THE REPRESENTATION THEORY, GEOMETRY, AND COMBINATORICS OF BRANCHED COVERS

BRIAN OSSERMAN

Abstract. The study of branched covers of the Riemann sphere has connections to many fields. We recall the classical relationship between branched covers and group theory via the Riemann existence theorem, which then leads to represention-theoretic formulas for Hurwitz numbers, counting the number of branched covers of prescribed types. We also review the Hurwitz spaces parametrizing branched covers as we allow the branch points to move, and the relationship between the components of a Hurwitz space and the orbits of a certain braid group action. Finally, we present two new results in this field: a connectedness theorem for Hurwitz spaces in the classical setting, joint with Fu Liu, and a result analogous to the Riemann existence theorem describing certain tamely branched cover of the projective line in positive characteristic.

1. Branched covers of the Riemann sphere

Let $\mathbb{CP}^1 = \mathbb{C} \cup \{\infty\} \cong S^2$ be the Riemann sphere, or equivalently, the complex projective line. Our main topic of discussion will be **branched covers** of $\mathbb{CP}^1$; these are pairs (C, f) with C a connected, compact Riemann surface, and $f : C \to \mathbb{CP}^1$ a surjective holomorphic map. Of course, this may be stated equivalently in terms of smooth projective complex curves. The key properties of such a branched cover are:

- for every point $P \in C$, there is some $e_P \geq 1$ such that in a neighborhood of P, the map f looks like $z \mapsto z^{e_P}$ (with P corresponding to the origin);
- there is some $d \geq 1$ such that for any $Q \in \mathbb{CP}^1$, we have $\sum_{P \in f^{-1}(Q)} e_P = d$.

Definition 1.1. We say that the above d is the **degree** of f. We say that $P \in C$ is a **ramification point** of f (of **index** e_P) if $e_P > 1$. We then say that $f(P)$ is a branch point of f. If $Q \in \mathbb{CP}^1$ is a branch point, we say that the **branch type** of f at Q is the non-trivial partition of d obtained from the multiset of e_P for $P \in f^{-1}(P)$.

We observe the following consequences:

- since C is compact, there are only finitely many ramification points of f, and hence only finitely many branch points;
- if $Q \in \mathbb{CP}^1$ is not a branch point, $\# f^{-1}(Q) = d$.

Up to isomorphism, there are only finitely many covers of given type with given branch points (we will see later why this is true). This leads to a basic question in the field:

Question 1.2. If we fix $d \geq 1$, branch points $Q_1, \dots, Q_n \in \mathbb{CP}^1$, and branch types $T_1, \dots, T_n$, how many branched covers of $\mathbb{CP}^1$ of degree d, branched only over the Q_i, with branch type T_i over each Q_i?

Definition 1.3. The above number is the **Hurwitz number** for $(T_1, \dots, T_n)$.

Hurwitz used the idea of monodromy to translate this into a problem in combinatorial group theory. Before discussing this, we make one more definition of fundamental importance:

Definition 1.4. Given $d, T_1, \dots, T_n$, let $\mathcal{H}(T_1, \dots, T_n)$ be the **Hurwitz space** parametrizing branched covers (C, f) of $\mathbb{CP}^1$ of degree d together with distinct $Q_1, \dots, Q_n \in \mathbb{CP}^1$ such that f is branched only over the Q_i, with branch type T_i at each Q_i.

If $U_n \subseteq (\mathbb{CP}^1)^n$ denotes the open subset of n-tuples of distinct points, then $\mathcal{H}(T_1, \dots, T_n)$ is a cover of U_n, of degree equal to the Hurwitz number. We can now also ask:

Question 1.5. How many (connected) components does $\mathcal{H}(T_1, \dots, T_n)$ have?

There are certainly many cases where $\mathcal{H}(T_1, \dots, T_n)$ is disconnected, but situations in which it is connected, or at least in which one can develop a good understanding of the components, have been important in many applications, including the original proof of the connectedness of the moduli space of curves in algebraic geometry, and much progress on the inverse Galois problem.

Remark 1.6. It is more common to take a "stacky" point of view, and count covers weighted fractionally by the number of automorphisms. In a similar vein, the Hurwitz spaces should formally be defined as stacks. However, in the situation of our main results, the covers will always be automorphism-free, so neither of these points arises.

2. Monodromy of branched covers and Riemann's existence theorem

Let us first suppose we are given a branched cover (C, f) of $\mathbb{CP}^1$, and label the branch points $(Q_1, \dots, Q_n)$. Write $D := \mathbb{CP}^1 \smallsetminus \{Q_1, \dots, Q_n\}$. We also fix $Q \in \mathbb{CP}^1$.

Suppose we have a loop $\gamma \in \pi_1(D, Q)$. Since Q is not a branch point of f, we have $\#f^{-1}(Q) = d$. Let P be a point of $f^{-1}(Q)$. Since γ doesn't pass through any branch points of f, it can be lifted uniquely to a path starting at P; this path will end in $f^{-1}(Q)$, but not necessarily at P. Thus, for each $P \in f^{-1}(Q)$, the loop γ gives us a $\gamma(P) \in f^{-1}(Q)$, so we obtain a permutation in $\mathrm{Sym}(f^{-1}(Q))$, called the **monodromy** of f around γ. It is not difficult to see that this in fact defines a homomorphism

$$\mu : \pi^1(D, Q) \to \mathrm{Sym}(f^{-1}(Q)),$$

which we call the **monodromy map**. If we choose a labelling of $f^{-1}(Q)$, we obtain an isomorphism $\mathrm{Sym}(f^{-1}(Q)) \cong S_d$, and can therefore think of the monodromy map as taking values in S_d.

Next, it is a classical fact that we choose $(\gamma_1, \dots, \gamma_n)$ by taking small loops (say, clockwise) around the Q_i, and connecting each to Q appropriately, that we will have:

- the γ_i generate $\pi_1(D, Q)$;
- $\gamma_1 \cdots \gamma_n = 1$;
- each γ_i is homotopic in D to a small loop around Q_i.

We call such a tuple $(\gamma_1, \ldots, \gamma_n)$ a **local generating system**, and we suppose we fix such a system.

Returning to our cover (C, f), if we assume we have labeled $f^{-1}(Q)$, we may use the monodromy map to obtain $\sigma_i := \mu(\gamma_i) \in S_d$. We let T_i be the branch type at Q_i for each i. We claim that the tuple $(\sigma_1, \ldots, \sigma_n)$ has the following properties:

- the σ_i generate a transitive subgroup of S_d;
- $\sigma_1 \cdots \sigma_n = 1$;
- for each i, the partition of d obtained from the disjoint cycle representation of σ_i is T_i.

The last assertion boils down to seeing that a ramification point P of index e over Q_i contributes an e-cycle to the monodromy around a small loop around Q_i. This follows from the fact that f looks like $z \mapsto z^e$ near P. Given this, it is not too difficult to see that the first assertion is in fact equivalent to the connectedness hypothesis on C. The second assertion follows from the fact that μ is a homomorphism.

Given any partitions T_i of d, we call a tuple $(\sigma_1, \ldots, \sigma_n)$ satisfying the above properties a **Hurwitz factorization** for $(T_1, \ldots, T_n)$. If we don't wish to specify the T_i, we simply say that $(\sigma_1, \ldots, \sigma_n)$ is a Hurwitz factorization if it satisfies the first two properties. We say two Hurwitz factorizations are **equivalent** if one can be obtained from the other by relabelling.

Since we have to choose a labeling of $f^{-1}(Q)$ to get the σ_i, each cover only naturally gives us an equivalence class of Hurwitz factorizations. A special case of Riemann's existence theorem asserts:

Theorem 2.1. *Fix $(Q_1, \ldots, Q_n)$, Q, d, and a local generating system $(\gamma_1, \ldots, \gamma_n)$. Given a Hurwitz factorization $(\sigma_1, \ldots, \sigma_n)$, there exists a unique cover (C, f) with monodromy σ_i around each γ_i.*

We immediately conclude:

Corollary 2.2. *The Hurwitz number for $(T_1, \ldots, T_n)$ is the number of Hurwitz factorizations for $(T_1, \ldots, T_n)$, up to equivalence.*

This then expresses Question 1.2 in terms of combinatorial group theory, as promised. However, such numbers are not at all trivial to compute, let alone to produce explicit formulas for, so the question remains a topic of active research.

We promised some mention of the relationship to representation theory, so we next discuss this briefly. Because our branch types T_i are simply conjugacy classes in S_d, Hurwitz numbers are visibly related to structure constants for S_d, which count numbers of tuples in given conjugacy classes whose product is 1. As is well-known in representation theory, structure constants can be computed rather easily from the character table.

However, this doesn't immediately allow translation between Hurwitz numbers and structure constants. The structure constants aren't counting up to relabeling, but this typically only involves dividing by a factor of $d!$. The more substantive issue is that they don't say anything about whether the tuple generates a transitive subgroup, so any formula going from Hurwitz numbers to structure constants or vice versa typically involves a messy recursive sum over all subgroups of S_d. That said, in doing explicit computations involving Hurwitz factorizations, using the character table to compute structure constants is frequently an important first step.

3. The braid group and connected components

We continue with our analysis of how to understand Hurwitz spaces in terms of group theory by bringing the braid group into the picture. The **braid group** B_n is a certain quotient of the free group F_{n-1}; its precise structure is not important, but the key point is that given a group G, we have that B_n acts on n-tuples of elements of G with trivial product. The ith generator φ_i replaces $(\sigma_1, \dots, \sigma_n)$ with $(\sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \sigma_{i+1}^{-1}\sigma_i\sigma_{i+1}, \sigma_{i+2}, \dots, \sigma_n)$; note that the product remains unchanged. We also have a natural map $B_n \to S_n$ which sends φ_i to the transposition $(i, i+1)$. The **pure braid group** PB_n is then the kernel of this map. We note that an element of PB_n not only preserves the trivial-product condition, but sends each σ_i to a conjugate of itself.

It therefore follows that PB_n acts both on local generating systems, and on Hurwitz factorizations.

A very classical fact is the following:

Proposition 3.1. *Any two choices of local generating system $(\gamma_1, \dots, \gamma_n)$ for $\pi_1(D, Q)$ are related by the action of PB_n.*

Thus, we can study how the monodromy of a given cover depends on our choice of local generating system by looking at the pure braid group action on the resulting Hurwitz factorization.

Moreover, we can realize the pure braid group action on local generating systems by moving the $Q_1, \dots, Q_n$ around one another and bringing them back to the same place. Based on this observation, it is not too hard to see that:

Corollary 3.2. *The connected components of $\mathcal{H}(T_1, \dots, T_n)$ are in one-to-one correspondence with the pure braid orbits of equivalence classes of Hurwitz factorizations for $(T_1, \dots, T_n)$.*

Thus, we also have a purely group-theoretic interpretation of Question 1.5.

4. New results

In general, both of Questions 1.2 and 1.5 are quite difficult to answer directly. We describe some new results on these questions, as well as on the following question:

Question 4.1. What is the correct version of Riemann's existence theorem for tame covers in positive characteristic?

It turns out that for tame covers, one can still develop the theories of the fundamental group and monodromy around branch points, but Riemann's existence theorem fails, and very little is known about when a cover does or doesn't exist with the given monodromy.

All our results will restrict to the following situation:

Situation 4.2. We consider covers such that each branch type consists of a single cycle, of length e_i, and with $2d - 2 = \sum_i (e_i - 1)$, or equivalently, the cover C is also isomorphic to $\mathbb{CP}^1$.

We say that the first condition is the **pure-cycle** case, so that we are studying genus 0, pure-cycle Hurwitz spaces. Our first results are joint with Fu Liu:

Theorem 4.3. *(Liu-Osserman) In situation 4.2, the Hurwitz space $\mathcal{H}(T_1, \dots, T_n)$ is always connected.*

Theorem 4.4. *(Liu-Osserman) If further $n = 4$, we have an explicit description of the Hurwitz factorizations, and in particular we show that the Hurwitz number is* $\min\{e_i(d+1-e_i)\}$.

In the positive characteristic case, we assume that all e_i are prime to p. In the situation that either $n = 3$, or we have $e_i < p$ for each i, we define a purely group-theoretic notion of **p-admissibility** for a Hurwitz factorization $(\sigma_1, \dots, \sigma_n)$ (in fact, it turns out that the condition only depends on the e_i, and not the given tuple). We then show:

Theorem 4.5. *Let $Q_1, \dots, Q_n$ be general points on $\mathbb{P}^1_k$, for k an algebraically closed field of characteristic p. Choose a base point $Q \neq Q_i$, and let $(\gamma_1, \dots, \gamma_n)$ be a local generating system. Fix also d and $e_1, \dots, e_n$ with $2d - 2 = \sum_i (e_i - 1)$ and either $n = 3$ and e_i prime to p, or $e_i < p$ for all i. Suppose we are given a Hurwitz factorization $(\sigma_1, \dots, \sigma_n)$ of $(e_1, \dots, e_n)$.*

Then $(\sigma_1, \dots, \sigma_n)$ is p-admissible if and only if there exists a cover (C, f) of $\mathbb{P}^1_k$, and a pure-braid transformation $(\gamma'_1, \dots, \gamma'_n)$ of $(\gamma_1, \dots, \gamma_n)$, such that the monodromy of f around each γ'_i is σ_i.

We remark that the necessity of taking the pure braid transformation of the local generating system is not an artifact of the proof, but can be shown to be necessary in examples. In particular, unlike the classical situation, which Hurwitz factorizations occur as monodromy of a cover around a local generating system actually depends on the choice of local generating system.

Our initial motivation for proving Theorem 4.3 was to apply it to Theorem 4.5, both strengthening the statement and greatly simplifying the proof. However, our result, and particularly the sharper work in the case $n = 4$, has turned out to have applications of a more arithmetic nature. Using our description as a starting point, Fried has been able to check his main conjecture for modular towers in an infinite family of new examples. This conjecture states that certain towers of Hurwitz spaces should not have rational points sufficiently high up in the tower, which roughly says that in the inverse Galois problem, as groups get more complicated, one has to allow ramification at more and more points to realize the groups as Galois groups. There is also some potential for positive progress on the inverse Galois problem using these spaces, as in some cases we are producing Hurwitz spaces with many rational points, and these should at a minimum provide simpler realizations of certain Galois groups than had previously been possible, and might even produce realizations of groups not already known.

5. Sketch of proof: classical results

We begin by sketching the proofs of Theorems 4.3 and Theorem 4.4. We proceed in two steps: we first give a sufficiently explicit description of the possible Hurwitz factorizations in the case $n = 4$ to prove everything directly in terms of group theory in that case, and we then use a geometric argument to reduce the general statement of Theorem 4.3 to the case $n = 4$.

It turns out that the $n = 4$ case lies at an important boundary: the $n = 3$ case is quite simple, but does not seem able to serve as a base case for induction, which for $n > 4$ the situation seems too complicated to work out explicitly in terms of group theory. The $n = 4$ case, while quite complicated, turns out to be tractable, and therefore gives us the necessary foothold for the general result.

We do not describe the situation too explicitly here, but we do sketch the basic idea. We have four cycles $\sigma_1, \sigma_2, \sigma_3, \sigma_4$, with trivial product, lengths e_1, e_2, e_3, e_4, and generating a transitive subgroup of S_d. Since $2d - 2 = \sum_i (e_i - 1)$, we have $e_1 + e_2 + e_3 + e_4 = 2d + 2$. By the transitivity, every number in $\{1, \dots, d\}$ must appear in some σ_i, and by the trivial product condition, every number must appear at least twice. That leaves two numbers left for additional repetition: we conclude that either two numbers appear in three of the σ_i, or one number appears in all four. This then allows us to show that $\sigma_1 \sigma_2$ is always a product of at most two disjoint cycles, and we are able to describe all the possibilities via direct and elementary, albeit rather intricate, analysis.

The reduction step has a very different flavor. If one attempts to construct an induction argument directly, one might try to analyze tuples by looking at $(\sigma_1, \dots, \sigma_{n-2}, \sigma_{n-1} \cdot \sigma_n)$ and $((\sigma_{n-1} \cdot \sigma_n)^{-1}, \sigma_{n-1}, \sigma_n)$ (this is equivalent to a geometric degeneration from the point of view of branched covers). However, this approach runs into the fatal flaw that we do not stay within our hypotheses: even if the σ_i are all cycles, the product $\sigma_{n-1} \cdot \sigma_n$ need not be a cycle. We get around this problem by not only working in the geometric setting, but by switching perspective again, from branched covers to **linear series**; in this case, we simply mean that rather than fixing branch points on the base and working up to automorphism of the cover, we fix ramification points on the source, and work up to automorphism of the image. From this point of view, if we carry out degeneration arguments (using the Eisenbud-Harris theory of limit linear series), we obtain a situation which is in fact inductive. The intuition is that we degenerate by bringing together ramification points on the source, so the curve is breaking at a single point in the source, and this is the only place with new ramification. In contrast, if we degenerate the base, then there is only one node in the limit on the base, but there are d points over the node, and it is difficult to control what happens to all of them. Thus, using limit linear series, we are able to show that it is enough to understand the case $n = 4$, and we conclude Theorem 4.3.

6. Sketch of proof: positive characteristic

As in the geometric portion of the argument for Theorem 4.3, the proof of Theorem 4.5 relies on the idea of exploiting the interplay between the branched cover point of view and the linear series point of view, and using degeneration arguments involving limit linear series. The proof has four key ingredients:

- In the $n = 3$ case, we show that one can understand existence of separable maps with given ramification in terms of non-existence of inseparable maps, interpreted suitably. This leads to a combinatorial criterion.
- In order to degenerate, it is necessary to show that separable maps will not become inseparable under degeneration. This is in general an extremely difficult problem; in this case, we are able to produce an argument which is ultimately self-contained, although discovered originally via a relationship to Mochizuki's "dormant torally indigenous bundles", which are a certain class of $\mathbb{P}^1$-bundles with logarithmic connection on curves.
- In order to compare the linear series point of view to the branched cover point of view, one needs to know that, roughly speaking, it is the same to consider general ramification points or general branch points. In order to show this, we use an elementary statement on the finiteness of rational

functions with given ramification points. However, despite the elementary nature of the statement, the proof of this finiteness result appears to require using the full machinery of the relationship to dormant torally indigenous bundles, and the application of a finiteness theorem of Mochizuki.

- Finally, to go from statements in terms of existence of branched covers with given ramification indices to statements in terms of the monodromy around branch points, we use the group-theoretic form of Theorem 4.3, which asserts that any two Hurwitz factorizations lie in the same pure braid orbit. This allows us to get a good enough handle on the situation to phrase Theorem 4.5 in the Riemann-existence-like form in which it appears.