ILLINOIS JOURNAL OF MATHEMATICS
Volume 28, Number 1, Spring 1984

ON THE CLASSIFICATION OF GENERIC BRANCHED COVERINGS OF SURFACES[1]

BY

ISRAEL BERSTEIN AND ALLAN L. EDMONDS

1. Introduction

Almost a century ago, A. Hurwitz showed, in principle, how to classify the branched coverings between two surfaces [4]. In this paper we pursue his approach and prove a new uniqueness theorem for a wide class of "generic" branched coverings. Among all branched coverings between two given surfaces the *simple* ones are generic: A branched covering $\phi : M \rightarrow N$ of degree d is simple if $\# \phi^{-1}(y) \geq d-1$ for all $y \in N$; any branched covering between surfaces can be approximated by a simple branched covering; and any branched covering close enough to a simple branched covering is itself simple. See [1] for an indication of proofs of these facts. Two branched coverings ϕ, $\psi : M \rightarrow N$ are said to be *equivalent* if there are homeomorphisms $f: M \rightarrow M$ and $g : N \rightarrow N$ such that $g \circ \phi = \psi \circ f$.

The classical function-theorist Clebsch [2], extending work of Luroth [6], proved that simple branched coverings of the sphere S^2 are uniquely determined up to equivalence by their domain and degree. See also [4]. A proof is sketched in [1]. We shall also reprove this result in Section 4, for the sake of completeness.

The theory of ordinary covering spaces intervenes in the classification of branched coverings of surfaces of higher genus which do not induce surjections between fundamental groups. Therefore we single out for study here the *primitive* branched coverings which do induce surjections between fundamental groups, and hence cannot be factored as a branched covering followed by an ordinary covering.

(1.1) UNIQUENESS CONJECTURE. Any two primitive, simple branched coverings of degree d between closed, orientable, connected surfaces are equivalent.

At this time (1.1) has not been proved in complete generality. Nonetheless we prove it in several significant cases here. In Section 5 we prove (1.1) for all primitive, simple branched coverings of the torus.

We say that a simple branched covering $\phi : M \rightarrow N$ of degree d is *metastable* if $d\chi(N) - \chi(M) > d/2$, or, equivalently, if ϕ has more than $d/2$ branch

Received October 15, 1981.

[1]Supported in part by grants from the National Science Foundation.

points. Note, for example, that a primitive, simple branched covering of degree ≤ 3 is metastable. (For the notion of a stable branched covering, see (4.6).) In Section 6 we prove (1.1) for metastable branched coverings. See Section 6 for other cases in which we can prove (1.1).

As further supporting evidence for (1.1) we prove in Section 8 a uniqueness result for "primitive, symplectic maps of degree d" between nonsingular symplectic inner product spaces over the integers. In particular, (1.1) holds on the level of homology.

The remaining Section 7 is devoted to applications of the main uniqueness results. These applications are all based upon the results of [3] characterizing those maps of surfaces which are homotopic to branched coverings. One such consequence of (1.1) is that up to equivalence there is at most one primitive map of given nonzero degree between two given surfaces. Another is an affirmative solution of the following, modulo a proof of a weaker version of (1.1).

(1.2) Simple Loop Conjecture. If $f : M \to N$ is a map of closed surfaces such that $f_* : \pi_1(M) \to \pi_1(N)$ is not injective, then there is a nontrivial simple loop $C \subset M$ such that $f|C$ is nullhomotopic.

This problem was first brought to our attention in 1977 by T. Tucker and was partially answered in [3]. We note that the analog of (1.2) for bounded planar surfaces is known to be false.

As a final application we are able to characterize those primitive maps of prime degree between surfaces which are homotopic to regular branched coverings, that is, to orbit maps for $\mathbf{Z}/p$ actions.

In the preliminary Sections 2 and 3 we develop the basic techniques for dealing with branched coverings for application in the remainder of the paper.

2. The Hurwitz classification of branched coverings

For the purposes of this paper a *branched covering* is a finite-to-one, open map $\phi : M \to N$ between compact, orientable surfaces which is an ordinary covering over the complement of a finite set in int N. The *singular set* of ϕ is the set Σ_ϕ of points $x \in M$ near which ϕ fails to be a local homeomorphism. The *branch set* is $B_\phi = \phi\Sigma_\phi$. The *degree of* ϕ is given by

$$\deg \phi = \max\{\# \phi^{-1}(y) : y \in N\},$$

which is easily seen to be the absolute value of the homological degree of ϕ with respect to chosen orientations of M and N.

A branched covering $\phi : M \to N$ of degree d is uniquely determined as the extension to end compactifications of the associated unbranched covering $\phi_0 : M_0 \to N_0$, where $N_0 = N - B_\phi$ and $M_0 = M - \phi^{-1}B_\phi$. Alternatively think of "coning off" the ideal boundary components of M_0 and N_0. The covering projection ϕ_0 is determined by a homomorphism $\varrho(\phi) : \pi_1(N_0,*) \to \mathcal{S}_d$. where $* \in N_0$ is a base point and $\mathcal{S}_d$ denotes the symmetric group on d symbols. The representation ϱ is determined up to an inner automorphism of $\mathscr{S}_d$ by choosing a one-to-one correspondence between $\phi^{-1}(*)$ and $\{1,2,\ldots,d\}$ and assigning to a loop γ in N_0, based at $*$, the permutation of $\{1,2,\ldots,d\}$ induced by transporting $\phi^{-1}(*)$ around γ using the path lifting property.

In this modern terminology we have the following two interpretations of fundamental results of Hurwitz [4].

(2.1) CLASSIFICATION THEOREM. *Two branched coverings of degree d,*

$$\phi_1 : M_1 \to N \quad \text{and} \quad \phi_2 : M_2 \to N,$$

over a compact, connected, orientable surface N, are equivalent if and only if there exist a homeomorphism

$$h : (N, B_{\phi_1}, *) \to (N, B_{\phi_2}, *)$$

and an inner automorphism $\mu : \mathscr{S}_d \to \mathscr{S}_d$ such that

$$\mu \circ \varrho(\phi_1) = \varrho(\phi_2) \circ h_*. \ \square$$

(2.2) EXISTENCE THEOREM. *For any compact, connected orientable surface N, finite set $B \subset \operatorname{int} N$, and representation*

$$\varrho : \pi_1(N - B, *) \to \mathscr{S}_d,$$

such that ϱ is nontrivial on each class represented by a small loop about any single point of B, there is a branched covering $\phi : M \to N$ with $B_\phi = B$ and $\varrho(\phi) = \varrho$. $\square$

Of course (2.1) and (2.2) are virtually immediate consequences of the classification of ordinary coverings via the fundamental group. They were originally explained by Hurwitz using "cut and paste" techniques. Their utility is that they reduce questions of existence and uniqueness to virtually combinatorial problems, in a way which we now formulate.

Fix an orientation of the compact, connected, orientable surface N, let $B \subset \operatorname{int} N$ be a finite set of cardinality n, and let $* \in N - B$ be a base point. Let $D \subset \operatorname{int} N - B$ be a small disk centered at $*$. The orientation of N induces one on $N - D$, and hence one on ∂D, using the convention that the orientation of a boundary component, followed by an inward normal, should coincide with the given orientation of N. Similarly each component of ∂N is oriented. Let $c_1, c_2, \ldots, c_k$ denote the oriented boundary components.

Let $a_1, b_1, \ldots, a_g, b_g$ be a maximal family of simple closed curves in $\operatorname{int} N$ such that $a_i \cap a_j = b_i \cap b_j = a_i \cap b_j = \emptyset$ if $i \neq j$, and $a_i \cap b_i$ is a single point of transverse intersection for each i. Orient each of these curves so that the orientation of a_i followed by that of b_i corresponds to the orientation of N at $a_i \cap b_i$, and so that the induced orientation on a curve representing the commutator $a_i b_i a_i^{-1} b_i^{-1}$ is the same as that induced by $N - a_i \cup b_i$ using our preceding convention.

Now choose $n + g + k$ simple arcs r_i with disjoint interiors running from $*$ to the points of B, to the points $a_1 \cap b_1, \ldots, a_g \cap b_g$, and to the boundary components $c_1, c_2, \ldots, c_k$, in order. We choose and index $\{r_i\}$ so that in the disk D about $*$ they correspond to distinct radii, indexed cyclically, consistent with the orientation of ∂D. We choose each arc r_j to $a_i \cap b_i$ so that at $a_i \cap b_i$ the orientation of a_i followed by the backward orientation of r_j toward $*$ is positive, while the orientation of b_i followed by this backward orientation is negative.

Around each point of B choose a small disk D_i whose boundary is oriented by $N - D_i$.

These choices identify $\pi_1(N-B,*)$ as a free group with $n+2g+k$ generators $w_1, \ldots, w_n, x_1, y_1, \ldots, x_g, y_g, z_1, \ldots, z_k$, subject to the single relation $w_1 \ldots w_n[x_1,y_1] \ldots [x_g,y_g]z_1 \ldots z_k = 1$, where $[x_i,y_i]$ denotes the commutator $x_i y_i x_i^{-1} y_i^{-1}$.

Now a representation $\varrho : \pi_1(N-B,*) \to \mathcal{S}_d$ determines, and is determined by, a sequence

$$(\sigma_1, \ldots, \sigma_n; \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g; \gamma_1, \ldots, \gamma_h)$$

of permutations in $\mathcal{S}_d$, subject only to the requirement that

$$\sigma_1 \ldots \sigma_n[\alpha_1,\beta_1] \ldots [\alpha_g,\beta_g]\gamma_1 \ldots \gamma_k = 1.$$

Note that, motivated by path multiplication in $\pi_1(N-B,*)$, we are adopting the convention that permutations shall be multiplied from left to right, as opposed to functional multiplication. Nevertheless, we shall occasionally denote the action of $\tau \in \mathcal{S}_d$ on $j \in \{1,2,\ldots,d\}$ by $\tau(j)$, as long as no confusion arises. Also

$$\alpha^\beta = \beta\alpha\beta^{-1} \quad \text{and} \quad [\alpha,\beta] = \alpha\beta\alpha^{-1}\beta^{-1}.$$

Suppose that N and $B \subset \operatorname{int} N$ are fixed and that two different systems of arcs and simple closed curves are chosen as described above. Then it easily follows from the classification of surfaces that there is a homeomorphism of N which takes one system onto the other. Suppose in addition that to the two systems of arcs and simple closed curves the same sequences of permutations of $\mathcal{S}_d$ are assigned. Then it follows from the theory of covering spaces that the homeomorphism which moves one system to the other can be lifted to give an equivalence of the corresponding branched coverings constructed using (2.2). If $\phi : M \to N$ is a branched covering of degree d between compact, connected, orientable surfaces corresponding, to a representation $\varrho : \pi_1(N-B_\phi,*) \to \mathcal{S}_d$, then the sequence

$$\mathcal{H} = (\sigma_1, \ldots, \sigma_n; \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g; \gamma_1, \ldots, \gamma_k)$$

of permutations corresponding to a set of arcs and simple closed curves, as above, will be called a *Hurwitz system* for ϕ.

In summary we have the following interpretation of (2.1).

(2.3) THEOREM. *Two branched coverings of degree d over a given com-*

pact, connected, orientable surface are equivalent if and only if they have Hurwitz systems which are conjugate by an element of $\mathscr{S}_d$. □

If $\phi : M \to N$ is a branched covering and N is connected, then M is *connected* if and only if the Hurwitz representation is *transitive*, i.e., $\varrho\pi_1(N - B_\phi,*) < \mathscr{S}_d$ acts transitively on $\{1,2,\ldots,d\}$. Note also that ϕ is *simple* if and only if the permutations $\sigma_1,\ldots,\sigma_n$ corresponding to branch points in a Hurwitz system are *transpositions.* Then the uniqueness theorem of Luroth and Clebsch stated in Section 1 is equivalent to the assertion that a degree d simple branched covering of S^2 has a Hurwitz system of the form

$$((1,2),\ldots,(1,2)\ ,\ (1,3)\ ,\ (1,3)\ ,\ (1,4)\ ,\ (1,4),\ldots,(1,d)\ ,\ (1,d)).$$

We note in passing that for a simple branched covering $\phi : M \to N$, the Riemann-Hurwitz formula takes the form

$$\chi(M) = d\chi(N) - n \quad \text{where } n = \# B_\phi.$$

The remainder of this section is devoted to an interpretation of primitivity of simple branched coverings in terms of the Hurwitz representation.

(2.4) LEMMA. *Let G be a transitive subgroup of the symmetric group $\mathscr{S}_d$ which contains a transposition. Then either $G = \mathscr{S}_d$ or the action of G preserves a nontrivial partition of the symbols* $1,2,\ldots,d$.

Proof. Let $H < G$ be the largest possible symmetric group on a subset of $\{1,2,\ldots,d\}$. After appropriate relabeling we may suppose that $H = \mathscr{S}_r$, the symmetric group on $\{1,2,\ldots,r\}$. Since G contains a transposition, $r \geq 2$. Suppose that $G \neq \mathscr{S}_d$, so that $r < d$.

For any $\tau \in G$ it follows that either

$$\tau\{1,\ldots,r\} = \{1,\ldots,r\}$$

or

$$\tau\{1,\ldots,r\} \cap \{1,\ldots,r\} = \emptyset.$$

For if not, then there would be some $\tau \in G$ such that, after appropriate relabeling, $\tau(1) = 1$ and $\tau(r+1) = r$. But this implies that $\tau(1,r)\tau^{-1} = (1,r+1) \in G$, which implies that $G > S_{r+1}$, contradicting the maximality of H. It follows that the G translates of $\{1,\ldots,r\}$ yield the desired partition of $\{1,\ldots,d\}$ into d/r subsets since G is transitive. □

Similar and more general results appear in [8; §1.7].

(2.5) PROPOSITION. *Let $\phi : M \to N$ be a simple branched covering of degree d between connected surfaces with $B_\phi \neq \emptyset$. Then ϕ is primitive if and only if the associated Hurwitz representation $\varrho : \pi_1(N - B_\phi,*) \to S_d$ is surjective.*

Proof. If ϕ is not primitive, then ϕ factors as $\pi\circ\psi$, where $\pi : N' \to N$ is a nontrivial ordinary covering and $\psi : M \to N'$ is a branched covering. Moreover deg $\psi \geq 2$ since $B_\phi \neq \emptyset$. Let $H < K < G$ be the sequence of groups

$$\pi_1(M - \phi^{-1}B_\phi,*) \to \pi_1(N' - \pi^{-1}B_\phi,*) \to \pi_1(N - B_\phi,*).$$

Then the Hurwitz representation ϱ is defined by the action of G on G/H by left translation. But then the G action respects the nontrivial grouping of H-cosets into K-cosets, which means ϱ cannot be surjective.

Conversely suppose that ϱ is not surjective and let $H < G$ be the inclusion of groups $\pi_1(M - \phi^{-1}B_\phi *) \to \pi_1(N - B_\phi,*)$. Because $\varrho(G)$ is a transitive subgroup of $\mathscr{S}_d$ which contains a transposition, (2.4) implies that the action of G on G/H preserves some nontrivial partition of the H-cosets. Let K be the union of the H-cosets which lie in the portion of the partition containing H. In fact K is the G stabilizer of K, so that K is a subgroup of G, properly between H and G.

Then there is a branched covering $\pi : N' \to N$ arising from the covering of $N - B_\phi$ corresponding to $K < \pi_1(N - B_\phi,*)$, and there is a lift $\psi : M \to N'$ of ϕ so that $\phi = \pi \circ \psi$. Since ϕ is simple and $B_\pi \subset B$, it follows that $B_\pi = \emptyset$; for otherwise some fiber of ϕ would contain more than one singular point. Thus π is a covering and $\pi_* : \pi_1(N',*) \to \pi_1(N,*)$ is proper injection. This implies that ϕ not primitive. □

We remark that (2.5) is true for branched coverings in any dimension, by the same proof.

3. Basic alterations of Hurwitz systems

Fix a simple branched covering $\phi : M \to N$ between closed, orientable surfaces and let $\mathscr{H} = (\sigma_1, \ldots, \sigma_n; \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g)$ be a Hurwitz system for ϕ. Here we record some simple ways of altering $\mathscr{H}$ to obtain new Hurwitz systems for ϕ. This amounts to studying the action of (the generators of) the group of isotopy classes of homeomorphisms of the surface N which preserve the set of branch points. Strictly speaking, we require in addition that a base point be fixed throughout.

For each alteration below we indicate the entries in the Hurwitz system which are to be changed, together with a suggestion to the reader for constructing the appropriate change in diagrams.

(3.1) $(\sigma_i, \sigma_{i+1}) \to (\sigma_i\sigma_{i+1}\sigma_i, \sigma_i)$. Simply twist a disk containing 2 successive branch points 180°. Note that the product $\sigma_i\sigma_{i+1}$ is preserved.

(3.2) $(\sigma_i, \sigma_{i+1}) \to (\sigma_{i+1}, \sigma_{i+1}\sigma_i\sigma_{i+1})$. This is the inverse of (3.1). (3.1) and (3.2) correspond to the standard generators of the braid group of the sphere.

Taken together, (3.1) and (3.2) imply that a given transposition σ_i can be moved to any other position in the sequence of transpositions at the expense of conjugating the intervening entries by σ_i.

(3.3) $(\alpha_i, \beta_i) \to (\alpha_i \beta_i, \beta_i)$. Perform a Dehn twist about β_i. Note that the commutator $[\alpha_i, \beta_i]$ is preserved.

(3.4) $(\alpha_i, \beta_i) \to (\alpha_i, \beta_i \alpha_i)$. Perform a Dehn twist about α_i. Note that $[\alpha_i, \beta_i]$ is preserved.

(3.5)

$$(\alpha_i, \beta_i, \alpha_{i+1}, \beta_{i+1}) \to (\alpha_i \gamma, \gamma^{-1} \beta_i \gamma, \gamma^{-1} \alpha_{i+1}, \beta_{i+1})$$

where $\gamma = \beta_i^{-1} \alpha_{i+1} \beta_{i+1} \alpha_{i+1}^{-1}$. Perform a Dehn twist about an appropriate simple closed curve representing γ.

We note that (3.3)–(3.5) correspond to the generators of the group of isotopy classes of homeomorphisms of N. Actually, in the present work we shall not have occasion to use (3.5). Here is one other (redundant) useful alteration involving only α's and β's, analogous to (3.1).

(3.6) $(\alpha_i, \beta_i, \alpha_{i+1}, \beta_{i+1}) \to (\alpha_{i+1}^{[\alpha_i, \beta_i]}, \beta_{i+1}^{[\alpha_i, \beta_i]}, \alpha_i, \beta_i)$. This alteration, together with its inverse, says that a pair α_i, β_i can be moved elsewhere in the sequence of α, β pairs, at the expense of conjugating intervening α, β pairs by the commutator

$$[\alpha_i, \beta_i] = \alpha_i \beta_i \alpha_i^{-1} \beta_i^{-1},$$

or its inverse.

Finally, we single out two alterations in which the two parts of the Hurwitz system interact.

(3.7) $(\sigma_n; \alpha_1, \beta_1) \to (\sigma_n^{\sigma_n \alpha_1 \beta_1 \alpha_1^{-1}}; \sigma_n \alpha_1, \beta_1)$. Pull the a_1 curve across the branch point. Thus one may multiply α_1 by σ_n without affecting the remainder of the second half of the system. When this is combined with (3.1) and (3.6), one may multiply any α_j by any σ_i at the expense of conjugating intervening terms.

(3.8) $(\sigma_n; \alpha_1, \beta_1) \to (\sigma_n^{\sigma_n \alpha_1}; \alpha_1^{\sigma_n}, \alpha_1 \sigma_n)$. Redraw the arc from $*$ to $a_1 \cap b_1$ around the branch point corresponding to σ_n; drag b_1 across the branch point; then redraw the arc to the branch point around α_1. This says that β_1 can be multiplied (on the right) by σ_n; or that α_1 can be conjugated by σ_n. Note that both the products $\sigma_n \alpha_1$ and $\sigma_n [\alpha_1, \beta_1]$ are preserved by this alteration. In particular, iteration of (3.8) yields

$$(\sigma_n^{(\sigma_n \alpha_1)^k}; \alpha_1^{(\sigma_n \alpha_1)^k}, \beta').$$

We conclude this section with a lemma which emphasizes the usefulness of "doubles"—that is, pairs of identical successive transpositions.

(3.9) PROPOSITION. (i) $(\sigma, \sigma, \tau) \to (\sigma^\tau, \sigma^\tau, \tau)$.

(ii) $(\sigma, \sigma; \alpha, \beta) \to (\sigma^\alpha, \sigma^\alpha; \alpha, \beta)$.

Proof. In neither case is it necessary that the transposition τ or the α, β pair be adjacent to the double σ. For (i), using (3.1) we may move σ,σ to the position adjacent to τ without changing the other entries, since they would be conjugated twice by the transposition σ. Now apply (3.2) twice to get $(\tau,\sigma^\tau,\sigma^\tau)$ and then apply (3.2) twice again to get $(\sigma^\tau,\tau^{\sigma^\tau},\sigma^\tau)$ and then $(\sigma^\tau,\sigma^\tau,\tau)$. Finally, one moves the new double back to its original position.

For (ii) using (3.1) and the analog of (3.6) for transpositions in place of the pair α_i,β_i, we temporarily move σ,σ to the position adjacent to α,β without changing the other entries. Now (3.8) yields $(\tau,\tau^{\tau\alpha};\alpha^\tau,\beta\tau)$. (3.1) yields

$$(\tau^\alpha,\tau;\alpha^\tau,\beta\tau).$$

Finally, (3.8) again yields

$$(\tau^\alpha,\tau^{\tau\alpha^\tau};\alpha^{\tau^2},\beta\tau^2) = (\tau^\alpha,\tau^\alpha;\alpha,\beta),$$

since $\tau^2 = 1$. □

4. Uniqueness over the sphere

In this section we give an exposition of the proof of the Uniqueness Conjecture (1.1) in the classical case of branched coverings of the sphere.

(4.1) Theorem (Clebsch). *A connected, simple branched covering $\phi : M \to S^2$ of degree d has a Hurwitz system of the form*

$$((1,2)\,,\,(1,2),\ldots,(1,2)\,,\,(2,3)\,,\,(2,3)\,,\,(3,4)\,,\,(3,4),\ldots,(d-1,d)\,,\,(d-1,d)).$$

The proof consists of using the operations described in Section 3 to put an arbitrary Hurwitz system for ϕ into the asserted form. We begin with a lemma which is essentially due to Luroth [6]. In the following $n(i,j)$ denotes a sequence of n transpositions (i,j).

(4.2) Lemma (Luroth). *Let $\mathscr{H} = (\sigma_1,\sigma_2,\ldots,\sigma_k)$ be a sequence of transpositions in $\mathscr{S}_d$. Then $\mathscr{H}$ is equivalent, using operations* (3.1) *and* (3.2) *and a conjugation of the entire sequence by an element of $\mathscr{S}_d$, to a sequence of the form*

$$(n_1(1,2),n_2(2,3),\ldots,n_{d-1}(d-1,d)).$$

Proof. We proceed by induction on d, and prove the slightly stronger statement that the desired form can be achieved using (3.1), (3.2) and a conjugation which fixes d. The result is clearly true for $d = 2$. Assume that $d > 2$ and that the result is true for $d-1$. Among all sequences which can be obtained from $\mathscr{H}$ by successive applications of (3.1) and (3.2) only, choose one which can be written as a concatenation $\mathscr{H}_1\mathscr{H}_2$ of two substrings $\mathscr{H}_1$ and $\mathscr{H}_2$, in which $\mathscr{H}_1$ does not involve d and is as large as possible subject to this constraint. We claim that $\mathscr{H}_2$ must then have the form $n(j,d)$ for some $j < d$. Certainly, by (3.2) and the maximality of $\mathscr{H}_1$, each element of $\mathscr{H}_2$ moves d. Suppose that $\mathscr{H}_2$ contains (i,d) and (j,d) for $i \neq j$. By (3.1) we may assume they are adjacent in $\mathscr{H}_2$. But then, applying (3.1) again, $((i,d)\ ,\ (j,d))$ yields $((i,j)\ ,\ (i,d))$. It follows that we can increase the size of $\mathscr{H}_1$, contradicting maximality. Conjugating the whole sequence by $(j,d-1)$, if necessary, we may assume that $j = d-1$.

Now examine $\mathscr{H}_1$. By the inductive hypothesis $\mathscr{H}_1$ can be put into the form $(n_1(1,2)\ ,\ n_2(2,3),\ldots,n_{d-2}(d-2,d-1))$ using (3.1) , (3.2), and a conjugation fixing $d-1$ (and d). These operations do not alter $n(d-1,d)$, and hence the desired form has been achieved. □

(4.3) LEMMA. *If* $\mathscr{H} = (\sigma_1,\sigma_2,\ldots,\sigma_k)$ *is a sequence of transpositions in* $\mathscr{S}_d$ *which generate a transitive subgroup of* $\mathscr{S}_d$, *then* $\mathscr{H}$ *can be put into the form*

$$(n_1(1,2)\ ,\ n_2(2,3),\ldots,n_{d-1}(d-1,d)),$$

where $n_i \geq 1$ *for all* i *and* $n_i \leq 2$ *for* $i \geq 2$, *using* (3.1), (3.2), *and a conjugation by an element of* $\mathscr{S}_d$.

Proof. By (4.2) we may assume that

$$\mathscr{H} = (n_1(1,2)\ ,\ n_2(2,3),\ldots,n_{d-1}(d-1,d)).$$

The transitivity assumption immediately implies that each $n_i \geq 1$. By application of an appropriate inductive hypothesis we may assume that $n_i \leq 2$ for $i \geq 3$. Now we examine the part

$$(n_1(1,2)\ ,\ n_2(2,3))$$

and show that if $n_2 > 2$ it can be decreased by 2. Consider a substring of the form $((1,2)\ ,\ (2,3)\ ,\ (2,3)\ ,\ (2,3))$. By the Doubles Lemma (3.9) we can conjugate the last two entries by (1,2) and then by (2,3), achieving

$$((1,2)\ ,\ (2,3)\ ,\ (1,2)\ ,\ (1,2)).$$

An application of (3.2) then yields $((1,2)\ ,\ (1,2)\ ,\ (1,2)\ ,\ (2,3))$. Repetition of these operations completes the proof of the lemma. □

Proof of (4.1). The transpositions in any Hurwitz system for ϕ generate a transitive subgroup of $\mathscr{S}_d$ since M is connected. By (4.3), ϕ has a Hurwitz system of the form

$$(n_1(1,2),n_2(2,3),\ldots,n_{d-1}(d-1,d))$$

where each $n_i \geq 1$ and $n_i \leq 2$ for $i \geq 2$. But the product of all the entries in the Hurwitz system must be 1. It follows that $n_i = 2$ for $i \geq 2$ and that n_i is even. □

(4.4) *Remark.* The appropriate classification statement in the case of connected, simple branched coverings of the disk D^2 is that such branched coverings are determined up to equivalence by degree d and the conjugacy class in $\mathscr{S}_d$ determined by the boundary curve, corresponding to the product of all the transpositions in a Hurwitz system. We leave the details to the reader.

(4.5) *Remark.* It is possible to push this approach somewhat further to prove the Uniqueness Conjecture for primitive, simple branched coverings $\phi : M \to N$ in what we call the *stable range*, where $d\,\chi(N) - \chi(M) \geq 2(d-1)$; that is,

$$\# B_\phi \geq 2(d-1).$$

In this case one can show that one can arrange a Hurwitz system to have $\alpha_1 = \beta_1 = \ldots = \alpha_g = \beta_g = (1)$, thus concentrating the branched covering within a disk; one then applies (4.1) to achieve a unique form which has an inviting geometric interpretation. Since this follows from the more general results of the next two sections we omit all details of the direct argument. After proving uniqueness in the stable range several years ago, we learned that a similar result appears in the unpublished Princeton thesis of R. Hamilton.

5. Uniqueness over the torus

In this section we prove the uniqueness conjecture (1.1) for primitive, simple branched coverings of the torus. We begin with a useful general criterion for producing doubles in a Hurwitz system for a simple branched covering.

Consider a triple $(\sigma,\tau;\alpha)$ of elements of $\mathcal{S}_d$, where σ and τ are transpositions, which we view as a segment $(\sigma_{n-1},\sigma_n;\alpha_1)$ of a Hurwitz system for a simple branched covering. We will use the two operations

$$(3.2) \quad (\sigma,\tau;\alpha) \to (\tau,\sigma^\tau;\alpha)$$

and

$$(3.8) \quad (\sigma,\tau;\alpha) \to (\sigma,\tau^{\tau\alpha};\alpha^\tau) = (\sigma,\tau^{\tau\alpha};\alpha^{\tau\alpha}).$$

Recall that the quantity "$\tau\alpha$" is preserved by (3.8). For $\gamma \in \mathcal{S}_d$, $|\gamma|$ denotes the *support* of γ,

$$\{x : 1 \leq x \leq d, \gamma(x) \neq x\}.$$

(5.1) Proposition. *Suppose that σ and τ are transpositions in $\mathcal{S}_d$ and $\alpha \in S_d$ such that $|\sigma|$ and $|\tau|$ are contained in the support of a single cycle of α. Then, using* (3.2) *and* (3.8), $(\sigma,\tau;\alpha)$ *can be converted into a triple $(\sigma',\tau';\alpha')$ such that $\sigma' = \tau'$ if and only if $\sigma\tau\alpha$ is conjugate to α.*

Proof. Note that under the given hypotheses on supports either $\sigma\tau\alpha$ contains two more components than α does in its decomposition into disjoint cycles, or $\sigma\tau\alpha$ is conjugate to α by an element of $\mathcal{S}_d$ supported in the cycle of α which supports σ and τ. In particular we may assume that α is a d-cycle.

Since (3.2) and (3.8) alter neither the product $\sigma\tau\alpha$ nor the conjugacy class of α, the condition that $\sigma\tau\alpha$ be conjugate to α is surely necessary.

For the converse, then, we assume that $\sigma\tau\alpha$ is also a d-cycle. By applying (3.8) a suitable number of times we may assume that

$$|\sigma| \cap |\tau| \neq \emptyset.$$

By relabeling we may assume that $(\sigma,\tau;\alpha)$ is

$$((1,i)\ ,\ (1,j)\ ;\ (1,2,\ldots,d)).$$

The condition that $\sigma\tau\alpha$ is a d-cycle is precisely that $i \leq j$. If $i = j$, then there is nothing further to prove. Therefore we assume $i < j$. We proceed by induction on the quantity $d-j$.

First suppose $d-j = 0$, so that $1 < i < j = d$. Applying (3.8) $d-i$ times we obtain $((1,i)\ ,\ (1,i)\ ;\ \alpha^{(\tau\alpha)^{d-i}})$.

Now inductively consider the case when $1 < i < j < d$. We have

$$\tau\alpha = (1,j+1,\ldots,d)\ (2,3,\ldots,j).$$

Apply (3.8) $j-i$ times. There are then two cases to consider.

Case 1. $(d-j)\,|\,(j-i)$. In this case $(\tau\alpha^{j-i})$ fixes 1. and we thus have obtained $((1,i)\ ,\ (1,i)\ ;\ \alpha^{(\tau\alpha)^{j-i}})$.

Case 2. $(d-j)\nmid(j-i)$. In this case we obtain

$$((1,i)\ ,\ (j+r,i)\ ;\ (\overset{1}{j+r},\overset{2}{i+1},\overset{3}{i+2},\ldots,j,2,\ldots,\overset{j}{i},\overset{j+1}{j+r+1},\ldots,\overset{j+s}{1},\ldots,\overset{d}{j+t})).$$

(The superscribed labels simply indicate position.) Here

$$j+1 \leq j+r \leq d \quad\text{and}\quad j+1 \leq j+s \leq d$$

for some $r,s \geq 1$. Now apply (3.2) to obtain the transpositions $(j+r,i)$, $(j+r,1)$. The indicated relabeling yields

$$((1,j)\ ,\ (1,j+s)\ ;\ (1,2,\ldots,d)).$$

Since $d-(j+s) < d-j$, induction applies to produce the desired double. $\square$

If $(\sigma_1,\ldots,\sigma_n;\alpha_1,\beta_1,\ldots,\alpha_g,\beta_g)$ is a Hurwitz system for a branched covering ϕ, then we say that α_1 is *maximal* if for any other Hurwitz system $(\sigma'_1,\ldots,\sigma'_n;\alpha'_1,\beta'_1,\ldots,\alpha'_g,\beta'_g)$ for ϕ,α'_1 has at least as many disjoint cycles (including trivial ones) as α_1 does. Alternatively, the corresponding simple closed curve a_1 is covered by the fewest simple closed curves under ϕ.

(5.2) Proposition. *Let $\phi : M \to T$ be a simple branched covering of the torus $T \cong S^1 \times S^1$. Then ϕ has a Hurwitz system*

$$(\sigma_1,\ldots,\sigma_n;\alpha,\beta)$$

in which α is maximal and $\sigma_{2i-1} = \sigma_{2i}$ for $1 \leq i \leq n/2$.

Proof. We may suppose that α is maximal. We then show that we can create the required doubles without destroying the maximality of α. Let $\gamma_1\gamma_1\ldots\gamma_r$ be the expression of α as a product of disjoint cycles. Maximality of α implies that for each i, $1 \leq i \leq n$, there is a $j(i)$, $1 \leq j(i) \leq r$, such that

$|\sigma_i| \subset |\gamma_{j(i)}|$. For otherwise, operation (3.7) which replaces α by $\sigma_n\alpha$ can be used, after applying (3.1) an appropriate number of times, to contradict maximality.

We claim that there exist indices $i_1 < i_2$ and j such that

$$|\sigma_{i_1}| \subset |\sigma_{i_2}| \subset |\gamma_j|$$

and $\sigma_{i_1}\sigma_{i_2}\alpha$ is conjugate to α. It is here that we crucially use the fact that T is the torus. We know that

$$\sigma_1\sigma_2\ldots\sigma_n\alpha\beta\alpha^{-1}\beta^{-1} = 1,$$

so that $\sigma_1\sigma_2\ldots\sigma_n\alpha = \beta\alpha\beta^{-1}$. By maximality $\sigma_n\alpha$ has one more component in its cycle structure than α has — we say σ_n *splits* α. Now either $\sigma_{n-1}\sigma_n\alpha$ is conjugate to α, and we have our desired pair, or σ_{n-1} splits $\sigma_n\alpha$. Continuing in this way, either σ_{i-1} splits a component of

$$\sigma_i\sigma_{i+1}\ldots\sigma_n\alpha,$$

or σ_{i-1} combines two components of a previously split component of α. Since $\sigma_1\sigma_2\ldots\sigma_n\alpha$ is completely recombined, it follows that one eventually encounters a σ_{i_1} which recombines two pieces of a component of α previously split apart by some σ_{i_2}, $i_1 < i_2$. Thus $\sigma_{i_1}\sigma_{i_2}\alpha$ is conjugate to α. Applying (3.1) if necessary we may assume $\sigma_{n-1}\sigma_n\alpha$ is conjugate to α. By (5.1) we may arrange so that $\sigma_{n-1} = \sigma_n$ without destroying the maximality of α.

Now simply formally delete σ_{n-1} and σ_n, and repeat the argument until all transpositions appear as doubles. □

(5.3) Theorem. *Let $\phi : M \to T$ be a primitive, simple branched covering of degree d over the torus T. Then ϕ has a Hurwitz system of the form*

$$((1,2)\ ,\ (1,2)\ ,\ldots,\ (1,2)\ ;\ (1,2,\ldots,d)\ ,\ (1)).$$

Proof. By (5.2), ϕ has a Hurwitz system $(\sigma_1, \sigma_2, \ldots, \sigma_n; \alpha, \beta)$ in which α is maximal and $\sigma_{2i-1} = \sigma_{2i}$ for $1 \leq i \leq n/2$. Then

$$[\alpha,\beta] = \alpha\beta\alpha^{-1}\beta^{-1} = (1).$$

Primitivity and maximality then imply that α must be a d-cycle. Since $[\alpha,\beta] = (1)$, β respects the orbits of α, i.e.,

$$\beta(\alpha(x)) = \alpha(\beta(x))\ ;$$

by maximality of α, each σ_i respects the orbits of α; therefore primitivity implies that α has only one orbit, and hence is a d-cycle.

Since β commutes with the d-cycle α, it follows that $\beta = \alpha^k$ for some k. (Proof: Say $\alpha = (1,2,\ldots,d)$ so that

$$\alpha(x) = x + 1 \pmod{d};$$

suppose $\beta(1) = 1 + k$ for some k; then reasoning inductively $\beta(x) = \beta\alpha(x-1) = \alpha\beta(x-1) = \alpha(x-1+k) = x+k$). Therefore, by applying the

Dehn twist operation (3.4), a suitable number of times, we may arrange that $\beta = (1)$.

Thus we have achieved a Hurwitz system $(\sigma_1, \sigma_2, \ldots, \sigma_n; \alpha, (1))$ for ϕ in which $\sigma_{2i-1} = \sigma_{2i}$ for $1 \leq i \leq n/2$ and $\alpha = (1, 2, \ldots, d)$, after appropriate relabeling. By (5.4) below we may arrange in addition that $\sigma_1 = \sigma_2 = \ldots = \sigma_n = (1,m)$ for some m. Primitivity then implies that $(1,m)$ and $(1, 2, \ldots, d)$ then generate $\mathscr{S}_d$. By (5.5) below, $\mathrm{GCD}(m-1,d) = 1$. Therefore, using the Dehn twist operations (3.3) and (3.4) we may replace α by the d-cycle

$$\alpha^{m-1} : (\alpha,(1)) \to (\alpha,\alpha) \to (\alpha^{m-1},\alpha) \to (\alpha^{m-1},(1)),$$

since α^{-1} is a power of α^{m-1}. Then the entries 1 and m are adjacent in α^{m-1}. Appropriate relabeling yields the required form.

It remains to prove two lemmas.

(5.4) LEMMA. *A Hurwitz system*

$$(2\sigma_1, \ldots, 2\sigma_n ; \alpha , \beta)$$

in $\mathscr{S}_d$, in which the transpositions appear as doubles and

$$\alpha = (1, 2, \ldots, d),$$

can be put into the form $((1,m), \ldots, (1,m) ; \alpha , \beta)$ for some m.

Proof. By (3.9) we can make the following two families of alterations of doubles:

$$A_{ij} : 2\sigma_j \to 2(\sigma_j^{\sigma_i}) \quad \text{and} \quad B_j : 2\sigma_j \to 2(\sigma_j^{\alpha}).$$

(All unspecified entries remain unaltered.)

Then by B_j moves we may arrange that $\sigma_j = (1,m_j)$ for some m_j, $1 \leq j \leq k$. Among Hurwitz systems of this form we choose one such that Σm_j is minimal. If some $m_i < m_j$, we can reduce Σm_j as follows: By A_{ij} we replace $2\sigma_j$ with $2(m_i,m_j)$. Application of B_j a suitable number of times replaces $2(m_i,m_j)$ with $2(1, m_j - m_i + 1)$. It follows that when Σm_j is minimal we have the required form. □

(5.5) LEMMA. $(1,m)$ *and* $(1, 2, \ldots, d)$ *generate* $\mathscr{S}_d$ *iff* $\mathrm{GCD}(m-1,d) = 1$.

Proof. It is well known that $(1,2)$ and $(1, 2, \ldots, d)$ generate $\mathscr{S}_d$. Suppose $\mathrm{GCD}(m-1,d) = 1$. Then $(1, 2, \ldots, d)^{m-1}$ is again a d-cycle generating the same cyclic group as $(1, 2, \ldots, d)$. Since 1 and m are adjacent in $(1, 2, \ldots, d)^{m-1}$, it follows that $(1,m)$ and $(1, 2, \ldots, d)^{m-1}$ generate $\mathscr{S}_d$; hence so do $(1,m)$ and $(1, 2, \ldots, d)$.

Now suppose $\mathrm{GCD}(m-1,d) = n$, $n > 1$. then $d = rn$ and $m = 1 + sn$ for some $r, s > 1$. Let

$$X_i = \{j : 1 \leq j \leq d \text{ and } j = i \bmod n\}.$$

Then $X_1, \ldots, X_n$ form a partition of $\{1, 2, \ldots, d\}$ which is preserved by $(1,m)$ and $(1, 2, \ldots, d)$. Therefore these two permutations cannot generate $\mathscr{S}_d$. □

6. Uniqueness in the metastable range

Define a simple branched covering $\phi : M \to N$ of degree d, where N has positive genus, to be *metastable* if $\# B_\phi > d/2$, or, equivalently, if $d\chi(N) - \chi(M) > d/2$. For example, if ϕ is primitive and $d \leq 3$, then ϕ is metastable if N is closed. Recall the notion of maximality introduced above (5.2).

(6.1) PROPOSITION. *Let $\phi : M \to N$ be a metastable, simple branched covering. Then ϕ has a Hurwitz system*

$$(\sigma_1, \ldots, \sigma_n; \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g)$$

in which α_1 is maximal and $\sigma_{n-1} = \sigma_n$.

(6.2) *Remark.* All one needs is that $n = \# B_\phi$ is greater than the number of nontrivial orbits of α_1; for example (6.1) also holds without metastability if α_1 is a d-cycle.

Proof of (6.1). Begin with any Hurwitz system

$$(\sigma_1, \ldots, \sigma_n; \alpha_1, \beta_1, \ldots)$$

in which α_1 is maximal. By maximality, operation (3.7) implies that each transposition σ_i is supported in an orbit of α_1. Metastability implies that at least two of the transpositions are supported in the same orbit of α_1. Thus by applying (3.8) a suitable number of times we may assume that

$$|\sigma_{n-1}| \cap |\sigma_n| \neq \emptyset.$$

Let $\alpha^* = \beta_1\alpha_1^{-1}\beta_1^{-1}$ and $\beta^* = \beta_1^{\{\beta_1\cdot\alpha_1\}}$. Then

$$(\sigma_1, \ldots, \sigma_n; \alpha^*, \beta^*, \alpha_2, \beta_2, \ldots)$$

is also a Hurwitz system for ϕ, and α^* is clearly maximal. Since σ_{n-1} and σ_n overlap it follows that σ_{n-1} and σ_n are supported in the same cycle γ of α^*. Let $\gamma = (1, 2, \ldots, k)$.

By applying operations (3.1), (3.2), and (3.8), *none of which change* α^*, and appropriate relabeling, we may assume that $\sigma_{n-1} = (1,i)$, $\sigma_n = (1,j)$ and $i \leq j$. Therefore (5.1) applied to the Hurwitz system

$$(\sigma_1, \ldots, \sigma_{n-1}, \sigma_n; \alpha^*, \beta^*, \alpha_2, \beta_2, \ldots)$$

produces the required double. Since (5.1) only uses operations (3.2) and (3.8), the conjugacy type, and hence maximality, of α^* is unchanged. □

(6.3) THEOREM. *Let $\phi : M \to N$ be a primitive, metastable, simple branched covering of degree d, where N is a closed orientable surface of positive genus g.*

Then ϕ has a Hurwitz system of the form

$$((1,2)\ ,\ (1,2),\ldots,(1,2)\ ;\ (1,2,\ldots,d)\ ,\ (1),\ldots,(1)\ ,\ (1)).$$

Proof. It suffices to show that ϕ has a Hurwitz system

$$(\sigma_1,\ldots,\sigma_n;\alpha_1,\beta_1,\ldots)$$

in which $\alpha_2 = \beta_2 = \ldots = \alpha_g = \beta_g = (1)$; for then the torus case (5.3) applies to create the desired Hurwitz system. We suppose, therefore, that we have chosen a Hurwitz system for ϕ such that α_1 is maximal and the total number of disjoint cycles (including trivial cycles) in α_2 , $\beta_2,\ldots,\alpha_g$, β_g is as large as possible. Suppose some α_i or β_i , $i \geq 2$, is nontrivial. Using standard operations we may suppose $\alpha_2 \neq (1)$. We shall show how to increase the number of disjoint cycles in α_2, without altering the conjugacy type of any α_i $(i \neq 2)$ or any β_i $(i \geq 2)$.

By (6.1) we may suppose $\sigma_{n-1} = \sigma_n$. Let G denote the subgroup

$$<\sigma_1,\ldots,\sigma_{n-2};\alpha_1,\beta_1,\ldots,\alpha_g\ ,\ \beta_g>$$

of S_d and H denote the subgroup $<\sigma_n^\gamma : \gamma \in G>$ of S_d. Now by primitivity $<\sigma_n,G> = S_d$. On the other hand H is a normal subgroup of $<\sigma_n,G> = S_d$ which contains a transposition. Therefore $H = \mathscr{S}_d$. In particular there is a $\gamma \in G$ such that $|\sigma_n^\gamma|$ is contained in the support of a cycle of α_2. By the Doubles Proposition (3.9) we may arrange so that σ_n itself is contained in the support of a cycle of α_2. Now apply (3.6) to move α_2 , β_2 to the first position in the second half of the Hurwitz system and a conjugate of α_1 , β_1 to the second position. Now apply (3.7) to replace α_2 , β_2 by $\sigma_n\alpha_2$, β_2. Then apply (3.6) to move $\sigma_n\alpha_2$, β_2 back to the second position. Now $\sigma_n\alpha_2$ has one more orbit than α_2 has, and all other entries in the Hurwitz system have at most been conjugated. Thus we may systematically increase the total number of cycles in α_2 , $\beta_2,\ldots,\alpha_g$, β_g until $\alpha_2 = \beta_2 = \ldots = \alpha_g = \beta_g = (1)$ as required. □

(6.5) *Remark.* Extending (6.2), we see that the preceding arguments work equally well for branched coverings which have a Hurwitz system

$$(\sigma_1,\ldots,\sigma_n;\alpha_1,\beta_1,\ldots,\alpha_g\ ,\ \beta_g)$$

in which n is larger than the number of nontrivial orbits in α_1. By analyzing the possibilities for a maximal α_1 when d is small, one can thus verify the Uniqueness Conjecture for $d \leq 5$. We leave the details of these case by case considerations to the sufficiently motivated reader. More generally, to prove the Uniqueness Conjecture in general it would suffice to show that α_1 can be chosen to be a d-cycle. Further along these lines we note that it follows that any primitive, simple branched covering can be stabilized by appropriate connected sum with a suitable branched covering of S^2 or with a cyclic covering of the torus so that the resulting branched covering can be put in standard form.

7. Applications

In this section we sketch a few applications of the results of the preceding sections. We begin by recalling the following results from [3] and [1].

(7.1) THEOREM [3]. *A primitive map $M \to N$ of degree $d \geq 2$ between closed, orientable surfaces is homotopic to a branched covering.* □

(7.2) THEOREM [1]. *A branched covering between surfaces is homotopic to a simple branched covering.* □

Let

$$\mathrm{Epi}^d(\pi_1(M)\ ,\ \pi_1(N))$$

denote the set of epimorphisms $\pi_1(M) \to \pi_1(N)$ which correspond to (primitive) maps $M \to N$ of degree d. As is well known, each element of $\mathrm{Hom}(\pi_1(M), \pi_1(N))$ is induced by a map $M \to N$, and each element of $\mathrm{Aut}\ \pi_1(M)$ is induced by a homeomorphism. Both $\mathrm{Aut}\ \pi_1(M)$ and $\mathrm{Aut}\ \pi_1(N)$ naturally act on $\mathrm{Epi}^d(\pi_1(M)\ ,\ \pi_1(N))$, and the two actions commute.

(7.3) COROLLARY. *Let M and N be closed, orientable surfaces and $d \geq 2$. Then the Uniqueness Conjecture* (1.1) *implies that the double coset space* $\mathrm{Aut}\ \pi_1(M) \setminus \mathrm{Epi}^d(\pi_1(M)\ ,\ \pi_1(N)) / \mathrm{Aut}\,\pi_1(N)$ *consists of at most one element.* □

(7.4) COROLLARY. *Suppose $d\chi(N) - \chi(M) > d/2$, that is,*

$$\text{genus }(M) > d \text{ genus }(N) - d/2 + 1.$$

Then $\mathrm{Aut}\ \pi_1(M) \setminus \mathrm{Epi}^d(\pi_1(M)\ ,\ \pi_1(N)) / \mathrm{Aut}\ \pi_1(N)$ *consists of exactly one element.* □

In another vein we examine the simple loop conjecture (1.2).

(7.5) PROPOSITION. *If* (1.1) *is true, then for any map $f : M \to N$ between closed orientable surfaces such that*

$$f_* : \pi_1(M) \to \pi_1(N)$$

is not injective, there is a homotopically nontrivial simple closed curve $C \subset M$ such that $f|C$ is nullhomotopic.

Proof. By proper choice of orientations, we may assume $\deg(f) \geq 0$. The cases $\deg(f) \leq 1$ were dealt with in [3]; so we may assume that $\deg(f) \geq 2$. By lifting f to an appropriate covering space if necessary we may also assume that f is primitive. Thus we may assume that f is a simple branched covering by (7.1) and (7.2). By (1.1), f has a standard Hurwitz system of the form

$$((1,2)\ ,\ (1,2),\ldots,(1,2)\ ;\ (1,2,\ldots,d)\ ,\ (1)\ ,\ (1),\ldots,(1)).$$

In particular there is an embedded arc $A \subset N$ connecting two branch points with the property that $f^{-1}(A)$ consists of $d-2$ arcs, each mapped homeomorphically to A and a single simple closed curve C. Clearly $f|C$ is nullhomotopic.

We are done if C is homotopically nontrivial. So suppose C is nullhomotopic. Then C bounds a disk D in M. It follows that $D - f^{-1}(A) \cap D$ is a branched covering of $N - A$. Let D^* denote D with each component of $f^{-1}(A) \cap D$ collapsed to a point. Then $D^* \cong S^2$, and f yields a map of positive degree $D^* \to N/A$. Now $N/A \cong N$; and it is a standard consequence of Poincaré duality that the manifold image of a sphere under a map of nonzero degree must be a rational homology sphere. It follows that $N \cong S^2$. Therefore if $C \simeq 0$, then any nontrivial simple closed curve in M will suffice. □

(7.6) COROLLARY. *The simple loop conjecture is true for a primitive map $f : M \to N$ of positive degree d, provided*

$$d\chi(N) - \chi(M) > d/2. \ \square$$

Remark. A reasonably straightforward argument shows that to solve the simple loop conjecture in general for surfaces of positive genus, it suffices to solve it for the special case of simple branched coverings of the once-punctured torus, with exactly two branch points.

As a final application we characterize those primitive surface maps of prime degree which are homotopic to orbit maps for $\mathbf{Z}/p$ actions.

(7.7) THEOREM. *Let $f : M \to N$ be a primitive map of prime degree p between closed orientable surfaces. Then up to homotopy f can be identified with the orbit map for a $\mathbf{Z}/p$ action on M if and only if*

$$p\chi(N) - \chi(M) \equiv 0 \bmod(p-1) \quad \textit{and} \quad p\chi(N) - \chi(M) \geq 2(p-1).$$

Proof. Suppose f is such an orbit map. Then the Riemann-Hurwitz formula shows that $p\chi(N) - \chi(M) = \#(F)(p-1)$, where F is the fixed point set of the action. Primitivity implies that $F \neq \emptyset$. As is well known (and easy to prove for surfaces) such a $\mathbf{Z}/p$ action cannot have just one fixed point. Thus

$$p\chi(N) - \chi(M) \geq 2(p-1).$$

Conversely, suppose $m = p\chi(N) - \chi(M) \equiv 0 \bmod (p-1)$ and $m \geq 2(p-1)$. Now f is homotopic to a simple branched covering $\phi : M \to N$ by (7.1) and (7.2). On the other hand one easily constructs a $\mathbf{Z}/p$-branched covering $\psi' : M' \to N$ with $n = m/(p-1)$ branch (fixed) points. The Riemann-Hurwitz formula shows that $\chi(M') = \chi(M)$, so that $M' \cong M$. Now ψ' is homotopic to a simple branched covering $\psi : M \to N$ with $m \geq 2(p-1)$ branch points. Therefore (6.3) implies that there are homeomorphisms $g : M \to M$ and $h : N \to N$ such that

$$\phi = h^{-1} \circ \psi \circ g.$$

Now $f \simeq \phi$ and $\psi \simeq \psi'$ imply that $f \simeq h^{-1} \circ \psi' \circ g$, the orbit map for an obvious $\mathbf{Z}/p$ action. □

8. Homological uniqueness

In this section we show that any two primitive maps $M \to N$ of nonzero degree d between closed orientable surfaces determine equivalent homomorphisms $H^1(N;\mathbf{Z}) \to H^1(M;\mathbf{Z})$.

A *symplectic inner product space* V over $\mathbf{Z}$ is a finitely generated, free $\mathbf{Z}$-module with a nonsingular skew symmetric (i.e., alternating) bilinear form $\beta : V \times V \to \mathbf{Z}$. We denote $\beta(x,y)$ by $x \cdot y$. Such a V has a *symplectic basis*, that is, a basis

$$a_1, \ldots, a_n, b_1, \ldots, b_n$$

such that $a_i \cdot a_j = b_i \cdot b_j = 0$ and $a_i \cdot b_j = 1$ or 0 as $i = j$ or $i \neq j$. For facts about symplectic inner product spaces, see [7], for example.

Now let V and W be symplectic inner product spaces over $\mathbf{Z}$. Define a homomorphism $\phi : V \to W$ to be *symplectic* of degree d if $\phi(x) \cdot \phi(y) = d\, x \cdot y$ for all $x, y \in V$. Define ϕ to be *primitive* if ϕ is a split monomorphism.

(8.1) THEOREM. *Let V and W be symplectic inner product spaces over $\mathbf{Z}$ and let $\phi : V \to W$ be a primitive symplectic homomorphism of degree $d \geq 2$. Let $a_1, \ldots, a_n, b_1, \ldots, b_n$ be a symplectic basis for V. Then there exists a symplectic basis $e_1, \ldots, e_m, f_1, \ldots, f_m$ for W such that $\phi(a_i) = e_i$ and $\phi(b_i) = df_i + f_{n+i}$ for $1 \leq i \leq n$.* (In particular $\dim W \geq 2 \dim V$.)

(8.2) COROLLARY. *Let $f, g : M \to N$ be primitive maps of degree $d \geq 2$ between closed orientable surfaces. Then there is a homeomorphism $h : M \to M$ such that the following diagram commutes:*

$$\begin{tikzcd} & H^1(N;\mathbf{Z}) \arrow[dl, "f^*"'] \arrow[dr, "g^*"] & \\ H^1(M;\mathbf{Z}) \arrow[rr, "h^*"] & & H^1(M;\mathbf{Z}). \end{tikzcd}$$

Proof. Both $H^1(M;\mathbf{Z})$ and $H^1(N;\mathbf{Z})$ are naturally symplectic inner product spaces over $\mathbf{Z}$, and f^* and g^* are primitive symplectic homomorphisms of degree d. Fix a symplectic basis for $H^1(N;\mathbf{Z})$. By (8.1), there are symplectic bases for $H^1(M;\mathbf{Z})$ with respect to which f^* and g^* have the same matrix. The corresponding change of symplectic basis matrix defines a symplectic automorphism of $H^1(M;\mathbf{Z})$ and is therefore well known to be induced by a homeomorphism $h : M \to M$, as required. □

Proof of (8.1). Let

$$A = \operatorname{span}\{a_1, \ldots, a_n\} \quad \text{and} \quad B = \operatorname{span}\{b_1, \ldots, b_n\}.$$

Then $V = A \oplus B$ and A and B are maximal totally isotropic subspaces of V. Now, in W, $\phi(A)$ and $\phi(B)$ are totally isotropic direct summands. Expand $\phi(A)$ and $\phi(B)$ to maximal totally isotropic subspaces E and F of W so that $W = E \oplus F$, and the nonsingular form defines an isomorphism $F \to \mathrm{Hom}(E; \mathbf{Z})$. In particular, any basis for E can be uniquely expanded to a symplectic basis of W by appending elements of F, and vice-versa.

By primitivity, $\phi(a_1), \ldots, \phi(a_n)$ form part of a basis of E. Let $f_1, \ldots, f_n$ be the corresponding dual elements of F. Now $\phi(b_1), \ldots, \phi(b_n)$ is a basis for the summand ϕB in F. Observe that $\phi(B) \cap \mathrm{span}\{f_1, \ldots, f_n\} = \{0\}$: If not there is an indivisible $f \in \phi(B) \cap \mathrm{span}\{f_1, \ldots, f_n\}$; since

$$f \in \mathrm{span}\{f_1, \ldots, f_n\}$$

there is $e \in \phi(A)$ such that $e \cdot f = 1$; but $e \in \phi(A)$ and $f \in \phi(B)$ imply that $e \cdot f \equiv 0 \mod d$.

Define $f_{n+i} = \phi(b_i) - df_i$, for $1 \leq i \leq n$. It follows that $f_1, \ldots, f_{2n}$ form a basis for a summand of F. Extend $f_1, \ldots, f_{2n}$, if necessary, to a basis $f_1, \ldots, f_m$ of F, by adding elements orthogonal to $\phi(A)$. Let $e_1, \ldots, e_m$ be the corresponding basis of E. It easily follows that $e_i = \phi(a_i)$ and that $\phi(b_i) = df_i + f_{n+i}$ for $1 \leq i \leq n$, as required. □

Remark (1) Of course (8.2) (or at least the weaker version allowing an automorphism of $H^1(N; \mathbf{Z})$) would be an immediate consequence of (1.1), when coupled with (7.1) and (7.2).

(2) Not all primitive symplectic homomorphisms $\phi : V \to W$ as in (8.1) correspond to surface maps. It follows easily from Kneser's theorem [5] and standard facts that ϕ is induced by a surface map if and only if $\dim W > d \dim V + 1 - d$. Thus (1.1) would not imply (8.1) entirely.

References

1. I. Berstein and A. Edmonds, *On the construction of branched coverings of low-dimensional manifolds,* Trans. Amer. Math. Soc., vol. 247(1979), pp. 87-124.
2. A. Clebsch, *Zur Theorie der Riemann'schen Flächen,* Math. Ann., vol. 6(1873), pp. 1-15.
3. A. Edmonds, *Deformation of maps to branched coverings in dimension two,* Ann. of Math., vol. 110(1979), pp. 113-125.
4. A. Hurwitz, *Uber Riemann'sche Flächen mit gegebenen Verzweigungspunkten,* Math. Ann., vol. 39(1891), pp. 1-60.
5. H. Kneser, *Die kleinste Bedeckungszahl innerhalb einer Klasse von Flächen abbildungen,* Math. Ann., vol. 103(1930), pp. 347-358.
6. J. Lüroth, *Note über Verzweigungsschnitte und Querschnitte in einer Reimann'schen Fläche,* Math. Ann., vol. 4(1871), pp. 181-184.
7. J. Milnor and D. Husemoller, *Symmetric Bilinear Forms,* Springer-Verlag, New York, 1973.
8. H. Wielandt, *Finite Permutation Groups,* Academic Press, New York, 1964.

Cornell University
Ithaca, New York

Indiana University
Bloomington, Indiana