overset{\circ}{B}$, which by freeness has trace zero.

Now we prove (ii). Let $\mathscr{M}$ be the von Neumann algebra on the left-hand side of (ii), and let τ be its trace. We will identify A with $A \otimes 1$ and B with $B \otimes 1$ as in the proof of (i). Let p be a projection in $1 \otimes M_2$ (commuting with A) of trace $\frac{1}{2}$ and q a projection in $1 \otimes L(\mathbf{Z}_2)$ (commuting with B) of trace $\frac{1}{2}$. Let $\mathscr{N}_0 = \{p, q\}''$ and let $x, y, w, z, b \in \mathscr{N}_0$ be as in the proof of (i). Let $u \in 1 \otimes M_2$ be a partial isometry from p to $1-p$. Then

$$p\mathscr{M}p = (\{pqp, x^*u\} \cup pA \cup w^*qBw \cup w^*y^*Byw)'',$$

and we shall show that x^*u is a Haar unitary (i.e. a unitary such that $(x^*u)^n$ has trace zero $\forall n \in \mathbf{Z}\backslash\{0\}$) and that $\{\{pqp\}, \{x^*u\}, pA, w^*qBw, w^*y^*Byw\}$ is $*$-free in $p\mathscr{M}p$. This will in turn prove (ii). For $n > 0$, $r = (x^*u)^n$ is a nontrivial alternating product in $\{x^*\}$ and $\{u\}$, and x^* is the s.o.-limit of a bounded sequence in $\operatorname{span}(\Lambda(\{a\}, \{b\}) \backslash \{1, a\})$, so to show $\tau(r) = 0$ it suffices to show that a nontrivial alternating product in $\operatorname{span}(\Lambda(\{a\}, \{b\}) \backslash \{1, a\})$ and $\{u\}$ has trace zero. Regrouping gives a nontrivial alternating product in $\{a, u\}$ and $\{b\}$, which by freeness has trace zero. Hence we have shown that x^*u is a Haar unitary in $p\mathscr{M}p$.

Now we show that x^*u and pqp are $*$-free in $p\mathscr{M}p$. Let g_k $(k \geq 1)$ be as in the proof of (i). It suffices to show that a nontrivial alternating product in $\{(x^*u)^n | n \in \mathbf{Z}\backslash\{0\}\}$ and $\{g_k | k \geq 1\}$ has trace zero. Regrouping gives an alternating product in Ω_0 and $\{u, u^*\}$, where Ω_0 is as in the proof of (i), which, proceeding as we did above, we see has trace zero. Similarly, we can show that letting $\widetilde{\mathscr{N}_0} = \{pqp, x^*u\}''$, $\{\widetilde{\mathscr{N}}, pA\}$ is free in $p\mathscr{M}p$, and that letting $\widetilde{\mathscr{N}_1} = (\widetilde{\mathscr{N}_0} \cup A)''$, $\{w^*\widetilde{\mathscr{N}_1}w, qB, y^*By\}$ is free in $q\mathscr{M}q$, thus proving (ii).

To prove (iii), let p and u in $1 \otimes M_2$ commuting with A be as above, let $q \in 1 \otimes M_2$ commuting with B be a projection of trace $\frac{1}{2}$ and $v \in 1 \otimes M_2$ commuting with B a partial isometry from q to $1-q$. Let $x, y, w \in \mathscr{N}_0 = \{p, q\}''$ be as above. Then we similarly show that x^*u and y^*v are Haar unitaries and that $\{\{pqp\}, \{x^*u\}, pA, \{w^*y^*vw\}, w^*qBw\}$, is $*$-free in $p\mathscr{M}p$ (and notice that these taken together generate $p\mathscr{M}p$), which proves (iii). $\square$

COROLLARY 3.6. *Let R and $\widetilde{R}$ be copies of the hyperfinite* II_1-*factor. Then*

$$R * \widetilde{R} \cong R * L(\mathbf{Z}),$$

with an isomorphism which when restricted is the identity map from R to R.

Proof. Write $R = (pRp) \otimes M_2$ and $\widetilde{R} = (\tilde{p}\widetilde{R}\tilde{p}) \otimes M_2$, where p and $\tilde{p}$ are projections of trace $\frac{1}{2}$ in R and respectively $\widetilde{R}$. Then by (iii) and the proof of (iii),

$$p(R * \widetilde{R})p \cong (pRp) * (\tilde{p}\widetilde{R}\tilde{p}) * L(\mathbf{F}_3),$$

and the isomorphism when restricted to $pRp \subset p(R * \widetilde{R})p$ is the identity map from pRp to pRp. Similarly, writing also $L(\mathbf{Z}) \cong L(\mathbf{Z}) \otimes L(\mathbf{Z}_2)$, we have from (ii) and the proof of (ii) that

$$p(R * L(\mathbf{Z}))p \cong (pRp) * L(\mathbf{F}_4),$$

and the isomorphism, when restricted to $pRp \subset p(R * L(\mathbf{Z}))p$, is the identity map from pRp to pRp. Considering the isomorphism (3), we get an isomorphism from $p(R * \widetilde{R})p$ to $p(R * L(\mathbf{Z}))p$ which when restricted is the identity map on pRp. Now tensor with M_2. □

4. The addition formula for free products.

THEOREM 4.1. $L(\mathbf{F}_r) * L(\mathbf{F}_{r'}) = L(\mathbf{F}_{r+r'})$ *for* $1 < r,\ r' \leq \infty$.

Proof. (Please see the comments at the end of the introduction.) In a W*-probability space $(\mathscr{M}, \tau)$ where τ is a trace, let R and $\widetilde{R}$ be copies of the hyperfinite II_1-factor and let $\nu = \{X^t | t \in T\}$ be a semicircular family such that $\{R, \widetilde{R}, \nu\}$ is free. Let

$$L(\mathbf{F}_r) = \mathscr{A} = (R \cup \{p_s X^s p_s | s \in S\})'',$$
$$L(\mathbf{F}_{r'}) = \mathscr{B} = (\widetilde{R} \cup \{q_s X^s q_s | s \in S'\})'',$$

where S and S' are disjoint subsets of T, $p_s \in R$, $q_s \in \widetilde{R}$ are projections and where $1 + \sum_{s\in S} \tau(p_s)^2 = r$, $1 + \sum_{s \in S'} \tau(q_s)^2 = r'$. Then $\mathscr{A}$ and $\mathscr{B}$ are free in $(\mathscr{M}, \tau)$, so

$$L(\mathbf{F}_r) * L(\mathbf{F}_{r'}) \cong \mathscr{N} = (R \cup \widetilde{R} \cup \{p_s X^s p_s | s \in S\} \cup \{q_s X^s q_s | s \in S'\})''.$$

By Corollary 3.6, there exists a semicircular element $Y \in \mathscr{N}_0 = (R \cup \widetilde{R})''$ such that R and $\{Y\}$ are free and together they generate

$\mathcal{N}_0$. Moreover, for $s \in S'$ let $U_s \in \mathcal{N}_0$ be a unitary such that $U_s q_s U_s^* = f_s \in R$. Then

$$\mathcal{N} = (R \cup \{Y\} \cup \{p_s X^s p_s | s \in S\} \cup \{f_s(U_s X^s U_s^*) f_s | s \in S'\})''.$$

To prove the theorem, it suffices to observe that $\{R, \{Y\}, (\{X^s\})_{s\in S}, (\{U_s X^s U_s^*\})_{s\in S'}\}$ is free in $\mathcal{M}$. □

Let us recall [4] that the fundamental group of a II_1-factor $\mathcal{M}$ is defined to be the set of positive real numbers γ such that $\mathcal{M}_\gamma \cong \mathcal{M}$. Murray and von Neumann [4] showed that the fundamental group of the hyperfinite II_1-factor is $\mathbf{R}_+$, and recently Rădulescu [5] has shown that the fundamental group of $L(\mathbf{F}_\infty)$ is also $\mathbf{R}_+$. A. Connes [1] has shown that the fundamental group of $L(G)$ where G is a group with property T of Kazhdan must be countable, but no other examples are known for fundamental groups of II_1-factors.

Equation (2) shows that the isomorphism question for (interpolated) free group factors is equivalent to the fundamental group question. Combined with the addition formula for free prdoucts, we now see that we must have one of two extremes.

COROLLARY 4.2. *We must have either*

(I) $L(\mathbf{F}_r) \cong L(\mathbf{F}_{r'})$ *for all* $1 < r, r' < \infty$ *and the fundamental group of* $L(\mathbf{F}_r)$ *is* $\mathbf{R}_+$ *for all* $1 < r < \infty$, *or*

(II) $L(\mathbf{F}_r) \not\cong L(\mathbf{F}_{r'})$ *for all* $1 < r < r' < \infty$ *and the fundamental group of* $L(\mathbf{F}_r)$ *is* $\{1\}$ *for all* $1 < r < \infty$.

Proof. Using formulas (1) and (2) we can show that if $L(\mathbf{F}_r) = L(\mathbf{F}_{r'})$ for some $r \neq r'$, then we have $L(\mathbf{F}_r) = L(\mathbf{F}_{r''})$ for r'' in some open interval, hence that the fundamental group of $L(\mathbf{F}_r)$ contains an open interval, thus is all of $\mathbf{R}_+$. □

Acknowledgments. I would like to thank Dan Voiculescu, my advisor, for helpful discussions and for suggesting I look at free products such as $M_2(\mathbf{C}) * M_2(\mathbf{C})$.

REFERENCES

[1] A. Connes, *A factor of type* II_1 *with countable fundamental group*, J. Operator Theory, **4** (1980), 151–153.

[2] K. Dykema, *On certain free product factors via an extended matrix model*, J. Funct. Anal., **112** (1993), 31–60.

[3] ——, *Free products of hyperfinite von Neumann algebras and free dimension*, Duke Math. J., **69** (1993), 97–119.

[4] F. J. Murray and J. von Neumann, *Rings of operators.* IV, Ann. of Math., **44** (1943), 716–808.
[5] F. Rădulescu, *The fundamental group of the von Neumann algebra of a free group with infinitely many generators is* $\mathbf{R}_+\backslash\{0\}$, J. Amer. Math. Soc., **5** (1992), 517–532.
[6] ——, *Stable isomorphism of the weak closure of free group convolution algebras,* Inst. Hautes Études Sci. Publ. Math., December, 1991, preprint.
[7] ——, *Random matrices, amalgamated free products and subfactors of the von Neumann algebra of a free group,* Inst. Hautes Études Sci. Publ. Math., December, 1991, preprint.
[8] D. Voiculescu, *Symmetries of some reduced free product C^*-algebras,* Operator Algebras and Their Connections with Topology and Ergodic Theory, Lecture Notes in Math., vol. 1132, Springer-Verlag, 1985, pp. 556–588.
[9] ——, *Multiplication of certain non-commuting random variables,* J. Operator Theory, **18** (1987), 223–235.
[10] ——, *Noncommutative random variables and spectral problems in free product C^*-algebras,* Rocky Mountain J. Math., **20** (1990), 263–283.
[11] ——, *Circular and semicircular systems and free product factors,* Operator Algebras, Unitary Representations, Enveloping Algebras, and Invariant Theory, Progress in Mathematics, vol. 92, Birkhäuser, Boston, 1990.
[12] ——, *Limit laws for Random matrices and free products,* Invent. Math., **104** (1991), 201–220.
[13] ——, *Free non-commutative random variables, random matrices and the* II_1-*factors of free groups,* Quantum Probability and Related Topics VI, L. Accardi, ed., World Scientific, Singapore, 1991, pp. 473–487.

Received April 13, 1992 and in revised form August 21, 1992. Studies and research supported by the Fannie and John Hertz Foundation.

University of California
Berkeley, CA 94720
E-mail address: dykema@math.berkeley.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

TWO-POINT DISTORTION THEOREMS FOR UNIVALENT FUNCTIONS

SEONG-A KIM AND DAVID MINDA

We establish a one-parameter family of symmetric, linearly invariant two-point distortion theorems for univalent functions defined on the unit disk. The weakest theorem in the family is a symmetric, linearly invariant form of a classical distortion theorem of Koebe, while another special case is a distortion theorem of Blatter. All of these distortion theorems are necessary and sufficient for univalence. Each of these distortion theorems can be expressed as a two-point comparison theorem between euclidean and hyperbolic geometry on a simply connected region; however, none of these comparison theorems characterize simply connected regions. We obtain analogous results for convex univalent functions and convex regions, except that in this context the two-point comparison theorems do characterize convex regions.

1. Introduction. We begin by recalling some basic information about the hyperbolic metric and related material. The hyperbolic metric on the unit disk $\mathbb{D} = \{z : |z| < 1\}$ is given by

$$\lambda_{\mathbb{D}}(z)|dz| = \frac{|dz|}{1 - |z|^2}.$$

It is normalized to have constant Gaussian curvature -4. A region Ω in the complex plane $\mathbb{C}$ is called hyperbolic if $\mathbb{C}\backslash\Omega$ contains at least two points. The density of the hyperbolic metric on a hyperbolic region Ω is obtained from

$$\lambda_{\Omega}(f(z))|f'(z)| = \lambda_{\mathbb{D}}(z),$$

where $f : \mathbb{D} \to \Omega$ is any holomorphic universal covering projection of $\mathbb{D}$ onto Ω. The density is independent of the choice of the covering projection of $\mathbb{D}$ onto Ω. The hyperbolic metric on Ω induces the hyperbolic distance function d_{Ω} as follows:

$$d_{\Omega}(a, b) = \inf \int_{\gamma} \lambda_{\Omega}(w)|dw|,$$

where the infimum is taken over all paths γ in Ω joining a and b. The infimum is actually a minimum; there always exists a path δ in

Ω connecting a and b such that

$$d_\Omega(a, b) = \int_\delta \lambda_\Omega(w)|dw|.$$

Any such path δ is called a hyperbolic geodesic joining a and b. There may be more than one hyperbolic geodesic joining a and b when Ω is not simply connected. Recall that

$$d_{\mathbb{D}}(a, b) = \operatorname{artanh}\left|\frac{b-a}{1-\bar{a}b}\right|.$$

Both the hyperbolic metric and the hyperbolic distance are conformally invariant.

Blatter [1] commented that a classical distortion theorem of Koebe for normalized univalent functions $g(z) = z + a_2z^2 + a_3z^3 + \cdots$, namely,

$$|g(z)| \geq \frac{|z|}{(1+|z|)^2}, \qquad z \in \mathbb{D},$$

was necessary, but not sufficient, for univalence. Recall that equality holds at $z \neq 0$ if and only if g is a rotation of the Koebe function $k(z) = z/(1-z)^2$ [3, p. 33]. Koebe's distortion theorem is a consequence of the coefficient bound $|a_2| \leq 2$ for normalized univalent functions. Blatter inquired whether there were distortion theorems for univalent functions that were also sufficient for univalence. He established the following two-point distortion theorem which is both necessary and sufficient for univalence [1]. There is no normalization on the univalent function.

Blatter's distortion theorem. *Suppose f is univalent in $\mathbb{D}$ and $a, b \in \mathbb{D}$. Then*

$$|f(a) - f(b)|^2 \geq \frac{\sinh^2(2d_{\mathbb{D}}(a, b))}{8\cosh(4d_{\mathbb{D}}(a, b))}([(1-|a|^2)|f'(a)|]^2 + [(1-|b|^2)|f'(b)|]^2).$$

Equality holds for distinct points $a, b \in \mathbb{D}$ if and only if $f = S \circ k \circ T$, where S is a conformal automorphism of $\mathbb{C}$, k is the Koebe function and T is a conformal automorphism of $\mathbb{D}$, and a and b lie on the axis of symmetry of f. Conversely, if a nonconstant holomorphic function f satisfies this inequality, then f is univalent on $\mathbb{D}$.

The square on the term $\sinh^2(2d_{\mathbb{D}}(a, b))$ is missing in the statement, but not in the proof, of this result in Blatter's paper. The proof of Blatter's distortion theorem is more sophisticated than the proof of Koebe's distortion theorem; it requires three coefficient inequalities for normalized univalent functions: $|a_2| \leq 2$, $|a_3| \leq 3$, and $|a_3 - a_2^2| \leq 1$. Blatter's distortion theorem is symmetric in a and b and linearly invariant. In this context, linear invariance means that if f is replaced in the inequality by $\tilde{f} = S \circ f \circ T$, where S is a conformal automorphism of $\mathbb{C}$ and T is a conformal automorphism of $\mathbb{D}$, then the new inequality has exactly the same form, except that f is replaced by $\tilde{f}$. This is closely related to the notion of linear invariance introduced by Pommerenke **[13]**. We shall establish a one-parameter family of symmetric, linearly invariant two-point distortion theorems for univalent functions; each of these distortion theorems characterizes univalence. The method of proof is an extension of Blatter's technique. The weakest two-point distortion theorem in the family is a symmetric, linearly invariant version of Koebe's distortion theorem. Blatter's distortion theorem is stronger than the symmetric, linearly invariant version of Koebe's distortion theorem, but is not the strongest one in the family.

Blatter's distortion theorem can easily be formulated as a two-point comparison theorem between euclidean and hyperbolic geometry on a simply connected region. It relates the euclidean distance between two points to their hyperbolic distance and the density of the hyperbolic metric at the points. This formulation asserts that if Ω is a simply connected hyperbolic region in $\mathbb{C}$ and $A, B \in \Omega$, then

$$|A-B|^2 \geq \frac{\sinh^2(2d_\Omega(A, B))}{8\cosh(4d_\Omega(A, B))}\left(\frac{1}{\lambda_\Omega^2(A)} + \frac{1}{\lambda_\Omega^2(B)}\right).$$

Equality holds if and only if Ω is a slit plane and A and B lie on the extension of the slit into Ω. This two-point comparison theorem can be viewed as an extension of the inequality $\lambda_\Omega \geq 1/(4\delta_\Omega)$ for simply connected regions **[6**, p. 45], where $\delta_\Omega(z)$ is the euclidean distance from z to $\partial\Omega$, since this inequality is a limiting case. Because Blatter's distortion theorem characterizes univalence, it is natural to inquire whether this comparison inequality characterizes simply connected regions. The answer is negative. In fact, there is a one-parameter family of similar two-point comparison theorems and not even the strongest comparison theorem in the family characterizes simple connectivity. Narrow annuli also satisfy these comparison inequalities.

Finally, we consider analogs of these results for both convex univalent functions and convex regions. The case of convex univalent functions parallels the univalent function situation. There is a one-parameter family of two-point distortion theorems for convex univalent functions, the weakest of which is the symmetric, linearly invariant version of a classical distortion theorem. These distortion theorems all characterize convex univalent functions. There is an associated one-parameter family of two-point comparison theorems for euclidean and hyperbolic geometry on convex regions. These comparison theorems characterize convex regions and are refinements of the inequality $\lambda_\Omega \geq 1/(2\delta_\Omega)$ **[10]** for convex regions.

We would like to thank Wancang Ma for several helpful conversations regarding univalent functions and the referee for useful comments, including the idea which led to Theorem 1.

2. Preliminaries. We first recall some results from Blatter's paper **[1]**. Some of these are reformulated in invariant terms here, while others are stated in more generality. We do not prove these generalizations if the proofs given in **[1]** immediately extend.

Minimum Principle. Suppose that a function $u\colon [-L, L] \to \mathbb{R}$ satisfies the following two conditions:

(i) $|u'| \leq q$,
(ii) $u'' \leq p(q^2 - (u')^2)$,

where p and q are positive constants. If v is the solution of the inequality $|y'| \leq q$ and the differential equation $y'' = p(q^2 - (y')^2)$ which satisfies the boundary conditions $v(L) = u(L)$ and $v(-L) = u(-L)$, then $u(s) \geq v(s)$ for all $s \in [-L, L]$. Moreover, if strict inequality holds in both (i) and (ii), then $u(s) > v(s)$ for all $s \in (-L, L)$.

The solution v can be expressed in elementary form:

$$v(s) = \frac{1}{p}\log[\cosh(pqs) + \tau\sinh(pqs)] + \log C,$$

where the constants $\tau \in [-1, 1]$ and $C > 0$ are determined by the boundary conditions. In fact,

$$C = \left(\frac{\exp(pu(L)) + \exp(pu(-L))}{2\cosh(pqL)}\right)^{1/p}.$$

LEMMA 1. *For $p > 1$, $q > 0$ and $\tau \in [-1, 1]$ let*

$$B(\tau) = \int_{-L}^{L} (\cosh(pqs) + \tau\sinh(pqs))^{1/p}\, ds.$$

Then for $\tau \in (-1, 1)$

$$B(\tau) > B(\pm 1) = \frac{2}{q}\sinh(qL).$$

Proof. Now,

$$B'(\tau) = \frac{1}{p}\int_{-L}^{L} \sinh(pqs)(\cosh(pqs) + \tau\sinh(pqs))^{(1-p)/p}\,ds$$

and

$$B''(\tau) = \frac{1-p}{p^2}\int_{-L}^{L} \sinh^2(pqs)\,(\cosh(pqs) + \tau\sinh(pqs))^{(1-2p)/p}\,ds.$$

Thus, $B''(\tau) < 0$ since $p > 1$, so $B(\tau)$ is strictly concave on $[-1, 1]$. This implies that the minimum value of $B(\tau)$ is either $B(1)$ or $B(-1)$. Because

$$B(1) = B(-1) = \frac{2}{q}\sinh(qL),$$

the proof is complete.

REMARKS. (i) When $p = 1$ the function $B(\tau)$ is the constant $\frac{2}{q}\sinh(qL)$.

(ii) If u and v are as in the statement of the minimum principle, then

$$\int_{-L}^{L} \exp(u(s))\,ds \geq \int_{-L}^{L} \exp(v(s))\,ds = CB(\tau) \geq C\frac{2}{q}\sinh(qL),$$

with equality if and only if $\exp u(s) = C\exp(\pm qs)$.

Next, we want to recall some differential geometric formulas relating to locally schlicht holomorphic functions. Before stating these formulas, it is convenient to introduce several invariant differential operators which were also considered in **[3]** and **[8]**. For a holomorphic function f defined on $\mathbb{D}$, let

$$\begin{aligned} D_1 f(z) &= (1-|z|^2)f'(z), \\ D_2 f(z) &= (1-|z|^2)^2 f''(z) - 2\bar{z}(1-|z|^2)f'(z), \\ D_3 f(z) &= (1-|z|^2)^3 f'''(z) - 6\bar{z}(1-|z|^2)^2 f''(z) \\ &\quad + 6\bar{z}^2(1-|z|^2)f'(z). \end{aligned}$$

If $T(z) = (z+a)/(1+\bar{a}z)$, then $D_j f(a) = (f\circ T)^{(j)}(0)$ for $j = 1, 2, 3$. In particular, $D_j f(0)$ is just the ordinary jth derivative at the origin. These differential operators are invariant in the sense that

$$|D_j(S\circ f\circ T)| = |D_j(f)|\circ T \qquad (j = 1, 2, 3),$$

where T is any conformal automorphism of $\mathbb{D}$ and S is any euclidean motion of $\mathbb{C}$ **[8]**. Observe that for a locally schlicht function f

$$\frac{D_3 f(z)}{D_1 f(z)} - \frac{3}{2}\left(\frac{D_2 f(z)}{D_1 f(z)}\right)^2 = (1-|z|^2)^2 S_f(z),$$

where

$$S_f(z) = \frac{f'''(z)}{f'(z)} - \frac{3}{2}\left(\frac{f''(z)}{f'(z)}\right)^2$$

denotes the Schwarzian derivative of f. For a locally schlicht holomorphic function f defined on the unit disk it is useful to introduce the abbreviation

$$Q_f(z) = \frac{D_2 f(z)}{D_1 f(z)} = (1-|z|)^2 \frac{f''(z)}{f'(z)} - 2\bar{z}.$$

Now, we establish some notation that will be in force for the remainder of the paper. Suppose f is a locally schlicht holomorphic function defined on the unit disk $\mathbb{D}$. We assume that there is a Jordan arc γ in $\mathbb{D}$ with finite hyperbolic length $2L$ joining a and b such that f maps γ injectively onto the euclidean segment $[f(a), f(b)] = [A, B]$. Suppose the arc γ is parametrized by hyperbolic arc length, say $\gamma\colon z = z(s)$, $s \in [-L, L]$. This implies $z'(s) = (1-|z(s)|^2)e^{i\theta(s)}$, where $\theta(s) = \arg z'(s)$. The hyperbolic curvature of γ is

$$\begin{aligned}\kappa_h(z(s), \gamma) &= (1-|z(s)|^2)\kappa_e(z(s), \gamma) + \operatorname{Im}\left\{\frac{2\overline{z(s)}z'(s)}{|z'(s)|}\right\} \\ &= (1-|z(s)|^2)\kappa_e(z(s), \gamma) + \operatorname{Im}\{2\overline{z(s)}e^{i\theta(s)}\}.\end{aligned}$$

Here $\kappa_e(z(s), \gamma)$ is the euclidean curvature of γ at $z(s)$; explicitly,

$$\kappa_e(z(s), \gamma) = \frac{1}{|z'(s)|}\operatorname{Im}\left\{\frac{z''(s)}{z'(s)}\right\}.$$

The formula which relates the euclidean curvature of $f\circ\gamma$ to the hyperbolic curvature of γ is

$$\kappa_e(f(z(s)), f\circ\gamma)|D_1 f(z(s))| = \kappa_h(z(s), \gamma) + \operatorname{Im}\left\{Q_f(z(s))\frac{z'(s)}{|z'(s)|}\right\}.$$

When $f\circ\gamma$ is a euclidean line segment, this simplifies to

$$\kappa_h(z(s), \gamma) = -\operatorname{Im}\left\{Q_f(z(s))\frac{z'(s)}{|z'(s)|}\right\}.$$

The rate of change of the euclidean curvature of $f \circ \gamma$ is related to the rate of change of the hyperbolic curvature of γ by

$$\frac{d\kappa_e(f(z(s)), f\circ\gamma)}{ds}|D_1 f(z(s))|$$
$$= \frac{d\kappa_h(z(s), \gamma)}{ds} + \operatorname{Im}\left\{(1-|z(s)|^2)^2 S_f(z(s))\left(\frac{z'(s)}{|z'(s)|}\right)^2\right\}.$$

When $f\circ\gamma$ is a euclidean line segment, this becomes

$$\frac{d\kappa_h(z(s), \gamma)}{ds} = -\operatorname{Im}\left\{(1-|z(s)|^2)^2 S_f(z(s))\left(\frac{z'(s)}{|z'(s)|}\right)^2\right\}.$$

Set

$$u(s) = \log|D_1 f(z(s))|.$$

Then

$$u'(s) = \operatorname{Re}\{Q_f(z(s))e^{i\theta(s)}\},$$

so that

$$|u'(s)| \le |Q_f(z(s))|$$

and

$$(u')^2(s) = \frac{1}{2}\operatorname{Re}\{(Q_f(z(s)))^2 e^{2i\theta(s)}\} + \frac{1}{2}|Q_f(z(s))|^2.$$

Also,

$$u''(s) = \operatorname{Re}\{(1-|z(s)|^2)^2 S_f(z(s))e^{2i\theta(s)}\} + \frac{1}{2}|Q_f(z(s))|^2 - 2.$$

By making use of some of these formulas, we obtain the identity

$$u''(s) + p(u')^2(s)$$
$$= \operatorname{Re}\left\{\left[(1-|z(s)|^2)^2 S_f(z(s)) + \frac{p}{2}(Q_f(z(s)))^2\right]e^{2i\theta(s)}\right\}$$
$$+ \frac{p+1}{2}|Q_f(z(s))|^2 - 2,$$

and so the differential inequality

$$u''(s) + p(u')^2(s) \le \left|(1-|z(s)|^2)^2 S_f(z(s)) + \frac{p}{2}\left(Q_f(z(s))\right)^2\right|$$
$$+ \frac{p+1}{2}|Q_f(z(s))|^2 - 2.$$

3. Univalent functions and simply connected regions. We establish symmetric, linearly invariant, two-point distortion theorems for univalent functions and consider the associated two-point comparison theorems between euclidean and hyperbolic geometry on simply connected regions.

INVARIANT KOEBE DISTORTION THEOREM. *Suppose f is univalent on $\mathbb{D}$. Then for all $a, b \in \mathbb{D}$,*

$$|f(a) - f(b)| \geq \frac{\sinh(2d_{\mathbb{D}}(a, b))}{2\exp(2d_{\mathbb{D}}(a, b))} \max\{|D_1 f(a)|, |D_1 f(b)|\}.$$

Equality holds for distinct points $a, b \in \mathbb{D}$ if and only if $f = S \circ k \circ T$, where S is a conformal automorphism of $\mathbb{C}$, k is the Koebe function and T is a conformal automorphism of $\mathbb{D}$, and a and b lie on the axis of symmetry of f. Conversely, if a nonconstant holomorhpic function f satisfies this inequality, then f is univalent on $\mathbb{D}$.

Proof. First, note that Koebe's classical distortion theorem can be written in the form

$$|g(z)| \geq \frac{|z|}{(1+|z|)^2} = \frac{\sinh(2d_{\mathbb{D}}(0, z))}{2\exp(2d_{\mathbb{D}}(0, z))}.$$

Here g is a normalized univalent function.

Now, assume f is univalent (not necessarily normalized) in $\mathbb{D}$ and $a, b \in \mathbb{D}$. Set $T(z) = (z+a)/(1+\bar{a}z)$; T is a conformal automorphism of $\mathbb{D}$ which sends 0 to a. Then

$$g(z) = [f \circ T(z) - f \circ T(0)]/(f \circ T)'(0)$$

is a normalized univalent function. If we apply the classical Koebe distortion theorem to g and use the fact that hyperbolic distance is conformally invariant, then we obtain

$$|f(a) - f(b)| \geq \frac{\sinh(2d_{\mathbb{D}}(a, b))}{2\exp(2d_{\mathbb{D}}(a, b))} |D_1 f(a)|.$$

We obtain a similar inequality when we interchange the roles of a and b. The final formula is obtained by taking the maximum value of these two lower bounds on $|f(a) - f(b)|$. The necessary and sufficient conditions for equality follow from the conditions for equality in the classical Koebe distortion theorem.

The fact that the condition is sufficient for univalence is elementary, but we give the details here and then omit them in subsequent related theorems. Suppose f is a nonconstant holomorphic function defined on $\mathbb{D}$ which satisfies the inequality. Assume $f(a) = f(b)$ for distinct points $a, b \in \mathbb{D}$. The inequality implies that $f'(a) = f'(b) = 0$. Then f is not univalent in any neighborhood of a (or b), so there exist two sequences $\{a_n\}$ and $\{b_n\}$ of distinct points such that $a_n \to a$, $b_n \to a$ and $f(a_n) = f(b_n)$ for all n. This gives $f'(a_n) = 0$ for all n which contradicts the fact that f is nonconstant since this implies f' must have an isolated zero at a. Hence, f is univalent on $\mathbb{D}$.

Thus, the invariant form of Koebe's distortion theorem is sufficient for univalence, so it provides an elementary answer to the question raised by Blatter. Theorem 2 will provide a connection between the invariant form of Koebe's distortion theorem and Blatter's distortion theorem. But first we need to establish a result for normalized univalent functions.

THEOREM 1. *If* $g(z) = z+a_2z^2+a_3z^3+\cdots$ *is a normalized univalent function on* $\mathbb{D}$, *then*

$$\left|a_3 - \left(\frac{3-p}{3}\right)a_2^2\right| + \frac{p}{3}|a_2|^2 \leq \begin{cases} 1+2\exp\left(\dfrac{2p-3}{p}\right), & 0 < p < \dfrac{3}{2}, \\ \dfrac{8p-3}{3}, & \dfrac{3}{2} \leq p. \end{cases}$$

This inequality is sharp for all $p > 0$. *For* $p \geq 3/2$, *equality holds if and only if* g *is a rotation of the Koebe function.*

Proof. It suffices to obtain the sharp upper bound on the functional

$$\begin{aligned} L_p(g) &= \operatorname{Re}\left\{a_3 - \left(\frac{3-p}{3}\right)a_2^2\right\} + \frac{p}{3}|a_2|^2 \\ &= \operatorname{Re}\{a_3\} - \left(\frac{3-2p}{3}\right)(\operatorname{Re} a_2)^2 + (\operatorname{Im} a_2)^2 \end{aligned}$$

over the family of normalized univalent functions. Because replacing $g(z)$ by $-g(-z)$ does not change the value of $L_p(g)$, we may assume that $\operatorname{Re}\{a_2\} \geq 0$ without loss of generality. Since $0 \leq \operatorname{Re}\{a_2\} \leq 2$, there is a unique $\lambda \in [0, 2]$ with $\operatorname{Re}\{a_2\} = \lambda(1+\log\frac{2}{\lambda})$.

Jenkins [5] obtained the sharp relationship between the second and third coefficients of a normalized univalent function. We shall use the version of this result from [14, p. 120]; specifically, we need inequality (12) of this reference which states

$$\operatorname{Re}\{a_3\} \leq (\operatorname{Re} a_2)^2 - (\operatorname{Im} a_2)^2 - 2\lambda \operatorname{Re} a_2 + \lambda^2 \log\frac{2}{\lambda} + \frac{3}{2}\lambda^2 + 1.$$

From this inequality we obtain

$$\begin{aligned} L_p(g) &\le \frac{2p}{3}(\operatorname{Re} a_2)^2 - 2\lambda(\operatorname{Re} a_2) + \lambda^2 \log\frac{2}{\lambda} + \frac{3}{2}\lambda^2 + 1 \\ &= \left(\frac{4p-3}{6}\right)\lambda^2 + \left(\frac{4p-3}{3}\right)\lambda^2 \log\frac{2}{\lambda} \\ &\quad + \frac{2p}{3}\lambda^2\left(\log\frac{2}{\lambda}\right)^2 + 1 = H(\lambda). \end{aligned}$$

Note that $H(0) = 1$, $H(2) = (8p-3)/3$ and

$$H'(\lambda) = \frac{2\lambda}{3}\log\left(\frac{2}{\lambda}\right)\left[2p - 3 + 2p\log\left(\frac{2}{\lambda}\right)\right].$$

For $p \ge 3/2$, $H'(\lambda)$ has no roots in (0, 2), so $H(\lambda)$ is strictly increasing in this case with maximum value $(8p-3)/3$ attained uniquely at $\lambda = 2$. This produces the sharp upper bound on $L_p(g)$ when $p \ge 3/2$, and implies that equality holds only if g is a rotation of the Koebe function. It is trivial that equality holds for a rotation of the Koebe function. When $0 < p < 3/2$, $H'(\lambda)$ has a root at $\lambda_0 = 2\exp((2p-3)/(2p)) \in (0, 2)$ and $H(\lambda)$ is increasing on $(0, \lambda_0)$ and decreasing on $(\lambda_0, 1)$. Thus, $H(\lambda)$ has maximum value $H(\lambda_0) = 1 + 2\exp((2p-3)/p)$ when $0 < p < 3/2$. The sharpness of the inequality in this case follows from the work of Jenkins; note that the Koebe function is not extremal.

COROLLARY. *If $g(z) = z + a_2 z^2 + a_3 z^3 + \cdots$ is a normalized univalent function on $\mathbb{D}$, then*

$$\left|a_3 - \frac{1}{2}a_2^2\right| + \frac{5}{6}|a_2|^2 \le \frac{13}{3}$$

with equality if and only if f is a rotation of the Koebe function.

Proof. By making use of the theorem with $p = 3/2$ and $|a_2| \le 2$, we get

$$\left|a_3 - \frac{1}{2}a_2^2\right| + \frac{5}{6}|a_2|^2 \le \left|a_3 - \frac{1}{2}a_2^2\right| + \frac{1}{2}|a_2|^2 + \frac{1}{3}|a_2|^2 \le 3 + \frac{4}{3} = \frac{13}{3}.$$

THEOREM 2. *Suppose f is univalent in $\mathbb{D}$. There is a constant $P \in (1, 3/2]$ such that for any $p \ge P$ and all $a, b \in \mathbb{D}$,*

$$|f(a) - f(b)| \ge \frac{\sinh(2d_{\mathbb{D}}(a, b))}{2\left[2\cosh(2p d_{\mathbb{D}}(a, b))\right]^{1/p}}\left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}.$$

Equality holds for distinct points $a, b \in \mathbb{D}$ if and only if $f = S \circ k \circ T$, where S is a conformal automorphism of $\mathbb{C}$, k is the Koebe function and T is a conformal automorphism of $\mathbb{D}$, and a and b lie on the axis of symmetry of f. Conversely, if a nonconstant holomorphic function f satisfies this inequality, then f is univalent on $\mathbb{D}$.

Proof. The sufficiency for univalence follows exactly as in the proof of the invariant form of the Koebe distortion theorem.

For the necessity, we make use of the notation established in §2. Because f is univalent, we know that $|u'(s)| \le 4$; this is the invariant version of the sharp classical coefficient bound $|a_2| \le 2$ for normalized univalent functions [2, p. 32]. We will make use of some of the results from §2 with $q = 4$. Suppose $p \ge 1$ is any number such that

$$(1) \qquad \left|(1-|z|^2)^2 S_f(z) + \frac{p}{2}\left(Q_f(z)\right)^2\right| + \frac{p+1}{2}|Q_f(z)|^2 - 2 \le 16p$$

for every univalent function f defined on $\mathbb{D}$ and all $z \in \mathbb{D}$. Then the results of §2 with $q = 4$ give

$$u''(s) + p(u')^2(s) \le 16p.$$

Therefore, we get

$$\begin{aligned} |f(a) - f(b)| &= \int_{-L}^{L} |f'(z(s))||dz(s)| \\ &= \int_{-L}^{L} (1-|z(s)|^2)|f'(z(s))| \frac{|dz(s)|}{1-|z(s)|^2} \\ &= \int_{-L}^{L} \exp u(s)\, ds \ge \int_{-L}^{L} \exp v(s)\, ds \ge \frac{C\sinh(4L)}{2}, \end{aligned}$$

with equality if and only if $\exp u(s) = C\exp(\pm 4s)$, where

$$C = \left(\frac{|D_1 f(a)|^p + |D_1 f(b)|^p}{2\cosh(4pL)}\right)^{1/p}.$$

Thus,

$$|f(a) - f(b)| \ge \frac{\sinh(4L)}{2[2\cosh(4pL)]^{1/p}} \left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}.$$

Since the function $h(t) = \sinh(t)/[2\cosh(pt)]^{1/p}$ is increasing and $2d_{\mathbb{D}}(a, b) \le 4L$, we obtain

$$|f(a) - f(b)| \ge \frac{\sinh(2d_{\mathbb{D}}(a,b))}{2[2\cosh(2pd_{\mathbb{D}}(a,b))]^{1/p}} \left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}.$$

This establishes the lower bound when $[f(a), f(b)]$ is contained in $f(\mathbb{D})$. If equality holds, then $d_{\mathbb{D}}(a, b) = 2L$ and so γ must be a hyperbolic geodesic.

We require a limiting form of this inequality. Set $\Omega = f(\mathbb{D})$. Suppose $\alpha \in \partial\Omega$ and $[f(a), \alpha) \subset \Omega$. Then for any $b \in \mathbb{D}$ with $f(b) \in [f(a), \alpha)$, the preceding inequality gives

$$|f(a) - f(b)| \geq \frac{\sinh(2d_{\mathbb{D}}(a, b))}{2[2\cosh(2pd_{\mathbb{D}}(a, b))]^{1/p}}|D_1 f(a)|.$$

When $f(b) \to \partial\Omega$ along the segment $[f(a), \alpha)$, then $b \to \partial\mathbb{D}$ and so $d_{\mathbb{D}}(a, b) \to \infty$. Since $h(\infty) = 1/2$, we get

$$|f(a) - \alpha| \geq \frac{1}{4}|D_1 f(a)|.$$

This is just an invariant form of the Koebe 1/4-theorem.

Now, suppose $[f(a), f(b)]$ does not lie in Ω. Then there exist points $\alpha, \beta \in \partial\Omega$ such that the half-open intervals $[f(a), \alpha)$ and $(\beta, f(b)]$ are disjoint, lie in Ω and their union is contained in $[f(a), f(b)]$. The preceding inequality implies that

$$|f(a) - \alpha| \geq \frac{1}{4}|D_1 f(a)| \quad \text{and} \quad |f(b) - \beta| \geq \frac{1}{4}|D_1 f(b)|.$$

Hence,

$$\begin{aligned} |f(a) - f(b)| &\geq |f(a) - \alpha| + |f(b) - \beta| \geq \frac{1}{4}\left(|D_1 f(a)| + |D_1 f(b)|\right) \\ &\geq \frac{1}{4}\left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}. \end{aligned}$$

Since $h(\infty) = 1/2$ and h is strictly increasing, we obtain

$$|f(a) - f(b)| > \frac{\sinh(2d_{\mathbb{D}}(a, b))}{2[2\cosh(2pd_{\mathbb{D}}(a, b))]^{1/p}}\left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}.$$

This establishes the lower bound in all cases.

Next, we determine necessary and sufficient conditions for equality. If equality holds, then γ must be a hyperbolic geodesic in $\mathbb{D}$. By performing a conformal automorphism of $\mathbb{D}$ if necessary, we may assume that $\gamma \subset (-1, 1)$ and is symmetric about the origin. There is no harm in assuming $[f(a), f(b)] \subset \mathbb{R}$ and is symmetric about the origin with $f(a) < 0$ and $f(b) = -f(a)$; if this were not true just compose f with a conformal automorphism of $\mathbb{C}$. Then the hyperbolic arc length parametrization of γ is $z(s) = \tanh(s)$ and $f'(z(s)) > 0$ for $s \in [-L, L]$. Symmetry implies $f(0) = 0$. Equality forces $\exp(u) = C\exp(\pm 4s)$. We consider the plus sign; the case of the minus sign is similar. We have

$$(1 - z(s)^2)f'(z(s)) = C\exp(4s).$$

Since

$$s = \operatorname{artanh} z = \frac{1}{2}\log\frac{1+z}{1-z}$$

holds on γ, we obtain

$$(1 - z^2)f'(z) = C\left(\frac{1+z}{1-z}\right)^2$$

or

$$f'(z) = C\frac{1+z}{(1-z)^3}$$

for z on γ. The identity theorem implies that this holds for all z in $\mathbb{D}$. Since $f(0) = 0$, we get $f(z) = Ck(z)$. This demonstrates that if equality holds then $f = S \circ k \circ T$, where S is a conformal automorphism of $\mathbb{C}$, k is the Koebe function and T is a conformal automorphism of $\mathbb{D}$, and a and b lie on the axis of symmetry of f. Conversely, if f has this form, then it is straightforward to show that equality holds for all points on the axis of symmetry of f, or equivalently, equality holds for all pairs of points on $(-1, 1)$ for the Koebe function itself.

Finally, we show that inequality (1) holds for all $p \geq P$, where P is some constant in $(1, 3/2]$. It is elementary to verify that if inequality (1) holds for one value of $p \geq 1$, then it also holds for all larger values of p. Let P be the minimum of all $p \geq 1$ such that inequality (1) holds for all univalent functions f defined on $\mathbb{D}$. Since the class of univalent functions is linearly invariant, it suffices to establish inequality (1) for $z = 0$ and normalized univalent functions. Thus, we want to find the smallest value of p such that

$$\left|a_3 - \left(\frac{3-p}{3}\right)a_2^2\right| + \left(\frac{p+1}{3}\right)|a_2|^2 - \frac{1}{3} \leq \frac{8p}{3}.$$

The corollary to Theorem 1 shows that this inequality is valid for $p = 3/2$. It might seem plausible that $P = 1$; this is equivalent to the coefficient inequality

$$\left|a_3 - \frac{2}{3}a_2^2\right| \leq 3 - \frac{2}{3}|a_2|^2$$

for a normalized univalent function. However, Ruscheweyh **[15]**, with the use of a computer, has shown that this inequality is false for the full class S of normalized univalent functions and that the best result for the class S is about

$$\left|a_3 - \frac{2}{3}a_2^2\right| + \frac{2}{3}|a_2|^2 < 3.0031896592.$$

Thus, $P > 1$.

REMARKS. (i) What is the best value of P in Theorem 2?

(ii) The right-hand side of the inequality in Theorem 2 is a decreasing function of p for $p \geq 1$. Consequently, the weakest necessary condition for univalence that Theorem 2 yields is the case $p = \infty$, or more precisely, $p \to \infty$. This is the invariant version of Koebe's distortion theorem. The case $p = 2$ is Blatter's distortion theorem, but it is not the strongest two-point distortion theorem contained in Theorem 2.

COROLLARY. *Let Ω be a simply connected hyperbolic region in $\mathbb{C}$. Then for any $p \geq P$ and all $A, B \in \Omega$,*

$$|A - B| \geq \frac{\sinh(2d_\Omega(A, B))}{2[2\cosh(2pd_\Omega(A, B))]^{1/p}} \left(\frac{1}{\lambda_\Omega^p(A)} + \frac{1}{\lambda_\Omega^p(B)}\right)^{1/p}.$$

Equality holds if and only if Ω is a slit plane A and B lie on the extension of the slit into Ω.

Proof. Apply the theorem to a conformal map f of $\mathbb{D}$ onto Ω and make use of the facts that f is an isometry from the hyperbolic metric on $\mathbb{D}$ to the hyperbolic metric on Ω and $|D_1 f(z)| = 1/\lambda_\Omega(f(z))$.

REMARK. Suppose Ω is any region which satisfies the inequality in the corollary for some $p \geq P$. Fix $A \in \Omega$. Select $\alpha \in \partial\Omega$ so that $|A - \alpha| = \delta_\Omega(A)$. Let $B \in \Omega$ tend to α along the half-open segment $[A, \alpha)$. Then $d_\Omega(A, B) \to \infty$ since the hyperbolic distance is complete and $\lambda_\Omega(B) \to \infty$ [**12**] so the inequality in the corollary yields $\lambda_\Omega \geq 1/(4\delta_\Omega)$. For simply connected regions this inequality is equivalent to the Koebe 1/4-theorem for univalent functions [**6**, p. 45].

EXAMPLE. Let $\Omega = \Omega(\delta) = \{z\colon \exp(-\pi\delta/2) < |z| < \exp(\pi\delta/2)\}$ for $\delta > 0$. We shall show that if $\delta > 0$ is sufficiently small, then for $A, B \in \Omega$

$$|A - B| \geq \frac{1}{4}\tanh[2d_\Omega(A, B)]\left(\frac{1}{\lambda_\Omega(A)} + \frac{1}{\lambda_\Omega(B)}\right).$$

This inequality corresponds to the case $p = 1$; it is the strongest possible lower bound in the corollary and shows that no comparison theorem in the corollary can characterize simply connected regions.

A holomorphic universal covering projection of $\mathbb{D}$ onto Ω is $f(z) = [(1+z)/(1-z)]^{i\delta}$. Then [**13**, p. 128]

$$\sup\{|Q_f(z)| : z \in \mathbb{D}\} = 2\sqrt{1+\delta^2}$$

and [**11**]

$$\sup\{(1-|z|^2)^2|S_f(z)| : z \in \mathbb{D}\} = 2(1+\delta^2).$$

We shall show that

$$|u'(s)| \leq 4$$

and

$$u''(s) + (u')^2(s) \leq 16$$

for δ sufficiently small. This is the case $p = 1$ and $q = 4$ in §2. Note that

$$|u'(s)| \leq |Q_f(z(s))| \leq 2\sqrt{1+\delta^2},$$

so the desired bound on $|u'(s)|$ will hold when $\delta \leq \sqrt{3}$. The other differential inequality will hold if

$$\left|(1-|z|^2)^2 S_f(z) + \frac{1}{2}\left(Q_f(z)\right)^2\right| + \left|Q_f(z)\right|^2 \leq 18,$$

which is weaker than

$$(1-|z|^2)^2|S_f(z)| + \frac{3}{2}|Q_f(z)|^2 \leq 18.$$

The preceding bounds show that this inequality will hold if $8(1+\delta^2) \leq 18$, that is, provided $\delta \leq \sqrt{5}/2$. Thus, both needed inequalities hold when $\delta \leq \sqrt{5}/2$.

The proof of Theorem 2 shows that if $[f(a), f(b)] \subset \Omega$, then

$$|f(a) - f(b)| \geq \frac{1}{4}(\tanh(4L))(|D_1 f(a)| + |D_1 f(b)|).$$

Since $\tanh(t)$ is an increasing function and $d_\Omega(f(a), f(b)) \leq 2L$, this gives

$$|f(a) - f(b)| \geq \frac{1}{4}(\tanh(2d_\Omega(f(a), f(b))))(|D_1 f(a)| + |D_1 f(b)|),$$

or equivalently,

$$|f(a) - f(b)| \geq \frac{1}{4}(\tanh(2d_\Omega(f(a), f(b))))\left(\frac{1}{\lambda_\Omega(f(a))} + \frac{1}{\lambda_\Omega(f(b))}\right).$$

This is the desired result when $[A, B] = [f(a), f(b)] \subset \Omega$. Then, just as in the proof of Theorem 2, this inequality holds even if $[f(a), f(b)]$ does not lie entirely in Ω. In fact, strict inequality holds in this case.

REMARK. If $g(z) = z + a_2z^2 + a_3z^3 + \cdots$ is a normalized close-to-convex function on $\mathbb{D}$, then Wancang Ma [7] has shown

$$\left|a_3 - \frac{2}{3}a_2^2\right| \leq 3 - \frac{2}{3}|a_2|^2$$

with equality if and only if f is a rotation of the Koebe function. Thus, if f is a close-to-convex univalent function, then the inequality in Theorem 2 holds for all $p \geq 1$. Does the inequality in Theorem 2 for $p = 1$ characterize close-to-convex univalent functions? Similarly, the inequality in the corollary to Theorem 2 holds for $p \geq 1$ if the region Ω is close-to-convex.

4. Convex univalent functions and convex regions. We now turn our attention to convex hyperbolic regions and convex univalent functions.

THEOREM 3. *Suppose Ω is a convex hyperbolic region. Then for any $p \geq 1$ and all $A, B \in \Omega$,*

$$|A - B| \geq \frac{\sinh(d_\Omega(A, B))}{[2\cosh(pd_\Omega(A, B))]^{1/p}}\left(\frac{1}{\lambda_\Omega^p(A)} + \frac{1}{\lambda_\Omega^p(B)}\right)^{1/p}.$$

Equality holds if and only if Ω is a half-plane and A and B lie on a line perpendicular to the edge of the half-plane. Conversely, if Ω is a hyperbolic region in $\mathbb{C}$ and the preceding inequality holds for some $p \geq 1$ and all $A, B \in \Omega$, then Ω is convex.

Proof. We first show that a hyperbolic region which satisfies the inequality must be convex. Fix $A \in \Omega$. As in the remark after the corollary to Theorem 2, select $\alpha \in \partial\Omega$ so that $|A - \alpha| = \delta_\Omega(A)$. Let $B \in \Omega$ tend to α along the half-open segment $[A, \alpha)$. Then $d_\Omega(A, B) \to \infty$ and $\lambda_\Omega(B) \to \infty$, so the inequality in the theorem yields $\lambda_\Omega \geq 1/(2\delta_\Omega)$. This inequality characterizes convex regions ([**4**], [**9**]).

Now, we turn to the proof of the inequality when Ω is convex. The proof is very similar to that of Theorem 2. If f is a conformal mapping of $\mathbb{D}$ onto Ω, then $|u'(s)| \leq 2$ is the invariant form of the coefficient bound $|a_2| \leq 1$ for a normalized convex univalent function [**2**, p. 45]. Therefore, we want to use the results from §2 with $q = 2$, so we wish to determine all $p \geq 1$ such that

$$\left|(1-|z|^2)^2 S_f(z) + \frac{p}{2}\left(Q_f(z)\right)^2\right| + \frac{p+1}{2}\left|Q_f(z)\right|^2 - 2 \leq 4p$$

for any convex univalent function f defined on $\mathbb{D}$ and all $z \in \mathbb{D}$. It is easy to verify that if this inequality holds for some value of p, then it also holds for all larger values of p. We shall establish it when $p = 1$:

$$\left|(1-|z|^2)^2 S_f(z) + \frac{1}{2}\left(Q_f(z)\right)^2\right| + \left|Q_f(z)\right|^2 \leq 6. \tag{2}$$

Trimble **[16]** established the following inequality for convex functions when $z = 0$; this was rediscovered and established in invariant form by Harmelin **[3]**:

$$(1-|z|^2)^2\left|S_f(z)\right| + \frac{1}{2}\left|Q_f(z)\right|^2 \leq 2.$$

It is now clear that (2) holds.

Then from §2 with $q = 2$, we have

$$u''(s) + p(u')^2(s) \leq 4p.$$

Given $A, B \in \Omega$, select $a, b \in \mathbb{D}$ with $f(a) = A$ and $f(b) = B$. Since Ω is convex, the straight line segment $[f(a), f(b)]$ always lies in Ω. Then we get

$$|f(a) - f(b)| = \int_{-L}^{L} \exp u(s)\,ds \geq \int_{-L}^{L} \exp v(s)\,ds \geq C\sinh(2L),$$

where

$$C = \left(\frac{|D_1 f(a)|^p + |D_1 f(b)|^p}{2\cosh(2pL)}\right)^{1/p}.$$

Thus,

$$|f(a) - f(b)| \geq \frac{\sinh(2L)}{[2\cosh(2pL)]^{1/p}}\left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p},$$

or

$$|A - B| \geq \frac{\sinh(2L)}{[2\cosh(2pL)]^{1/p}}\left(\frac{1}{\lambda_\Omega^p(A)} + \frac{1}{\lambda_\Omega^p(B)}\right)^{1/p}.$$

Recall that $2L$ denotes the hyperbolic length (relative to Ω) of the segment $[A, B]$. Since the function $h(t) = \sinh(t)/[2\cosh(pt)]^{1/p}$ is increasing and $d_\Omega(A, B) \leq 2L$, we obtain

$$|A - B| \geq \frac{\sinh(d_\Omega(A, B))}{[2\cosh(p d_\Omega(A, B))]^{1/p}}\left(\frac{1}{\lambda_\Omega^p(A)} + \frac{1}{\lambda_\Omega^p(B)}\right)^{1/p}.$$

This establishes the lower bound.

Finally, we determine when equality holds. First, suppose $p > 1$. If equality holds, then $[A, B]$ must be a hyperbolic geodesic. There is no harm in assuming $[A, B] \subset \mathbb{R}$ and is symmetric about the origin with $A < 0$ and $B = -A$; if this were not true, apply a conformal automorphism of $\mathbb{C}$ to Ω. Now, γ is a hyperbolic geodesic in $\mathbb{D}$; by performing a conformal automorphism of $\mathbb{D}$ if necessary, we may assume that $\gamma \subset (-1, 1)$ and is symmetric about the origin. The hyperbolic arclength parametrization of the path γ is $z(s) = \tanh(s)$ and $f'(z(s)) > 0$ for $s \in [-L, L]$. Symmetry implies $f(0) = 0$. Equality forces $\exp(u) = C\exp(\pm 2s)$. We consider the plus sign; the case of the minus sign is similar. As in the proof of Theorem 2, we obtain

$$f'(z) = \frac{C}{(1-z)^2}.$$

Since $f(0) = 0$, $f(z) = CK(z)$, where $K(z) = z/(1-z)$. In this situation $\Omega = f(\mathbb{D})$ is a half-plane and the segment $[A, B]$ is orthogonal to the edge of the half-plane. Conversely, if Ω is a half-plane, it is straightforward to show that equality holds whenever $[A, B]$ is orthogonal to the edge of the half-plane. It is sufficient to verify this for the special case of the upper half-plane $\mathbb{H} = \{z: \operatorname{Im} z > 0\}$. In this case,

$$d_{\mathbb{H}}(A, B) = \operatorname{artanh}\left|\frac{A-B}{A-\overline{B}}\right| \quad \text{and} \quad \lambda_{\mathbb{H}}(z) = \frac{1}{2\operatorname{Im}(z)}.$$

We omit the details.

It remains to consider the case of equality when $p = 1$. In this situation Lemma 1 does not apply, so we use a different method. If Ω is not a half-plane, then $|u'(s)| < 2$ and $u''(s) + (u')^2(s) < 4$. These strict inequalities imply that equality cannot hold in this case. Thus, we need only determine necessary and sufficient conditions for equality when Ω is a half-plane. Because of the invariance of the inequality under conformal automorphisms of $\mathbb{C}$, we may assume Ω is the upper half-plane $\mathbb{H} = \{z: \operatorname{Im} z > 0\}$. We need to determine when equality holds in

$$(3) \qquad |A-B| \geq \frac{1}{2}\tanh(d_{\mathbb{H}}(A, B))\left(\frac{1}{\lambda_{\mathbb{H}}(A)} + \frac{1}{\lambda_{\mathbb{H}}(B)}\right).$$

Inequality (3) is equivalent to

$$|A-\overline{B}| \geq \operatorname{Im}(A) + \operatorname{Im}(B).$$

But trivially

$$|A-\overline{B}| \geq \operatorname{Im}(A-\overline{B}) = \operatorname{Im}(A) + \operatorname{Im}(B)$$

with equality if and only if $\operatorname{Re}(A-\overline{B}) = 0$, that is, $\operatorname{Re} A = \operatorname{Re}\overline{B} = \operatorname{Re} B$. In geometric terms this necessary and sufficient condition for equality is that $[A, B]$ be orthogonal to the real axis, the edge of $\mathbb{H}$.

COROLLARY. *Suppose f is univalent in $\mathbb{D}$ and $f(\mathbb{D})$ is a convex region. Then for $p \geq 1$ and all $a, b \in \mathbb{D}$,*

(4)

$$|f(a)-f(b)| \geq \frac{\sinh(d_{\mathbb{D}}(a,b))}{[2\cosh(p d_{\mathbb{D}}(a,b))]^{1/p}} \left(|D_1 f(a)|^p + |D_1 f(b)|^p\right)^{1/p}.$$

Equality holds for distinct $a, b \in \mathbb{D}$ if and only if $f = S \circ K \circ T$, where S is a conformal automorphism of $\mathbb{C}$, $K(z) = z/(1-z)$ and T is a conformal automorphism of $\mathbb{D}$, and a and b lie on any axis of symmetry of f. Conversely, if a nonconstant holomorphic function f defined on $\mathbb{D}$ satisfies this inequality for some $p \geq 1$, then f is univalent on $\mathbb{D}$ and $f(\mathbb{D})$ is a convex region.

Proof. Suppose f is convex univalent in $\mathbb{D}$. Set $\Omega = f(\mathbb{D})$. Then the inequality and the necessary and sufficient conditions for equality follow from applying Theorem 3 to Ω and the points $A = f(a)$ and $B = f(b)$.

Conversely, suppose f is a nonconstant holomorphic function defined on $\mathbb{D}$ which satisfies the inequality. As in the proof of the invariant form of the Koebe distortion theorem, we conclude that f is univalent on $\mathbb{D}$. Set $\Omega = f(\mathbb{D})$. Since f is a conformal map of $\mathbb{D}$ onto Ω and hyperbolic distance is preserved, inequality (4) implies that the inequality in the theorem holds. Hence, Ω is convex, so f is convex univalent.

REMARK. The right-hand side of the inequality in the corollary is a decreasing function of p for $p \geq 1$. Therefore, the strongest necessary condition for a convex univalent function that the corollary produces is the case $p = 1$:

$$|f(a)-f(b)| \geq \frac{1}{2}\tanh(d_{\mathbb{D}}(a,b))\left(|D_1 f(a)| + |D_1 f(b)|\right).$$

The weakest sufficient condition for convex univalence that the corollary yields is $p = \infty$ (or more precisely, the limit as $p \to \infty$):

$$|f(a)-f(b)| \geq \frac{\sinh(d_{\mathbb{D}}(a,b))}{\exp(d_{\mathbb{D}}(a,b))} \max\{|D_1 f(a)|, |D_1 f(b)|\}.$$

This is the symmetric, linearly invariant form of the classical distortion theorem

$$|g(z)| \geq \frac{|z|}{1+|z|}, \qquad z \in \mathbb{D},$$

for a normalized convex univalent function g [**2**, p. 70].

5. Comments. The method of Blatter that we have employed in this paper uses certain differential geometric ideas in conjunction with coefficient bounds for univalent functions to produce symmetric, linearly invariant two-point distortion theorems for (convex) univalent functions which characterize (convex) univalence. Can these results be established in a purely differential geometric fashion without using coefficient bounds? In the convex case our results characterize convex regions so it is plausible that, at least in this setting, a purely differential geometric proof might be available.

REFERENCES

[1] C. Blatter *Ein Verzerrungssatz für schlichte Funktionen*, Comment. Math. Helv., **53** (1978), 651–659.

[2] P. Duren, *Univalent Functions*, Grundlehren der math. Wissenschaften, vol. 259, Springer-Verlag, New York, 1983.

[3] R. Harmelin, *Locally convex functions and the Schwarzian derivative*, Israel J. Math., **67** (1989), 367–379.

[4] J. R. Hilditch, *The hyperbolic metric and the distance to the boundary in plane domains*, unpublished manuscript, circa 1985.

[5] J. A. Jenkins, *On certain coefficients of univalent functions*, Analytic Functions, Princeton University Press, 1960, pp. 159–164.

[6] I. Kra, *Automorphic Functions and Kleinian Groups*, W. A. Benjamin, Reading, Mass., 1972.

[7] W. Ma, personal communication.

[8] W. Ma and D. Minda, *Euclidean linear invariance and uniform local convexity*, J. Australian Math. Soc., **52** (1992), 401–418.

[9] D. Mejia and D. Minda, *Hyperbolic geometry in k-convex regions*, Pacific J. Math., **141** (1990), 333–354.

[10] D. Minda *Lower bounds for the hyperbolic metric in convex regions*, Rocky Mountain J. Math., **13** (1983), 61–69.

[11] ——, *The Schwarzian derivative and univalence criteria*, Contemp. Math., **38** (1985), 43–52.

[12] D. Minda and D. J. Wright, *Univalence criteria and the hyperbolic metric*, Rocky Mountain J. Math., **12** (1982), 471–479.

[13] Chr. Pommerenke, *Linear-invariante Familien analytischer Funktionen* I, Math. Ann., **155** (1964), 108–154.

[14] ——, *Univalent Functions*, Vandenhoeck and Ruprecht, Göttingen, 1975.

[15] S. Ruscheweyh, personal communication.
[16] S. Y. Trimble, *A coefficient inequality for convex univalent functions*, Proc. Amer. Math. Soc., **48** (1975), 266–267.

Received March 15, 1992 and in revised form January 6, 1993. Research supported in part by National Science Foundation grant DMS-9008051. Research of the first author supported in part by the Global Analysis Research Center, Seoul National University, Seoul, Korea.

UNIVERSITY OF CINCINNATI
CINCINNATI, OH 45221-0025

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

VALUE DISTRIBUTION OF THE GAUSS MAP AND THE TOTAL CURVATURE OF COMPLETE MINIMAL SURFACE IN R^m

XIAOKANG MO

The aim of this paper is to prove the following

THEOREM. ***Let S be a complete non-degenerate minimal surface in R^m such that its generalized Gauss map f intersects only a finite number of times the hyperplanes $A_1, \dots, A_q$ in CP^{m-1} in general position. If $q > m(m+1)/2$, then S must have finite total curvature.***

1. Introduction. The study of the value distribution property of Gauss map of minimal surface began with a series of papers by Osserman [**9**], [**11**] and the results can be summarized in the following

THEOREM (*R. Osserman*). *Let S be a complete minimal surface in R^3. Then*

S has infinite total curvature $\Leftrightarrow$ the Gauss map of S takes on all directions infinitely often with the exception of at most a set of logarithmic capacity zero;

S has finite non-zero total curvature $\Leftrightarrow$ the Gauss map of S takes on all directions a finite number of times, omitting at most three directions;

S has zero total curvature $\Leftrightarrow$ S is a plane.

For a long time, the above theorem had been the best result on this direction. But all the known examples indicated that the exceptional set of logarithmic capacity should be a finite set. In 1981, Xavier made a surprising breakthrough by proving the following result, using a result of Yau about a differential equation on complete Riemannian manifold.

THEOREM (*F. Xavier* [**13**]). *Let S be a complete minimal surface in R^3. Then its Gauss map can omit at most six directions unless it is a plane.*

In 1988, Fujimoto finally found a way to arrive at the best possible number 4.

THEOREM (*H. Fujimoto* [4]). *Let S be a complete minimal surface in R^3. Then its Gauss map can omit at most* 4 *directions unless it is a plane.*

A combination of Osserman's early study and Fujimoto's above work gives the following

THEOREM (*X. Mo and R. Osserman* [8]). *Let S be a complete minimal surface in R^3 with infinite total curvature. Then its Gauss map must take every direction infinitely often except at most 4 directions.*

For a surface in R^m there is the following

THEOREM (*H. Fujimoto* [5]). *Let S be a complete minimal surface in R^m with nondegenerate Gauss map. Then the image of S under the Gauss map cannot fail to intersect more than $m(m+1)/2$ hyperplanes in general position in CP^{m-1}.*

And the result of this paper mentioned at the beginning of this section is the infinite covering property corresponding to the above theorem.

An oriented minimal surface S in R^m may be described by a conformal immersion

$$X\colon M \to R^m, \quad X = (x_1, \dots, x_m),$$

where M is a Riemann surface and each x_k is a harmonic function on M.

By definition, the generalized Gauss map of S is the map that assigns to each point of S the tangent plane of S at that point. Because the tangent space of R^m at every point is naturally identified with R^m itself, the range of the Gauss map is the Grassmannian manifold consisting of all the oriented 2-subspaces of R^m. We can further identify the 2-plane spanned by the orthonormal basis X, Y with the line in C^m generated by $(X - iY)/2$. So the range of the Gauss map can be thought of as $P^{m-1}(C)$.

Let $z = u + iv$ be a holomorphic local coordinate of M. Denote

$$\frac{\partial x}{\partial z} = \frac{1}{2}\left(\frac{\partial x_1}{\partial u} - i\frac{\partial x_1}{\partial v}, \dots, \frac{\partial x_m}{\partial u} - i\frac{\partial x_m}{\partial v}\right),$$

by $F = (f_0, \dots, f_n)$, where $n = m - 1$; $f = (f_0 : f_1 : \cdots : f_n)$ is the point in CP^n represented by $(f_0, \dots, f_n)$ in C^m. Then the holomorphic map f represents the Gauss map, and the metric on M as a minimal surface is

$$ds^2 = 2|F(z)|^2|dz|^2,$$

where $|F|^2 = |f_0|^2 + \cdots + |f_n|^2$.

In this way, we turn the problem of the Gauss map partly into a problem on holomorphic curves. The value distribution property of the holomorphic curve may lead to corresponding results about the Gauss map.

In §2 we will summarize some of the basic ideas and notation of holomorphic curves. We will also introduce an important construction of Cowen and Griffiths [2] on holomorphic curves in CP^n which was the basis of their remarkable proof of Ahlfors' defect relation. In §3 we will present the proof of Fujimoto's inequality in such a way that will clarify the relation between Cowen and Griffiths' construction and Fujimoto's. Fujimoto's inequality is the key to both the proof of his theorem mentioned above and the proof of our result. In §4, we will give the proof of our result.

2. Some properties of holomorphic curves. Value distribution properties of holomorphic curves have been studied since the end of the 19th century. The central problem was to generalize the Picard theorem and the Nevanlinna defect relation for entire functions to the case of holomorphic curves. This was finally achieved in 1941 by L. Ahlfors, overcoming great technical difficulties.

In 1976, M. Cowen and P. Griffiths [2] gave a much simpler proof of Ahlfors' result using what they called a "negatively curved collection of metrics". Using their result, H. Fujimoto [5] was able to construct a single metric of negative curvature under certain conditions. Then by the Schwarz-Pick lemma, he derived an inequality which is the key to the study of the value distribution property of the Gauss map of minimal surface. In this section, we will give an outline of Cowen and Griffiths' result.

Let $\Delta_R = \{z \,|\, |z| < R\}$ be a disk in the complex plane, $f\colon \Delta_R \to P^n(C)$ be a holomorphic curve derived from a holomorphic map $F\colon \Delta_R \to C^{n+1}$ through homogeneous coordinates. $F(z) = (f_0(z), \ldots, f_n(z))$, $f_0, \ldots, f_n$ are holomorphic functions on Δ_R. We write $f = (f_0 : \cdots : f_n)$ and define $|F| = (\sum_{i=1}^n |f_i|^2)^{1/2}$; for our purposes, we assume that $|F| \neq 0$.

Take the l-th derivative:

$$F^{(l)}(z) = (f_0^{(l)}(z), \ldots, f_n^{(l)}(z)).$$

Define $F_k = F^{(0)} \wedge F^{(l)} \wedge \cdots \wedge F^{(k)}$: $\Delta_R \to \bigwedge^{k+1} C^{k+1} \subset G(n, k)$, where $G(n, k)$ is the Grassmannian manifold. By the Plücker embedding $G(n, k) \subset P^N(C)$, $N = \binom{n+1}{k+1} - 1$, F_k induces a map f_k: $\Delta_R \to P^N(C)$, called the kth derived curve of f.

We can define $|F_k|$ in a natural way. Let $e_0, \ldots, e_n$ be the standard basis of C^{n+1},

$$F(z) = F^{(0)} \wedge \cdots \wedge F^{(k)}(z) = \sum_{i_0 < \cdots i_k} F_{i_0 < \cdots < i_k} e_{i_0} \wedge \cdots \wedge e_{i_k},$$

and we define

$$|F_k(z)| = \left(\sum_{i_0 < \cdots < i_k} |F_{i_0 \cdots i_k}(z)|^2 \right)^{1/2}.$$

Now the Fubini-Study metrics on P^n and P^N naturally induce metrics on Δ_R by pulling back:

$$\Omega_0 = dd^c \log |F_0|^{1/2} = dd^c \log |F|^2,$$
$$\Omega_k = dd^c \log |F_k|^2, \qquad k = 1, \ldots, n,$$

where $d^c = (\sqrt{-1}/4\pi)(\bar{\partial} - \partial)$. Because F_n is just a holomorphic function, $\Omega_n = 0$. We also set $|F_{-1}| = 1$ for convenience, so $\Omega_{-1} = 0$.

The metrics Ω_k will be used later to construct the negatively curved collection of metrics.

Let $a = (a_0, \ldots, a_n) \in C^{n+1}$, $|a| = (\sum_{i=1}^n |a_i|^2)^{1/2} = 1$. Then

$$a_0 w_0 + \cdots + a_n w_n = 0$$

defines a hyperplane A in both C^{n+1} and P_n, and

$$F(A) = a_0 f_0 + \cdots + a_n f_n$$

measures the distance from $F(z)$ to A; in a similar way

$$|F_k(A)|^2 = \sum_{i_1 < \cdots < i_k} \left| \sum_{j \neq i_1, \ldots, i_k} a_j F_{j_{i_1 \cdots i_k}} \right|^2$$

measures how far F_k is from A. Here $F_{j_{i_1 \cdots i_k}} = \text{sign}(\sigma) F_{j_0 \cdots j_k}$, σ is the permutation

$$\sigma = \begin{pmatrix} j & i_1 \cdots i_k \\ j_0 & j_1 \cdots j_k \end{pmatrix}.$$

In fact $|F_k(A)(z_0)| = 0$ means $F(z_0), F^{(1)}(z_0), \ldots, F^{(k)}(z_0)$ all lie in the hyperplane

$$a_0 w_0 + \cdots + a_n w_n = 0.$$

The corresponding quantities for the holomorphic curve f in P^n are

$$\phi_0(A) = \frac{|F(A)|^2}{|F_k|^2}, \qquad \phi_k(A) = \frac{|F_k(A)|^2}{|F_k|^2},$$

and if $\phi_k(A)(z_0) = 0$, the curve f is said to have contact of order $k+1$ with A at z_0.

Now if the holomorphic curve $f\colon \Delta_R \to P^n$ omits a certain number of hyperplanes $A_1, A_2, \ldots, A_q$, we want to construct a metric or a collection of metrics that is negatively curved.

If $n = 1$, $A_1, \ldots, A_q$ are points on P^1, we can just pull back the Poincaré metric of $P^1 - \{A_1, \ldots, A_q\}$. To be more explicit, let us take a local coordinate ζ of P^1 around a neighborhood of A_1 (or any other A_i, $i = 1, \ldots, q$), with $\zeta = 0$ at A^1. Then the Poincaré metric is asymptotically

$$\frac{d\zeta \wedge d\bar{\zeta}}{|\zeta|^2 \log^2(1/|\zeta|^2)}$$

around the point A_1. Cowen and Griffiths [2] found a way to generalize this construction to the case when $n \geq 2$. In that case, it becomes necessary to consider not only f but all of its derived curves f_k. The quantity $|\zeta|^2$ for A_1 will be replaced by $\phi_k(A_1)$ as defined above.

Let $\omega = (\sqrt{-1}/2\pi) h(z)\, dz \wedge d\bar{z}$ be a metric. Then the Ricci form is defined by $\operatorname{Ric}\omega = dd^c \log h(z)$, and $\operatorname{Ric}\omega \geq \omega$ is equivalent to the fact that the curvature of ω is less than -1.

Let $A_1, \ldots, A_q$ be hyperplanes in general position in P^n and $q \geq n+2$. For $i = 0, \ldots, n-1$, following the indication of the Poincaré metric, define

$$\omega_i = c_i \prod_{\nu=1}^{q} \left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu) \log^2(\mu/\phi_i(A_\nu))} \right)^{1/(n-1)} \Omega_i.$$

Cowen and Griffiths [2] proved the following

PROPOSITION. *Given $\varepsilon > 0$, for a suitable choice of constants c_i, and μ, we have*

$$\sum_{i=0}^{n-1} (n-i) \operatorname{Ric}\omega_i \geq (q-(n+1))\Omega_0 + \sum_{i=0}^{n-1} \omega_i - \varepsilon \left(\sum_{i=0}^{n-1} \Omega_i \right).$$

Aside from the term with the ε, this inequality illustrates what is meant by saying that the collection of metrics $\{\omega_i\}$ is negatively curved. Based on this, Fujimoto constructed a single metric with negative curvature under some additional assumptions. The next section will give a detailed presentation of Fujimoto's construction.

3. Fujimoto's inequality. This section will be centered around curvature computations. For this purpose, a few lemmas from **[2]** are collected here for convenience.

We have defined $\Omega_k = dd^c \log |F_k|^2$,

LEMMA 1.

$$\Omega_k = \frac{\sqrt{-1}}{2\pi} \frac{|F_{k-1}|^2 |F_{k+1}|^2}{|F_k|^4} dz \wedge d\overline{z}.$$

LEMMA 2. *Define*

$$h_k = \frac{|F_{k-1}^2| \, |F_{k+1}|^2}{|F_k|^4};$$

then

$$\operatorname{Ric} \Omega_k = dd^c \log h_k = \Omega_{k+1} + \Omega_{k-1} - 2\Omega_k.$$

In the process of computation, we will use these two lemmas whenever necessary without referring to them explicitly.

To help understanding, we give here an outline of the idea of the proof of this section. The motivation is to construct a single metric of negative curvature out of a collection of negatively curved metrics.

Let $\omega_i = (\sqrt{-1}/2\pi) h_i(z) \, dz \wedge d\overline{z}$, and suppose

$$\sum \operatorname{Ric} \omega_i \geq \sum \omega_i.$$

Then

$$\sum_i dd^c \log h_i \geq \sum h_i dz \wedge d\overline{z},$$

$$dd^c \log \left(\prod h_i\right) \geq \left(\sum h_i\right) dz \wedge d\overline{z} \geq n \left(\prod h_i\right)^{1/n} dz \wedge d\overline{z},$$

$$dd^c \log \left(\prod h_i\right)^{1/n} \geq \left(\prod h_i\right)^{1/n} dz \wedge d\overline{z},$$

so $\omega = (\prod h_i)^{1/n} dz \wedge d\overline{z}$ satisfies $\operatorname{Ric} \omega \geq \omega$ and ω is the desired metric. In our situation, there are two other factors that complicate the proof. One is that in the proposition of the last section, the collection of metrics is not strictly negatively curved; the term with ε will cause some complications. The other factor is that there are many computations and cancellations due to the special form of metrics that we have. Let us start with the inequality

$$\sum_{i=0}^{n-1}(n-i)\mathrm{Ric}\,\omega_i \geq (q-(n+1))\Omega_0 + \sum_{i=0}^{n-1}\omega_i - \varepsilon\left(\sum_{i=0}^{n-1}\Omega_i\right),$$

where

$$\omega_i = c_i \prod_{\nu=1}^{q}\left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))}\right)^{1(n-1)}\Omega_i.$$

We want to compute each term of the inequality explicitly.

Step 1.

$$\begin{aligned}
&\sum_{i=0}^{n-1}(n-i)\mathrm{Ric}\,\omega_i \\
&\quad= \sum_{i=0}^{n-1}(n-i)dd^c\log\prod_{\nu=1}^{q}\left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))}\right)^{1/(n-1)} \\
&\quad\quad+\sum_{i=0}^{n-1}(n-i)\mathrm{Ric}\,\Omega_i \\
&\quad= \sum_{i=0}^{n-1}dd^c\log\prod_{\nu=1}^{q}\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))} \\
&\quad\quad+\sum_{i=0}^{n-1}(\Omega_{i+1}+\Omega_{i-1}-2\Omega_i) \\
&\quad= dd^c\prod_{\nu=1}^{q}\frac{\phi_n(A_\nu)}{\phi_0(A_\nu)\prod_i\log^2(\mu/\phi_i(A_\nu))} - (n+1)\Omega_0;
\end{aligned}$$

but $\phi_0(A_\nu)=1$, $\phi_0(A_\nu)=|F(A_\nu)|^2/|F|^2$, so

$$\begin{aligned}
\sum_{i=0}^{n-1}(n-i)\mathrm{Ric}\,\omega_i &= dd^c\log\prod_{\nu=1}^{q}\left(\frac{|F|^2}{|F(A_\nu)|^2\prod_i\log^2(\mu/\phi_i(A_\nu))}\right) \\
&\quad-(n+1)dd^c\log|F|^2 \\
&= dd^c\log\left(\frac{|F|^{2q}}{\prod_{\nu=1}^{q}(|F(A_\nu)|^2\prod_i\log^2(\mu/\phi_i(A_\nu)))}\right) \\
&\quad- dd^c\log|F|^{2(n+1)} \\
&= dd^c\log\left(\frac{|F|^{2(q-(n+1))}}{\prod_{\nu=1}^{q}(|F(A_\nu)|^2\prod_i\log^2(\mu/\phi_i(A_\nu)))}\right).
\end{aligned}$$

Step 2.

$$\sum_{i=0}^{n-1}\omega_i=\sum_{i=0}^{n-1}c_i\prod_{\nu=1}^{q}\left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))}\right)^{1(n-1)}\Omega_i$$
$$=\sum_{i=0}^{n-1}c_i\prod_{\nu=1}^{q}\left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))}h_i^{n-i}\right)^{1/(n-1)}dz\wedge d^cz\,,$$

where $\Omega_i=h_i dz\wedge d^cz$. Using the inequality

$$a_1x_1+\cdots+a_nx_n\geq(a_1+\cdots+a_n)(x_1^{a_1}\cdots x_n^{a_n})^{1/(a_1+\cdots+a_n)}$$

with $a_i=n-i$, $\sum_{i=0}^{n-1}a_i=n(n+1)/2$, we have

$$\sum_{i=0}^{n-1}\omega_i\geq C\left(\prod_{i=1}^{n-1}\left(\prod_{\nu=1}^{q}\left(\frac{\phi_{i+1}(A_\nu)}{\phi_i(A_\nu)\log^2(\mu/\phi_i(A_\nu))}\right)h_i^{n-i}\right)^{2/n(n+1)}\right)dz\wedge d^cz$$
$$=C\left(\prod_{\nu=1}^{q}\left(\frac{\phi_n(A_\nu)}{\phi_0(A_\nu)\prod_{i=1}^{n-1}\log^2(\mu/\phi_i(A_\nu))}\right)\cdot\prod_{i=0}^{n-1}\left(\frac{|F_{i-1}|^2|F_{i+1}|^2}{|F_i|^4}\right)^{n-1}\right)^{2/n(n+1)}$$
$$=C\left(\prod_{\nu=1}^{q}\left(\frac{|F|^2}{|F(A_\nu)|^2\prod_{i=1}^{n-1}\log^2(\mu/\phi_i(A_\nu))}\right)\cdot\prod_{i=0}^{n-1}\left(\frac{|F_n|^2}{|F_0|^{2(n+1)}}\right)\right)^{2/n(n+1)},$$

but $|F_0|=|F|$, so

$$\sum_{i=0}^{n-1}\omega_i\geq C\left(\frac{|F|^{2(q-(n+1))|F_n|^2}}{\prod_{\nu=1}^{q}(|F(A_\nu)|^2\prod_{i=1}^{n-1}\log^2(\mu/\phi_i(A_\nu)))}\right)^{2/n(n+1)}.$$

Step 3.

$$\varepsilon\left(\sum_{i=1}^{n-1}\Omega_i\right)=\varepsilon\sum_{i=1}^{n-1}dd^c\log|F_k|^2=dd^c\log|F_o|^{2\varepsilon}\cdots|F_{n-1}|^{2\varepsilon}\,,$$
$$(q-(n+1))\Omega_0=(q-(n+1))dd^c\log|F_0|^2=dd^c\log|F|^{2(q-(n+1))}.$$

Step 4. Combining the results of Steps 1, 2, 3, we have

$$dd^c \log \frac{|F_0|^{2\varepsilon} \cdots |F_{n-1}|^{2\varepsilon}}{\prod_{\nu=1}^{q}(|F(A_\nu)|^2 \prod_{i=1}^{n-1} \log^2(\mu/\phi_i(A_\nu)))} \geq C \left(\frac{|F|^{2(q-(n+1))}|F_n|^2}{\prod_{\nu=1}^{q}(|F(A_\nu)|^2 \prod_{i=1}^{n-1} \log^2(\mu/\phi_i(A_\nu)))} \right)^{2/n(n+1)} dz \wedge d^c z.$$

Setting $G = \prod_{\nu=1}^{q}(|F(A_\nu)|^2 \prod_{i=1}^{n-1} \log^2(\mu/\phi_i(A_\nu)))$, we have

$$dd^c \log |F_0|^{2\varepsilon} \cdots |F_{n-1}|^{2\varepsilon} + dd^c \log(1/G^2) \geq C \left(\frac{|F|^{2(q-(n+1))}|F_n|^2}{G^2} \right)^{2/n(n+1)} dz \wedge d^c z.$$

Step 5. Notice that F_n is a holomorphic function, so $dd^c \log |F| = 0$; also $\log |F|^2$ is subharmonic, so $dd^c \log |F|^2$, the -4ε in the exponent is necessary and we will see the reason in the arguments later. With $\eta = (|F|^{2(q-(n+1))}|F_n|^2)/G^2$, we have

$$\varepsilon dd^c \log |F_0|^2 \cdots |F_{n-1}|^2 + dd^c \log \frac{\eta}{|F|^{4\varepsilon P_{n+1}}} \geq C \eta^{2/n(n+1)} dz \wedge d^c z.$$

Step 6. Let $P_n = n(n+1)/2$, $Q_n = \sum_{k=1}^{n} P_k$. Then

$$\begin{aligned} P_n dd^c \log |F_0|^2 \cdots |F_{n-1}|^2 &\geq \left(P_n \frac{|F_1|^2}{|F_0|^4} + P_{n-1} \frac{|F_0|^2|F_2|^2}{|F_1|^4} + \cdots + P_1 \frac{|F_{n-2}|^2|F_n|^2}{|F_{n-1}|^4} \right) dz \wedge d^c z \\ &\geq Q_n \left(\left(\frac{|F_1|^2}{|F_0|^4} \right)^{P_n} \left(\frac{|F_0|^2|F_2|^2}{|F_1|^4} \right)^{P_{n-1}} \cdots \left(\frac{|F_{n-2}|^2|F_n|^2}{|F_{n-1}|^4} \right)^{P_1} \right)^{1/Q_n} dz \wedge d^c z \\ &= Q_n \left(\frac{|F_1|^2 \cdots |F_{n-1}|^2|F_n|^2}{|F_0|^{n^2+3n}} \right)^{1/Q_n} dz \wedge d^c z, \end{aligned}$$

so

$$\varepsilon dd^c \log |F_0|^2 \cdots |F_{n-1}|^2 \geq \varepsilon \frac{Q_n}{P_n} \left(\frac{|F_0|^2 \cdots |F_n|^2}{|F_0|^{2P_{n+1}}} \right)^{1/Q_n} dz \wedge d^c z.$$

Step 7. Add up the results of Steps 5 and 6, replace the ε (which is arbitrary) with $\varepsilon/2$ and notice that $dd^c \log|F_n|^2 = 0$, we have

$$\varepsilon dd^c \log|F_0|^2 \cdots |F_n|^2 + dd^c \log \frac{\eta}{|F|^{2\varepsilon P_{n+1}}}$$
$$\geq \left(C\eta^{1/P_n} + \frac{\varepsilon}{2}\frac{Q_n}{P_n}\left(\frac{|F_0|^2 \cdots |F_n|^2}{|F_0|^{2P_{n+1}}}\right)^{1/Q_n}\right) dz \wedge d^c z,$$

using $a_1 x_1 + a_2 x_2 \geq (a_1 + a_2)(x_1^{a_1} x_2^{a_2})^{1/(a_1+a_2)}$ with $a_1 = P_n$, $a_2 = \varepsilon Q_n$, we have

$$dd^c \log \frac{|F_0|^{2\varepsilon} \cdots |F_n|^{2\varepsilon}}{|F|^{2\varepsilon P_{n+1}}}\eta \geq C_1 \left(\frac{|F_0|^{2\varepsilon} \cdots |F_n|^{2\varepsilon}\eta}{|F|^{2\varepsilon P_{n+1}}}\right)^{1/(P_n+\varepsilon Q_n)} dz \wedge d^c z.$$

Set

$$h = \left(\frac{|F_0|^{2\varepsilon} \cdots |F_n|^{2\varepsilon}\eta}{|F|^{2\varepsilon M_{n+1}}}\right)^{1/(P_n+\varepsilon Q_n)};$$

then

$$dd^c h \geq C_2 h dz \wedge d^c z,$$

so $hdz \wedge d^c <$ is the desired metric.

Step 8. By the Schwarz-Pick lemma, we have a constant C_3 such that

$$h(z) \leq C_3 \frac{2R}{R^2 - |z|^2},$$

where $\frac{2R}{R^2-|z|^2} dz \wedge d^c z$ is the Poincaré metric of the disk $\{z||z| < R\}$. Writing out everything explicitly, we have

$$\frac{|F|^{q-(n+1)-\varepsilon P_{n+1}}|F_0|^{\varepsilon} \cdots |F_{n-1}|^{\varepsilon}|F_n|^{\varepsilon}}{\prod_{\nu=1}^{q}\prod_{i=1}^{n}\log(\mu/\phi_i(A_\nu))} \leq C_4 \left(\frac{2R}{R^2 - |z|^2}\right)^{P_n+Q_n\varepsilon}.$$

Step 9. We would like to get rid of the log terms. Knowing that

$$K = \sup_{0 \leq x \leq 1} x^{\varepsilon/2q} \log \frac{\mu}{x} < +\infty \quad \text{for } \mu > 1,$$

we have

$$\frac{1}{\log(\mu/\phi_k(A_\nu))} \geq \frac{1}{K}\phi_k(A_\nu)^{\varepsilon/2q} = \frac{1}{K}\frac{|F_k(A_\nu)|^{\varepsilon/q}}{|F_k|^{\varepsilon/q}},$$

substituting this into the result of Step 8, we have

PROPOSITION (*Fujimoto's inequality* [5]). *Let* $\Delta_R = \{z||z| < R\}$ *be a disk in the complex plane,* $f\colon \Delta_R \to CP^n$ *be a holomorphic curve* *derived from a holomorphic map* $F\colon \Delta_R \to C^{n+1}$, *using the notations introduced in the previous section, we have the following statement. For any* $\varepsilon > 0$, *there is a* $C > 0$, *such that*

$$\frac{|F|^{q-(n+1)-\varepsilon P_{n+1}} \prod_{i=1}^{n} (\prod_{\nu=1}^{q} |F_k(A_\nu)|)^{\varepsilon/q} |F_n|^{1+\varepsilon}}{\prod_{\nu=1}^{q} |F(A_\nu)|} \leq C \left(\frac{2R}{R^2 - |z|^2} \right)^{P_n + Q_n \varepsilon}.$$

4. Minimal surfaces in R^m. We assume that all surfaces are orientable, since analogous theorems for non-orientable surfaces are easily formulated by taking the two sheeted orientable covering surface and applying the theorem to it. Following the notation of the previous section, we will prove the following

THEOREM. *Let* S *be a complete non-degenerate minimal surface in* R^m *such that the Gauss map* $f = (f_0 \colon \cdots \colon f_n)$ (*here* $n = m = 1$) *intersects only a finite number of times the hyperplanes* $A_1, \ldots, A_q$ (*in* CP^n) *in general position. If* $q > m(m+1)/2 = (n+1) + n(n+1)/2$, *then* S *must have finite total curvature.*

REMARK. If S is a generalized minimal surface with a finite number of branch points, all the arguments of our proof will not be affected. So the theorem is also true for the somewhat more general class of surfaces. This also applies to the similar theorem for surfaces in R^3 by Mo and Osserman [**8**].

It was already observed by Osserman (see R. Osserman, *A survey of minimal surfaces*, second edition, 1986, p. 73) that his classic results on the value distribution of Gauss map is true for simply connected surfaces with a finite number of branch points. An observation of Ahlfors implies that they are still true if a certain condition on the distribution of the branching points is satisfied. But there exist complete generalized minimal surfaces in R^3, not lying in a plane, whose Gauss map lies in an arbitrarily small neighborhood on the sphere. So the results are not true for arbitrary generalized minimal surfaces. The method of our proof is similar to the method of [**8**].

Proof. Step 1. Since f is non-degenerate, none of the $F_k(A_\nu)$ vanishes identically, where $\nu = 1, \ldots, q$, $k = 0, \ldots, n$. Let A be given by the equations

$$a_{\nu_0} z_0 + \cdots + a_{\nu_n} z_n = 0,$$

$$F_k = \sum_{i_0 < \cdots < i_k} F_{i_0 \cdots i_k} e_{i_0} \wedge \cdots \wedge e_{i_k},$$

$$F_{j_0 \cdots j_k} = \operatorname{sign}\begin{pmatrix} i_0 \cdots i_k \\ j_0 \cdots j_k \end{pmatrix} F_{i_0 \cdots i_k},$$

then for each pair (ν, k), there is $i_1, \ldots, i_k$ such that

$$\psi_{\nu k} = \sum_{l \neq i_1, \ldots, i_k} a_{\nu l} F_{l_{i_1 \cdots i_k}}$$

does not vanish identically. Apparently $\psi_{\nu_0} = F(A_\nu)$, $\psi_{\nu n} = F_n$. Every $\psi_{\nu k}$ is holomorphic, so they have only isolated zeros.

Step 2. The hypothesis of the theorem implies that outside of a compact set D in S, f does not intersect any of the $A_1, \ldots, A_q$; therefore $F(A_\nu) \neq 0$. Let

$$S' = \{p \in S \backslash D \colon \psi_{\nu k} \neq 0 \text{ for any } (\nu, k)\}.$$

On S' we define a new metric

$$d\tilde{s}^2 = \left| \frac{\prod_{\nu=1}^q F(A_\nu)}{|F_n|^{1+\varepsilon} \prod_{\nu, k} |\psi_{\nu k}|^{\varepsilon/q}} \right|^{2p^*} |dz|^2$$

where

$$p^* = \frac{1}{(q-(n+1)-P_{n+1}\varepsilon) - (P_n + Q_n \varepsilon)},$$

$$\frac{q-(n+1)-P_n}{P_{n+1}+Q_n} > \varepsilon > \frac{q-(n+1)-Q_n}{P_{n+1}+Q_n+1/q};$$

the last inequality is equivalent to $\varepsilon p^*/q > 1$.

Here the definition of $d\tilde{s}^2$ would be valid if S' has a global coordinate z. Take a hyperplane A (out of $A_1, \ldots, A_q$). Then on S', f does not intersect A, namely

$$a_0 \frac{\partial x_1}{\partial z} + \cdots + a_n \frac{\partial x_n}{\partial z} \neq 0;$$

this means that if $\xi = a_0 x_1(z) + \cdots + a_m x_m(z)$ is a global coordinate on S', call it z, then $d\tilde{s}^2$ is well defined.

Step 3. Since $F(A_\nu)$, F_n and $\psi_{\nu k}$ are all holomorphic, the metric $d\tilde{s}^2$ is flat, and it can be smoothly extended over D. We thus obtain a metric, still call it $d\tilde{s}^2$, on

$$S'' = S' \cup D$$

that is flat outside the compact set D. The key to our proof is showing that S'' is complete in that metric.

Step 4. We proceed by contradiction. If S'' is not complete, then there is a divergent curve $\gamma(t)$ on S'' with finite length. By removing an initial segment, if necessary, we may assume that there is a positive distance d between the curve γ and the compact set D. Thus $\gamma\colon [0, 1) \to S'$, and since γ is divergent on S'', with finite length, it follows that from the point of view of S, either $\gamma(t)$ tends to a point z_0 where

$$|F_n|^{1+\varepsilon} \prod_{\nu, k} |\psi_{\nu k}|^{\varepsilon/q} = 0,$$

or else $\gamma(t)$ tends to the boundary of S as $t \to 1$. But the former case cannot occur, because if

$$|F_n(z_0)|^{1+\varepsilon} \prod_{\nu, k} |\psi_{\nu k}(z_0)|^{\varepsilon/q} = 0,$$

then by the fact that $\varepsilon p^*/q > 1$ (here q is the number of hyperplanes) we have

$$|d\tilde{s}| \sim \frac{c}{|z - z_0|^{\delta_0}} dz$$

around z_0 where $c > 0$, $\delta_0 > 1$. Thus

$$\int_0^1 d\tilde{s} = \infty,$$

contradicting the finite length of γ.

Step 5. We conclude that $\gamma(t)$ must tend to the boundary of S when $t \to 1$. Choose t_0 such that

$$\int_{t_0}^1 d\tilde{s} < \frac{d}{3};$$

that is, the length of $\gamma([t_0, 1))$ is less than $d/3$. Consider a small disk Δ with center $\gamma(t_0)$. Since $d\tilde{s}^2$ is flat, Δ is isometric to an ordinary disk in the plane. Let G be an isometry of $|w| < \eta$ onto Δ with $G(0) = \gamma(t_0)$. Extend G, as a local isometry into S', to the largest disk possible, say $|w| < R$. (Note that G may be viewed simply as the exponential map to S'' at $\gamma(t_0)$.) In view of $\int_{t_0}^1 d\tilde{s} < \frac{d}{3}$, and the fact that γ is a divergent curve on S, we have $R \leq d/3$. Hence the image under G must be bounded away from D by a distance of at least $2d/3$. Thus, the reason that the map G cannot be extended to a larger disk must be that the image goes to the boundary of S''. Since the zeros of $|F_n|^{1+\varepsilon} \prod_{\nu, k} |\psi_{\nu k}|^{\varepsilon/q}$ have been shown to be infinitely far away in the metric, the image must actually go to the boundary of S. More specifically, there must be a point w_0 with $|w_0| = r$, such that the image under G of the line segment from 0 to w_0 is a divergent curve Γ on S. Our goal is to show that Γ has finite length in the *original* metric ds^2 on S, contradicting the completeness of the original surface.

Step 6. We know that

$$|dw| = |d\tilde{s}| = \left| \frac{\prod_{\nu=1}^{q} F(A_\nu)}{|F_n|^{1+\varepsilon} \prod_{\nu,k} |\psi_{\nu k}|^{\varepsilon/q}} \right|^{p^*} |dz|.$$

Instead of z, we change to the coordinate w for the right-hand side of the above expression. Precisely speaking, we let

$\overline{F}(w) = (\overline{f}_0(w), \dots, \overline{f}_n(w)) = (f_0(z(w)), \dots, f_n(z(w))) = F(z(w))$,

and let $\overline{\psi}_{\nu k}(w)$ be defined from $\overline{F}(z)$ in the same way the $\psi_{\nu k}$ was defined from $F(z)$. Then a little computation shows that

$$\left|\frac{dw}{dz}\right| = \left| \frac{\prod_{\nu=1}^{q} \overline{F}(A_\nu)}{|\overline{f}_n(\frac{dw}{dz})^{P_N}|^{1+\varepsilon} \prod_{\nu,k} |\overline{\psi}_{\nu k}(\frac{dw}{dz})^{P_k}|^{\varepsilon/q}} \right|^{P^*}$$

$$= \left| \frac{\prod_{\nu=1}^{q} \overline{F}(A_\nu)}{|\overline{F}_n|^{1+\varepsilon} \prod_{\nu,k} |\overline{\psi}_{\nu k}|^{\varepsilon/q}} \right|^{p^*} \frac{1}{|\frac{dw}{dz}|^{p^*(P_n+\varepsilon Q_n)}},$$

$$\left|\frac{dw}{dz}\right|^{1+p^*(P_n+\varepsilon Q_n)} = \left| \frac{\prod_{\nu=1}^{q} \overline{F}(A_\nu)}{|\overline{F}_n|^{1+\varepsilon} \prod_{\nu,k} |\overline{\psi}_{\nu k}|^{\varepsilon/q}} \right|^{p^*}$$

by

$$p^* = \frac{1}{(q-(n+1)-P_{n+1}\varepsilon) - (P_n + Q_n\varepsilon)};$$

we have

$$\left|\frac{dw}{dz}\right| = \left| \frac{\prod_{\nu=1}^{q} \overline{F}(A_\nu)}{|\overline{F}_{n+1}|^{1+\varepsilon} \prod_{\nu,k} |\overline{\psi}_{\nu k}|^{\varepsilon/q}} \right|^{1/(q-(n+1)-P_{n+1}\varepsilon)}.$$

Step 7. We now denote by C the line segment from 0 to w_0, and by Γ, the image of C on S. Then for the length L of Γ, we have

$$\begin{aligned} L &= 2\int_C |F(z(w))|\,|dz(w)| \\ &= 2\int_C |\overline{F}(w)| \left|\frac{dz}{dw}\right| |dw| \\ &= 2\int \left| \frac{|\overline{F}|^{q-(n+1)-P_{n+1}\varepsilon} |\overline{F}_{n+1}|^{1+\varepsilon} \prod_{\nu,k} |\overline{\psi}_{\nu k}|^{\varepsilon/q}}{\prod_{\nu=1}^{q} |\overline{F}(A_\nu)|} \right|^{1/(q-(n+1)-\varepsilon P_{n+1})}. \end{aligned}$$

By the definition of $\overline{\Psi}_{\nu k}$, $|\overline{\Psi}_{\nu k}| \leq |\overline{F}_k(A_\nu)|$, and using the proposition of the previous section, the Fujimoto inequality, we have

$$L \leq 2C \int \left(\frac{2R}{R^2 - |w|^2}\right)^{(P_n+Q_n\varepsilon)/(q-(n+1)-P_{n+1}\varepsilon)} |dw|.$$

Because $0 < (P_n + Q_n\varepsilon)/(q - (n+1) - Q_{n+1}\varepsilon) < 1$, L is finite.

Step 8. To sum up, we have shown that if the surface S'' were not complete, then we could find a divergent curve on S with finite length in the original metric, so that S would not be complete. We therefore conclude that S'' is complete. Since the metric on S'' is flat outside of a compact set, we are in a familiar situation (see **[11]** p. 3564, or Osserman, *A survey...*, p. 81). By a theorem of Huber [7], the fact that S'' has finite total curvature implies that S'' is finitely connected. We conclude first that $|F_n|^{1+\varepsilon} \prod_{\nu,k} |\psi_{\nu k}|^{\varepsilon/q}$ can have only a finite number of zeros, and second, that the original surface S is finitely connected. Further, by [**10**, Theorem 2.1] (or the argument in [**11**, pp. 354]) each annular end of S'', hence of S, is conformally equivalent to a punctured disk. Thus, the Riemann surface M on which S is based must be conformally equivalent to a compact Riemann surface $\overline{M}$ with a finite number of points removed. In a neighborhood of each of those points the Gauss map f does not intersect $q \geq n(n-1)/2 + 1 \geq n + 2$ hyperplanes. By a generalized Picard theorem (see [2, p. 136]), the Gauss map f can be extended to a holomorphic map from $\overline{M}$ to $P^n(C)$. If the homology class represented by the image of $f\colon \overline{M} \to P^n(C)$ is m times the fundamental homology class of $P^n(C)$, then we have

$$\iint K\, dA = -2\pi m$$

as the total curvature of S. This proves the theorem.

Acknowledgment. This paper is the first part of my Ph.D. thesis at Stanford University written under the guidance of Professor R. Osserman. He helped me most generously with his insight into our topics. His exceptional kindness towards a student has encouraged me to complete my work.

REFERENCES

[1] S.-S. Chern and R. Osserman, *Complete minimal surfaces in euclidean n-space*, J. Analyse Math., **19** (1967), 15–34.
[2] M. J. Cowen and P. Griffiths, *Holomorphic curves and metrics of negative curvature*, J. Analyse Math., **29** (1976), 13–153.
[3] H. Fujimoto, *On the Gauss map of a complete minimal surface in* R^m, J. Math. Soc. Japan, **35** (1983), 279–288.
[4] ——, *On the number of exceptional values of the Gauss map of minimal surfaces*, J. Math. Soc. Japan, **40** (1988), 235–247.
[5] ——, *Modified defect relations for the Gauss map of minimal surfaces* II, J. Differential Geom., **31** (1990), 365–385.
[6] D. Hoffman and R. Osserman, *The geometry of the generalized Gauss map*, Mem. Amer. Math. Soc., No. 236, November 1980.
[7] A. Huber, *On subharmonic functions and differential geometry in large*, Comment. Math. Helv., **32** (1961), 13–72.
[8] X. Mo and R. Osserman, *On the Gauss map and total curvature of complete minimal surfaces and an extension of Fujimoto's theorem*, J. Differential Geom., **31** (1990), 343–355.
[9] R. Osserman, *Proof of a conjecture of Nirenberg*, Comm. Pure Appl. Math., **12** (1959), 229–232.
[10] ——, *On complete minimal surfaces*, Arch. Rational Mech. Anal., **13** (1963), 392–404.
[11] ——, *Global properties of minimal surfaces in* E^3 *and* E^n, Ann. of Math., **80** (1964), 340–364.
[12] ——, *A Survey of Minimal Surfaces*, 2nd edition, Dover, New York, 1986.
[13] F. Xavier, *The Gauss map of a complete non-flat minimal surface cannot omit 7 points on the sphere*, Ann. of Math., **113** (1981), 211–214.

Received December 3, 1990.

WASHINGTON UNIVERSITY
ST. LOUIS, MO 63139

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

ON COMPLETE RIEMANNIAN MANIFOLDS WITH COLLAPSED ENDS

ZHONGMIN SHEN

We show that if a complete open manifold with bounded curvature and sufficiently small ends, then each end is an infranilend. Conversely, an open manifold with finitely many infranilends admits a complete metric with bounded curvature and arbitrarily small ends.

1. Introduction. It is well known that if a complete open manifold M of finite volume has bounded negative sectional curvature, i.e.

$$-\Lambda_2 \leq \sec(M) \leq -\Lambda_1,$$

where Λ_i are positive constants, then M has finite topological type (see **[G1]**). In particular, M has finitely many ends. Moreover, each end collapses, i.e., for any end E and any point $p \in M$,

$$\lim_{r\to\infty} \operatorname{diam}(E \cap S(p, r)) = 0,$$

where $S(p, r) = \{x \in M; d(p, x) = r\}$ denotes the geodesic sphere of radius r around p. Further, each end is topologically of the form $N \times (0, \infty)$ for some infranilmanifold N. See **[E]** and **[Sc]** for details. See also **[K]** in the case $\Lambda_2/\Lambda_1 < 4$.

An open manifold M is said to have N ends, if there is a compact subset K such that for any compact subset $K \subset K' \subset M$, $M\backslash K'$ contains exactly N unbounded components. Simply we call any such component an end of M.

In **[Sh]** we studied complete open Riemannian manifolds M with sectional curvature bounded from below and small ends. In order to state the result, we need to introduce some notations. For $r > 0$, the connected components, Σ, of $\partial(M\backslash\overline{B(p, r)})$, are called the boundary components, where $B(p, r)$ denotes the open geodesic ball of radius r around p. Following **[C]** (compare **[AG]**), we define the essential diameter $\mathscr{D}(p, r)$ at distance r from p by

$$(p, r) = \sup_{\Sigma} \operatorname{diam}(\Sigma),$$

where the supremum is taken over all boundary components Σ of $M\backslash\overline{B(p, r)}$ with $\Sigma \cap R(p, r) \neq \varnothing$, where $R(p, r) = \{\gamma(r); \gamma$ *is a ray from* $p\} \subset S(p, r)$. Notice that in the definition of $\mathscr{D}(p, r)$ we do not assume that M has finitely many ends. In **[Sh]** we prove the following

THEOREM 1 (**[Sh]**). *Let* M *be complete with* $\sec(M) \geq -1$. *Suppose that*

$$\overline{\lim_{r\to\infty}}\, \mathscr{D}(p, r) < \ln 2. \tag{1}$$

Then M *is homeomorphic to the interior of a compact manifold with boundary. In particular,* M *has finitely many ends.*

We do not know what is the best constant on the right side of (1). This theorem is proved by applying the Morse theory to the distance function $d_p(x) = d(p, x)$. Of course, d_p is not of C^1 in general. But we still have a notion of critical points of d_p. We say a point q is a critical point in the sense of Grove-Shiohama **[GS]** if for any vector $v \in T_qM$, there is a minimal geodesic σ from q to p, making an angle $\angle(v, \dot{\sigma}(0)) \leq \frac{\pi}{2}$. For the further discussion in §2, we would like to outline the proof of Theorem 1 here (see **[Sh]** and **[G]** for details). Suppose M is as in Theorem 1. Let e_p denote the excess function on M, which is defined by $e_p(x) = \lim_{r\to+\infty} d_p(x) + d(x, S(p, r)) - r$. It follows from Toponogov's comparison theorem that if $q \in S(p, r)$ is a critical point of d_p, then

$$e_p(q) \geq \ln \frac{e^r}{\cosh r}.$$

Let Σ be any boundary component of $M\backslash\overline{B(p, r)}$ with $\Sigma \cap R(p, r) \neq \varnothing$. An elementary argument shows that for any $x \in \Sigma$,

$$e_p(x) \leq \operatorname{diam}(\Sigma) \leq \mathscr{D}(p, r).$$

Notice that $\ln(e^r/\cosh r) \to \ln 2$ as $r \to +\infty$. Thus there is a large R_0 such that if $r \geq R_0$, Σ contains no critical points of d_p. Let E be an unbounded component of $M\backslash\overline{B(p, R_0)}$, and let γ be a ray from p such that $\gamma(R_0, \infty) \subset E$. Denote by Σ_r the boundary component of $M\backslash\overline{B(p, r)}$ with $\gamma(r) \in \Sigma_r$, $r \geq R_0$. Since all Σ_r, $r \geq R_0$, contain no critical points of d_p, one can show that all Σ_r are homeomorphic, and E is homeomorphic to $\Sigma_{R_0} \times (R_0, \infty)$. Notice that $M\backslash\overline{B(p, R_0)}$ has finitely many unbounded components. Thus M has finite topological type in the sense of Theorem 1. In this case, it is also easy to see

$$\mathscr{D}(p, r) = \max_E \operatorname{diam}(E \cap S(p, r)), \qquad r \geq R_0, \tag{2}$$

where the maximum is taken over all unbounded components of $M\backslash\overline{B(p, R_0)}$. Readers can also refer to [C] for further discussion.

In this paper we will study the structure of small ends for complete open manifolds with bounded sectional curvature. Riemannian manifolds under consideration may have infinite volume and are not required to be negatively curved.

THEOREM 2. *Let M be a complete open n-manifold with $p \in M$ fixed. Suppose that $|\sec(M)| \leq 1$. There is a small $\varepsilon(n) > 0$, such that if*

$$\overline{\lim_{r\to\infty}} \mathscr{D}(p, r) < \varepsilon(n),$$

then there is a compact subset $K \subset M$, such that each unbounded component of $M\backslash K$ is diffeomorphic to $N \times (0, \infty)$ for some infranilmanifold N.

A manifold N is called an infranilmanifold if it is diffeomorphic to a compact space G/Γ, where G is a nilpotent Lie group and Γ is a discrete group of affine transformations of G satisfying $[\Gamma : G \cap \Gamma] < \infty$. Here we have put the left invariant connection D on G for which left invariant vector fields are parallel and G is regarded as a group of affine transformations on G by left translations (see [**R**]). An end of an open manifold, which is diffeomorphic to $N \times (r, \infty)$ for some infranilmanifold, is called an infranilend.

Suppose M is an open n-manifold with finitely many infranilends. A natural question is whether or not M admits a complete metric g such that $|\sec(M, g)| \leq 1$ and each one $E \cong N \times (0, \infty)$ collapses, i.e.

$$\overline{\lim_{r\to\infty}} \operatorname{diam}(N \times \{r\}) = 0. \tag{3}$$

The answer is affirmative.

THEOREM 3. *If M is an open manifold with finitely many infranilends, then M admits a complete Riemannian metric g satisfying $|\sec(M)| \leq 1$ and* (3) *for each end $E \cong N \times (0, \infty)$.*

The construction of the metric is not trivial, and is given in §3. One needs to know precisely the structure of an infranilmanifold. Our construction is inspired by [**W1**].

2. Proof of Theorem 2. The proof of Theorem 2 is simple. We are going to apply a theorem of Fukaya [**F1**] and [**F2**] to our case.

For two metric spaces X, Y and $\varepsilon > 0$, a map $h: X \to Y$ is said to be an ε-Hausdorff approximation if for any points $x_1, x_2 \in X$,

$$|d(x_1, x_2) - d(h(x_1), h(x_2))| < \varepsilon$$

and the image of h, $h(X) \subset Y$, is ε-dense in Y. Let M_1^n, M_2^m, $m \leq n$, be complete with $|\sec(M_1)| \leq 1$, $|\sec(M_2)| \leq 1$ and $\operatorname{inj}(M_2) \geq i_0$. Fukaya's theorem says that there is a small number $\varepsilon = \varepsilon(n, i_0)$ so that if $h: M_1 \to M_2$ is an ε-Hausdorff approximation, then there is a fibration $f: M_1 \to M_2$ such that $f^{-1}(y)$, $y \in M_2$, is an infranilmanifold. Fukaya's theorem is an important generalization of Gromov's result **[G2]** and **[R]** on almost flat manifolds.

Now we let M be a complete open n-manifold with $p \in M$ fixed. First of all we only assume that

$$\mathscr{D}(p, r) < \delta, \qquad r \geq R_0,$$

for some $\delta < \ln 2$. As mentioned in §1, taking a larger number R_0 if necessary, one can show that any unbounded component, E, of $M \backslash \overline{B(p, R_0)}$ is homeomorphic to $N \times (0, \infty)$, where N is a compact manifold without boundary (see **[Sh]** for details). Furthermore, by (2)

$$\mathscr{D}(p, r) = \max_E \operatorname{diam}(E \cap S(p, r)) < \delta, \qquad r \geq R_0, \tag{4}$$

where the maximum is taken over all unbounded components, E, of $M \backslash \overline{B(p, R_0)}$.

Fix an unbounded component E of $M \backslash \overline{B(p, R_0)}$. Define

$$\pi: E \to (R_0, \infty),$$
$$x \in E \cap S(p, r) \to r \in (R_0, \infty).$$

It is easy to see by (4) that for any $x_1, x_2 \in E$,

$$|d(x_1, x_2) - |\pi(x_1) - \pi(x_2)|| < 2\delta. \tag{5}$$

Thus π is an 2δ-Hausdorff approximation.

Although Fukaya's theorem is stated for complete manifolds, his argument can be carried over to our case. Let $E_1 = \{x \in E, d(x, \partial E) > 1\}$. By **[F1]** and **[F2]** there is a small number $\varepsilon(n) > 0$ such that if (5) holds for some $\delta < \varepsilon(n)$, then there is an open neighborhood U, $E_1 \subset U \subset E$, and a fibration $f: U \to (R_1, \infty)$, $R_1 \geq R_0$, with fibres $f^{-1}(r)$ diffeomorphic to an infranilmanifold N by ϕ_r. One then defines a diffeomorphism $\phi: U \to N \times (R_1, \infty)$ by $\phi(x) = (\phi_r(x), r)$ for $x \in f^{-1}(r)$.

One can also follow **[CFG, §2]** to construct the above fibration f by the center of mass techniques. This completes the proof.

Let M be a complete n-manifold with N ends. Let $E_1, \dots, E_N$ denote all unbounded connected components of $M\backslash K$, where K is some compact subset. Let $p \in M$. Set

$$\operatorname{diam}(p, M) = \overline{\lim_{r\to\infty}} \max_{1\le i\le N} \operatorname{diam}(E_i \cap S(p, r)).$$

Clearly, $\operatorname{diam}(p, M)$ is independent of K. The following corollary is a direct consequence of the above argument.

COROLLARY 1. *Let M be complete n-manifold with finitely many ends. Suppose that $|\sec(M)| \le 1$. There is a small $\varepsilon(n)$, if for some point $p \in M$,*

$$\operatorname{diam}(p, M) < \varepsilon(n),$$

then there is a compact subset K such that each unbounded component E of $M\backslash K$ is diffeomorphic to $N \times (0, \infty)$, where N is an infranil-manifold.

3. Construction of the metrics. Let M be an open manifold with finitely many infranilends, say, $E_1, \dots, E_N$, which are unbounded connected components of $M\backslash K$ for some compact subset K. By definition, each end E_i is diffeomorphic to $N_i \times (0, \infty)$ for some infranilmanifold N_i. Fix an end $E \cong N \times (0, \infty)$. In order to construct a complete metric on M satisfying (3) in Theorem 3, it suffices to construct a metric g on $E \cong N \times (0, \infty)$ such that $|\sec(E, g)| \le 1$ and (3) holds.

Recall that N is an infranilmanifold if $N = G/\Gamma$, where G is a nilpotent Lie group and Γ is a discrete group of affine transformations of G satisfying $[\Gamma, G \cap \Gamma] < \infty$. Here we have put the left invariant connection D on G for which left invariant vector fields are parallel, and G is regarded as a group of affine transformations on G by left translations. Let L denote the Lie algebra of G, the space of left-invariant vector fields on G. One has the following stratification

$$L = L_0 \supset L_1 \supset \cdots \supset L_k \supset L_{k+1} = 0 \tag{6}$$

where $L_{i+1} = [L, L_i]$. Notice that $H := \Gamma/(G \cap \Gamma)$ acts on L and preserves the stratification (6). One can choose an H-invariant inner product $\langle\ ,\ \rangle_0$ on L. Let $F_i = \{X \in L_i,\ \langle X, Y\rangle_0 = 0, Y \in L_{i+1}\}$. Then $L = F_0 \oplus \cdots \oplus F_k$. One can define an H-invariant inner product $\langle\ ,\ \rangle_r$ on L by

$$\langle X, Y\rangle_r = h_i(r)^2 \langle X, Y\rangle_0, \qquad X, Y \in F_i,$$

and $\langle X, Y\rangle_r = 0$ if $X \in F_i, Y \in F_j, i \neq j$, where h_i are some positive functions which are to be determined later. $\langle\ ,\ \rangle_r$ defines an H-invariant metric on G, which is then Γ-invariant. Hence it induces a Riemannian metric g_r on G/Γ. There is a bound C depending only on G and $\langle\ ,\ \rangle_0$ such that

$$\|[X, Y]\|_0 \leq C\|X\|_0\|Y\|_0$$

for all $X, Y \in L$. For a left-invariant vector $X \in L$, we denote it by $Z = \sum_i Z_i$, where Z_i denotes the component of Z in F_i. Then for any $X, Y \in L$, one has

$$\|[X, Y]_i\|_0 \leq \|[X, Y]\|_0 \leq C\|X\|_0\|Y\|_0.$$

We choose h_i in such a way that the following inequalities hold:

$$\sum_{s=i+1}^{k} h_s(r) \leq h_i(r)^2 \leq h_{i-1}(r)^2. \tag{7}$$

Then for $X = \sum_{i=0}^k X_i$ and $Y = \sum_{i=0}^k Y_i$,

$$\begin{aligned} \|[X, Y]\|_r &= \left\|\sum_{ij}[X_i, Y_j]\right\|_r \leq \sum_{ij}\|[X_i, Y_j]\|_r \\ &\leq \sum_{ij}\sum_{s>\max(ij)} h_s(r)\|[X_i.Y_j]_s\|_0 \\ &\leq C\sum_{ij}\sum_{s>\max(ij)} h_s(r)\|X_i\|_0\|Y_j\|_0 \\ &\leq C\sum_{ij} h_i(r)h_j(r)\|X_i\|_0\|Y_j\|_0 \leq (k+1)C\|X\|_r\|Y\|_r. \end{aligned} \tag{8}$$

Let $\widehat{\nabla}$ denote the Levi-Civita connection of g_r, and $\widehat{R}$ its curvature tensor. It follows from (8) and formulas in [CE, Proposition 3.18], that

$$\|\widehat{\nabla}_X Y\|_r \leq \frac{3}{2}(k+1)C, \tag{9}$$

$$\|\widehat{R}(X, Y)Z\|_r \leq 6(k+1)^2C^2 \tag{10}$$

for any orthonormal left-invariant vector fields X, Y and Z in L with respect to $\langle\ ,\ \rangle_r$.

We define a warped product metric g on $E = N \times (0, \infty)$ by

$$g = g_r \oplus dr^2.$$

Let $H = \frac{\partial}{\partial r}$ and X, Y, Z, etc. be left-invariant vector fields on G. Let ∇ denote the Levi-Civita connection of g. Then it is easy to get the following

$$\begin{aligned}
&\nabla_H H = 0, \\
&\nabla_H X = \nabla_X H = \frac{h_i'(r)}{h_i(r)} X, \\
&\nabla_X Y = \widehat{\nabla}_X Y - \frac{h_i'(r)}{h_i(r)} \langle X, Y \rangle_r H
\end{aligned}$$

for any left-invariant vector fields $X \in F_i$ and $Y \in F_j$.

We now take $h_i(r) = e^{-\alpha_i(r+1)}$, $\alpha_i = 2^i - 1$. Then (7) holds. It is easy to see that g has bounded curvature on $E \cong N \times (0, \infty)$. More precisely, if

$$\left| \frac{h_i'(r)}{h_i(r)} \right| \le \alpha_k, \quad \left| \frac{h_i''(r)}{h_i(r)} \right| \le \alpha_k^2,$$

then by (9) (10) there is constant $K(k, C)$ such that

$$\begin{aligned}
&\|R(X, H)H\| \le K(k, C), \\
&\|R(X, Y)H\| \le K(k, C), \\
&\|R(X, H)Y\| \le K(k, C), \\
&\|R(X, Y)Z\| \le K(k, C)
\end{aligned}$$

for any orthonormal left-invariant vector fields $X \in F_i$, $Y \in F_j$ and $Z \in F_m$.

In particular, one has (compare [W2])

$$\begin{aligned}
g(R(X, H)H, X) &= -\frac{h_i''(r)}{h_i(r)}, \\
g(R(X, Y)Y, X) &= g_r(\widehat{R}(X, Y)Y, X) - \frac{h_i'(r)h_j'(r)}{h_i(r)h_j(r)}
\end{aligned}$$

for orthonormal left-invariant vector fields $X \in F_i$, $Y \in F_j$.

Observe that $g_r \le e^{-2(r+1)} g_0$. One concludes that

$$\operatorname{diam}(N \times \{r\}, g_r) \le e^{-r-1} \cdot \operatorname{diam}(N, g_0) \to 0$$

as $r \to \infty$. This completes the proof.

REMARK. Taking $h_i(r) = e^{-\alpha_i(r+1)}$, $\alpha_i = \delta \cdot 2^i - 1$, one can make the curvature negatively pinched on each end, provided that δ is sufficiently large.

References

[AG] U. Abresch and D. Gromoll, *On complete manifolds with nonnegative Ricci curvature*, J. Amer. Math. Soc., **3** (1990), 355–374.

[C] J. Cheeger, *Critical points of distance functions and applications to geometry*, preprint (1990).

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North-Holland, Amsterdam, 1975.

[CFG] J. Cheeger, K. Fukaya and M. Gromov, *Nilpotent structure and invariant metrics on collapsed manifolds*, preprint.

[E] P. Eberlein, *Lattices in spaces of nonpositive curvature*, Ann. of Math., **111** (1980), 435–476.

[F1] K. Fukaya, *Collapsing Riemannian manifolds to ones of lower dimension*, J. Differential Geom., **25** (1987), 137–156.

[F2] ——, *Collapsing Riemannian manifolds to ones of lower dimension*, II, J. Math. Soc. Japan, **41** (1989), 333–356.

[G1] M. Gromov, *Manifolds of negative curvature*, J. Differential Geom., **13** (1978), 223–230.

[G2] ——, *Almost flat manifolds*, J. Differential Geom., **13** (1978), 231–242.

[G] K. Grove, *Critical point theory for distance functions*, preprint.

[GS] K. Grove and K. Shiohama, *A generalized sphere theorem*, Ann. of Math., **106** (1977), 201–211.

[K] M. Kanai, *A pinching theorem for cusps of negatively curved manifolds with finite volume*, preprint (1990).

[R] E. Ruh, *Almost flat manifolds*, J. Differential Geom., **17** (1982), 1–14.

[Sc] V. Schroeder, *Finite volume and fundamental group on manifolds of negative curvature*, J. Differential Geom., **20** (1984), 175–183.

[Sh] Z. Shen, *On complete manifolds of nonnegative kth-Ricci curvature*, Trans. Amer. Math. Soc., **338** (1993), 289–310.

[W1] G. Wei, *Examples of complete manifolds of positive Ricci curvature with nilpotent isometry groups*, Bull. Amer. Math. Soc., **19** (1988), 311–313.

[W2] ——, *Aspects of positively Ricci curved spaces: New examples and the fundamental group*, Ph.D. thesis at SUNY at Stony Brook, 1989.

Received April 23, 1992.

The University of Michigan
Ann Arbor, MI 48109
E-mail address: zhongmin@math.lsa.umich.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

CURVATURE CHARACTERIZATION OF CERTAIN BOUNDED DOMAINS OF HOLOMORPHY

FANGYANG ZHENG

In this note, we study the relation between the existence of a negatively curved complete hermitian metric on a complex manifold M and the product structure of (or contained in) M. We introduce the concept of geometric ranks and give a curvature characterization of the rank one manifolds, which generalizes the previous results of P. Yang and N. Mok (see below). In the proof, we used the old techniques of Yau's Schwartz lemma and Cheng-Yau's result on the existence of Kähler-Einstein metrics.

1. Introduction and statement of results. Let $M = M_1 \times M_2$ be the product of two complex manifolds. Then it is generally believed that M does not admit any complete Kähler metric with bisectional curvature bounded between two negative constants. When M is compact, this is certainly true since the cotangent bundle T_M^* is not ample. In the noncompact case, the first result toward this direction was obtained by Paul Yang in 1976:

THEOREM ([Y]). *For any $n \geq 2$, there exists no complete Kähler metric on the polydisc C^n with bisectional curvature bounded between two negative constants.*

In [M], as an application of his metric rigidity theory, Mok generalized the above to give an interesting curvature characterization of the rank one bounded symmetric domains:

THEOREM ([M]). *If Ω is a bounded symmetric domain of rank ≥ 2, then there exists no complete hermitian metric on Ω with bounded torsion and with bisectional curvature bounded between two negative constants.*

Mok's proof is a constructive one. It used the existence of a uniform lattice Γ on Ω, as well as the integral formula on Ω/Γ discovered by Mok (cf. Proposition (3.2) in [M]). This proof is very interesting by itself. However, we noticed that Yang's approach can be used to give a more straightforward proof of Mok's result, and the conclusion holds for a larger class of manifolds (since one avoids the use of uniform lattice). Intuitively speaking, the reason for the non-existence of the above negatively curved metrics on D^n or Ω is the product structure on or nicely contained in the manifold (by the polydisc theorem, there is a totally geodesic proper embedding $D^r \to \Omega$, $r = \text{rank}(\Omega)$).

First let us fix some notations. From now on, we shall say that a hermitian manifold (M, h) is *negatively curved*, if it is complete, of bounded torsion, and with bisectional curvature bounded between two negative constants.

Now let Ω be a bounded domain of holomorphy in $\mathbf{C}^n$. By the results of Cheng-Yau **[C-Y]** and Mok-Yau **[M-Y]**, there exists a unique complete Kähler-Einstein metric on Ω with Ricci curvature -1. Denote it by g. Again let D be the unit disc in $\mathbf{C}$.

DEFINITION. Ω is said to be of *geometric rank* ≥ 2, if there is a complete Kähler manifold (M, g_0) with Ricci curvature bounded from below, and a holomorphic embedding $f: D \times M \to \Omega$ such that $f_t^*(g) \geq g_0$ for each $t \in D$, where $f_t = f(t, \cdot)$.

In other words, Ω is of geometric rank ≥ 2 if it contains a product manifold with bounded second fundamental forms. It is obvious that one can define the actual geometric rank of Ω; however, in this note we shall only be interested in the distinction between the rank one case and the higher rank cases.

For bounded symmetric domains, the polydisc theorem implies that the usual rank dominates the geometric rank.

In §2, we shall prove the following generalization to the above cited result of Mok:

THEOREM A. *Let Ω be a bounded domain of holomorphy. If it is of geometric rank ≥ 2, then it cannot be negatively curved.*

We shall also give partial answers in §3 to the question that product manifolds cannot be negatively curved:

THEOREM B. *Let $M = M_1 \times M_2$ be the product of two complex manifolds, with M_1 compact. Then there is no (not necessarily complete) hermitian metric on M with bisectional curvature ≤ -1.*

THEOREM C. *Let $M = M_1 \times M_2$ be the product of two complex manifolds. Suppose that both M_1 and M_2 admit complete Kähler metrics with Ricci curvature bounded between two negative constants. Then M cannot be negatively curved.*

COROLLARY. *If both M_1 and M_2 are relatively compact open subsets of some Stein manifolds, then $M = M_1 \times M_2$ cannot be negatively curved.*

In particular, the product of two bounded domains cannot be negatively curved.

2. Bounded domains of holomorphy. First let us recall the generalized Schwarz lemma of Yau [Y1] and Chen-Yang [C-Y1].

PROPOSITION 1 ([Y1]). *Let (M, g) be a complete Kähler manifold with Ricci curvature $\geq -K_1$, and (N, h) be a hermitian manifold with bounded torsion and with Ricci curvature $\leq -K_2 < 0$. If* $\dim(M) = \dim(N)$, *and $f: M \to N$ is holomorphic, then $f^* dv_h \leq (K_1/K_2)\, dv_g$.*

PROPOSITION 2 ([C-Y1]). *Suppose (M, g) is a complete hermitian manifold with bounded torsion, and with second Ricci curvature $\geq -K_1$. Let (N, h) be a hermitian manifold with nonpositive bisectional curvature and with holomorphic sectional curvature $\leq -K_2 < 0$. Then for any holomorphic map $f: M \to N$, one has $f^*(h) \leq (K_1/K_2) g$.*

We shall also need the following generalized maximum principle of Yau:

PROPOSITION 3 ([Y2]). *If (M, g) is a complete Kähler manifold with Ricci curvature bounded from below, and φ is a C^2 function on M bounded from above. Then for any $\varepsilon > 0$, there exists $x \in M$ such that: $\varphi(x) > \sup \varphi(M) - \varepsilon$, $|\nabla \varphi(x)| < \varepsilon$, $\Delta \varphi(x) < \varepsilon$.*

Now we are ready to prove Theorem A. The idea comes from Yang's proof in [Y] and the basic tool is the Schwarz lemma.

Proof of Theorem A. Let Ω be a bounded domain of holomorphy, with geometric rank ≥ 2. Let g be the complete Kähler-Einstein metric on it. By definition, there is a complete Kähler manifold (M, g_0) with Ricci curvature bounded from below, and a holomorphic embedding $f: D \times M \to \Omega$ such that $f_t^* g \geq g_0$ for each $t \in D$.

Assume that Ω admits a negatively curved metric h. Applying Proposition 2 to the identity map id: $(\Omega, g) \to (\Omega, h)$, one gets $g \geq c' h$, while by Proposition 1 to id: $(\Omega, h) \to (\Omega, g)$ one gets $dv_g \leq c'' dv_h$, with c, c'' some positive constants. Therefore, g and h dominate each other. Hence $h \geq cg$ with $c > 0$.

Now let ρ be a nonnegative smooth function with compact support in D. For $z \in M$, define

$$\varphi(z) = \int_D \rho(t) \cdot f_z^*(\omega_h)$$

where ω_h is the Kähler form on h, and $f_z = f(\cdot\,, z)\colon D \to \Omega$. Then φ is a positive smooth function on M. It is also bounded from above, since by Schwarz lemma, for any $z \in M$, $f_z^*(\omega_h)$ is dominated by the Poincaré metric on D.

Let $z = (z_1, \dots, z_k)$ be a local holomorphic coordinate on M, $t = z_0$ a local coordinate on D, and $(t, z, z_{k+1}, \dots, z_{n-1})$ a coordinate on Ω. Let $-c' < 0$ be an upper bound for the bisectional curvatures of h. Then we have that for each $1 \leq i \leq k$:

$$h_{t\bar{t},i\bar{i}} \geq -R_{t\bar{t}i\bar{i}}(h) \geq c' h_{i\bar{i}} h_{t\bar{t}} \geq cc'(g_0)_{i\bar{i}} h_{t\bar{t}}.$$

Therefore,

$$\frac{\partial^2 \varphi}{\partial z_i \partial \overline{z}_i} = \int_D \rho(t) h_{t\bar{t},i\bar{i}} \frac{\sqrt{-1}}{2}\, dt \wedge d\bar{t} \geq cc'(g_0)_{i\bar{i}}\varphi\,;$$

hence $\Delta\varphi \geq cc'\varphi$, where the Laplacian is with respect to g_0. Let $u = \log\varphi$; then the inequality becomes

$$\Delta u + |\nabla u|^2 \geq cc' > 0.$$

Since u is also bounded from above, by Proposition 3, we get a contradiction. So we conclude that Ω cannot be negatively curved. □

REMARK. From the proof it is clear that the bounded domain (Ω, g) can be replaced by any complete hermitian manifold with bounded torsion and with Ricci curvature bounded between two negative constants (or, Ricci $\leq -c < 0$ and second Ricci $\geq -c'$), as long as we keep the same condition on the geometric rank. One may also replace the Kählerness of g_0 by *hermitian with bounded torsion*, since Proposition 3 (hence Proposition 1) remains valid under such a replacement; here Δ is the complex Laplacian.

3. Noncompact product manifolds. Let $M = M_1 \times M_2$ be the product of two complex manifolds. In this section we shall verify that M cannot be negatively curved under the additional assumptions. First let us quote the following result due to Cheng-Yau **[C-Y]** and Mok-Yau (cf. **[M-Y]**, (3.1)):

PROPOSITION 4 ([C-Y], [M-Y]). *Suppose X is a Stein manifold, $M \subseteq X$ is a relatively compact open subset which is also Stein. If there exists a hermitian metric h on M with holomorphic sectional curvature ≤ -1 and $h \geq g$ for some Kähler metric g on X. Then M admits a complete Kähler-Einstein metric with negative Ricci curvature.*

Proof of Theorem B. Assume the contrary, namely assume that there is a hermitian metric on M with bisectional curvature ≤ -1. Take a small disc $D \subseteq M_2$ and a cut off function ρ in D. Then there exists a positive constant c such that $h|_{M_1\times\{t\}} \geq c \cdot h|_{M_1\times\{0\}}$ for each $t \in \mathrm{Supp}(\rho)$. Since M_1 is compact, the proof of Theorem A gives a contradiction. □

Proof of Theorem C. Assume the contrary: there is a negatively curved metric h on M. For $i = 1, 2$, let g_i be the complete Kähler metric on M_i with Ricci curvature bounded between two negative constants, and $g = g_1 \times g_2$. Then by applying Propositions 1 and 2 to the identity map on M we get that h and g are dominated by each other. Hence for each $y \in M_2$, $h|_{M_1\times\{y\}} \geq c \cdot g|_{M_1\times\{y\}} = c \cdot g_1$, where $c > 0$ is a constant. Take any small disc $D \subseteq M_2$, and the proof of Theorem A goes through. □

REMARK. It is also clear that in Theorem C above one can lose the Kählerness assumption on g_i to the weaker *hermitian with bounded torsion*.

Proof of Corollary. Again assume the contrary: there is a negatively curved metric h on $M = M_1 \times M_2$. Then $h|_{M_1\times\{y\}}$ and $h|_{\{x\}\times M_2}$ give complete hermitian metrics with non-positive holomorphic sectional curvature on M_1 and M_2, respectively. By [G] or [S], we know that both M_1 and M_2 are holomorphically convex, hence Stein as they are contained in Stein manifolds. By Proposition 4, they admit complete Kähler-Einstein metrics with negative Ricci curvature. Hence Theorem C applies and one gets a contradiction. □

REMARK. In proving the non-existence of negatively curved metrics on a general product manifold, the main difficulty comes from the fact that on a submanifold with restricted metric, the curvature is not necessarily bounded from below even if the ambient manifold is so. Or equivalently the second fundamental form need not be bounded. While the above line of argument depends on the Schwarz lemma, or eventually the generalized maximum principle, which requires a lower bound on the Ricci curvature, it should be interesting to know whether or not the following holds:

Question. On the bidisc $D \times D$, is there any complete Kähler metric with bisectional curvature ≤ -1?

References

[C-Y] S. Y. Cheng and S. T. Yau, *On the existence of a complete Kähler metric on non-compact complex manifolds and the regularity of Fefferman's equation*, Comm. Pure Appl. Math., **33** (1980), 507–544.

[C-Y1] Z. H. Chen and H. C. Yang, *On the Schwarz lemma for complete hermitian manifolds*, Proceedings of the 1981 Hangzhou Conference (ed. by Kohn-Lu-Remmert-Siu), Birkhäuser, pp. 99–116.

[G] P. A. Griffiths, *Two theorems on extension of holomorphic mappings*, Invent. Math., **14** (1971), 27–62.

[M] N. Mok, *Uniqueness theorems of hermitian metrics of seminegative curvature on quotients of bounded symmetric domains*, Ann. of Math., **125** (1987), 105–152.

[M-Y] N. Mok and S. T. Yau, *Completeness of the Kähler-Einstein metric on bounded domains and the curvature characterization of domains of holomorphy by curvature conditions*, Proc. Sympos. Pure Math., vol. 39, Amer. Math. Soc., Providence, RI, 1983, pp. 41–59.

[S] B. Shiffman, *Extension of holomorphic maps into hermitian manifolds*, Math. Ann., **94** (1971), 249–258.

[Y] P. C. Yang, *On Kähler manifolds with negative holomorphic bisectional curvature*, Duke Math. J., **43** (1976), 871–874.

[Y1] S. T. Yau, *A general Schwarz lemma for Kähler manifolds*, Amer. J. Math., **100** (1978), 197–203.

[Y2] ——, *Harmonic functions on complete Riemannian manifolds*, Comm. Pure Appl. Math., **28** (1975), 201–228.

Received March 18, 1992. Research supported by NSF Grant 91-05185 and Duke University.

Duke University
Durham, NC 27706

PACIFIC JOURNAL OF MATHEMATICS
Vol. 163, No. 1, 1994

THE CLASSIFICATION OF COMPLETE LOCALLY CONFORMALLY FLAT MANIFOLDS OF NONNEGATIVE RICCI CURVATURE

SHUNHUI ZHU

The main purpose of this note is to give a classification of complete locally conformally flat manifolds of nonnegative Ricci curvature. Such classification for the compact case has been obtained by various authors in the past decade.

1. Introduction. Recall that an n-dimensional Riemannian manifold (M^n, g) is said to be locally conformally flat if it admits a coordinate covering $\{U_\alpha, \phi_\alpha\}$ such that the map $\phi_\alpha: (U_\alpha, g_\alpha) \to (S^n, g_0)$ is a conformal map, where g_0 is the standard metric on S^n. It follows from this definition that the Weyl tensor of g vanishes. In particular, the full curvature tensor of g can be recovered from the Ricci tensor of g (an alternating sum). Thus conditions on the Ricci tensor of such manifolds impose very strong restrictions on their metrics. In the first part of this note we confirm this by showing,

THEOREM 1. *If (M^n, g) is a complete locally conformally flat Riemannian manifold with $\mathrm{Ric}(g) \geq 0$, then the universal cover $\widetilde{M}$ of M with the pulled-back metric is either conformally equivalent to S^n, R^n or is isometric to $R \times S^{n-1}$. If M itself is compact, then $\widetilde{M}$ is either conformally equivalent to S^n or isometric to R^n, $R \times S^{n-1}$, where S^n and S^{n-1} are spheres of constant curvature.*

The second part of Theorem 1 was obtained by various authors as consequences of investigating more general classes of manifolds, see the work of Schoen and Yau ([**SY**]) for references. An elementary proof for this case was also given recently by Noronha ([**No**]).

We remark that although the validity of Theorem 1 is not surprising, many similar problems in Riemannian geometry still remain open in the noncompact case, while the compact case has long been solved. The difficulty usually lies in the lack of analytic techniques for noncompact manifolds. The analysis in our case does carry through ([**SY**]) essentially because of the developing map as outlined below.

Our argument for the complete case uses heavily the results of Schoen and Yau. Let us outline the idea here. In **[SY]**, Schoen and Yau proved that the developing map for locally conformally flat manifolds with nonnegative scalar curvature is injective, thus exhibiting them as quotients of domains in the sphere by Kleinian groups. Just as in the study of Kleinian groups in the works of Patterson (**[Pa]**) and Sullivan (**[Su]**), Schoen-Yau studied the Hausdorff dimension of the complement of the image of $\widetilde{M}$ under the developing map, and proved that it can be controlled by properties of Green's functions of the conformal Laplacian on $\widetilde{M}$. Our observation is that under the condition of $\mathrm{Ric} \geq 0$, Green's function gives a much stronger control on the Hausdorff dimension than in the case of nonnegative scalar curvature. In fact, we will show that the Hausdorff dimension is zero. Theorem 1 is a consequence of this fact and the splitting theorem of Cheeger-Gromoll (**[CG]**).

In the second part, we study locally conformally flat manifolds under the more general condition of $\mathrm{Ric} \geq -\Lambda^2$, and prove,

THEOREM 2. *If (M^n, g) is a compact locally conformally flat manifold with*

$$\mathrm{Ric} \geq -\Lambda^2, \quad \mathrm{diam}(M) \leq D,$$

then $b_i(M, R) \leq C(n, \Lambda D)$ for any i, where $C(n, \Lambda D)$ is a constant depending only on n and ΛD.

Theorem 2 is a consequence of a general result about elliptic inequalities based on Moser iteration and P. Li's lemma. This line of thought was initiated by P. Li and later developed by Gallot, Besson and Berard, among others (see **[Be]**). Theorem 2 is basically known without being explicitly stated; we find it illuminating to put it here since it gives a parallel to Gromov's famous estimate for Betti numbers for Riemannian manifolds with lower sectional curvature and diameter bounds. And together with a corollary of Theorem 1, it gives strong evidence to the validity of the following conjecture, which was the author's initial motivation for studying locally conformally flat manifolds.

Conjecture. There are only finitely many homotopy (homeomorphism, diffeomorphism) types of locally conformally flat manifolds satisfying

$$\mathrm{Ric}(M) \geq -\Lambda^2, \quad \mathrm{diam}(M) \leq D, \quad \mathrm{vol}(M) \geq V.$$

2. Nonnegative Ricci curvature. As pointed out in the introduction, we will use heavily the results from **[SY]**. Since **[SY]** is a long paper with many results, we will summarize here what is needed for our argument.

By the definition of locally conformally flat manifolds and a standard monodromy argument (as in the proof of analytic continuation), it is easy to construct a conformal map $\Phi: \widetilde{M} \to S^n$ which is unique up to conformal transformations of S^n. Φ is called the developing map. It is an easy consequence of the existence of the developing map that any compact simply connected locally conformally flat manifold is conformally equivalent to S^n (originally due to Kuiper (**[Ku1]**, **[Ku2]**)). In the general case, the significance of the developing map is at least twofold. Firstly, it gives in a natural way a compactification for $\widetilde{M}$ which makes the analysis easier when $\widetilde{M}$ is not compact. Secondly, when Φ is injective, it gives a uniformization for locally conformally flat manifolds, exhibiting them as quotients of domains in the sphere by Kleinian groups. The major result of **[SY]** is to find a class of manifolds for which the developing maps are injective. In order to state the results from **[SY]**, we need to consider the conformal Laplacian L_g, which, when acting on a function ϕ, is defined as

$$L_g\phi = \Delta\phi - \frac{n-2}{4(n-1)}R(g)\phi,$$

where $R(g)$ is the scalar curvature of g and Δ is the usual (negative) Laplacian. L_g is conformally invariant in the sense that for any conformal metric $g_* = u^{4/(n-2)}g$, we have

$$L_{g_*}(\phi) = u^{-(n+2)/(n-2)}L_g(u\phi). \tag{1}$$

Letting $\phi = 1$, we get the Yamabe equation:

$$\Delta u - \frac{n-2}{4(n-1)}Ru = -\frac{n-2}{4(n-1)}u^{(n+2)/(n-2)}R(g_*). \tag{2}$$

By the help of the developing map, it is quite standard to show that the conformal Laplacian L_g of $\widetilde{M}$ has a minimal Green's function on $\widetilde{M}$, denoted by G_p, where p is the pole. We will now state the result we need from **[SY]** as the following lemma.

LEMMA 1 ([**SY**]). *Let (M^n, g) be a complete locally conformally flat Riemannian manifold with nonnegative scalar curvature and $\Phi: \widetilde{M} \to S^n$ the developing map. Then,*

(1) *Φ is injective (Theorem 4.5 in* [**SY**]).

(2) *$\partial\Phi(\widetilde{M}) \subset S^n$ is of codimension at least two (Propositions 3.3 and 4.4 in* [**SY**]).

(3) *For any $\varepsilon > 0$, any open set O containing p (Proposition 2.4(iii) in* [**SY**]),

$$\int_{\widetilde{M}\setminus O} G^{(n+\varepsilon)/(n-2)}\, dv_g < \infty.$$

As pointed out in the introduction, our strategy is to give a good estimate for $\dim(\partial\Phi(\widetilde{M}))$, where dim is the Hausdorff dimension. The idea in [**SY**] is to consider the quantity

$$d(M) = \inf\left\{ r \,\middle|\, \int_{M\setminus O} G^{2r/(n-2)}\, dv_g < \infty \right\},$$

and proved $\dim(\partial\Phi(\widetilde{M})) \le d(\widetilde{M})$. The starting point of our investigation is that this inequality is not sharp for the following trivial example, and in trying to give a sharp estimate for this example we obtained a proof of Theorem 1.

EXAMPLE. Consider (R^n, ω_0) and (S^n, g_0) where the metrics are the standard metrics. Let $\Psi: (R^n, \omega_0) \to (S^n, g_0)$ be the stereographic projection (which is the developing map for (R^n, ω_0)) defined as

$$\Psi(y_1, \dots, y_n) = \left(\frac{2y_1}{1+|y|^2}, \dots, \frac{2y_n}{1+|y|^2}, \frac{|y|^2-1}{|y|^2+1} \right),$$
$$\Psi^{-1}(x_1, \dots, x_n, \xi) = \left(\frac{x_1}{1-\xi}, \dots, \frac{x_n}{1-\xi} \right).$$

Then,

$$(\Psi^{-1})^*(\omega_0) = \frac{1}{(1-\xi)^2} g_0 = u^{4/(n-2)} g_0,$$
$$\Psi^*(g_0) = \frac{4}{(1+|y|^2)^2}\omega_0,$$

where $u = 1/(1-\xi)^{(n-2)/2}$. The Green's function for (R^n, ω) at 0 is

$$G_0(y) = \frac{1}{(n-2)\omega_{n-1}}|y|^{2-n} = c_n|y|^{2-n}.$$

The Green's function for (S^n, g_0) at S (the south pole) is

$$\begin{aligned} H_S(x_1, \dots, x_n, \xi) &= u(0, \dots, -1)^{-(n+2)/(n-2)} u \\ &\quad \cdot (\Psi^{-1})^*(G_0)(x_1, \dots, x_n, \xi) \\ &= 2^{(n+2)/2} c_n (1+\xi)^{(2-n)/2}. \end{aligned}$$

Similarly, for the north pole N,

$$H_N = 2^{(n+2)/2} c_n (1-\xi)^{(2-n)/2}.$$

From the formula for G_0, we see that $d(R^n, \omega_0) = \frac{n}{2}$. This shows that (3) of Lemma 1 is sharp. But obviously $\dim(\partial\Psi(R^n)) = 0$. Thus the inequality $\dim(\partial\Phi(\widetilde{M})) \leq d(\widetilde{M})$ is not sharp when $M = R^n$.

Since the functions in the above examples are explicit, it is not hard to give an analytic proof that $\dim(\partial\Psi(R^n)) = 0$. Because this proof illustrates the idea for the proof of Theorem 1, we will first give a proof in this case.

To this end, as in **[SY]**, we consider the concept of capacity, which is easier to handle analytically than the Hausdorff dimension.

Definition. For a subset $S \subset (M^n, g)$, we define

$$C_p(S) = \inf_{\phi}\left\{\int_M |\nabla_g \phi|^p\, dx \colon \phi \in C_0^\infty,\ \phi|_O = 1\right\},$$

where O is some open set containing S.

The relation between capacity and the Hausdorff dimension is that if $C_p(S) = 0$, then $\dim(S) \leq n - p$ (**[AM]**).

Example (continued). We now give an analytic proof that $\dim(\partial\Psi(R^n)) = 0$. In fact, choose a function $\phi_a \colon R^n \to R$ such that

$$\phi_a(y_1, \dots, y_n) = \begin{cases} 0, & |y| \leq a, \\ 1, & |y| \geq 2a, \end{cases}$$

and $|\nabla_{\omega_0}\phi_a| \leq 2/a$. Note that

$$|\nabla_{\Psi^*(g_0)}\phi_a| = |\nabla_{4\omega_0/(1+|y|^2)^2}\phi_a| = \frac{1+|y|^2}{2}|\nabla_{\omega_0}\phi_a|.$$

Thus,

$$\begin{aligned}\int_{S^n\backslash N} |\nabla_{g_0}(\Psi^{-1})^*(\phi_a)|^{n-\varepsilon}\, dv_{g_0} &= \int_{R^n} |\nabla_{\Psi^*(g_0)}\phi_a|^{n-\varepsilon}\, dv_{\Psi^*(g_0)} \\ &= \int_{R^n}\left(\frac{1+|y|^2}{2}|\nabla_{\omega_0}\phi_a|\right)^{n-\varepsilon}\left(\frac{2}{1+|y|^2}\right)^n dv_{\omega_0} \\ &\le \frac{2^n}{a^{n-\varepsilon}}\int_{a\le|y|\le 2a}(1+|y|^2)^{-\varepsilon}\, dv_{\omega_0} \le \frac{2^n}{a^{n-\varepsilon}}(1+a^2)^{-\varepsilon}\int_{a\le|y|\le 2a} dv_{\omega_0} \\ &\le \frac{2^n}{a^{n-\varepsilon}}(1+a^2)^{-\varepsilon}\cdot\omega_n(2a)^n = \frac{4^n\omega_n a^\varepsilon}{(1+a^2)^\varepsilon} \to 0 \quad \text{as } a\to\infty.\end{aligned}$$

Thus $C_{n-\varepsilon}(\partial\Psi(R^n)) = 0$ for any $\varepsilon > 0$. Hence $\dim(\partial\Psi(R^n)) = 0$.

We are now ready to prove the following main lemma.

LEMMA 2. *Let (M^n, g) be a locally conformally flat manifold with nonnegative Ricci curvature. Let Φ: $(\widetilde{M}, g) \to (S^n, g_0)$ be the developing map. Then,*

$$\dim(\partial\Phi(\widetilde{M})) = 0.$$

Proof. By Lemma 1, Φ is injective, thus we can view $\widetilde{M}$ as a subset of S^n, and there is a function u: $\widetilde{M} \to R^+$ such that $\Phi^* g_0 = u^{-4/(n-2)} g$. Without loss of generality, we assume $\Phi(p) = N$. By equation (1), we calculate,

$$\begin{aligned}L_g(u^{-1}\cdot\Phi^*(H_N)) &= L_{(\Phi^{-1})^*g}((\Phi^{-1})^*(u^{-1})\cdot H_N) \\ &= L_{[(\Phi^{-1})^*(u)]^{4/(n-2)}g_0}((\Phi^{-1})^*(u^{-1})\cdot H_N) \\ &= [(\Phi^{-1})^*(u)]^{-(n+2)/(n-2)}L_{g_0}(H_N) \\ &= u(p)^{-(n+2)/(n-2)}\delta_N\end{aligned}$$

thus $L_g(u(p)^{(n+2)/(n-2)}\cdot u^{-1}\cdot\Phi^*(H_N)) = \delta_p$. Using (2) of Lemma 1 and the minimality of G_p, it is standard to conclude that $G_p = u(p)^{(n+2)/(n-2)}\cdot u^{-1}\cdot\Phi^*(H_N)$ (see [SY], p. 55). Therefore, the integrability condition in (3) in Lemma 1 is equivalent to

$$\int_{\widetilde{M}\backslash O} u^{-(n+\varepsilon)/(n-2)}\, dv_g < \infty. \tag{3}$$

(Note that H_N is bounded in $S^n\backslash O$.) Now for any $a > 0$, we choose, as in the example, a function ϕ_a on $\widetilde{M}$, such that

$$\phi_a(x) = \begin{cases} 0, & d_g(x, p) \le a, \\ 1, & d_g(x, p) \ge 2a,\end{cases}$$

and $|\nabla_g \phi_a| \le 2/a$. Then

$$\begin{aligned}
\int_{\Phi(\widetilde{M})} |\nabla_{g_0}(\Phi^{-1})^*(\phi_a)|^{n-\varepsilon}\, dv_{g_0} &= \int_{\widetilde{M}} |\nabla_{\Phi^*(g_0)}\phi_a|^{n-\varepsilon}\, dv_{\Phi^*(g_0)} \\
&= \int_{\widetilde{M}} |(\nabla_g \phi_a) u^{2/(n-2)}|^{n-\varepsilon} u^{-2n/(n-2)}\, dv_g \\
&\le \left(\frac{2}{a}\right)^{n-\varepsilon} \int_{a\le d(x,p)\le 2a} u^{-2\varepsilon/(n-2)}\, dv_g \\
&\le \left(\frac{2}{a}\right)^{n-\varepsilon} \left(\int_{a\le d(x,p)\le 2a} u^{-(n+\varepsilon)/(n-2)}\, dv_g\right)^{2\varepsilon/(n+\varepsilon)} \\
&\quad \cdot \left(\int_{a\le d(x,p)\le 2a} dv_g\right)^{(n-\varepsilon)/(n+\varepsilon)} \\
&\le \left(\frac{2}{a}\right)^{n-\varepsilon} \left(\int_{\widetilde{M}\setminus O} u^{-(n+\varepsilon)/(n-2)}\, dv_g\right)^{2\varepsilon/(n+\varepsilon)} \\
&\quad \cdot (\mathrm{vol}_{\widetilde{M}}(B_p(2a)))^{(n-\varepsilon)/(n+\varepsilon)} \\
&\le C \cdot \frac{1}{a^{n-\varepsilon}} [\omega_n (2a)^n]^{(n-\varepsilon)/(n+\varepsilon)} \\
&= C a^{-\varepsilon(n-\varepsilon)/(n+\varepsilon)} \to 0 \quad (\text{as } a \to +\infty),
\end{aligned}$$

where in the last inequality we have used (3) and the Bishop volume comparison theorem. Thus $C_{n-\varepsilon}(\partial\Phi(\widetilde{M})) = 0$ for any $\varepsilon > 0$. Hence $\dim(\partial\Phi(\widetilde{M})) = 0$. □

Proof of Theorem 1. Since any manifold with more than one end contains a line, it follows from the Cheeger-Gromoll splitting theorem that a manifold of nonnegative Ricci curvature has at most two ends. Consider the developing map $\Phi: \widetilde{M} \to S^n$. Each end of $\widetilde{M}$ gives a connected component of $\partial\Phi(\widetilde{M})$; therefore, $\partial\Phi(\widetilde{M})$ has at most two connected components. By Lemma 2, $\partial\Phi(\widetilde{M})$ consists of at most two points. We therefore have the following three cases.

(1) If $\partial\Phi(\widetilde{M})$ is empty, then $\widetilde{M}$ is conformally equivalent to S^n.

(2) If $\partial\Phi(\widetilde{M})$ has only one point, then $\widetilde{M}$ is conformally equivalent to R^n.

(3) If $\partial\Phi(\widetilde{M})$ has two points, by composing Φ with a conformal transformation of S^n, we can assume $\partial\Phi(\widetilde{M}) = \{S, N\}$. Writing the metric of S^n in polar coordinates, we have $g = u(t, x)(dt^2 + \sin^2 t\, d\sigma)$ where $d\sigma$ is the standard metric on S^{n-1}. On the other hand, by the splitting theorem, $\widetilde{M}$ is isometric to $R \times N$ with N closed and

simply connected, hence conformally equivalent to S^{n-1}. Therefore, the metric g can be written as $g = dr^2 + f^2(x)\, d\sigma$. It follows that the function u is independent of x. By a change of the parameter t, we conclude that $\widetilde{M}$ is isometric to $R \times S^{n-1}$ with S^{n-1} of constant curvature.

In the case when M is compact, we only need to show that in case (2) $\widetilde{M}$ is actually isometric to R^n. In fact, from (2), there is a positive function u on M with $\omega_0 = u^{4/(n-2)} g$. The Yamabe equation (2) implies

$$\Delta u - \frac{n-2}{4(n-1)} R u = 0.$$

Thus u satisfies the maximal principle. Since M is compact, u is a constant. This shows $\widetilde{M}$ is isometric to R^n. □

COROLLARY. *If* (M^n, g) *is an open locally conformally flat manifold with*

$$\mathrm{Ric} \geq 0, \quad \mathrm{vol}(B_p(r)) \geq c r^n$$

for some point $p \in M$ *and some constant* $c > 0$, *where* $B_p(r)$ *is the geodesic ball of radius* r *around* p, *then* M^n *is conformally equivalent to* R^n.

Proof. It is well known that $\pi_1(M)$ is finite and $\widetilde{M}$ has only one end; thus $\widetilde{M}$ is conformally equivalent to R^n. This implies that $\pi_1(M)$ is torsion free, hence trivial. Therefore, M is conformally equivalent to R^n. □

REMARK. This corollary says that the local model in the sense of M. Anderson ([**An**]) for the class in the conjecture in §1 is conformally equivalent to R^n. This gives evidence that the conjecture is correct.

We end this section with a family of examples of conformally flat metrics on R^n with nonnegative Ricci curvature and various volume growth.

EXAMPLE. Let (R^n, ω_0) be the standard flat metric on R^n. Consider $g = (r^2+1)^{-2\alpha}\omega_0$, a globally conformally flat metric. It follows easily from a direct computation that

$$\mathrm{Ric}_{ii} = \frac{4\alpha(n-2)(1-\alpha)(r^2 - x_i^2)}{(r^2+1)^2} + \frac{4(n-1)\alpha}{(r^2+1)^2}.$$

Thus when $0 \leq \alpha \leq 1$, we have $\mathrm{Ric} \geq 0$. It's also easy to see,

(a) $\frac{1}{2} < \alpha \leq 1$: $\mathrm{Ric} \geq 0$, noncomplete;

(b) $\alpha = \frac{1}{2}$: $\mathrm{Ric} \geq 0$, complete, $\mathrm{vol}(B(r)) = c_n r$;
(c) $0 \leq \alpha < \frac{1}{2}$: $\mathrm{Ric} \geq 0$, complete, $\mathrm{vol}(B(r)) = c_n r^n$.

3. Estimating Betti numbers. As pointed out in §1, Theorem 2 is a consequence of a general result stated in **[Be]** and the following well-known Weizenböck formula. Since the proof is simple, for completeness, we will give a detailed proof of Theorem 2 here. The first part of the proof is the standard Moser iteration. The second part is what is known as Peter Li's lemma.

LEMMA 3 (**[G1]**). *Let (M^n, g) be a compact locally conformally flat Riemannian manifold and ϕ a harmonic p-form. Then*

$$\Delta|\phi|^2 = 2|\nabla\phi|^2 + \frac{2p(n-2p)}{n-2} R_{ij}\phi^{ii_2\cdots i_p}\phi^{j}_{i_2\cdots i_p} + \frac{2p(p-1)\cdot p!}{(n-1)(n-2)} R|\phi|^2,$$

where R_{ij} is the Ricci tensor of g.

Proof of Theorem 2. Let us assume that $\mathrm{Ric} \geq -\Lambda^2$ and $\mathrm{diam}(M) = D$. Then $R \geq -n(n-1)\Lambda^2$. It follows from Lemma 3 that

$$\Delta|\phi|^2 \geq 2|\nabla\phi|^2 - c(n, p)\Lambda^2|\phi|^2,$$

where $c(n, p)$ is a constant depending only on n and p. In what follows constants will always be denoted in this way, while their values may change. From the definition of the Laplacian, we have

$$\Delta|\phi|^2 = 2|\nabla|\phi||^2 + 2|\phi|\Delta|\phi|.$$

Thus,

$$|\phi|\Delta|\phi| \geq |\nabla\phi|^2 - |\nabla|\phi||^2 - c(n, p)\Lambda^2|\phi|^2.$$

By the Schwarz inequality, it is easy to see $|\nabla\phi|^2 \geq |\nabla|\phi||^2$; therefore,

$$-\Delta|\phi| \leq c(n, p)\Lambda^2|\phi|.$$

Multiply both sides by $|\phi|^{2k-1}$ for $k > 1/2$, and integrate by parts,

$$\int \nabla(|\phi|^{2k-1}) \cdot \nabla|\phi| \leq c(n, p)\Lambda^2 \int |\phi|^{2k},$$

that is,

$$\|\nabla|\phi|^k\|_2 \leq \frac{k\Lambda c(n, p)}{\sqrt{2k-1}} \| |\phi|^k\|_2.$$

Recall the Sobolev inequality for a Riemannian manifold says (**[Be]**),

$$\|f\|_{2n/(n-2)} \leq \frac{1}{V^{1/n}}(c(n, D\Lambda) \cdot D\|\nabla f\|_2 + \|f\|_2)$$

for any $f \in W^{1,2}$, where $V = \mathrm{vol}(M)$. Using the Sobolev inequality, we continue the previous inequality,

$$\|\phi\|^2_{2nk/(n-2)} \le \left(1 + c(n, p, D\Lambda)\frac{k}{\sqrt{2k-1}}\right)^{2/k} V^{-2/nk}\|\phi\|^2_{2k}.$$

Let $k = (\frac{n}{n-2})^i$, and multiply all inequalities with $i = 0, 1, \dots$, we deduce,

$$\|\phi\|^2_\infty \le \prod_{i=0}^{\infty}\left(1 + c(n, p, D\Lambda)\frac{\tau^i}{\sqrt{2\tau^i - 1}}\right)^{2/\tau^i} V^{-1}\|\phi\|^2_2,$$

where we have denoted $\tau = \frac{n}{n-2}$. It is easy to see that the product in the above inequality converges.

Let H^p be the space of harmonic p-forms with the L^2 inner product. By the Hodge theory, $\dim(H^p) = b_p$. Let $\psi_1, \dots, \psi_{b_p}$ be an orthonormal basis for H^p. Consider the following function on M,

$$f(x) = \sum_{i=1}^{b_p}\frac{|\psi_i(x)|^2}{\int_M |\psi_i|^2}.$$

Note f is independent of the choice of orthonormal basis. Let $f(x_0) = \max f$. Define a map $H^p \xrightarrow{s} \bigwedge^p(T^*_{x_0}M)$ by $s(\psi) = \psi(x_0)$. Then $H^p = \mathrm{Ker}\, s \oplus (\mathrm{Ker}\, s)^\perp$. Let $\{\phi_i\}$ be an orthonormal basis adapted to this decomposition, there are at most $\dim(\mathrm{Ker}\, s)^\perp$ of the ϕ_i's with $\phi_i(x_0) \ne 0$. Thus,

$$f(x_0) \le \dim(\mathrm{Ker}\, s)^\perp \cdot \max_i \frac{\sup|\phi_i|^2}{\int_M |\phi_i|^2} \le \binom{n}{p}\cdot \sup_\phi \frac{\|\phi\|^2_\infty}{\|\phi\|^2_2}.$$

Therefore,

$$\begin{aligned} b_p(M, R) = \int_M f(x)\, dv_g \le f(x_0)\cdot V &\le \binom{n}{p} \sup_\phi \frac{\|\phi\|^2_\infty}{\|\phi\|^2_2}\cdot V \\ &\le \binom{n}{p}\prod_{i=0}^{\infty}\left(1 + c(n, p, D\Lambda)\frac{\tau^i}{\sqrt{2\tau^i - 1}}\right)^{2/\tau^i} \\ &\le C(n, D\Lambda). \end{aligned}$$

□

References

[AM] D. Adams and N. Meyers, *Bessel potentials, including relations among classes of exceptional sets*, Indiana Univ. Math. J., **22** (1973), 873–905.

[An] M. Anderson, *Convergence and rigidity of manifolds under Ricci curvature bounds*, Invent. Math., **102** (1990), 429–445.

[Be] P. Berard, *From vanishing theorems to estimating theorems: the Bochner technique revisited*, Bull. Amer. Math. Soc., **19** (1988), 371–406.

[CG] J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of non-negative Ricci curvature*, J. Differential Geom., **6** (1971), 119–129.

[Fr] D. Fried, *Closed similarity manifolds*, Comment. Math. Helv., **55** (1980), 576–582.

[Gl] S. Goldberg, *Curvature and Homology*, Dover Publications, New York, 1982.

[Go] W. Goldman, *Conformally flat manifolds with nilpotent holonomy and the uniformization problem for* 3*-manifolds*, Trans. Amer. Math. Soc., **278** (1983), 573–583.

[Gr] M. Gromov, *Curvature, diameter and Betti numbers*, Comment. Math. Helv., **56** (1981), 179–195.

[Ka] Y. Kamishima, *Conformally flat manifolds whose developing maps are not surjective*, Trans. Amer. Math. Soc., **294** (1986), 609–623.

[Ku1] N. Kuiper, *On conformally flat manifolds in the large*, Ann. of Math., **50** (1949), 916–924.

[Ku2] ——, *On compact conformally Euclidean spaces of dimension* > 2, Ann. of Math., **52** (1950), 478–490.

[Li] P. Li, *On the Sobolev constant and the* p*-spectrum of a compact Riemannian manifold*, Ann. Sci. Ecole. Norm. Sup., **13** (1980), 451–469.

[No] M. Horonha, *Some compact conformally flat manifolds with nonnegative scalar curvature*, preprint.

[Pa] S. Patterson, *The limit set of a fuchsian group*, Acta Math., **136** (1976), 241–273.

[SY] R. Schoen and S.-T. Yau, *Conformally flat manifolds, Kleinian groups and scalar curvature*, Invent. Math., **92** (1988), 47–71.

[Su] D. Sullivan, *The density at infinity of a discrete group of hyperbolic motions*, Inst. Hautes Études Sci. Publ. Math., **50** (1979), 171–202.

Received April 8, 1992. Research supported by an NSF grant.

UNIVERSITY OF CALIFORNIA
LOS ANGELES, CA 90024-1555

PACIFIC JOURNAL OF MATHEMATICS

Volume 163 No. 1 March 1994

0030-8730(1994)163:1;1-J